

ORIGINAL ARTICLE OPEN ACCESS

Inference on Common Trends in a Cointegrated Nonlinear SVAR

James A. Duffy¹  | Xiyu Jiao²¹Department of Economics, University of Oxford, Oxford, UK | ²Department of Economics, University of Gothenburg, Gothenburg, Sweden**Correspondence:** James A. Duffy (james.duffy@economics.ox.ac.uk)**Received:** 8 April 2026 | **Accepted:** 13 April 2026**Keywords:** cointegration | common trends | nonlinear SVAR | variance ratio test

ABSTRACT

We consider the problem of performing inference on the number of common stochastic trends when data is generated by a cointegrated CKSVAR (a two-regime, piecewise affine SVAR; Mavroudis, 2021), using a modified version of the Breitung (2002) multivariate variance ratio test that is robust to the presence of nonlinear cointegration (of a known form). To derive the asymptotics of our test statistic, we prove a fundamental LLN-type result for a class of stable but nonstationary autoregressive processes, using a novel dual linear process approximation. We show that our modified test yields correct inferences regarding the number of common trends in such a system, whereas the unmodified test tends to infer a higher number of common trends than are actually present, when cointegrating relations are nonlinear.

1 | Introduction

For almost half a century, the structural vector autoregression (SVAR) has been the workhorse model of empirical macroeconomics. In addition to providing a tractable framework for the identification of causal relationships in the presence of simultaneity, the model succeeds in capturing many of the characteristic properties of macroeconomic time series: their temporal dependence, their trending and random wandering behaviour, and the tendency of related series to move together. In this regard, the emergence of the theory of cointegration [1, 2] was of major significance. For by formalising that co-movement in terms of common stochastic trends, it made it possible to identify the precise conditions under which an SVAR could generate such common trends, as per the Granger–Johansen representation theorem (GJRT; [3, 4]). This result has, in turn, provided the basis for a rich and fruitful theory of asymptotic inference in cointegrated SVARs, concerning the number of common stochastic trends in the system (or equivalently, the cointegrating

rank), the coefficients on the cointegrating relations, and the model parameters (and implied impulse responses, etc.).

In its original conception, cointegration was inherently linear; there have since been multifarious efforts to extend it in a nonlinear direction, as reviewed by Tjøstheim [5]. Paralleling those efforts has been the burgeoning of a literature on nonlinear SVARs, but which has been confined almost entirely to the modelling of stationary time series (see, e.g., [6, 7]; for the exceptional case of ‘nonlinear VECM’ models, see [8]). This unfortunately precludes the application of these nonlinear SVARs to settings where, for economic reasons, the nonlinearities relate to the level of a stochastically trending series, so that reformulating the model in terms of the (more approximately stationary) differenced series is not appropriate. A leading example arises in the context of the zero lower bound (ZLB) constraint on nominal interest rates, which refers to the level of a highly persistent—and arguably integrated—series, rather than to its first differences.

This paper is for inclusion in the *Oxford Bulletin of Economics & Statistics* Special Issue, ‘Celebrating 40 Years of Unit Roots and Cointegration’.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.

The development of a new class of ‘endogenous regime switching’ piecewise affine SVARs—and their successful application to highly persistent series that are subject to occasionally binding constraints [9–11]—has recently foregrounded the question of whether, and how, one can accommodate stochastic trends in *nonlinear* SVARs. By way of an answer, Duffy, Mavroeidis and Wycherley [12] and Duffy and Mavroeidis [13] provide extensions of the GJRT to a broad class of nonlinear SVARs: in the former, to a two-regime piecewise affine SVAR (the ‘CKSVAR’), and in the latter, to more general, additively time-separable nonlinear SVARs of the form

$$f_0(z_t) = c + \sum_{i=1}^k f_i(z_{t-i}) + u_t \quad (1.1)$$

where z_t and u_t are respectively the observed series and the innovations, both of which are \mathbb{R}^p -valued, and $f_i : \mathbb{R}^p \rightarrow \mathbb{R}^p$. Their results demonstrate that, alongside linear cointegration, nonlinear SVARs of the form (1.1) are capable of accommodating much richer varieties of long-run behaviour than are linear SVARs, including *nonlinear* common stochastic trends and *nonlinear* cointegrating relations.

There remains the question of how to perform inference in the setting of (1.1), in the presence of (linear or nonlinear) cointegration. In this paper, we consider this problem when (1.1) is specialised to the two-regime piecewise affine model of Duffy et al. [12], as per

$$\phi_0^+ y_t^+ + \phi_0^- y_t^- + \Phi_0^x x_t = c + \sum_{i=1}^k [\phi_i^+ y_{t-i}^+ + \phi_i^- y_{t-i}^- + \Phi_i^x x_{t-i}] + u_t, \quad (1.2)$$

where we have partitioned $z_t = (y_t, x_t^\top)^\top$ such that y_t is \mathbb{R} -valued and x_t is \mathbb{R}^{p-1} -valued, and $y_t^+ = \max\{y_t, 0\}$ and $y_t^- = \min\{y_t, 0\}$ respectively denote the positive and negative parts of y_t . We further suppose that this model is configured such that the cointegrating rank, r , is invariant to the sign of y_t , while permitting those r cointegrating relations to be nonlinear: what is termed ‘case (ii)’ in the typology of Duffy et al. [12]; see Section 2 for a discussion. Even in this case, asymptotic inference is complicated by the fact that the processes generated by the model do not readily fall within any class previously considered in econometrics. Although $\{z_t\}$ behaves similarly, in large samples, to a (linear) integrated process, in the sense that $n^{-1/2} z_{\lfloor n\lambda \rfloor}$ converges weakly to a nondegenerate limiting process $Z(\lambda)$, neither its first differences nor the equilibrium errors will be stationary, but instead follow a (stable) time-varying autoregressive process, whose coefficients depend on the sign of the integrated process $\{y_t\}$. This renders any existing LLN-type results for ‘weakly dependent’ processes inapplicable.

In this paper, we take the first steps towards the development of valid asymptotic inference in the model (1.2), in the presence of cointegration. We do so by considering the simpler problem of inference on the cointegrating rank of (1.2), using a form of the Breitung [14] multivariate variance ratio test statistic, modified so as to accommodate the possibility of nonlinear cointegration. This motivates the main technical contribution of the paper: a new LLN-type result for the class of time-varying, stable, but nonstationary autoregressive processes that may be generated by (1.2), which is provided in Section 3 along with the asymptotics

of our test statistic. This result is fundamental to the asymptotics of estimators of the parameters of (1.2), the derivation of which is the subject of the authors’ ongoing research. The finite-sample performance of our proposed test is investigated through simulation exercises reported in Section 4, where it is shown that the conventional (i.e., unmodified) Breitung [14] test tends to incorrectly interpret the presence of *nonlinear* cointegration as evidence in favour of additional stochastic trends being present in the data, a problem that is avoided by our proposed test. Section 5 concludes.

Notation. $e_{m,i}$ denotes the i th column of the $m \times m$ identity matrix I_m ; when m is clear from the context, we write this simply as e_i . In a statement such as $f(a^\pm, b^\pm) = 0$, the notation ‘ \pm ’ signifies that both $f(a^+, b^+) = 0$ and $f(a^-, b^-) = 0$ hold; similarly, ‘ $a^\pm \in A$ ’ denotes that both a^+ and a^- are elements of A . All limits are taken as $n \rightarrow \infty$ unless otherwise stated. \xrightarrow{p} and \rightsquigarrow respectively denote convergence in probability and in distribution (weak convergence). We write ‘ $X_n(\lambda) \rightsquigarrow X(\lambda)$ on $D_{\mathbb{R}^m}[0, 1]$ ’ to denote that $\{X_n\}$ converges weakly to X , where these are considered as random elements of $D_{\mathbb{R}^m}[0, 1]$, the space of cadlag functions $[0, 1] \rightarrow \mathbb{R}^m$, equipped with the uniform topology; we denote this as $D[0, 1]$ whenever the value of m is clear from the context. $\|\cdot\|$ denotes the Euclidean norm on \mathbb{R}^m , and the matrix norm that it induces. For X a random vector and $p \geq 1$, $\|X\|_p := (\mathbb{E}\|X\|^p)^{1/p}$. C, C_1 , etc., denote generic constants that may take different values at different places of the same proof.

2 | Model: The Censored and Kinked SVAR

2.1 | Framework

We consider a structural VAR(k) model in p variables, in which one series, y_t , enters with coefficients that differ according to whether it is above or below a time-invariant threshold b , while the other $p - 1$ series, collected in x_t , enter linearly [9, 12]. Defining

$$y_t^+ := \max\{y_t, b\} \quad y_t^- := \min\{y_t, b\}, \quad (2.1)$$

we specify that $z_t = (y_t, x_t^\top)^\top$ follow

$$\phi_0^+ y_t^+ + \phi_0^- y_t^- + \Phi_0^x x_t = c + \sum_{i=1}^k [\phi_i^+ y_{t-i}^+ + \phi_i^- y_{t-i}^- + \Phi_i^x x_{t-i}] + u_t \quad (2.2)$$

or, more compactly,

$$\phi^+(L)y_t^+ + \phi^-(L)y_t^- + \Phi^x(L)x_t = c + u_t, \quad (2.3)$$

where

$$\phi^\pm(L) := \phi_0^\pm - \sum_{i=1}^k \phi_i^\pm L^i, \quad \Phi^x(L) := \Phi_0^x - \sum_{i=1}^k \Phi_i^x L^i,$$

for $\phi_i^\pm \in \mathbb{R}^{p \times 1}$ and $\Phi_i^x \in \mathbb{R}^{p \times (p-1)}$, and L denotes the lag operator. Through an appropriate redefinition of y_t and c , we may take b (which we treat here as being known) to be zero without loss of generality, and will do so throughout the sequel. In this case, y_t^+ and y_t^- respectively equal the positive and negative parts of y_t , and $y_t = y_t^+ + y_t^-$.¹ Following Mavroeidis [9], we term this model the ‘censored and kinked SVAR’ (CKSVAR), even though we here

suppose that y_t is observed on *both* sides of zero, rather than being subject to censoring.

We follow Mavroeidis [9] and Aruoba et al. [10] in maintaining the following conditions, which are necessary and sufficient to ensure that (2.3) has a unique solution for (y_t, x_t) , for all possible values of u_t . Define

$$\Phi_0 := \begin{bmatrix} \phi_0^+ & \phi_0^- & \Phi_0^x \end{bmatrix} = \begin{bmatrix} \phi_{0,yy}^+ & \phi_{0,yy}^- & \phi_{0,yx}^T \\ \phi_{0,xy}^+ & \phi_{0,xy}^- & \Phi_{0,xx} \end{bmatrix},$$

$$\Phi_0^+ := [\phi_0^+, \Phi_0^x] \text{ and } \Phi_0^- := [\phi_0^-, \Phi_0^x].$$

Assumption DGP.

1. $\{(y_t, x_t)\}$ are generated according to (2.1) to (2.3) with $b = 0$, with (possibly random) initial values (y_i, x_i) , for $i \in \{-k + 1, \dots, 0\}$;
2. $\text{sgn}(\det \Phi_0^+) = \text{sgn}(\det \Phi_0^-) \neq 0$.
3. $\Phi_{0,xx}$ is invertible, and $\text{sgn}\{\phi_{0,yy}^+ - \phi_{0,yx}^T \Phi_{0,xx}^{-1} \phi_{0,xy}^+\} = \text{sgn}\{\phi_{0,yy}^- - \phi_{0,yx}^T \Phi_{0,xx}^{-1} \phi_{0,xy}^-\} > 0$.
4. $\{u_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence in \mathbb{R}^p with $\mathbb{E}u_t = 0$, $\mathbb{E}u_t u_t^T = \Sigma_u$ positive definite, and $\|u_t\|_{2+\delta_u} < \infty$ for some $\delta_u > 0$.

As discussed in Duffy et al. [15, Rem. 2.1(i)], DGP.3 may be maintained without loss of generality, when the invertibility condition DGP.2 holds. Let $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ denote an underlying filtration to which the preceding processes are all adapted. When we say that a sequence is i.i.d., as per $\{u_t\}_{t \in \mathbb{Z}}$ in DGP.4, we mean that this sequence is $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ -adapted, and additionally that u_s is independent of \mathcal{F}_t for $s > t$. An immediate implication of DGP.4 is that

$$U_n(\lambda) := n^{-1/2} \sum_{t=1}^{\lfloor n\lambda \rfloor} u_t \rightsquigarrow U(\lambda) \tag{2.4}$$

on $D[0, 1]$, where U is a p -dimensional Brownian motion with variance Σ_u . All the weak convergences that are stated in this paper hold jointly with (2.4).

2.2 | Canonical Form

In the terminology of Duffy et al. [15] and Duffy et al. [12], we designate a CKSVAR as *canonical* if

$$\Phi_0 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & I_{p-1} \end{bmatrix} =: I_p^*. \tag{2.5}$$

While it is not always the case that the reduced form of (2.3) corresponds directly to a canonical CKSVAR, by defining the canonical variables

$$\begin{bmatrix} \tilde{y}_t^+ \\ \tilde{y}_t^- \\ \tilde{x}_t \end{bmatrix} := \begin{bmatrix} \phi_{0,yy}^+ & 0 & 0 \\ 0 & \phi_{0,yy}^- & 0 \\ \phi_{0,xy}^+ & \phi_{0,xy}^- & \Phi_{0,xx} \end{bmatrix} \begin{bmatrix} y_t^+ \\ y_t^- \\ x_t \end{bmatrix} =: P^{-1} \begin{bmatrix} y_t^+ \\ y_t^- \\ x_t \end{bmatrix}, \tag{2.6}$$

where $\tilde{\phi}_{0,yy}^\pm := \phi_{0,yy}^\pm - \phi_{0,yx}^T \Phi_{0,xx}^{-1} \phi_{0,xy}^\pm > 0$ and P^{-1} is invertible under DGP; and setting

$$\begin{bmatrix} \tilde{\phi}^+(\beta) & \tilde{\phi}^-(\beta) & \tilde{\Phi}^x(\beta) \end{bmatrix} := Q \begin{bmatrix} \phi^+(\beta) & \phi^-(\beta) & \Phi^x(\beta) \end{bmatrix} P, \tag{2.7}$$

for $\beta \in \mathbb{C}$, where

$$Q := \begin{bmatrix} 1 & -\phi_{0,yx}^T \Phi_{0,xx}^{-1} \\ 0 & I_{p-1} \end{bmatrix}, \tag{2.8}$$

we obtain a canonical CKSVAR for $\tilde{z}_t := (\tilde{y}_t, \tilde{x}_t^T)^T$ (see Proposition 2.1 in [15]).

To distinguish between a general CKSVAR in which possibly $\Phi_0 \neq I_p^*$, and its associated canonical form, we shall refer to the former as the ‘structural form’ of the CKSVAR. Since the time series properties of a general CKSVAR are largely inherited from its derived canonical form, we shall occasionally work with this more convenient representation of the system, and indicate this as follows.

Assumption DGP*. $\{(y_t, x_t)\}$ are generated by a canonical CKSVAR, i.e., DGP holds with $\Phi_0 = [\phi_0^+, \phi_0^-, \Phi_0^x] = I_p^*$, so that (2.2) may be equivalently written as

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = c + \sum_{i=1}^k \begin{bmatrix} \phi_i^+ & \phi_i^- & \Phi_i^x \end{bmatrix} \begin{bmatrix} y_{t-i}^+ \\ y_{t-i}^- \\ x_{t-i} \end{bmatrix} + u_t. \tag{2.9}$$

2.3 | The Cointegrated CKSVAR

Duffy et al. [12], henceforth DMW25, develop conditions under which the CKSVAR is capable of generating cointegrated time series. Their work identifies three cases, which may be distinguished according to whether stochastic trends are imparted: (i) to y_t^+ only (or equivalently to y_t^- only); (ii) to both y_t^+ and y_t^- ; and (iii) to neither y_t^+ nor y_t^- . Here, our focus is on case (ii), which entails that the system has a well-defined cointegrating rank r , but permits the r cointegrating relationships that eliminate the $(p - r = q)$ common trends to be nonlinear. The assumptions that characterise how the model needs to be configured for case (ii) are given below. To state these, define the autoregressive polynomials

$$\Phi^\pm(\beta) := \begin{bmatrix} \phi^\pm(\beta) & \Phi^x(\beta) \end{bmatrix},$$

and let $\Gamma_i^\pm := -\sum_{j=i+1}^k \Phi_j^\pm =: [\gamma_i^\pm, \Gamma_i^x]$ for $i \in \{1, \dots, k - 1\}$, so that $\Gamma^\pm(\beta) := \Phi_0^\pm - \sum_{i=1}^{k-1} \Gamma_i^\pm \beta^i$ is such that

$$\Phi^\pm(\beta) = \Phi^\pm(1)\beta + \Gamma^\pm(\beta)(1 - \beta).$$

We further define

$$\Pi^\pm := -\Phi^\pm(1) = -[\phi^\pm(1), \Phi^x(1)] =: [\pi^\pm, \Pi^x].$$

Assumption CVAR.

1. $\det \Phi^\pm(\beta)$ has $q^\pm \in \{1, \dots, p\}$ roots at real unity, and all others outside the unit circle; and
2. $\text{rk} \Pi^\pm = r^\pm = p - q^\pm$.

The preceding conditions are common to all three cases noted above. To specialise to case (ii), which has a constant cointegrating rank $r = r^+ = r^-$, with a stochastic trend being present in y_t ,

we must additionally suppose that $\text{rk}\Pi^x = r$, so that Π^\pm may be written as

$$\Pi^\pm = \Pi^x \begin{bmatrix} \theta^\pm & I_{p-1} \end{bmatrix} = \alpha \begin{bmatrix} \beta_y^\pm & \beta_x^\pm \end{bmatrix} =: \alpha \beta^{\pm\top},$$

where $\alpha \in \mathbb{R}^{p \times r}$, $\beta_x \in \mathbb{R}^{(p-1) \times r}$ and $\beta^\pm \in \mathbb{R}^{p \times r}$ have rank r , and $\theta^\pm \in \mathbb{R}^{p-1}$ is such that $\Pi^x \theta^\pm = \pi^\pm$ (see Section 4.2 of DMW25). Letting $\mathbf{1}^+(y) := \mathbf{1}\{y \geq 0\}$ and $\mathbf{1}^-(y) := \mathbf{1}\{y < 0\}$, the (possibly nonlinear) r cointegrating relationships among the elements of z_t are given by

$$\beta(y) := \beta^+ \mathbf{1}^+(y) + \beta^- \mathbf{1}^-(y).$$

Let $\alpha_\perp \in \mathbb{R}^{p \times q}$ be such that $\alpha_\perp^\top \alpha = 0$, and $[\alpha, \alpha_\perp]$ is nonsingular. The limiting form of the stochastic trends will be a kind of (regime-dependent) projection of the p -dimensional Brownian motion U onto a manifold of dimension $q = p - r$, where this projection is defined in terms of

$$P_{\beta_\perp}(y) := \beta_\perp(y) [\alpha_\perp^\top \Gamma(1; y) \beta_\perp(y)]^{-1} \alpha_\perp^\top, \quad (2.10)$$

$$\beta_\perp(y) := \begin{bmatrix} 1 & 0 \\ -\theta(y) & \beta_{x,\perp} \end{bmatrix}, \quad \Gamma(1; y) := \Gamma^+(1) \mathbf{1}^+(y) + \Gamma^-(1) \mathbf{1}^-(y), \quad (2.11)$$

for $\theta(y) := \mathbf{1}^+(y)\theta^+ + \mathbf{1}^-(y)\theta^-$. (Such objects as $P_{\beta_\perp}(y)$ take only two distinct values, depending on the sign of y , and we routinely use the notation $P_{\beta_\perp}(+1)$ and $P_{\beta_\perp}(-1)$ to indicate these.) Define $\alpha, \beta(y) \in \mathbb{R}^{[k(p+1)-1] \times [r+(k-1)(p+1)]}$ as

$$\alpha := \begin{bmatrix} \alpha & \Gamma_1 & \Gamma_2 & \cdots & \Gamma_{k-1} \\ & I_{p+1} & & & \\ & & I_{p+1} & & \\ & & & \ddots & \\ & & & & I_{p+1} \end{bmatrix}, \quad \beta(y)^\top := \begin{bmatrix} \beta(y)^\top \\ S_p(y) & -I_{p+1} \\ & I_{p+1} & -I_{p+1} \\ & & \ddots & \\ & & & I_{p+1} & -I_{p+1} \end{bmatrix}, \quad (2.12)$$

where $\Gamma_i := [\gamma_i^+, \gamma_i^-, \Gamma_i^x]$ for $i \in \{1, \dots, k-1\}$, and

$$S_p(y) := \begin{bmatrix} \mathbf{1}^+(y) & \mathbf{1}^-(y) & 0 \\ 0 & 0 & I_{p-1} \end{bmatrix}^\top. \quad (2.13)$$

Finally, let $\rho(M)$ denote the spectral radius of $M \in \mathbb{R}^{m \times m}$, and for $\mathcal{A} \subset \mathbb{R}^{m \times m}$ a bounded collection of matrices, let

$$\rho_{\text{JSR}}(\mathcal{A}) := \limsup_{t \rightarrow \infty} \sup_{B \in \mathcal{A}^t} \rho(B)^{1/t}$$

denote its joint spectral radius (JSR; e.g., [16], Defn. 1.1), where $\mathcal{A}^t := \{\prod_{s=1}^t M_s \mid M_s \in \mathcal{A}\}$ is the set of t -fold products of matrices in \mathcal{A} .

Assumption CO(ii).

1. $r^+ = r^- = \text{rk}\Pi^x = r$, for some $r \in \{0, 1, \dots, p-1\}$.
2. $\rho_{\text{JSR}}(\{I + \tilde{\beta}(+1)^\top \tilde{\alpha}, I + \tilde{\beta}(-1)^\top \tilde{\alpha}\}) < 1$.
3. $\text{sgn det } \alpha_\perp^\top \Gamma(1; +1) \beta_\perp(+1) = \text{sgn det } \alpha_\perp^\top \Gamma(1; -1) \beta_\perp(-1) \neq 0$.
4.
 - a. $\beta(y_t)^\top z_t$ and Δz_t have uniformly bounded $2 + \delta_u$ moments, for $t \in \{-k+1, \dots, 0\}$.
 - b. $n^{-1/2} z_0 \xrightarrow{p} \mathcal{Z}_0 = \begin{bmatrix} \mathcal{Y}_0 \\ \mathcal{X}_0 \end{bmatrix}$, where \mathcal{Z}_0 is non-random, and satisfies $\beta(\mathcal{Y}_0)^\top \mathcal{Z}_0 = 0$.

Condition CO(ii).2 is stated slightly differently from the form given in DMW25, so as to more directly accommodate the case of a general (i.e., non-canonical) CKSVAR. In particular, $\tilde{\beta}(y)$ and $\tilde{\alpha}$ refer to the counterparts of (2.12) constructed from the parameters of the canonical form of the CKSVAR, derived via the mapping (2.7). (So if the CKSVAR is in fact canonical, the tildes are redundant.) See Remark 4.2(i) of DMW25 for further details. Regarding the history of the process prior to time $t = -k + 1$, we henceforth adopt the (innocuous) convention that

$$\Delta z_t = 0, \quad \forall t \leq -k; \quad (2.14)$$

or equivalently that $z_t = z_{-k}$ for all $t \leq -k$.

Finally, for the purposes of developing the asymptotics of our rank test (Theorem 3.2 below), we shall maintain that the intercept c is such that no deterministic trends are present in any of the model variables, as per

Assumption DET. $c \in \text{sp}\Pi^+ \cap \text{sp}\Pi^-$.

Under the preceding conditions (DGP, CVAR, CO(ii) and DET), it follows by Theorem 4.2 in DMW25 that

$$n^{-1/2} \begin{bmatrix} Y_{[n\lambda]} \\ X_{[n\lambda]} \end{bmatrix} = n^{-1/2} z_{[n\lambda]} \rightsquigarrow P_{\beta_\perp} [Y(\lambda)] U_0(\lambda) =: Z(\lambda) = \begin{bmatrix} Y(\lambda) \\ X(\lambda) \end{bmatrix}, \quad (2.15)$$

where $U_0(\lambda) = \Gamma(1; \mathcal{Y}_0) \mathcal{Z}_0 + U(\lambda)$. (For a further heuristic discussion of the convergence in (2.15) and the properties of the limiting process $Z(\lambda)$, see Section 3.3 of DMW25.) Since $P_{\beta_\perp}(\pm 1)$ are rank q (oblique projection) matrices, we may regard $\{z_t\}$ as having q common (stochastic) trends, and r cointegrating relationships given by the columns of $\beta(y)$, that eliminate those trends (since $\beta(y)^\top P_{\beta_\perp}(y) = 0$).

On the basis of (2.15), DMW25 (see their Defn. 3.1) classify $\{z_t\}$ as $I^*(1)$, because $n^{-1/2} z_{[n\lambda]}$ converges weakly to a non-degenerate process. By contrast, since the equilibrium errors $\xi_t := \beta(y_t)^\top z_t$ are purged of the common trends in z_t , these satisfy $\max_{1 \leq t \leq n} \|\xi_t\| = o_p(n^{1/2})$, and so are of strictly smaller order than $\{z_t\}$; they accordingly classify $\{\xi_t\}$ as $I^*(0)$. These notions of $I^*(0)$ and $I^*(1)$ processes provide a means of distinguishing between processes whose magnitudes differ, because of the presence or absence of stochastic trends, in a setting where the usual definitions of $I(0)$ and $I(1)$ processes do not apply—because in

general neither ξ_t nor Δz_t will be stationary under the foregoing assumptions.

Although (2.15) implies that Z is not ‘globally’ a linear projection of U_0 onto a q -dimensional linear subspace, the following relationships hold ‘locally’, depending on the sign of the first component, Y , of Z :

$$Y(\lambda) > 0 \Rightarrow \beta^{+\top} Z(\lambda) = 0, \quad Y(\lambda) < 0 \Rightarrow \beta^{-\top} Z(\lambda) = 0.$$

But in general neither $\beta^{+\top} Z(\lambda)$ nor $\beta^{-\top} Z(\lambda)$ will be identically zero for all $\lambda \in [0, 1]$, unless $\beta^+ = \beta^-$. The fact that there may be no rank $r = p - q$ matrix whose columns force Z to be identically zero significantly complicates the problem of inference on the cointegrating rank, and motivates our development of a modified form of the Breitung [14] test below.

3 | The Modified Breitung [14] Test

3.1 | Fundamental Ideas

We seek to develop an (asymptotically valid) test on the cointegrating rank r —or equivalently, the number of common trends q —that is able to accommodate the possibility of data generated by a CKSVAR configured as per case (ii), by adapting the approach of Breitung [14, Sec. 5]. Henceforth, as per the discussion following (2.1) above, the threshold b that delineates the two regimes is assumed to be known, and normalised to zero: so that what we have denoted as y_t^+ and y_t^- may be regarded as directly observed, rather than depending on some prior estimator of b . Estimation of b may be undertaken in conjunction with the estimation of the other parameters of the SVAR (2.2), e.g., by maximum likelihood, the asymptotics of which are deferred to future work. (We anticipate that use of a consistent estimator of b would yield a test statistic with an identical null limiting distribution to that derived below: due to y_t being integrated under the null, any misclassification that results from $\hat{b}_n \neq b$ would affect at most $o_p(n^{1/2})$ observations.)

The mathematical underpinnings of Breitung’s [14] test, itself a multivariate generalisation of the variance ratio test, may be conveniently summarised as follows. (The proof of which, together with those of all other results given in this section, appears in Appendix B.)

Proposition 3.1. *Suppose that $\{w_{n,t}\}_{t=1}^n$ is a triangular array, taking values in \mathbb{R}^{d_w} , such that*

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} w_{n,t} \rightsquigarrow \int_0^\lambda \begin{bmatrix} \mathbb{W}(s) \\ 0_{d_w-\ell} \end{bmatrix} ds =: \begin{bmatrix} \mathbb{V}(\lambda) \\ 0_{d_w-\ell} \end{bmatrix} \quad (3.1)$$

on $D_{\mathbb{R}^{d_w}}[0, 1]$, where \mathbb{W} is a random element of $D_{\mathbb{R}^\ell}[0, 1]$ and

$$\frac{1}{n} \sum_{t=1}^n w_{n,t} w_{n,t}^\top \rightsquigarrow \begin{bmatrix} \int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top ds & 0 \\ 0 & \Omega \end{bmatrix} \quad (3.2)$$

where $\int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top ds$, $\int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top ds$ and $\Omega \in \mathbb{R}^{(d_w-\ell) \times (d_w-\ell)}$ are a.s. positive definite. Let $\{\lambda_{n,i}\}_{i=1}^{d_w}$ denote the solutions to

$$\det(\lambda \mathbb{B}_n - \mathbb{A}_n) = 0$$

ordered as $\lambda_{n,1} \leq \lambda_{n,2} \leq \dots \leq \lambda_{n,d_w}$, for

$$\mathbb{A}_n := \sum_{t=1}^n w_{n,t} w_{n,t}^\top, \quad \mathbb{B}_n := \sum_{t=1}^n \sum_{i=1}^l w_{n,t} \sum_{j=1}^l w_{n,t}^\top.$$

Then

i. if $\ell_0 = \ell$,

$$n^2 \sum_{i=1}^{\ell_0} \lambda_{n,i} \rightsquigarrow \text{tr} \left[\int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top ds \left(\int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top ds \right)^{-1} \right];$$

ii. if $\ell_0 > \ell$, $n^2 \sum_{i=1}^{\ell_0} \lambda_{n,i} \xrightarrow{P} \infty$.

To illustrate how Proposition 3.1 provides the basis for a test of cointegrating rank, let us suppose initially that $\{z_t\}$ is generated by a linearly cointegrated SVAR with q common trends, or more generally by a CKSVAR satisfying the conditions above (DGP, CVAR, CO(ii) and DET), but for which $c = 0$, $\beta^+ = \beta^- = \beta$ and $\Gamma^+(1) = \Gamma^-(1) = \Gamma(1)$. Then $P_{\beta_\perp}(y)$ no longer depends on (the sign of) y , and (2.15) reduces to

$$n^{-1/2} z_{\lfloor n\lambda \rfloor} \rightsquigarrow \beta_\perp [\alpha_\perp^\top \Gamma(1) \beta_\perp]^{-1} \alpha_\perp^\top U_0(\lambda).$$

It follows that by taking

$$w_{n,t} := \begin{bmatrix} n^{-1/2} \beta_\perp^\top \\ \beta^\top \end{bmatrix} z_t$$

we may linearly separate z_t into its q ‘integrated’ (i.e., $I^*(1)$) and $r = p - q$ ‘weakly dependent’ (i.e., $I^*(0)$) components, with the result that the first q components of $\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} w_{n,t}$ will converge weakly to a (nondegenerate) limiting process, whereas the final r components will converge to zero, exactly as in the manner of (3.1). $\frac{1}{n} \sum_{t=1}^n w_{n,t} w_{n,t}^\top$ then also converges to an (invertible) block diagonal matrix, as in (3.2).

By Proposition 3.1, the sum of the first q_0 generalised eigenvalues of \mathbb{A}_n with respect to \mathbb{B}_n will then exhibit divergent asymptotic behaviour, depending on whether q_0 is equal to or strictly greater than q . This provides the basis for the use of this quantity as a statistic for testing hypotheses regarding the value of q , exactly as proposed in Breitung [14]. Since these generalised eigenvalues are invariant to common linear transformations of \mathbb{A}_n and \mathbb{B}_n , and $w_{n,t}$ is a linear transformation of z_t , they may be computed without knowledge of $[\beta_\perp, \beta]$, simply by replacing each instance of $w_{n,t}$ by z_t in the definitions of those matrices.

3.2 | Extension to Nonlinearly Cointegrated Series

Suppose that we now permit $\beta^+ \neq \beta^-$ and/or $\Gamma^+(1) \neq \Gamma^-(1)$. In this case, $P_{\beta_\perp}(-1)$ and $P_{\beta_\perp}(+1)$ each have rank q , but may differ by a rank one matrix, and as a result, there may only be $r - 1$ distinct linear combinations of z_t that will be $I^*(0)$. Accordingly, applying the usual Breitung test to $\{z_t\}$ directly would tend to yield the incorrect conclusion that there are $q + 1$ common trends, rather than only q . (Thus, e.g., in a bivariate nonlinear SVAR with one

common nonlinear trend, this test may tend to conclude that there are *two* common trends and *no* cointegrating relations.)

To address this problem, here we utilise the fact that the non-linearity in the CKSVAR is entirely a function of the sign of the first component of $z_t = (y_t, x_t^\top)^\top$, such that the nonlinear cointegrating relationships $\beta(y)$ can be rewritten as *linear* cointegrating relationships between the elements of

$$z_t^* := \begin{bmatrix} y_t^+ \\ y_t^- \\ x_t \end{bmatrix} = \begin{bmatrix} \mathbf{1}^+(y_t) & 0 \\ \mathbf{1}^-(y_t) & 0 \\ 0 & I_{p-1} \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = S_p(y_t)z_t$$

via

$$\beta(y) = \begin{bmatrix} \beta_y^{+\top} \mathbf{1}^+(y) + \beta_y^{-\top} \mathbf{1}^-(y) \\ \beta_x \end{bmatrix} = \begin{bmatrix} \mathbf{1}^+(y) & \mathbf{1}^-(y) & 0 \\ 0 & 0 & I_{p-1} \end{bmatrix} \begin{bmatrix} \beta_y^{+\top} \\ \beta_y^{-\top} \\ \beta_x \end{bmatrix} \\ =: S_p(y)^\top \beta^* \tag{3.3}$$

from which it follows that

$$\beta(y_t)^\top z_t = \beta^{*\top} S_p(y_t)z_t = \beta^{*\top} z_t^*$$

since $z_t^* = S_p(y_t)z_t$; the r.h.s. thus gives the r linear relationships that render $\beta^{*\top} z_t^* \sim I^*(0)$. As a corollary, there will be $q + 1$ (linearly independent) vectors in \mathbb{R}^{p+1} that extract distinct $I^*(1)$ components from z_t^* . We obtain an additional $I^*(1)$ component, because under case (ii) the common trends are present in both y_t^+ and y_t^- , which appear separately as the first two components of z_t^* .

In extracting those common trends, we are free to choose any $(q + 1)$ -dimensional basis in \mathbb{R}^{p+1} whose span does not (non-trivially) intersect with $\text{sp}\beta^*$. Here we take this basis to be the columns of the following $(p + 1) \times (q + 1)$ matrix

$$\tau^* := \begin{bmatrix} 1 & 0 & \tau_{xy}^{+\top} \\ 0 & 1 & \tau_{xy}^{-\top} \\ 0 & 0 & \beta_{x,\perp} \end{bmatrix}, \tag{3.4}$$

where the columns of $\beta_{x,\perp} \in \mathbb{R}^{(p-1) \times (q-1)}$ span the orthogonal complement of $\text{sp}\beta_x$ in \mathbb{R}^{p-1} , and as shown in the proof of Theorem 3.2 (see Lemma A4, in particular), we are free to choose $\tau_{xy}^\pm \in \mathbb{R}^{q-1}$ so as to facilitate the convergence of our test statistic to a pivotal limiting distribution. The matrix τ^* plainly has rank $q + 1$; moreover the $(p + 1) \times (p + 1)$ matrix $[\beta^*, \tau^*]$ is nonsingular, irrespective of the values of τ_{xy}^\pm (see Lemma A3).

Thus, the linear transformation

$$T_n^\top z_t^* := \begin{bmatrix} n^{-1/2} \tau^{*\top} \\ \beta^{*\top} \end{bmatrix} z_t^* = \begin{bmatrix} n^{-1/2} \tau^{*\top} z_t^* \\ \beta^{*\top} z_t^* \end{bmatrix} = \begin{bmatrix} n^{-1/2} \tau^{*\top} z_t^* \\ \beta(y_t)^\top z_t \end{bmatrix} \\ =: \begin{bmatrix} n^{-1/2} \varrho_t \\ \xi_t \end{bmatrix} =: \begin{bmatrix} \varrho_{n,t} \\ \xi_t \end{bmatrix} \tag{3.5}$$

exhaustively separates z_t^* into its $I^*(0)$ and (appropriately standardised) $I^*(1)$ components, and so renders the process $\{z_t^*\}$ into

a form conformable with (3.1) above. The decomposition (3.5) provides the basis for applying what we term our *modified Breitung* (MB) test to the data generated by a cointegrated CKSVAR, under case (ii), ‘modified’ in the sense that the test statistic will be constructed from z_t^* rather than z_t . Indeed, if $c = 0$, then it will follow from our results below that $\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t \rightsquigarrow 0$ on $D[0, 1]$, and so the test could be applied directly to z_t^* in this case. More generally, when $c \neq 0$, we need to first extract any deterministic components whose presence would otherwise distort the distribution of the test statistic. If we suppose that DET holds, then no deterministic trends are present in z_t , and by analogy with the approach taken in the linear setting, we may project out any constant deterministic terms by applying the test not to z_t^* but rather to

$$\bar{z}_t^* := z_t^* - \hat{\mu}_{n,z^*}$$

where $\hat{\mu}_{n,z^*} := \frac{1}{n} \sum_{t=1}^n z_t^*$, so that now

$$T_n^\top \bar{z}_t^* = T_n^\top (z_t^* - \hat{\mu}_{n,z^*}) = \begin{bmatrix} \varrho_{n,t} - \hat{\mu}_{n,\varrho} \\ \xi_t - \hat{\mu}_{n,\xi} \end{bmatrix} =: \begin{bmatrix} \bar{\varrho}_{n,t} \\ \bar{\xi}_t \end{bmatrix} \tag{3.6}$$

where $\hat{\mu}_{n,\xi} := \frac{1}{n} \sum_{t=1}^n \xi_t$ and $\hat{\mu}_{n,\varrho} := \frac{1}{n} \sum_{t=1}^n \varrho_{n,t}$.

To obtain the limiting distribution of our proposed test, we shall verify that $w_{n,t} = T_n^\top \bar{z}_t^*$ satisfies the requirements of Proposition 3.1 above. For (3.6) to conform with (3.1), we must show that

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{\xi}_t = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t - \lambda \hat{\mu}_{n,\xi} = o_p(1)$$

uniformly in $\lambda \in [0, 1]$. Similarly, for the purposes of (3.2), we require that $\frac{1}{n} \sum_{t=1}^n \bar{\xi}_t \bar{\xi}_t^\top$ converges weakly to an (a.s.) positive definite matrix. In other words, we require a fundamental law of large numbers (LLN) for sample averages of the form $\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(\xi_t)$. Since $\{\xi_t\}$ is not, in general, a stationary process, existing results do not apply here, and this motivates the development of the novel LLN given as Theorem 3.1 below.

3.3 | LLN for Regime-Switching Processes

To illustrate the essential ideas, suppose for simplicity of exposition that $k = 1$, and that the CKSVAR is canonical. Then by Lemma B.2 of DMW25, $\xi_t = \beta(y_t)^\top z_t$ admits the time-varying autoregressive representation

$$\xi_t = \beta_t^\top c + (I_r + \beta_t^\top \alpha) \xi_{t-1} + \beta_t^\top u_t, \tag{3.7}$$

where $\{\beta_t\}$ is a random sequence that in general depends, nonlinearly, on the values of y_t and y_{t-1} . Under CO(ii).2, which implies that $I_r + \beta_t^\top \alpha$ is drawn from a set of matrices whose joint spectral radius is strictly bounded by unity, $\{\xi_t\}$ will be a ‘stable’ process in the sense that it is stochastically bounded; but the dependence of β_t on y_t prevents $\{\xi_t\}$ from being stationary.

Since $\beta_t = \beta^+$ whenever $y_{t-1} > 0$ and $y_t > 0$, it follows that if $y_s > 0$ for all $s \in \{t - m, \dots, t\}$, then

$$\xi_t = (I_r + \beta^{+\top} \alpha)^m \xi_{t-m} + \sum_{\ell=0}^{m-1} (I_r + \beta^{+\top} \alpha)^\ell \beta^{+\top} (c + u_{t-\ell}).$$

Since $\{y_t\}$ has a stochastic trend, it will tend to make lengthy sojourns above the origin, during which periods ξ_t will be well approximated by the stationary linear process,

$$\xi_t^+ := -(\beta^{+\top} \alpha)^{-1} \beta^{+\top} c + \sum_{\ell=0}^{\infty} (I_r + \beta^{+\top} \alpha)^{\ell} \beta^{+\top} u_{t-\ell} =: \mu_{\xi}^+ + w_t^+.$$

On the other hand, $\{y_t\}$ will also tend to spend lengthy epochs below the origin, permitting ξ_t to then be approximated by

$$\xi_t^- := -(\beta^{-\top} \alpha)^{-1} \beta^{-\top} c + \sum_{\ell=0}^{\infty} (I_r + \beta^{-\top} \alpha)^{\ell} \beta^{-\top} u_{t-\ell} =: \mu_{\xi}^- + w_t^-.$$

This reasoning suggests a kind of ‘dual linear process’ approximation to ξ_t , leading to an argument along the lines of

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(\xi_t) \mathbf{1}^+(y_t) &= \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(\xi_t^+) \mathbf{1}^+(y_t) + o_p(1) \\ &= [\mathbb{E}g(\xi_0^+)] \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \mathbf{1}^+(y_t) + o_p(1) \\ &\rightsquigarrow [\mathbb{E}g(\xi_0^+)] \int_0^{\lambda} \mathbf{1}^+[Y(s)] ds =: [\mathbb{E}g(\xi_0^+)] m_Y^+(\lambda) \end{aligned}$$

where $m_Y^+(\lambda)$ measures the fraction of the interval $[0, \lambda]$ for which $Y(s) \geq 0$. We thus arrive at

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(\xi_t) &= \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(\xi_t) [\mathbf{1}^+(y_t) + \mathbf{1}^-(y_t)] \\ &\rightsquigarrow [\mathbb{E}g(\xi_0^+)] m_Y^+(\lambda) + [\mathbb{E}g(\xi_0^-)] m_Y^-(\lambda), \end{aligned}$$

which will, in general, be random (so that the convergence is merely in distribution), except in the special case where $\mathbb{E}g(\xi_0^+) = \mathbb{E}g(\xi_0^-) = \mu_g$ —whereupon the r.h.s. collapses to $\lambda \mu_g$, since $m_Y^+(\lambda) + m_Y^-(\lambda) = \lambda$. (Importantly for the purposes of our test, such a case systematically arises under our assumptions, when $g(\xi) = \xi$.) The randomness of the limit provides another manifestation of the non-ergodicity of $\{\xi_t\}$, as induced by the dependence of its law of motion on the level of y_t .

Such arguments, in the more general setting of a (not necessarily canonical) CKSVAR(k), lead to the main technical contribution of this paper, a LLN-type result for additive functionals of a class of time-varying autoregressive processes, of which (3.7) is a special case. To facilitate its use in other contexts, we prove this result supposing that the following weaker condition holds in place of DET.

Assumption DET’. $e_1^\top P_{\beta_\perp} (+1)c = 0$.

The preceding permits the model to impart deterministic trends to x_t (but not to y_t), and leads us to consider the linearly detrended process

$$\begin{bmatrix} y_t^d \\ x_t^d \end{bmatrix} = z_t^d := z_t - [P_{\beta_\perp} (+1)c]t, \quad t \geq 1$$

in place of z_t , with the convention that $z_t^d := z_t$ for $t \leq 0$; note that $y_t^d = y_t$ (see Section 4.4 in DMW25). Recall that, as per the remarks following the statement of DGP above, there is an

underlying filtration $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ to which $\{u_t\}$ and $\{z_t\}$ are adapted, and that an i.i.d. process $\{v_t\}$ is one that is both \mathcal{F}_t -adapted, and such that v_s is independent of \mathcal{F}_t for $s > t$.

Theorem 3.1. *Suppose DGP, CVAR, CO(ii) and DET’ hold. Let $\{A_t\}$, $\{B_t\}$ and $\{c_t\}$ be random sequences adapted to $\{\mathcal{F}_t\}$, respectively taking values in $\mathbb{R}^{d_w \times d_w}$, $\mathbb{R}^{d_w \times d_v}$ and \mathbb{R}^{d_w} , where $t \in \mathbb{Z}$. Suppose $\{v_t\}$ is i.i.d with $\mathbb{E}v_t = 0$, and that $\{w_t\}$ satisfies*

$$w_t = c_t + A_t w_{t-1} + B_t v_t \tag{3.8}$$

for $t \geq 1$ and some given (random) w_0 (with $w_t := 0$ for all $t \leq -1$); and:

- i. $A_t \in \mathcal{A}$, $B_t \in \mathcal{B}$ and $c_t \in \mathcal{C}$ for all $t \in \mathbb{N}$, where \mathcal{A} , \mathcal{B} and \mathcal{C} are bounded subsets of $\mathbb{R}^{d_w \times d_w}$, $\mathbb{R}^{d_w \times d_v}$ and \mathbb{R}^{d_w} respectively, and $\rho_{\text{JSR}}(\mathcal{A}) < 1$;
- ii. there exist $A^\pm \in \mathcal{A}$, $B^\pm \in \mathcal{B}$ and $c^\pm \in \mathcal{C}$ such that
 - $y_{t-1} > 0$ and $y_t > 0 \Rightarrow A_t = A^+, B_t = B^+, c_t = c^+$,
 - $y_{t-1} < 0$ and $y_t < 0 \Rightarrow A_t = A^-, B_t = B^-, c_t = c^-$;
- iii. $m_0 \geq 1$ is such that $\|w_0\|_{m_0} + \|v_0\|_{m_0} < \infty$.
- iv. $g : \mathbb{R}^{d_w} \rightarrow \mathbb{R}^{d_s}$ is a continuous function satisfying

$$\|g(w) - g(w')\| \leq C(1 + \|w\|^{\ell_0} + \|w'\|^{\ell_0}) \|w - w'\| \tag{3.9}$$

for all $w, w' \in \mathbb{R}^{d_w}$, for some $0 \leq \ell_0 < m_0 - 1$.

Then $\mathbb{E}\|g(w_0^\pm)\| < \infty$, and on $D[0, 1]$,

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) \mathbf{1}^\pm(y_t) \rightsquigarrow [\mathbb{E}g(w_0^\pm)] \int_0^\lambda \mathbf{1}^\pm[Y(\mu)] d\mu, \tag{3.10}$$

where

$$w_0^\pm = (I_{d_w} - A^\pm)^{-1} c^\pm + \sum_{\ell=0}^{\infty} (A^\pm)^\ell B^\pm v_{-\ell}. \tag{3.11}$$

Moreover,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{\lfloor n\lambda \rfloor} [g(w_t) \otimes z_t^d] \mathbf{1}^\pm(y_t) \rightsquigarrow [\mathbb{E}g(w_0^\pm)] \otimes \int_0^\lambda Z(\mu) \mathbf{1}^\pm[Y(\mu)] d\mu, \tag{3.12}$$

jointly with $U_n \rightsquigarrow U$.

3.4 | Limiting Distribution and Consistency

Using Theorem 3.1 and the representation theory of DMW25, we are able to derive the limiting distribution of our modified Breitung (MB) statistic for testing the null of q_0 common trends (and $r_0 = p - q_0$ cointegrating relations), which is defined as

$$\Lambda_{n,q_0} := n^2 \sum_{i=1}^{q_0+1} \lambda_{n,i} \tag{3.13}$$

where $\{\lambda_{n,i}\}_{i=1}^{p+1}$ are the solutions to

$$\det(\lambda \mathbf{B}_n - \mathbf{A}_n) = 0 \tag{3.14}$$

ordered as $\lambda_{n,1} \leq \lambda_{n,2} \leq \dots \leq \lambda_{n,p+1}$, for

$$\mathbf{A}_n := \sum_{i=1}^n \bar{z}_i^* \bar{z}_i^{*\top}, \quad \mathbf{B}_n := \sum_{i=1}^n \sum_{j=1}^i \bar{z}_i^* \bar{z}_j^{*\top}. \quad (3.15)$$

This statistic has the same form as that considered in Proposition 3.1, though note that for testing the null of q_0 common trends we sum over the first $q_0 + 1$ generalised eigenvalues $\{\lambda_{n,i}\}_{i=1}^{q_0+1}$, reflecting the fact that y_i^+ and y_i^- separately enter z_i^* .

To state the limiting distribution of the test statistic, define

$$W_0(\lambda) := \mathcal{W}_0 e_{q,1} + W(\lambda), \quad (3.16)$$

where $\mathcal{W}_0 \in \mathbb{R}$ is nonrandom, and W is a q -dimensional standard Brownian motion. Define the $(q + 1)$ -dimensional process

$$W_0^*(\lambda) := S_q [e_1^\top W_0(\lambda)] W_0(\lambda) =: \begin{bmatrix} [W_{0,1}(\lambda)]_+ \\ [W_{0,1}(\lambda)]_- \\ W_{0,-1}(\lambda) \end{bmatrix} \quad (3.17)$$

and define $\bar{W}_0^*(\lambda)$ to be the residual from the pathwise $L^2[0, 1]$ projection of each element of W_0^* onto a constant. Let $\bar{V}_0^*(\lambda) := \int_0^\lambda \bar{W}_0^*(\mu) d\mu$ denote the cumulation of \bar{W}_0^* .

We only provide limit theory here for the case where $y_0 = o_p(n^{1/2})$. This simplifies the asymptotics of the testing problems in two respects: (i) it ensures that the limiting process visits both regimes (positive and negative) with probability one, so that the relevant matrices are positive definite a.s.; (ii) it yields a distribution for the test statistic that (upon demeaning) is nuisance parameter free, being invariant to $X(0) = \mathcal{X}_0$. (Possible extensions to handle the case where $n^{-1/2}y_0 \xrightarrow{p} \mathcal{Y}_0 \neq 0$ are discussed below.) In the following statement, q denotes the actual (i.e., the true) number of common trends in the system, whereas q_0 denotes the null hypothesised value, i.e., the number used to compute the test statistic.

Theorem 3.2. *Suppose DGP, CVAR, CO(ii) and DET hold, with $y_0 = o_p(n^{1/2})$. Then for W_0 as defined in (3.16), with $\mathcal{W}_0 = 0$:*

i. if $q_0 = q$,

$$\begin{aligned} \Lambda_{n,q_0} &= \Lambda_{n,q} \\ &\rightsquigarrow \text{tr} \left[\int_0^1 \bar{W}_0^*(s) \bar{W}_0^*(s)^\top ds \left(\int_0^1 \bar{V}_0^*(s) \bar{V}_0^*(s)^\top ds \right)^{-1} \right] \\ &=: \Lambda_q \end{aligned} \quad (3.18)$$

ii. if $q_0 < q$, the weak limit of Λ_{n,q_0} is stochastically dominated by Λ_q ; and

iii. if $q_0 > q$, $\Lambda_{n,q_0} \xrightarrow{p} \infty$.

Moreover, the convergence in (3.18) holds jointly with $U_n \rightsquigarrow U$, and with

$$n^{-1/2}y_{[n\lambda]} \rightsquigarrow Y(\lambda) = \omega^+ [e_1^\top W_0(\lambda)]_+ + \omega^- [e_1^\top W_0(\lambda)]_-, \quad (3.19)$$

where the latter convergence also holds if $n^{-1/2}y_0 \xrightarrow{p} \mathcal{Y}_0$ with \mathcal{Y}_0 possibly nonzero.

Part (i) of the preceding implies that valid asymptotic critical values for $H_0 : q = q_0$ can be drawn from the distribution of Λ_{q_0} (which equals Λ_q under H_0); these may be computed by simulation. Part (ii) implies that Λ_{n,q_0} is stochastically bounded when the true number of common trends (q) is greater than the hypothesised number (q_0), such that a test of $H_0 : q = q_0$ will not be consistent against the alternative $H_1 : q > q_0$. On the other hand, by part (iii), it will be consistent against $H_1 : q < q_0$. This suggests that the estimation of q may be effected via a stepwise testing procedure, starting with the null $H_0 : q = p$ of no cointegration, and progressing downwards (i.e., testing $H_0 : q = p - 1$ if the preceding null is rejected, etc., and stopping at the first q_0 for which $H_0 : q = q_0$ is not rejected).

3.5 | Extensions

Once we allow that $n^{-1/2}y_0 \xrightarrow{p} \mathcal{Y}_0$, with \mathcal{Y}_0 possibly nonzero, the preceding runs into certain difficulties. If $\mathcal{Y}_0 = 0$, then $\mathcal{W}_0 = 0$ also, and so $W_{0,1}$ visits both sides of the origin at some point during $[0, 1]$ (indeed, during any subinterval $[0, \lambda]$) with probability one. But if $\mathcal{Y}_0 \neq 0$ then $\mathcal{W}_0 \neq 0$, and this event is no longer guaranteed to occur, with the consequence that $\int \bar{W}_0^* \bar{W}_0^{*\top}$ and $\int \bar{V}_0^* \bar{V}_0^{*\top}$ are no longer positive definite with probability one. In a sense, this is merely a technical rather than a practical problem, because the failure of $W_{0,1}$ to visit both sides of the origin is the large-sample counterpart of the possibility that $\{y_i\}$ itself may not visit both sides of the origin, and were it to fail to do so, the observed data would be well (indeed, perfectly) approximated by a linearly cointegrated system, with cointegrating relations given by either β^+ or β^- (depending on whether $\{y_i\}_{i=1}^n$ was always positive or negative, respectively).

The fact that we would only contemplate conducting (the modified version of) the test in cases where $\{y_i\}$ spends an appreciable amount of time in both regimes also suggests a remedy for this problem. Namely, that we should refer the test statistic Λ_{n,q_0} not to the quantiles of its unconditional limiting distribution, but to those of its distribution conditional on $\{y_i\}$ (and therefore $W_{0,1}$) spending more than a certain fraction of the sample in each regime; this thereby avoids the rank deficiency problem. That is, letting $\mathcal{M} := \min\{m_{W_{0,1}}^+(1), m_{W_{0,1}}^-(1)\}$, we propose to compare Λ_{n,q_0} with the $1 - \alpha$ quantile of the distribution of Λ_{q_0} conditional on $\mathcal{M} \geq \tau$, i.e., choosing a critical value $c_{\alpha,q_0}(\tau)$ such that

$$\mathbb{P}\{\Lambda_{q_0} \geq c_{\alpha,q_0}(\tau) | \mathcal{M} \geq \tau\} = \alpha \quad (3.20)$$

where $\tau \in (0, 0.5)$ is some user-specified value (say, 10% or 15%).

The preceding remains well defined when $\mathcal{Y}_0 \neq 0$, but in that case the (conditional) distribution of Λ_{q_0} will depend on the unknown nuisance parameter \mathcal{W}_0 . Since the sign of y_0 and therefore $Y(0) = \mathcal{Y}_0$ is known, \mathcal{W}_0 may be estimated when (say) $y_0 > 0$ on the basis of the representation (3.19) as $(\hat{\omega}_n^+)^{-1}(n^{-1/2}y_0)$, where

$$\begin{aligned} \hat{\omega}_n^+ &:= \left(\sum_{\ell=-L_n}^{L_n} K(\ell/L_n) \hat{\gamma}_\ell^+ \right)^{1/2} \\ \hat{\gamma}_\ell^+ &:= \frac{1}{\sum_{t=1}^n \mathbf{1}^+(y_t)} \sum_{t=\ell+1}^n \Delta y_t \Delta y_{t-\ell} \mathbf{1}^+(y_t) \end{aligned}$$

TABLE 1 | Rejection rates; nominal level 10%.

Design	Linear ($\beta^+ = \beta^-, q = 1$)				Nonlinear ($\beta^+ \neq \beta^-, q = 1$)				Stationary ($q = 0$)	
	$q = 1$		$q = 2$		$q = 1$		$q = 2$		$q = 1$	
	SB	MB	SB	MB	SB	MB	SB	MB	SB	MB
$H_0 : n$										
200	0.09	0.06	0.94	0.68	0.06	0.02	0.57	0.36	0.40	0.38
500	0.09	0.09	1.00	0.95	0.08	0.05	0.64	0.75	0.71	0.81
1000	0.10	0.10	1.00	1.00	0.08	0.08	0.61	0.94	0.91	0.98
1500	0.10	0.10	1.00	1.00	0.08	0.08	0.58	0.98	0.97	1.00

denotes a long-run variance estimator, with kernel K and lag truncation sequence $L_n \rightarrow \infty$. (If on the other hand $y_0 < 0$, then an estimator $\hat{\omega}_n^-$ of ω^- would be constructed analogously.)

4 | Finite-Sample Performance

Here we report the results of Monte Carlo simulations conducted to evaluate the performance of the proposed test. We generate data from a bivariate (i.e., $p = 2$) cointegrated CKSVAR with $q = 1$ common trends (and so $r = 1$ cointegration relations),

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = c + \alpha \beta^{*\top} \begin{bmatrix} y_{t-1}^+ \\ y_{t-1}^- \\ x_{t-1} \end{bmatrix} + u_t,$$

where $\alpha = (0.5, 0.1)^\top$, $\beta^* = (\beta_y^+, \beta_y^-, 1)^\top$, $c = 2\alpha$, $z_0 = (y_0, x_0)^\top = 0$ and $u_t \sim_{i.i.d.} N[0, I_2]$. We set $\beta_y^+ = -1$, and consider a *linear* design in which $\beta_y^- = -1$, and a *nonlinear* design in which $\beta_y^- = -0.5$. The implied cointegrating vectors are $\beta^+ = \beta^- = (-1, 1)^\top$ in the former, and $\beta^+ = (-1, 1)^\top$ and $\beta^- = (-0.5, 1)^\top$ in the latter. In both cases, the assumptions of Theorem 3.2 are satisfied; e.g., it may be verified that $|1 + \beta^{\pm\top} \alpha| < 1$, so that the stability condition CO(ii).2 holds. The sample size ranges over $n \in \{200, 500, 1000, 1500\}$. We only retain samples in which $\{y_t\}$ spends at least $0.15n$ observations both above and below zero.

For each dataset thus generated, we test the null that $H_0 : q = q_0$ using the following test statistics:

- The *standard Breitung* (SB) test is that given in Breitung [14, Sec. 5]. In this case, \mathbf{A}_n and \mathbf{B}_n in (3.15) are computed on the basis of \bar{z}_t , rather than \bar{z}_t^* ;
- The *modified Breitung* (MB) test is our proposed test statistic, based on \bar{z}_t^* , and using a ‘partially conditional’ critical value $c_{\alpha, q_0}(\tau)$ as in (3.20) with $\tau = 0.15$.

(Note that to test the null that $H_0 : q = q_0$, SB sums over the first q_0 generalised eigenvalues of a p -dimensional system, whereas MB sums over the first $q_0 + 1$ generalised eigenvalues of a $(p + 1)$ -dimensional system.) Let q denote the true number of common trends. Since the true number of common trends $q = 1$ in the foregoing designs, we test $H_0 : q = 1$ to evaluate size and $H_0 : q = 2$ to evaluate power, with a nominal significance level of 10 per cent. (We run 10,000 Monte Carlo replications for every design.)

The results are displayed in the first eight columns of Table 1. In line with our expectations, the standard Breitung test performs poorly in the nonlinear design, having a noticeable tendency to incorrectly find that $q = 2$. This problem is remedied by the modified Breitung test, at least for sufficiently large sample sizes, at the cost of the test being somewhat conservative in small samples. Both tests appear to be approximately correctly sized for testing $H_0 : q = 1$, and both (as expected) perform well in the linear design.

As an additional check on the performance of these tests, per a request from a referee, we also evaluated their power to reject $H_0 : q = 1$ using data generated under the following *stationary* ($q = 0$) nonlinear design,

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = c + \begin{bmatrix} \pi^+ & \pi^- & \Pi^x \end{bmatrix} \begin{bmatrix} y_{t-1}^+ \\ y_{t-1}^- \\ x_{t-1} \end{bmatrix} + u_t \\ = \begin{bmatrix} -0.2 & -0.5 & 0.3 \\ 0.1 & 0.1 & -0.2 \end{bmatrix} \begin{bmatrix} y_{t-1}^+ \\ y_{t-1}^- \\ x_{t-1} \end{bmatrix} + u_t.$$

The results are reported in the final two columns of Table 1, and show that both the SB and MB tests have substantial power in this direction. Since the ‘cointegrating space’ is $\{0\}$ in this case, and so trivially linear, the similar performance of the two tests is not surprising.

5 | Conclusion

This paper has considered the problem of testing the cointegrating rank in a CKSVAR, proposing a modified version of the Breitung [14] test that is robust to the forms of nonlinear cointegration that may be generated by that model. En route to deriving the asymptotics of this test, we have proved a novel LLN-type result for a class of stable but nonstationary autoregressive processes. This result underpins the development of the asymptotics of likelihood-based estimators of the cointegrated CKSVAR, our results on which will be reported elsewhere.

Funding

The second author gratefully acknowledges financial support from the Tore Browaldh Foundation.

Endnote

¹Throughout the following, the notation ‘ a^\pm ’ connotes a^+ and a^- as objects associated respectively with y_t^+ and y_t^- , or their lags. If we want to instead denote the positive and negative parts of some $a \in \mathbb{R}$, we shall do so by writing $[a]_+ := \max\{a, 0\}$ or $[a]_- := \min\{a, 0\}$.

References

1. C. W. J. Granger, “Developments in the Study of Cointegrated Economic Variables,” *Oxford Bulletin of Economics and Statistics* 48 (1986): 213–228.
2. R. F. Engle and C. W. J. Granger, “Co-Integration and Error Correction: Representation, Estimation, and Testing,” *Econometrica* 55 (1987): 251–276.
3. S. Johansen, “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica* 59 (1991): 1551–1580.
4. S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press, 1995).
5. D. Tjøstheim, “Some Notes on Nonlinear Cointegration: A Partial Review With Some Novel Perspectives,” *Econometric Reviews* 39 (2020): 655–673.
6. H. Tong, *Non-Linear Time Series: A Dynamical System Approach* (Oxford University Press, 1990).
7. T. Teräsvirta, D. Tjøstheim, and C. W. J. Granger, *Modelling Nonlinear Economic Time Series* (Oxford University Press, 2010).
8. D. Kristensen and A. Rahbek, “Likelihood-Based Inference for Cointegration With Nonlinear Error-Correction,” *Journal of Econometrics* 158 (2010): 78–94.
9. S. Mavroeidis, “Identification at the Zero Lower Bound,” *Econometrica* 89 (2021): 2855–2885.
10. S. B. Aruoba, M. Mlikota, F. Schorfheide, and S. Villalvazo, “SVARs With Occasionally-Binding Constraints,” *Journal of Econometrics* 231 (2022): 477–499.
11. D. Ikeda, S. Li, S. Mavroeidis, and F. Zanetti, “Testing the Effectiveness of Unconventional Monetary Policy in Japan and the United States,” *American Economic Journal: Macroeconomics* 16 (2024): 250–286.
12. J. A. Duffy, S. Mavroeidis, and S. Wycherley, “Cointegration With Occasionally Binding Constraints,” *Journal of Econometrics* 252 (2025): 106103.
13. J. A. Duffy and S. Mavroeidis, “Common Trends and Long-Run Identification in Nonlinear Structural VARs,” arXiv:2404.05349, (2024).
14. J. Breitung, “Nonparametric Tests for Unit Roots and Cointegration,” *Journal of Econometrics* 108 (2002): 343–363.
15. J. A. Duffy, S. Mavroeidis, and S. Wycherley, “Stationarity With Occasionally Binding Constraints,” arXiv:2307.06190, (2023).
16. R. M. Jungers, *The Joint Spectral Radius: Theory and Applications* (Springer, 2009).
17. I. Berkes and L. Horváth, “Convergence of Integral Functionals of Stochastic Processes,” *Econometric Theory* 22 (2006): 304–322.
18. P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application* (Academic Press, 1980).
19. D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed. (Springer, 1999).

Appendix A

Auxiliary Lemmas

We here collect the fundamental technical results that are needed for the proof of Theorems 3.1 and 3.2. These are all stated for a CKSVAR in

canonical form, i.e., supposing that DGP* holds. For a general CKSVAR, i.e., one satisfying DGP rather than DGP*, Proposition 2.1 in DMW25 establishes that there is a linear mapping between z_t^* and a derived canonical process \tilde{z}_t^* satisfying DGP*. Because $\Lambda_{n,q}$ is invariant to (common) linear transformations of \mathbf{A}_n and \mathbf{B}_n , as defined in (3.15), the asymptotics of the canonical process accordingly govern the large-sample behaviour of our test statistic.

We first recall that under DGP*, CVAR, CO(ii) and DET’, it follows by Theorems 4.2 and 4.4 of DMW25 that

$$n^{-1/2} \begin{bmatrix} Y_{[n\lambda]} \\ X_{[n\lambda]}^d \end{bmatrix} = n^{-1/2} z_{[n\lambda]}^d \rightsquigarrow P_{\beta_1} [Y(\lambda)] U_0(\lambda) = \begin{bmatrix} Y(\lambda) \\ X(\lambda) \end{bmatrix} = Z(\lambda) \quad (A1)$$

on $D[0, 1]$ with the further implication (via Lemma B.3 of DMW25) that

$$Y(\lambda) = h[\vartheta^\top U_0(\lambda)] \vartheta^\top U_0(\lambda) \quad (A2)$$

where $h(u) = \mathbf{1}^+(u)h^+ + \mathbf{1}^-(u)h^-$ for $h^+ = 1$ and $h^- > 0$, and $\vartheta^\top := e^\top P_{\beta_1} (+1) \neq 0$. We note that as a consequence of (A1), CO(ii).4 and our (innocuous) convention that $\Delta z_i = 0$ for $i \leq -k$ (as per (2.14) above) that

$$n^{-1/2} \sup_{s \leq n} \|z_s^d\| = O_p(1), \quad n^{-1/2} \sup_{s \leq n} \|\Delta z_s^d\| = o_p(1). \quad (A3)$$

Indeed, it follows by Lemmas A.1 and B.2 of DMW25 that

$$\sup_{s \in \mathbb{Z}} \|\Delta z_s^d\|_{2+\delta_n} < \infty. \quad (A4)$$

(Recall that for X a random vector, and $p \geq 1$, $\|X\|_p := (\mathbb{E}\|X\|^p)^{1/p}$.)

Lemma A1. *Suppose DGP*, CVAR, CO(ii) and DET’ hold. Then:*

i. *as $n \rightarrow \infty$ and then $\delta \rightarrow 0$*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{n^{-1/2}|y_i| \leq \delta\} \xrightarrow{p} 0;$$

ii. *on $D[0, 1]$ jointly with $U_n \rightsquigarrow U$,*

$$\frac{1}{n} \sum_{i=1}^{[n\lambda]} \mathbf{1}^\pm(y_i) \begin{bmatrix} 1 \\ n^{-1/2} z_i^d \end{bmatrix} \rightsquigarrow \int_0^\lambda \mathbf{1}^\pm[Y(\mu)] \begin{bmatrix} 1 \\ Z(\mu) \end{bmatrix} d\mu.$$

The following is a slightly restricted counterpart of Theorem 3.1, which holds under DGP* rather than DGP. It will in turn be used to prove Theorem 3.1 in Appendix B.

Lemma A2. *Suppose DGP*, CVAR, CO(ii) and DET’ hold. Then the conclusions of Theorem 3.1 hold.*

For the next two results, we specialise from DET’ to DET, so that no deterministic trends are present in any components of z_t , which is identically equal to z_t^d . Recall the definitions of $\bar{\rho}_{n,t}$ and $\bar{\xi}_t$ given in (3.6). We note also that as an immediate consequence of (A1) and the continuous mapping theorem, on $D[0, 1]$,

$$n^{-1/2} z_{[n\lambda]}^* = S_p(n^{-1/2} y_{[n\lambda]}) n^{-1/2} z_{[n\lambda]} \rightsquigarrow S_p[Y(\lambda)] Z(\lambda) =: Z^*(\lambda) \quad (A5)$$

for $S_p(y)$ as in (2.13), and hence

$$\varrho_{n,[n\lambda]} = \tau^{*\top} n^{-1/2} z_{[n\lambda]}^* \rightsquigarrow \tau^{*\top} Z^*(\lambda) =: R(\lambda), \quad (A6)$$

for τ^* as in (3.4). Since $z_t^* = (y_t^+, y_t^-, x_t^{\top})^\top$ can be written as a linear function of elements of $(y_t^+, x_t^{\top})^\top$ and $(y_t^-, x_t^{\top})^\top$, it follows from Lemma A2 and the continuous mapping theorem, under the conditions of Lemma A2 and DET that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{[n\lambda]} \left(g(w_i) \otimes \begin{bmatrix} 1 \\ n^{-1/2} z_i^* \end{bmatrix} \right) \mathbf{1}^\pm(y_i) \\ & \rightsquigarrow [\mathbb{E}g(w_0^*)] \otimes \int_0^\lambda \begin{bmatrix} 1 \\ Z^*(\mu) \end{bmatrix} \mathbf{1}^\pm[Y(\mu)] d\mu \end{aligned} \quad (A7)$$

on $D[0, 1]$, jointly with $U_n \rightsquigarrow U$.

Lemma A3. Suppose DGP^* , $CVAR$, $CO(ii)$ and DET hold. Then

- i. for all $\tau_{xy}^\pm \in \mathbb{R}^{q-1}$, the matrix $[\beta^*, \tau^*]$ is nonsingular;
- ii. on $D[0, 1]$,

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \begin{bmatrix} \bar{Q}_{n,t} \\ \bar{\xi}_t \end{bmatrix} \rightsquigarrow \begin{bmatrix} \int_0^\lambda \bar{R}(s) \, ds \\ 0 \end{bmatrix} \quad (A8)$$

where

$$\bar{R}(s) := R(s) - \int_0^1 R(\lambda) \, d\lambda;$$

- iii. there exist positive definite matrices Σ_{ξ^+} and Σ_{ξ^-} such that

$$\frac{1}{n} \sum_{t=1}^n \begin{bmatrix} \bar{Q}_{n,t} \bar{Q}_{n,t}^\top & \bar{Q}_{n,t} \bar{\xi}_t^\top \\ \bar{\xi}_t \bar{Q}_{n,t}^\top & \bar{\xi}_t \bar{\xi}_t^\top \end{bmatrix} \rightsquigarrow \begin{bmatrix} \int_0^1 \bar{R}(s) \bar{R}(s)^\top \, ds & 0 \\ 0 & \Sigma_{\xi^+} m_Y^+(1) + \Sigma_{\xi^-} m_Y^-(1) \end{bmatrix}$$

and the r.h.s. is positive definite a.s.

Recall the definition of the q -dimensional standard (up to initialisation) Brownian motion W_0 given in (3.16).

Lemma A4. Suppose DGP^* , $CVAR$, $CO(ii)$ and DET hold. Then there exist $\tau_{xy}^\pm \in \mathbb{R}^{q-1}$, and an invertible $(q+1) \times (q+1)$ matrix Q such that

$$QR(\lambda) = S_q [e_1^\top W_0(\lambda)] W_0(\lambda) = W_0^*(\lambda).$$

Moreover, there exist $\omega^\pm > 0$ such that

$$Y(\lambda) = \omega^+ [e_1^\top W_0(\lambda)]_+ + \omega^- [e_1^\top W_0(\lambda)]_- \quad (A9)$$

We note further that because the mapping between $z_t = (y_t, x_t^\top)^\top$ and its derived canonical form $\bar{z}_t = (\bar{y}_t, \bar{x}_t^\top)^\top$ is such that \bar{y}_t^+ and \bar{y}_t^- are respectively positive scalar multiples of y_t^+ and y_t^- , a representation of the form (A9) also obtains when DGP holds in place of DGP^* .

Lemma A5. Suppose $W_0 = 0$ in (3.16). Then the matrices

$$\bar{S}_W^* := \int_0^1 \bar{W}_0^*(s) \bar{W}_0^*(s)^\top \, ds, \quad \bar{S}_V^* := \int_0^1 \bar{V}_0^*(s) \bar{V}_0^*(s)^\top \, ds,$$

are positive definite a.s.

Appendix B

Proofs of Main Results

Proof of Proposition 3.1

Since \mathbb{A}_n and \mathbb{B}_n are positive definite with probability approaching one (w.p.a.1.), the eigenvalues $\{\lambda_{n,i}\}_{i=1}^{d_w}$ of $\mathbb{A}_n \mathbb{B}_n^{-1}$ are well defined, real and positive w.p.a.1. By our assumptions and the continuous mapping theorem (CMT),

$$n^{-1} \mathbb{A}_n = \frac{1}{n} \sum_{t=1}^n w_{n,t} w_{n,t}^\top \rightsquigarrow \begin{bmatrix} \int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top \, ds & 0 \\ 0 & \Omega \end{bmatrix},$$

and

$$n^{-3} \mathbb{B}_n = \frac{1}{n} \sum_{t=1}^n \left(\frac{1}{n} \sum_{i=1}^t w_{n,i} \right) \left(\frac{1}{n} \sum_{j=1}^t w_{n,j} \right)^\top \rightsquigarrow \begin{bmatrix} \int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top \, ds & 0 \\ 0 & 0 \end{bmatrix}.$$

Let $\{\mu_{n,i}\}_{i=1}^{d_w}$ denote the eigenvalues of $\mathbb{B}_n \mathbb{A}_n^{-1}$ ordered as $\mu_{n,1} \leq \mu_{n,2} \leq \dots \leq \mu_{n,d_w}$, so that $\lambda_{n,i} = \mu_{n,d_w+1-i}^{-1}$ for $1 \leq i \leq d_w$. By the CMT and the a.s. invertibility of Ω ,

$$\begin{aligned} n^{-2} \mathbb{B}_n \mathbb{A}_n^{-1} &= (n^{-3} \mathbb{B}_n) (n^{-1} \mathbb{A}_n)^{-1} \\ &\rightsquigarrow \begin{bmatrix} \int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top \, ds & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top \, ds & 0 \\ 0 & \Omega \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top \, ds \left(\int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top \, ds \right)^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned} \quad (B1)$$

For the above limiting matrix, let $\{\mu_i^*\}_{i=1}^{d_w}$ denote its eigenvalues ordered as $\mu_1^* \leq \mu_2^* \leq \dots \leq \mu_{d_w}^*$. The first $d_w - \ell$ eigenvalues are zero, i.e., $\mu_i^* = 0$ for $1 \leq i \leq d_w - \ell$. The remaining ℓ eigenvalues $\{\mu_i^*\}_{i=d_w-\ell+1}^{d_w}$ are real and positive since they are the eigenvalues of

$$\int_0^1 \mathbb{V}(s) \mathbb{V}(s)^\top \, ds \left(\int_0^1 \mathbb{W}(s) \mathbb{W}(s)^\top \, ds \right)^{-1} =: \mathcal{V} \mathcal{W}^{-1},$$

where \mathcal{W} and \mathcal{V} are positive definite almost surely. By (B1), the continuity of eigenvalues and the CMT, then:

- i. for $1 \leq i \leq \ell$,

$$\begin{aligned} n^2 \lambda_{n,i} &= (n^{-2} \mu_{n,d_w+1-i})^{-1} \rightsquigarrow (\mu_{d_w+1-i}^*)^{-1} \\ &= (\mu_{(d_w-\ell)+(\ell+1-i)}^*)^{-1} < \infty, \end{aligned}$$

where $\mu_{(d_w-\ell)+(\ell+1-i)}^* > 0$ is the $(\ell+1-i)$ th eigenvalue of $\mathcal{V} \mathcal{W}^{-1}$; and

- ii. for $\ell+1 \leq i \leq d_w$,

$$n^2 \lambda_{n,i} = (n^{-2} \mu_{n,d_w+1-i})^{-1} \xrightarrow{p} \infty,$$

since $n^{-2} \mu_{n,d_w+1-i} \xrightarrow{p} \mu_{d_w+1-i}^* = 0$.

Therefore, if $\ell_0 = \ell$

$$\begin{aligned} n^2 \sum_{i=1}^{\ell_0} \lambda_{n,i} &= \sum_{i=1}^{\ell_0} (n^{-2} \mu_{n,d_w+1-i})^{-1} \\ &\rightsquigarrow \sum_{i=1}^{\ell_0} (\mu_{(d_w-\ell)+(\ell+1-i)}^*)^{-1} = \text{tr}[(\mathcal{V} \mathcal{W}^{-1})^{-1}] = \text{tr}(\mathcal{W} \mathcal{V}^{-1}), \end{aligned}$$

where the penultimate equality holds since the trace of a matrix equals the sum of its eigenvalues; and if $\ell_0 > \ell$,

$$\begin{aligned} n^2 \sum_{i=1}^{\ell_0} \lambda_{n,i} &= n^2 \sum_{i=1}^{\ell} \lambda_{n,i} + n^2 \sum_{i=\ell+1}^{\ell_0} \lambda_{n,i} \\ &= n^2 \sum_{i=1}^{\ell} \lambda_{n,i} + \sum_{i=\ell+1}^{\ell_0} (n^{-2} \mu_{n,d_w+1-i})^{-1} \xrightarrow{p} \infty, \end{aligned}$$

since $n^2 \sum_{i=1}^{\ell} \lambda_{n,i} = O_p(1)$ and the second term diverges in probability. \square

Proof of Theorem 3.1

As noted in the proof of Theorem 4.4 in DMW25, the process $\{\bar{z}_t\}$ obtained via the mapping (2.6) satisfies both DGP^* and DET' . Thus $\{\bar{z}_t\}$ satisfies the requirements of Lemma A2. The convergence (3.10) follows immediately, since $\text{sgn } \bar{y}_t = \text{sgn } y_t$ by Proposition 2.1 of Duffy et al. [15].

We next establish the convergence (3.12) holds in the '+' case; the proof in the '-' case is analogous. As per (D.3) of DMW25, define

$$P(\pm 1) := \begin{bmatrix} \bar{\phi}_{0,yy}^\pm & 0 \\ \bar{\phi}_{0,xy}^\pm & \Phi_{0,xx} \end{bmatrix}^{-1}$$

and set $P(y) = P(+1)\mathbf{1}^+(y) + P(-1)\mathbf{1}^-(y)$. It follows from (D.18) in DMW25 that

$$z_t^d = P(\tilde{y}_t)z_t^d,$$

and therefore

$$\mathbf{1}^+(y_t)z_t^d = \mathbf{1}^+(\tilde{y}_t)P(\tilde{y}_t)z_t^d = \mathbf{1}^+(\tilde{y}_t)P(+1)z_t^d.$$

Since (3.12) obtains for $\{z_t\}$ by Lemma A2, it follows that

$$\begin{aligned} \frac{1}{n^{3/2}} \sum_{i=1}^{\lfloor n\lambda \rfloor} [g(w_i) \otimes z_i^d] \mathbf{1}^+(y_i) &= [I_{d_g} \otimes P(+1)] \frac{1}{n^{3/2}} \sum_{i=1}^{\lfloor n\lambda \rfloor} [g(w_i) \otimes z_i^d] \mathbf{1}^+(\tilde{y}_i) \\ &\rightsquigarrow [I_{d_g} \otimes P(+1)] \\ &\quad \times [\mathbb{E}g(w_0^+)] \otimes \int_0^\lambda \tilde{Z}(\mu) \mathbf{1}^+[\tilde{Y}(\mu)] d\mu \\ &= [\mathbb{E}g(w_0^+)] \otimes \int_0^\lambda Z(\mu) \mathbf{1}^+[Y(\mu)] d\mu, \end{aligned}$$

where we have used that

$$\mathbf{1}^+[\tilde{Y}(\mu)]P(+1)\tilde{Z}(\mu) = \mathbf{1}^+[\tilde{Y}(\mu)]P[\tilde{Y}(\mu)]\tilde{Z}(\mu) = \mathbf{1}^+[Y(\mu)]Z(\mu)$$

as per (D.13) of DMW25. \square

Proof of Theorem 3.2

We now seek to verify the conditions of Proposition 3.1. As discussed in Section 2.2, by Proposition 2.1 in Duffy et al. [15] there exists an invertible $P \in \mathbb{R}^{(p+1) \times (p+1)}$ such that

$$\tilde{z}_t^* = \begin{bmatrix} \tilde{y}_t^+ \\ \tilde{y}_t^- \\ \tilde{x}_t \end{bmatrix} := P^{-1} \begin{bmatrix} y_t^+ \\ y_t^- \\ x_t \end{bmatrix} = P^{-1} z_t^*, \quad (\text{B2})$$

where $\text{sgn } \tilde{y}_t = \text{sgn } y_t$. As noted in Remark 4.2(i) of DMW25, $\{\tilde{z}_t\}$ follows—in view of our assumptions, in particular of the form taken by CO(ii).2—a canonical CKSVAR satisfying DGP*, CVAR, CO(ii) and DET. Because of the invariance properties of generalised eigenvalues, Λ_{n,q_0} is invariant to the pre- and/or post-multiplication of \mathbf{A}_n and \mathbf{B}_n by common matrices, and so it follows from (B2) that Λ_{n,q_0} computed on $\{z_t^*\}$ is identical to that computed on $\{\tilde{z}_t^*\}$. We may therefore suppose, without loss of generality, that $\{z_t\}$ follows a canonical CKSVAR, i.e., that DGP* holds in place of DGP.

By those same invariance properties of generalised eigenvalues, we may further replace $(\mathbf{A}_n, \mathbf{B}_n)$ by

$$\begin{aligned} \mathbb{A}_n &:= \bar{Q}(T_n^\top \mathbf{A}_n T_n) \bar{Q}^\top = \sum_{i=1}^n w_{n,i} w_{n,i}^\top \\ \mathbb{B}_n &:= \bar{Q}(T_n^\top \mathbf{B}_n T_n) \bar{Q}^\top = \sum_{i=1}^n \sum_{j=1}^i w_{n,i} w_{n,j}^\top \end{aligned}$$

where $\bar{Q} := \text{diag}\{Q, I_r\}$, for Q as in Lemma A4, and as per (3.6),

$$w_{n,t} := \bar{Q}(T_n^\top \tilde{z}_t^*) = \begin{bmatrix} Q\bar{Q}_{n,t} \\ \tilde{\xi}_t \end{bmatrix}.$$

By Lemmas A3 and A4, $\{w_{n,t}\}$ satisfies the requirements of Proposition 3.1, with

$$\mathbb{W}(s) = Q\bar{R}(s) = \bar{W}_0^*(s), \quad \Omega = \Sigma_{\xi^+} m_Y^+(1) + \Sigma_{\xi^-} m_Y^-(1), \quad (\text{B3})$$

and $\mathbb{V}(s) = \bar{V}_0^*(s) = \int_0^s \bar{W}_0^*(\lambda) d\lambda$, with the a.s. positive definiteness of $\int_0^1 \mathbb{W}(s)\mathbb{W}(s)^\top ds$ and $\int_0^1 \mathbb{V}(s)\mathbb{V}(s)^\top ds$ following by Lemma A5.

An application of Proposition 3.1 (with $\ell = q + 1$ and $\ell_0 = q_0 + 1$) then yields the conclusions of parts (i) and (iii). Part (ii) follows immediately from the result of part (i), noting that $\Lambda_{n,q_0} \leq \Lambda_{n,q}$ for all n , in this case, and $\Lambda_{n,q} \rightsquigarrow \Lambda_q$. Under DGP*, the convergence in (3.19) is an immediate consequence of Lemma A4; if instead DGP holds, then this follows from the fact that y_t^+ and y_t^- are respectively scalar multiples of the canonical variables \tilde{y}_t^+ and \tilde{y}_t^- , by (2.6). \square

Appendix C

Proofs of Auxiliary Lemmas

Proof of Lemma A1. (i). We have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{n^{-1/2}|y_i| \leq \delta\} &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{-\delta \leq n^{-1/2}y_i < 0\} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{0 \leq n^{-1/2}y_i \leq \delta\}. \end{aligned}$$

We will show that the second r.h.s. term is $o_p(1)$ as $n \rightarrow \infty$ and then $\delta \rightarrow 0$; the proof for the first r.h.s. term is analogous. Similarly to the proof of Theorem 4.2 in DMW25, define $f(y) := h(y)^{-1}y$. Then $f(y) = y$ for all $y \geq 0$, and it follows from (2.15) and (A2) above that

$$\begin{aligned} f(n^{-1/2}y_{\lfloor n\lambda \rfloor}) &\rightsquigarrow f[Y(\lambda)] = \vartheta^\top U_0(\lambda) \\ &= \vartheta^\top \Gamma(1; \mathcal{Y}_0) \mathcal{Z}_0 + \vartheta^\top U(\lambda) =: B_0 + B(\lambda) \end{aligned}$$

where B is a (scalar) Brownian motion, and $B_0 \in \mathbb{R}$ is non-random. Since $x \mapsto \mathbf{1}\{0 < x \leq \delta\}$ is Riemann integrable, it follows by Theorem 2.3 and Remark 2.2 in Berkes and Horváth [17] that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{0 \leq n^{-1/2}y_i \leq \delta\} &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{0 \leq f(n^{-1/2}y_i) \leq \delta\} \\ &\rightsquigarrow \int_0^1 \mathbf{1}\{0 \leq B_0 + B(\lambda) \leq \delta\} d\lambda \xrightarrow{p} 0 \end{aligned}$$

as $n \rightarrow \infty$ and then $\delta \rightarrow 0$, since B has a (Lebesgue) local time density.

(ii). By the Cramér–Wold device, it suffices to show that, on $D[0, 1]$,

$$\frac{1}{n} \sum_{i=1}^{\lfloor n\lambda \rfloor} \mathbf{1}^\pm(y_i)(a_0 + a^\top z_{n,t}^d) \rightsquigarrow \int_0^\lambda \mathbf{1}^\pm[Y(\mu)](a_0 + a^\top Z(\mu)) d\mu$$

for $a_0 \in \mathbb{R}$ and $a \in \mathbb{R}^p$, where $z_{n,t}^d := n^{-1/2}z_t^d$. We give the proof here for $\mathbf{1}^+$; the proof for $\mathbf{1}^-$ is analogous. To that end, define

$$T(\lambda) := \int_0^\lambda \mathbf{1}^+[Y(\mu)](a_0 + a^\top Z(\mu)) d\mu.$$

Letting $y_{n,t} := n^{-1/2}y_t$, we have

$$\frac{1}{n} \sum_{i=1}^{\lfloor n\lambda \rfloor} \mathbf{1}^+(y_i)(a_0 + a^\top z_{n,t}^d) = \frac{1}{n} \sum_{i=1}^{\lfloor n\lambda \rfloor} \mathbf{1}\{y_{n,t} \geq 0\}(a_0 + a^\top z_{n,t}^d) =: T_n(\lambda)$$

For $\epsilon > 0$, define a continuous function

$$f_\epsilon(y) := \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{\epsilon}y & \text{if } y \in [0, \epsilon) \\ 1 & \text{if } y \geq \epsilon, \end{cases}$$

so that by CMT and (A1),

$$\begin{aligned} T_{n,\epsilon}(\lambda) &:= \frac{1}{n} \sum_{i=1}^{\lfloor n\lambda \rfloor} f_\epsilon(y_{n,t})(a_0 + a^\top z_{n,t}^d) \\ &\rightsquigarrow \int_0^\lambda f_\epsilon[Y(\mu)](a_0 + a^\top Z(\mu)) d\mu =: T_\epsilon(\lambda) \end{aligned}$$

as $n \rightarrow \infty$. It then follows by arguments given in the proof of part (i) that, for some $C < \infty$ (depending on a and a_0),

$$\begin{aligned} & |T_\epsilon(\lambda) - T(\lambda)| \\ & \leq C \left(1 + \sup_{\lambda \in [0,1]} \|Z(\lambda)\| \right) \int_0^1 \mathbf{1}\{0 \leq Y(\mu) \leq \epsilon\} d\mu \\ & = C \left(1 + \sup_{\lambda \in [0,1]} \|Z(\lambda)\| \right) \int_0^1 \mathbf{1}\{0 \leq B_0 + B(\mu) \leq \epsilon\} d\mu \xrightarrow{p} 0 \end{aligned}$$

as $\epsilon \rightarrow 0$. Moreover, by the result of part (i), and (A3),

$$\begin{aligned} |T_{n,\epsilon}(\lambda) - T_n(\lambda)| & \leq \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} |f_\epsilon(y_{n,t}) - \mathbf{1}^+(y_{n,t})| |a_0 + a^\top z_{n,t}^d| \\ & \leq C \left(1 + \sup_{1 \leq s \leq n} \|z_{n,s}^d\| \right) \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{0 \leq y_{n,t} \leq \epsilon\} \xrightarrow{p} 0 \end{aligned}$$

as $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$. The preceding three convergences thus yield the result. \square

Proof of Lemma A2. By the Cramér–Wold device, it suffices to consider the case where $d_g = 1$. We note that the r.h.s. of (3.11) is well defined since $\rho(A^\pm) \leq \rho_{\text{JSR}}(A) < 1$. Here we shall prove the results only in the ‘+’ case; the proof in the ‘-’ case follows by identical arguments. We also only give the proof of (3.12), since (3.10) is essentially a simpler case of (3.12) in which $n^{-1/2} z_t^d =: z_{n,t}^d$ has been replaced by 1. The proof proceeds in the following five steps.

- i. Reduction to the case where g is bounded.
- ii. Disentangling of weakly dependent and integrated components:

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t}^d \mathbf{1}^+(y_t) \\ & = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} + o_p(1) \end{aligned} \quad (\text{C1})$$

as $n \rightarrow \infty, m \rightarrow \infty$ and then $\delta \rightarrow 0$, uniformly over $\lambda \in [0, 1]$.

- iii. Approximation of w_t : for each $m \in \mathbb{N}$ and $\delta > 0$,

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \\ & = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_{m,t}^+) z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} + o_p(1) \end{aligned} \quad (\text{C2})$$

as $n \rightarrow \infty$, uniformly over $\lambda \in [0, 1]$, where

$$w_{m,t}^+ := \sum_{\ell=0}^{m-1} (A^+)^{\ell} (c^+ + B^+ v_{t-\ell}). \quad (\text{C3})$$

- iv. Recentring of $g(w_{m,t}^+)$: for each $m \in \mathbb{N}$ and $\delta > 0$,

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_{m,t}^+) z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \\ & = [\mathbb{E}g(w_{m,0}^+)] \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} + o_p(1) \end{aligned}$$

as $n \rightarrow \infty$, uniformly over $\lambda \in [0, 1]$.

- v. Computing the limit:

$$\begin{aligned} & [\mathbb{E}g(w_{m,0}^+)] \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \\ & \rightsquigarrow [\mathbb{E}g(w_0^+)] \int_0^\lambda Z(\mu) \mathbf{1}^+[Y(\mu)] d\mu \end{aligned}$$

on $D[0, 1]$, as $n \rightarrow \infty, m \rightarrow \infty$ and then $\delta \rightarrow 0$.

- (i) Reduction to the case where g is bounded.

It follows directly from the local Lipschitz condition on g that

$$|g(w)| \leq |g(0)| + C(1 + \|w\|^{\ell_0}) \|w\| \leq C_1(1 + \|w\|^{\ell_0+1}) \quad (\text{C4})$$

for all $w \in \mathbb{R}^{d_w}$, and hence for some $\eta_0 \in (0, m_0/(\ell_0 + 1) - 1]$, which exists since $m_0 > \ell_0 + 1$,

$$|g(w)|^{1+\eta_0} \leq C_2(1 + \|w\|^{(\ell_0+1)(1+\eta_0)}) \leq C_3(1 + \|w\|^{m_0}).$$

Since $\sup_{t \in \mathbb{Z}} \|w_t\|_{m_0} < \infty$ by Lemma A.1 in DMW25, it follows immediately that $\sup_{t \in \mathbb{Z}} \|g(w_t)\|_{1+\eta_0} < \infty$. Moreover, since

$$\|w_0^+\|_{m_0} \leq \|(I_{d_w} - A^+)^{-1} c^+ + \sum_{\ell=0}^{\infty} \|(A^+)^{\ell}\| \|B^+\| \|v_{-\ell}\|_{m_0} < \infty, \quad (\text{C5})$$

it follows that $\mathbb{E}|g(w_0^+)|^{1+\eta_0} < \infty$, so that the r.h.s. of (3.12) is indeed well defined.

Now decompose

$$\begin{aligned} g(w) & = g(w) \mathbf{1}\{|g(w)| \leq M\} + g(w) \mathbf{1}\{|g(w)| > M\} \\ & =: g_M^{(\leq)}(w) + g_M^{(>)}(w). \end{aligned}$$

Recalling $z_{n,t}^d =: n^{-1/2} z_t^d$, we have

$$\left| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g_M^{(>)}(w_t) z_{n,t}^d \mathbf{1}^+(y_t) \right| \leq \sup_{s \leq n} \|z_{n,s}^d\| \frac{1}{n} \sum_{t=1}^n |g_M^{(>)}(w_t)| \xrightarrow{p} 0$$

as $n \rightarrow \infty$ and then $M \rightarrow \infty$, since $\sup_{s \leq n} \|z_{n,s}^d\| = O_p(1)$ as per (A3) above, and by Chebyshev’s inequality,

$$\sup_{t \in \mathbb{Z}} \mathbb{E}|g_M^{(>)}(w_t)| \leq \frac{\sup_{t \in \mathbb{Z}} \mathbb{E}|g(w_t)|^{1+\eta_0}}{M^{\eta_0}} \rightarrow 0$$

as $M \rightarrow \infty$. Since $\mathbb{E}g_M^{(>)}(w_0^+) \rightarrow 0$ as $M \rightarrow \infty$ by dominated convergence, it suffices to prove the result with $g_M^{(\leq)}$ in place of g . Moreover, since $g_M^{(\leq)}$ satisfies the same local Lipschitz condition as does g , we may henceforth suppose that g itself is bounded by some constant $C_g < \infty$, without loss of generality.

- (ii) Disentangling of weakly dependent and integrated components.

Let $m \in \mathbb{N}$. Since $d_g = 1$, we have that $g(w_t) \otimes z_t^d = g(w_t) z_t^d$. The l.h.s. of (3.12) may be written as

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t}^d \mathbf{1}^+(y_t) & = \sum_{i=0}^{m-1} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) \Delta z_{n,t-i}^d \mathbf{1}^+(y_t) \\ & \quad + \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t-m}^d \mathbf{1}^+(y_t), \end{aligned} \quad (\text{C6})$$

where we recall the convention that $\Delta z_i = \Delta z_i^d = 0$ for all $i \leq -k$ and that therefore $z_i^d = z_i = z_{-k} = z_{-k}^d$ for all $i \leq -k$, as per (2.14) above. For each $i \in \{0, \dots, m-1\}$, we have

$$\left\| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) \Delta z_{n,t-i}^d \mathbf{1}^+(y_t) \right\| \leq C_g \sup_{s \leq n} \|\Delta z_{n,s}^d\| \xrightarrow{p} 0 \quad (\text{C7})$$

as $n \rightarrow \infty$, since $\sup_{s \leq n} \|\Delta z_{n,s}^d\| = o_p(1)$ by (A3). Deduce that the first r.h.s. term in (C6) is $o_p(1)$ as $n \rightarrow \infty$, uniformly in $\lambda \in [0, 1]$.

This leaves the second r.h.s. term in (C6); to complete the proof of (C1), we need to replace $\mathbf{1}^+(y_t) = \mathbf{1}\{y_t \geq 0\}$ by $\mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\}$. Therefore consider

$$\begin{aligned} & |\mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} - \mathbf{1}^+(y_t)| \\ & = \mathbf{1}\{y_t \leq 0, y_{t-m} \geq n^{1/2}\delta\} + \mathbf{1}\{y_t \geq 0, y_{t-m} \leq n^{1/2}\delta\} \end{aligned}$$

$$\begin{aligned} &\leq \mathbf{1}\{y_t \leq 0, y_{t-m} \geq n^{1/2}\delta\} \\ &\quad + \mathbf{1}\{y_t \geq 0, y_{t-m} \leq -n^{1/2}\delta\} + \mathbf{1}\{|y_{t-m}| < n^{1/2}\delta\} \\ &=: \kappa_{1t} + \kappa_{2t} + \kappa_{3t} \end{aligned}$$

Using that $y_t - y_{t-m} = \sum_{\ell=0}^{m-1} \Delta y_{t-\ell}$, we have

$$y_t \leq 0 \text{ and } y_{t-m} \geq n^{1/2}\delta \Rightarrow \left| \sum_{\ell=0}^{m-1} \Delta y_{t-\ell} \right| \geq n^{1/2}\delta. \quad (\text{C8})$$

Hence

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \kappa_{1t} &\leq \frac{1}{n} \sum_{t=1}^n \mathbf{1} \left\{ \left| \sum_{\ell=0}^{m-1} \Delta y_{t-\ell} \right| \geq n^{1/2}\delta \right\} \\ &\leq \sum_{\ell=0}^{m-1} \frac{1}{n} \sum_{t=1}^n \mathbf{1} \{ |\Delta y_{t-\ell}| \geq n^{1/2}m^{-1}\delta \} \end{aligned}$$

where the second inequality holds since if $am \leq |\sum_{\ell=0}^{m-1} \Delta y_{t-\ell}| \leq \sum_{\ell=0}^{m-1} |\Delta y_{t-\ell}|$, then $|\Delta y_{t-\ell}| \geq a$ for some $\ell \in \{0, \dots, m-1\}$. By Chebyshev's inequality,

$$\max_{t \leq n} \mathbb{P}\{|\Delta y_t| \geq n^{1/2}m^{-1}\delta\} \leq n^{-1/2}\delta^{-1}m \max_{t \leq n} \mathbb{E}|\Delta y_t| \rightarrow 0 \quad (\text{C9})$$

as $n \rightarrow \infty$, since $\max_{t \leq n} \mathbb{E}|\Delta y_t| < \infty$ in view of (A4). Deduce that

$$\left\| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t-m}^d \kappa_{1t} \right\| \leq C_g \sup_{s \leq n} \|z_{n,s}^d\| \frac{1}{n} \sum_{t=1}^n \kappa_{1t} \xrightarrow{p} 0. \quad (\text{C10})$$

By a symmetric argument, the preceding also holds with κ_{2t} in place of κ_{1t} . Finally, it follows from Lemma A1(i) that

$$\left\| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_t) z_{n,t-m}^d \kappa_{3t} \right\| \leq C_g \sup_{s \leq n} \|z_{n,s}^d\| \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{|y_{t-m}| < n^{1/2}\delta\} \xrightarrow{p} 0 \quad (\text{C11})$$

as $n \rightarrow \infty$ and then $\delta \rightarrow 0$. Thus (C1) follows from (C10) and (C11).

(iii) Approximation of w_t .

We begin by decomposing

$$g(w_t) = g(w_{m,t}^+) + [g(w_t) - g(w_{m,t}^+)] =: g(w_{m,t}^+) + \nabla_{m,t}.$$

Since g is bounded, and $\sup_{s \leq n} \|z_{n,s}^d\| = O_p(1)$ as per (A3) above, the first m summands on the l.h.s. of (C2) are $o_p(1)$. Thus, to prove (C2), it suffices to establish the asymptotic negligibility of

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{t=m+1}^{\lfloor n\lambda \rfloor} \nabla_{m,t} z_{n,t-m}^d \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \right\| \\ &\leq \sup_{s \leq n} \|z_{n,s}^d\| \frac{1}{n} \sum_{t=m+1}^n |\nabla_{m,t}| \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\}. \end{aligned}$$

To handle the sum on the r.h.s., define

$$\mathbf{1}_{m,t} := \{y_s > 0, \forall s \in \{t-m, \dots, t\}\}.$$

If $\mathbf{1}_{m,t} = 1$, then $y_s > 0$ for all $s \in \{t-m, \dots, t\}$, and so $(A_s, B_s, c_s) = (A^+, B^+, c^+)$ for all $s \in \{t-m+1, \dots, t\}$, whence recursive substitution applied to (3.8) yields

$$w_t = (A^+)^m w_{t-m} + \sum_{\ell=0}^{m-1} (A^+)^\ell (c^+ + B^+ v_{t-\ell}) = (A^+)^m w_{t-m} + w_{m,t}^+.$$

In other words, when $\mathbf{1}_{m,t} = 1$ holds w_t may be approximated by $w_{m,t}^+$, and so $\nabla_{m,t}$ should be small. Indeed,

$$\begin{aligned} |\nabla_{m,t}| \mathbf{1}_{m,t} &= |g(w_t) - g(w_{m,t}^+)| \mathbf{1}_{m,t} \\ &= |g((A^+)^m w_{t-m} + w_{m,t}^+) - g(w_{m,t}^+)| \mathbf{1}_{m,t} \\ &\leq C_1 \min\{1, \|(A^+)^m\| \|w_{t-m}\| (1 + \|w_t\|^{\ell_0} + \|w_{m,t}^+\|^{\ell_0})\} \\ &\leq C_2 \min\{1, \|(A^+)^m\| \|w_{t-m}\| (1 + \|w_{t-m}\|^{\ell_0} + \|w_{m,t}^+\|^{\ell_0})\} \end{aligned}$$

for some $C_1, C_2 < \infty$, using the local Lipschitz condition (3.9), and the boundedness of g . By Lemma A.1 of DMW25, for $\gamma \in (\rho_{JSR}(A), 1)$,

$$\|w_t\| \leq C_3 \left[\sum_{s=0}^{t-1} \gamma^s (1 + \|v_{t-s}\|) + \gamma^t \|w_0\| \right]$$

for some $C_3 < \infty$. Therefore, for $t \geq m+1$, the distribution of $\|w_{t-m}\|$ is stochastically dominated by that of

$$C_3 \left[\sum_{\ell=1}^{\infty} \gamma^{\ell-1} (1 + \|v_\ell\|) + \|w_0\| \right] =: \bar{w}_0$$

while the distribution of $\|w_{m,t}^+\|$ is stochastically dominated by that of

$$\sum_{\ell=0}^{\infty} \|(A^+)^{\ell} (\|c^+\| + \|B^+\| \|v_\ell\|) =: \bar{w}_0^+$$

Since $w_{m,t}^+$ depends only on $\{v_s\}_{s=t-m+1}^t$, it is independent of w_{t-m} . Therefore, taking (\bar{w}_0, \bar{w}_0^+) to be such that \bar{w}_0 and \bar{w}_0^+ are independent, with (marginally) $\bar{w}_0 =_d \bar{w}_0$ and $\bar{w}_0^+ =_d \bar{w}_0^+$, we have that

$$\begin{aligned} &\max_{m+1 \leq t \leq n} \mathbb{E}|\nabla_{m,t}| \mathbf{1}_{m,t} \\ &\leq \max_{m+1 \leq t \leq n} C_2 \mathbb{E} \min\{1, \|(A^+)^m\| \|w_{t-m}\| \\ &\quad \times (1 + \|w_{t-m}\|^{\ell_0} + \|w_{m,t}^+\|^{\ell_0})\} \\ &\leq C_2 \mathbb{E} \min\{1, \|(A^+)^m\| \|\bar{w}_0\| \\ &\quad \times (1 + \|\bar{w}_0\|^{\ell_0} + \|\bar{w}_0^+\|^{\ell_0})\} \\ &\rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$, by dominated convergence. Deduce

$$\begin{aligned} &\frac{1}{n} \sum_{t=m+1}^n |\nabla_{m,t}| \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \\ &= \frac{1}{n} \sum_{t=m+1}^n |\nabla_{m,t}| \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} [\mathbf{1}_{m,t} + (1 - \mathbf{1}_{m,t})] \\ &= \frac{1}{n} \sum_{t=m+1}^n |\nabla_{m,t}| \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} (1 - \mathbf{1}_{m,t}) + o_p(1). \quad (\text{C12}) \end{aligned}$$

as $n \rightarrow \infty$ and then $m \rightarrow \infty$.

It remains to show that the first r.h.s. term in (C12) is also asymptotically negligible. We note that the summands are nonzero only if $\mathbf{1}_{m,t} = 0$, in which case, there must exist an $i \in \{0, \dots, m\}$ such that $y_{t-i} \leq 0$. Using a similar argument to that which follows (C8) above, since $y_{t-i} = y_{t-m} + \sum_{j=i}^{m-1} \Delta y_{t-j}$ we have that

$$y_{t-i} \leq 0 \text{ and } y_{t-m} \geq n^{1/2}\delta \Rightarrow \left| \sum_{j=i}^{m-1} \Delta y_{t-j} \right| \geq n^{1/2}\delta.$$

Hence

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} (1 - \mathbf{1}_{m,t}) \\ &= \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \mathbf{1}\{\exists i \in \{0, \dots, m\} \text{ s.t. } y_{t-i} \leq 0\} \\ &\leq \sum_{i=0}^{m-1} \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{y_{t-m} \geq n^{1/2}\delta\} \mathbf{1}\{y_{t-i} \leq 0\} \end{aligned}$$

$$\begin{aligned} &\leq \sum_{i=0}^{m-1} \frac{1}{n} \sum_{t=1}^n \mathbf{1} \left\{ \left| \sum_{j=i}^{m-1} \Delta y_{t-j} \right| \geq n^{1/2} \delta \right\} \\ &\leq \sum_{i=0}^{m-1} \sum_{j=i}^{m-1} \frac{1}{n} \sum_{t=1}^n \mathbf{1} \{ |\Delta y_{t-j}| \geq n^{1/2} (m-i)^{-1} \delta \} \end{aligned} \quad (C13)$$

with the expectation of the summands being bounded by the l.h.s. of (C9), modulo the replacement of m by $m-i$ there. Since g is bounded, deduce that

$$\frac{1}{n} \sum_{t=1}^n |\nabla_{m,t}| \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} (1 - \mathbf{1}_{m,t}) \xrightarrow{p} 0$$

as $n \rightarrow \infty$, as required.

(iv) Recentering of $g(w_{m,t}^+)$.

Defining

$$\bar{g}(w_{m,t}^+) := g(w_{m,t}^+) - \mathbb{E}g(w_{m,t}^+) = g(w_{m,t}^+) - \mathbb{E}g(w_{m,0}^+)$$

we may write

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} g(w_{m,t}^+) z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \\ &= [\mathbb{E}g(w_{m,0}^+)] \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \\ &\quad + \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{g}(w_{m,t}^+) z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \}. \end{aligned} \quad (C14)$$

We must show that the second r.h.s. term in (C14) is negligible. We first note that

$$\begin{aligned} &\mathbb{E} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{g}(w_{m,t}^+) z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \mathbf{1} \{ \|z_{n,t-m}^d\| > M \} \right\| \\ &\leq C \mathbb{P} \left\{ \sup_{1 \leq t \leq n} \|z_{n,t}^d\| > M \right\} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ and then $M \rightarrow \infty$, since $\sup_{1 \leq t \leq n} \|z_{n,t}^d\| = O_p(1)$. Therefore, letting $h_M(z) := z \mathbf{1} \{ \|z\| \leq M \}$, it suffices to show that

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{g}(w_{m,t}^+) h_M(z_{n,t-m}^d) \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \xrightarrow{p} 0$$

as $n \rightarrow \infty$, for each $M > 0$.

In view of (C3), $w_{m,t}^+$ is a function only of $\{v_{t-m+1}, \dots, v_t\}$, and is therefore independent of \mathcal{F}_{t-m}^- . $\bar{g}(w_{m,t}^+)$ admits the telescoping sum decomposition

$$\begin{aligned} \bar{g}(w_{m,t}^+) &= g(w_{m,t}^+) - \mathbb{E}g(w_{m,t}^+) = \sum_{\ell=0}^{m-1} [\mathbb{E}_{t-\ell} g(w_{m,t}^+) - \mathbb{E}_{t-\ell-1} g(w_{m,t}^+)] \\ &=: \sum_{\ell=0}^{m-1} \zeta_{\ell,m,t}, \end{aligned}$$

where $\mathbb{E}_s[\cdot] := \mathbb{E}[\cdot | \mathcal{F}_s]$, and we have used the fact that $\mathbb{E}_{t-m} g(w_{m,t}^+) = \mathbb{E}g(w_{m,t}^+)$. For every $\ell \in \{0, \dots, m-1\}$, $\{\zeta_{\ell,m,t}\}_{t \in \mathbb{N}}$ defines a bounded martingale difference sequence. Rewriting

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{g}(w_{m,t}^+) h_M(z_{n,t-m}^d) \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \\ &= \frac{1}{n^{1/2}} \sum_{\ell=0}^{m-1} \sum_{t=1}^{\lfloor n\lambda \rfloor} \frac{\zeta_{\ell,m,t}}{n^{1/2}} h_M(z_{n,t-m}^d) \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \\ &=: \frac{1}{n^{1/2}} \sum_{\ell=0}^{m-1} S_{\ell,m,n}(\lambda). \end{aligned} \quad (C15)$$

Applying Theorem 2.11 in Hall and Heyde [18, with $p=2$] to each element of the martingale $S_{\ell,m,n}(\lambda)$, it follows that there exists a $C < \infty$ such that

$$\mathbb{E} \sup_{\lambda \in [0,1]} \|S_{\ell,m,n}(\lambda)\|^2 \leq C(1 + n^{-1} M^2),$$

and hence

$$\frac{1}{n^{1/2}} \sum_{\ell=0}^{m-1} S_{\ell,m,n}(\lambda) \xrightarrow{p} 0$$

uniformly in $\lambda \in [0, 1]$, as $n \rightarrow \infty$.

(v) Computing the limit.

Finally, regarding the first r.h.s. term in (C14), we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} &= \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1}^+(y_{t-m}) \\ &\quad - \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ 0 \leq y_{t-m} < n^{1/2} \delta \} \end{aligned}$$

and by Lemma A1(i),

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ 0 \leq y_{t-m} < n^{1/2} \delta \} \right\| \\ &\leq \max_{s \leq n} \|z_{n,s}^d\| \frac{1}{n} \sum_{t=1}^n \mathbf{1} \{ |y_{t-m}| < n^{1/2} \delta \} \\ &\leq \max_{s \leq n} \|z_{n,s}^d\| \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1} \{ |y_t| < n^{1/2} \delta \} + o_p(1) \right) \xrightarrow{p} 0 \end{aligned}$$

as $n \rightarrow \infty$, $m \rightarrow \infty$ and then $\delta \rightarrow 0$. Hence by Lemma A1(ii),

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} &= \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t}^d \mathbf{1}^+(y_t) + o_p(1) \\ &\rightsquigarrow \int_0^\lambda \mathbf{1}^+[Y(s)] Z(s) ds. \end{aligned}$$

as $n \rightarrow \infty$, $m \rightarrow \infty$ and then $\delta \rightarrow 0$. Since g is bounded and continuous, and

$$w_{m,0}^+ = \sum_{\ell=0}^{m-1} (A^+)^{\ell} (c^+ + B^+ v_{-\ell}) \xrightarrow{\text{a.s.}} \sum_{\ell=0}^{\infty} (A^+)^{\ell} (c^+ + B^+ v_{-\ell}) = w_0^+,$$

it follows by the dominated convergence theorem that $\mathbb{E}g(w_{m,0}^+) \rightarrow \mathbb{E}g(w_0^+)$ as $m \rightarrow \infty$. Hence

$$\begin{aligned} \mathbb{E}g(w_{m,0}^+) \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} z_{n,t-m}^d \mathbf{1} \{ y_{t-m} \geq n^{1/2} \delta \} \\ \rightsquigarrow \mathbb{E}g(w_0^+) \int_0^\lambda \mathbf{1}^+[Y(s)] Z(s) ds \end{aligned}$$

as $n \rightarrow \infty$, $m \rightarrow \infty$ and then $\delta \rightarrow 0$. □

Proof of Lemma A3. (i). Recall from (3.3) and (3.4) that

$$\tau^* = \begin{bmatrix} 1 & 0 & \tau_{xy}^{+\top} \\ 0 & 1 & \tau_{xy}^{-\top} \\ 0 & 0 & \beta_{x,\perp} \end{bmatrix} \quad \beta^* = \begin{bmatrix} \beta_y^{+\top} \\ \beta_y^{-\top} \\ \beta_x \end{bmatrix}.$$

Let $a = (a_1, a_2, a_3^{\top})^{\top} \in \mathbb{R}^{q+1}$ and $b \in \mathbb{R}^r$ be such that

$$0 = \begin{bmatrix} \tau^* & \beta^* \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a_1 + \tau_{xy}^{+\top} a_3 + \beta_y^{+\top} b \\ a_2 + \tau_{xy}^{-\top} a_3 + \beta_y^{-\top} b \\ \beta_{x,\perp} a_3 + \beta_x b \end{bmatrix}$$

where $a_{(3)} \in \mathbb{R}^{q-1}$. Since $[\beta_{x,\lambda}, \beta_x]$ has rank $p-1$, it follows that $a_{(3)} = 0$ and $b = 0$. Hence $a_1 = a_2 = 0$, i.e., $a = 0$ also.

(ii). Regarding $\bar{\varrho}_{n,t}$, we have by (A6) that

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{\varrho}_{n,t} &= \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \varrho_{n,t} - \lambda \frac{1}{n} \sum_{t=1}^n \varrho_{n,t} \\ &\rightsquigarrow \int_0^\lambda R(s) \, ds - \lambda \int_0^1 R(s) \, ds = \int_0^\lambda \bar{R}(s) \, ds \end{aligned}$$

on $D[0, 1]$ jointly with $U_n \rightsquigarrow U$. We next consider $\bar{\xi}_t$, for which we similarly have

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{\xi}_t = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t - \lambda \frac{1}{n} \sum_{t=1}^n \xi_t. \quad (C16)$$

To determine the weak limits of the various components on the r.h.s., we apply Lemma A2. To that end, define

$$\xi_t := \beta(y_t)^\top z_t = (\bar{\xi}_t^\top, \Delta z_t^{*\top}, \dots, \Delta z_{t-k+2}^{*\top})^\top$$

where as per (2.12),

$$\alpha := \begin{bmatrix} \alpha & \Gamma_1 & \Gamma_2 & \dots & \Gamma_{k-1} \\ & I_{p+1} & & & \\ & & I_{p+1} & & \\ & & & \ddots & \\ & & & & I_{p+1} \end{bmatrix}, \quad \beta(y)^\top := \begin{bmatrix} \beta(y)^\top \\ S_p(y) & -I_{p+1} \\ & I_{p+1} & -I_{p+1} \\ & & \ddots & \ddots \\ & & & I_{p+1} & -I_{p+1} \end{bmatrix}, \quad (C17)$$

and

$$c := \begin{bmatrix} c \\ 0_{(p+1)(k-1)} \end{bmatrix} \quad u_t := \begin{bmatrix} u_t \\ 0_{(p+1)(k-1)} \end{bmatrix} \quad (C18)$$

it follows from Lemma B.2 and the arguments subsequently given in the proof of Theorem 4.2 in DMW25, that $w_t = \xi_t$ follows (for $t \geq 1$) an autoregressive process satisfying the requirements of Lemma A2 above (see the statement of Theorem 3.1), with in particular

$$\begin{aligned} c^\pm &= \beta(\pm 1)^\top c \\ A^\pm &= I_{r+(k-1)(p+1)} + \beta(\pm 1)^\top \alpha \\ B^\pm &= \beta(\pm 1)^\top \\ v_t &= u_t. \end{aligned}$$

Hence by that result, with $g(w) = w$ and noting that $\|v_t\|_{2+\delta_u} < \infty$,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{\xi}_t &= E_r^\top \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t = E_r^\top \left[\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t \mathbf{1}^+(y_t) + \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t \mathbf{1}^-(y_t) \right] \\ &\rightsquigarrow E_r^\top [(\mathbb{E} \xi_0^+) m_Y^+(\lambda) + (\mathbb{E} \xi_0^-) m_Y^-(\lambda)] \\ &= \mu_\xi^+ m_Y^+(\lambda) + \mu_\xi^- m_Y^-(\lambda) = \lambda \mu_\xi, \end{aligned} \quad (C19)$$

where E_r denotes the first r columns of $I_{r+(k-1)(p+1)}$,

$$\begin{aligned} \xi_0^\pm &:= -[\beta(\pm 1)^\top \alpha]^{-1} \beta(\pm 1)^\top c \\ &+ \sum_{\ell=0}^{\infty} [J_{r+(k-1)(p+1)} + \beta(\pm 1)^\top \alpha]^\ell \beta(\pm 1)^\top u_{-\ell}, \end{aligned} \quad (C20)$$

and for $\xi_0^\pm := E_r^\top \xi_0^\pm$,

$$\mu_\xi^\pm := \mathbb{E} \xi_0^\pm = -E_r^\top [\beta(\pm 1)^\top \alpha]^{-1} \beta(\pm 1)^\top c = \mu_\xi \quad (C21)$$

because by DET there exists a $\mu_\xi \in \mathbb{R}^r$ such that $c = -\alpha \mu_\xi$, and therefore $c = -\alpha \mu_\xi$ for $\mu_\xi := (\mu_\xi^\top, 0_{(p+1)(k-1)}^\top)^\top$. Hence it follows from (C16) and (C19) that

$$\frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \bar{\xi}_t = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \xi_t - \lambda \frac{1}{n} \sum_{t=1}^n \xi_t \xrightarrow{p} \lambda \mu_\xi - \lambda \mu_\xi = 0. \quad (C22)$$

on $D[0, 1]$.

(iii). Observe that because $\bar{\varrho}_{n,t}$ and $\bar{\xi}_t$ have zero sample mean,

$$\frac{1}{n} \sum_{t=1}^n \begin{bmatrix} \bar{\varrho}_{n,t} \bar{\varrho}_{n,t}^\top & \bar{\varrho}_{n,t} \bar{\xi}_t^\top \\ \bar{\xi}_t \bar{\varrho}_{n,t}^\top & \bar{\xi}_t \bar{\xi}_t^\top \end{bmatrix} = \frac{1}{n} \sum_{t=1}^n \begin{bmatrix} \bar{\varrho}_{n,t} \varrho_{n,t}^\top & \varrho_{n,t} \bar{\xi}_t^\top \\ \bar{\xi}_t \varrho_{n,t}^\top & \bar{\xi}_t \bar{\xi}_t^\top \end{bmatrix}. \quad (C23)$$

For the upper left block of (C23), we have directly from (A6) that

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \bar{\varrho}_{n,t} \varrho_{n,t}^\top &= \frac{1}{n} \sum_{t=1}^n \varrho_{n,t} \varrho_{n,t}^\top - \hat{\mu}_{n,\varrho} \hat{\mu}_{n,\varrho}^\top \\ &\rightsquigarrow \int_0^1 R(s) R(s)^\top \, ds - \left(\int_0^1 R(s) \, ds \right) \left(\int_0^1 R(s) \, ds \right)^\top \\ &= \int_0^1 \bar{R}(s) \bar{R}(s)^\top \, ds, \end{aligned}$$

where $\hat{\mu}_{n,\varrho} = \frac{1}{n} \sum_{t=1}^{\lfloor n\lambda \rfloor} \varrho_{n,t}$.

We next consider the off-diagonal block, for which

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \bar{\xi}_t \varrho_{n,t}^\top &= \frac{1}{n} \sum_{t=1}^n (\xi_t - \hat{\mu}_{n,\xi}) \varrho_{n,t}^\top \\ &= \frac{1}{n} \sum_{t=1}^n (\xi_t - \hat{\mu}_{n,\xi}) \varrho_{n,t}^\top \mathbf{1}^+(y_t) + \frac{1}{n} \sum_{t=1}^n (\xi_t - \hat{\mu}_{n,\xi}) \varrho_{n,t}^\top \mathbf{1}^-(y_t) \end{aligned}$$

since $\mathbf{1}^+(y_t) + \mathbf{1}^-(y_t) = 1$, where $\hat{\mu}_{n,\xi} = \frac{1}{n} \sum_{t=1}^n \xi_t$. Using, as noted in the proof of part (ii), that $\xi_t = E_r^\top \bar{\xi}_t$, it follows from (A6) and (A7) (itself an implication of Lemma A2) and (C21) that

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \bar{\xi}_t \varrho_{n,t}^\top \mathbf{1}^\pm(y_t) &= E_r^\top \left[\frac{1}{n^{3/2}} \sum_{t=1}^n \mathbf{1}^\pm(y_t) \xi_t z_t^{*\top} \right] \tau^* \\ &\rightsquigarrow E_r^\top [\mathbb{E} \xi_0^\pm] \left[\int_0^1 Z^*(s) \mathbf{1}^\pm[Y(s)] \, ds \right]^\top \tau^* \\ &= (\mathbb{E} \xi_0^\pm) \int_0^1 R(s)^\top \mathbf{1}^\pm[Y(s)] \, ds \\ &= \mu_\xi \int_0^1 R(s)^\top \mathbf{1}^\pm[Y(s)] \, ds \end{aligned}$$

while by another application of Lemma A2, and (C19) above (with $\lambda = 1$)

$$\hat{\mu}_{n,\xi} \frac{1}{n} \sum_{t=1}^n \varrho_{n,t}^\top \mathbf{1}^\pm(y_t) \rightsquigarrow \mu_\xi \int_0^1 R(s)^\top \mathbf{1}^\pm[Y(s)] \, ds.$$

Deduce that

$$\frac{1}{n} \sum_{t=1}^n (\xi_t - \hat{\mu}_{n,\xi}) \varrho_{n,t}^\top \mathbf{1}^\pm(y_t) \xrightarrow{p} 0,$$

and thus $\frac{1}{n} \sum_{t=1}^n \bar{\xi}_t \varrho_{n,t}^\top \xrightarrow{p} 0$, as required.

We finally come to the lower right block of (C23). We have

$$\frac{1}{n} \sum_{t=1}^n \bar{\xi}_t \bar{\xi}_t^\top = \frac{1}{n} \sum_{t=1}^n (\xi_t - \hat{\mu}_{n,\xi}) \bar{\xi}_t^\top = \frac{1}{n} \sum_{t=1}^n \xi_t \bar{\xi}_t^\top - \hat{\mu}_{n,\xi} \hat{\mu}_{n,\xi}^\top \quad (C24)$$

where $\hat{\mu}_{n,\xi} \xrightarrow{p} \mu_\xi$ per (C19) above. Similarly to (C19), we also have by Lemma A2 (in this instance with $g(u) = wu^\top$, and noting that $\|v_t\|_{2+\delta_t} < \infty$) that

$$\frac{1}{n} \sum_{t=1}^n \xi_t \xi_t^\top \mathbf{1}^\pm(y_t) = E_r^\top \left[\frac{1}{n} \sum_{t=1}^n \xi_t \xi_t^\top \mathbf{1}^\pm(y_t) \right] E_r \rightsquigarrow (\mathbb{E} \xi_0^\pm \xi_0^{\pm\top}) m_Y^\pm(1). \quad (\text{C25})$$

By (C20) and (C21) above,

$$\begin{aligned} \xi_0^\pm - \mu_\xi &= E_r^\top [\xi_0^\pm - \mathbb{E} \xi_0^\pm] \\ &= E_r^\top \sum_{\ell=0}^{\infty} (J_{r+(k-1)(p+1)} + \beta(\pm 1)^\top \alpha^\ell \beta(\pm 1)^\top u_{-\ell}). \end{aligned}$$

Recalling the definitions of $\beta(y)$ and u , in (C17) and (C18) above, the first term on the r.h.s. series is

$$E_r^\top \beta(\pm 1)^\top u_0 = \beta(\pm 1)^\top u_0,$$

which has nonsingular matrix variance $\beta(\pm 1)^\top \Sigma_u \beta(\pm 1)$. It follows that $\Sigma_\xi^\pm := \text{var}(\xi_0^\pm)$ is positive definite, and since

$$\mathbb{E} \xi_0^\pm \xi_0^{\pm\top} = \Sigma_\xi^\pm + \mu_\xi \mu_\xi^\top$$

we deduce from (C24) and (C25) that

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \xi_t \xi_t^\top &\rightsquigarrow (\Sigma_\xi^+ + \mu_\xi \mu_\xi^\top) m_Y^+(1) + (\Sigma_\xi^- + \mu_\xi \mu_\xi^\top) m_Y^-(1) - \mu_\xi \mu_\xi^\top \\ &= \Sigma_\xi^+ m_Y^+(1) + \Sigma_\xi^- m_Y^-(1). \end{aligned}$$

Since $m_Y^+(1) + m_Y^-(1) = 1$, this is positive definite as the convex combination of two positive definite matrices. \square

Proof of Lemma A4. In view of (A1), (A5) and (A6), we have

$$\begin{aligned} R(\lambda) &= \tau^{*\top} Z^*(\lambda) = \tau^{*\top} S_p[Y(\lambda)] Z(\lambda) \\ &= \tau^{*\top} S_p[Y(\lambda)] P_{\beta_\perp} [Y(\lambda)] U_0(\lambda). \end{aligned} \quad (\text{C26})$$

As in Lemma B.3 in DMW25, define $g(y, u) := P_{\beta_\perp}(y)u$ and $\vartheta^\top := e_1^\top P_{\beta_\perp}(+1) \neq 0$. It follows from Theorem 4.2 in DMW25 that $\text{sgn } Y(\lambda) = \text{sgn } \vartheta^\top U_0(\lambda)$, and therefore

$$\begin{aligned} Z^*(\lambda) &= S_p[Y(\lambda)] P_{\beta_\perp} [Y(\lambda)] U_0(\lambda) \\ &= S_p[\vartheta^\top U_0(\lambda)] P_{\beta_\perp} [\vartheta^\top U_0(\lambda)] U_0(\lambda). \end{aligned} \quad (\text{C27})$$

The r.h.s. is a (continuous) function of a p -dimensional Brownian motion $U_0(\lambda)$; our objective is to rewrite it in terms of a (known) function of a q -dimensional standard (up to initialisation) Brownian motion W_0 . The chief obstacle here (relative to the linear case) lies in the nonlinearity with which U_0 enters the r.h.s.; we therefore first seek to obtain an expression for Z^* in terms of a p -dimensional Brownian motion, such that only the first component of that Brownian motion enters $Z^*(\lambda)$ nonlinearly.

To that end, define $\theta := \|\vartheta\|^{-1} \vartheta$, and let $\Theta := [\theta, \theta_\perp]$ be a $p \times p$ orthonormal matrix. Then for any $y \in \mathbb{R}$ and $u \in \mathbb{R}^p$,

$$g(y, u) = P_{\beta_\perp}(y)u = P_{\beta_\perp}(y)\Theta\theta^\top u = \begin{bmatrix} P_{\beta_\perp}(y)\theta & P_{\beta_\perp}(y)\theta_\perp \end{bmatrix} \begin{bmatrix} \theta^\top u \\ \theta_\perp^\top u \end{bmatrix},$$

and note that $\vartheta^\top \theta_\perp = 0$ by construction. Therefore applying Lemma B.3(ii) in DMW25 to each column of $P_{\beta_\perp}(y)\theta_\perp$, we obtain

$$P_{\beta_\perp}(+1)\theta_\perp = P_{\beta_\perp}(-1)\theta_\perp$$

whence

$$g(y, u) = P_{\beta_\perp}(y)\theta[\theta^\top u] + P_{\beta_\perp}(+1)\theta_\perp[\theta_\perp^\top u].$$

This allows us to confine the nonlinearity in the function to the scalar variable $\theta^\top u$, with the remaining $p - 1$ variables $\theta_\perp^\top u$ entering the r.h.s. linearly. In view of (C27), which because $\text{sgn } \vartheta^\top u = \text{sgn } \theta^\top u$ may be written as

$$\begin{aligned} Z^*(\lambda) &= S_p[\theta^\top U_0(\lambda)] P_{\beta_\perp} [\theta^\top U_0(\lambda)] U_0(\lambda) \\ &= S_p[\theta^\top U_0(\lambda)] g[\theta^\top U_0(\lambda), U_0(\lambda)], \end{aligned} \quad (\text{C28})$$

we are only interested in the case where $\text{sgn } y = \text{sgn } \theta^\top u$, for which

$$\begin{aligned} g(\theta^\top u, u) &= P_{\beta_\perp}(\theta^\top u)\theta[\theta^\top u] + P_{\beta_\perp}(+1)\theta_\perp[\theta_\perp^\top u] \\ &= P_{\beta_\perp}(+1)\theta[\theta^\top u]_+ + P_{\beta_\perp}(-1)\theta[\theta^\top u]_- + P_{\beta_\perp}(+1)\theta_\perp[\theta_\perp^\top u] \\ &=: \psi^+[\theta^\top u]_+ + \psi^-[\theta^\top u]_- + \Psi^x[\theta_\perp^\top u]. \end{aligned} \quad (\text{C29})$$

By Lemma B.3(i) in DMW25,

$$e_1^\top \psi^+ = e_1^\top P_{\beta_\perp}(+1)\theta = \frac{\vartheta^\top \vartheta}{\|\vartheta\|} = \|\vartheta\| > 0$$

and also $e_1^\top \psi^- > 0$, while

$$e_1^\top \Psi^x = e_1^\top P_{\beta_\perp}(+1)\theta_\perp = \vartheta^\top \theta_\perp = 0, \quad (\text{C30})$$

and thus we may write $\Psi^x = [0_{q-1}^\top, \Psi_{xx}^\top]^\top$ for some $\Psi_{xx} \in \mathbb{R}^{(p-1) \times (q-1)}$. Partitioning $\psi^\pm = (\psi_y^\pm, \psi_x^\pm)^\top$, where $\psi_y^\pm := e_1^\top \psi^\pm$, we obtain from (C28) and (C29) the representation

$$\begin{aligned} Z^*(\lambda) &= \begin{bmatrix} \mathbf{1}^+[\theta^\top U_0(\lambda)] & 0 \\ \mathbf{1}^-[\theta^\top U_0(\lambda)] & 0 \\ 0 & I_{p-1} \end{bmatrix} \begin{bmatrix} \psi_y^+ & \psi_y^- & 0 \\ \psi_x^+ & \psi_x^- & \Psi_{xx} \end{bmatrix} \begin{bmatrix} [\theta^\top U_0(\lambda)]_+ \\ [\theta^\top U_0(\lambda)]_- \\ \theta_\perp^\top U_0(\lambda) \end{bmatrix} \\ &= \begin{bmatrix} \psi_y^+ & 0 & 0 \\ 0 & \psi_y^- & 0 \\ \psi_x^+ & \psi_x^- & \Psi_{xx} \end{bmatrix} \begin{bmatrix} [\theta^\top U_0(\lambda)]_+ \\ [\theta^\top U_0(\lambda)]_- \\ \theta_\perp^\top U_0(\lambda) \end{bmatrix} =: \Psi^* S_p[\theta^\top U_0(\lambda)] \Theta^\top U_0(\lambda) \end{aligned} \quad (\text{C31})$$

where we have used the fact that $\psi_y^\pm > 0$. We have thus represented Z^* in terms of a p -dimensional Brownian motion $\Theta^\top U_0$, where only the first component $e_1^\top \Theta^\top U_0(\lambda) = \theta^\top U_0(\lambda)$ enters $Z^*(\lambda)$ nonlinearly.

The next step is to collapse the $(p + 1)$ -dimensional process $Z^*(\lambda)$ into the $(q + 1)$ -dimensional process $R(\lambda)$. From (C26) and (C31), we have

$$\begin{aligned} R(\lambda) &= \tau^{*\top} Z^*(\lambda) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \tau_{xy}^+ & \tau_{xy}^- & \beta_{x,\perp}^\top \end{bmatrix} \begin{bmatrix} \psi_y^+ & 0 & 0 \\ 0 & \psi_y^- & 0 \\ \psi_x^+ & \psi_x^- & \Psi_{xx} \end{bmatrix} S_p[\theta^\top U_0(\lambda)] \Theta^\top U_0(\lambda) \end{aligned}$$

where we are entirely free to choose $\tau_{xy}^\pm \in \mathbb{R}^{q-1}$, in view of Lemma A3. (Note that the corresponding choice of τ_{xy}^\pm is then embedded into the definition of $R(\lambda)$.) In particular, if we take

$$\tau_{xy}^\pm := -\beta_{x,\perp}^\top \psi_x^\pm (\psi_y^\pm)^{-1},$$

as is permitted since $\psi_y^\pm \neq 0$, then it will follow that

$$\begin{aligned} R(\lambda) &= \tau^{*\top} \Psi^* \begin{bmatrix} [\theta^\top U_0(\lambda)]_+ \\ [\theta^\top U_0(\lambda)]_- \\ \theta_\perp^\top U_0(\lambda) \end{bmatrix} = \begin{bmatrix} \psi_y^+ & 0 & 0 \\ 0 & \psi_y^- & 0 \\ 0 & 0 & \beta_{x,\perp}^\top \Psi_{xx} \end{bmatrix} \begin{bmatrix} [\theta^\top U_0(\lambda)]_+ \\ [\theta^\top U_0(\lambda)]_- \\ \theta_\perp^\top U_0(\lambda) \end{bmatrix} \\ &= \begin{bmatrix} \psi_y^+ & 0 & 0 \\ 0 & \psi_y^- & 0 \\ 0 & 0 & I_{q-1} \end{bmatrix} \begin{bmatrix} [\theta^\top U_0(\lambda)]_+ \\ [\theta^\top U_0(\lambda)]_- \\ \beta_{x,\perp}^\top \Psi_{xx} \theta_\perp^\top U_0(\lambda) \end{bmatrix}. \end{aligned}$$

Defining

$$B_0(\lambda) := \begin{bmatrix} \theta^\top \\ \beta_{x,\perp}^\top \Psi_{xx} \theta_\perp^\top \end{bmatrix} U_0(\lambda)$$

we thus obtain a q -dimensional Brownian motion. To show that it has full rank variance matrix, since $\theta \neq 0$ and $\beta_{x,\perp}^\top \Psi_{xx} \theta_\perp^\top = 0$, it suffices to show that $\text{rk} \beta_{x,\perp}^\top \Psi_{xx} = q - 1$.

To that end, we first note that by the remark following (C30) above,

$$\beta_{x,\perp}^\top \Psi_{xx} = [0_{q-1}, \beta_{x,\perp}^\top] \Psi^x = [0_{q-1}, \beta_{x,\perp}^\top] P_{\beta_\perp} (+1) \theta_\perp, \quad (\text{C32})$$

The columns of $\Psi^x = P_{\beta_\perp} (+1) \theta_\perp$ are orthogonal to those of $e_{p,1}$ (by (C30) above) and of $\beta(+1)$; while $\text{rk}[e_{p,1}, \beta(+1)] = r + 1$, because $e_{p,1}^\top \beta_\perp(+1) \neq 0$ (by Lemma B.3(i) in DMW25) and so cannot be contained in the span of $\beta(+1)$. It follows that the $(p - 1)$ columns of Ψ^x span the $(q - 1)$ -dimensional subspace of \mathbb{R}^p that is orthogonal to $[e_{p,1}, \beta(+1)]$. Since the $(q - 1)$ columns of

$$\begin{bmatrix} 0_{q-1}^\top \\ \beta_{x,\perp}^\top \end{bmatrix} \in \mathbb{R}^{p \times (q-1)}$$

also span that subspace, it follows from (C32) that $[0_{q-1}, \beta_{x,\perp}^\top] \Psi^x = \beta_{x,\perp}^\top \Psi_{xx}$ has rank $q - 1$. Letting $D_\psi := \text{diag}\{\psi_y^+, \psi_y^-, I_{q-1}\}$, we have thus obtained

$$R(\lambda) = D_\psi S_q [e_1^\top B_0(\lambda)] B_0(\lambda),$$

where B_0 is a q -dimensional Brownian motion.

The final step is to recognise that, despite the nonlinearity on the r.h.s., we may still render this in terms of a standard (up to initialisation) Brownian motion by means of the usual Cholesky factorisation. Let Σ_B denote the variance of B_0 , and let L denote the (lower triangular) Cholesky root of Σ_B^{-1} , so that

$$W_0(\lambda) := LB_0(\lambda)$$

is a q -dimensional standard (up to initialisation) Brownian motion. Partitioning L and defining L^* as

$$L = \begin{bmatrix} \ell_1 & 0 \\ \ell_{(2),1} & L_{(2)} \end{bmatrix}, \quad L^* := \begin{bmatrix} \ell_1 & 0 & 0 \\ 0 & \ell_1 & 0 \\ \ell_{(2),1} & \ell_{(2),1} & L_{(2)} \end{bmatrix},$$

where $\ell_{(1)} > 0$ is scalar, and $L_{(2)} \in \mathbb{R}^{(q-1) \times (q-1)}$, and partitioning $I_q = [e_{q,1}, E_{q-1}]$, we obtain

$$\begin{aligned} L^* D_\psi^{-1} R(\lambda) &= L^* S_q [e_1^\top B_0(\lambda)] B_0(\lambda) \\ &= \begin{bmatrix} \ell_1 & 0 & 0 \\ 0 & \ell_1 & 0 \\ \ell_{(2),1} & \ell_{(2),1} & L_{(2)} \end{bmatrix} \begin{bmatrix} [e_{q,1}^\top B_0(\lambda)]_+ \\ [e_{q,1}^\top B_0(\lambda)]_- \\ E_{q-1}^\top B_0(\lambda) \end{bmatrix} \\ &= \begin{bmatrix} [\ell_1 e_{q,1}^\top B_0(\lambda)]_+ \\ [\ell_1 e_{q,1}^\top B_0(\lambda)]_- \\ (\ell_{(2),1} e_{q,1}^\top + L_{(2)} E_{q-1}^\top) B_0(\lambda) \end{bmatrix} \\ &= \begin{bmatrix} [e_{q,1}^\top W_0(\lambda)]_+ \\ [e_{q,1}^\top W_0(\lambda)]_- \\ E_{q-1}^\top W_0(\lambda) \end{bmatrix} = S_q [e_{q,1}^\top W_0(\lambda)] W_0(\lambda) = W_0^*(\lambda). \quad (\text{C33}) \end{aligned}$$

Hence the result for R holds with $Q = L^* D_\psi^{-1}$.

To obtain the desired representation for Y , we first invert (C33) to write

$$\tau^{*\top} Z^*(\lambda) = R(\lambda) = D_\psi (L^*)^{-1} W_0^*(\lambda).$$

Let $E_{d,2}$ denote the first two columns of I_d . Because the first two rows of each of $(L^*)^{-1}$, D_ψ and $\tau^{*\top}$ are zero everywhere except for the (1, 1) and (2, 2) elements, we have

$$E_{q+1,2}^\top D_\psi (L^*)^{-1} = \begin{bmatrix} \ell_1^{-1} \psi_y^+ & 0 & 0_{1 \times (q-1)} \\ 0 & \ell_1^{-1} \psi_y^- & 0_{1 \times (q-1)} \end{bmatrix}$$

and $E_{q+1,2}^\top \tau^{*\top} = E_{p+1,2}^\top$. Hence

$$\begin{aligned} \begin{bmatrix} [Y(\lambda)]_+ \\ [Y(\lambda)]_- \end{bmatrix} &= E_{p+1,2}^\top Z^*(\lambda) = E_{q+1,2}^\top \tau^{*\top} Z^*(\lambda) \\ &= E_{q+1,2}^\top R(\lambda) = \begin{bmatrix} \ell_1^{-1} \psi_y^+ [e_{q,1}^\top W_0(\lambda)]_+ \\ \ell_1^{-1} \psi_y^- [e_{q,1}^\top W_0(\lambda)]_- \end{bmatrix} \end{aligned}$$

whence the claim follows with $\omega^\pm = \ell_1^{-1} \psi_y^\pm > 0$. \square

Proof of Lemma A5. Since $\mathcal{W}_0 = 0$, we have $W_0 = W$, a q -dimensional standard Brownian motion (initialised at zero). To reduce the notational clutter, we will drop the '0' subscript from \overline{W}_0^* and \overline{V}_0^* throughout what follows.

We first consider \overline{S}_V^* . We note that a realisation of the positive semi-definite matrix \overline{S}_V^* is rank deficient if and only if there exists (for that realisation) an $a \in \mathbb{R}^{q+1}$ such that

$$0 = a^\top \overline{S}_V^* a = \int_0^1 [a^\top \overline{V}^*(s)]^2 ds.$$

Since $\overline{V}^*(s) = \int_0^s \overline{W}^*(\lambda) d\lambda$ has continuous paths, the preceding implies that

$$0 = a^\top \overline{V}^*(s) = a^\top \int_0^s \overline{W}^*(\lambda) d\lambda$$

for all $s \in [0, 1]$; and hence, differentiating with respect to s , that

$$a^\top \overline{W}^*(\lambda) = 0$$

for all $\lambda \in [0, 1]$. Since \overline{W}^* itself has continuous paths, a realisation of \overline{S}_W^* is rank deficient only if there exists an a such that the preceding condition holds. Hence, it suffices to show that

$$\mathbb{P}\{\exists a \in \mathbb{R}^{q+1} \text{ s.t. } a^\top \overline{W}^*(\lambda) = 0, \forall \lambda \in [0, 1]\} = 0. \quad (\text{C34})$$

Since \overline{W}^* is the residual from an $L^2([0, 1])$ projection of (each element of) the $(q + 1)$ -dimensional process

$$W^*(\lambda) = \begin{bmatrix} [W_1(\lambda)]_+ \\ [W_1(\lambda)]_- \\ W_{-1}(\lambda) \end{bmatrix}$$

onto a constant, the event referred to in (C34) holds only if (for a given realisation) there exists a $b = (b_1, b_{-1})^\top \in \mathbb{R}^{q+2}$ such that

$$0 = b^\top \begin{bmatrix} 1 \\ W^*(\lambda) \end{bmatrix} = b_1 + b_{-1}^\top W^*(\lambda)$$

for all $\lambda \in [0, 1]$. Taking $\lambda = 0$, we see this implies $b_1 = 0$. Hence it suffices for (C34) to show that

$$\mathbb{P}\{\exists a \in \mathbb{R}^{q+1} \text{ s.t. } a^\top W^*(\lambda) = 0, \forall \lambda \in [0, 1]\} = 0. \quad (\text{C35})$$

To that end, we note that by Tanaka's formula (Theorem VI.1.2 in [19]) that

$$[W_1(\lambda)]_\pm = \int_0^\lambda \mathbf{1}^\pm [W_1(s)] ds + \frac{1}{2} L_{W_1}(\lambda, 0)$$

where $L_{W_1}(\lambda, x)$ denotes the local time of W_1 at time $\lambda \in [0, 1]$ and spatial point $x \in \mathbb{R}$, which is a continuous increasing process (for each x fixed). It follows that W^* is a vector semimartingale, with quadratic variation process

$$Q(\lambda) := \begin{bmatrix} \int_0^\lambda \mathbf{1}^+[W_1(s)] \, ds & 0 & 0 \\ 0 & \int_0^\lambda \mathbf{1}^-[W_1(s)] \, ds & 0 \\ 0 & 0 & \lambda I_{q-1} \end{bmatrix}.$$

We note that $Q(1)$ is rank deficient only if one of its first two diagonal entries is zero, which in turn requires that either $\min_{\lambda \in [0,1]} W_1(\lambda) \geq 0$ or $\max_{\lambda \in [0,1]} W_1(\lambda) \leq 0$. But since W_1 is a standard Brownian motion (initialised at zero), both of these events have zero probability. It follows by a standard characterisation of quadratic variation (Definition IV.1.20 in [19]) that for $\Delta_{m,i} W^* := W^*(\frac{i}{m}) - W^*(\frac{i-1}{m})$

$$Q_m(1) := \sum_{i=1}^m \Delta_{m,i} W^* (\Delta_{m,i} W^*)^\top \xrightarrow{p} Q(1)$$

as $m \rightarrow \infty$ and thus, since $W^*(0) = 0$, that

$$\begin{aligned} & \mathbb{P}\{\exists a \in \mathbb{R}^{q+1} \text{ s.t. } a^\top W^*(\lambda) = 0, \forall \lambda \in [0, 1]\} \\ & \leq \mathbb{P}\{\exists a \in \mathbb{R}^{q+1} \text{ s.t. } a^\top \Delta_{m,i} W^* = 0, \forall i \in \{1, \dots, m\}\} \\ & = \mathbb{P}\{\text{rk} Q_m(1) < q + 1\} \\ & \rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$. Thus (C35) holds. □