Article

Modeling I(2) Processes Using Vector Autoregressions Where the Lag Length Increases with the Sample Size

Yuanyuan Li [ID] and Dietmar Bauer *[ID]

Faculty of Business Administration and Economics, Bielefeld University, Universitätsstrasse 25,
D-33615 Bielefeld, Germany; yuan_yuan.li@uni-bielefeld.de
* Correspondence: Dietmar.Bauer@uni-bielefeld.de

Abstract: In this paper the theory on the estimation of vector autoregressive (VAR) models for I(2) processes is extended to the case of long VAR approximation of more general processes. Hereby the order of the autoregression is allowed to tend to infinity at a certain rate depending on the sample size. We deal with unrestricted OLS estimators (in the model formulated in levels as well as in vector error correction form) as well as with two stage estimation (2SI2) in the vector error correction model (VECM) formulation. Our main results are analogous to the I(1) case: We show that the long VAR approximation leads to consistent estimates of the long and short run dynamics. Furthermore, tests on the autoregressive coefficients follow standard asymptotics. The pseudo likelihood ratio tests on the cointegrating ranks (using the Gaussian likelihood) used in the 2SI2 algorithm show under the null hypothesis the same distributions as in the case of data generating processes following finite order VARs. The same holds true for the asymptotic distribution of the long run dynamics both in the unrestricted VECM estimation and the reduced rank regression in the 2SI2 algorithm. Building on these results we show that if the data is generated by an invertible VARMA process, the VAR approximation can be used in order to derive a consistent initial estimator for subsequent pseudo likelihood optimization in the VARMA model.

Keywords: vector autoregressions; vector error correction model; integrated processes of order two

---

1. Introduction

Many macroeconomic variables have been found to exhibit trend-like behaviour that can be modelled by using vector autoregressions (VARs). Katarina Juselius (2006) states that empirical modelling led to the development of I(1) and I(2) models since certain features of the datasets considered required including first and second differences in order to obtain stationary time series. Additionally cointegrating relations were found in the corresponding analyses. Similar findings have reoccurred numerous times in the literature for example related to money demand Johansen (1992b); Juselius (1994), inflation Banerjee et al. (2001); Georgoutsos and Kouretas (2004), interest rates and real exchange rates Johansen et al. (2007); Juselius and Assenmacher (2017); Juselius and Stillwagon (2018); Stillwagon (2018) to mention only a few sources.

The predominant methodological approach to model integration and cointegration in the I(1) and the I(2) case in the vector autoregressive (VAR) framework has been established mainly by Søren Johansen and Katarina Juselius together with a number of coauthors (see the lists of references in Johansen (1995); Juselius (2006) for details) building on vector error correction models (see Engle and Granger (1987) for early comments on the history of using error correction models for co-integrated processes). Extending the main ideas for cointegration modeling for the I(1) setting Johansen (1997) see, e.g., Johansen (1992a) suggested a representation for the I(2) case. Johansen (1997) established asymptotic distributions for the suggested two step I(2) estimator (2SI2) as an approximation to pseudo maximum likelihood

estimation involving numerical optimization. Asymptotics for the corresponding likelihood ratio tests has been developed in Paruolo (1994, 1996), its asymptotic equivalence to pseudo likelihood (using the Gaussian distribution) optimization (and hence in a certain sense statistical efficiency) is shown in Paruolo (2000). However, Nielsen and Rahbek (2007) shows that in finite samples the likelihood ratio test has size advantages. The testing of restrictions on the parameters has been investigated by Boswijk and Doornik (2004); Boswijk and Paruolo (2017); Johansen and Lütkepohl (2005). Due to the implicit vector error correction (VECM) modeling, deterministic terms in the VECM produce complex deterministic terms in the solutions processes. In the I(2) context Nielsen and Rahbek (2007); Paruolo (1994, 2006); Rahbek et al. (1999); Kurita et al. (2011) discuss the impacts of deterministic terms.

As the VECM representation includes the representation of reduced rank matrices by a product of two matrices, identification conditions are of particular importance, see Juselius (2006); Mosconi and Paruolo (2013, 2017). In this context also weak exogeneity has been studied Kurita (2012); Paruolo and Rahbek (1999).

The main idea underlying the VECM approach for estimating VAR models in the I(2) context is to reparameterize the problem such that integration and cointegration properties relate to the rank of two matrices. Assuming the data generating process to be a VAR of known finite order, the rank of matrices can be tested using (pseudo) likelihood ratio tests.

Sometimes the assumption of known order is not justified. For example it is known that a subset of variables that are generated using a finite order VAR cannot be described by a finite order VAR, but instead requires a vector autoregressive moving average (VARMA) model. However, the class of VARs provides flexibility in the sense that a VAR of infinite order can represent a large set of linear dynamical systems including all invertible VARMA systems. For stationary processes Berk (1974) and Lewis and Reinsel (1985) show that by letting the order of the VAR tend to infinity at a suitable function of the sample size, consistent estimation of the underlying transfer function can be achieved for data generating processes that can be described by a VAR(∞) subject to mild assumptions on the summability of the VAR coefficients. Additionally Lewis and Reinsel (1985) also establishes asymptotic normality (in a very specific sense) of linear combinations of the estimated autoregressive coefficients. Hannan and Deistler (1988) make the concepts operational by showing that in the case of a VARMA process generating the dataset the required rate of letting the order tend to infinity can be estimated using `BIC` model selection.

In the case of I(1) processes the estimation theory for long VAR approximations to VARMA processes has been extended based on the techniques in the stationary case of Lewis and Reinsel in a series of papers by Saikkonen and coauthors Saikkonen (1991, 1992); Lütkepohl and Saikkonen (1997); Saikkonen and Lütkepohl (1996); Saikkonen and Luukkonen (1997). Additionally also the Johansen framework of rank restricted estimation in the VECM model has been extended to the long VAR approximations by Saikkonen and Luukkonen (1997). Bauer and Wagner (2004) provide extensions to the multi frequency I(1) case where unit roots may occur at the seasonal frequencies.

For the I(2) case no such extensions are currently known. This is the research gap this paper tries to fill: First we establish consistency and asymptotic normality of estimated autoregressive coefficients (in the sense of Lewis and Reinsel) for unrestricted ordinary least squares (OLS) estimation in the VECM representation. This can be used in order to derive Wald type tests of linear restrictions on the autoregressive parameters. Secondly, we extend the rank restricted regression techniques in the I(2) case to the long VAR approximations showing that the asymptotics (for estimated cointegrating relations, likelihood ratio tests and the two step estimation procedures) are identical in the case of long VAR approximations and VARs of finite known order. Third, we show that if the data generating process is an invertible VARMA process the long VAR system estimator can be used in order to obtain consistent initial estimators for subsequent pseudo likelihood maximization in the VARMA model class. In all results we limit ourselves to the case of no deterministic terms being included in the VECM representation. The inclusion of deterministic terms requires changing the test distribution, compare the theory contained for example in Rahbek et al. (1999).

The paper is organized as follows: In the next section the data generating process and the main assumptions are described. Section 3 then provides the results for the unrestricted estimation. Section 4 deals with rank restricted regression in the 2SI2 procedure, while Section 5 investigates the initial guess in the VARMA setting for subsequent pseudo likelihood maximization. Finally Section 6 concludes the paper. Proofs are relegated to an appendix.

Throughout the paper we will use the notation introduced by Johansen (1997): For a matrix $C \in \mathbb{R}^{p \times s}, s < p$, of full column rank we use the notation $\bar{C} = C(C'C)^{-1}$. Furthermore, $C_{\perp}$ denotes a full column rank matrix of dimension $p \times (p - s)$ such that $C'_{\perp} C = 0$. Whenever this notation is used the particular choice of $C_{\perp}$ is not of importance. For a matrix $C = (C_{i,j}) \in \mathbb{R}^{p \times s}$ we let $\|C\|$ denote the Frobenius norm $\|C\| = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{s} C_{i,j}^2}$.

2. Data Generating Process and Assumptions

In this paper we use the following assumptions on the data generating process:

Assumption 1 (DGP). *The process $(y_t)_{t \in \mathbb{Z}}, y_t \in \mathbb{R}^p$, is generated from the difference equation for $t \in \mathbb{Z}$:*

$$\Delta^2 y_t = \alpha \beta' y_{t-1} + \Gamma \Delta y_{t-1} + \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} + \varepsilon_t \tag{1}$$

where $\alpha, \beta \in \mathbb{R}^{p \times r}, 0 \leq r < p$ are full column rank matrices, $\Delta = (1 - L)$ with L denoting the backward shift operator such that $L(y_t)_{t \in \mathbb{Z}} = (y_{t-1})_{t \in \mathbb{Z}}$. The matrix function $A(z) = (1 - z)^2 I_p - \alpha \beta' z - \Gamma z(1 - z) - \sum_{j=1}^{\infty} \Pi_j (1 - z)^2 z^j$ fulfills the special marginal stability condition that

$$|A(z)| = 0 \quad \text{implies that} \quad |z| > 1 \quad \text{or} \quad z = 1. \tag{2}$$

Furthermore, there exists a real $\delta > 0$ such that the power series defining $A(z)$ converges absolutely for $|z| < 1 + \delta$. Define $\beta_2 = \beta_{\perp} \eta_{\perp}, \alpha_2 = \alpha_{\perp} \zeta_{\perp}$ where $\alpha'_{\perp} \Gamma \beta_{\perp} = \zeta \eta', \eta, \zeta \in \mathbb{R}^{(p-r) \times s}$ are of full column rank $s < p - r$. Then it is assumed that the matrix

$$\alpha'_2 (I_p + \Gamma \bar{\beta} \bar{\alpha}' \Gamma - \sum_{j=1}^{\infty} \Pi_j) \beta_2 \tag{3}$$

is nonsingular.

Furthermore, the process $(\varepsilon_t)_{t \in \mathbb{Z}}$ denotes independent identically distributed (iid) white noise with mean zero and variance $\Sigma_\epsilon > 0$.

It is well known that the conditions (2) and (3) are necessary and sufficient for the existence of solutions to the difference equation that are I(2) processes, see for example Johansen (1992a). Moreover, note that the assumption of absolute convergence of $A(z)$ for $|z| < 1 + \delta$ implies that $\sum_{j=0}^{\infty} j^k \|\Pi_j\| < \infty$ for every $k \in \mathbb{N}$. In particular $\sum_{j=0}^{\infty} j^2 \|\Pi_j\| < \infty$ follows as will be used frequently below.

Every vector autoregressive function $A(z)$ corresponding to the autoregression $A(L)y_t = \varepsilon_t$, that fulfills Assumption 1, allows a representation as $A(z) = (1 - z)^2 I_p - \alpha \beta' z - \Gamma z(1 - z) - \sum_{j=1}^{\infty} \Pi_j (1 - z)^2 z^j = \tilde{g}(z) \tilde{B}(z), \tilde{B}(z) = (1 - z)^2 I_p - \tilde{\Pi}z - \tilde{\Gamma}z(1 - z), \tilde{g}(z) = I_p + \sum_{j=1}^{\infty} G_j z^j$. This can be seen as follows:

$$\begin{aligned}
\varepsilon_t = A(L)y_t &= (A(1) - \dot{A}(1)\Delta + A^*(L)\Delta^2)y_t = (A(1) - \dot{A}(1)\Delta + A^*(L)\Delta^2)\mathcal{B}\mathcal{B}'y_t \\
&= \left([-\alpha, 0, 0] + [\alpha, 0, 0]\Delta - \Gamma\mathcal{B}\Delta + A^*(L)\mathcal{B}\Delta^2\right)\mathcal{B}'y_t \\
&= \left([-\alpha, -\Gamma\beta_1, -\Gamma\beta_2] + [\alpha - \Gamma\beta, A_1^*(L), A_2^*(L)]\Delta + \left[A_0^*(L), 0, 0\right]\Delta^2\right)\begin{pmatrix} \beta' \\ \beta_1'\Delta \\ \beta_2'\Delta \end{pmatrix}y_t \\
&= \left([-\alpha, -\Gamma\beta_1, -\Gamma\beta_2 + \alpha\bar{\alpha}'\Gamma\beta_2] + [\alpha - \Gamma\beta, A_1^*(L), \tilde{A}_2^*(L)]\Delta + \left[A_0^*(L), 0, -A_0^*(L)\bar{\alpha}'\Gamma\beta_2\right]\Delta^2\right)\begin{pmatrix} \beta' + \bar{\alpha}'\Gamma\beta_2\beta_2'\Delta \\ \beta_1'\Delta \\ \beta_2'\Delta \end{pmatrix}y_t \\
&= \left([-\alpha, -\Gamma\beta_1, \tilde{A}_2^*(L)] + [\alpha - \Gamma\beta, A_1^*(L), -A_0^*(L)\bar{\alpha}'\Gamma\beta_2]\Delta + \left[A_0^*(L), 0, 0\right]\Delta^2\right)\begin{pmatrix} \beta' + \bar{\alpha}'\Gamma\beta_2\beta_2'\Delta \\ \beta_1'\Delta \\ \beta_2'\Delta^2 \end{pmatrix}y_t \\
&= g(L)B(L)y_t
\end{aligned}$$

where $\mathcal{B} = [\beta, \beta_1, \beta_2], \beta_1 = \beta_\perp \eta$, is without restriction of generality assumed to be an orthonormal matrix, $A^*(L)\mathcal{B} = [A_0^*(L), A_1^*(L), A_2^*(L)], A(1) = -\alpha\beta', \dot{A}(1) = -\alpha\beta' + \Gamma$ and where we use that

$$\Gamma\beta_2 - \alpha\bar{\alpha}'\Gamma\beta_2 = (I_p - \alpha\bar{\alpha}')\Gamma\beta_2 = \bar{\alpha}_\perp\alpha_\perp'\Gamma\beta_\perp\eta_\perp = 0.$$

Here

$$B(L) = \begin{pmatrix} \beta' + \bar{\alpha}'\Gamma\beta_2\beta_2'\Delta \\ \beta_1'\Delta \\ \beta_2'\Delta^2 \end{pmatrix}.$$

In this representation

$$g(1) = \left[-\alpha, -\Gamma\beta_1, \tilde{A}_2^*(1)\right]$$

is nonsingular due to assumption (3). Furthermore, $g(z) = \sum_{j=0}^\infty G_j z^j$ is a transfer function with $\sum_{j=0}^\infty \|G_j\| j^2 < \infty$ since $\sum_{j=1}^\infty \|\Pi_j\| j^2 < \infty$ and thus the same holds for the power series coefficients $A^*(L)$. Since $|B(z)| \neq 0, z \neq 1$ it follows that $|g(z)| \neq 0, |z| \leq 1$. Therefore

$$B(L)y_t = u_t, \quad g(L)u_t = \varepsilon_t \tag{4}$$

is a VAR process. Note, however, that $g(0) = G_0 \neq I_p$ in general. This constitutes a triangular representation of the process denoting $y_{1,t} = \beta'y_t \in \mathbb{R}^{p_1}, y_{2,t} = \beta_1'y_t \in \mathbb{R}^{p_2}, y_{3,t} = \beta_2'y_t \in \mathbb{R}^{p_3}$ such that

$$\begin{aligned}
y_{1,t} &= -\bar{\alpha}'\Gamma\beta_2\Delta y_{3,t} + u_{1,t} = A\Delta y_{3,t} + u_{1,t} \qquad A : p_1 \times p_3 \\
\Delta y_{2,t} &= u_{2,t}, \\
\Delta^2 y_{3,t} &= u_{3,t}
\end{aligned}$$

where $u_t = [u_{1,t}', u_{2,t}', u_{3,t}']'$ has a VAR(∞) representation. Furthermore, defining

$$\tilde{B}(L) = \mathcal{B}\begin{pmatrix} I_{p_1} & 0 & -\bar{\alpha}'\Gamma\beta_2 \\ 0 & I_{p_2} & 0 \\ 0 & 0 & I_{p_3} \end{pmatrix}B(L) = \Delta^2 I_p + \beta\beta'L + (\beta\beta' + \beta\bar{\alpha}'\Gamma\beta_2\beta_2' + \beta_1\beta_1')L\Delta,$$

$$\tilde{g}(L) = g(L)\left(\mathcal{B}\begin{pmatrix} I_{p_1} & 0 & -\bar{\alpha}'\Gamma\beta_2 \\ 0 & I_{p_2} & 0 \\ 0 & 0 & I_{p_3} \end{pmatrix}\right)^{-1}$$

we obtain $A(L) = g(L)B(L) = \tilde{g}(L)\tilde{B}(L)$ such that

$$\tilde{B}(L)y_t = \Delta^2 y_t + \tilde{\Pi}y_{t-1} + \tilde{\Gamma}\Delta y_{t-1} = v_t, \quad \tilde{g}(L)v_t = \varepsilon_t$$

is another representation of the process $(y_t)_{t \in \mathbb{Z}}$ with $\tilde{B}(0) = I_p$. It follows that the triangular representation can be seen as a special case where one has partial information on the matrices β, β_1, β_2. For estimation the VECM representation is approximated using a finite order h:

$$\Delta^2 y_t = \Phi y_{t-1} + \Psi \Delta y_{t-1} + \sum_{j=1}^{h-2} \Pi_j \Delta^2 y_{t-j} + e_t$$

where $e_t = \varepsilon_t + e_{1t}, e_{1t} = \sum_{j=h-1}^{\infty} \Pi_j \Delta^2 y_{t-j}$. As in the VECM representation the dimensions of β, β_1, β_2 are linked to the rank of the matrices Φ and $\alpha'_\perp \Psi \beta_\perp$. Restricting these matrices to be of particular rank is simpler than imposing the equivalent restrictions in the VAR(h) representation directly.

In the following we will first investigate the unrestricted ordinary least squares estimator in the VECM representation without taking rank restrictions into account. In the second step the 2SI2 procedure as presented in Paruolo (2000) for imposing the two rank restrictions in two steps is investigated.

For both procedures the selection of the order h is of importance. In this respect the following assumption will be used:

Assumption 2 (Lag order h). *The order h is chosen subject to the following restrictions:*

1. $h = o(T^{1/5})$.
2. $T^{1/2} \sum_{j=h+1}^{\infty} \|\Pi_j\| \to 0$ *as* $T, h \to \infty$.

This condition defines an upper bound for the order which is usually directly assured during order selection using for example information criteria. The upper bound is smaller than the usual rate $T^{1/3}$ for technical reasons. The stronger bound is not needed for all results. However, the implications for practical applications are minor as for example in the range $1 \le T \le 950$ we have $2.5T^{1/5} > T^{1/3}$. The second condition of Assumption 2 implies a lower bound for the increase of h as a function of the sample size. Clearly $\sum_{j=h+1}^{\infty} \|\Pi_j\| \to 0$ for $h \to \infty$. The bound implies that for $h = h(T)$ this convergence needs to be fast enough such that $T^{1/2} \sum_{j=h(T)+1}^{\infty} \|\Pi_j\|$ still converges to zero. The lower bound depends on the underlying true parameters. For invertible VARMA processes – which can be seen as the leading case – $\|\Pi_j\| \le C\rho_0^j$ for some $0 \le \rho_0 < 1$. Hannan and Deistler (1988) show that for an invertible stationary VARMA process the lower bound (in this case proportional to $\log T$) can be achieved asymptotically by using BIC as the order selection procedure. Thus in this case also the stronger condition ($h = o(T^{1/5})$) is satisfied. Bauer and Wagner (2004) extend this result to the multi frequency I(1) setting. For the I(2) case no analogous result is known, although the developments of Bauer and Wagner (2004) suggest that a similar result holds also there. This is left for future research.

Therefore the difference between the 'usual' rates and the ones assumed above are deemed to be of minor practical consequences. Thus we are not explicit in the main text as to which results hold true under the less restrictive set of results and which do not. In the appendix, we will comment on this point, however.

3. Unrestricted Estimation

In this section the results of Lewis and Reinsel (1985) and Saikkonen and Lütkepohl (1996) are extended to the I(2) case. To simplify notation define $\langle a_t, b_t \rangle = \sum_{t=h+1}^{T} a_t b_t'$ for sequences $a_t, b_t, t = 1, \ldots, T$.[1] Then the unrestricted least squares estimator in the finite VECM model uses the regressor vector $Z_{t,h} = [y'_{t-1}, \Delta y'_{t-1}, \Delta^2 y'_{t-1}, \ldots, \Delta^2 y'_{t-h+2}]' \in \mathbb{R}^{ph}$. The corresponding ordinary least squares estimator is given as

---

[1]   Here somewhat sloppily we use the same symbols for processes and their realizations.

$$[\hat{\Phi}, \hat{\Psi}, \hat{\Pi}_1, \ldots, \hat{\Pi}_{h-2}] = \left[\langle \Delta^2 y_t, y_{t-1} \rangle, \langle \Delta^2 y_t, \Delta y_{t-1} \rangle, \langle \Delta^2 y_t, \Delta^2 y_{t-1} \rangle, \ldots, \langle \Delta^2 y_t, \Delta^2 y_{t-h+2} \rangle \right] \langle Z_{t,h}, Z_{t,h} \rangle^{-1}$$
$$= \langle \Delta^2 y_t, Z_{t,h} \rangle \langle Z_{t,h}, Z_{t,h} \rangle^{-1}.$$

The noise covariance is estimated from the residuals as usual as

$$\hat{\Sigma}_\epsilon = N^{-1} \langle \hat{e}_t, \hat{e}_t \rangle, \quad \hat{e}_t = \Delta^2 y_t - \hat{\Phi} y_{t-1} - \hat{\Psi} \Delta y_{t-1} - \sum_{j=1}^{h-2} \hat{\Pi}_j \Delta^2 y_{t-j} \tag{5}$$

where $N = T - h$ denotes the effective sample size.

3.1. Estimation in the Triangular VECM Representation

As typical for the cointegration framework, analysis is easier in the triangular representation which separates stationary components from I(1) and I(2) processes: Let $y_t = [y'_{1,t}, y'_{2,t}, y'_{3,t}]' \in \mathbb{R}^p$ where $y_{i,t} \in \mathbb{R}^{p_i}$ is such that

$$y_{1,t} = A\Delta y_{3,t} + u_{1,t},$$
$$\Delta y_{2,t} = u_{2,t},$$
$$\Delta^2 y_{3,t} = u_{3,t}$$

where $u_t = [u'_{1,t}, u'_{2,t}, u'_{3,t}]'$ has a VAR(∞) representation $g(L)u_t = \varepsilon_t$ where

$$g(0) = \begin{pmatrix} I & 0 & A \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Note, however, that using the triangular representation implies that the matrix $B(L)$ is known up the value of the matrix A. For applications this is the case only seldom.

Thus letting $g(z) = g(1) + g^*(z)\Delta$ we obtain

$$\varepsilon_t = g(L) \begin{pmatrix} y_{1,t} - A\Delta y_{3,t} \\ \Delta y_{2,t} \\ \Delta^2 y_{3,t} \end{pmatrix} = g(L) \begin{pmatrix} \Delta^2 y_{1,t} + \Delta y_{1,t-1} + y_{1,t-1} - A\Delta^2 y_{3,t} - A\Delta y_{3,t-1} \\ \Delta^2 y_{2,t} + \Delta y_{2,t-1} \\ \Delta^2 y_{3,t} \end{pmatrix}$$

$$= g(L) \begin{pmatrix} I & 0 & -A \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \Delta^2 y_t + g(L) \begin{pmatrix} y_{1,t-1} \\ 0 \\ 0 \end{pmatrix} + g(L) \begin{pmatrix} \Delta y_{1,t-1} - A\Delta y_{3,t-1} \\ \Delta y_{2,t-1} \\ 0 \end{pmatrix}$$

$$= \tilde{g}(L)\Delta^2 y_t + [g(1) + g^*(L)\Delta] \begin{pmatrix} y_{1,t-1} \\ 0 \\ 0 \end{pmatrix} + g(1) \begin{pmatrix} \Delta y_{1,t-1} - A\Delta y_{3,t-1} \\ \Delta y_{2,t-1} \\ 0 \end{pmatrix}$$

$$= \pi(L)\Delta^2 y_t + g(1) \begin{pmatrix} y_{1,t-1} \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} \mathcal{G}_1 + \mathcal{G}_1^* & \mathcal{G}_2 & -\mathcal{G}_1 A \end{bmatrix} \begin{pmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta y_{3,t-1} \end{pmatrix}$$

$$= \pi(L)\Delta^2 y_t + \begin{bmatrix} \mathcal{G}_1 & 0 & 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} \mathcal{G}_1 + \mathcal{G}_1^* & \mathcal{G}_2 & -\mathcal{G}_1 A \end{bmatrix} \Delta y_{t-1}$$

with $\pi(L) = I_p - \sum_{j=1}^{\infty} \Pi_j L^j$ leads to the corresponding VECM representation:

$$\Delta^2 y_t = \Phi y_{t-1} + \Psi \Delta y_{t-1} + \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} + \varepsilon_t.$$

Here $\mathcal{G} := g(1) = \sum_{j=0}^{\infty} G_j = [\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3]$, where $\mathcal{G}_i$ is $p \times p_i$ for $i = 1, 2, 3$: Similarly, $\mathcal{G}^* := g^*(1) = -\sum_{j=0}^{\infty} jG_j = [\mathcal{G}_1^*, \mathcal{G}_2^*, \mathcal{G}_3^*]$, where $\mathcal{G}_i^*$ is $p \times p_i$ for $i = 1, 2, 3$. The sums exists since $\sum_{j=1}^{\infty} \|G_j\| j^2 < \infty$ by assumption. Similarly, we partition Φ, Ψ and Π_j into $[\Phi_1, \Phi_2, \Phi_3]$, $[\Psi_1, \Psi_2, \Psi_3]$ and $[\Pi_{j1}, \Pi_{j2}, \Pi_{j3}]$, respectively. The analogous partitioning is used for estimates.

Then $\Phi = -[\mathcal{G}_1, 0, 0]$, $\Psi = [-\mathcal{G}_1^* - \mathcal{G}_1, -\mathcal{G}_2, \mathcal{G}_1 A]$. Therefore $\Psi_3 = -\Phi_1 A$. Note that in this notation the I(2) components on the right hand side are $y_{t-1,3}$, the I(1) components are $y_{t-1,1}, y_{t-1,2}, \Delta y_{t-1,3}$, where $y_{t-1,1} - A\Delta y_{t-1,3}$ is stationary. Thus in order to separate regressors of different integration orders in the proof (as is usually done in the literature) we use a transformation using the unknown matrix A such that the regressor $y_{t-1,1}$ is replaced by $y_{t-1,1} - A\Delta y_{t-1,3}$. Consequently the estimate $\hat{\Psi}_3$ of Ψ_3 is replaced by the estimate $\hat{\Theta} = \hat{\Psi}_3 + \hat{\Phi}_1 A$ of $\Theta = \Psi_3 + \Phi_1 A = 0$.

Based on the estimates $\hat{\Psi}$ and $\hat{\Phi}$ then A can be estimated as

$$\hat{A} = -(\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1)^{-1} \hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Psi}_3. \tag{6}$$

Here the insertion of $\hat{\Sigma}_\epsilon^{-1}$ appears somewhat arbitrary. A motivation for this choice in the I(1) case can be found in Saikkonen (1992) equation (12). However, any other positive definite matrix could be used as well. Currently there is no knowledge on the optimality of the choice suggested above.

In the asymptotic distribution of the estimation error Brownian motions occur relating to the process $(u_t)_{t \in \mathbb{Z}}$: Under Assumption 1 we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t \Rightarrow B(r) = [B_1(r)', B_c(r)']' = [B_1(r)', B_2(r)', B_3(r)]'$$

where $B(r), 0 \le r \le 1$, denotes a Brownian motion with corresponding variance

$$\Omega = \left[\begin{array}{c|c} \Omega_{11} & \Omega_{1c} \\ \hline \Omega_{c1} & \Omega_{cc} \end{array} \right] = \left[\begin{array}{c|cc} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \hline \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{array} \right] = g(1)^{-1} \Sigma_\epsilon \, (g(1)')^{-1},$$

where $B_{1.c}(r) = B_1(r) - \Omega_{1c}\Omega_{cc}^{-1}B_c(r)$ is a p_1-dimensional Brownian motion, which is independent of $B_c(r)$, with covariance

$$\Omega_{1.c} = \Omega_{11} - \Omega_{1c}\Omega_{cc}^{-1}\Omega_{c1}.$$

An estimator of $\Omega_{1.c}$ is given by[2]

$$\hat{\Omega}_{1.c} = (\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1)^{-1}. \tag{7}$$

With these definitions we can state our first result of the paper (which is proved in Appendix B):

Theorem 1. *Under Assumptions 1 and 2 for the triangular VECM representation we have:*
(A) Consistency:

(i) $\hat{\Phi} \xrightarrow{p} \Phi$; *(ii)* $\hat{\Sigma}_\epsilon \xrightarrow{p} \Sigma_\epsilon$; *(iii)* $\hat{\Omega}_{1.c} \xrightarrow{p} \Omega_{1.c}$; *(iv)* $\hat{\Psi} \xrightarrow{p} \Psi$; *(v)* $\hat{\Theta} \xrightarrow{p} 0$; *(vi)* $\hat{A} \xrightarrow{p} A$.

(B) Asymptotic distribution of coefficients to nonstationary regressors: Under Assumptions 1 and 2 we have
($N = T - h$):

$$(i)[N\hat{\Phi}_2, N\hat{\Theta}, N^2\hat{\Phi}_3] \xrightarrow{d} g(1) \int_0^1 dBF' \left(\int_0^1 FF' \right)^{-1}, \quad (ii) \quad N(\hat{A} - A) \xrightarrow{d} \int_0^1 dB_{1.c}L' \left(\int_0^1 LL' \right)^{-1} \tag{8}$$

---

2 Note that $\alpha = [I_{p_1}, 0]'$, and thus $\Omega_{1.c} = ([\Omega^{-1}]_{11})^{-1} = (\alpha'\Omega^{-1}\alpha)^{-1} = (\alpha'g(1)'\Sigma_\epsilon^{-1}g(1)\alpha)^{-1} = (\Phi_1'\Sigma_\epsilon^{-1}\Phi_1)^{-1}$.

where $F(u) = \begin{bmatrix} B_c(u) \\ \int_0^u B_3(v)dv \end{bmatrix}$, $F_a(u) = \begin{bmatrix} B_2(u) \\ \int_0^u B_3(v)dv \end{bmatrix}$ *and* $L(u) = B_3(u) - \int_0^1 B_3 F_a'(\int_0^1 F_a F_a')^{-1} F_a(u)$.

(C) Asymptotic distribution of coefficients to stationary regressors: Let L_h be a sequence of $(p^2(h-2) + p(2p_1 + p_2)) \times J$ matrices such that $L_h'(\Gamma_{ECM}^{-1} \otimes \Sigma_\epsilon) L_h \to M > 0$ where $\Gamma_{ECM} = \mathbb{E}(X_t X_t')$ with $X_t := [u_{1,t-1}', \Delta y_{1,t-1}', \Delta y_{2,t-1}', \Delta^2 y_{t-1}', \dots, \Delta^2 y_{t-h+2}']'$.

   Let

$$\underline{\Pi} = \begin{bmatrix} \Phi_1 & \Psi_1 & \Psi_2 & \Pi_1 & \dots & \Pi_{h-2} \end{bmatrix}.$$

   Then

$$N^{\frac{1}{2}} L_h' \text{vec}(\hat{\Pi} - \underline{\Pi}) \xrightarrow{d} N(0, M).$$

(D) Asymptotic distribution on Wald type tests: Finally letting

$$\hat{\Gamma}_{ECM} = N^{-1}(\langle \tilde{X}_t, \tilde{X}_t \rangle - \langle \tilde{X}_t, \Delta y_{3,t-1} \rangle \langle \Delta y_{3,t-1}, \Delta y_{3,t-1} \rangle^{-1} \langle \Delta y_{3,t-1}, \tilde{X}_t \rangle)$$

where $\tilde{X}_t = [y_{1,t-1}', \Delta y_{1,t-1}', \Delta y_{2,t-1}', \Delta^2 y_{t-1}', \dots, \Delta^2 y_{t-h+2}']'$, the Wald test for the null hypothesis $H_0 : L_h' vec(\underline{\Pi}) = l_h$ is given by

$$\hat{\lambda}_{Wald} = N \left(L_h' \text{vec}(\hat{\Pi}) - l_h \right)' \left(L_h'(\hat{\Gamma}_{ECM}^{-1} \otimes \hat{\Sigma}_\epsilon) L_h \right)^{-1} \left(L_h' \text{vec}(\hat{\Pi}) - l_h \right).$$

   Then if L_h is such that $L_h'(\Gamma_{ECM}^{-1} \otimes \Sigma_\epsilon) L_h \to M > 0$, under the null hypothesis $\hat{\lambda}_{Wald} \xrightarrow{d} \chi^2(J)$.

The theorem provides the asymptotic distributions of the OLS estimates in the triangular system. Note that in this somewhat special case the properties of the regressor components (stationary or not) are known such that for each entry the convergence speed is known. Correspondingly the definition of the regressor vector $\tilde{X}_t$ involves only lags of y_t but omits all nonstationary regressors except the ones cointegrated with $\Delta y_{3,t-1}$.

The assumptions on L_h are more restrictive than needed. Lewis and Reinsel (1985) and Saikkonen and Lütkepohl (1996) only require that L_h has full column rank when deriving the normalized convergence to normal distribution with unit variance as the limit for

$$N^{\frac{1}{2}} (L_h'(\Gamma_{ECM}^{-1} \otimes \Sigma_\epsilon) L_h)^{-1/2} L_h' vec(\hat{\Pi} - \underline{\Pi}).$$

Similar arguments could be used here.

3.2. Estimation in the General VECM Representation

The previous section dealt with the special case that a triangular representation is used and hence knowledge on the matrices $[\beta, \beta_1, \beta_2]$ is given. This section provides a result for the general case, which, however, is limited to the coefficients to the stationary components. Since a general process generated according to Assumption 1 can be rewritten into a triangular representation using the knowledge of $[\beta, \beta_1, \beta_2]$, some asymptotic properties of the unrestricted OLS estimators can be derived from Theorem 1 for the general case (which is proved in Appendix C):

Theorem 2. *Let the regressor vector $Z_{t,h} = [y_{t-1}', \Delta y_{t-1}', \Delta^2 y_{t-1}', \dots, \Delta^2 y_{t-h+2}']$ and define*

$$\underline{\Lambda} = \begin{bmatrix} \Phi & \Psi & \Pi_1 & \dots & \Pi_{h-2} \end{bmatrix}, \quad \tilde{\Lambda} = \langle \Delta^2 y_t, Z_{t,h} \rangle \langle Z_{t,h}, Z_{t,h} \rangle^{-1}, \quad \tilde{\Gamma}_{ECM} = N^{-1} \langle Z_{t,h}, Z_{t,h} \rangle.$$

   Then under Assumptions 1 and 2 it follows that $\tilde{\Lambda} - \underline{\Lambda} = o_P(1)$.

Furthermore, let $L_h \in \mathbb{R}^{p^2(h+2) \times J}$ be such that $L_h'(\tilde{\Gamma}_{ECM}^{-1} \otimes \Sigma_\epsilon)L_h \to M > 0$. Then

$$N^{\frac{1}{2}} L_h' \mathrm{vec}(\tilde{\Lambda} - \underline{\Lambda}) \xrightarrow{d} N(0, M).$$

Beside consistency the theorem implies that linear combination of OLS estimators show asymptotic normality and hence standard inference, if the asymptotic variance is nonsingular. One application of such results consists in the so called 'surplus lag' formulation in the context of Granger causality testing, see Bauer and Maynard (2012); Dolado and Lütkepohl (1996).

Finally note that this section does not contain results with regard to the cointegrating rank or the cointegrating space. The theorem above merely allows to test coefficients corresponding to stationary regressors. Therefore the usage is limited to somewhat special situations like the surplus-lag causality tests. However, it is also relevant for impulse response analysis, compare Inoue and Kilian (2020).

4. Rank Restricted Regression

The previous sections show that for the estimators discussed in that sections full inference on all coefficients is only possible when information on the matrices β, β_1 and β_2 exists. The dimensions of the matrices relate to the ranks of the matrices $\Phi = \alpha\beta'$ and, conditional on this, to the rank of $\bar{\alpha}_\perp' \Psi \bar{\beta}_\perp$. The two rank restrictions make estimation and specification more complex than in the I(1) case.

Johansen (1995) provides the two-step approach 2SI2 that can be used for estimation and specification of the two integer valued parameters p_1 and p_2. Paruolo and Rahbek (1999) extend the 2SI2 procedure suggested in section 8 of Johansen (1997). Paruolo (2000) shows that this 2SI2 procedure achieves the same asymptotic distribution as pseudo maximum likelihood estimation which could be performed subsequent to 2SI2 estimation. This makes the procedure attractive from a practical point of view. In this section we show that these approaches extend naturally to the long VAR case. The main focus here lies on the derivation of the asymptotic properties of the rank tests.

Recall the long VAR approximation given as

$$\Delta^2 y_t = \Phi y_{t-1} + \Psi \Delta y_{t-1} + \sum_{j=1}^{h-2} \Pi_j \Delta^2 y_{t-j} + e_t \tag{9}$$

where $\Phi = \alpha\beta'$ has reduced rank $r < p$ and $\bar{\alpha}_\perp' \Psi \bar{\beta}_\perp = \zeta\eta'$ has reduced rank $s < p - r$. In this notation the 2SI2 procedure works as follows: In the first step the rank constraint on $\bar{\alpha}_\perp' \Psi \bar{\beta}_\perp$ is neglected estimating α and β by using reduced-rank regression (RRR). Then in the second step the reduced rank of $\bar{\alpha}_\perp' \Psi \bar{\beta}_\perp$ is imposed using RRR in a transformed equation.

In more detail using the Johansen notation we denote with R_{0t}, R_{1t} and R_{2t} the residuals of regressing $\Delta^2 y_t, \Delta y_{t-1}$ and y_{t-1} on $\Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-h+2}$, respectively; then we can rewrite (9) as

$$R_{0t} = \alpha\beta' R_{2t} + \Psi R_{1t} + \tilde{e}_t. \tag{10}$$

Concentrating out R_{1t} and denoting the residuals as $R_{0.1t}$ and $R_{2.1t}$ we obtain with $S_{ij.1} = \langle R_{it}, R_{jt} \rangle - \langle R_{it}, R_{1t} \rangle \langle R_{1t}, R_{1t} \rangle^{-1} \langle R_{1t}, R_{jt} \rangle$ the solution to the RRR problem from solving the eigenvalue problem

$$|\lambda S_{22.1} - S_{20.1} S_{00.1}^{-1} S_{02.1}| = 0, \tag{11}$$

with solutions $1 > \hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p > 0$ ordered with decreasing size and corresponding vectors $V = (v_1, \ldots, v_p)$. Then as usual the trace statistic of testing the model H_r with $\mathrm{rank}(\Phi) \leq r, r < p$, in the model H_p with $\mathrm{rank}(\Phi) \leq p$, is given as

$$Q_r = -2 \log Q(H_r | H_p) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i). \tag{12}$$

The optimizers for α, β are given by

$$\hat{\beta} = (v_1, \ldots, v_r), \qquad \hat{\alpha} = S_{02.1}\hat{\beta}, \qquad \hat{\Sigma}_\epsilon = S_{00.1} - \hat{\alpha}\hat{\alpha}'. \tag{13}$$

In the second step, given α and β known, we can obtain by multiplying (10) by $\bar{\alpha}'_\perp$ that

$$\bar{\alpha}'_\perp R_{0t} = \bar{\alpha}'_\perp \Psi(\bar{\beta}_\perp \beta'_\perp + \bar{\beta}\beta')R_{1t} + \bar{\alpha}'_\perp \tilde{e}_t = \zeta\eta'(\beta'_\perp R_{1t}) + C(\beta' R_{1t}) + \bar{\alpha}'_\perp \tilde{e}_t. \tag{14}$$

Note that $\beta' R_{1t}$ is stationary. Thus concentrating out C and denoting the residuals as $R_{\bar{\alpha}_\perp.\beta,t}$ and $R_{\beta_\perp.\beta,t}$, respectively, we can define $S_{ab.\beta} := \langle R_{a.\beta,t}, R_{b.\beta,t} \rangle$, for $a, b = \bar{\alpha}_\perp$ or $\beta_\perp$. Then the likelihood ratio test of the model $H_{r,s}$ with $\mathrm{rank}(\zeta\eta') \leq s$, $s < p - r$ in the model H_r^0 with $\mathrm{rank}(\bar{\alpha}'_\perp \Psi \bar{\beta}_\perp) = p - r$ is given by

$$Q_{r,s} = -2\log Q(H_{r,s}|H_r^0) = -T \sum_{i=s+1}^{p-r} \log(1 - \hat{\rho}_i). \tag{15}$$

where $1 > \hat{\rho}_1 \geq \ldots \geq \hat{\rho}_{p-r} > 0$ are the solutions of the eigenvalue problem

$$|\rho S_{\beta_\perp \beta_\perp.\beta} - S_{\beta_\perp \bar{\alpha}_\perp.\beta} S_{\bar{\alpha}_\perp \bar{\alpha}_\perp.\beta}^{-1} S_{\bar{\alpha}_\perp \beta_\perp.\beta}| = 0, \tag{16}$$

and the corresponding eigenvectors are $W = (w_1, \ldots, w_{p-r})$. Estimators of ζ and η are given by

$$\hat{\eta} = (w_1, \ldots, w_s), \qquad \hat{\zeta} = S_{\bar{\alpha}_\perp \beta_\perp.\beta} \, \hat{\eta}. \tag{17}$$

For the 2SI2 procedure in this second step the first step estimates $\hat{\alpha}$ and $\hat{\beta}$ are used in place of the unknown true quantities. Then we obtain the following analogon to the results in the finite order VAR framework (the proof is given in Appendix D):

Theorem 3. *Let the data be generated according to Assumption 1 and let the VAR order fulfill Assumption 2. Then the following asymptotic results hold:*
(A) The asymptotic distribution of the likelihood ratio statistic Q_r under the null hypothesis H_r is given by

$$Q_r \xrightarrow{d} \mathrm{tr}\left\{ \int_0^1 dW_{\dagger} F'_{\dagger} \left(\int_0^1 F_{\dagger} F'_{\dagger} du \right)^{-1} \int_0^1 F_{\dagger} dW'_{\dagger} \right\}. \tag{18}$$

where $W_{\dagger} = (\alpha'_\perp \Sigma_\epsilon \alpha_\perp)^{-1/2} \alpha'_\perp W, F_a(u) = \begin{bmatrix} B_2(u) \\ \int_0^u B_3(v)dv \end{bmatrix}$ and $F_{\dagger}(u) = F_a(u) - \int_0^1 F_a B'_3 (\int_0^1 B_3 B'_3)^{-1} B_3(u)$.
This is identical to the distribution achieved in the finite VAR case.
(B) The asymptotic distribution of the likelihood ratio statistic $Q_{r,s}$ under the null hypothesis $H_{r,s}$ is given by

$$Q_{r,s} \xrightarrow{d} \mathrm{tr}\left\{ \int_0^1 dW'_2 B'_3 \left(\int_0^1 B_3 B'_3 du \right)^{-1} \int_0^1 B_3 dW'_2 \right\}. \tag{19}$$

where $W_2(u) = (\alpha'_2 \Sigma_\epsilon \alpha_2)^{-1/2} \alpha'_2 W(u)$.
(C) The asymptotic distribution of the test statistic $S_{r,s} = Q_r + Q_{r,s}$ under the null hypothesis $H_{r,s}$ is given by

$$S_{r,s} \xrightarrow{d} \mathrm{tr}\left\{ \int_0^1 dW_{\dagger} F'_{\dagger} \left(\int_0^1 F_{\dagger} F'_{\dagger} du \right)^{-1} \int_0^1 F_{\dagger} dW'_{\dagger} \right\} + \mathrm{tr}\left\{ \int_0^1 dW_2 B'_3 \left(\int_0^1 B_3 B'_3 du \right)^{-1} \int_0^1 B_3 dW'_2 \right\}. \tag{20}$$

(D) Using suitable normalizations all estimators are consistent: $\hat{\alpha}(c'_\alpha \hat{\alpha})^{-1} \xrightarrow{p} \alpha, \hat{\beta}(c'_\beta \hat{\beta})^{-1} \xrightarrow{p} \beta, \hat{\zeta}(c'_\zeta \hat{\zeta})^{-1} \xrightarrow{p} \zeta, \hat{\eta}(c'_\eta \hat{\eta})^{-1} \xrightarrow{p} \eta, \hat{\Psi} \xrightarrow{p} \Psi, \hat{\Pi}_j \xrightarrow{p} \Pi_j$ where for example $c'_\alpha \alpha = I_r$.
(E) The asymptotic distributions of the coefficients to the nonstationary regressors are identical to the ones in the

finite order VAR case stated in Paruolo (2000). *The asymptotic distribution of the coefficients* $\hat{\Pi}_j$ *are identical to the ones in Theorem* 1.

The main message of the theorem is that the 2SI2 procedure shows the same asymptotic properties including the rank tests as in the finite order VAR case. As usual also restricting the coefficients for the non-stationary regressors does not influence the asymptotics for the coefficients corresponding to the stationary regressors.

Note that Paruolo (2000) shows that in the finite VAR case 2SI2 estimates have the same asymptotic distribution as pseudo maximum likelihood (pML) estimates maximizing the Gaussian likelihood. The first order conditions for the pML estimates of the coefficients to the non-stationary regressors provided in the first display on p. 548 in Paruolo (2000) depend on the data only via the matrices S_{ij} defined above. These matrices depend on the lag length of the VECM only via the concentration step. The proof of our Theorem 3 shows that these terms have the same asymptotic distributions for the finite order VAR and the long VAR. Theorem 4.3 of Paruolo (2000) shows that the asymptotic distribution of the coefficients due to stationary regressors does not depend on the distribution of the coefficients corresponding to the non-stationary regressors as long as they are estimated super-consistently. Thus our results imply that also in the long VAR case the asymptotic distribution of all estimates for the 2SI2 and the pML approach is identical.

5. Initial Guess for VARMA Estimation

One usage of long VAR approximations is as preliminary estimate for VARMA model estimation. Hannan and Kavalieris (1986) provide properties of such an approach in the stationary case, Lütkepohl and Claessen (1997) extend the procedure to the I(1) case. Here we extend this idea to the I(2) case.

The goal is to provide a consistent initial guess for the estimation of a VARMA model for I(2) processes. In this respect we assume the following data generating process:

Assumption 3 (VARMA dgp). *The process* $(y_t)_{t \in \mathbb{Z}}$ *is generated as the solution to the state space equations*

$$y_t = Cx_t + \varepsilon_t, \qquad x_{t+1} = Ax_t + B\varepsilon_t \tag{21}$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ *denotes white noise subject to the same assumptions as in Assumption* 1.

Here $x_t \in \mathbb{R}^n$ *is the unobserved state process. The system* (A, B, C) *is assumed to be minimal and in the canonical form of* Bauer and Wagner (2012), *that is*

$$A = \begin{bmatrix} I_c & I_c & 0 & 0 \\ 0 & I_c & 0 & 0 \\ 0 & 0 & I_d & 0 \\ 0 & 0 & 0 & A_\bullet \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_\bullet \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 & C_3 & C_\bullet \end{bmatrix},$$

where $|\lambda_{max}(A_\bullet)| < 1$ *(the matrix* $A_\bullet$ *is stable),* $C_1'C_1 = I_c, C_3'C_3 = I_d, C_1'C_3 = 0, C_1'C_2 = 0, C_2'C_3 = 0.$ *Furthermore, the system is strictly minimum-phase, that is* $\rho_0 = |\lambda_{max}(A - BC)| < 1$. *Finally the matrix* $\bar{A} = A - BC$ *is nonsingular.*

At time $t = 0$ *the state* $x_0 = [x_{0,u}', x_\bullet']', x_{0,u} \in \mathbb{R}^{2c+d}$, *is such that* $x_{0,u}$ *is deterministic and* $x_{0,\bullet} = \sum_{j=1}^{\infty} A_\bullet^{j-1} B_\bullet \varepsilon_{-j}$ *denotes the stationary solution to the stable part of the system.*

In this situation it follows that $(y_t)_{t \in \mathbb{Z}}$ is an I(2) process in the definition of Bauer and Wagner (2012), that is its second difference is a stationary VARMA process. The integers c and d are connected to the integers p_1, p_2, p_3 via $c = p_3, d = p_2$ such that $p_1 = p - c - d$. It can furthermore be shown that a process generated using Assumption 3 possesses a VAR(h) approximation:

$$y_t + \sum_{j=1}^{h} A_j y_{t-j} = \varepsilon_t + C(A - BC)^h x_{t-h}$$

where $A_j = -C(A - BC)^{j-1}B, \|A_j\| \leq \mu \rho^j$ $(0 \leq \rho_0 < \rho < 1)$ converges to zero exponentially fast for $j \to \infty$ due to the strict minimum-phase condition. Letting $h \to \infty$ then implies the existence of a VAR(∞) representation. It follows that for such systems $A(z)$ converges absolutely for $|z| < \rho^{-1}$ where $1 < \rho^{-1}$.

From the autoregressive representation the VECM representation can be obtained:

$$a(z) = I_p + \sum_{j=1}^{\infty} A_j z^j = I_p - \sum_{j=1}^{\infty} C\bar{A}^{j-1}Bz^j = (1-z)^2 I_p - \Phi z - \Psi z(1-z) - (1-z)^2 \sum_{j=1}^{\infty} \Pi_j z^j$$

where $\bar{A} = A - BC$ such that

$$\Delta^2 y_t = \Phi y_{t-1} + \Psi \Delta y_{t-1} + \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} + \varepsilon_t.$$

A comparison of power series coefficients provides the identities:

$$\Phi = -I_p + C(I - \bar{A})^{-1}B,$$
$$\Psi = -I_p - C(I - \bar{A})^{-2}\bar{A}B,$$
$$\Pi_j = [C\bar{A}^2(I - \bar{A})^{-2}]\bar{A}^{j-1}B = D\bar{A}^{j-1}B, j = 1, 2, \ldots$$

It follows that the coefficients $\Pi_j, j = 1, 2, \ldots$ form the impulse response of a rational transfer function of order smaller or equal to n. If $\bar{A}$ is nonsingular then the order equals n and the system $(\bar{A}, B, D)$ is minimal. Furthermore, it follows that for arbitrary Φ and Ψ the transfer function

$$a(z) = (1-z)^2 I_p - \Phi z - \Psi z(1-z) - (1-z)^2 z D(I - z\bar{A})^{-1}B$$

is a rational transfer function with the additional property that

$$a(1) = -\Phi = -\alpha\beta', \quad \bar{\alpha}'_\perp \dot{a}(1)\bar{\beta}_\perp = \bar{\alpha}'_\perp(-\Phi + \Psi)\bar{\beta}_\perp = -\bar{\alpha}'_\perp C(I - \bar{A})^{-2}B\bar{\beta}_\perp = \zeta\eta'.$$

Consequently Φ and Ψ determine the integration properties of processes generated using $a(z)$. Conversely whenever the constraints

$$-I_p + C(I - \bar{A})^{-1}B = \alpha\beta', \quad -\bar{\alpha}'_\perp C(I - \bar{A})^{-2}B\bar{\beta}_\perp = \zeta\eta'$$

hold the corresponding triple (A, B, C) corresponds to an I(2) process (if the eigenvalues of A are in the closed unit disc). Defining $C_* = \bar{\alpha}'C, C_\dagger = \bar{\alpha}'_\perp C$ we obtain

$$-\bar{\alpha}' + C_*(I - \bar{A})^{-1}B = \beta', \quad -\bar{\alpha}'_\perp + C_\dagger(I - \bar{A})^{-1}B = 0, \quad -C_\dagger(I - \bar{A})^{-2}B\bar{\beta}_\perp = \zeta\eta'. \quad (22)$$

The third equation does not have a solution for fixed $B\bar{\beta}_\perp, \zeta, \eta$, if the row space of $B\bar{\beta}_\perp$ does not contain the space spanned by the rows of η'. In this case row-wise projection of η' onto the space spanned by the rows of $B\bar{\beta}_\perp$ allows for (not necessarily unique) solutions in $C_\dagger$. In the limit no projection is needed. Consequently for large enough T the projected matrix will have full row rank. The second equation then determines $\bar{\alpha}_\perp$ which in turn determines $\bar{\alpha}$ up to the choice of the basis such that $\bar{\alpha}' = T_C \bar{\alpha}_o'$ for some full row rank matrix $\bar{\alpha}_o' \in \mathbb{R}^{r \times p}, \bar{\alpha}_o' \bar{\alpha}_\perp = 0$. The first equation then can be rewritten as

$$[T_C, C_*] \underbrace{\begin{bmatrix} -\bar{\alpha}_o' \\ (I - \bar{A})^{-1}B \end{bmatrix}}_{R_1} = \beta'.$$

The second equation shows that the row space of $(I - \bar{A})^{-1}B$ contains the row space of $\bar{\alpha}_\perp'$. Thus the matrix R_1 has full row rank. It follows that this equation has solutions.

Having obtained a solution for $C_*, C_\dagger, \bar{\alpha}, \bar{\alpha}_\perp$ then C is obtained from

$$C = \begin{bmatrix} \alpha & \alpha_\perp \end{bmatrix} \begin{bmatrix} C_* \\ C_\dagger \end{bmatrix}.$$

A unique solution then can be obtained from adding the restrictions $\Pi_j = C(I - \bar{A})^{-2}\bar{A}^{j+1}B, j = 1, 2, \ldots, 2n$ which for the estimates are to be solved in a least squares sense among all solutions to equations (22).

It then follows that for the true matrices Φ, Ψ, Π_j the only solution for given $\bar{A}, B$ consists in the corresponding true C. These facts therefore can be used in order to develop an initial guess for subsequent pseudo likelihood maximization using the parameterization of I(2) processes in state space representation: Given the integer valued parameters n, c and d:

1. Obtain a long VAR approximation $\hat{\Phi}, \hat{\Psi}, \hat{\Pi}_j, j = 1, 2, \ldots$, including $\hat{\Phi} = \hat{\alpha}\hat{\beta}'$ and $\hat{\zeta}\hat{\eta}' = \widehat{\bar{\alpha}_\perp}'\hat{\Psi}\widehat{\bar{\beta}_\perp}$ using the 2SI2 approach.
2. Choose the integer $f \geq n$. Use the algorithm described in Appendix F to obtain estimates $(\hat{\bar{A}}, \hat{B}, \hat{D})$ realizing the impulse response $\hat{\Pi}_j, j = 1, \ldots, 2f$ from the Hankel matrix with f block columns and f block rows.
3. Project rows of $\hat{\eta}'$ onto the space spanned by the rows of $\hat{B}\widehat{\bar{\beta}_\perp}$ to obtain $\tilde{\eta}'$.
4. Obtain a unique solution $\hat{C}$ solving (22) such that the matrices $\tilde{\Pi}_j = \hat{C}(I - \hat{\bar{A}})^{-2}\hat{\bar{A}}^{j+1}\hat{B}, j = 1, 2, \ldots, 2n$ have minimal Euclidean distance to $\hat{\Pi}_j, j = 1, 2, \ldots, 2n$.
5. Transform the corresponding system $(\hat{\bar{A}} + \hat{B}\hat{C}, \hat{B}, \hat{C})$ to the canonical form of Bauer and Wagner (2012) to obtain the estimate $(\tilde{A}, \tilde{B}, \tilde{C})$.

The algorithm obtains a minimal state space system of order n in the canonical form for I(2) processes given in Bauer and Wagner (2012) and hence can be used as an initial guess for subsequent pseudo-likelihood optimization in the set $M_n(r, s)$ of all order n rational transfer functions corresponding to I(2) processes with state space unit root structure $((0, (c, c + d)))$.

Theorem 4 (Consistent initial guess). *Let $(y_t)_{t \in \mathbb{Z}}$ denote a process generated using the system (A_0, B_0, C_0) according to Assumption 3 and let the system $(\tilde{A}, \tilde{B}, \tilde{C})$ be estimated based on the long VAR approximation with lag order chosen according to Assumption 2. Then $(\tilde{A}, \tilde{B}, \tilde{C})$ is a weakly consistent estimator of the data generating system (A_0, B_0, C_0) in the sense that $\tilde{C}\tilde{A}^j\tilde{B} \xrightarrow{p} C_0 A_0^j B_0, j = 0, 1, \ldots$ and hence the corresponding transfer functions converge in pointwise topology.*

The proof of this theorem can be found in Appendix E.

6. Conclusions

In this paper the theory on long VAR approximation of general linear dynamical processes is extended to the case of I(2) processes. We find that we need slightly narrower upper and lower bounds in the approximations. The tighter bounds are not needed for all results and appear not very restrictive for applications.

The main results are completely analogous to the I(1) case: The asymptotics in many respects is identical to the finite order VAR case. Asymptotic distributions for the coefficients to non- stationary variables are the same as in the finite order VAR case. This holds true both for unrestricted OLS

estimates as well as the 2SI2 approach in the Johansen framework. Tests on cointegrating ranks show identical asymptotic distributions under the null as in the finite order VAR case and hence do not require other tables. In this respect the main conclusion is that the usual procedure of estimating the lag order in the first step and then applying the Johansen procedure for estimated lag order is justified also for processes generated from a VAR(∞) that is approximated with a choice of the lag order lying within the prescribed bounds.

Additionally in the VARMA case the long VAR approximation can be used in order to derive consistent initial guesses that can be used in subsequent pseudo likelihood estimation.

Thus the paper provides both a full extension of results that have been achieved in the I(1) case as well as a useful starting point for subsequent VARMA modeling which might be preferable in situations which require a high VAR order or show a large number of variables to be modeled, a situation where VARMA models can be more parsimonious than VAR models.

Author Contributions: The two authors of the paper have contributed equally, via joint efforts, regarding both ideas, research, and writing. Conceptualization, Y.L. and D.B.; methodology, Y.L. and D.B.; software, not applicable; validation, not applicable; formal analysis, Y.L. and D.B.; investigation, Y.L. and D.B.; resources, not applicable; writing–original draft preparation, Y.L. and D.B.; writing–review and editing, Y.L. and D.B.; visualization, not applicable; supervision, not applicable; project administration, D.B.; funding acquisition, D.B. Both authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Preliminaries

The theory in this paper follows closely the arguments in Lewis and Reinsel (1985) and its extension to the I(1) case in Saikkonen and Lütkepohl (1996). To this end consider the finite order VECM approximation:

$$\Delta^2 y_t = \Phi y_{t-1} + \Psi \Delta y_{t-1} + \sum_{j=1}^{h} \Pi_j \Delta^2 y_{t-j} + e_t. \tag{A1}$$

The properties of the various estimators heavily use the following rewriting of the approximation using the triangular representation of y_t:

$$
\begin{aligned}
\Delta^2 y_t = [\Phi_1,\ \Phi_2,\ \Phi_3] &\begin{bmatrix} A\Delta y_{3,t-1} + u_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + [\Psi_1,\ \Psi_2,\ \Psi_3] \begin{bmatrix} Au_{3,t-1} + \Delta u_{1,t-1} \\ u_{2,t-1} \\ \Delta y_{3,t-1} \end{bmatrix} \\
&+ \sum_{j=1}^{h} [\Pi_{j,1},\ \Pi_{j,2},\ \Pi_{j,3}] \begin{bmatrix} A\Delta u_{3,t-j} + \Delta^2 u_{1,t-j} \\ \Delta u_{2,t-j} \\ u_{3,t-j} \end{bmatrix} + e_t \\
&= \Phi_2 y_{2,t-1} + \Phi_3 y_{3,t-1} + \Theta \Delta y_{3,t-1} + \sum_{j=1}^{h} \Xi_j u_{t-j} + [\Xi_{h+1,1},\ \Xi_{h+1,2}] \begin{bmatrix} u_{1,t-h-1} \\ u_{2,t-h-1} \end{bmatrix} + \Xi_{h+2,1} \tilde{u}_{1,t-h-2} + e_t,
\end{aligned} \tag{A2}
$$

where $\tilde{u}_{1,t-h-2} := u_{1,t-h-2} - Au_{3,t-h-1}$ and $\Phi_2 = \Phi_3 = 0$, $\Theta = \Phi_1 A + \Psi_3 = 0$, and

$$
\begin{aligned}
\Xi_1 &= [\Phi_1 + \Psi_1 + \Pi_{1,1},\ \Psi_2 + \Pi_{1,2},\ (\Psi_1 + \Pi_{1,1})A + \Pi_{1,3}], \\
\Xi_2 &= [-\Psi_1 + \Pi_{2,1} - 2\Pi_{1,1},\ \Pi_{2,2} - \Pi_{1,2},\ (\Pi_{2,1} - \Pi_{1,1})A + \Pi_{2,3}], \\
\Xi_j &= [\Pi_{j,1} - 2\Pi_{j-1,1} + \Pi_{j-2,1},\ \Pi_{j,2} - \Pi_{j-1,2},\ (\Pi_{j,1} - \Pi_{j-1,1})A + \Pi_{j,3}],\quad j = 3,\dots,h, \\
\Xi_{h+1,1} &= -2\Pi_{h,1} + \Pi_{h-1,1},\qquad \Xi_{h+1,2} = -\Pi_{h,2},\qquad \Xi_{h+2,1} = \Pi_{h,1}.
\end{aligned}
$$

Furthermore, we can see that $\sum_{j=1}^{h+2} \Xi_{j,1} = \Phi_1$, $\sum_{j=1}^{h+1} \Xi_{j,2} = \Psi_2$, and $\sum_{j=1}^{h} \Xi_{j,3} = \Psi_1 A + \sum_{j=1}^{h} \Pi_{j,3}$. Finally $\Psi_1 = -\sum_{j=2}^{h+2} (j-1)\Xi_{j,1}$.

Note that in the reparametrization (A2), the I(1) components, $y_{c,t} := (y_{2,t}', \Delta y_{3,t}')'$, as well as the I(2) components, $y_{3,t-1}$, are isolated from the stationary ones, u_{t-j}, and have coefficients equal to zero, which facilitates the derivation of the asymptotic properties.

In the reparameterized setting define ^{3} $\underline{\Xi} := [\Xi_1, \ldots, \Xi_h, \Xi_{h+1,1}, \Xi_{h+1,2}, \Xi_{h+2,1}]$, $\quad p \times (ph + 2p_1 + p_2)$,
$U_t := [u_{t-1}', \ldots, u_{t-h}', u_{1,t-h-1}', u_{2,t-h-1}', \tilde{u}_{1,t-h-2}']'$, $\quad (ph + 2p_1 + p_2) \times 1$,
$\underline{\Lambda} := [\underline{\Xi}, \Phi_2, \Theta, \Phi_3] = [\underline{\Xi}, 0]$, $\quad p \times p(h+2)$,
$W_t := [U_t', y_{c,t-1}', y_{3,t-1}']'$, $\quad p(h+2) \times 1$.

we have

$$\Delta^2 y_t = \underline{\Lambda} W_t + e_t, \tag{A3}$$

and correspondingly,

$$\Delta^2 y_t = \hat{\Lambda} W_t + \tilde{e}_t$$

where

$$\hat{\Lambda} = [\hat{\underline{\Xi}}, \hat{\Phi}_2, \hat{\Theta}, \hat{\Phi}_3] = \langle \Delta^2 y_t, W_t \rangle \langle W_t, W_t \rangle^{-1}$$

is the OLS estimator of $\underline{\Lambda}$. Here $\langle X_t, Z_t \rangle := \sum_{t=h+3}^{T} X_t Z_t'$.
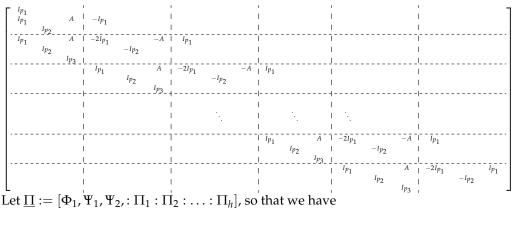
Note that W_t and the regressors in (A1) are in one-one correspondence. In the original Equation (A1) beside the nonstationary regressors $y_{c,t-1}$ and $y_{3,t-1}$ the regressor vector

$$\tilde{X}_t = [y_{1,t-1}', \Delta y_{1,t-1}', u_{2,t-1}', \Delta^2 y_{t-1}', \ldots, \Delta^2 y_{t-h}']' \in \mathbb{R}^{2p_1 + p_2 + ph}$$

occurs which cointegrates with $\Delta y_{3,t-1}$ such that

$$X_t = \tilde{X}_t - [A', 0]' \Delta y_{3,t-1} = T_h U_t \tag{A4}$$

is stationary. Here the nonsingular matrix $T_h \in \mathbb{R}^{(ph+2p_1+p_2) \times (ph+2p_1+p_2)}$ is defined as:

$$\begin{bmatrix}
I_{p_1} & & & & & & & \\
I_{p_1} & & A & -I_{p_1} & & & & \\
& I_{p_2} & & & & & & \\
\hline
I_{p_1} & & A & -2I_{p_1} & & -A & I_{p_1} & \\
& I_{p_2} & & -I_{p_2} & & & & \\
& & I_{p_3} & & & & & \\
\hline
& & & I_{p_1} & A & -2I_{p_1} & -A & I_{p_1} \\
& & & & I_{p_2} & & -I_{p_2} & \\
& & & & I_{p_3} & & & \\
\hline
& & & & & \ddots & & \ddots & & \ddots \\
\hline
& & & & & & & I_{p_1} & A & -2I_{p_1} & -A & I_{p_1} \\
& & & & & & & & I_{p_2} & & -I_{p_2} \\
& & & & & & & & I_{p_3} \\
\hline
& & & & & & & & & I_{p_1} & A & -2I_{p_1} & I_{p_1} \\
& & & & & & & & & & I_{p_2} & -I_{p_2} \\
& & & & & & & & & I_{p_3} \\
\end{bmatrix}$$

Let $\underline{\Pi} := [\Phi_1, \Psi_1, \Psi_2, : \Pi_1 : \Pi_2 : \ldots : \Pi_h]$, so that we have

$$\underline{\Xi} = \underline{\Pi} T_h. \tag{A5}$$

It can be verified that T_h is invertible. The asymptotic properties of $\hat{\Lambda} - \underline{\Lambda}$ are clarified in the next lemma:

Lemma A1. *Under the assumptions of Theorem 1 using $N = T - h - 2$ as the effective sample size*

---

[3] In this appendix processes whose dimension depends on the choice of h are denoted using upper case letters neglecting the dependence on h in the notation otherwise for simplicity.

$$N^{\frac{1}{2}}(\hat{\Xi} - \Xi) = N^{\frac{1}{2}}\langle \varepsilon_t, U_t\rangle(\mathbb{E}U_t U_t')^{-1} + o_P(h^{\frac{1}{2}}),$$

$$[N\hat{\Phi}_2, N\hat{\Theta}, N^2\hat{\Phi}_3] \Rightarrow g(1)\left[\begin{array}{cc}\int_0^1 dBB_c' & \int_0^1 dBH_3'\end{array}\right]\left[\begin{array}{cc}\int_0^1 B_c B_c' & \int_0^1 B_c H_3' \\ \int_0^1 H_3 B_c' & \int_0^1 H_3 H_3'\end{array}\right]^{-1}$$

where $H_3(u) = \int_0^u B_3(s)ds.$

Proof. The proof essentially shows that the coefficients corresponding to the stationary regressors and the ones corresponding to the integrated regressors asymptotically can be dealt with separately. Let $D_T := \text{diag}[N^{-\frac{1}{2}}I_{ph+2p_1+p_2}, \ N^{-1}I_{p_2+p_3}, \ N^{-2}I_{p_3}]$. Note that $N^{\frac{1}{2}}(\hat{\Xi} - \Xi)$, $N[\hat{\Phi}_2, \hat{\Theta}]$, and $N^2\hat{\Phi}_3$ are the 1st, 2nd and 3rd column blocks of $(\hat{\Lambda} - \underline{\Lambda})D_T^{-1}$, respectively. Moreover, we have

$$(\hat{\Lambda} - \underline{\Lambda})D_T^{-1} = \langle e_t, W_t\rangle D_T (D_T\langle W_t, W_t\rangle D_T)^{-1}.$$

Let $\hat{R} := D_T\langle W_t, W_t\rangle D_T$, and define $R := \text{diag}[\Gamma_u, \ R_2]$, where $\Gamma_u = \mathbb{E}[U_t U_t']$, and

$$R_2 := \left[\begin{array}{cc} N^{-2}\langle y_{c,t-1}, y_{c,t-1}\rangle & N^{-3}\langle y_{c,t-1}, y_{3,t-1}\rangle \\ N^{-3}\langle y_{3,t-1}, y_{c,t-1}\rangle & N^{-4}\langle y_{3,t-1}, y_{3,t-1}\rangle \end{array}\right].$$

Note that each block of the matrix R_2 is of order $O_p(1)$, and moreover, both R_2 and its limit are almost surely invertible, as there is no cointegration between $y_{c,t-1}$ and $y_{3,t-1}$ (see Lemma 3.1.1 in Chan and Wei (1988), and Sims et al. (1990)). Note that

$$(\hat{\Lambda} - \underline{\Lambda})D_T^{-1} - \langle \varepsilon_t, W_t\rangle D_T R^{-1} = \underbrace{\langle e_{1t}, W_t\rangle D_T R^{-1}}_{=:E_1} + \underbrace{\langle e_{1t}, W_t\rangle D_T(\hat{R}^{-1} - R^{-1})}_{=:E_2} + \underbrace{\langle \varepsilon_t, W_t\rangle D_T(\hat{R}^{-1} - R^{-1})}_{=:E_3}.$$

Here $\langle \varepsilon_t, W_t\rangle D_T R^{-1}$ has the limits stated in the lemma since:

$$N^{-1}\langle \varepsilon_t, y_{c,t-1}\rangle \Rightarrow g(1)\int_0^1 dBB_c', \quad N^{-2}\langle \varepsilon_t, y_{3,t-1}\rangle \Rightarrow g(1)\int_0^1 dBH_3',$$

$$\left[\begin{array}{cc} N^{-2}\langle y_{c,t-1}, y_{c,t-1}\rangle & N^{-3}\langle y_{c,t-1}, y_{3,t-1}\rangle \\ N^{-3}\langle y_{3,t-1}, y_{c,t-1}\rangle & N^{-4}\langle y_{3,t-1}, y_{3,t-1}\rangle \end{array}\right] \Rightarrow \left[\begin{array}{cc} \int_0^1 B_c B_c' & \int_0^1 B_c H_3' \\ \int_0^1 H_3 B_c' & \int_0^1 H_3 H_3' \end{array}\right].$$

The lemma therefore holds, if $E_1 = [o_P(h^{1/2}), o_P(1), o_P(1)]$, $E_2 = o_P(1)$, $E_3 = o_P(1)$ can be shown (where the blocks in E_1 correspond to the partitioning of W_t into stationary, I(1) and I(2) components). For this it is sufficient to show:

(I)　$\|\hat{R}^{-1} - R^{-1}\|_1 = O_P(h/N^{\frac{1}{2}})$

(II)　$\|\langle e_{1t}, W_t\rangle D_T\| = o_P(h^{1/2})$ where $N^{-1}\langle e_{1t}, y_{c,t-1}\rangle = o_P(1)$ and $N^{-2}\langle e_{1t}, y_{3,t-1}\rangle = o_P(1)$
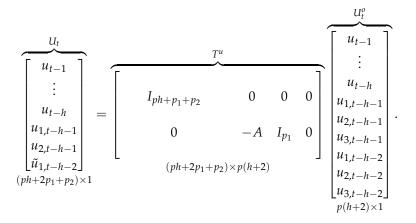
(III)　$\|\langle \varepsilon_t, W_t\rangle D_T\| = O_P(h^{1/2}).$

Here $\|.\|_1$ denotes the spectral norm of a matrix while $\|.\|$ denotes the Frobenius norm.
(I) To see $\|\hat{R}^{-1} - R^{-1}\|_1 = O_p(h/N^{\frac{1}{2}})$, according to Lewis and Reinsel (1985), it is sufficient to show $\|\hat{R} - R\|_1 = O_p(h/N^{\frac{1}{2}})$, $\|R^{-1}\|_1 = O_p(1)$. Note that

$$\hat{R} - R = \left[\begin{array}{ccc} N^{-1}\langle U_t, U_t\rangle - \Gamma_u & N^{-\frac{3}{2}}\langle U_t, y_{c,t-1}\rangle & N^{-\frac{5}{2}}\langle U_t, y_{3,t-1}\rangle \\ N^{-\frac{3}{2}}\langle y_{c,t-1}, U_t\rangle & 0 & 0 \\ N^{-\frac{5}{2}}\langle y_{3,t-1}, U_t\rangle & 0 & 0 \end{array}\right] =: \left[\begin{array}{ccc} \hat{Q} & \hat{P}_{12} & \hat{P}_{13} \\ \hat{P}_{21} & 0 & 0 \\ \hat{P}_{31} & 0 & 0 \end{array}\right],$$

then we have $\mathbb{E}\|\hat{R} - R\|_1^2 \leq \mathbb{E}\|\hat{R} - R\|^2 = \mathbb{E}\|\hat{Q}\|^2 + 2(\mathbb{E}\|\hat{P}_{12}\|^2 + \mathbb{E}\|\hat{P}_{13}\|^2).$

Now let $U_t^o := [u'_{t-1}, \ldots, u'_{t-h-2}]'$, then there exists a transformation T^u of full row rank, such that $U_t = T^u U_t^o$, where T^u is a $(ph + 2p_1 + p_2) \times p(h+2)$ matrix:

$$
\underbrace{\begin{bmatrix} u_{t-1} \\ \vdots \\ u_{t-h} \\ u_{1,t-h-1} \\ u_{2,t-h-1} \\ \tilde{u}_{1,t-h-2} \end{bmatrix}}_{\substack{U_t \\ (ph+2p_1+p_2)\times 1}} = \underbrace{\begin{bmatrix} I_{ph+p_1+p_2} & 0 & 0 & 0 \\ & & & \\ 0 & & -A & I_{p_1} & 0 \end{bmatrix}}_{\substack{T^u \\ (ph+2p_1+p_2)\times p(h+2)}} \underbrace{\begin{bmatrix} u_{t-1} \\ \vdots \\ u_{t-h} \\ u_{1,t-h-1} \\ u_{2,t-h-1} \\ u_{3,t-h-1} \\ u_{1,t-h-2} \\ u_{2,t-h-2} \\ u_{3,t-h-2} \end{bmatrix}}_{\substack{U_t^o \\ p(h+2)\times 1}}.
$$

Then, we have $\hat{Q} = T^u \hat{Q}^o T^{u'}$, where $\hat{Q}^o = \frac{1}{N}\langle U_t^o, U_t^o \rangle - \mathbb{E}[U_t^o U_t^{o'}]$; moreover, $\hat{P}_{1i} = T^u \hat{P}_{1i}^o$ for $i = 2, 3$, where $\hat{P}_{12}^o = N^{-\frac{3}{2}}\langle U_t^o, y_{c,t-1}\rangle, \hat{P}_{13}^o = N^{-\frac{5}{2}}\langle U_t^o, y_{3,t-1}\rangle$. Since $\|T^u\|_1 = O(1)$, $\hat{Q}$ and $\hat{P}_{1i}$ have the same rate of convergence as $\hat{Q}^o$ and $\hat{P}_{1i}^o$, respectively. From Saikkonen (1991) Lemma A.2. we know $\mathbb{E}\|\hat{Q}^o\|^2 = O(h^2/N)$ and $\mathbb{E}\|\hat{P}_{12}^o\|^2 = O(h/N)$ by direct calculation.

For $\hat{P}_{13}^o$ note that

$$
\mathbb{E}\|y_{3,t-1}\|^2 = \mathbb{E}\left\|\sum_{j=1}^{t-1}\sum_{i=1}^{j} u_{3,i}\right\|^2 = \mathbb{E}\left\|\sum_{i=1}^{t-1} i\, u_{3,t-1-i}\right\|^2 = O(t^3).
$$

Then analogous calculation as for $\hat{P}_{12}^o$ show that $\mathbb{E}\|\hat{P}_{13}^o\|^2 = O(h/N)$. Concluding we obtain $\mathbb{E}\|\hat{R} - R\|_1^2 = O(h^2/N)$ such that $\|\hat{R} - R\|_1 = O_P(h/N^{\frac{1}{2}})$.

To show $\|R^{-1}\|_1 = O_P(1)$ note that $R^{-1} = \text{diag}\{\Gamma_u^{-1}, R_2^{-1}\}$ where $\|\Gamma_u^{-1}\|_1 = O(1)$ (see Lewis and Reinsel (1985), p. 397) and $\|R_2^{-1}\|_1 = O_P(1)$, since R_2 is a.s. invertible and converges in distribution to an almost surely nonsingular random matrix.

(II) With respect to $\|\langle e_{1t}, W_t\rangle D_T\| = o_P(h^{1/2})$ note that

$$
\|\langle e_{1t}, W_t\rangle D_T\| \leq \left\|N^{-\frac{1}{2}}\langle e_{1t}, U_t\rangle\right\| + \left\|N^{-1}\langle e_{1t}, y_{c,t-1}\rangle\right\| + \left\|N^{-2}\langle e_{1t}, y_{3,t-1}\rangle\right\|.
$$

From Saikkonen (1991) Lemma A.5 we have $\left\|N^{-\frac{1}{2}}\langle e_{1t}, U_t\rangle\right\| = o_P(h^{\frac{1}{2}})$, and $\left\|N^{-1}\langle e_{1t}, y_{c,t-1}\rangle\right\| = o_P(1)$. Then $\mathbb{E}\|y_{3,t-1}\|^2 = O(t^3)$ and $\mathbb{E}\|e_{1t}\|^2 = o(N^{-1})$ imply

$$
\mathbb{E}\left\|N^{-2}\langle e_{1t}, y_{3,t-1}\rangle\right\| \leq N^{-2}\sum_{t=h+3}^{T}\left(\mathbb{E}\|e_{1t}\|^2\mathbb{E}\|y_{3,t-1}\|^2\right)^{\frac{1}{2}} = o(N^{-2}NN^{-1/2}N^{3/2}) = o(1).
$$

(III) To show $\|\langle \varepsilon_t, W_t\rangle D_T\| = O_P(h^{1/2})$ note that $N^{-\frac{1}{2}}\langle \varepsilon_t, U_t\rangle = O_P(h^{1/2}), N^{-1}\langle \varepsilon_t, y_{c,t-1}\rangle = O_P(1)$ according to (A.7) of Saikkonen (1992). Moreover $N^{-2}\sum_{t=h+3}^{T}\varepsilon_t y'_{3,t-1} \Rightarrow g(1)\int_0^1 dBH_3'$ implies $N^{-2}\langle \varepsilon_t, y_{3,t-1}\rangle = O_P(1)$. $\square$

Note that for the lemma to hold we only need $h^3/N \to 0$ and $N^{1/2}\sum_{j=h+1}^{\infty}\|\Pi_j\| = o(1)$.

Appendix B. Proof of Theorem 1

Appendix B.1. (A) Consistency

(i) Lemma A1 implies $\hat{\Phi}_2 \to 0 = \Phi_2, \hat{\Phi}_3 \to 0 = \Phi_3$. Furthermore, the reparameterization implies $\Phi_1 = \sum_{j=1}^{h+2}\Xi_{j1}$ and thus $\hat{\Phi}_1 = \sum_{j=1}^{h+2}\hat{\Xi}_{j,1}$ leading to

$$\|\hat{\Phi}_1 - \underline{\Phi}_1\| \leq \|\sum_{j=1}^{h+2} \hat{\Xi}_{j,1} - \sum_{j=1}^{h+2} \underline{\Xi}_{j,1}\|$$

$$\leq \sum_{j=1}^{h+2} \|\hat{\Xi}_{j,1} - \underline{\Xi}_{j,1}\| \leq \|\hat{\Xi} - \underline{\Xi}\| = O_P(h^{3/2}/N^{1/2})$$

where the last inequality holds due to $\langle \varepsilon_t, u_{t-j} \rangle = O_P(N^{1/2})$ in combination with Lemma A1.

(ii) Note that

$$\hat{\Sigma}_\epsilon = N^{-1}\langle \Delta^2 y_t - \hat{\Lambda} W_t, \Delta^2 y_t - \hat{\Lambda} W_t \rangle = N^{-1}\langle e_t + (\underline{\Lambda} - \hat{\Lambda})W_t, e_t + (\underline{\Lambda} - \hat{\Lambda})W_t \rangle.$$

Now

$$\langle (\underline{\Lambda} - \hat{\Lambda})W_t, (\underline{\Lambda} - \hat{\Lambda})W_t \rangle = (\underline{\Lambda} - \hat{\Lambda})D_T^{-1}D_T\langle W_t, W_t \rangle D_T D_T^{-1}(\underline{\Lambda} - \hat{\Lambda})'$$

where $\hat{R} = D_T \langle W_t, W_t \rangle D_T$ such that $\|\hat{R}\|_1 = O_P(1)$ and $\|(\underline{\Lambda} - \hat{\Lambda})D_T^{-1}\| = O_P(h^{1/2})$. Consequently

$$N^{-1}\langle (\underline{\Lambda} - \hat{\Lambda})W_t, (\underline{\Lambda} - \hat{\Lambda})W_t \rangle = O_P(h/N) \to 0.$$

Next, from the definition of e_t, we can show that

$$N^{-1}\langle \varepsilon_t + e_{1t}, \varepsilon_t + e_{1t} \rangle = N^{-1}\langle \varepsilon_t, \varepsilon_t \rangle + o_P(1) = \Sigma_\epsilon + o_P(1),$$

where the last equality follows the law of large numbers and the first equality is implied by the fact that $\|e_{1t}\|^2 = o_P(T^{-1})$ and $\|\varepsilon_t\|^2 = O_P(1)$.

(iii) From (i) and (ii), $\hat{\Omega}_{1.c} = (\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1)^{-1} = (\Phi_1' \Sigma_\epsilon^{-1} \Phi_1)^{-1} + o_P(1) = \Omega_{1.c} + o_P(1)$ directly follows.

(iv) With respect to $\hat{\Psi}$ recall that

$$\Psi_1 = -\sum_{j=2}^{h+2}(j-1)\Xi_{j,1}, \quad \Psi_2 = \sum_{j=1}^{h+1}\Xi_{j,2}.$$

Then Lemma A1 shows that each entry of $\hat{\Xi} - \underline{\Xi}$ is of order $O_P(h^{1/2}/N^{1/2})$. Then

$$\|\hat{\Psi}_1 - \Psi_1\| \leq \sum_{j=2}^{h+2}(j-1)\|\hat{\Xi}_{j,1} - \Xi_{j,1}\| = O_P(\sum_{j=2}^{h+2}(j-1)h^{1/2}/N^{1/2}) = O_P(h^{5/2}/N^{1/2})$$

which converges to zero for $h^5/T \to 0$. Similarly $\hat{\Psi}_2 - \Psi_2 = O_P(h^{3/2}/N^{1/2})$.

For $\hat{\Psi}_3$ note that $\Theta = \Phi_1 A + \Psi_3$. Thus $\hat{\Psi}_3 = \hat{\Theta} - \hat{\Phi}_1 A$ such that $\hat{\Psi}_3 \to \Psi_3$ from (i) and Lemma A1.

(v) is contained in Lemma A1.

(vi) From (6), and the definition $\hat{\Omega}_{1.c} = (\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1)^{-1}$, we have

$$\hat{A} - A = -(\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1)^{-1} \hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Psi}_3 - A$$

$$= -\hat{\Omega}_{1.c} \hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Psi}_3 - \hat{\Omega}_{1.c}\hat{\Omega}_{1.c}^{-1} A = -\hat{\Omega}_{1.c}\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Psi}_3 - \hat{\Omega}_{1.c}\,\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1} \hat{\Phi}_1\, A$$

$$= -\hat{\Omega}_{1.c}\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1}(\hat{\Psi}_3 + \hat{\Phi}_1 A) = -\hat{\Omega}_{1.c}\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1}\hat{\Theta}.$$

Then (i-iii, v) show the result.

Appendix B.2. (B) Asymptotic Distribution of Coefficients to Nonstationary Regressors

(i) The distribution of the coefficients due to the nonstationary components is contained in Lemma A1.

(ii) With respect to the cointegrating relation note that from the proof of Theorem 1 we have

$$N(\hat{A} - A) = -N\hat{\Omega}_{1.c}\hat{\Phi}_1' \hat{\Sigma}_\epsilon^{-1}\hat{\Theta} = -\Omega_{1.c}\Phi_1' \Sigma_\epsilon^{-1} \cdot N\hat{\Theta} + o_P(1).$$

Note that $N\hat{\Theta} = [N\hat{\Phi}_2, N\hat{\Theta}, N^2\hat{\Phi}_3]\eta$, where $\eta = [0_{p_3 \times p_2}, I_{p_3}, 0_{p_3 \times p_3}]'$. Then by Lemma A1, we have

$$N(\hat{A} - A) \Rightarrow -\Omega_{1.c}\Phi_1'\Sigma_\epsilon^{-1} \cdot g(1) \int_0^1 dBF' \left(\int_0^1 FF'\right)^{-1} \eta = -\Omega_{1.c}\Phi_1'\Sigma_\epsilon^{-1} \cdot g(1) \int_0^1 dBL' \left(\int_0^1 LL'\right)^{-1}.$$

Note that $\Phi_1 = g(1)\alpha$, and by definition $\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{1c} \\ \Omega_{c1} & \Omega_{cc} \end{bmatrix} = g(1)^{-1}\Sigma_\epsilon g(1)'^{-1}$, we have

$$\begin{aligned} -\Omega_{1.c}\Phi_1'\Sigma_\epsilon^{-1}g(1)B &= -\Omega_{1.c}\alpha'g(1)'\Sigma_\epsilon^{-1}g(1)\,B = -\Omega_{1.c}\,[I_{p_1}\ 0]\,\Omega^{-1}B \\ &= \Omega_{1.c}\,[(\Omega^{-1})_{11}\ (\Omega^{-1})_{1c}]\,B = \Omega_{1.c}\,[\,\Omega_{1.c}^{-1}\quad -\Omega_{1.c}^{-1}\Omega_{1c}\Omega_{cc}^{-1}\,]\,B \\ &= [I_{p_1}\quad -\Omega_{1c}\Omega_{cc}^{-1}]\begin{bmatrix} B_1 \\ B_c \end{bmatrix} = B_1 - \Omega_{1c}\Omega_{cc}^{-1}B_c = B_{1.c}. \end{aligned}$$

Therefore, we have

$$N(\hat{A} - A) \Rightarrow \int_0^1 dB_{1.c}L' \left(\int_0^1 LL'\right)^{-1}.$$

Appendix B.3. (C) Asymptotic Distribution of Coefficients to Stationary Regressors

Since the regressor vector U_t is stationary, the asymptotic distribution of $N^{1/2}L_h'vec(\hat{\Xi} - \Xi)$ follows from Lewis and Reinsel (1985) in combination with uniform boundedness of the maximal and the minimal eigenvalue of $\Gamma_u = \mathbb{E}U_tU_t'$, see above. Analogously the result for the coefficients corresponding to the regressor vector X_t are shown as $X_t = T_hU_t$ for nonsingular matrix T_h.

Appendix B.4. (D) Asymptotic Distribution of Wald Type Tests

For the Wald test in addition to (C) note that the variance Γ_{ECM} is replaced by an estimate $\hat{\Gamma}_{ECM}$. For

$$L_h'(\Gamma_{ECM}^{-1} \otimes \Sigma_\epsilon)L_h - L_h'(\hat{\Gamma}_{ECM}^{-1} \otimes \hat{\Sigma}_\epsilon)L_h$$

note that $\hat{\Sigma}_\epsilon - \Sigma_\epsilon = o_P(1)$ due to (A) (ii). The regressor vectors $\tilde{X}_t$ and X_t differ only in the first block where $y_{1,t-1} = u_{1,t-1} + A\Delta y_{3,t-1}$ replaces $u_{1,t-1}$. Regressing out $\Delta y_{3,t-1}$ eliminates this difference. Then $\|\hat{\Gamma}_{ECM} - \Gamma_{ECM}\|_1 = O_P(h/N^{1/2})$ according to (Saikkonen and Lütkepohl 1996, p. 835, l. 3). There also invertibility of Γ_{ECM} is shown. Using Lemma A.2 of Saikkonen and Lütkepohl (1996) this implies $\|\hat{\Gamma}_{ECM}^{-1} - \Gamma_{ECM}^{-1}\|_1 = O_P(h/N^{1/2})$.

The rest then follows as the proof of Theorem 4 in Saikkonen and Lütkepohl (1996).

Appendix C. Proof of Theorem 2

Consistency follows directly from Theorem 1 as the general representation can be transformed into a triangular representation using the matrix $\mathcal{B} = [\beta, \beta_1, \beta_2]$, see (4).

With respect to the asymptotic distribution following the proof of Theorem 1 there exists a nonsingular transformation matrix S_h such that $W_t = S_hZ_{t,h}$. From $\|\hat{R}^{-1} - R^{-1}\| = O_P(h/N^{1/2})$ it follows that

$$(N^{-1}\langle W_t, W_t\rangle)^{-1} = \begin{bmatrix} (\Gamma_u)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + o_P(h/N^{1/2}).$$

Therefore it follows that the blocks corresponding to the nonstationary regressors do not contribute to the asymptotic distribution. Then standard arguments for the stationary part of the regressor vector can be used.

Appendix D. Proofs for Theorem 3

The proof combines the ideas of Saikkonen and Luukkonen (1997) (in the following S&L) with the asymptotics of 2SI2 of Paruolo (2000) (in the following P). In the proof we will work without restriction of generality with the triangular representation.

The key to the asymptotic properties of the estimators obtained from the 2SI2 algorithm lies in the results of P Lemma A.4 and Lemma A.5 in the appendix. These lemmas deal with the limits of various moment matrices of the form $N^{-a}\langle R_{it}, R_{jt}\rangle$ corrected for the stationary components $\Delta^2 y_{t-j}, j = 1, \ldots, h-2$. The correction involves a regressor vector growing in dimension with sample size. This is dealt with in S&L.

In this respect let $S_t = [\Delta^2 y'_{t-1}, \ldots, \Delta^2 y'_{t-h+2}]'$ which according to (A4) is a linear function of U_t such that $S_t = \mathcal{T}_s U_t$. The definition of $\hat{U}_t$ implies $\hat{Q} = N^{-1}\langle U_t, U_t\rangle - \mathbb{E}U_t U'_t = O_P(h/N^{1/2})$. On p. 543 in P the matrices $\Sigma_{ij}, i, j \in \{Y, U, 0\}$ are defined as limits of second moment matrices. Here $'U'$ refers to $\beta'_1 \Delta y_{t-1} = u_{2,t-1}$ in the triangular representation, $'Y'$ refers to $\beta' y_{t-1} + \delta\beta'_2 \Delta y_{t-1} = y_{1,t-1} - A\Delta y_{3,t-1} = u_{1,t-1}$ and $'0'$ refers to $\Delta^2 y_t$. These are all stationary processes and linear functions of u_t, u_{t-1}, u_{t-2}. Additional to S_t also $\beta' \Delta y_{t-1} = \Delta u_{1,t-1} + A u_{3,t-1}$ is corrected for in the second stage.

The arguments on p. 114 and 115 of S&L deal with terms of the form

$$N^{-1}\langle u_{1,t-1}, u_{1,t-1}\rangle - N^{-1}\langle u_{1,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, u_{1,t-1}\rangle.$$

Analogous arguments to S&L(A.12) show that this equals (up to terms of order $o_P(1)$)

$$C_{11} = \mathbb{E}u_{1,t-1}u'_{1,t-1} - \mathbb{E}u_{1,t-1}S'_t(\mathbb{E}S_t S'_t)^{-1}\mathbb{E}S_t u'_{1,t-1}.$$

S&L state that this is bounded from above and bounded away from zero. The second claim actually is wrong. If $(u_{1,t})_{t\in\mathbb{Z}}$ is univariate white noise with unit variance then $C_{11} = \frac{1}{h}$ is achieved by predicting $u_{1,t-1}$ by

$$\sum_{j=1}^h \frac{h-j}{h}\Delta u_{1,t-j} = u_{1,t-1} - \frac{1}{h}\sum_{j=1}^h u_{1,t-j}$$

including integration of the regressors in the form of the summation. This does not change the remaining arguments in S&L, it only implies that the separation of the eigenvalues corresponding to the stationary regressors and the ones corresponding to the non-stationary ones is weaker.

In the current case one can show that for

$$N^{-1}\langle u_{1,t-1}, u_{1,t-1}\rangle - N^{-1}\langle u_{1,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, u_{1,t-1}\rangle$$

where S_t contains $\Delta u_{1,t-1}$ and $\Delta^2 u_{1,t-j}, j = 1, \ldots, h$ for the corresponding limit C_{11} the lower bound $h C_{11} \geq cI$ holds for some $0 < c$. The order of the lower bound is achieved by including a double integration of the regressors. For

$$N^{-1}\langle \Delta u_{1,t-1}, \Delta u_{t,t-1}\rangle - N^{-1}\langle \Delta u_{1,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, \Delta u_{1,t-1}\rangle = C_{\Delta\Delta} + o_p(1)$$

we have $h^3 C_{\Delta\Delta} \geq cI$. Here the arguments from above can be applied to the process $(\Delta u_t)_{t\in\mathbb{Z}}$. For a differenced process the smallest eigenvalue of the matrix

$$\mathbb{E}\delta U_t \delta U'_t, \quad \delta U'_t = [\Delta u'_t, \Delta u'_{t-1}, \ldots, \Delta u'_{t-h}]$$

is of order h^{-2}, compare Theorem 2 of Palma and Bondon (2003).

Since $N^{-1}\langle S_t, y_{c,t-1}\rangle = O_P(h^{1/2})$ and $N^{-2}\langle S_t, y_{3,t-1}\rangle = O_P(h^{1/2})$ it follows that

$$N^{-1}(\langle u_{1,t-1}, y_{c,t-1}\rangle - \langle u_{1,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, y_{c,t-1}\rangle) = O_P(h^{1/2}),$$
$$N^{-2}(\langle y_{c,t-1}, y_{c,t-1}\rangle - \langle y_{c,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, y_{c,t-1}\rangle) = N^{-2}\langle y_{c,t-1}, y_{c,t-1}\rangle + o_P((h/N)^{1/2})$$

as well as

$$N^{-2}(\langle u_{1,t-1}, y_{3,t-1}\rangle - \langle u_{1,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, y_{3,t-1}\rangle) = O_P(h^{1/2}),$$
$$N^{-3}(\langle y_{c,t-1}, y_{3,t-1}\rangle - \langle y_{c,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, y_{3,t-1}\rangle) = N^{-3}\langle y_{c,t-1}, y_{3,t-1}\rangle + o_P((h/N)^{1/2}),$$
$$N^{-4}(\langle y_{3,t-1}, y_{3,t-1}\rangle - \langle y_{3,t-1}, S_t\rangle\langle S_t, S_t\rangle^{-1}\langle S_t, y_{3,t-1}\rangle) = N^{-4}\langle y_{3,t-1}, y_{3,t-1}\rangle + o_P((h/N)^{1/2}).$$

Therefore the limits of the moment matrices M_{ij} are not affected by the correction using stationary terms even if $h \to \infty$ except for the terms involving the orders $O_P(h^{1/2})$. For all stationary terms we find convergence to the corresponding limits denoted Σ_{ij} in P.

The first step in the 2SI2 procedure then uses RRR in the equation

$$\Delta^2 y_t = \Psi\Delta y_{t-1} + \alpha\beta' y_{t-1} + \underline{\Pi}S_t + e_t.$$

Then R_{0t} denotes $\Delta^2 y_t$ corrected for S_t, $R_{1,t}$ denotes Δy_{t-1} corrected for S_t and $R_{2,t}$ denotes y_{t-1} corrected for S_t. Lemma A.4 of P derives the limits of different directions of $M_{ij.k}$ defined as

$$M_{ij.k} = M_{ij} - M_{ik}M_{kk}^{-1}M_{kj}, \quad M_{ij} = N^{-1}\langle R_{i,t}, R_{j,t}\rangle$$

where $i, j \in \{0, 1, 2, \varepsilon, \beta\}$. Here $R_{\varepsilon,t}$ equals e_t correct for S_t and $R_{\beta,t} = \beta'R_{1,t}$. Further P uses the notation $A_T = [\bar{\beta}_1, T^{-1}\bar{\beta}_2]$ and $\bar{\beta}_{2,T} = \bar{\beta}_2$. Here and below we assume without restriction of generality that $[\beta, \beta_1, \beta_2]$ is an orthonormal matrix. Consequently $\bar{\beta} = \beta, \bar{\beta}_1 = \beta_1, \bar{\beta}_2 = \beta_2$. Then the results above imply all results of Lemma A.4. of P except that now $A_T'M_{20.1} = O_P(h^{1/2})$.

In particular we obtain the following limits:

$$A_T'M_{2\varepsilon.1} \xrightarrow{d} \int_0^1 F_+(dW)' \quad, \quad A_T'M_{22.1}A_T \xrightarrow{d} \int_0^1 F_+F_+',$$
$$\beta_2'M_{1\varepsilon.\beta} \xrightarrow{d} \int_0^1 B_3(dW)' \quad, \quad T^{-1}\beta_2'M_{11.\beta}\beta_2 \xrightarrow{d} \int_0^1 B_3B_3',$$
$$\beta_2'M_{1\varepsilon.b} \xrightarrow{d} \int_0^1 L(dW)' \quad, \quad T^{-1}\beta_2'M_{11.b}\beta_2 \xrightarrow{d} \int_0^1 LL'.$$

Here $W = g(1)B$ denotes the Brownian motion corresponding to $(\varepsilon_t)_{t\in\mathbb{Z}}$, F_+ denotes the Brownian motion corresponding to R_{2t} (equaling y_{t-1} corrected for S_t) corrected for R_{1t} (Δy_{t-1} whose only nonstationary component equals $\Delta y_{3,t-1}$ with corresponding Brownian motion B_3). Thus we obtain the following definitions (where L is as in Theorem 1):

$$F_a(u) = \begin{bmatrix} B_2(u) \\ \int_0^u B_3(v)dv \end{bmatrix}, \quad F_+(u) = F_a(u) - \int_0^1 F_aB_3'(\int_0^1 B_3B_3')^{-1}B_3(u),$$
$$L(u) = B_3(u) - \int_0^1 B_3F_a'(\int_0^1 F_aF_a')^{-1}F_a(u).$$

The above arguments show that in the current setting $U_{t-1} = u_{2,t-1}$ and $Y_{t-1} = u_{1,t-1}$ are contained in the space spanned by S_t for $h \to \infty$. Therefore $\Sigma_{ij} = 0$ for $i, j \in \{U, Y\}$. The subscript 'b' refers to correcting for $\beta'_{\perp}R_{2t}$ used in the second stage of 2SI2.

Let $\tilde{\Sigma}_{YY}$ denote the limit of $h\langle Y_{t-1}, Y_{t-1}\rangle$ and analogously define $\tilde{\Sigma}_{YU}, \tilde{\Sigma}_{UU}, \tilde{\Sigma}_{0Y}$ and $\tilde{\Sigma}_{0U}$. For the latter two note that $\tilde{\Sigma}_{0Y}$ denotes the limit of

$$h\langle\Delta^2 y_t, Y_{t-1}\rangle = h\alpha\langle Y_{t-1}, Y_{t-1}\rangle + h\langle\zeta U_{t-1}, Y_{t-1}\rangle + h\langle\zeta_2\beta'\Delta y_{t-1}, Y_{t-1}\rangle + h\underline{\Pi}\langle S_t, Y_{t-1}\rangle + h\langle e_t, Y_{t-1}\rangle$$

corrected for S_t and $\beta' \Delta y_{t-1}$. Since Y_{t-1} is stationary the last term is of order $O_P((h^3/N)^{1/2}) = o_P(1)$. Therefore it follows that $\tilde{\Sigma}_{0Y} = \alpha \tilde{\Sigma}_{YY} + \zeta \tilde{\Sigma}_{UY}$. Then the results of Lemma A.5 of P hold where in (A.11) and (A.14) Σ_{ij} can be replaced by $\tilde{\Sigma}_{ij}$.

The asymptotic analysis below will heavily use the Johansen approach of investigating the solutions to eigenvalue problems in order to maximize the pseudo-likelihood corresponding to the reduced rank regression problem. In order to use the corresponding local analysis one has to first clarify consistency for the various estimators as well as rates of convergence.

The main tool in this respect is Theorem A.1 of Johansen (1997) which establishes in the I(2) setting for the regression $y_t = \theta' Z_t + \varepsilon_t$ (Z_t being composed of stationary, I(1) and I(2) components) where $D_T \langle Z_t, Z_t \rangle D_T = O_P(1)$ and $D_T \langle Z_t, \varepsilon_t \rangle = o_P(1)$ that $D_T^{-1}(\hat{\theta} - \theta) = o_P(1)$ where $\hat{\theta}$ denotes the pseudo likelihood estimator over some closed parameter set Θ.

It is straightforward to see that analogous results hold in the present setting when first concentrating out the stationary components: Consider $y_t = \theta_1' z_t + \theta_2' Z_t + e_t$. Then $\hat{\theta}_2(\theta_1)$ is obtained from the concentration step and the pseudo likelihood involves $\langle R_{t,y} - \theta_1' R_{t,z}, R_{t,y} - \theta_1' R_{t,z} \rangle$ where again the processes $R_{t,y}$ and $R_{t,z}$ denote the processes y_t and z_t with the corresponding stationary regressors Z_t regressed out. These concentrated quantities now can be used in the proof of Theorem A.1 of Johansen (1997) essentially without changes to show consistency for $\hat{\theta}_1$. Consistency of $\hat{\theta}_2(\hat{\theta}_1)$ then follows from the unrestricted estimation as contained in Theorem 2. As shown above the rates of convergence as well as the limits are unchanged for the coefficients corresponding to the non-stationary components of the regressors for the long VAR case compared to the finite VAR case.

Note that these results hold for general closed parameter space Θ, thus including the unrestricted as well as the rank-reduced problem. This shows that we can always reduce the asymptotic analysis of the eigenvalue problems to a neighborhood of the true value as is done in P.

The first step in the proof of Theorem 4.1. of P consists in the investigation of the solutions to the

equation ($\tilde{\beta} = \beta H + \beta_1 H_1 + \beta_2 H_2$, letting $B_T' = \begin{bmatrix} \beta' \\ T^{-1/2} \beta_1' \\ T^{-3/2} \beta_2' \end{bmatrix}$)

$$B_T' M_{22.1} B_T \begin{bmatrix} H \\ T^{1/2} H_1 \\ T^{3/2} H_2 \end{bmatrix} \Lambda = B_T' M_{20.1} M_{00.1}^{-1} M_{02.1} B_T \begin{bmatrix} H \\ T^{1/2} H_1 \\ T^{3/2} H_2 \end{bmatrix}. \tag{A6}$$

Now Lemma A.4 implies that the matrix $B_T' M_{22.1} B_T$ on the left hand side converges to $\mathrm{diag}(\Sigma_{YY.U}, \int_0^1 F_+ F_+')$. $B_T' M_{20.1} = \begin{bmatrix} \Sigma_{Y0.U} \\ 0 \end{bmatrix} + O_P(h^{1/2} T^{-1/2})$, $M_{00.1} = \Sigma_{00.U} + O_P(T^{-1/2})$. Multiplying the equation by h^2 we obtain the limiting eigenvalue problem

$$\begin{bmatrix} \tilde{\Sigma}_{YY.U} & O_P(T^{-1/2} h^{3/2}) \\ O_P(T^{-1/2} h^{3/2}) & h \int_0^1 F_+ F_+' \end{bmatrix} \begin{bmatrix} H \\ T^{1/2} H_1 \\ T^{3/2} H_2 \end{bmatrix} h\Lambda = \begin{bmatrix} \tilde{\Sigma}_{Y0.U} \Sigma_{00.U}^{-1} \tilde{\Sigma}_{0Y.U} & O_P(T^{-1/2} h^{5/2}) \\ O_P(T^{-1/2} h^{5/2}) & O_P(T^{-1} h^3) \end{bmatrix} \begin{bmatrix} H \\ T^{1/2} H_1 \\ T^{3/2} H_2 \end{bmatrix}.$$

equation

Therefore asymptotically the first $p - r$ eigenvalues of $h\Lambda$ are positive, the remaining ones tending to zero. Likewise the eigenvectors converge at the same speed as the matrices. Thus $H_1 = O_P(h^{5/2}/T), H_2 = O_P(h^{5/2}/T^2)$ from which

$$\beta' M_{22.1} \beta H \Lambda H^{-1} = \beta' M_{20.1} M_{00.1}^{-1} M_{02.1} \beta + O_P(h^4/T)$$

and thus using (A.11)

$$H \Lambda H^{-1} = \tilde{\Sigma}_{YY.U}^{-1} \tilde{\Sigma}_{Y0.U} \Sigma_{00.U}^{-1} \tilde{\Sigma}_{0Y.U} / h + O_P(hT^{-1/2}) = \alpha' \Sigma_{00.U}^{-1} \alpha \tilde{\Sigma}_{YY.U} / h + O_P(hT^{-1/2})$$

follows. Then as in P we have[4]

$$M_{22.1}\underline{\tilde{\beta}} = M_{20.1}(\Sigma_{00.U}^{-1}\tilde{\Sigma}_{0Y.U}(hH\Lambda H^{-1})^{-1} + O_P(hT^{-1/2})) = M_{22.1}\beta + M_{2\varepsilon.1}\Sigma_{\epsilon}^{-1}\alpha(\alpha'\Sigma_{\epsilon}^{-1}\alpha)^{-1} + a_1$$

where $a_1 = M_{20.1}O_P(h^2T^{-1/2}) = o_P(1)$ and $\tilde{\underline{\beta}} = \tilde{\beta}H^{-1}$. Then the remaining arguments on p. 546 of P show that the asymptotic distribution of $(T\beta_1, T^2\beta_2)'(\tilde{\underline{\beta}} - \beta)$ is identical for the long VAR case as in the finite VAR case.

From these arguments the distribution of the likelihood ratio test of H_r versus H_p can be shown: Define $S_1(\lambda) := \lambda M_{22.1} - M_{20.1}M_{00.1}^{-1}M_{02.1}$, $A_T := (\beta_1, T^{-1}\beta_2)$ and $\tilde{B}_T := (\beta, A_T) = (\beta, \beta_1, T^{-1}\beta_2)$. Note that $\tilde{B}_T$ is of full rank, (11) is equivalent to $|\tilde{B}_T'S_1(\lambda)\tilde{B}_T| = 0$; that is,

$$\left| \begin{pmatrix} \beta' \\ \beta_1' \\ T^{-1}\beta_2' \end{pmatrix} S_1(\lambda)(\beta, \beta_1, T^{-1}\beta_2) \right| = |\beta'S_1(\lambda)\beta| \cdot \left| A_T'\left(S_1(\lambda) - S_1(\lambda)\beta\,(\beta'S_1(\lambda)\beta)^{-1}\,\beta'S_1(\lambda)\right)A_T \right| = 0. \qquad \text{(A7)}$$

Let $\delta_1 = T\lambda$, so that for every δ_1 we have that $\lambda \to 0$, as $T \to \infty$. By the above arguments we have that

$$h^2\left|\beta'S_1(\lambda)\beta\right| = \left| \delta_1\frac{h^2}{T}\beta'M_{22.1}\beta - h^2\beta'M_{20.1}M_{00.1}^{-1}M_{02.1}\beta \right| \xrightarrow{p} \left| -\tilde{\Sigma}_{Y0.U}\Sigma_{00.U}^{-1}\tilde{\Sigma}_{0Y.U} \right| \neq 0,$$

which has no zero root. Moreover, we have

$$hA_T'S_1(\lambda)\beta = h\lambda A_T'M_{22.1}\beta - hA_T'M_{20.1}M_{00.1}^{-1}M_{02.1}\beta = -A_T'M_{20.1}\Sigma_{00.U}^{-1}\tilde{\Sigma}_{0Y.U} + o_P(1),$$

which yields that

$$\left| A_T'\left(S_1(\lambda) - S_1(\lambda)\beta\,(\beta'S_1(\lambda)\beta)^{-1}\,\beta'S_1(\lambda)\right)A_T \right|$$
$$= \left| \left(\delta_1\frac{1}{T}A_T'M_{22.1}A_T - A_T'M_{20.1}M_{00.1}^{-1}M_{02.1}A_T\right) - A_T'S_1(\lambda)\beta\,(\beta'S_1(\lambda)\beta)^{-1}\,\beta'S_1(\lambda)A_T \right|$$
$$= \left| \left(\delta_1\frac{1}{T}A_T'M_{22.1}A_T\right) - A_T'M_{20.1}\left(M_{00.1}^{-1} - \Sigma_{00.U}^{-1}\tilde{\Sigma}_{0Y.U}\left(\tilde{\Sigma}_{Y0.U}\Sigma_{00.U}^{-1}\tilde{\Sigma}_{0Y.U}\right)^{-1}\tilde{\Sigma}_{Y0.U}\Sigma_{00.U}^{-1} + o_P(1)\right)M_{02.1}A_T \right|$$
$$= \left| \left(\delta_1\frac{1}{T}A_T'M_{22.1}A_T\right) - A_T'M_{20.1}\left(\Sigma_{00.U}^{-1} - \Sigma_{00.U}^{-1}\alpha\left(\alpha\Sigma_{\epsilon}^{-1}\alpha\right)^{-1}\alpha'\Sigma_{00.U}^{-1} + o_P(1)\right)M_{02.1}A_T \right|$$
$$\xrightarrow{d} \left| \delta_1\int_0^1 F_{\dagger}F_{\dagger}' - \int_0^1 F_{\dagger}dW'\alpha_{\perp}(\alpha_{\perp}'\Sigma_{\epsilon}\alpha_{\perp})^{-1}\alpha_{\perp}'\int_0^1 dWF_{\dagger}' \right| = \left| \delta_1\int_0^1 F_{\dagger}F_{\dagger}' - (\int_0^1 F_{\dagger}dW_{\dagger}')(\int_0^1 dW_{\dagger}F_{\dagger}') \right|$$

where $W_{\dagger} = (\alpha_{\perp}'\Sigma_{\epsilon}\alpha_{\perp})^{-1/2}\alpha_{\perp}'W$. Thus, the smallest $(p-r)$ solutions of (11) converge in distribution to the solutions of $\left| \delta_1\int_0^1 F_{\dagger}F_{\dagger}' - (\int_0^1 F_{\dagger}dW_{\dagger}')(\int_0^1 dW_{\dagger}F_{\dagger}') \right| = 0$, which implies that the test statistic Q_r has the following limiting distribution,

$$Q_r = \sum_{i=r+1}^{p} \delta_{1,i} + o_P(1) \xrightarrow{d} tr\left((\int_0^1 dW_{\dagger}F_{\dagger}')\left(\int_0^1 F_{\dagger}F_{\dagger}'\right)^{-1}(\int_0^1 F_{\dagger}dW_{\dagger}') \right).$$

For the second stage the arguments are very similar. The eigenvalue problem solved here is the following:

$$\overline{\tilde{\beta}}_{\perp}'M_{11.\tilde{\beta}}\overline{\tilde{\beta}}_{\perp}\tilde{\eta}Y = \overline{\tilde{\beta}}_{\perp}'M_{1\tilde{\alpha}_{\perp}.\tilde{\beta}}M_{\tilde{\alpha}_{\perp}\tilde{\alpha}_{\perp}.\tilde{\beta}}^{-1}M_{\tilde{\alpha}_{\perp}1.\tilde{\beta}}\overline{\tilde{\beta}}_{\perp}\tilde{\eta}.$$

---

4   Contrary to the usual Johansen notation we use Σ_{ϵ} as the noise covariance and Ω as the variance of the Brownian motion corresponding to $(u_t)_{t\in\mathbb{Z}}$. Thus some of the formulas in this part show 'unusual' form.

This formula uses $\tilde{\alpha}_{\perp}$, the ortho-complement of

$$\tilde{\alpha} = M_{02.1}\tilde{\beta}(\tilde{\beta}'M_{22.1}\tilde{\beta})^{-1}$$

From the above results noting that $h\tilde{\beta}'M_{22.1}\tilde{\beta} \to \tilde{\Sigma}_{YY.U}$ and $hM_{02.1}\tilde{\beta} \to \alpha\tilde{\Sigma}_{YY.U}$ according to Lemma A.4 we have $\tilde{\alpha} \to \alpha$. Considering the order of convergence we obtain $\tilde{\alpha} - \alpha = O_P(hT^{-1/2})$. As in P this implies $\tilde{\alpha}_{\perp} - \alpha_{\perp} = O_P(hT^{-1/2})$. Using $\tilde{\beta} - \beta = O_P(h^{5/2}/T)$ from stage 1 one observes that in the eigenvalue problem estimates can be replaced by true quantities introducing an error of order $o_P(hT^{-1/2})$:

$$\overline{\beta}'_{\perp}M_{11.\beta}\tilde{\beta}_1 Y = \overline{\beta}'_{\perp}M_{1\alpha_{\perp}.\beta}M^{-1}_{\alpha_{\perp}\alpha_{\perp}.\beta}M_{\alpha_{\perp}1.\beta}\tilde{\beta}_1 + o_P(hT^{-1/2}).$$

Then as in P consider $\tilde{\beta}_1 = \beta H + \beta_1 H_1 + \overline{\beta}_2 H_2$, reusing the symbols H, H_1, H_2 here for $\tilde{\beta}_1$ in place of $\tilde{\beta}$ as before. Identical arguments as around (A6) show that $H_1 = O_P(1)$ and $H_2 = O_P(h^2/T)$. Then combining the arguments around (A6) with the developments in P, p. 546 and 547 we obtain (A.21) of P:

$$\overline{\beta}'_{\perp}M_{11.\beta}(\tilde{\beta}_1 - \beta_1) = \overline{\beta}'_{\perp}M_{1\varepsilon.\beta}\alpha_{\perp}\Sigma^{-1}_{\alpha_{\perp}\alpha_{\perp}}\zeta(\zeta'\Sigma^{-1}_{\alpha_{\perp}\alpha_{\perp}}\zeta)^{-1} + o_P(1).$$

The rest of the proof of (4.3a) and (4.3b) of P follows as in P.

With respect to the second likelihood ratio test consider

$$\tilde{S}_2(\rho) = \rho\overline{\tilde{\beta}}'_{\perp}M_{11.\tilde{\beta}}\overline{\tilde{\beta}}_{\perp} - \overline{\tilde{\beta}}'_{\perp}M_{1\tilde{\alpha}_{\perp}.\tilde{\beta}}M^{-1}_{\tilde{\alpha}_{\perp}\tilde{\alpha}_{\perp}.\tilde{\beta}}M_{\tilde{\alpha}_{\perp}1.\tilde{\beta}}\overline{\tilde{\beta}}_{\perp}.$$

The results above imply that $\tilde{S}_2(\rho)$ has uniformly in $|\rho| < C$ (for every $0 < C < \infty$) distance to $S_2(\rho)$ of order $O_P(hT^{-1/2})$ where

$$S_2(\rho) = \rho\overline{\beta}'_{\perp}M_{11.\beta}\overline{\beta}_{\perp} - \overline{\beta}'_{\perp}M_{1\alpha_{\perp}.\beta}M^{-1}_{\alpha_{\perp}\alpha_{\perp}.\beta}M_{\alpha_{\perp}1.\beta}\overline{\beta}_{\perp}.$$

Note that since $(\eta, \eta_{\perp})$ is of full rank, (16) is equivalent to

$$\left|\begin{pmatrix}\eta' \\ \eta'_{\perp}\end{pmatrix}S_2(\rho)(\eta, \eta_{\perp})\right| = |\eta'S_2(\rho)\eta| \cdot \left|\eta'_{\perp}\left(S_2(\rho) - S_2(\rho)\eta\,(\eta'S_2(\rho)\eta)^{-1}\,\eta'S_2(\rho)\right)\eta_{\perp}\right| = 0. \quad \text{(A8)}$$

Let $\delta_2 = T\rho$, so that $\rho \to 0$, as $T \to \infty$. As above it can be seen that

$$h^2\left|\eta'S_2(\tfrac{\delta_2}{T})\eta\right| = h^2\left|\tfrac{\delta_2}{T}\beta'_1 M_{11.\beta}\beta_1 - \beta'_1 M_{1\tilde{\alpha}_{\perp}.\beta}M^{-1}_{\tilde{\alpha}_{\perp}\tilde{\alpha}_{\perp}.\beta}M_{\tilde{\alpha}_{\perp}1.\beta}\beta_1\right| \xrightarrow{p} \left|-\tilde{\Sigma}_{U0}\alpha_{\perp}(\alpha'_{\perp}\Sigma_{00}\alpha_{\perp})^{-1}\alpha'_{\perp}\tilde{\Sigma}_{0U}\right| \neq 0.$$

This shows that the s larger roots of $S_2(\rho)$ tend to zero slower than $O(1/T)$. Moreover, we have

$$h\eta'_{\perp}S_2(\tfrac{\delta_2}{T})\eta = h\left(\tfrac{\delta_2}{T}\beta'_2 M_{11.\beta}\beta_1 - \beta'_2 M_{1\tilde{\alpha}_{\perp}.\beta}M^{-1}_{\tilde{\alpha}_{\perp}\tilde{\alpha}_{\perp}.\beta}M_{\tilde{\alpha}_{\perp}1.\beta}\beta_1\right) = -\beta'_2 M_{1\tilde{\alpha}_{\perp}.\beta}(\alpha'_{\perp}\Sigma_{00}\alpha_{\perp})^{-1}\alpha'_{\perp}\tilde{\Sigma}_{0U} + o_P(1),$$

which yields that (using $P_M := (\alpha'_{\perp}\Sigma_{00}\alpha_{\perp})^{-1}\alpha'_{\perp}\tilde{\Sigma}_{0U}(\tilde{\Sigma}_{U0}\alpha_{\perp}(\alpha'_{\perp}\Sigma_{00}\alpha_{\perp})^{-1}\alpha'_{\perp}\tilde{\Sigma}_{0U})^{-1}\tilde{\Sigma}_{U0}\alpha_{\perp}(\alpha'_{\perp}\Sigma_{00}\alpha_{\perp})^{-1}$)

$$\left|\eta'_{\perp}\left(S_2(\tfrac{\delta_2}{T}) - S_2(\tfrac{\delta_2}{T})\eta\,(\eta'S_2(\tfrac{\delta_2}{T})\eta)^{-1}\,\eta'S_2(\tfrac{\delta_2}{T})\right)\eta_{\perp}\right|$$

$$= \left|\left(\delta_2\tfrac{1}{T}\beta'_2 M_{11.\beta}\beta_2 - \beta'_2 M_{1\tilde{\alpha}_{\perp}.\beta}M^{-1}_{\tilde{\alpha}_{\perp}\tilde{\alpha}_{\perp}.\beta}M_{\tilde{\alpha}_{\perp}1.\beta}\beta_2\right) - h\eta'_{\perp}S_2(\tfrac{\delta_2}{T})\eta\left(h^2\eta'S_2(\tfrac{\delta_2}{T})\eta\right)^{-1}h\eta'S_1(\tfrac{\delta_2}{T})\eta_{\perp}\right|$$

$$= \left|\left(\delta_2\tfrac{1}{T}\beta'_2 M_{11.\beta}\beta_2 - \beta'_2 M_{1\alpha_{\perp}.\beta}(\alpha_{\perp}'\Sigma_{00}\alpha_{\perp})^{-1}M_{\alpha_{\perp}1.\beta}\beta_2\right) + \beta'_2 M_{1\alpha_{\perp}.\beta}P_M M_{\alpha_{\perp}1.\beta}\beta_2\right| + o_P(1)$$

$$\xrightarrow{d} \left|\delta_2\int_0^1 B_3 B'_3 - \int_0^1 B_3 dW'\,\alpha_2(\alpha'_2\Sigma_{\epsilon}\alpha_2)^{-1}\alpha'_2\int_0^1 dW B'_3\right| = \left|\delta_2\int_0^1 B_3 B'_3 - (\int_0^1 B_3 dW'_2)(\int_0^1 dW_2 B'_3)\right|$$

using the results of Lemma A.5 of P. and (A.18) of Paruolo (1996) as an expression for

$$(\alpha'_\perp \Sigma_{00}\alpha_\perp)^{-1} - P_M$$

where $W_2 = (\alpha'_2 \Sigma_\epsilon \alpha_2)^{-1/2} \alpha'_2 W$.

Thus, the smallest $(p - r - s)$ solutions of (16) converge in distribution to the solutions of

$$\left| \delta_2 \int_0^1 B_3 B'_3 - \left(\int_0^1 B_3 dW'_2 \right) \left(\int_0^1 dW_2 B'_3 \right) \right| = 0,$$

which shows that the test statistic $Q_{r,s}$ has the following limiting distribution,

$$Q_{r,s} = \sum_{i=s+1}^{p-r} \delta_{2,i} + o_P(1) \xrightarrow{d} tr \left(\int_0^1 dW_2 B'_3 \left(\int_0^1 B_3 B'_3 \right)^{-1} \int_0^1 B_3 dW'_2 \right).$$

It follows also that the sum $S_{r,s} = Q_s + Q_{r,s}$ converges in distribution showing (C).

The rest of the proof of relations (4.3a, b) of P follow exactly as in P. In P (4.4) the order of convergence is replaced by $o_P(T^{-1})$, in (4.5) the error term can be shown to be $o_P(T^{-1/2})$ and in (4.6) instead of the term $O_P(T^{-2})$ we achieve $o_P(1)$.

These terms show consistency for $\tilde{\beta}, \tilde{\eta}$. Using the results of Lemma A.4 of P then consistency for $\tilde{\alpha}, \tilde{\zeta}$ follow.

Following the proof of Theorem 4.2. on pp. 548+549 of P we can show consistency for $\tilde{\psi}$ of P. The only changes refer to the orders of convergence where our setting introduces orders of h into the arguments. Jointly this proves consistency of $\tilde{\Psi}$ and $\tilde{\Gamma}$. Consistency for the coefficients to the stationary terms $\Delta^2 y_{t-j}$ follows as usual from the consistency of the estimates for the coefficients to non-stationary regressors. This completes the proof of (D).

With respect to (E) note that the results above show that the asymptotics for the two eigenvalue problems to be solved converge to the same quantities as in the finite VAR case. This shows that the results of P in this respect hold also in the case of long VARs.

Finally for the matrices Π_j note that Theorem 4.3. of P shows that the asymptotic distribution for all quantities corresponding to stationary regressors are identical for every super-consistent estimator for the coefficients to the non-stationary components.

Appendix E. Proof of Theorem 4

From Theorem 3 it follows that $\hat{\Phi} = \hat{\alpha}\hat{\beta}' \to \Phi, \hat{\Psi} \to \Psi, \hat{\Pi}_j \to \Pi_j, j = 1, 2, \ldots, 2f - 1$. Therefore the Hankel matrix of impulse response coefficients $\hat{\Pi}_j$ converges to the Hankel matrix corresponding to the Π'_js. As $(\bar{A}, B)$ is controllable, (A, B, C) is minimal and $\bar{A}$ is nonsingular according to the assumptions, this Hankel matrix has rank n. This implies that the stochastic realisation algorithm of Appendix F provides consistent estimates $(\hat{A}, \hat{B}, \hat{D}) \to (\bar{A}, B, D)$. This implies

$$\hat{a}(z) = (1 - z)^2 I_p - \hat{\Phi} z - \hat{\Psi} z (1 - z) - (1 - z)^2 z \hat{D} (I_n - z\hat{A})^{-1} \hat{B} \to a(z).$$

For details see Appendix F.

$\hat{a}(z)$ does not necessarily correspond to a rational transfer function of order n. It does so, however, if the additional restrictions (22) hold. Step 3 and 4 of the proposed algorithm achieve this. Here step 3 ascertains that solutions to the third equation exist. The second equation explicitly provides a solution $\bar{\alpha}_\perp$ for given C_+. This solution not necessarily is of full row rank. As in the limit this is the case, it also holds for large enough T. The first equation always admits solutions. Thus for large enough T the set of all solutions is defined by polynomial restrictions. Adding the least squares distance to the estimated impulse response sequence then leads to a quadratic problem under non-linear differentiable constraints, which in the limit has a unique solution. Thus the solution is unique for large enough T.

Consistency of the estimates in combination with continuity of the solution of step 4 implies consistency for the system $(\hat{A}, \hat{B}, \hat{C})$. This implies consistency for the inverse system $(\hat{A}, \hat{B}, \hat{C})$ in the sense of converging impulse response coefficients and hence consistency for the transfer function estimator in the pointwise topology. The fulfillment of restrictions (22) ensures the structure of the corresponding matrix $\hat{A}$ according to state space unit root structure $((0, (c, c + d)))$.

Appendix F. Stochastic Realization Using Overlapping Echelon Forms

This section describes the approximate realization of the first f coefficients $G_j, j = 1, \ldots, 2f$ of an impulse response sequence using a rational transfer function of order n where $f \geq n$. More details can be found in Section 2.6. of Hannan and Deistler (1988).

Define the Hankel matrix

$$
\mathcal{H}_{f,f} = \begin{bmatrix} G_1 & G_2 & G_3 & \ldots & G_f \\ G_2 & G_3 & \ldots & & \\ G_3 & \ldots & & & \\ \vdots & & & & \vdots \\ G_f & G_{f+1} & \ldots & & G_{2f-1} \end{bmatrix} = \begin{bmatrix} h(1,1) \\ h(1,2) \\ \vdots \\ h(1,p) \\ h(2,1) \\ \vdots \\ h(f,p) \end{bmatrix}.
$$

Here $h(i, j)$ denotes the j-th row in the i-th block row. Let $\alpha = (n_1, \ldots, n_p)$ define a *nice selection* of rows[5] of $\mathcal{H}$ such that $\mathcal{H}_\alpha \in \mathbb{R}^{n \times fp}$, the submatrix of $\mathcal{H}$ containing the rows $h(i, j), i \leq n_j$, is of full row rank. If the impulse response corresponds to a transfer function of order at least n there exists such a nice selection α. Finally let $\mathcal{H}_{\alpha+1} \in \mathbb{R}^{n \times fp}$ denote the matrix $\mathcal{H}_\alpha$ shifted down one block row (that is in each row where $\mathcal{H}_\alpha$ contains $h(i, j), \mathcal{H}_{\alpha+1}$ contains $h(i + 1, j)$).

Then it is derived in Hannan and Deistler (1988), Theorem 2.6.2. that if G_j corresponds to a transfer function $k(z) = \sum_{j=1}^{\infty} G_j z^{-j}$ of order exactly n such that the corresponding $\mathcal{H}_\alpha$ is formed using a nice selection, then a system (A, B, C) can be defined using the following formulas

$$
A\mathcal{H}_\alpha = \mathcal{H}_{\alpha+1}, \quad B = \mathcal{H}_\alpha \begin{bmatrix} I_p \\ 0 \end{bmatrix}, \quad C\mathcal{H}_\alpha = \begin{bmatrix} G_1 & G_2 & \ldots & G_f \end{bmatrix} \tag{A9}
$$

such that $G_j = CA^{j-1}B, j = 1, 2, \ldots$.

If the order of the transfer function is larger than n, then the equations for A and C can be solved using least squares. If a sequence of impulse responses $\hat{G}_j \to G_j, j = 1, \ldots, 2f - 1$, and the limit G_j corresponds to a transfer function where the rank of $\mathcal{H}_\alpha$ equals n, it is obvious that the resulting systems $(\hat{A}, \hat{B}, \hat{C}) \to (A, B, C)$ since in this case the least squares solution depends continuously on the matrix $\mathcal{H}$.

References

Banerjee, Anindya, Lynne Cockerell, and Bill Russell. 2001. An I(2) analysis of inflation and the markup. *Journal of Applied Econometrics* 16: 221–40. [CrossRef]

Bauer, Dietmar, and Alex Maynard. 2012. Persistence-robust surplus-lag Granger causality testing. *Journal of Econometrics* 169: 293–300. [CrossRef]

Bauer, Dietmar, and Martin Wagner. 2004. Autoregressive Approximations to MFI(1) Processes. Working Paper No. 174, Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS), Vienna, Austria.

---

[5] A nice selection is such that if $h(i, j)$ is contained in the selection, then also $h(l, j)$ are contained for all $0 < l < i$.

Bauer, Dietmar, and Martin Wagner. 2012. A state space canonical form for unit root processes. *Econometric Theory* 28: 1313–49. [CrossRef]

Berk, Kenneth N. 1974. Consistent autoregressive spectral estimates. *The Annals of Statistics* 2: 489–502. [CrossRef]

Boswijk, H. Peter, and Jurgen A. Doornik. 2004. Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica* 58: 440–65. [CrossRef]

Boswijk, H. Peter, and Paolo Paruolo. 2017. Likelihood ratio tests of restrictions on common trends loading matrices in I(2) VAR systems. *Econometrics* 5: 28. [CrossRef]

Chan, Ngai Hang, and Ching Zong Wei. 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. *The Annals of Statistics* 16: 367–401. [CrossRef]

Dolado, Juan J., and Helmut Lütkepohl. 1996. Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* 15: 369–86. [CrossRef]

Engle, Robert F., and Clive W.J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55: 251–76. [CrossRef]

Georgoutsos, Dimitris A., and Georgios P. Kouretas. 2004. A Multivariate I (2) cointegration analysis of German hyperinflation. *Applied Financial Economics* 14: 29–41. [CrossRef]

Hannan, Edward James, and Manfred Deistler. 1988. *The Statistical Theory of Linear Systems*. New York: John Wiley.

Hannan, Edward James, and Laimonis Kavalieris. 1986. Regression, autoregression models. *Journal of Time Series Analysis* 7: 27–49. [CrossRef]

Inoue, Atsushi, and Lutz Kilian. 2020. The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics* 215: 450–72. [CrossRef]

Johansen, Søren, and Helmut Lütkepohl. 2005. A note on testing restrictions for the cointegration parameters of a VAR with I(2) variables. *Econometric Theory* 21: 653–58. [CrossRef]

Johansen, Søren, Katarina Juselius, Roman Frydman, and Michael Goldberg. 2007. Testing hypotheses in an I (2) model with applications to the persistent long swings in the Dmk/$ rate. *Journal of Econometrics* 158: 1–35. [CrossRef]

Johansen, Søren. 1992a. A representation of vector autoregresive processes integrated of order 2. *Econometric Theory* 8: 188–202. [CrossRef]

Johansen, Søren. 1992b. Testing weak exogeneity and the order of cointegration in UK money demand data. *Journal of Policy Modeling* 14: 313–34. [CrossRef]

Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford: Oxford University Press.

Johansen, Søren. 1997. Likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24: 433–62. [CrossRef]

Juselius, Katarina, and Katrin Assenmacher. 2017. Real exchange rate persistence and the excess return puzzle: The case of Switzerland versus the US. *Journal of Applied Econometrics* 32: 1145–55. [CrossRef]

Juselius, Katarina, and Josh R. Stillwagon. 2018. Are outcomes driving expectations or the other way around? An I(2) CVAR analysis of interest rate expectations in the dollar/pound market. *Journal of International Money and Finance* 83: 93–105. [CrossRef]

Juselius, Katarina. 1994. On the duality between long-run relations and common trends in the i(1) versus (2) model. an application to aggregate money holdings. *Econometric Reviews* 13: 151–78. [CrossRef]

Juselius, Katarina. 2006. *The Cointegrated VAR Model*. Oxford: Oxford University Press.

Kurita, Takamitsu, Heino Bohn Nielsen, and Anders Rahbek. 2011. An I(2) cointegration model with piecewise linear trends. *Econometrics Journal* 14: 131–55. [CrossRef]

Kurita, Takamitsu. 2012. Likelihood-based inference for weak exogeneity in I(2) cointegrated VAR models. *Econometric Reviews* 31: 325–60. [CrossRef]

Lewis, Richard, and Gregory C. Reinsel. 1985. Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16: 393–411. [CrossRef]

Lütkepohl, Helmut, and Holger Claessen. 1997. Analysis of cointegrated VARMA processes. *Journal of Econometrics* 80: 223–29. [CrossRef]

Lütkepohl, Helmut, and Pentti Saikkonen. 1997. Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics* 81: 127–57. [CrossRef]

Mosconi, Rocco, and Paolo Paruolo. 2013. *Identification of Cointegrating Relations in I(2) Vector Autoregressive Models*. Rome: PRIN Workshop Forecasting Economic and Financial Time Series, Research Publications at Politecnico di Milano, pp. 1–35.

Mosconi, Rocco, and Paolo Paruolo. 2017. Identification conditions in simultaneous systems of cointegrating equations with integrated variables of higher order. *Journal of Econometrics* 198: 271–76. [CrossRef]

Nielsen, Heino Bohn, and Anders Rahbek. 2007. The likelihood ratio test for cointegration ranks in the I(2) model. *Econometric Theory* 23: 615–37. [CrossRef]

Palma, Wilfredo, and Pascal Bondon. 2003. On the eigenstructure of generalized fractional processes. *Statistics and Probability Letters* 65: 93–101. [CrossRef]

Paruolo, Paolo, and Anders Rahbek. 1999. Weak exogeneity in I(2) VAR systems. *Journal of Econometrics* 93: 281–308. [CrossRef]

Paruolo, Paolo. 1994. The role of the drift in I(2) systems. *Journal of the Italian Statistical Society* 3: 93–123. [CrossRef]

Paruolo, Paolo. 1996. On the determination of integration indices in I(2) systems. *Journal of Econometrics* 72: 313–56. [CrossRef]

Paruolo, Paolo. 2000. Asymptotic efficiency of the two stage estimator in I(2) systems. *Econometric Theory* 16: 524–50. [CrossRef]

Paruolo, Paolo. 2006. Common trends and cycles in I(2) VAR systems. *Journal of Econometrics* 132: 143–68. [CrossRef]

Rahbek, Anders, Hans Christian Kongsted, and Clara Jørgensen. 1999. Trend stationarity in the I(2) cointegration model. *Journal of Econometrics* 90: 265–289. [CrossRef]

Saikkonen, Pentti, and Helmut Lütkepohl. 1996. Infinite-order cointegrated vector autoregressive processes : estimation and inference. *Econometric Theory* 12: 814–44. [CrossRef]

Saikkonen, Pentti, and Ritva Luukkonen. 1997. Testing cointegration in infinite order vector autoregressive processes. *Journal of Econometrics* 81: 93–126. [CrossRef]

Saikkonen, Pentti. 1991. Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7: 1–21. [CrossRef]

Saikkonen, Pentti. 1992. Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8: 1–27. [CrossRef]

Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. Inference in linear time series models with some unit roots. *Econometrica* 58: 113–44. [CrossRef]

Stillwagon, Josh R. 2018. Are risk premia related to real exchange rate swings? Evidence from I(2) CVARs with survey expectations. *Macroeconomic Dynamics* 22: 255–78. [CrossRef]