UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Cointegration without Unit Roots

James A. Duffy
Jerome R. Simons

Abstract

It has been known since Elliott (1998) that standard methods of inference on cointegrating relationships break down entirely when autoregressive roots are near but not exactly equal to unity. We consider this problem within the framework of a structural VAR, arguing this it is as much a problem of identification failure as it is of inference. We develop a characterisation of cointegration based on the impulse response function, which allows long-run equilibrium relationships to remain identified even in the absence of exact unit roots. Our approach also provides a framework in which the structural shocks driving the common persistent components continue to be identified via long-run restrictions, just as in an SVAR with exact unit roots. We show that inference on the cointegrating relationships is affected by nuisance parameters, in a manner familiar from predictive regression; indeed the two problems are asymptotically equivalent. By adapting the approach of Elliott, Müller and Watson (2015) to our setting, we develop tests that robustly control size while sacrificing little power (relative to tests that are efficient in the presence of exact unit roots).

Reference Details

CWPE	2332
Published	12 April 2023
Key Words	Cointegration, inference, near-integrated processes, least-favourable distributions, nuisance parameters, power
JEL Codes	C01, C32, C40, C80
Website	www.econ.cam.ac.uk/cwpe

Cointegration without Unit Roots

James A. Duffy* — University of Oxford

Jerome R. Simons† — University of Cambridge

April 2023

Abstract

It has been known since Elliott (1998) that standard methods of inference on cointegrating relationships break down entirely when autoregressive roots are near but not exactly equal to unity. We consider this problem within the framework of a structural VAR, arguing this it is as much a problem of identification failure as it is of inference. We develop a characterisation of cointegration based on the impulse response function, which allows long-run equilibrium relationships to remain identified even in the absence of exact unit roots. Our approach also provides a framework in which the structural shocks driving the common persistent components continue to be identified via long-run restrictions, just as in an SVAR with exact unit roots. We show that inference on the cointegrating relationships is affected by nuisance parameters, in a manner familiar from predictive regression; indeed the two problems are asymptotically equivalent. By adapting the approach of Elliott, Müller and Watson (2015) to our setting, we develop tests that robustly control size while sacrificing little power (relative to tests that are efficient in the presence of exact unit roots).

JEL: C01, C32, C40, C80

Key words: Cointegration, inference, near-integrated processes, least-favourable distributions, nuisance parameters, power

The authors thank G. Bårdsen, V. Berenguer-Rico, S. Bond, P. Boswijk, V. Carvalho, G. Chevillon, A. Harvey, O. Linton, S. Mavroeidis, U. Müller, B. Nielsen, A. Onatskiy and participants at seminars at Amsterdam, NTNU (Trondheim), ESSEC (Cergy), Essex, Southampton, Sydney, UNSW and Oxford for helpful comments on earlier drafts of this work. Jerome R. Simons gratefully acknowledges funding from the Keynes Fund.

*Department of Economics and Corpus Christi College; `james.duffy@economics.ox.ac.uk`

†Faculty of Economics and St Edmund's College; `jrs89@cam.ac.uk`

Contents

1 Introduction

Since its development in the late 1980s, the cointegrated vector autoregressive model has been widely applied to the modelling of macroeconomic time series – a testament to its ability to account for both the short- and long-run dynamics of these series in a unified way (Hendry, 1986; Engle and Granger, 1987; Johansen, 1995). By allowing for one or more autoregressive roots at unity, the model is able to match two key features of these series: their high degree of persistence, which gives rise to their characteristically 'random wandering' behaviour, and the tendency for economically related series to move together, such that certain linear combinations of these series – given by the cointegrating relationships – are markedly less persistent than the series themselves. Dual to this, the model provides a framework for identifying the structural shocks whose permanent effects generate these patterns of co-movement (Blanchard and Quah, 1989; King, Plosser, Stock, and Watson, 1991), which has been widely used in empirical studies.

Cointegrating relationships are often of intrinsic interest because macroeconomic theories make definite predictions about the existence and magnitude of the long-run equilibrium relationships that these embody. (See e.g. Fuhrer and Moore, 1995; Maccini, Moore, and Schaller, 2004; Shiue and Keller, 2007.) For the purposes of estimating these relationships, and thus of testing the predictions of such theories, a variety of efficient methods exist, such as FM-OLS (Phillips and Hansen, 1990), DOLS (Stock and Watson, 1993), and (rank-imposed) maximum likelihood estimation of the VAR itself (Johansen, 1995). However, all these methods rely on a common assumption that the data is generated by a VAR with a certain number of exact unit roots. Elliott (1998) showed that should this assumption fail only slightly – such that some roots are merely 'close' but not exactly equal to unity – then inferences based on these methods can suffer from severe size distortions. His findings are particularly disturbing because this problem arises even in a VAR with roots that are 'nearly' unity, in the sense of lying within an $O(n^{-1})$ neighbourhood thereof, which is practically indistinguishable from the same model with exact unit roots.

The present work addresses the problem posed by Elliott (1998): how can one perform valid inference on the cointegrating relationships in an (S)VAR, when the dominant roots in that model may not be exactly unity? In view of the significance of Elliott's findings, it is perhaps surprising that only a few previous contributions have also attempted to respond to them: most notably Wright (2000), Magdalinos and Phillips (2009), Müller and Watson (2013), Franchi and Johansen (2017), and Hwang and Valdés (2023). The approach taken in this paper is quite different from that taken in those previous works, which have largely followed Elliott (1998) in framing the problem as an inferential one, which might be solved merely by using appropriately modified estimators and tests. Instead, our view is that the problem is at least as much one of identification failure as it is of inference. Indeed, the usual definition of cointegration – in terms of linear combinations of series that eliminate their common integrated components – becomes meaningless as soon as the largest characteristic root in a VAR departs even slightly from unity. (See Section 2.5 below for a further discussion of how our contribution relates to those previous works.)

Our first task is thus to develop a characterisation of cointegration, based on the impulse response function implied by the VAR, that remains meaningful in a model with some roots near but not necessarily equal to unity. In a p-dimensional VAR with q roots near unity, one can always identify a $p - q = r$-dimensional subspace S_r, such that the decay of the impulse response function in the directions contained in S_r is more rapid than it is in all other directions (Sections 2.1–2.3). We term S_r the *quasi-cointegrating space* (QCS) of the VAR. When those q roots are exactly unity, the QCS coincides exactly with the cointegrating space – and when they are modelled as being local to unity, i.e. lying within a $O(n^{-1})$ neighbourhood of unity, the quasi-cointegrating relations are exactly those that eliminate the common near stochastic trends from the system.

While quasi-cointegration is not the only conceivable way of extending cointegration to a wider domain, our approach has the further advantage of maintaining the duality, that exists in an SVAR with exact unit roots, between the identification of the long-run equilibrium relationships between the series, and of the subvector of structural shocks whose common permanent effects underpin those relationships. In this way, we simultaneously extend both cointegration, and the use of long-run identifying restrictions, to an SVAR without exact unit roots, by allowing that a subset of the structural shocks may have effects that are highly persistent, rather than permanent, where persistence is understood in terms of the (relative) decay rate of the impact of those shocks (Section 2.4). We thereby show how these long-run restrictions, which are often thought to be available only in the case of exact unit roots (as noted e.g. in Kilian and Lütkepohl, 2017, Sec. 10.5.1), remain a viable approach to identification even without this auxiliary assumption.

Inference on the QCS is complicated by the presence of nuisance parameters, which measure the proximity of the dominant roots of the VAR to unity (Section 3). This problem is similar to that which arises in predictive regressions, when the regressors have an unknown but possibly high degree of persistence, such as has been studied e.g. by Cavanagh, Elliott, and Stock (1995), Campbell and Yogo (2006), Jansson and Moreira (2006), Phillips and Lee (2013), Phillips (2014), and Kostakis, Magdalinos, and Stamatogiannis (2015). In fact, we show that the problem of inference on the QCS is asymptotically equivalent to that of inference in a predictive regression: both converge to a common limiting experiment under an appropriate local parametrisation. This equivalence permits methods that have been developed for predictive regression – for which there are a great many – to be transposed the present setting. Our problem also fits within the general framework of Elliott, Müller, and Watson (2015), and by adapting their approach to the present setting, we obtain tests and confidence intervals that are effectively free of any size distortions, while sacrificing little power relative to the efficient estimators, even when the data is generated with exact unit roots. We also extend the mixed normality of the maximum-likelihood estimator, when the correct number of unit roots are imposed, to the case where the correct values of the dominant roots are imposed. This provides efficient likelihood-based tests and confidence intervals for cases where one is willing to take a stand on the values of these parameters.

The finite-sample performance of our procedure is evaluated through a series of simulation exercises, and illustrated with an empirical application to the expectations theory of the term structure (Section 4). Proofs of all technical results appear in the appendices.

Notation. All limits are taken as $n \to \infty$ 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 $D[0,1]$' to denote that $\{X_n\}$ converges weakly to X, where these are considered as random elements of $D[0,1]$, the space of cadlag functions $[0,1] \to \mathbb{R}^m$, equipped with the uniform topology. $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$; all matrix norms are induced by the corresponding vector norms. For X a random variable and $p \geq 1$, $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$. $M^{1/2}$ denotes the principal square root of a positive semidefinite matrix M.

2 'Cointegration' in a VAR without unit roots

2.1 Model and assumptions

The data generating process (DGP) for the observed series $\{y_t\}_{t=1}^n$ is a kth order vector autoregressive (VAR) model, written in unobserved components form as

$$y_t = \mu + \delta t + x_t \qquad\qquad x_t = \sum_{i=1}^{k} \Phi_i x_{t-i} + \varepsilon_t \tag{2.1}$$

where ε_t, x_t and y_t are p-dimensional random vectors. The reduced-form shocks $\{\varepsilon_t\}$ depend on an underlying (p-dimensional) vector of i.i.d. and mutually uncorrelated structural shocks $\{w_t\}$, via

$$\varepsilon_t = \Upsilon w_t, \tag{2.2}$$

so that (2.1)–(2.2) comprise a structural VAR (or SVAR). Let $\Phi(\lambda) := I\lambda^k - \sum_{i=1}^k \Phi_i \lambda^{k-i}$ denote the characteristic polynomial associated to (2.1); we shall refer to any λ for which $\det \Phi(\lambda) = 0$ as a 'root of Φ'. Let $\mathbf{\Phi} := (\Phi_1, \Phi_2, \ldots, \Phi_k) \in \mathbb{R}^{p \times kp}$. We generally maintain the following.

Assumption DGP. *$\{y_t\}_{t=1}^n$ and $\{x_t\}_{t=1}^n$ are generated under (2.1)–(2.2), where:*

DGP1 $\det \Phi(\lambda) \neq 0$ *for all* $|\lambda| > 1$*;*

DGP2 $\{w_t\}$ *is i.i.d. with* $\mathbb{E}w_t = 0$ *and* $\mathbb{E}w_t w_t^\mathsf{T} = I_p$*, and* $\Sigma := \Upsilon\Upsilon^\mathsf{T}$ *is positive definite;*

DGP3 $x_0 = x_{-1} = \cdots = x_{-k+1} = 0$.

We say that a d_z-dimensional process $\{z_t\}$ is integrated of order zero, denoted $z_t \sim I(0)$, if there exists a deterministic process $\{\mu_t\}$ such that $n^{-1/2} \sum_{s=1}^{\lfloor nr \rfloor} (z_s - \mu_s) \rightsquigarrow B(r)$, for B a d_z-dimensional Brownian motion. Letting Δ^d denote the dth order temporal differencing operator, we say that z_t is *integrated of order* d, denoted $z_t \sim I(d)$, if $\Delta^d z_t \sim I(0)$. We say $\{z_t\}$ is *nearly integrated* if $n^{-1/2}(z_{\lfloor nr \rfloor} - \mu_{\lfloor nr \rfloor}) \rightsquigarrow \int_0^r \mathrm{e}^{C(r-s)} \mathrm{d}B(s)$ for some $C \in \mathbb{R}^{d_z \times d_z}$.

2.2 Cointegration: the model with unit roots

Cointegration analysis is concerned with how linear combinations of $I(d)$ processes can yield processes that are themselves only $I(d-b)$ for some $0 < b \leq d$. The reduced persistence and more rapid mean reversion of the latter is interpreted as evidence of a long-run equilibrium relationship between the original processes. Here, we focus exclusively on the special but practically important case of $I(1)$ processes having linear combinations that are $I(0)$, reserving the term *cointegration* exclusively for this case. As is well-known, the VAR model (2.1) is able to generate cointegrated

$I(1)$ processes under the following assumptions, which define the $I(0)/I(1)$ cointegrated VAR (CVAR) model.

Assumption CV.

CV1 *Φ has q roots at (real) unity, and all others strictly inside the unit circle.*

CV2 $\operatorname{rk}\Phi(1) = p - q =: r$

By the Granger–Johansen representation theorem (GJRT; see e.g. Johansen 1995, Thm 4.2 and Cor. 4.3), the preceding is necessary and sufficient for $y_t \sim I(1)$, and for there to exist a rank r matrix $\beta \in \mathbb{R}^{p\times r}$ of cointegrating relationships, such that $\beta^{\mathsf{T}} y_t \sim I(0)$. The matrix β is identified only up to its column space, $\mathrm{CS} := \operatorname{sp}\beta$, termed the *cointegrating space* (CS). Two equivalent characterisations of the cointegrating space, the first of which is definitional and the second of which follows immediately from the GJRT, are

(C.i) $b^{\mathsf{T}} y_t \sim I(0)$ if and only if $b \in \mathrm{CS}$; and

(C.ii) $\mathrm{CS} = \operatorname{sp}\Phi(1)^{\mathsf{T}} = \{\ker \Phi(1)\}^{\perp}$.

Our objective in this paper is to estimate the CS, or a space sharing its key properties, in a setting more general than that of CV. For this purpose, we next recall two further characterisations of the CS that extend beyond the setting of CV, in a way that the preceding do not.[1] The third characterisation is in terms of the impulse response function of $\{y_t\}$ with respect to the reduced-form or structural disturbances (i.e. $\{\varepsilon_s\}$ or $\{w_s\}$), denoted

$$\mathrm{IRF}_s^{\varepsilon} := \frac{\partial y_{t+s}}{\partial \varepsilon_t} = \frac{\partial x_{t+s}}{\partial \varepsilon_t} \qquad\qquad \mathrm{IRF}_s^{w} := \frac{\partial y_{t+s}}{\partial w_t} = \mathrm{IRF}_s^{\varepsilon}\frac{\partial w_t}{\partial \varepsilon_t} = \mathrm{IRF}_s^{\varepsilon}\Upsilon,$$

For a given $b \in \mathbb{R}^p$, the product $b^{\mathsf{T}}\mathrm{IRF}_s^{w}$ gives the response of the linear combination $b^{\mathsf{T}} y_{t+s}$ to w_s. The rate at which $b^{\mathsf{T}}\mathrm{IRF}_s^{w}$ (or $b^{\mathsf{T}}\mathrm{IRF}_s^{\varepsilon}$) decays as the horizon s diverges provides a measure of the persistence of the series $\{b^{\mathsf{T}} y_t\}$. Now let $m < p$, and define $S_m \subset \mathbb{R}^p$ to be an m-dimensional linear subspace such that for every $b \in S_m$ and $c \notin S_m$,

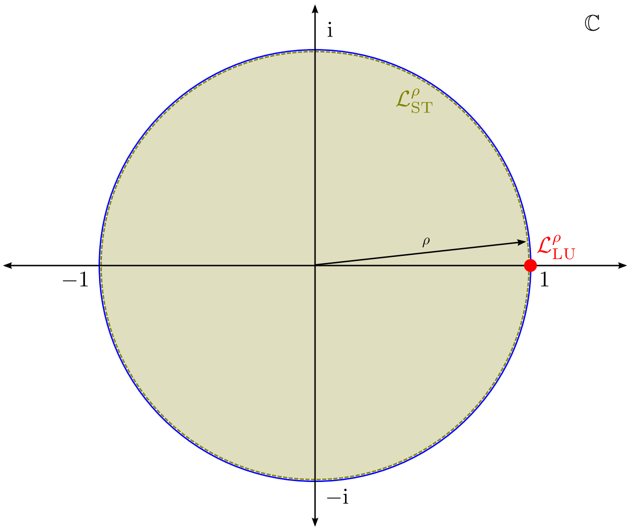
$$\lim_{s\to\infty} \frac{\|b^{\mathsf{T}}\mathrm{IRF}_s^{w}\|}{\|c^{\mathsf{T}}\mathrm{IRF}_s^{w}\|} = 0. \tag{2.3}$$

When it exists, S_m collects those m linear combinations of y_t that are, in the sense of (2.3), the least persistent. While IRF_s^{w} evidently depends on Υ, the subspace S_m itself is *invariant* to Υ, and hence to the scheme used to identify the structural shocks; indeed, we may equivalently characterise S_m in terms of the reduced-form impulse responses, with $\mathrm{IRF}_s^{\varepsilon}$ taking the place of IRF_s^{w} in (2.3). Under CV, S_m with $m = r$ exists and is unique, and moreover

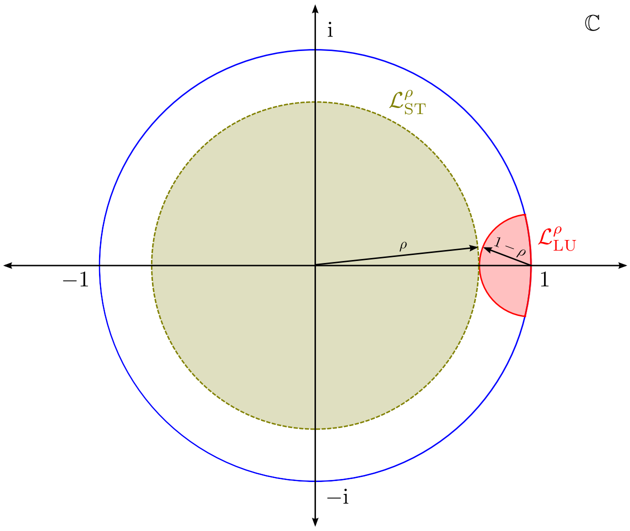
(C.iii) $\mathrm{CS} = S_r$

(see Lemma A.3). In other words, the cointegrating space is spanned by the vectors giving the r least persistent linear combinations of y_t.

[1] While the arguments that lead to these characterisations may be familiar to the reader, for completeness formal statements and proofs of the results underlying the discussion that follows appear in Appendix A.



(a) Cointegration ($\rho = 1$)



(b) Quasi-cointegration ($\rho < 1$)

Figure 1: $\mathcal{L}^{\rho}_{\text{LU}}$ and $\mathcal{L}^{\rho}_{\text{ST}}$, from (2.4), in the complex plane

Our final characterisation of the cointegrating space provides the basis for its estimation in settings more general than CV; it derives from the application of a spectral decomposition to the companion form representation of (2.1) (see Lemma A.1). Define the disjoint sets

$$\mathcal{L}^{\rho}_{\text{LU}} := \{z \in \mathbb{C} \mid |z| \leq 1 \text{ and } |z-1| \leq 1-\rho\} \qquad \mathcal{L}^{\rho}_{\text{ST}} := \{z \in \mathbb{C} \mid |z| < \rho\}, \tag{2.4}$$

so that for a given $\rho \leq 1$ (but close to unity), $\mathcal{L}^{\rho}_{\text{LU}}$ defines a closed set of points on or inside the unit complex circle, within a distance $1-\rho$ of real unity, and $\mathcal{L}^{\rho}_{\text{ST}}$ defines an open ball of radius ρ centred at zero, as depicted in Figure 1 separately for the cases where $\rho < 1$ and $\rho = 1$. Now suppose that Φ has q roots in $\mathcal{L}^{\rho}_{\text{LU}}$ and all others in $\mathcal{L}^{\rho}_{\text{ST}}$ for some $\rho \leq 1$. Under CV1 this setup holds with $\rho = 1$, so that these sets are as in Figure 1(a). Since $\mathcal{L}^{\rho}_{\text{LU}}$ and $\mathcal{L}^{\rho}_{\text{ST}}$ are disjoint, there exist real matrices

$$\underset{(p\times kp)}{R} := [\underset{(p\times q)}{R_{\text{LU}}} \quad R_{\text{ST}}] \qquad \underset{(p\times kp)}{L} := [\underset{(p\times q)}{L_{\text{LU}}} \quad L_{\text{ST}}] \qquad \underset{(kp\times kp)}{\Lambda} := \operatorname{diag}\{\underset{(q\times q)}{\Lambda_{\text{LU}}},\ \Lambda_{\text{ST}}\} \tag{2.5}$$

such that: (a) the eigenvalues of Λ_{LU} and Λ_{ST} correspond to the roots of Φ, and lie in $\mathcal{L}^{\rho}_{\text{LU}}$ and $\mathcal{L}^{\rho}_{\text{ST}}$ respectively; (b) the triple $(R_{\text{LU}}, \Lambda_{\text{LU}}, L_{\text{LU}})$ satisfies

$$R_{\text{LU}}\Lambda^{k}_{\text{LU}} - \sum_{i=1}^{k} \Phi_i R_{\text{LU}}\Lambda^{k-i}_{\text{LU}} = 0 \qquad \Lambda^{k}_{\text{LU}}L^{\mathsf{T}}_{\text{LU}} - \sum_{i=1}^{k} \Lambda^{k-i}_{\text{LU}}L^{\mathsf{T}}_{\text{LU}}\Phi_i = 0; \tag{2.6}$$

and (c) the (reduced-form) impulse response function of y_t is can be written as

$$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \text{IRF}^{\varepsilon}_{s} = R\Lambda^{k-1+s}L^{\mathsf{T}} = R_{\text{LU}}\Lambda^{k-1+s}_{\text{LU}}L^{\mathsf{T}}_{\text{LU}} + R_{\text{ST}}\Lambda^{k-1+s}_{\text{ST}}L^{\mathsf{T}}_{\text{ST}}, \tag{2.7}$$

(see Lemma A.1 and the subsequent remarks). Under CV, we have $\Lambda_{\text{LU}} = I_q$ and $\operatorname{rk} R_{\text{LU}} = \operatorname{rk} L_{\text{LU}} = q$ (see Lemma A.3); it follows that the limits of the structural and reduced-form IRFs as the horizon $s \to \infty$ are

$$\lim_{s\to\infty} \mathrm{IRF}^{w}_{s} = R_{\mathrm{LU}} L_{\mathrm{LU}}^{\mathsf{T}} \Upsilon, \qquad \lim_{s\to\infty} \mathrm{IRF}^{\varepsilon}_{s} = R_{\mathrm{LU}} L_{\mathrm{LU}}^{\mathsf{T}}, \tag{2.8}$$

yielding our final characterisation of the cointegrating space as

(C.iv) $\mathrm{CS} = (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$.

Consistent with the discussion of S_r above, the matrix Υ – and thus the identification of the structural shocks – plays no role here; the cointegrating space depends only on the column space of the matrices appearing on the r.h.s. of (2.8), and so is invariant to post-multiplication by any full-rank matrix.[2]

2.3 Cointegration without unit roots

Having thoroughly characterised cointegration in a VAR with q exact unit roots, we may now return to our motivating problem: that of inference on the cointegrating relationships when those q roots are allowed to be merely 'near' to (real) unity. As shown by Elliott (1998), even if we consider the apparently favourable case of a sequence of models whose roots drift towards unity as per $\Lambda_{\mathrm{LU}} = I + n^{-1}C$, standard efficient estimators of the cointegrating relationships (such as FM-OLS, DOLS and ML) are in general asymptotically biased, and the associated inferences severely size distorted. This lack of robustness is particularly disturbing because it arises in VARs that cannot be consistently distinguished from those with exact unit roots, preventing the extent of this problem from being empirically evaluated.

Our view is that this problem is fundamentally one of identification, whose resolution demands a characterisation of cointegration that retains its meaning over a wider domain than merely a VAR with exact unit roots. Neither (C.i) nor (C.ii), which are implicitly utilised by the standard estimators, are fit for this purpose. For if the largest q roots of Φ were strictly inside the unit circle – as would now be permitted – then *all* linear combinations of y_t would be $I(0)$, $\Phi(1)$ would have full rank, and hence both characterisations would identify the cointegrating space with the whole of $\mathbb{R}^p$. In contrast, both (C.iii) and (C.iv) would continue to identify that (r-dimensional) subspace of linear combinations of $\{y_t\}$ having the least persistence, and thereby continue to capture the long-run equilibrium relationships between these series. Accordingly, they provide a sound basis on which to extend 'cointegration' to VARs without exact unit roots.

To this end, we now consider relaxing CV above as follows

Assumption QC. *Let $\rho \leq 1$ be given.*

QC1 *Φ has q roots in $\mathcal{L}^{\rho}_{\mathrm{LU}}$, and all others in $\mathcal{L}^{\rho}_{\mathrm{ST}}$.*

Let Λ_{LU} denote a real $(q \times q)$ matrix whose eigenvalues correspond to the roots of Φ that are in $\mathcal{L}^{\rho}_{\mathrm{LU}}$, and let R_{LU} and L_{LU} be $p \times q$ matrices that satisfy (2.5)–(2.7).

QC2 $\operatorname{rk} R_{\mathrm{LU}} = \operatorname{rk} L_{\mathrm{LU}} = q$ *and Λ_{LU} is diagonalisable.*

[2] As Example 2.2 below illustrates, if $\{\varepsilon_t\}$ follows a finite-order MA process – such as typically arises when $\{y_t\}$ is generated by a linear state-space model – then the autoregressive coefficients, and the associated characteristic polynomial, alone carry all the information required to recover CS (and similarly for the quasi-cointegrating space introduced in the next section).

QC1 is plainly the analogue of CV1: whereas we previously assumed q roots at unity, we now allow for q roots in the vicinity of unity; indeed CV is a special case of QC with $\rho = 1$ (Lemma A.3).[3] By allowing for the possibility that $\rho < 1$, we move from panel (a) of Figure 1 to panel (b). The requirement that Λ_{LU} be diagonalisable purposely rules out series that are integrated of order two or higher (d'Autume, 1992), since these are also excluded under CV (which as noted above implies $\Lambda_{\mathrm{LU}} = I_q$). For $\rho < 1$ but 'close' to unity, a model satisfying QC will thus inherit the main qualitative features of the cointegrated VAR model: the high persistence of $\{y_t\}$, and the lesser persistence of r linear combinations of $\{y_t\}$, understood in terms of (2.3) above.

Accordingly, the subspace S_r spanned by the r 'least persistent' linear combinations of y_t retains an interpretation akin to that of the cointegrating space. These two objects coincide exactly in a VAR with unit roots (recall (C.iii) above), but only the former remains meaningfully interpretable when these roots are allowed to be merely near to unity, as entertained by QC. That S_r is always well defined under our assumptions is guaranteed by the following, the proof of which appears in Appendix A.

Proposition 2.1. *Suppose* DGP *and* QC *hold. Then* $S_r = (\mathrm{sp}\, R_{\mathrm{LU}})^{\perp}$.

We henceforth term the elements of S_r the *quasi-cointegrating relationships*, and refer to S_r itself as the *quasi-cointegrating space* (QCS), denoted

$$\mathrm{QCS} \coloneqq S_r = (\mathrm{sp}\, R_{\mathrm{LU}})^{\perp}.$$

We continue to reserve the term *cointegration* for the VAR with q exact unit roots. We also let $\beta \in \mathbb{R}^{p\times r}$ denote a matrix of rank r whose columns span the QCS, and which therefore has the property that $\beta^{\mathsf{T}} R_{\mathrm{LU}} = 0$.

There remains the question of how ρ might be chosen in practice as opposed to merely fixing it at unity. To build intuition on the choice of ρ, we consider the reduced-form IRF (2.7), and the allied decomposition

$$y_t - \mu - \delta t = x_t = \Phi_{\mathrm{LU}} z_{\mathrm{LU},t-1} + \Phi_{\mathrm{ST}} z_{\mathrm{ST},t-1} + \varepsilon_t \tag{2.9}$$

where $\Phi_{\mathrm{LU}} = R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^k$ and $\Phi_{\mathrm{ST}} = R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^k$, and the 'common trend' $z_{\mathrm{LU},t} \in \mathbb{R}^q$ and 'transitory' $z_{\mathrm{ST},t} \in \mathbb{R}^{kp-q}$ components follow

$$z_{\mathrm{LU},t} = \Lambda_{\mathrm{LU}} z_{\mathrm{LU},t-1} + \varepsilon_{\mathrm{LU},t} \qquad \varepsilon_{\mathrm{LU},t} \coloneqq L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t \tag{2.10a}$$

$$z_{\mathrm{ST},t} = \Lambda_{\mathrm{ST}} z_{\mathrm{ST},t-1} + \varepsilon_{\mathrm{ST},t} \qquad \varepsilon_{\mathrm{ST},t} \coloneqq L_{\mathrm{ST}}^{\mathsf{T}}\varepsilon_t \tag{2.10b}$$

under QC (see Lemma A.4). By imposing a lower bound on the eigenvalues of Λ_{LU}, the value ρ regulates the persistence of $z_{\mathrm{LU},t}$. Equivalently, via (2.7), ρ is interpretable in terms of the minimum half-life of the most persistent reduced-form shocks, $\varepsilon_{\mathrm{LU},t} = L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$, that drive y_t,

[3] The construction of $\mathcal{L}_{\mathrm{LU}}^{\rho}$ requires the q roots to be on or inside the unit circle. Though the theory developed here could also accommodate explosive roots – simply by redefining $\mathcal{L}_{\mathrm{LU}}^{\rho}$ as $\{z \in \mathbb{C} \mid |z-1| \leq 1-\rho\}$ – we have deliberately excluded such roots to be consistent with the greater empirical relevance of stationary departures from unit roots for most applications.

as being $\underline{h} := -\log 2/\log \rho$ periods. As discussed in the next section, $L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$ may be given a structural interpretation, in terms of the subset of the structural shocks w_t that are identified (by the relevant macroeconomic theory) as having highly persistent, and possibly permanent, effects.

In the extreme case where $\rho = 1$, $z_{\mathrm{LU},t}$ is an integrated process and these shocks will have permanent effects, i.e. $\underline{h} = \infty$; but in general reasonable finite choices for $\underline{h}$ will be available, with smaller values of $\underline{h}$ (and hence of ρ) affording greater robustness to departures from exact unit roots. This choice will itself depend on the application at hand. For example, in a macroeconomic context it would be appropriate to allow that the most persistent shocks to $\{y_t\}$ may not have permanent effects, but still have a half-life longer than the average duration of the business cycle: with postwar US data of annual frequency, this corresponds to setting $\underline{h} = 8$ and thus $\rho = 2^{-1/\underline{h}} = 0.917$; or for quarterly data, $\rho = 0.979$.

2.4 Implications for long-run identifying restrictions

Up to this point, we have motivated quasi-cointegration in essentially descriptive terms, as a means of extending the key time-series properties of cointegrated systems to a wider domain. Of course, other characterisations of 'cointegration' or 'long-run equilibria' in time series models might be developed, and used to extend the concept in alternative directions. From the point of view of empirical macroeconomics, however, a particularly advantageous feature of quasi-cointegration is that it is grounded in the SVAR, what is arguably the workhorse model in this field (Kilian and Lütkepohl, 2017). Accordingly, as we shall now discuss, quasi-cointegration both provides a framework for identifying structural shocks via long-run restrictions, and a means of extracting testable long-run predictions from macroeconomic theories, that does not require one to take a stand on the presence or absence of exact unit roots in the underlying SVAR – a matter on which economic theory is largely silent. Indeed, it does so more generally in the VARMA, or approximate VAR, representations implied by linear state-space models, and is thus relevant to a broad class of (linearised) structural macroeconomic models.

2.4.1 Structural impulse response functions

A long-standing approach to the identification of structural IRFs involves 'long-run restrictions', which demarcate shocks according to whether they are permitted to have permanent effects (e.g. Blanchard and Quah, 1989; King et al., 1991; Gali, 1999; Christiano, Eichenbaum, and Vigfusson, 2006). These typically derive from an underlying theoretical model in which one or more state variables – such as total factor productivity or the natural rate of interest – are assumed to have a stochastic trend. In equilibrium, these trends are imparted to some of the endogenous variables, upon which the driving structural shocks must therefore have permanent effects. When formulated in the setting of an SVAR, q stochastic trends manifest as q unit roots and limiting impulse response matrices (2.8) of reduced rank, which identify the relevant subset of the structural shocks as $L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$.

Being expressed in terms of limiting impulse response matrices, these restrictions are generally thought to require the presence of exact unit roots to provide a viable approach to identification (see e.g. the discussion in Kilian and Lütkepohl, 2017, Sec. 10.5.1). However, since these restrictions are dual to the cointegrating relations – the former relates to the column span, the latter to the row span, of the long-run IRF (2.8) – the former are just as amenable to being extended beyond the setting of exact unit roots as is the latter. As the examples below illustrate, if certain state variables are permitted to, more generally, follow a highly persistent but not exactly integrated autoregressive process, we obtain a SVAR(MA) process with q roots near unity (see also Campbell, 1994, for a discussion with $q = 1$). The decomposition (2.7), which is generally available under QC, then provides a means of isolating the driving structural shocks from those having comparatively transient effects, with $L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$ yielding q linear combinations of the former.

Example 2.1. Gali (1999) develops a stylised DSGE model of labour market dynamics in the presence of a nominal rigidity, with two (i.i.d. and mutually independent) structural shocks: one, η_t, to the underlying technology process $\{Z_t\}$, which evolves as

$$\log Z_t = \rho_z \log Z_{t-1} + \eta_t$$

with $\rho_z = 1$, i.e. as a random walk, and the other, ξ_t, to the growth rate of the money supply. The model implies the following VAR(1) representation for log productivity a_t and hours n_t,

$$\begin{aligned} x_t \coloneqq \begin{bmatrix} a_t \\ n_t \end{bmatrix} &= c + \begin{bmatrix} \rho_z & \rho_z(1-\varphi) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ n_{t-1} \end{bmatrix} + \varphi^{-1} \begin{bmatrix} \varphi - 1 & \gamma(\varphi-1)+1 \\ 1 & -(1-\gamma) \end{bmatrix} \begin{bmatrix} \xi_t \\ \eta_t \end{bmatrix} \\ &\eqqcolon c + \Phi x_{t-1} + \varepsilon_t. \end{aligned} \tag{2.11}$$

When $\rho_z = 1$, only η_t has a permanent effect on the (log) level of productivity, which justifies an empirical strategy of identifying the technology shock, in a bivariate SVAR of productivity and hours, from the restriction that only it may have a permanent effect on productivity. However, such a strategy is also viable when ρ_z is merely near unity, as can be seen by applying the decomposition (2.7) to the VAR (2.11), which for $h \geq 1$ yields

$$\frac{\partial x_{t+h}}{\partial \varepsilon_t} = \Phi^h = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rho_z^h \begin{bmatrix} 1 & 1-\varphi \end{bmatrix} = \lambda_{\mathrm{LU}}^h r_{\mathrm{LU}} l_{\mathrm{LU}}^{\mathsf{T}} \tag{2.12}$$

with $\lambda_{\mathrm{LU}} = \rho_z$. In particular, irrespective of whether $\rho_z = 1$, $l_{\mathrm{LU}}^{\mathsf{T}}$ recovers the technology shock, since

$$l_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t = \varphi^{-1} \begin{bmatrix} 1 & 1-\varphi \end{bmatrix} \begin{bmatrix} \varphi - 1 & \gamma(\varphi-1)-1 \\ 1 & -(1-\gamma) \end{bmatrix} \begin{bmatrix} \xi_t \\ \eta_t \end{bmatrix} = \eta_t. \qquad \square$$

From (2.12), we see that the implied quasi-cointegrating relationship is $\beta = (0,1)^{\mathsf{T}} \in (\operatorname{sp} r_{\mathrm{LU}})^{\perp}$, i.e. that $\beta^{\mathsf{T}} x_t = n_t$, consistent with the implication of the model that technology shocks only have a long-lived effects on productivity, and not on hours.

The preceding example yields a VAR(1) with a reduced rank autoregressive matrix (with eigenvalues at zero and ρ_z). However, the same points may be made more generally for linear structural models that can be written in state-space form, which nests a wide range of structural macroeconomic models, as follows. For simplicity, we consider a model with a first-order state equation, but the conclusions carry over straightforwardly to higher-order processes.

Example 2.2. Consider the state-space model

$$x_t = Ax_{t-1} + Bw_t \tag{2.13a}$$

$$y_t = Cx_{t-1} + Dw_t \tag{2.13b}$$

in which each of y_t, x_t and w_t are p-dimensional (cf. Fernández-Villaverde, Rubio-Ramirez, Sargent, and Watson, 2007). The dynamics are governed by the state equation (2.13a): if one or more of the state variables x_t are integrated, then A will have (say) q unit eigenvalues and the long-run IRF for x_t, with respect to w_t, will have rank q. Let us partition $w_t = (w_{1t}^{\mathsf{T}}, w_{2t}^{\mathsf{T}})^{\mathsf{T}}$, where w_{1t} is the q-dimensional subvector of shocks that have permanent effects on x_t. In light of the preceding discussion, it is not necessary to maintain that the persistence in the state variables is generated by q exact unit roots, but only that the weaker requirement of QC should hold, with A having q eigenvalues in $\mathcal{L}^{\rho}_{\mathrm{LU}}$. Regardless of the specific values of these roots, by the analogue of decomposition (2.5)–(2.7) for the state equation, we have

$$\frac{\partial x_{t+s}}{\partial w_t} = R_{A,\mathrm{LU}}\Lambda_{A,\mathrm{LU}}^{k-1+s}L_{A,\mathrm{LU}}^{\mathsf{T}} + R_{A,\mathrm{ST}}\Lambda_{A,\mathrm{ST}}^{k-1+s}L_{A,\mathrm{ST}}^{\mathsf{T}},$$

and hence $L_{A,\mathrm{LU}}^{\mathsf{T}}Bw_t = Mw_{1t}$, for some $M \in \mathbb{R}^{q\times q}$ having full rank, since only the impact of these shocks decay at the slower rate regulated by the eigenvalues of $\Lambda_{A,\mathrm{LU}}$.

Under this weaker assumption, the implied VAR(MA) representation of the model yields the same long-run identifying restrictions as when exact unit roots are present. Provided C and D are invertible, (2.13) implies that

$$y_t = \Phi y_{t-1} + \varepsilon_t - \Psi\varepsilon_{t-1} \tag{2.14}$$

where $\Phi = CAC^{-1}$, $\Psi = \Phi - CBD^{-1}$, and $\varepsilon_t \coloneqq Dw_t$ are the reduced-form shocks. Because Φ and A are similar, the characteristic roots in the state equation (2.13a) coincide exactly with those in (2.14); in particular, both systems are characterised by q roots in $\mathcal{L}^{\rho}_{\mathrm{LU}}$, and $L_{\mathrm{LU}} = (C^{-1})^{\mathsf{T}}L_{A,\mathrm{LU}}$. Because of the MA component, the reduced-form IRF takes the modified form $\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \Phi^{h-1}(\Phi - \Psi)$, and it is no longer the case that $L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$ recovers the shocks w_{1t} driving the common persistent components; instead, we have

$$L_{\mathrm{LU}}^{\mathsf{T}}(\Phi - \Psi)\varepsilon_t = L_{A,\mathrm{LU}}^{\mathsf{T}}C^{-1}[CBD^{-1}]Dw_t = L_{A,\mathrm{LU}}^{\mathsf{T}}Bw_t = Mw_{1t}, \tag{2.15}$$

where $L_{\mathrm{LU}}^{\mathsf{T}}(\Phi - \Psi)$ depends only on the reduced-form parameters.

In practice, (2.14) is rarely estimated; the usual approach is to approximate by a finite-order VAR, truncating the l.h.s. of the representation

$$\sum_{i=0}^{\infty}\Psi^i(y_{t-i} - \Phi y_{t-1-i}) = \varepsilon_t \tag{2.16}$$

at some finite k.[4] Suppose that the eigenvalues of Ψ lie in $\mathcal{L}^{\rho}_{\mathrm{ST}}$, so that these may be distinguished

[4]If Ψ, or equivalently $A - BD^{-1}C$, is a nilpotent matrix, then there exists an exact finite order VAR representation for the system, of some order k^*, and the following claims – in particular (2.17) – hold for any $k \geq k^*$, rather than merely in the limit (cf. Ravenna, 2007, Cor. 2.2).

from the q dominant eigenvalues of Φ. Letting $\Psi_{k-1}(\lambda) := I_p\lambda^{k-1} + \Psi\lambda^{k-2} + \cdots + \Psi^{k-1}$, the truncated VAR($k$) has characteristic polynomial $\Gamma(\lambda) := \Psi_k(\lambda)(I_p\lambda - \Phi)$, whose roots are the eigenvalues of Φ (and therefore of A), and otherwise complex rotations of the eigenvalues of Ψ. Then the truncated VAR satisfies QC, and if we apply the decomposition (2.7) to the truncated VAR, then

$$L_{k,\mathrm{LU}}^{\mathsf{T}}\varepsilon_t \to L_{\mathrm{LU}}^{\mathsf{T}}(\Phi - \Psi)\varepsilon_t = Mw_{1t} \tag{2.17}$$

as $k \to \infty$, where the equality holds by (2.15), and thus the correct shocks are recovered in the limit, as the order of the VAR approximation grows. Remarkably, $R_{k,\mathrm{LU}} = R_{\mathrm{LU}}$ for all k, so the quasi-cointegrating relations are carried *exactly* by the truncated VAR. □

2.4.2 Long-run predictions of macroeconomic theories

A second respect in which quasi-contegration is empirically useful is in testing what might be termed the long-run predictions of economic theories, which in an (S)VAR with exact unit roots would be embodied in the cointegrating relations between the series. In a range of structural models, the dependence of the endogenous variables on a common set of state variables, combined with an elevated degree of persistence in the mechanisms generating some of those variables, manifests as a collection of long-run equilibrium relationships between those variables. If the theory underlying the structural model makes definite predictions about the coefficients parametrising these relationships, this provides a means of testing the theory, and cointegration analysis has been widely applied to this end (such as to testing business cycle models, theories of purchasing power parity, and the expectations theory of the term structure). Quasi-cointegration provides a means of continuing to conduct tests of this kind without having to maintain the auxiliary assumption of exact unit roots, by providing a means of expressing these long-run equilibrium relationships in a way that is robust to departures from this assumption. The following example illustrates this in detail, and we shall return to it in in our empirical application in Section 4.2.

Example 2.3. In its simplest form, the expectations theory of the term structure holds that the yield on a (zero-coupon) bond should be equal to the sum of the expected future yields on a shorter-dated bond (see e.g. Ljungqvist and Sargent, 2004, Ch. 13, for a textbook derivation from a dynamic asset pricing model under risk neutrality). If the reduced-form process followed by the yields on these two bonds follows a bivariate VAR with a single unit root, then it has long been known that a major implication of this theory is that the the (annualised) rate of return on these bonds should be cointegrated, with $\beta = (1, -1)^{\mathsf{T}}$ (see e.g. Campbell and Shiller, 1987; Carriero, Favero, and Kaminska, 2006). However, as we shall now show, that implied one-for-one long-run equilibrium relationship is contingent on the assumption of an exact unit root – a contingency also noted previously, albeit in a simplified setting in which one of the yields follows a reduced-form AR(1) process, by Müller and Watson (2013).

Let $\varrho_{i,t}$ denote the (annualised) yield on a zero-coupon bond with i years to maturity (in year t), and suppose that we observe data on both a 1-year and an m-year bond, generated by a VAR(k) that satisfies QC for some $\rho \leq 1$, with $\delta = 0$ and reduced form errors $\{\varepsilon_t\}$. The loglinearised form the expectations theory implies that

$$\varrho_{m,t} = \frac{1}{m}\sum_{i=0}^{m-1}\mathbb{E}_t\varrho_{1,t+i} + \xi_t, \tag{2.18}$$

where $\{\xi_t\}$ captures the term premium. To keep (2.18) consistent with a VAR for $\varrho_t \coloneqq (\varrho_{m,t}, \varrho_{1,t})$, $\{\xi_t\}$ must be a linear process of the form $\xi_t = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i}$. In a setting with exact unit roots ($\rho = 1$), one would assume that $\xi_t \sim I(0)$. To allow the theory to retain predictive content in our more general setting ($\rho \leq 1$), we make the analogous assumption that ξ_t is strictly less persistent than ϱ_t itself, in the sense that $\rho^{-h}\psi_h \to 0$ as $h \to \infty$.

Recognising that $\varrho_{1,t+i} = \sum_{k=0}^{i-1}\lambda^k\Delta_\lambda\varrho_{1,t+i-k} + \lambda^i\varrho_{1,t}$, we may rewrite the preceding for any $\lambda \in [0,1]$ as

$$\varrho_{m,t} - a_m(\lambda)\varrho_{1,t} \coloneqq \varrho_{m,t} - \frac{1}{m}\frac{1-\lambda^m}{1-\lambda}\varrho_{1,t} = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{i-1}\lambda^j\mathbb{E}_t\Delta_\lambda\varrho_{1,t+i-j} + \xi_t. \tag{2.19}$$

In particular, if we take $\lambda = \lambda_{\mathrm{LU}}$ (where λ_{LU} corresponds to the root nearest to unity in the bivariate VAR representation of the yields), then by Lemma A.5 in Appendix A,

$$\frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{i-1}\lambda_{\mathrm{LU}}^j\mathbb{E}_t\Delta_{\lambda_{\mathrm{LU}}}\varrho_{1,t+i-j} = \gamma_0\beta^{\mathsf{T}}\varrho_t + \sum_{i=0}^{p-1}\gamma_i\Delta_{\lambda_{\mathrm{LU}}}\varrho_{t-i} \eqqcolon \varsigma_t$$

for some $\{\gamma_i\}_{i=0}^{p-1}$ that depends on m, λ_{LU} and the VAR coefficients. By that same result, $(\beta^{\mathsf{T}}\varrho_t, \Delta_{\lambda_{\mathrm{LU}}}\varrho_t)$ also follows a VAR, whose roots lie in $\mathcal{L}_{\mathrm{ST}}^{\rho}$. Thus $\rho^{-h}\frac{\partial\varsigma_{t+h}}{\partial\varepsilon_t} \to 0$ as $h \to \infty$, and hence the same is true of $\varrho_{m,t} - a_m(\lambda_{\mathrm{LU}})\varrho_{1,t}$. In this manner, one of the principal implications of the expectations theory may be generalised. Rather than merely implying that yields $\varrho_t \coloneqq (\varrho_{m,t}, \varrho_{1,t})$ should be cointegrated (with unit coefficient), in a VAR with an exact unit root, the theory more generally entails that these are quasi-cointegrated, with

$$\beta(\lambda_{\mathrm{LU}})^{\mathsf{T}} = [1 \quad -a_m(\lambda_{\mathrm{LU}})]. \qquad \square$$

2.5 Connections to the literature

As noted in the introduction, there have been relatively few attempts to address the problem identified by Elliott (1998), most notably Wright (2000), Magdalinos and Phillips (2009), Müller and Watson (2013), Franchi and Johansen (2017), and Hwang and Valdés (2023). Having now outlined our own approach to this problem, we may briefly explain how our work is situated relative to those contributions.

Insofar as they also consider a VAR model with some characteristic roots near unity, Franchi and Johansen (2017) relates closely to the present study. Their setting is a VAR model with one lag, written in error correction form as

$$\Delta x_t = (\alpha\beta^{\mathsf{T}} + \alpha_1\Gamma\beta_1^{\mathsf{T}})x_{t-1} + \varepsilon_t \eqqcolon \Pi x_{t-1} + \varepsilon_t, \tag{2.20}$$

where $\alpha, \beta \in \mathbb{R}^{p\times r}$ and $\alpha_1, \beta_1 \in \mathbb{R}^{p\times q}$ have full column rank, and $\Gamma \in \mathbb{R}^{q\times q}$. When $\Gamma = 0$, the model specialises exactly to the CVAR model of Section 2.2 with q unit roots and $\mathrm{CS} = \mathrm{sp}\,\beta$. If some elements of Γ are non-zero, the cointegrated model becomes one with some roots near but not equal to unity, and Π need no longer be of reduced rank. As the authors acknowledge, if each of α_1, β_1 and Γ are freely varying, then β is not identified. They accordingly treat α_1 and β_1 as known, which restores identification and facilitates likelihood-based inference on each of α, β and Γ. But while a priori knowledge of α_1 and β_1 may indeed be available in certain situations, this seems unlikely to be the case in general; whereas in our approach, the criterion (2.3) ensures that β remains identified even as the dominant roots depart from unity. Moreover, with Γ fixed (and nonzero) in this model, it is unclear how β in (2.20) could be interpreted in terms of long-run relationships between the elements of x_t.

Magdalinos and Phillips (2009) consider Elliott's (1998) problem in the context of the triangular model

$$x_{1t} = Ax_{2t} + u_{1t} \tag{2.21a}$$

$$x_{2t} = R_n x_{2,t-1} + u_{2t} \tag{2.21b}$$

where $u_t = (u_{1t}^{\mathsf{T}}, u_{2t}^{\mathsf{T}})^{\mathsf{T}}$ is a weakly dependent linear process. When $R_n = I_q$, this encompasses the $I(0)/I(1)$ CVAR model with q unit roots, but allows for a more general semiparametric treatment of the model's short-run dynamics. If R_n instead merely drifts towards I_q (possibly at a slower rate than n^{-1}) as $n \to \infty$, then the authors show that it is still possible to obtain an asymptotically mixed normal estimate of A, by using instruments that are constructed by filtering x_{2t} (what they term the 'IVX' estimator of A). However, the greater generality afforded by the triangular model comes at the price that $R_n \to I_q$ is now necessary for identification of A; if on the other hand R_n were fixed with eigenvalues strictly less than unity, then all linear combinations of x_t would be weakly dependent, leaving A unidentified. This is true generally of approaches that rely on the triangular form, because of its agnosticism about the dynamics of u_t; thus the same point may be made in the context of Hwang and Valdés (2023), who (when $R_n = I_q + n^{-1}C$) consider an augmented regression estimator of (2.21a) using low frequency transforms of the original data, and a Bonferroni-based approach to correct for the effect of C on its limiting distribution.

Finally, Müller and Watson (2013) consider a very general setting in which the 'common trends' in x_t are permitted to belong to a broad family of processes. A consequence of this generality is that the authors conceptualise 'cointegration' in terms different from quasi-cointegration, and the two definitions do not always agree. Essentially, Müller and Watson define x_t to be 'cointegrated' with cointegrating relations $\beta \in \mathbb{R}^{p\times r}$, if $n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} \beta^{\mathsf{T}} x_t$ converges weakly to a Brownian motion, while the common trends $n^{-1/2}\beta_\perp^{\mathsf{T}} x_{\lfloor nr\rfloor}$ converge weakly to a some cadlag process (where $\beta_\perp \in \mathbb{R}^{p\times q}$ has $\mathrm{rk}\,\beta_\perp = q$ and $\beta_\perp^{\mathsf{T}}\beta = 0$). In the context of our VAR model, where QC holds for some $\rho < 1$, $n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} x_t$ converges weakly to a Brownian motion if all the roots are strictly inside the unit circle; so in such a case there is no 'cointegration' in the foregoing sense, even though quasi-cointegrating relationships would be defined. (On the other hand, if the largest q roots of Φ are localised to unity at rate n^{-1}, though not more slowly, then their 'cointegrating' vectors would coincide with our quasi-cointegrating vectors.) Regarding inference, the authors construct a confidence set for β by inverting a stationarity test for $\beta^{\mathsf{T}} x_t$, extending a idea originally due to Wright (2000). For a comparison of their tests with ours, in terms of size and power, see the simulations in Section 4.1 below: these indicate that the price paid, in terms of power, for robustness to a broader class of trend generating mechanisms (than are permitted by the VAR), may be considerable.

Thus in relation to these papers, one of the major distinguishing contributions of our work is to provide a means of identifying long-run equilibrium relationships that is well-defined for a *fixed* parametrisation of the underlying (S)VAR; i.e. we do not rely on that VAR drifting towards a model with exact unit roots (as $n \to \infty$). Not only is our identifying criterion (2.3) readily interpretable in terms of the relative persistence of either structural or reduced-form impulse responses, but it also maintains the duality that exists, both with and without exact unit roots, between the identification of long-run equilibrium relationships, and of those structural shocks whose common persistent effects give rise to those relationships. As we will discuss subsequently, an added benefit is that it reduces Elliott's (1998) problem to one that is asymptotically equivalent to inference in a multivariate predictive regression, a canonical problem that has given rise to a rich literature.

3 Estimation and inference

3.1 Formulation of the problem

3.1.1 Model likelihood

As with the CS in a cointegrated VAR model, inference on the QCS in our more general setting will be based on the normal model likelihood (or quasi-likelihood, if ε_t is not in fact normally distributed). Recall that the model (2.1) may be rendered as

$$y_t = m + dt + \sum_{i=1}^{k} \Phi_i y_{t-i} + \varepsilon_t. \tag{3.1}$$

To facilitate the exposition, we focus on the case where the intercept and trend parameters (m, d) are unrestricted in (3.1), while maintaining that the data is generated under (2.1), so as to exclude the possibility of a quadratic trend in y_t. (For a discussion of alternative potential treatments of the deterministic terms, see Section 3.5 below.)

The loglikelihood with (m, d) concentrated out – or equivalently, expressed in terms of a maximal invariant for transformations of the form $\{y_t\} \mapsto \{m + dt + y_t\}$ – may be written as

$$\ell_n(\mathbf{\Phi}, \Sigma) \coloneqq -\frac{n}{2} \log(2\pi \det \Sigma) - \min_{m,d} \frac{1}{2} \sum_{t=1}^{n} \left\| y_t - m - dt - \sum_{i=1}^{k} \Phi_i y_{t-i} \right\|_{\Sigma^{-1}}^2,$$

where $\|x\|_W^2 \coloneqq x^\mathsf{T} W x$ for $x \in \mathbb{R}^p$ and $W \in \mathbb{R}^{p \times p}$ positive semidefinite. The QCS depends only on $\mathbf{\Phi}$, and the main (asymptotic) results of this paper are not sensitive to the choice of estimator for Σ, provided that it is consistent. To simplify our arguments, we shall therefore generally assume that the unrestricted ML estimator $\hat{\Sigma}_n$, i.e. the OLS variance estimator, is used. Henceforth, let $\ell_n^*(\mathbf{\Phi}) \coloneqq \ell_n(\mathbf{\Phi}, \hat{\Sigma}_n)$; for convenience we shall refer to maximisers of ℓ_n^* as 'maximum likelihood estimators'.

3.1.2 QCS as a functional of the VAR coefficients

Under QC, the QCS is well-defined and has dimension q. Since any basis $\beta \in \mathbb{R}^{p\times q}$ for the QCS is only identified up to its column space, and has rank q, it is convenient to maintain the normalisation

$$\beta^{\mathsf{T}} = [I_r \quad -A], \tag{3.2}$$

so that inference on the QCS reduces to inference on the elements of the matrix $A \in \mathbb{R}^{r\times q}$. (3.2) is not restrictive – i.e. it is indeed merely a normalisation of β – if the QCS does not contain any nonzero vectors whose first r elements are all zero, as will be the case if the elements of y_t are ordered appropriately. Since R_{LU} has rank q and $\beta^{\mathsf{T}} R_{\mathrm{LU}} = 0$, (3.2) is equivalent to

$$R_{\mathrm{LU}} = \begin{bmatrix} A \\ I_q \end{bmatrix}. \tag{3.3}$$

Since the q roots in $\mathcal{L}^{\rho}_{\mathrm{LU}}$ are separated from the $kp - q$ roots in $\mathcal{L}^{\rho}_{\mathrm{ST}}$, the column space of R_{LU} depends smoothly on the VAR coefficients. To express this formally, let $\lambda_i(\mathbf{\Phi})$ denote the ith root of the characteristic polynomial associated to the VAR with coefficients $\mathbf{\Phi}$, when these are placed in descending order of modulus; and $G^{\mathsf{T}} := [0_{q\times r}, I_q]$. Define $\mathscr{P} \subset \mathbb{R}^{p\times kp}$ to be the set of VAR coefficients such that: (i) $|\lambda_{q+1}(\mathbf{\Phi})| < |\lambda_q(\mathbf{\Phi})|$; (ii) there exist $R_{\mathrm{LU}} \in \mathbb{R}^{p\times q}$ and $\Lambda_{\mathrm{LU}} \in \mathbb{R}^{q\times q}$ such that the eigenvalues of Λ_{LU} are $\{\lambda_1(\mathbf{\Phi}), \ldots, \lambda_q(\mathbf{\Phi})\}$,

$$R_{\mathrm{LU}}\Lambda^{k}_{\mathrm{LU}} - \sum_{i=1}^{k} \Phi_i R_{\mathrm{LU}}\Lambda^{k-i}_{\mathrm{LU}} = 0; \tag{3.4}$$

and (iii) $\mathrm{rk}\{G^{\mathsf{T}} R_{\mathrm{LU}}\} = q$. Then $\mathscr{P}$ is open, and since $G^{\mathsf{T}} R_{\mathrm{LU}}$ has full rank, we may choose $(R_{\mathrm{LU}}, \Lambda_{\mathrm{LU}})$ to be consistent with the normalisation (3.3). The conditions defining $\mathscr{P}$, together with (3.3), implicitly define smooth (i.e. infinitely differentiable) maps $R_{\mathrm{LU}}(\mathbf{\Phi})$, $A(\mathbf{\Phi})$, and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$ on $\mathscr{P}$ (Lemma B.1). In light of this, inference on the QCS may be rephrased in terms of inference on parameters $A = A(\mathbf{\Phi})$ defined by a smooth transformation of the VAR coefficients.

3.1.3 Parameter space for the near-unit roots

In a model with exact unit roots, efficient estimation of A requires the model to be estimated under some of the restrictions implied by CV: for example, ML estimation of A proceeds under the assumption that $\mathrm{rk}\,\Phi(1) = r$, as per CV2. In a setting with only $I(1)$ series, this is equivalent to maintaining $q = p - r$ roots at real unity. Transposed to the present setting, with QC taking the place of CV, we now require the model be estimated with q roots lying in $\mathcal{L}^{\rho}_{\mathrm{LU}}$, as per QC1. This entails both a choice of ρ, and the specification of an appropriate parameter space $\mathscr{L} \subset \mathbb{R}^{q\times q}$ for Λ_{LU}, such that it is diagonalisable (as per QC2), with eigenvalues lying in $\mathcal{L}^{\rho}_{\mathrm{LU}}$.

When $q = 1$, Λ_{LU} is scalar, so $\mathscr{L} = [\rho, 1]$ and only a lower bound ρ on the largest root of Φ needs to be specified. A discussion of the considerations that might inform ρ were given at the end of Section 2.3 (see also the application in Section 4.2 below). When $q \geq 2$, $\mathscr{L}$ is instead a set of matrices with eigenvalues lying in the interval $[\rho, 1]$. While QC2 might suggest taking $\mathscr{L}$ to be the subset $\mathscr{L}_{\mathrm{d}}$ of real, diagonalisable $q \times q$ matrices, a potential difficulty with $\mathscr{L}_{\mathrm{d}}$ is that is that some non-diagonalisable matrices are in its closure, as can be seen e.g. by taking the limit of $\left[\begin{smallmatrix} \lambda+\epsilon & 1 \\ 0 & \lambda-\epsilon \end{smallmatrix}\right]$ as $\epsilon \to 0$. This would effectively permit departures from the $I(0)/I(1)$ cointegrated VAR model in the direction of a model with some $I(2)$ components, something that we wish to avoid here. The set of either normal ($\mathscr{L}_{\mathrm{n}}$) or symmetric ($\mathscr{L}_{\mathrm{s}}$) matrices with eigenvalues in $[\rho, 1]$ would thus be more appropriate choices for $\mathscr{L}$, since each give a closed subset of the set of diagonalisable matrices, with the principal difference between the two being that the former allows for complex eigenvalues.

3.1.4 Local-to-unity asymptotics

The QCS, and the associated coefficient matrix A, remain identified so long as the roots of Φ separate as prescribed by QC. In particular, there is no requirement that the roots in $\mathcal{L}_{\mathrm{LU}}^{\rho}$ should drift towards unity at any rate, as $n \to \infty$. However, as will become evident below, the proximity of those q largest roots to unity affects the distributions of estimators and test statistics, even in very large samples. We therefore need to work with a sequence of models that preserves this dependence in the limit, and avoids the discontinuities in the asymptotics that would otherwise arise at exact unit roots. We shall accordingly study the large-sample behaviour of the likelihood, and of derived estimators and test statistics, under

Assumption LOC. *$\{(y_t, x_t)\}_{t=1}^{n}$ is generated per DGP with $\mathbf{\Phi} = \mathbf{\Phi}_n$, for $\{\mathbf{\Phi}_n\} \subset \mathscr{P}$ such that:*

LOC1 *for some $C \in \mathbb{R}^{q \times q}$ with non-positive eigenvalues,*

$$\Lambda_{\mathrm{LU}}(\mathbf{\Phi}_n) = \Lambda_{n,\mathrm{LU}} \coloneqq I_q + n^{-1}C; \tag{3.5}$$

LOC2 *$R_{\mathrm{LU}}(\mathbf{\Phi}_n) = \left[\begin{smallmatrix} A \\ I_q \end{smallmatrix}\right]$ for some $A \in \mathbb{R}^{r \times q}$; and*

letting $R_{n,\mathrm{ST}}$, $\Lambda_{n,\mathrm{ST}}$ and $L_n = [L_{n,\mathrm{LU}}, L_{n,\mathrm{ST}}]$ be such that (2.5)–(2.7) hold for each n:

LOC3 *$R_{n,\mathrm{ST}} = R_{\mathrm{ST}}$ and $\Lambda_{n,\mathrm{ST}} = \Lambda_{\mathrm{ST}}$ are fixed, and the eigenvalues of the latter lie strictly inside the complex unit circle; and*

LOC4 *$L_{\mathrm{LU}} \coloneqq \lim_{n\to\infty} L_{n,\mathrm{LU}}$ has full column rank.*

Under LOC, we may choose a $\rho < 1$ such that QC holds for all n sufficiently large; LOC may thus be regarded as capturing (sequences of) VAR models that are both in the immediate vicinity of a cointegrated VAR with unit roots, while also satisfying those regularity conditions that ensure the QCS is well-defined. The localisation (3.5) moreover entails a sharper delineation between the common trend and transitory components appearing in the implied decomposition of y_t in (2.9), since we now have the joint weak convergences

$$n^{-1/2}\sum_{t=1}^{n} \varepsilon_t \rightsquigarrow E(r) \qquad\qquad n^{-1/2} z_{\mathrm{LU},\lfloor nr \rfloor} \rightsquigarrow \int_0^r \mathrm{e}^{C(r-s)} L_{\mathrm{LU}}^{\mathsf{T}} \mathrm{d}E(s) =: Z_C(r), \tag{3.6}$$

on $D[0,1]$, for E a p-dimensional Brownian motion with variance Σ; and thus

$$n^{-1/2}x_{\lfloor nr\rfloor} = \Phi_{n,\mathrm{LU}} n^{-1/2} z_{\mathrm{LU},\lfloor nr\rfloor} + o_p(1) =_d \Phi_{n,\mathrm{LU}} Z_C(r) + o_p(1) \tag{3.7}$$

so that $z_{\mathrm{LU},t}$ and x_t are nearly integrated. Although $\Phi_{n,\mathrm{LU}} = R_{\mathrm{LU}}\Lambda^k_{n,\mathrm{LU}}$ depends on n, its column space does not, and

$$\beta^{\mathsf{T}} x_t = \beta^{\mathsf{T}} \Phi_{\mathrm{ST}} z_{\mathrm{ST},t-1} + \beta^{\mathsf{T}}\varepsilon_t \sim I(0). \tag{3.8}$$

Thus, analogously to the GJRT, (2.9) decomposes x_t, and therefore also y_t (upon detrending), into the sum of a nearly integrated component and an $I(0)$ component; the quasi-cointegrating relations are precisely those that eliminate the nearly integrated common trends from y_t.

3.2 Asymptotics of the loglikelihood ratio process

We first consider the asymptotic behaviour of the loglikelihood ratio process, under a local reparametrisation of the VAR given by

$$\boldsymbol{\pi} := \boldsymbol{\pi}_n(\boldsymbol{\Phi}) := n \operatorname{vec} \begin{bmatrix} A(\boldsymbol{\Phi}) - A(\boldsymbol{\Phi}_n) \\ \Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}) - \Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}_n) \end{bmatrix}, \quad f := f_n(\boldsymbol{\Phi}) := n^{1/2} \operatorname{vec}\{(\boldsymbol{\Phi} - \boldsymbol{\Phi}_n)\mathbf{R}_{n,\mathrm{ST}}\}; \tag{3.9}$$

where $\{\boldsymbol{\Phi}_n\}$ is as in LOC, and $\mathbf{R}_{n,\mathrm{ST}}$ is a $kp \times (kp - q)$ matrix defined in Appendix C. (3.9) effectively isolates the signal from the nearly integrated and $I(0)$ components of x_t, as given in (3.7)–(3.8), with only the former carrying (asymptotically) information relevant to the estimation of A and Λ_{LU}. These parameters will thus enjoy elevated rate of convergence, relative to the other components of the VAR, just as is familiar from the VAR with exact unit roots. In connection with (3.9), let

$$K := \begin{bmatrix} \beta^{\mathsf{T}} R_{\mathrm{ST}}(I - \Lambda_{\mathrm{ST}})^{-1} L^{\mathsf{T}}_{\mathrm{ST}} \\ L^{\mathsf{T}}_{\mathrm{LU}} \end{bmatrix} =: \begin{bmatrix} \mathcal{J} \\ L^{\mathsf{T}}_{\mathrm{LU}} \end{bmatrix}. \tag{3.10}$$

denote a component of limiting Jacobian matrix for $\boldsymbol{\pi}_n(\boldsymbol{\Phi})$ at $\boldsymbol{\Phi} = \boldsymbol{\Phi}_n$.

The asymptotics reveal the close correspondence that exists between the problem of inference on $(A, \Lambda_{\mathrm{LU}})$ in a quasi-cointegrated VAR, and of inference in a predictive regression model with highly persistent regressors. This is of interest because the latter is a canonical problem that has received considerable attention in the literature, where a variety of inferential procedures with good size and power properties have been developed. A major implication of our next result is that such procedures should enjoy similar properties, once transposed to our setting. While we do not develop this possibility further here, the good properties found for the Elliott et al. (2015) procedure in Section 4.1 are of little surprise, in view of the similar performance enjoyed by that method in a predictive regression.

By a *predictive regression model*, we mean a model of the form

$$y_{\mathrm{PR},t} = m_y + d_y t + A_{\mathrm{PR}} z_{\mathrm{PR},t-1} + \xi_{y,t} \tag{3.11a}$$

$$z_{\mathrm{PR},t} = m_z + d_z t + \Lambda_{\mathrm{PR}} z_{\mathrm{PR},t-1} + \xi_{z,t} \tag{3.11b}$$

where $\{y_{\mathrm{PR},t}\}$ and $\{z_{\mathrm{PR},t}\}$ respectively take values in $\mathbb{R}^r$ and $\mathbb{R}^q$. To complete the specification of the model, suppose the following:

Assumption PR. *$\{y_{\mathrm{PR},t}\}_{t=1}^{n}$ and $\{z_{\mathrm{PR},t}\}_{t=1}^{n}$ are generated under (3.11), where*

PR1 $A_{\mathrm{PR}} \in \mathbb{R}^{r\times q}$ *and* $\Lambda_{\mathrm{PR}} = I_q + n^{-1}C$, *for* C *as in* LOC;

PR2 $z_{\mathrm{PR},0} = 0$, $m_y = d_y = 0$ *and* $m_z = d_z = 0$;

PR3 $\xi_t := (\xi_{y,t}^{\mathsf{T}}, \xi_{z,t}^{\mathsf{T}})^{\mathsf{T}} \sim_{\text{i.i.d.}} N[0,\Omega]$.

To simplify the discussion, the covariance matrix Ω is assumed to be known. Under PR3, the model loglikelihood is

$$\ell_n^{\mathrm{PR}}(A,\Lambda) := -\frac{n}{2}\log(2\pi\det\Omega) - \min_{m,d}\frac{1}{2}\sum_{t=1}^{n}\left\|\begin{bmatrix} y_{\mathrm{PR},t} \\ z_{\mathrm{PR},t}\end{bmatrix} - \begin{bmatrix} m_y \\ m_z\end{bmatrix} - \begin{bmatrix} d_y \\ d_t\end{bmatrix} t - \begin{bmatrix} A \\ \Lambda\end{bmatrix} z_{\mathrm{PR},t-1}\right\|_{\Omega^{-1}}^{2}.$$

Let $\bar{Z}_C(r)$ denote the residual of an $L^2[0,1]$ projection of each sample path of Z_C in (3.6) onto a constant and linear trend. The proof of the next result, and of all other theorems, appears in Appendix E.

Theorem 3.1. *Suppose that:*

(i) *$\{y_t\}$ is generated under* LOC. *Let $\ell_n(\boldsymbol{\pi}, f) := \ell_n(\boldsymbol{\Phi},\Sigma)$, where $\boldsymbol{\pi} = \boldsymbol{\pi}_n(\boldsymbol{\Phi})$ and $f = f_n(\boldsymbol{\Phi})$ as in (3.9), and let $\ell_n^*(\boldsymbol{\pi}) := \max_{\boldsymbol{\pi}(\boldsymbol{\Phi})=\boldsymbol{\pi}} \ell_n^*(\boldsymbol{\Phi})$.*

(ii) *$\{(y_{\mathrm{PR},t}, z_{\mathrm{PR},t})\}$ is generated under* PR, *with $\Omega = K\Sigma K^{\mathsf{T}}$. Let $\ell_n^{\mathrm{PR}}(\boldsymbol{\pi}) := \ell_n^{\mathrm{PR}}(A,\Lambda)$, where*

$$\boldsymbol{\pi} = n\operatorname{vec}\begin{bmatrix} A - A_{\mathrm{PR}} \\ \Lambda - (I_q + n^{-1}C)\end{bmatrix}.$$

Then the finite-dimensional distributions of each of $\{\ell_n(\boldsymbol{\pi}, f) - \ell_n(0,f)\}$, $\{\ell_n^*(\boldsymbol{\pi}) - \ell_n^*(0)\}$ and $\{\ell_n^{\mathrm{PR}}(\boldsymbol{\pi}) - \ell_n^{\mathrm{PR}}(0)\}$ converge to those of

$$S_{\boldsymbol{\pi}}^{\mathsf{T}}\boldsymbol{\pi} - \tfrac{1}{2}\boldsymbol{\pi}^{\mathsf{T}}H_{\boldsymbol{\pi}}\boldsymbol{\pi} \tag{3.12}$$

where, for W a p-dimensional standard Brownian motion on $[0,1]$,

$$S_{\boldsymbol{\pi}} := \int_0^1 [\bar{Z}_C(r)\otimes(K\Sigma K^{\mathsf{T}})^{-1/2}\mathrm{d}W(r)] \qquad H_{\boldsymbol{\pi}} := \int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\otimes(K\Sigma K^{\mathsf{T}})^{-1}. \tag{3.13}$$

Up to a term that depends only on f, the likelihood ratio processes $\{\ell_n(\boldsymbol{\pi}, f) - \ell_n(0,0)\}$ and $\{\ell_n^{\mathrm{PR}}(\boldsymbol{\pi}) - \ell_n^{\mathrm{PR}}(0)\}$ thus share the same distributional limit; since this limit obtains under all local-to-unity sequences permitted by LOC and PR, it is evident that these models converge to the same limiting experiment, in the sense of van der Vaart (1998, Def. 9.1). This substantiates the asymptotic equivalence between the problem of inference on $(A, \Lambda_{\mathrm{LU}})$ in the VAR (2.1)–(2.2), and of inference on $(A_{\mathrm{PR}}, \Lambda_{\mathrm{PR}})$ in a predictive regression, in the vicinity of unit roots. That the weak limit in (3.12) is also shared by $\{\ell_n^*(\boldsymbol{\pi}) - \ell_n(0)\}$ is of practical importance for the likelihood-based tests discussed below.

3.3 Likelihood-based inference

3.3.1 ML estimators

We next turn to the ML estimators of A and Λ_{LU}. Let the unrestricted and restricted estimators of $\mathbf{\Phi}$, the latter with $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0 \in \mathbb{R}^{q\times q}$ imposed, be denoted as

$$\hat{\mathbf{\Phi}}_n \coloneqq \operatorname*{argmax}_{\mathbf{\Phi}\in\mathbb{R}^{p\times kp}} \ell_n^*(\mathbf{\Phi}) \qquad \hat{\mathbf{\Phi}}_{n|\Lambda_0} \coloneqq \operatorname*{argmax}_{\{\mathbf{\Phi}\in\mathscr{P}|\Lambda_{\mathrm{LU}}(\mathbf{\Phi})=\Lambda_0\}} \ell_n^*(\mathbf{\Phi}).$$

(For details of how to compute $\hat{\mathbf{\Phi}}_{n|\Lambda_0}$ in practice, see Appendix F.) Then $\hat{A}_n \coloneqq A(\hat{\mathbf{\Phi}}_n)$ and $\hat{\Lambda}_{n,\mathrm{LU}} \coloneqq \Lambda_{\mathrm{LU}}(\hat{\mathbf{\Phi}}_n)$ are the associated unrestricted MLEs of A and Λ_{LU}, and $\hat{A}_{n|\Lambda_0} \coloneqq A(\hat{\mathbf{\Phi}}_{n|\Lambda_0})$ the MLE of A under $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0$.

To state our main result on these estimators, let $L_{\mathrm{LU},\perp}$ be any $p\times r$ matrix spanning $(\operatorname{sp} L_{\mathrm{LU}})^{\perp}$; one possible choice is $\alpha \coloneqq \lim_{n\to\infty} \Phi_n(1)\beta(\beta^{\mathsf{T}}\beta)^{-1}$. Recall that a random vector η is *mixed normal* with mean zero and conditional variance V, denoted $\eta \sim \mathrm{MN}[0, V]$, if $\mathbb{E}\mathrm{e}^{\mathrm{i}\tau^{\mathsf{T}}\eta} = \mathbb{E}\mathrm{e}^{-\frac{1}{2}\tau^{\mathsf{T}}V\tau}$.

Theorem 3.2. *Suppose* LOC *holds. Then*

(i) $\hat{A}_n \coloneqq A(\hat{\mathbf{\Phi}}_n)$ *and* $\hat{\Lambda}_{n,\mathrm{LU}} \coloneqq \Lambda_{\mathrm{LU}}(\hat{\mathbf{\Phi}}_n)$ *satisfy*[5]

$$n\begin{bmatrix} \hat{A}_n - A \\ \hat{\Lambda}_{n,\mathrm{LU}} - \Lambda_{n,\mathrm{LU}} \end{bmatrix} \rightsquigarrow K \int (\mathrm{d}E)\bar{Z}_C^{\mathsf{T}} \left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\right)^{-1}$$

(ii) $n \operatorname{vec}(\hat{A}_{n|\Lambda_{n,\mathrm{LU}}} - A) \rightsquigarrow \mathrm{MN}[0, V_{zz} \otimes V_{\varepsilon\varepsilon}]$, *where*

$$V_{zz} \otimes V_{\varepsilon\varepsilon} \coloneqq \left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\right)^{-1} \otimes \mathcal{J} L_{\mathrm{LU},\perp}(L_{\mathrm{LU},\perp}^{\mathsf{T}}\Sigma^{-1}L_{\mathrm{LU},\perp})^{-1}L_{\mathrm{LU},\perp}^{\mathsf{T}}\mathcal{J}^{\mathsf{T}} \tag{3.14}$$

$$= \left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\right)^{-1} \otimes (\alpha^{\mathsf{T}}\Sigma^{-1}\alpha)^{-1}. \tag{3.15}$$

The limiting distribution of the unrestricted ML estimator of A thus depends on C, which cannot be consistently estimated. However, if the *correct* value of Λ_{LU} is imposed, then the restricted ML estimator $\hat{A}_{n|\Lambda_{n,\mathrm{LU}}}$ is asymptotically mixed normal. In the special case where $\Lambda_{\mathrm{LU}} = I_q$, this exactly replicates the mixed normality of the ML estimates of the cointegrating relations, when the correct cointegrating rank is imposed (see e.g. Johansen, 1995, Thm. 13.3). In this manner, the preceding theorem generalises that mixed normality beyond the setting of a VAR with exact unit roots. Though we shall not give the proof here, it may be shown that with the correct Λ_{LU} imposed, the model loglikelihood is locally asymptotically mixed normal (LAMN), so that $\hat{A}_{n|\Lambda_{n,\mathrm{LU}}}$ also inherits the large-sample efficiency properties familiar from the case of exact unit roots (Phillips, 1991).

[5] Recall $\bar{Z}_C(r) = Z_C(r) - \mu_0 - \mu_1 r$, for $\mu_0 \coloneqq \int_0^1 (4-6s)B(s)\,\mathrm{d}s$ and $\mu_1 = \int_0^1(-6+12s)B(s)\,\mathrm{d}s$ (see e.g. Elliott, 1998, p. 151). Since $\bar{Z}_C$ is not adapted, an expression such as $\int \bar{Z}_C(\mathrm{d}E)^{\mathsf{T}}$ should be understood as a convenient shorthand for $\int Z_C(\mathrm{d}E)^{\mathsf{T}} - \mu_0 \int(\mathrm{d}E)^{\mathsf{T}} - \mu_1 \int r(\mathrm{d}E)^{\mathsf{T}}$.

3.3.2 Likelihood ratio tests

Though part (i) of the preceding provides a basis for inference on A using Wald-type statistics, there are some difficulties with this approach in practice, because there is no guarantee that the characteristic roots of the unrestrictedly estimated VAR will separate in the manner prescribed by QC. Since these roots come in conjugate pairs, it may well be the case that when ordered by their complex modulus (or proximity to real unity), the qth and $(q+1)$th roots will be complex conjugates, preventing us from isolating the first q roots from the rest – a problem exacerbated by typically imprecise estimation of these roots (Onatski and Uhlig, 2012). A superior approach therefore utilises (quasi-) likelihood ratio (LR) tests to perform inference on both Λ_{LU} and A; specifically the statistics

$$\mathcal{LR}_n(\Lambda_0) \coloneqq 2\left[\max_{\{\mathbf{\Phi}\in\mathscr{P} \mid \Lambda_{\mathrm{LU}}(\mathbf{\Phi})\in\mathscr{L}\}} \ell_n^*(\mathbf{\Phi}) - \max_{\{\mathbf{\Phi}\in\mathscr{P} \mid \Lambda_{\mathrm{LU}}(\mathbf{\Phi})=\Lambda_0\}} \ell_n^*(\mathbf{\Phi})\right] \tag{3.16a}$$

$$\mathcal{LR}_n(a_0;\Lambda_0) \coloneqq 2\left[\max_{\{\mathbf{\Phi}\in\mathscr{P} \mid \Lambda_{\mathrm{LU}}(\mathbf{\Phi})=\Lambda_0\}} \ell_n^*(\mathbf{\Phi}) - \max_{\{\Phi\in\mathscr{P} \mid \Lambda_{\mathrm{LU}}(\mathbf{\Phi})=\Lambda_0, a_{ij}(\mathbf{\Phi})=a_0\}} \ell_n^*(\mathbf{\Phi})\right], \tag{3.16b}$$

where $\mathscr{L}$ is the parameter space for Λ_{LU} (see Section 3.1.3 above). $\mathcal{LR}_n(\Lambda_0)$ is the usual likelihood ratio test for $H_0 : \Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0$, while $\mathcal{LR}_n(a_0;\Lambda_0)$ corresponds to the likelihood ratio test of $H_0 : a_{ij}(\mathbf{\Phi}) = a_0$, when $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0$ is maintained under both the null and the alternative. Our next result provides the asymptotic distributions of these test statistics; for given $C \in \mathbb{R}^{q\times q}$, let

$$C_* \coloneqq (L_{\mathrm{LU}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}})^{-1/2} C (L_{\mathrm{LU}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}})^{1/2}.$$

Theorem 3.3. *Suppose* LOC *holds. Then*

$$\mathcal{LR}_n(\Lambda_{n,\mathrm{LU}}) \rightsquigarrow \operatorname{tr}\left\{\int (\mathrm{d}W_*)\bar{Z}_{C_*}^{\mathsf{T}} \left(\int \bar{Z}_{C_*}\bar{Z}_{C_*}^{\mathsf{T}}\right)^{-1} \int \bar{Z}_{C_*}(\mathrm{d}W_*)^{\mathsf{T}}\right\} \tag{3.17}$$

where $W_* \sim \mathrm{BM}(I_q)$, $\bar{Z}_{C_*}$ *is the residual from an* $L^2[0,1]$ *projection of the sample paths of* $Z_{C_*}(r) \coloneqq \int_0^r \mathrm{e}^{C_*(r-s)}\,\mathrm{d}W_*(s)$ *onto a constant and linear trend; and*

$$\mathcal{LR}_n[a_{ij}(\mathbf{\Phi}_n);\Lambda_{n,\mathrm{LU}}] \rightsquigarrow \chi_1^2. \tag{3.18}$$

3.3.3 Nearly optimal tests

Elliott et al. (2015) consider hypothesis tests that are affected by nuisance parameters, in settings where the limiting experiment is not a Gaussian shift experiment (with an unrestricted parameter space), such the usual asymptotic optimality enjoyed by ML-based inference does not hold. In view of Theorem 3.1, our problem falls within their framework: the correspondence can be most easily seen when, analogously to their equation (1), we seek to test hypotheses of the form

$$H_0 : A = A_0,\ \Lambda_{\mathrm{LU}} \in \mathscr{L} \qquad \text{against} \qquad H_1 : A \neq A_0,\ \Lambda_{\mathrm{LU}} \in \mathscr{L},$$

so that A is the parameter of interest, and Λ_{LU} the nuisance parameter.[6] Their tests have the following Neyman–Pearson form,

$$\mathcal{NP}_n(A_0) \coloneqq \mathbf{1}\left\{\int_{\mathbb{R}^{r\times q}\times\mathscr{L}} \mathrm{e}^{\ell_n^*(A,\Lambda)} F_1(\mathrm{d}A, \mathrm{d}\Lambda) > \mathrm{cv}_\alpha \int_{\mathscr{L}} \mathrm{e}^{\ell_n^*(A_0,\Lambda)} F_0(A_0, \mathrm{d}\Lambda)\right\} \tag{3.19}$$

where $\ell_n^*(A,\Lambda) \coloneqq \max_{\{\Phi\in\mathscr{P}|\Lambda_{\text{LU}}(\boldsymbol{\Phi})=\Lambda, A(\boldsymbol{\Phi})=A\}} \ell_n^*(\boldsymbol{\Phi})$, and F_0 and F_1 are distributions that respectively concentrate on those subsets of the parameter space for (A,Λ) consistent with the null and the alternative.

As discussed by Elliott et al. (2015), the level α test that maximises weighted average power (WAP; against the F_1-weighted alternative) results when F_0 and cv_α are such that

$$\int_{\mathscr{L}} \mathbb{P}_{(A_0,\Lambda)}\{\mathcal{NP}_n(A_0) = 1\} F_0(\mathrm{d}\Lambda) = \alpha, \qquad \mathbb{P}_{(A_0,\Lambda)}\{\mathcal{NP}_n(A_0) = 1\} \leq \alpha,\ \forall\Lambda\in\mathscr{L}; \tag{3.20}$$

in which case F_0 is the least favourable distribution (LFD) for the testing problem (for the given F_1-weighed alternative). While exact calculation of the LFD is infeasible, the authors provide an algorithm that delivers an F_0 (and associated cv_α) that approximates the properties of the LFD, in the sense of their Definition 1; the associated test is thus 'nearly optimal' in the WAP sense against F_1.

Their methodology may be applied to the present setting, with Theorem 3.1 indicating the manner in which the limiting experiment depends on the model parameters $(\boldsymbol{\Phi}, \Sigma)$. In particular, Theorem 3.1 justifies simulating from an appropriately parametrised VAR(1) model to approximate the distribution of the likelihood ratio process, and particularly the probabilities in (3.20), which are the key input for the determination of F_0 and cv_α (see Appendix F for details).[7] While the application to our problem is in many respects similar to their application to a predictive regression model (Elliott et al., 2015, Sec. 5.3), with both problems sharing a common limiting experiment, an important difference arises in that we impose a lower bound ρ for the q largest roots, for reasons discussed in Section 2.3 above. Accordingly, we do not develop a 'switching test' of the form that would be needed to accommodate a parametrisation of the model lying deeper in the stationary region.

More generally, the same approach may be taken to testing hypotheses that restrict only a subset of the elements of A under the null, with the remaining elements being concentrated out. While the optimality properties of the resultant test are less clear in that case, the procedure still yields a test of asymptotic level α. For example, a test of $H_0 : a_{ij}(\boldsymbol{\Phi}) = a_0, \Lambda_{\text{LU}} \in \mathscr{L}$, for a

[6] More precisely, this correspondence emerges asymptotically, under the local parametrisation (3.9), with $\boldsymbol{\pi}$ playing the role of (A, Λ_{LU}). Though the other components of the VAR, represented by f, are technically also nuisance parameters, the separability of the limiting experiment in $\boldsymbol{\pi}$ and f entails that these asymptotically play no role in the testing problem. Concentrating these parameters out of the likelihood thus leads to test statistics with the same limiting distribution as if these other components were known *a priori*.

[7] Unlike Elliott et al. (2015), we have phrased the testing problem in terms of the 'original' model parameters, rather than the local parameters that appear in the limiting experiment. Since the approximate LFD and the weighted alternative are defined in terms of the local parameters, F_0 and F_1 in (3.19) should be indexed by n so as to correspond to (sample-size independent) distributions on the local parameter space. In our implementation, where F_0 is obtained by simulating from the finite-sample distribution of a suitably-chosen model, such details are handled implicitly. That this construction yields a test that is asymptotically of size α follows from Theorem 3.1 and the continuous mapping theorem (with the finiteness of the supports of F_0 and F_1 implying that the finite-dimensional convergence obtained in that result is sufficient here): a rigorous development is not given here, to keep the paper to a manageable length.

given (i, j), could be constructed as

$$\mathcal{NP}_n(a_0) \coloneqq \mathbf{1}\left\{\int_{\mathbb{R}\times\mathscr{L}} \mathrm{e}^{\ell_n^*(a,\Lambda)} F_1(\mathrm{d}a, \mathrm{d}\Lambda) > \mathrm{cv}_\alpha \int_{\mathscr{L}} \mathrm{e}^{\ell_n^*(a_0,\Lambda)} F_0(a_0, \mathrm{d}\Lambda)\right\} \tag{3.21}$$

where now $\ell_n^*(a, \Lambda) \coloneqq \max_{\{\Phi\in\mathscr{P}|\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi})=\Lambda, a_{ij}(\boldsymbol{\Phi})=a\}} \ell_n^*(\boldsymbol{\Phi})$. F_0 and F_1 concentrate respectively on subsets of the parameter space for a_{ij} and Λ_{LU} consistent with the null and the alternative, with the remaining elements of A replaced (in a finite sample) by consistent estimators.

3.4 Confidence intervals

Suppose now that interest centres on a given element a_{ij} of A. The preceding suggests a number of possible ways for constructing asymptotically valid $1 - \alpha$ confidence intervals for a_{ij}. Firstly, if Λ_{LU} were known to be some $\Lambda_0 \in \mathscr{L}$, then by Theorem 3.3 a confidence interval based on the efficient (likelihood ratio) test may be constructed conditionally on that Λ_0, as

$$\mathcal{C}_{a_{ij}|\Lambda_0}(\alpha) \coloneqq \{a_0 \in \mathbb{R} \mid \mathcal{LR}_n(a_0; \Lambda_0) \leq \chi^2_{1,1-\alpha}\}. \tag{3.22}$$

$\mathcal{C}_{a_{ij}|\Lambda_0}$ may also be of interest in cases where Λ_{LU} is not plausibly known a priori, insofar as a plot of these intervals illustrates the potential sensitivity (or robustness) of inferences on a_{ij} to departures from the assumption of exact unit roots. (This is particularly feasible when $q = 1$: see Section 4.2 for an illustration).

In the more realistic case that Λ_{LU} is unknown, a well-established approach to inference is based on Bonferroni's inequality, which involves constructing a first-stage confidence interval for Λ_{LU} on the basis of the LR test in Theorem 3.3, as per

$$\mathcal{C}_\Lambda(\alpha_1) \coloneqq \{\Lambda_0 \in \mathscr{L} \mid \mathcal{LR}_n(\Lambda_0) \leq c_{1-\alpha_1}[n(\Lambda_0 - I_q)]\}$$

where $c_\tau(C)$ denotes the τth quantile of the distribution of (3.17), under the local parameter C. By construction, for any $\alpha_1 + \alpha_2 \leq \alpha$, the set

$$\mathcal{C}_{\mathrm{B}}(\alpha_1, \alpha_2) \coloneqq \bigcup_{\Lambda_0\in\mathcal{C}_\Lambda(\alpha_1)} \mathcal{C}_{a_{ij}|\Lambda_0}(\alpha_2),$$

has asymptotic level α. Since this yields inferences on a_{ij} that are necessarily conservative, refinements along lines proposed by Cavanagh et al. (1995) and Campbell and Yogo (2006) in the context of predictive regression (an approach that has since been further extended by McCloskey, 2017) may also be considered here.

However, we have found in practice that the likelihood ratio test of Λ_{LU} generally lacks power, with the associated confidence set $\mathcal{C}_\Lambda(\alpha_1)$ being unsuitably wide. In general, much tighter intervals can be found on the basis of the nearly optimal test of Elliott et al. (2015),

$$\mathcal{C}_{\mathrm{NP}}(\alpha) \coloneqq \{a_0 \in \mathbb{R} \mid \mathcal{NP}_n(a_0) = 0\}.$$

where $\mathcal{NP}_n(a_0)$ is defined in (3.21) above.

3.5 Deterministic terms

For the cointegrated VAR with exact unit roots, Johansen (1995, Sec. 5.7) develops a hierarchy of models (in his notation, $H_2 \subset H_1^* \subset H_1 \subset H^* \subset H$) ordered according to their treatment of the deterministic terms in form (3.1) of the model. In our more general setting where $\Lambda_{\mathrm{LU}} = I_q$ is not required, these models take on an altered expression, and not all are realisable through restrictions on the model parameters. To discuss how we treat might handle deterministic terms in our setting, and their implications for inference, we first recall that the mapping from the DGP (2.1) to form (3.1) of the VAR implies

$$m = \Phi(1)\mu + \Psi\delta, \qquad\qquad d = \Phi(1)\delta,$$

where $\Psi := \sum_{i=1}^{k} i\Phi_i$. Three important cases are the following:[8]

(i) Both μ and δ are unrestricted. The reduced form VAR (3.1) should be estimated with (m, d) unrestricted (as per Johansen's model H). Our asymptotics assume that the DGP is the structural VAR (2.1), so that $d = \Phi(1)\delta$ holds even though this is not imposed in estimation. Indeed, it would not be possible to impose the restriction $d \in \operatorname{sp}\Phi(1)$ (as per Johansen's H^*) in the present setting, because whenever the largest roots of Φ are not exactly unity, $\Phi(1)$ has full rank. Thus $d = \Phi(1)\delta$ would be effectively unrestricted, and a model with exact unit roots and $d \notin \Phi(1)$ would lie in the closure of the parameter space.

(ii) $\delta = 0$, but μ is unrestricted. The VAR (3.1) should be estimated with only a constant (as in Johansen's model H_1). Under the assumption that the DGP is the structural VAR with $\delta = 0$, y_t has no drift. The limit theorems given in Theorems 3.1–3.3 must be amended in this case, by replacing each instance of $\bar{Z}_C$ with the demeaned diffusion process $Z_C(r) - \int_0^1 Z_C(s)\,\mathrm{d}s$. (Imposing the restriction that $m \in \operatorname{sp}\Phi(1)$, as per Johansen's model H_1^*, is impossible in our setting.)

(iii) $\mu = \delta = 0$. The VAR (3.1) should be estimated with $m = d = 0$ (as per Johansen's model H_2); in Theorems 3.1–3.3, $\bar{Z}_C$ is replaced by Z_C.

In light of this, our recommendation is to estimate the model with an unrestricted intercept and trend if there is a discernable linear trend in the data, and to otherwise estimate the model with only an intercept.

4 Finite-sample performance

4.1 Simulations

We conducted simulations to evaluate the finite-sample performance of $\mathcal{NP}_n(a)$, in terms of size and power. For this exercise, natural comparisons are with the likelihood ratio test $\mathcal{LR}_n(a; I_q)$

[8] There is a fourth case, which sits in between the first two, in which a linear trend is present in y_t but is assumed to be eliminated by the quasi-cointegrating relationships, whence $\beta^{\mathsf{T}}\delta = 0$. Since $\beta \in (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$, this is equivalent to requiring $d \in \operatorname{sp}\Phi(1)R_{\mathrm{LU}}$. If we assume exact unit roots, then $\Phi(1)R_{\mathrm{LU}} = 0$ (from (2.6) above) and this restriction can be imposed simply by estimating the VAR (3.1) without a trend (as in Johansen's model H_1). However, in our setting with non-unit roots this restriction cannot be so simply expressed, because $\Phi(1)$ may have full rank; all that can be said is that $d \in \operatorname{sp}\Phi(1)R_{\mathrm{LU}}$. Estimation under this restriction is accordingly more involved, and we leave the development of the asymptotics of our procedure in this case for future work.

corresponding to the efficient rank-imposed MLE (Johansen, 1995), and with the low-frequency stationarity test $\mathcal{ST}_n(a)$ of Müller and Watson (2013), constructed with $b = 10/r^{1/2}$ in their equation (25), as per their recommendations.

The DGP in the simulation design is a bivariate ($p = 2$) VAR(2) with one root near unity ($q = 1$), is parametrised in terms of the underlying (R, Λ) matrices (see Appendix A) as

$$R_{\text{LU}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad R_{\text{ST}} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} \qquad \lambda_{\text{LU}} = \lambda_{\text{LU},0} \qquad \Lambda_{\text{ST}} = \operatorname{diag}\{0.5, 0.4, 0.3\}$$

from which the implied VAR coefficients may be recovered (via Lemma A.1(iv)). The implied quasi-cointegrating vector is $\beta_0 = [1, -a_0]^\mathsf{T}$ with $a_0 = 1$. Across the simulation designs, we vary:

(i) the largest root $\lambda_{\text{LU},0}$ over $\{0.96, 0.98, 1.00\}$; and

(ii) the matrix Σ such that $\Omega = K\Sigma K^\mathsf{T} = [\begin{smallmatrix} 1 & \omega_{\text{SL}} \\ \omega_{\text{SL}} & 1 \end{smallmatrix}]$, for $\omega_{\text{SL}} \in \{-0.9, -0.8, \ldots, 0.9\}$;

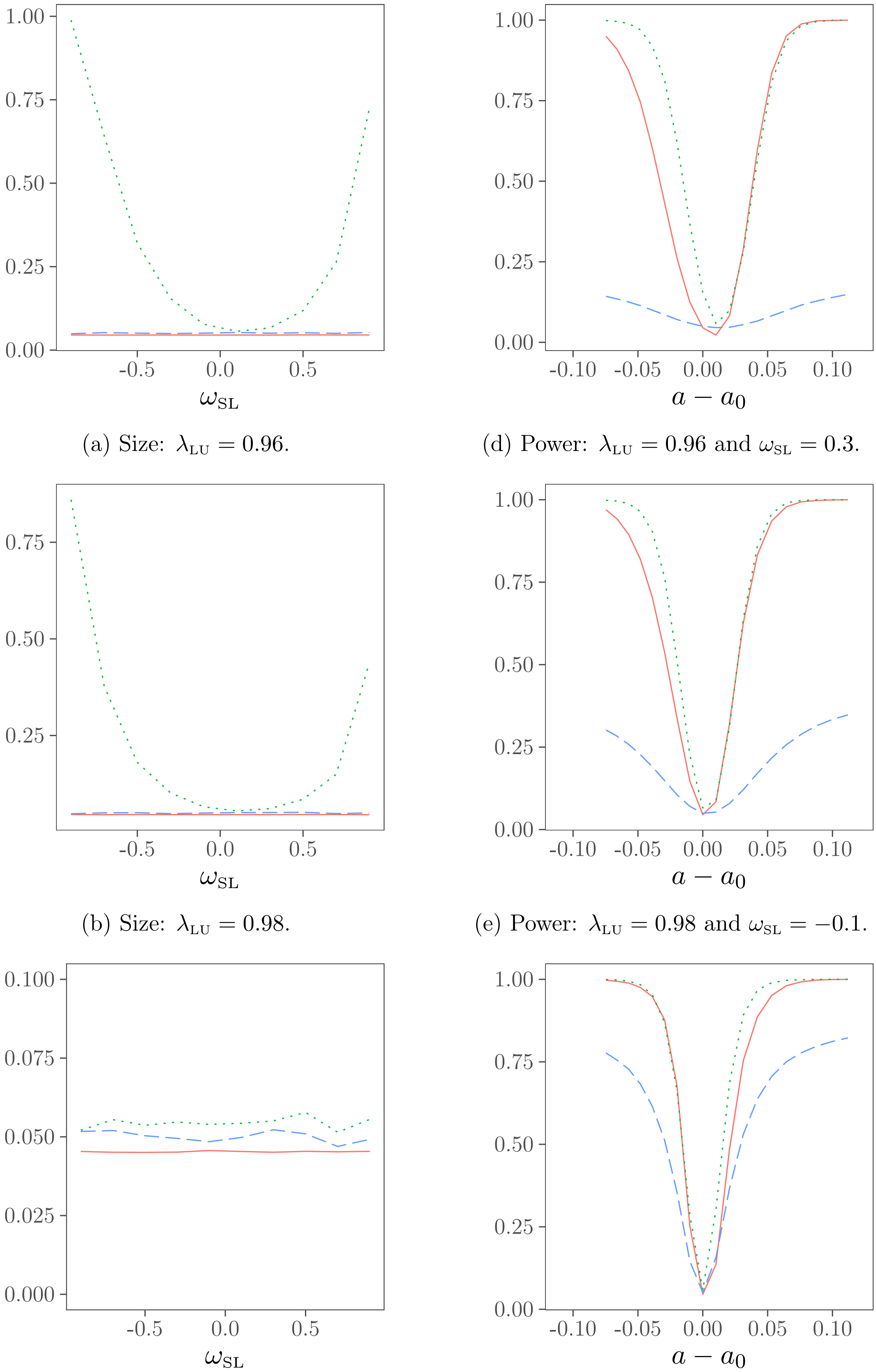
for each of which we generate 20,000 samples of length $n = 200$. In implementing $\mathcal{NP}_n(a)$, we set a lower bound of $\rho = 0.9$ on $\mathscr{L} = [\rho, 1]$, the parameter space for λ_{LU}. When estimating the VAR, we include a constant, and select the lag length by the Akaike information criterion (AIC; with a maximum lag order of 10). The nominal level of all tests is $\alpha = 0.05$.

The role of ω_{SL} in regulating the extent of the size distortion in $\mathcal{LR}_n(a; 1)$ is clearly evident in panels (a)–(c) of Figure 2. That the size distortion disappears at $\omega_{\text{SL}} = 0$ is entirely consistent with Theorem 3.1, which implies that in this case the Hessian of the limiting loglikelihood is diagonal, and that the part of the LR process that refers to a is LAMN. Relative to the test $\mathcal{LR}_n(a; I_q)$ that is efficient in the presence of a unit root, our test entirely avoids the size distortions that this test is prone to, while giving up very little power, as is evident in panels (d)–(f). $\mathcal{ST}_n(a)$ also exhibits perfect size control, but the power sacrificed for the sake of greater robustness – with respect to other departures from the VAR with unit roots – is clearly evidenced by its flatter power curves.

4.2 Application to the expectations theory of the term structure

In Section 2.4.2 above, we discussed the expectations theory of the term structure: in particular, how the predictions made about cointegrating relations in a VAR with exact unit roots, could be generalised to the quasi-cointegrating relations implied by a VAR with some roots near unity. To evaluate these predictions empirically, we estimated a bivariate VAR (with an intercept only) using quarterly data on 1- and 10-year US Treasury bond yields 1953:Q2–2011:Q3, as plotted in Figure 3(a). The Akaike, Bayesian, and Hannan-Quinn information criteria agreed on a lag length of 8, and the dominant characteristic root is estimated to be 0.983. Tests for cointegrating rank (using the trace test of Johansen, 1995) comfortably reject the null of a cointegrating rank of zero but do not reject a cointegrating rank of one, at conventional significance levels, which we take as evidence in favour of the presence of one root in the vicinity of unity.

Consistent with the discussion in Section 2.3, when specifying a parameter space $\mathscr{L} = [\rho, 1]$ for the dominant root, we take $\underline{h} = 8 \times 4 = 32$ quarters to be consistent with a lower estimate of the average duration of the US business cycle, which yields $\rho = 2^{-1/\underline{h}} = 0.979$. The upper panel



(a) Size: $\lambda_{\text{LU}} = 0.96$.

(b) Size: $\lambda_{\text{LU}} = 0.98$.

(c) Size: $\lambda_{\text{LU}} = 1$.

(d) Power: $\lambda_{\text{LU}} = 0.96$ and $\omega_{\text{SL}} = 0.3$.

(e) Power: $\lambda_{\text{LU}} = 0.98$ and $\omega_{\text{SL}} = -0.1$.

(f) Power: $\lambda_{\text{LU}} = 1$ and $\omega_{\text{SL}} = 0.5$.

Figure 2: Rejection probabilities under the simulation design of Section 4.1, for $\mathcal{NP}_n(a)$ (solid line), $\mathcal{ST}_n(a)$ (dashed line) and $\mathcal{LR}_n(a;1)$ (dotted line)

of Figure 3(b) plots the value of the maximised likelihood conditional on a range of values for the dominant root, λ_{LU}, over this interval. The lower panel plots values of the coefficient on the 1-year bond in the normalised quasi-cointegrating vector β with the coefficient on the 10-year bond being normalised to unity. The dashed line reports the model-implied value $a_{10}(\lambda_{\text{LU}})$ of this parameter (see (2.19) above). For each value of the root in the upper panel, the green error bars in the lower panel give the corresponding 95 per cent conditional confidence intervals $\mathcal{C}_{a|\lambda_{\text{LU}}}$ (see (3.22) above), which are based on the efficient test when the imposed value of λ_{LU} is correct; their midpoints are given by the dotted line. Both dashed and dotted lines lie remarkably close to each other, suggesting that the data accord well with the predictions of the expectations theory, irrespective of the actual value of λ_{LU}. Finally, using the nearly optimal test $\mathcal{NP}_n$ gives a 95 per cent confidence interval of $[0.87, 1.07]$, as compared with that of $[0.92, 1.13]$ when an exact unit root is imposed.

5 Conclusion

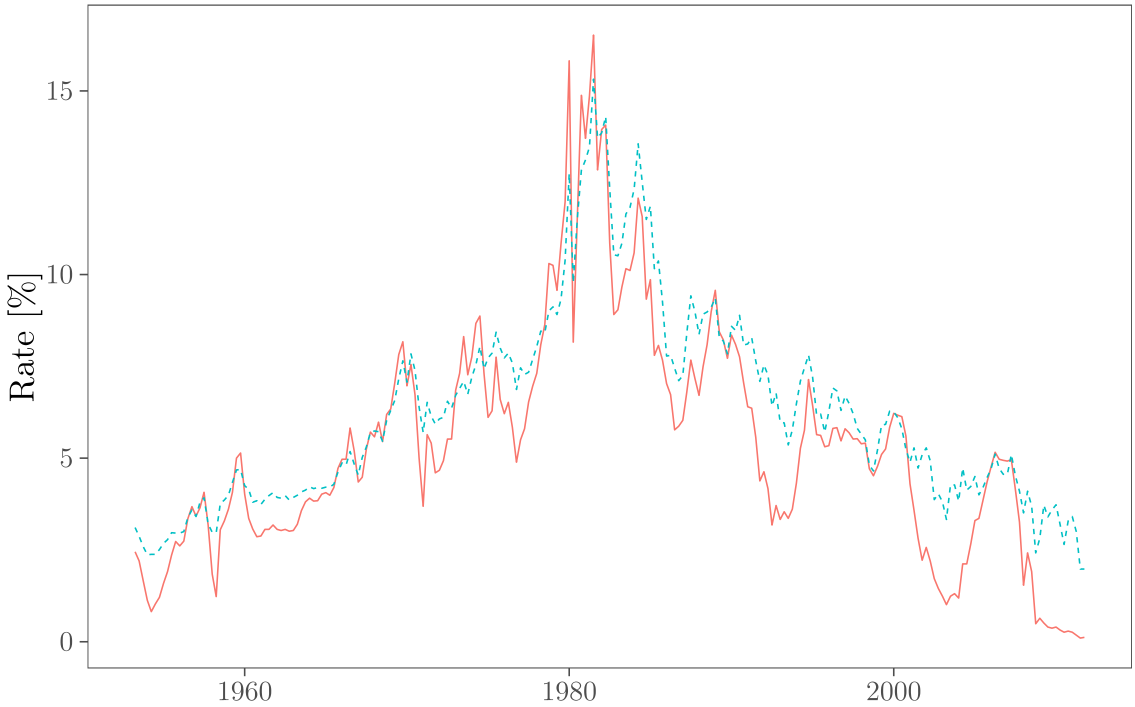
This paper was motivated by Elliott's (1998) finding that inference on cointegrating relationships are highly sensitive to departures from the assumption of exact unit roots. We have argued that this problem is as much one of (non-)identification as it is of inference, because of the manner in which the conventional definitions of cointegration break down in the absence of exact unit roots. We have therefore developed an alternative characterisation of cointegration in an SVAR, in which the long-run equilibrium relationships between the series are identified by those directions for which the impulse responses decay (relatively) most rapidly. With q roots at unity, this exactly recovers the r-dimensional cointegrating space – and when these roots merely near unity, there remains a well-defined q-dimensional quasi-cointegrating space. While this is not the only possible way of extending 'cointegration' to a wider domain, a conceptual advantage of the approach taken here is that it maintains the duality that exists, in an SVAR with exact unit roots, between the identification of the long-run equilibrium relationships between the series, and of the subvector of structural shocks whose common persistent effects underpin those relationships.

Likelihood-based inference on the (quasi-)cointegrating relationships is affected by nuisance parameters corresponding to the proximity of the dominant q roots to unity. We have shown that this problem is not merely reminiscent of inference in a predictive regression, but in fact asymptotically equivalent in the sense of sharing a common limiting experiment. Our problem also falls within the class of problems studied by Elliott et al. (2015), and we have found that, in practice, tests with excellent size and power properties can be developed by adapting their approach to the present setting.

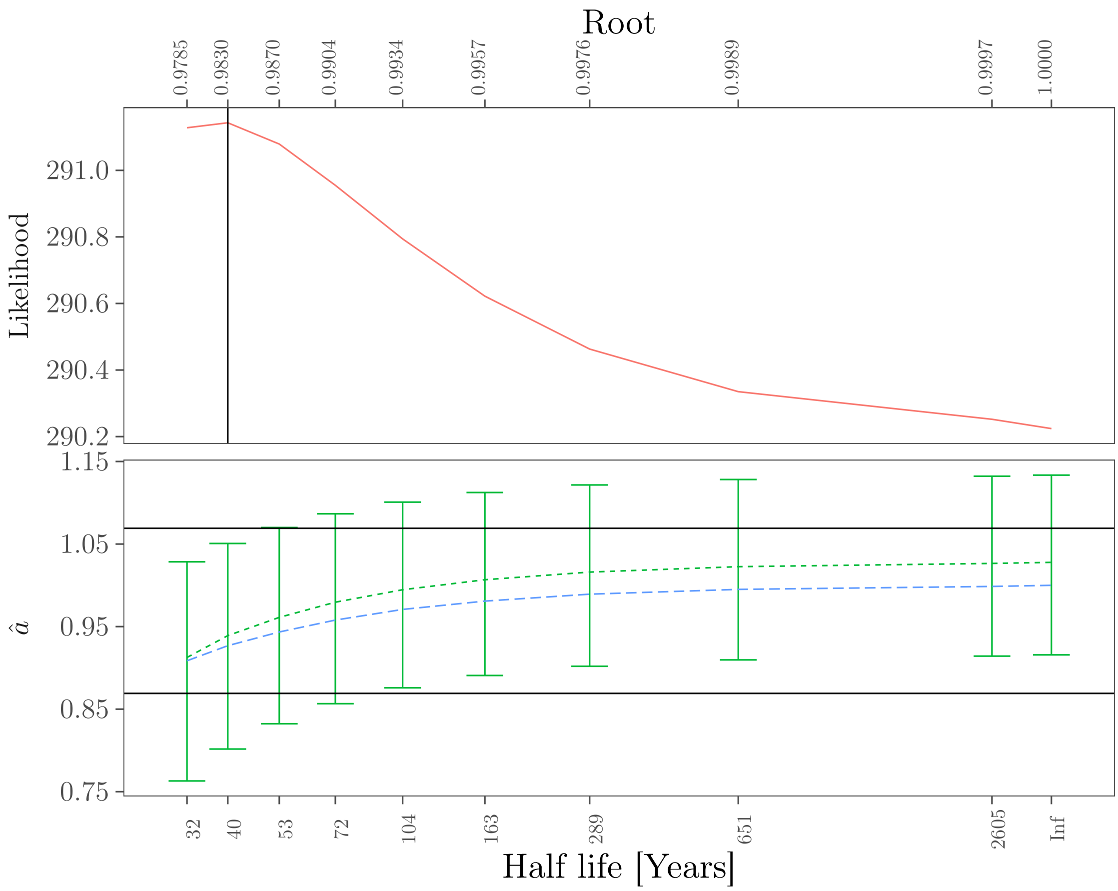
6 References

Blanchard, O. J. and D. Quah (1989): "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655–73.

Campbell, J. Y. (1994): "Inspecting the mechanism: An analytical approach to the stochastic growth model," *Journal of Monetary Economics*, 33, 463–506.



(a) US treasury bond yields: 1-year (solid line) and 10-year (dotted line)



(b) Upper panel: concentrated loglikelihood at imposed characteristic root λ_{LU}.
Bottom panel: 95% conditional intervals $\mathcal{C}_{a|\lambda_{\text{LU}}}$ (green error bars); $\mathcal{NP}_n$ interval (horizontal lines); midpoints of $\mathcal{C}_{a|\lambda_{\text{LU}}}$ (dotted green line); theory-implied value of $a_{10}(\lambda_{\text{LU}})$ (dashed blue line)

Figure 3: Expectations theory of term spread.

Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062–1088.

Campbell, J. Y. and M. Yogo (2006): "Efficient tests of stock return predictability," *Journal of Financial Economics*, 81, 27–60.

Carriero, A., C. A. Favero, and I. Kaminska (2006): "Financial factors, macroeconomic information and the expectations theory of the term structure of interest rates," *Journal of Econometrics*, 131, 339–358.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): "Inference in models with nearly integrated regressors," *Econometric Theory*, 11, 1131–1147.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2006): "Assessing Structural VARs," *NBER Macroeconomics Annual*, 21, 1–105.

d'Autume, A. (1992): "Deterministic dynamics and cointegration of higher orders," Tech. rep., Université de Paris I.

Elliott, G. (1998): "On the robustness of cointegration methods when regressors almost have unit roots," *Econometrica*, 66, 149–158.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly optimal test when a nuisance parameter is present under the null hypothesis," *Econometrica*, 83, 771–811.

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

Fernández-Villaverde, J., J. F. Rubio-Ramirez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of understanding VARs," *American economic review*, 97, 1021–1026.

Franchi, M. and S. Johansen (2017): "Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles," *Econometrics*, 5, 25–44.

Fuhrer, J. and G. Moore (1995): "Inflation persistence," *Quarterly Journal of Economics*, 110, 127–159.

Gali, J. (1999): "Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?" *American Economic Review*, 89, 249–271.

Gohberg, I., S. Lancaster, and L. Rodman (1982): *Matrix Polynomials*, London (UK): Academic Press.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, New York (USA): Academic Press.

Hendry, D. F. (1986): "Econometric modelling with cointegrated variables: an overview," *Oxford Bulletin of Economics and Statistics*, 48, 201–212.

Horn, R. A. and C. R. Johnson (2013): *Matrix Analysis*, New York (USA): C.U.P., 2nd ed.

Hwang, J. and G. Valdés (2023): "Low freqency cointegrating regression with local to unity regressors and unknown form of serial dependence," *Journal of Business & Economic Statistics*, in press, DOI:10.1080/07350015.2023.2166513.

Jansson, M. and M. J. Moreira (2006): "Optimal inference in regression models with nearly integrated regressors," *Econometrica*, 74, 681–714.

Johansen, S. (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, O.U.P.

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

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autorgressive Analysis*, C.U.P.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015): "Robust econometric inference for stock return predictability," *Review of Financial Studies*, 28, 1506–1553.

Lang, S. (1993): *Real and Functional Analysis*, New York (USA): Springer, 3rd ed.

Ljungqvist, L. and T. J. Sargent (2004): *Recursive Macroeconomic Theory*, MIT Press.

Lütkepohl, H. (2007): *New Introduction to Multiple Time Series Analysis*, Springer, 2nd ed.

Maccini, L. J., B. J. Moore, and H. Schaller (2004): "The interest rate, learning, and inventory investment," *American Economic Review*, 94, 1303–1327.

Magdalinos, T. and P. C. B. Phillips (2009): "Econometric inference in the vicinity of unity," CoFie Working Paper 7, Singapore Management University.

Magnus, J. R. and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Chichester (UK): Wiley, 3rd ed.

McCloskey, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200.

Müller, U. K. and M. W. Watson (2013): "Low-frequency robust cointegration testing," *Journal of Econometrics*, 174, 66–81.

Onatski, A. and H. Uhlig (2012): "Unit roots in white noise," *Econometric Theory*, 28, 485–508.

Phillips, P. C. B. (1988): "Regression theory for near-integrated time series," *Econometrica*, 56, 1021–1043.

——— (1991): "Optimal inference in cointegrated systems," *Econometrica*, 59, 283–306.

——— (2014): "On confidence intervals for autoregressive roots and predictive regression," *Econometrica*, 82, 1177–1195.

Phillips, P. C. B. and B. E. Hansen (1990): "Statistical inference in instrumental variables regression with $I(1)$ processes," *Review of Economic Studies*, 57, 99–125.

Phillips, P. C. B. and J. H. Lee (2013): "Predictive regression under various degrees of persistence and robust long-horizon regression," *Journal of Econometrics*, 177, 250–264.

Ravenna, F. (2007): "Vector autoregressions and reduced form representations of DSGE models," *Journal of Monetary Economics*, 54, 2048–2064.

Shiue, C. H. and W. Keller (2007): "Markets in China and Europe on the eve of the Industrial Revolution," *American Economic Review*, 97, 1189–1216.

Stewart, G. W. and J.-g. Sun (1990): *Matrix Perturbation Theory*, Boston (USA): Academic Press.

Stock, J. H. and M. W. Watson (1993): "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica*, 61, 783–820.

Sun, J.-g. (1991): "Perturbation expansions for invariant subspaces," *Linear Algebra and its Applications*, 153, 85–97.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, New York (USA): Cambridge University Press.

Wright, J. H. (2000): "Confidence sets for cointegrating coefficients based on stationarity tests," *Journal of Business and Economic Statistics*, 18, 211–222.

Appendices

Notation. For $x \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times p}$, $\|x\|$ denotes the Euclidean norm and $\|A\| := \sup_{\|x\|=1} \|Ax\|$ the induced matrix norm.

A Representation theory

This section provides results that support some of the assertions made in the course of Sections 2 and 3, and which are auxiliary to results proved in the following appendices. Some are well known, but are collected here for ease of reference. Proofs follow at the end of this appendix. DGP is maintained throughout.

For VAR coefficients $\mathbf{\Phi} := (\Phi_1, \ldots, \Phi_k) \in \mathbb{R}^{p \times kp}$, let

$$F := F(\mathbf{\Phi}) := \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{k-1} & \Phi_k \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \tag{A.1}$$

denote the associated companion form matrix. For a collection of $m \times n$ matrices $Z_1, \ldots, Z_k$, let

$$\operatorname{col}\{Z_i\}_{i=1}^k := \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \end{bmatrix},$$

so that taking $\mathbf{x}_t := \operatorname{col}\{x_{t-i}\}_{i=0}^{k-1}$, we may write (2.1) as

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + \begin{bmatrix} \varepsilon_t \\ 0_{(k-1)p \times 1} \end{bmatrix} =: F\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{A.2}$$

Let $\lambda_i(\mathbf{\Phi})$ denote the ith root of the characteristic polynomial associated to $\mathbf{\Phi}$, when these are placed in descending order of modulus.

Lemma A.1. *Suppose that $|\lambda_q(\mathbf{\Phi})| > |\lambda_{q+1}(\mathbf{\Phi})|$ for some $q \in \{1, \ldots, p\}$. Then there exist there matrices $R \in \mathbb{R}^{p \times kp}$, $\Lambda \in \mathbb{R}^{kp \times kp}$ and $L \in \mathbb{R}^{p \times kp}$ such that:*

(i) $\Lambda = \operatorname{diag}\{\Lambda_{\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$, *where the eigenvalues of $\Lambda_{\mathrm{LU}} \in \mathbb{R}^{q \times q}$ and Λ_{ST} are $\{\lambda_i(\mathbf{\Phi})\}_{i=1}^q$ and $\{\lambda_i(\mathbf{\Phi})\}_{i=q+1}^{kp}$ respectively;*

(ii) *the following hold:*

$$R\Lambda^k - \sum_{i=1}^k \Phi_i R \Lambda^{k-i} = 0 \qquad \Lambda^k L^\mathsf{T} - \sum_{i=1}^k \Lambda^{k-i} L^\mathsf{T} \Phi_i = 0. \tag{A.3}$$

(iii) $\mathbf{R} := \operatorname{col}\{R\Lambda^{k-i}\}_{i=1}^k$ *is invertible, and L equals the first p rows of $\mathbf{L} := (\mathbf{R}^{-1})^\mathsf{T}$;*

(iv) $F(\mathbf{\Phi}) = \mathbf{R}\Lambda\mathbf{L}^{\mathsf{T}}$; *and*

(v) *in the model (2.1),* $\mathrm{IRF}_s^{\varepsilon} \coloneqq \partial y_{t+s}/\partial\varepsilon_t = R\Lambda^{k-1+s}L^{\mathsf{T}}$ *for* $s \geq 1$.

Further, the matrices $R^* \in \mathbb{R}^{p\times kp}$, $\Lambda^* \in \mathbb{R}^{kp\times kp}$ *and* $L^* \in \mathbb{R}^{p\times kp}$ *satisfy conditions (i)–(v) if and only if there exists an invertible* $kp \times kp$ *matrix* $Q = \operatorname{diag}\{Q_{\mathrm{LU}}, Q_{\mathrm{ST}}\}$, *where* $Q_{\mathrm{LU}} \in \mathbb{R}^{q\times q}$, *such that* $R^* = RQ$, $\Lambda^* = Q^{-1}\Lambda Q$ *and* $L^* = L(Q^{\mathsf{T}})^{-1}$.

For a given $\mathbf{\Phi}$, and its associated companion form $F = F(\mathbf{\Phi})$, we shall routinely partition the matrices appearing in Lemma A.1 as

$$R \coloneqq [R_{\mathrm{LU}}, R_{\mathrm{ST}}] \qquad \mathbf{R} \coloneqq [\mathbf{R}_{\mathrm{LU}}, \mathbf{R}_{\mathrm{ST}}] \qquad L \coloneqq [L_{\mathrm{LU}}, L_{\mathrm{ST}}] \qquad \mathbf{L} \coloneqq [\mathbf{L}_{\mathrm{LU}}, \mathbf{L}_{\mathrm{ST}}] \tag{A.4}$$

where each of R_{LU}, $\mathbf{R}_{\mathrm{LU}}$, L_{LU} and $\mathbf{L}_{\mathrm{LU}}$ have q columns, i.e. the partitioning is conformable with that of $\Lambda = \operatorname{diag}\{\Lambda_{\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$. This partitioning, in conjunction with parts (ii) and (v) of the preceding lemma, yields (2.6) and (2.7) above. Moreover, we may write part (iv) as

$$F = \mathbf{R}\Lambda\mathbf{L}^{\mathsf{T}} = \mathbf{R}_{\mathrm{LU}}\Lambda_{\mathrm{LU}}\mathbf{L}_{\mathrm{LU}}^{\mathsf{T}} + \mathbf{R}_{\mathrm{ST}}\Lambda_{\mathrm{ST}}\mathbf{L}_{\mathrm{ST}}^{\mathsf{T}} \tag{A.5}$$

which decomposes F with respect to the invariant subspaces associated to the eigenvalues of Λ_{LU} and Λ_{ST}.

Lemma A.2. *Suppose that* $|\lambda_q(\mathbf{\Phi})| > |\lambda_{q+1}(\mathbf{\Phi})|$ *for some* $q \in \{1, \ldots, p\}$, *the eigenvalues of* $\Lambda_0 \in \mathbb{R}^{q\times q}$ *are all greater than* $|\lambda_{q+1}(\mathbf{\Phi})|$ *in modulus, and* $R_0 \in \mathbb{R}^{p\times q}$ *is a full column rank matrix such that*

$$R_0\Lambda_0^k - \sum_{i=1}^{k} \Phi_i R_0 \Lambda_0^{k-i} = 0. \tag{A.6}$$

Then there exist matrices $R = [R_{\mathrm{LU}}, R_{\mathrm{ST}}]$, $\Lambda = \operatorname{diag}\{\Lambda_{\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$ *and* L *satisfying the conditions of Lemma A.1, with* $R_{\mathrm{LU}} = R_0$ *and* $\Lambda_{\mathrm{LU}} = \Lambda_0$.

For the next result, recall the definition of S_r given in the context of (2.3) above. By showing that $S_r = (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$, we substantiate the claim made in the Section 2.2 that S_r is invariant to the identification of the structural shocks w_t. S_r may thus be equivalently defined with $\mathrm{IRF}_s^{\varepsilon}$ taking the place of IRF_s^{w} in (2.3).

Lemma A.3.

(i) *If* QC *holds for some* $\rho \in (0, 1]$, *then* $S_r = (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$.

(ii) *If* CV *holds, then* $\mathrm{CS} = S_r = (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$, *and* QC *holds with* $\rho = 1$.

Lemma A.4. *Suppose* QC *holds. Let* $\Lambda = \operatorname{diag}\{\Lambda_{\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$, $R = [R_{\mathrm{LU}}, R_{\mathrm{ST}}]$ *and* $\mathbf{L} = [\mathbf{L}_{\mathrm{LU}}, \mathbf{L}_{\mathrm{ST}}]$ *be as in Lemma A.1 and (A.4). Then (2.9)–(2.10) hold with* $\Phi_{\mathrm{LU}} = R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^k$, $\Phi_{\mathrm{ST}} = R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^k$, $z_{\mathrm{LU},t} \coloneqq \mathbf{L}_{\mathrm{LU}}^{\mathsf{T}}\mathbf{x}_t$ *and* $z_{\mathrm{ST},t} \coloneqq \mathbf{L}_{\mathrm{ST}}^{\mathsf{T}}\mathbf{x}_t$.

One aspect of the cointegrated VAR with exact unit roots, which does not readily translate to the quasi-cointegrated VAR, is the 'error correction' representation of the adjustments towards equilibrium. The only situation in which an analogue of that representation is available is when the q largest roots are identical, so that $\Lambda_{\mathrm{LU}} = \lambda_0 I_q$ for some $\lambda_0 \in \mathcal{L}^{\rho}_{\mathrm{LU}}$, as arises automatically if $q = 1$. (Note that λ_0 must be real in this case.) In that case, letting $\Delta_\lambda x_t \coloneqq x_t - \lambda x_{t-1}$, we have the following.

Lemma A.5. *Suppose* QC *holds with* $\Lambda_{\mathrm{LU}} = \lambda_0 I_q$. *Then* $\{x_t\}$ *satisfies*

$$\Delta_{\lambda_0} x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Psi_i \Delta_{\lambda_0} x_{t-i} + \varepsilon_t, \tag{A.7}$$

where $\Pi \coloneqq -\lambda_0^{-k+1}\Phi(\lambda_0) = \alpha\beta^{\mathsf{T}}$, *for some* $\alpha \in \mathbb{R}^{p\times r}$ *having full column rank. Moreover,* $(\Delta_{\lambda_0} x_t, \beta^{\mathsf{T}} x_t)$ *follow a VAR(*$k-1$*), all of whose characteristic roots lie in* $\mathcal{L}^{\rho}_{\mathrm{ST}}$.

Proof of Lemma A.1. Let J denote a $(kp \times kp)$ real Jordan matrix similar to F, each of whose diagonal blocks correspond to roots of $\Phi(\cdot)$, so that $P^{-1}FP = J$ for some $P \in \mathbb{R}^{kp\times kp}$. We may take the diagonal blocks of J to be ordered such that $J = \mathrm{diag}\{J_{\mathrm{LU}}, J_{\mathrm{ST}}\}$, where $J_{\mathrm{LU}} \in \mathbb{R}^{q\times q}$ has all its eigenvalues in $\mathcal{L}^{\rho}_{\mathrm{LU}}$. Letting

$$X \coloneqq [0_p \quad \cdots \quad 0_p \quad I_p]P$$

we have by Gohberg, Lancaster, and Rodman (1982, Thm. 1.24 and 1.25) that the matrices (X, J) form a standard pair for $\Phi(\cdot)$.[9] Therefore,

$$XJ^k - \sum_{i=1}^{k} \Phi_i X J^{k-i} = 0,$$

and $\mathrm{col}\{XJ^{k-i}\}_{i=1}^{k} = P$ is invertible, so that the matrix

$$Y \coloneqq [I_p \quad \cdots \quad 0_p \quad 0_p](P^{\mathsf{T}})^{-1} \tag{A.8}$$

is well defined. By Gohberg et al. (1982, Prop. 2.1), (Y, J) satisfy

$$J^k Y^{\mathsf{T}} - \sum_{i=1}^{k} J^{k-i} Y^{\mathsf{T}} \Phi_i = 0.$$

Parts (i)–(iv) of the lemma are thus satisfied with $(R, \Lambda, L^{\mathsf{T}}, \mathbf{R}, \mathbf{L}^{\mathsf{T}}) = (X, J, Y^{\mathsf{T}}, P, P^{-1})$. It further follows by recursive substitution that

$$\mathrm{IRF}_s = [F^s]_{11} = [\mathbf{R}\Lambda^s \mathbf{L}^{\mathsf{T}}]_{11} = R\Lambda^{k-1+s} L^{\mathsf{T}}$$

where $[A]_{11}$ denotes the upper left $p \times p$ block of the matrix A; thus part (v) is proved.

Finally, let $Q = \mathrm{diag}\{Q_{\mathrm{LU}}, Q_{\mathrm{ST}}\}$ be as in the final part of the lemma. It is easily verified that

$$\Lambda_* \coloneqq \mathrm{diag}\{Q_{\mathrm{LU}}^{-1}\Lambda_{\mathrm{LU}}Q_{\mathrm{LU}}, Q_{\mathrm{ST}}^{-1}\Lambda_{\mathrm{ST}}Q_{\mathrm{ST}}\} = Q^{-1}\Lambda Q,$$

$R_* \coloneqq RQ$ and $L_* \coloneqq L(Q^{\mathsf{T}})^{-1}$ have the required properties. Conversely, if both (R, Λ, L) and

[9] Note that the 'first companion form' matrix defined by these authors (C_1 on p. 13 of that work) equals F with the ordering of its rows and columns reversed, so our definitions of X (and below, Y) differ from theirs.

(R_*, Λ_*, L_*) satisfy conditions (i)–(v), then both Λ and Λ_* are block diagonal matrices similar to $J = \operatorname{diag}\{J_{\mathrm{LU}}, J_{\mathrm{ST}}\}$, whence there exists $Q = \operatorname{diag}\{Q_{\mathrm{LU}}, Q_{\mathrm{ST}}\}$ such that $\Lambda^* = Q^{-1}\Lambda Q$, etc. □

Proof of Lemma A.2. $\mathbf{R}_0 := \operatorname{col}\{R_0\Lambda_0^{k-i}\}_{i=1}^{k} \in \mathbb{R}^{kp\times q}$ has rank q, and (A.6) implies that $F\mathbf{R}_0 = \mathbf{R}_0\Lambda_0$, for $F := F(\mathbf{\Phi})$. Since the remaining $kp - q$ eigenvalues of F are distinct from the eigenvalues of Λ_0, $\mathbf{R}_0$ is a simple invariant subspace of F (Stewart and Sun, 1990, Defn V.1.2). Hence there exist $\mathbf{R}, \Lambda, \mathbf{L} \in \mathbb{R}^{kp\times kp}$ such that $F = \mathbf{R}\Lambda\mathbf{L}^{\mathsf{T}}$ and $\mathbf{L}^{\mathsf{T}}\mathbf{R} = I_{kp}$, and $\mathbf{R}$ and Λ can be partitioned as $\mathbf{R} = [\mathbf{R}_0, \mathbf{R}_{\mathrm{ST}}]$ and $\Lambda = \operatorname{diag}\{\Lambda_0, \Lambda_{\mathrm{ST}}\}$ (Stewart and Sun, 1990, Thm V.1.5). Since Λ_0 and Λ_{ST} must be similar to the blocks J_{LU} and J_{ST} of the real Jordan form of F, as introduced in the proof of Lemma A.1, the result then follows by the same arguments as were given in that proof. □

Proof of Lemma A.3. **(i).** By Lemma A.1(v), for any $b \in \mathbb{R}^p$,

$$b^{\mathsf{T}}\mathrm{IRF}_s^w = b^{\mathsf{T}}\mathrm{IRF}_s^{\varepsilon}\Upsilon = b^{\mathsf{T}}R\Lambda^{k-1+s}L^{\mathsf{T}}\Upsilon = b^{\mathsf{T}}R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^{k-1+s}L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon + b^{\mathsf{T}}R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^{k-1+s}L_{\mathrm{ST}}^{\mathsf{T}}\Upsilon. \tag{A.9}$$

Since the spectral radius of Λ_{ST} is strictly less than ρ, we have by Horn and Johnson (2013, Cor. 5.6.13) that

$$\Lambda_{\mathrm{ST}}^t/\rho^t \to 0 \tag{A.10}$$

as $t \to \infty$. Since Λ_{LU} is diagonalisable under QC2, by Lemma A.1 we may choose $(R_{\mathrm{LU}}, \Lambda_{\mathrm{LU}}, L_{\mathrm{LU}})$ such that Λ_{LU} is a real Jordan block diagonal matrix (as in Corollary 3.4.1.10 of Horn and Johnson, 2013). The eigenvalues of $\Lambda_{\mathrm{LU}}^{\mathsf{T}}\Lambda_{\mathrm{LU}} = \Lambda_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^{\mathsf{T}}$ are therefore of the form $|\lambda|^2$, for λ an eigenvalue of Λ_{LU}, and thus $\lambda_{\min}(\Lambda_{\mathrm{LU}}^{\mathsf{T}}\Lambda_{\mathrm{LU}}) \geq \rho^2$, where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of a positive-definite matrix M. Therefore letting $x := R_{\mathrm{LU}}^{\mathsf{T}}b$,

$$\begin{aligned}\|x^{\mathsf{T}}\Lambda_{\mathrm{LU}}^t L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon\|^2 &\geq \lambda_{\min}(L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon^{\mathsf{T}}\Upsilon L_{\mathrm{LU}})\|x^{\mathsf{T}}\Lambda_{\mathrm{LU}}^t\|^2 \\ &\geq \rho\lambda_{\min}(L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon^{\mathsf{T}}\Upsilon L_{\mathrm{LU}})\|\Lambda_{\mathrm{LU}}^{t-1}x\|^2 \geq \cdots \geq \rho^{2t}\lambda_{\min}(L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon^{\mathsf{T}}\Upsilon L_{\mathrm{LU}})\|x\|^2.\end{aligned}$$

$\lambda_{\min}(L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon^{\mathsf{T}}\Upsilon L_{\mathrm{LU}}) > 0$, since Υ is nonsingular, and L_{LU} has full column rank under QC2. Deduce that if $b^{\mathsf{T}}R_{\mathrm{LU}} \neq 0$, then

$$\liminf_{t\to\infty}\|b^{\mathsf{T}}R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^t L_{\mathrm{LU}}^{\mathsf{T}}\Upsilon\|/\rho^t > 0. \tag{A.11}$$

It follows from (A.9)–(A.11) that $b^{\mathsf{T}}\mathrm{IRF}_s^w/\rho^s \to 0$ as $s \to \infty$ if and only if $b \perp \operatorname{sp} R_{\mathrm{LU}}$. Thus $(\operatorname{sp} R_{\mathrm{LU}})^{\perp}$ gives the unique r-dimensional subspace of $\mathbb{R}^p$ satisfying the definition of S_r.

(ii). Since $\operatorname{rk}\Phi(1) = p - q$ under CV2, there exists $R_{\mathrm{LU}} \in \mathbb{R}^{p\times q}$ having rank q such that

$$0 = \Phi(1)R_{\mathrm{LU}} = R_{\mathrm{LU}} - \sum_{i=1}^{k}\Phi_i R_{\mathrm{LU}} =_{(1)} R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^k - \sum_{i=1}^{k}\Phi_i R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^{k-i} \tag{A.12}$$

where $=_{(1)}$ follows by taking $\Lambda_{\mathrm{LU}} = I_q$. By a similar argument, here exists a $L_{\mathrm{LU}} \in \mathbb{R}^{p\times q}$ with $\operatorname{rk} L_{\mathrm{LU}} = q$ and $L_{\mathrm{LU}}^{\mathsf{T}}\Phi(1) = 0$. CV is thus a special case of QC with $\rho = 1$. $S_r = (\operatorname{sp} R_{\mathrm{LU}})^{\perp}$ therefore follows immediately from part (i) of the lemma. Finally, recall from the second characterisation of the cointegrating space given in Section 2.2 that $\mathrm{CS} = \{\ker\Phi(1)\}^{\perp}$. By (A.12) this also coincides with $(\operatorname{sp} R_{\mathrm{LU}})^{\perp}$. □

Proof of Lemma A.4. By (A.2) and Lemma A.1,

$$\mathbf{L}^{\mathsf{T}}\mathbf{x}_t = \mathbf{L}^{\mathsf{T}}F\mathbf{x}_{t-1} + L^{\mathsf{T}}\varepsilon_t = \Lambda\mathbf{L}^{\mathsf{T}}\mathbf{x}_{t-1} + L^{\mathsf{T}}\varepsilon_t.$$

Since $\Lambda = \operatorname{diag}\{\Lambda_{\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$, it is clear that (2.10) holds for $z_{\mathrm{LU},t}$ and $z_{\mathrm{ST},t}$ as defined in the lemma. Further, taking the first p rows of (A.2) and using Lemma A.1 again yields

$$x_t = R\Lambda^k\mathbf{L}^{\mathsf{T}}\mathbf{x}_{t-1} + \varepsilon_t = R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^k\mathbf{L}_{\mathrm{LU}}^{\mathsf{T}}\mathbf{x}_{t-1} + R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^k\mathbf{L}_{\mathrm{ST}}^{\mathsf{T}}\mathbf{x}_{t-1} + \varepsilon_t. \qquad \square$$

Proof of Lemma A.5. The representation (A.7) follows directly from Theorem 1 in Johansen and Schaumburg (1999). By the same arguments as given in the proof of Corollary 4.3 in Johansen (1995), since Φ has q roots at λ_0, $\Phi(\lambda_0)$ must have rank at least equal to $r = p - q$; that it has rank equal to r then follows from QC2, which implies that $\Phi(\lambda_0)R_{\mathrm{LU}} = 0$. For the final claim, we note that under the maintained assumption that $\Lambda_{\mathrm{LU}} = \lambda_0 I_q$, we have from (A.5) that

$$\Delta_{\lambda_0}\mathbf{x}_t = \mathbf{x}_t - \lambda_0\mathbf{x}_{t-1} = (F - \lambda_0 I_{kp})\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t = \mathbf{R}_{\mathrm{ST}}(\Lambda_{\mathrm{ST}} - \lambda_0 I_{kp-q})\mathbf{L}_{\mathrm{ST}}^{\mathsf{T}}\mathbf{x}_t + \boldsymbol{\varepsilon}_t. \tag{A.13}$$

By Lemma A.1, $\mathbf{L}_{\mathrm{ST}} \in \mathbb{R}^{kp\times(kp-q)}$ is a full column rank matrix such that $\mathbf{L}_{\mathrm{ST}}^{\mathsf{T}}\mathbf{R}_{\mathrm{LU}} = 0$, where $\mathbf{R}_{\mathrm{LU}} = \operatorname{col}\{R_{\mathrm{LU}}\Lambda_{\mathrm{LU}}^{k-i}\}_{i=1}^k = \operatorname{col}\{\lambda_0^{k-i}R_{\mathrm{LU}}\}_{i=1}^k$. By considering the column span of $\mathbf{R}_{\mathrm{LU}}$, it follows that $\mathbf{L}_{\mathrm{ST}}$ must have the same column span as

$$\begin{bmatrix} \beta & I_p & & \\ & -\lambda_0 I_p & \ddots & \\ & & \ddots & I_p \\ & & & -\lambda_0 I_p \end{bmatrix}.$$

Hence there exists a full-rank $\Psi \in \mathbb{R}^{(kp-q)\times(kp-q)}$ such that

$$\mathbf{L}_{\mathrm{ST}}^{\mathsf{T}}\mathbf{x}_t = \Psi \begin{bmatrix} \beta^{\mathsf{T}}x_t \\ \Delta_{\lambda_0}x_t \\ \vdots \\ \Delta_{\lambda_0}x_{t-k+2} \end{bmatrix}. \tag{A.14}$$

We recall from Lemma A.4 that the l.h.s. is equal to $z_{\mathrm{ST},t}$, which by that result has a first-order autoregressive representation with characteristic roots that are the eigenvalues of Λ_{ST}, and hence in $\mathcal{L}_{\mathrm{ST}}^{\rho}$. Indeed, (2.10a) provides the companion form representation for autoregressive process followed by $(\Delta_{\lambda_0}x_t, \beta^{\mathsf{T}}x_t)$, which accordingly has the properties claimed. $\square$

Proof of Proposition 2.1. This is an immediate corollary of Lemma A.3. $\square$

B Perturbation theory

Recall the definition of $\mathscr{P} \subset \mathbb{R}^{p\times kp}$ given in Section 3.1.2. The normalisation (3.3) entails that

$$\begin{bmatrix} A \\ I_q \end{bmatrix} \Lambda_{\mathrm{LU}}^k - \sum_{i=1}^k \Phi_i \begin{bmatrix} A \\ I_q \end{bmatrix} \Lambda_{\mathrm{LU}}^{k-i} = 0 \tag{B.1}$$

which by Lemmas A.1 and A.2 uniquely determines $R_{\mathrm{LU}} = [\begin{smallmatrix} A \\ I_q \end{smallmatrix}]$ and Λ_{LU} as a function of $\mathbf{\Phi} \in \mathscr{P}$. As in Section 3.1.2, we shall denote the implied mappings by $R_{\mathrm{LU}}(\mathbf{\Phi})$, $A(\mathbf{\Phi})$, $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$, and $\mathbf{R}_{\mathrm{LU}}(\mathbf{\Phi}) \coloneqq \mathrm{col}\{R_{\mathrm{LU}}(\mathbf{\Phi})\Lambda_{\mathrm{LU}}^{k-i}(\mathbf{\Phi})\}_{i=1}^k$. Our first result is that these are smooth (i.e. infinitely differentiable); its proof and those of the subsequent lemmas appear at the end of this appendix.

Lemma B.1. *$\mathscr{P}$ is open; and $A(\mathbf{\Phi})$ and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$ are smooth on $\mathscr{P}$.*

Our next result gives the first derivatives of the maps $A(\mathbf{\Phi})$ and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$; it is closely related to Theorem 2.1 in Sun (1991). To express these derivatives more concisely, let

$$B(\mathbf{\Phi}) \coloneqq (I_q \otimes R_{\mathrm{ST}})[(\Lambda_{\mathrm{LU}}^{\mathsf{T}} \otimes I_{kp-q}) - (I_q \otimes \Lambda_{\mathrm{ST}})]^{-1}(I_q \otimes L_{\mathrm{ST}}^{\mathsf{T}}), \tag{B.2}$$

where we have suppressed the dependence of each of the r.h.s. quantities on $\mathbf{\Phi}$. The matrix in square brackets on the r.h.s. has eigenvalues of the form $\lambda - \mu$, where λ and μ are eigenvalues of Λ_{LU} and Λ_{ST} respectively; it is thus invertible for all $\mathbf{\Phi} \in \mathscr{P}$. Under the normalisation implied by (B.1), B is uniquely determined by $\mathbf{\Phi} \in \mathscr{P}$, even though R_{ST}, Λ_{ST} and L_{ST} individually are not (as follows from the final part of Lemma A.1).

Lemma B.2. *Let $\mathbf{\Phi}_0 \in \mathscr{P}$, $A_0 \coloneqq A(\mathbf{\Phi}_0)$, $\Lambda_{0,\mathrm{LU}} \coloneqq \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_0)$, $R_{0,\mathrm{LU}} \coloneqq [\begin{smallmatrix} A_0 \\ I_q \end{smallmatrix}]$ and $\mathbf{R}_{0,\mathrm{LU}} \coloneqq \mathrm{col}\{R_{0,\mathrm{LU}}\Lambda_{0,\mathrm{LU}}^{k-i}\}_{i=1}^k$. Then*

(i) *$A_0 = A(\mathbf{\Phi})$ and $\Lambda_{0,\mathrm{LU}} = \Lambda_{\mathrm{LU}}(\mathbf{\Phi})$ for all $\mathbf{\Phi} \in \mathscr{P}$ such that $(\mathbf{\Phi} - \mathbf{\Phi}_0)\mathbf{R}_{0,\mathrm{LU}} = 0$ and $|\lambda_{q+1}(\mathbf{\Phi})| < |\lambda_q(\mathbf{\Phi}_0)|$;*

(ii) *the first differentials of $A(\cdot)$ and $\Lambda_{\mathrm{LU}}(\cdot)$ at $\mathbf{\Phi} = \mathbf{\Phi}_0$ satisfy*[10]

$$\begin{bmatrix} \mathrm{vec}(\mathrm{d}A) \\ \mathrm{vec}(\mathrm{d}\Lambda_{\mathrm{LU}}) \end{bmatrix} = \begin{bmatrix} J_A(\mathbf{\Phi}_0) \\ J_\Lambda(\mathbf{\Phi}_0) \end{bmatrix} \mathrm{vec}\{(\mathrm{d}\mathbf{\Phi})\mathbf{R}_{0,\mathrm{LU}}\}$$

where

$$J(\mathbf{\Phi}) \coloneqq \begin{bmatrix} J_A(\mathbf{\Phi}) \\ J_\Lambda(\mathbf{\Phi}) \end{bmatrix} \coloneqq \begin{bmatrix} (I_q \otimes \beta^{\mathsf{T}})B \\ [(\Lambda_{\mathrm{LU}}^{\mathsf{T}} \otimes I_q) - (I_q \otimes \Lambda_{\mathrm{LU}})](I_q \otimes G^{\mathsf{T}})B + (I_q \otimes L_{\mathrm{LU}}^{\mathsf{T}}) \end{bmatrix} \tag{B.3}$$

for $G^{\mathsf{T}} \coloneqq [0_{q\times r}, I_q]$, $\beta^{\mathsf{T}} = [I_r, -A]$, and $\Lambda_{\mathrm{LU}} = \Lambda_{\mathrm{LU}}(\mathbf{\Phi})$, etc.; and

(iii) *$J(\mathbf{\Phi})$ is continuous.*

When $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = I_q$, the $pq \times pq$ matrix $J(\mathbf{\Phi})$ simplifies as follows.

[10] For a more compact notation, here and subsequently we express matrix derivatives in terms of differentials, in the manner of Magnus and Neudecker (2007).

Lemma B.3. *Suppose $\mathbf{\Phi} \in \mathscr{P}$ with $\Lambda_{\text{LU}}(\mathbf{\Phi}) = I_q$. Then $J(\mathbf{\Phi})$ is nonsingular, and*

$$\begin{bmatrix} J_A(\mathbf{\Phi}) \\ J_\Lambda(\mathbf{\Phi}) \end{bmatrix} = \begin{bmatrix} I_q \otimes \beta^\mathsf{T} R_{\text{ST}}(I_{kp-q} - \Lambda_{\text{ST}})^{-1} L_{\text{ST}}^\mathsf{T} \\ I_q \otimes L_{\text{LU}}^\mathsf{T} \end{bmatrix}.$$

Proof of Lemma B.1. We first prove $\mathscr{P}$ is open. For $F \in \mathbb{R}^{kp \times kp}$, let $\lambda_i(F)$ denote the ith eigenvalue of F, when these are placed in descending order of modulus. Let $\mathscr{F}$ denote the set of $kp \times kp$ matrices such that

(i) $|\lambda_{q+1}(F)| < |\lambda_q(F)|$; and

there exist $\Lambda_{\text{LU}} \in \mathbb{R}^{q \times q}$ and $\mathbf{R}_{\text{LU}} \in \mathbb{R}^{kp \times q}$ such that

(ii) the eigenvalues of Λ_{LU} are $\{\lambda_i(F)\}_{i=1}^q$, $F\mathbf{R}_{\text{LU}} = \mathbf{R}_{\text{LU}}\Lambda_{\text{LU}}$; and

(iii) $\text{rk}\{\mathbf{G}^\mathsf{T}\mathbf{R}_{\text{LU}}\} = q$, where $\mathbf{G}^\mathsf{T} \coloneqq [0_{q \times (kp-q)}, I_q] = [0_{q \times k(p-1)}, G^\mathsf{T}]$.

In view of Lemma A.1, $\mathbf{\Phi} \in \mathscr{P}$ if and only if the companion form matrix $F(\mathbf{\Phi})$ is in $\mathscr{F}$. Since $F(\cdot)$ is trivially continuous, it suffices to show that $\mathscr{F}$ is open.

To that end, fix $F_0 \in \mathscr{F}$, and let $\mathbf{R}_{0,\text{LU}}$ and $\Lambda_{0,\text{LU}}$ denote matrices satisfying (ii) and (iii) above. By the continuity of eigenvalues and simple invariant subspaces (Theorems IV.1.1 and V.2.8 in Stewart and Sun, 1990), for every $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $\|F - F_0\| < \delta$, F satisfies requirements (i) and (ii) above, with associated $\mathbf{R}_{\text{LU}}$ such that $\|\mathbf{R}_{\text{LU}} - \mathbf{R}_{0,\text{LU}}\| < \epsilon$. Since the set of full rank matrices is open, we may take $\epsilon > 0$ sufficiently small such that (iii) also holds. Thus $F \in \mathscr{F}$, and so F_0 is an interior point of $\mathscr{F}$; deduce $\mathscr{F}$ is open.

We turn next to the smoothness of $A(\mathbf{\Phi})$ and $\Lambda_{\text{LU}}(\mathbf{\Phi})$. For $F_0 \in \mathscr{F}$ we have the invariant subspace decomposition (as per (A.5) above)

$$F_0 = \mathbf{R}_{0,\text{LU}}\Lambda_{0,\text{LU}}\mathbf{L}_{0,\text{LU}}^\mathsf{T} + \mathbf{R}_{0,\text{ST}}\Lambda_{0,\text{ST}}\mathbf{L}_{0,\text{ST}}^\mathsf{T} \tag{B.4}$$

where $\mathbf{R}_{0,\text{LU}}$ and $\Lambda_{0,\text{LU}}$ satisfy (ii)–(iii) above. Since (iii) holds, we may choose $\mathbf{R}_{0,\text{LU}}$ such that $\mathbf{G}^\mathsf{T}\mathbf{R}_{0,\text{LU}} = I_q$; note that $\mathbf{L}_0^\mathsf{T}\mathbf{R}_0 = I_{kp}$ (as per Lemma A.1(iii)) implies $\mathbf{L}_{0,\text{LU}}^\mathsf{T}\mathbf{R}_{0,\text{LU}} = I_q$. Define the maps

$$H(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}}; F) \coloneqq \Big[\mathbf{R}_{\text{LU}}\Lambda_{\text{LU}} - F\mathbf{R}_{\text{LU}}; \quad \mathbf{G}^\mathsf{T}\mathbf{R}_{\text{LU}} - I_q\Big] \tag{B.5a}$$

$$H^*(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}}; F) \coloneqq \Big[\mathbf{R}_{\text{LU}}\Lambda_{\text{LU}} - F\mathbf{R}_{\text{LU}}; \quad \mathbf{L}_{0,\text{LU}}^\mathsf{T}\mathbf{R}_{\text{LU}} - I_q\Big], \tag{B.5b}$$

so that $H(\mathbf{R}_{0,\text{LU}}, \Lambda_{0,\text{LU}}; F_0) = H^*(\mathbf{R}_{0,\text{LU}}, \Lambda_{0,\text{LU}}; F_0) = 0$; but note that these maps need not otherwise agree, since they impose distinct normalisations on $\mathbf{R}_{\text{LU}}$. Once we have shown that the Jacobian of H^* with respect to $(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}})$ is nonsingular at $(\mathbf{R}_{0,\text{LU}}, \Lambda_{0,\text{LU}}; F_0)$, it will follow by the implicit mapping theorem (Lang, 1993, Thm. XIV.2.1) that there exists a neighbourhood $N \subset \mathscr{F}$ of F_0 and smooth functions $\mathbf{R}_{\text{LU}}^* : N \to \mathbb{R}^{kp \times q}$, $\Lambda_{\text{LU}}^* : N \to \mathbb{R}^{q \times q}$ such that

$$H^*[\mathbf{R}_{\text{LU}}^*(F), \Lambda_{\text{LU}}^*(F); F] = 0$$

for all $F \in N$; by the continuity of $\mathbf{R}_{\text{LU}}^*(\cdot)$, we may choose N such that $\text{rk}\{\mathbf{G}^\mathsf{T}\mathbf{R}_{\text{LU}}^*(F)\} = q$ for all $F \in N$. Thus

$$\mathbf{R}_{\text{LU}}(F) \coloneqq \mathbf{R}^*_{\text{LU}}(F)[\mathbf{G}^\top \mathbf{R}^*_{\text{LU}}(F)]^{-1} \tag{B.6}$$

$$\Lambda_{\text{LU}}(F) \coloneqq [\mathbf{G}^\top \mathbf{R}^*_{\text{LU}}(F)]\Lambda^*_{\text{LU}}(F)[\mathbf{G}^\top \mathbf{R}^*_{\text{LU}}(F)]^{-1} \tag{B.7}$$

are well defined for all $F \in N$, and have the property that

$$H[\mathbf{R}_{\text{LU}}(F), \Lambda_{\text{LU}}(F); F] = 0$$

for all $F \in N$. Since the $(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}})$ satisfying $H(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}}; F) = 0$ is unique, repeating this construction for every $F_0 \in \mathscr{F}$ allows the smooth maps $\mathbf{R}_{\text{LU}}(F)$ and $\Lambda_{\text{LU}}(F)$ to be extended to the whole of $\mathscr{F}$. The smoothness of $\Lambda_{\text{LU}}(\boldsymbol{\Phi}) \coloneqq \Lambda_{\text{LU}}[F(\boldsymbol{\Phi})]$ and $\mathbf{R}_{\text{LU}}(\boldsymbol{\Phi}) \coloneqq \mathbf{R}_{\text{LU}}[F(\boldsymbol{\Phi})]$ follows immediately, and that of $A(\boldsymbol{\Phi})$ by noting that it corresponds to rows $(k-1)p+1$ to $(k-1)p+r$ of $\mathbf{R}_{\text{LU}}(\boldsymbol{\Phi})$.

It thus remains to verify that the Jacobian of H^* with respect to $(\mathbf{R}_{\text{LU}}, \Lambda_{\text{LU}})$ is nonsingular at $(\mathbf{R}_{0,\text{LU}}, \Lambda_{0,\text{LU}}; F_0)$. Matrix differentiation gives

$$\mathrm{d}H^* = \left[\mathbf{R}_{0,\text{LU}}(\mathrm{d}\Lambda_{\text{LU}}) + (\mathrm{d}\mathbf{R}_{\text{LU}})\Lambda_{0,\text{LU}} - F_0(\mathrm{d}\mathbf{R}_{\text{LU}}); \quad \mathbf{L}^\top_{0,\text{LU}}(\mathrm{d}\mathbf{R}_{\text{LU}})\right] \eqqcolon \left[\mathrm{d}H^*_1; \quad \mathrm{d}H^*_2\right]$$

The Jacobian is nonsingular if $\mathrm{d}H^* = 0$ implies $\mathrm{d}\mathbf{R}_{\text{LU}} = 0$ and $\mathrm{d}\Lambda_{\text{LU}} = 0$. To that end, suppose $\mathrm{d}H^* = 0$. Then $0 = \mathrm{d}H^*_2 = \mathbf{L}^\top_{0,\text{LU}}(\mathrm{d}\mathbf{R}_{\text{LU}})$, and

$$\mathrm{d}\mathbf{R}_{\text{LU}} = (\mathbf{R}_0\mathbf{L}_0^\top)\mathrm{d}\mathbf{R}_{\text{LU}} = (\mathbf{R}_{0,\text{LU}}\mathbf{L}^\top_{0,\text{LU}} + \mathbf{R}_{0,\text{ST}}\mathbf{L}^\top_{0,\text{ST}})\mathrm{d}\mathbf{R}_{\text{LU}} = (\mathbf{R}_{0,\text{ST}}\mathbf{L}^\top_{0,\text{ST}})\mathrm{d}\mathbf{R}_{\text{LU}}$$

and similarly, by (B.4) above,

$$F_0(\mathrm{d}\mathbf{R}_{\text{LU}}) = (\mathbf{R}_{0,\text{LU}}\Lambda_{0,\text{LU}}\mathbf{L}^\top_{0,\text{LU}} + \mathbf{R}_{0,\text{ST}}\Lambda_{0,\text{ST}}\mathbf{L}^\top_{0,\text{ST}})\mathrm{d}\mathbf{R}_{\text{LU}} = \mathbf{R}_{0,\text{ST}}\Lambda_{0,\text{ST}}\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}}).$$

Hence

$$\begin{aligned}
\mathrm{d}H^*_1 &= \mathbf{R}_{0,\text{LU}}(\mathrm{d}\Lambda_{\text{LU}}) + \mathbf{R}_{0,\text{ST}}[\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}})\Lambda_{0,\text{LU}} - \Lambda_{0,\text{ST}}\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}})] \\
&= \begin{bmatrix}\mathbf{R}_{0,\text{LU}} & \mathbf{R}_{0,\text{ST}}\end{bmatrix}\begin{bmatrix}\mathrm{d}\Lambda_{\text{LU}} \\ \mathcal{T}[\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}})]\end{bmatrix},
\end{aligned}$$

where $\mathcal{T}(M) \coloneqq M\Lambda_{0,\text{LU}} - \Lambda_{0,\text{ST}}M$. Since $\mathbf{R}_0$ is nonsingular, $\mathrm{d}H^*_1 = 0$ implies that $\mathrm{d}\Lambda_{\text{LU}} = 0$ and $\mathcal{T}[\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}})] = 0$; but since $\Lambda_{0,\text{LU}}$ and $\Lambda_{0,\text{ST}}$ have no common eigenvalues, $\mathcal{T}(M) = 0$ if and only if $M = 0$ (Stewart and Sun, 1990, Thm V.1.3). Thus $\mathbf{L}^\top_{0,\text{ST}}(\mathrm{d}\mathbf{R}_{\text{LU}}) = 0$, whence

$$\begin{bmatrix}\mathbf{L}^\top_{0,\text{LU}} \\ \mathbf{L}^\top_{0,\text{ST}}\end{bmatrix}\mathrm{d}\mathbf{R}_{\text{LU}} = 0$$

from which it follows that $\mathrm{d}\mathbf{R}_{\text{LU}} = 0$, since $\mathbf{L}_0$ is nonsingular. □

Proof of Lemma B.2. **(i).** We have

$$R_{0,\mathrm{LU}}\Lambda_{0,\mathrm{LU}}^{k} - \sum_{i=1}^{k} \Phi_i R_{0,\mathrm{LU}}\Lambda_{0,\mathrm{LU}}^{k-i} = \boldsymbol{\Phi}\mathbf{R}_{0,\mathrm{LU}} =_{(1)} \boldsymbol{\Phi}_0\mathbf{R}_{0,\mathrm{LU}} = R_{0,\mathrm{LU}}\Lambda_{0,\mathrm{LU}}^{k} - \sum_{i=1}^{k} \Phi_{0,i} R_{0,\mathrm{LU}}\Lambda_{0,\mathrm{LU}}^{k-i} =_{(2)} 0$$

where $=_{(1)}$ is by hypothesis, and $=_{(2)}$ by Lemma A.1. Since $|\lambda_{q+1}(\boldsymbol{\Phi})| < |\lambda_q(\boldsymbol{\Phi}_0)| = |\lambda_q(\Lambda_{0,\mathrm{LU}})|$ and $\boldsymbol{\Phi} \in \mathscr{P}$, the result then follows by Lemma A.2.

(ii). Analogously to (B.5) above, define

$$\begin{aligned} H(\mathbf{R}_{\mathrm{LU}}, \Lambda_{\mathrm{LU}}; \boldsymbol{\Phi}) &\coloneqq \left[\mathbf{R}_{\mathrm{LU}}\Lambda_{\mathrm{LU}} - F(\boldsymbol{\Phi})\mathbf{R}_{\mathrm{LU}}; \quad \mathbf{G}^{\mathsf{T}}\mathbf{R}_{\mathrm{LU}} - I_q\right] \\ H^*(\mathbf{R}_{\mathrm{LU}}, \Lambda_{\mathrm{LU}}; \boldsymbol{\Phi}) &\coloneqq \left[\mathbf{R}_{\mathrm{LU}}\Lambda_{\mathrm{LU}} - F(\boldsymbol{\Phi})\mathbf{R}_{\mathrm{LU}}; \quad \mathbf{L}_{0,\mathrm{LU}}^{\mathsf{T}}\mathbf{R}_{\mathrm{LU}} - I_q\right]. \end{aligned}$$

By the argument given in the proof of Lemma B.1, there are smooth maps $\mathbf{R}_{\mathrm{LU}}(\boldsymbol{\Phi})$, $\mathbf{R}_{\mathrm{LU}}^*(\boldsymbol{\Phi})$, $\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi})$ and $\Lambda_{\mathrm{LU}}^*(\boldsymbol{\Phi})$ such that $H[\mathbf{R}_{\mathrm{LU}}(\boldsymbol{\Phi}), \Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}); \boldsymbol{\Phi}] = 0$ and $H^*[\mathbf{R}_{\mathrm{LU}}^*(\boldsymbol{\Phi}), \Lambda_{\mathrm{LU}}^*(\boldsymbol{\Phi}); \boldsymbol{\Phi}] = 0$ for all $\boldsymbol{\Phi} \in \mathscr{P}$. Since $G^{\mathsf{T}}R_{0,\mathrm{LU}} = I_q$ implies that $\mathbf{G}^{\mathsf{T}}\mathbf{R}_{0,\mathrm{LU}} = I_q$, we have $\mathbf{R}_{\mathrm{LU}}(\boldsymbol{\Phi}) = \mathbf{R}_{\mathrm{LU}}^*(\boldsymbol{\Phi}) = \mathbf{R}_{0,\mathrm{LU}}$ and $\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}) = \Lambda_{\mathrm{LU}}^*(\boldsymbol{\Phi}) = \Lambda_{0,\mathrm{LU}}$ when $\boldsymbol{\Phi} = \boldsymbol{\Phi}_0$, but otherwise these maps need not agree. Since the maps $\mathbf{R}_{\mathrm{LU}}^*(\boldsymbol{\Phi})$ and $\Lambda_{\mathrm{LU}}^*(\boldsymbol{\Phi})$ are easier to work with, we first obtain the derivatives of these, and subsequently those of $A(\boldsymbol{\Phi})$ and $\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi})$ via renormalisation, analogously to (B.6)–(B.7).

Setting the total differential of H^* to zero gives

$$0 = \mathrm{d}H^* = \left[\mathbf{R}_{0,\mathrm{LU}}(\mathrm{d}\Lambda_{\mathrm{LU}}^*) + (\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*)\Lambda_{0,\mathrm{LU}} - F_0(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*) - F(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}}; \quad \mathbf{L}_{0,\mathrm{LU}}^{\mathsf{T}}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*)\right] \tag{B.8}$$

where $F_0 \coloneqq F(\boldsymbol{\Phi})$, whence by similar arguments as were given in the proof of Lemma B.1,

$$F(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}} = \mathbf{R}_{0,\mathrm{LU}}(\mathrm{d}\Lambda_{\mathrm{LU}}^*) + \mathbf{R}_{0,\mathrm{ST}}\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*)\Lambda_{0,\mathrm{LU}} - \mathbf{R}_{0,\mathrm{ST}}\Lambda_{0,\mathrm{ST}}\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*). \tag{B.9}$$

Vectorising gives

$$\operatorname{vec}[F(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}}] = (I_q \otimes \mathbf{R}_{0,\mathrm{LU}}) \operatorname{vec}(\mathrm{d}\Lambda_{\mathrm{LU}}^*) + M \operatorname{vec}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*) \tag{B.10}$$

for $M \coloneqq (I_q \otimes \mathbf{R}_{0,\mathrm{ST}})[(\Lambda_{0,\mathrm{LU}}^{\mathsf{T}} \otimes I_{kp-q}) - (I_q \otimes \Lambda_{0,\mathrm{ST}})](I_q \otimes \mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}})$. Since $\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}\mathbf{R}_{0,\mathrm{LU}} = 0$ and $\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}\mathbf{R}_{0,\mathrm{ST}} = I_{kp-q}$, setting

$$M^\dagger \coloneqq (I_q \otimes \mathbf{R}_{0,\mathrm{ST}})[(\Lambda_{0,\mathrm{LU}}^{\mathsf{T}} \otimes I_{kp-q}) - (I_q \otimes \Lambda_{0,\mathrm{ST}})]^{-1}(I_q \otimes \mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}})$$

we have $M^\dagger(I_q \otimes \mathbf{R}_{0,\mathrm{LU}}) = 0$ and $M^\dagger M = I_q \otimes \mathbf{R}_{0,\mathrm{ST}}\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}$. Since $\mathbf{L}_{0,\mathrm{LU}}^{\mathsf{T}}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*) = 0$ by (B.8), it follows that

$$\mathrm{d}\mathbf{R}_{\mathrm{LU}}^* = (\mathbf{R}_{0,\mathrm{LU}}\mathbf{L}_{0,\mathrm{LU}}^{\mathsf{T}} + \mathbf{R}_{0,\mathrm{ST}}\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}})\mathrm{d}\mathbf{R}_{\mathrm{LU}}^* = (\mathbf{R}_{0,\mathrm{ST}}\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}})\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*$$

whence $M^\dagger M \operatorname{vec}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*) = \operatorname{vec}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*)$, and so premultiplying (B.10) by $M^\dagger$ gives

$$\operatorname{vec}(\mathrm{d}\mathbf{R}_{\mathrm{LU}}^*) = M^\dagger \operatorname{vec}[F(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}}^*].$$

By the structure of the companion form matrix, $\mathbf{L}_{0,\mathrm{ST}}^{\mathsf{T}}F(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}} = L_{0,\mathrm{ST}}^{\mathsf{T}}(\mathrm{d}\boldsymbol{\Phi})\mathbf{R}_{0,\mathrm{LU}}$. Since R

is given by the final p rows of $\mathbf{R}$, we have

$$\begin{aligned}\operatorname{vec}(\mathrm{d}R^*_{\mathrm{LU}}) &= (I_q \otimes R_{0,\mathrm{ST}})[(\Lambda^{\mathsf{T}}_{0,\mathrm{LU}} \otimes I_{kp-q}) - (I_q \otimes \Lambda_{0,\mathrm{ST}})]^{-1}(I_q \otimes L^{\mathsf{T}}_{0,\mathrm{ST}}) \operatorname{vec}\{(\mathrm{d}\mathbf{\Phi})\mathbf{R}_{0,\mathrm{LU}}\} \\ &= B(\mathbf{\Phi}_0) \operatorname{vec}\{(\mathrm{d}\mathbf{\Phi})\mathbf{R}_{0,\mathrm{LU}}\}. \end{aligned} \tag{B.11}$$

To compute the Jacobian of $A(\mathbf{\Phi})$, note that by partitioning the $p \times p$ identity matrix as

$$\begin{bmatrix} G_\perp & G \end{bmatrix} \coloneqq \begin{bmatrix} I_r & 0 \\ 0 & I_q \end{bmatrix}$$

we have $A(\mathbf{\Phi}) = G_\perp^{\mathsf{T}} R_{\mathrm{LU}}(\mathbf{\Phi}) = G_\perp^{\mathsf{T}} R^*_{\mathrm{LU}}(\mathbf{\Phi})[G^{\mathsf{T}} R^*_{\mathrm{LU}}(\mathbf{\Phi})]^{-1}$. From $R^*_{\mathrm{LU}}(\mathbf{\Phi}_0) = R_{0,\mathrm{LU}}$, $G^{\mathsf{T}} R_{0,\mathrm{LU}} = \mathbf{G}^{\mathsf{T}}\mathbf{R}_{0,\mathrm{LU}} = I_q$ and $G_\perp^{\mathsf{T}} R_{0,\mathrm{LU}} = A_0$, it follows that at $\mathbf{\Phi} = \mathbf{\Phi}_0$

$$\mathrm{d}A = G_\perp^{\mathsf{T}}(\mathrm{d}R^*_{\mathrm{LU}}) - (G_\perp^{\mathsf{T}} R_{0,\mathrm{LU}})G^{\mathsf{T}}(\mathrm{d}R^*_{\mathrm{LU}}) = (G_\perp^{\mathsf{T}} - A_0 G^{\mathsf{T}})\mathrm{d}R^*_{\mathrm{LU}} = \beta_0^{\mathsf{T}} \mathrm{d}R^*_{\mathrm{LU}} \tag{B.12}$$

for $\beta_0^{\mathsf{T}} = [I_r, -A_0]$. The first part of (B.3) follows immediately from (B.11) and (B.12). For the Jacobian of $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$, note that (as per (B.7) above)

$$\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = [\mathbf{G}^{\mathsf{T}}\mathbf{R}^*_{\mathrm{LU}}(\mathbf{\Phi})]\Lambda^*_{\mathrm{LU}}(\mathbf{\Phi})[\mathbf{G}^{\mathsf{T}}\mathbf{R}^*_{\mathrm{LU}}(\mathbf{\Phi})]^{-1}$$

whence at $\mathbf{\Phi} = \mathbf{\Phi}_0$,

$$\mathrm{d}\Lambda_{\mathrm{LU}} = \mathbf{G}^{\mathsf{T}}(\mathrm{d}\mathbf{R}^*_{\mathrm{LU}})\Lambda_{0,\mathrm{LU}} + \mathrm{d}\Lambda^*_{\mathrm{LU}} - \Lambda_{0,\mathrm{LU}}\mathbf{G}^{\mathsf{T}}(\mathrm{d}\mathbf{R}^*_{\mathrm{LU}}).$$

Recognising that $\mathbf{G}^{\mathsf{T}}(\mathrm{d}\mathbf{R}^*_{\mathrm{LU}}) = G^{\mathsf{T}}(\mathrm{d}R^*_{\mathrm{LU}})$ and vectorising, we have

$$\operatorname{vec}(\mathrm{d}\Lambda_{\mathrm{LU}}) = \{(\Lambda^{\mathsf{T}}_{0,\mathrm{LU}} \otimes I_q) - (I_q \otimes \Lambda_{0,\mathrm{LU}})\}(I_q \otimes G^{\mathsf{T}}) \operatorname{vec}(\mathrm{d}R^*_{\mathrm{LU}}) + \operatorname{vec}(\mathrm{d}\Lambda^*_{\mathrm{LU}}). \tag{B.13}$$

$\mathrm{d}R^*_{\mathrm{LU}}$ is given in (B.11) above. To obtain $\mathrm{d}\Lambda^*_{\mathrm{LU}}$, note that premultiplying (B.9) by $\mathbf{L}^{\mathsf{T}}_{0,\mathrm{LU}}$ yields

$$\mathrm{d}\Lambda^*_{\mathrm{LU}} = \mathbf{L}^{\mathsf{T}}_{0,\mathrm{LU}} F(\mathrm{d}\mathbf{\Phi})\mathbf{R}_{0,\mathrm{LU}} = L^{\mathsf{T}}_{0,\mathrm{LU}}(\mathrm{d}\mathbf{\Phi})\mathbf{R}_{0,\mathrm{LU}}. \tag{B.14}$$

Thus (B.11), (B.13) and (B.14) give the second part of (B.3).

(iii). Continuity of $J(\mathbf{\Phi})$ is immediate from $A(\mathbf{\Phi})$ and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi})$ being smooth. □

Proof of Lemma B.3. The stated expression for $J(\mathbf{\Phi})$ is immediate from (B.2), Lemma B.2, and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = I_q$. That $J(\mathbf{\Phi})$ is nonsingular will follow once we have shown that the $(p \times p)$ matrix

$$K \coloneqq \begin{bmatrix} \beta^{\mathsf{T}} R_{\mathrm{ST}}(I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1} L^{\mathsf{T}}_{\mathrm{ST}} \\ L^{\mathsf{T}}_{\mathrm{LU}} \end{bmatrix} \tag{B.15}$$

is nonsingular. We first note the following facts. Since $\mathbf{\Phi} \in \mathscr{P}$ with $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = I_q$, it follows from (B.1) that $\operatorname{rk}\Phi(1) \leq p - q$. Since $\Phi(\cdot)$ has exactly q roots at unity, the reverse inequality holds by Corollary 4.3 of Johansen (1995), whence $\operatorname{rk}\Phi(1) = p - q$. Thus CV holds: this implies that $\operatorname{sp}\beta = \operatorname{sp}\Phi(1)^{\mathsf{T}}$ and $\operatorname{rk} L_{\mathrm{LU}} = q$ (see Lemma A.3 and the characterisation of the CS discussed in Section 2.2).

Now let $c \in \mathbb{R}^p$ be such that $Kc = 0$, so that in particular $L_{\mathrm{LU}}^{\mathsf{T}} c = 0$. Since $\operatorname{rk} \Phi(1) + \operatorname{rk} L_{\mathrm{LU}} = p$, while (2.6) with $\Lambda_{\mathrm{LU}} = I_q$ implies $L_{\mathrm{LU}}^{\mathsf{T}} \Phi(1) = 0$, it follows that $c \in \operatorname{sp} \Phi(1)$, i.e. $c = \Phi(1)b$ for some $b \in \mathbb{R}^p$. By Gohberg et al. (1982, Thm 2.4), $\Phi(\mu)^{-1} = R(\mu I - \Lambda)^{-1} L^{\mathsf{T}}$ for any μ not a root of $\Phi(\cdot)$. Since the columns of the quasi-cointegrating matrix β are orthogonal to R_{LU}, we have

$$\beta^{\mathsf{T}} = \beta^{\mathsf{T}} R_{\mathrm{ST}}(\mu I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1} L_{\mathrm{ST}}^{\mathsf{T}} \Phi(\mu) \to \beta^{\mathsf{T}} R_{\mathrm{ST}}(I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1} L_{\mathrm{ST}}^{\mathsf{T}} \Phi(1) \tag{B.16}$$

by the continuity of the r.h.s., as $\mu \to 1$, since Λ_{ST} has no eigenvalues at unity. Hence

$$0 = Kc = \begin{bmatrix} \beta^{\mathsf{T}} R_{\mathrm{ST}}(I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1} L_{\mathrm{ST}}^{\mathsf{T}} \Phi(1) b \\ 0 \end{bmatrix} = \begin{bmatrix} \beta^{\mathsf{T}} b \\ 0 \end{bmatrix}$$

implying $\beta^{\mathsf{T}} b = 0$. But $\operatorname{sp} \beta = \operatorname{sp} \Phi(1)^{\mathsf{T}}$, so we must have $\Phi(1)b = 0$. Thus $c = 0$, from which it follows that K is nonsingular. □

C Asymptotics

The assumptions DGP and LOC are maintained throughout this appendix. We first recall some notation. Let $\mathbf{\Phi}_0 := \lim_{n\to\infty} \mathbf{\Phi}_n$, where $\{\mathbf{\Phi}_n\}$ is the sequence specified by LOC. Let $R_n := [R_{\mathrm{LU}}(\mathbf{\Phi}_n), R_{\mathrm{ST}}]$ and $\Lambda_n := \operatorname{diag}\{\Lambda_{n,\mathrm{LU}}, \Lambda_{\mathrm{ST}}\}$ be as in LOC. Take $\mathbf{R}_n := \operatorname{col}\{R_n \Lambda_n^{k-i}\}_{i=1}^k$ and $\mathbf{L}_n := (\mathbf{R}_n^{\mathsf{T}})^{-1}$ as in Lemma A.1, and partition these as $\mathbf{R}_n = [\mathbf{R}_{n,\mathrm{LU}}, \mathbf{R}_{n,\mathrm{ST}}]$ and $\mathbf{L}_n = [\mathbf{L}_{n,\mathrm{LU}}, \mathbf{L}_{n,\mathrm{ST}}]$ (as per (A.4)); note that both these matrices are convergent.

Let $z_{\mathrm{LU},t} := \mathbf{L}_{n,\mathrm{LU}}^{\mathsf{T}} \mathbf{x}_t$ and $z_{\mathrm{ST},t} = \mathbf{L}_{n,\mathrm{ST}}^{\mathsf{T}} \mathbf{x}_t$ be as in Lemma A.4 (for $\mathbf{\Phi} = \mathbf{\Phi}_n$); these follow the autoregressions given in (2.10). Recall $E \sim \mathrm{BM}(\Sigma)$ and $Z_C(r) := \int_0^r e^{C(r-s)} L_{\mathrm{LU}}^{\mathsf{T}} \mathrm{d}E(s)$ from (3.6). For $i \in \{\mathrm{LU}, \mathrm{ST}\}$, let $\bar{z}_{i,t}$ denote the residual from an OLS regression of $\{\bar{z}_{\mathrm{LU},t-1}\}_{t=1}^n$ onto a constant and linear trend. Recall that $\bar{Z}_C$ denotes the residual from an $L^2[0,1]$ projection of each sample path of Z_C onto a constant and linear trend. As in Section 3.1.1, let $\hat{\Sigma}_n$ denote the unrestricted MLE for Σ, i.e. the OLS residual variance matrix estimator.

Proofs of the following results appear at the end of this section.

Lemma C.1. *The following hold jointly:*

(i) $n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_t \rightsquigarrow E(r)$

(ii) $n^{-1/2} z_{\mathrm{LU},\lfloor nr \rfloor} \rightsquigarrow Z_C(r)$

(iii) $n^{-1/2} \bar{z}_{\mathrm{LU},\lfloor nr \rfloor} \rightsquigarrow \bar{Z}_C(r)$

as weak convergences on the space of right-continuous functions $[0,1] \to \mathbb{R}^m$ *(with respect to the uniform topology); and*

(iv) $n^{-1} \sum_{t=1}^n (\bar{z}_{\mathrm{LU},t-1} \otimes \varepsilon_t) \rightsquigarrow \int_0^1 [\bar{Z}_C(r) \otimes \mathrm{d}E(r)]\,\mathrm{d}r$

(v) $n^{-1/2} \sum_{t=1}^n (\bar{z}_{\mathrm{ST},t-1} \otimes \varepsilon_t) \rightsquigarrow \xi \sim \mathrm{N}[0, \Omega \otimes \Sigma]$

(vi) $\hat{\Sigma}_n \xrightarrow{p} \Sigma$,

where $\Omega := \lim_{n\to\infty} \operatorname{var}(z_{\mathrm{ST},n})$ *is positive definite, and* ξ *is independent of* E.

Now define the reparametrisation $\mathbf{\Phi} \mapsto \varphi$ by

$$\varphi := \begin{bmatrix} \varphi_{\mathrm{LU}} \\ \varphi_{\mathrm{ST}} \end{bmatrix} = \begin{bmatrix} \operatorname{vec}\{(\mathbf{\Phi} - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}\} \\ \operatorname{vec}\{(\mathbf{\Phi} - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{ST}}\} \end{bmatrix} = \operatorname{vec}\{(\mathbf{\Phi} - \mathbf{\Phi}_n)\mathbf{R}_n\}, \tag{C.1}$$

which is reversed by setting $\mathbf{\Phi} = \mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}$, where $\operatorname{vec}^{-1}(x)$ maps $x \in \mathbb{R}^{kp^2}$ to the matrix $X \in \mathbb{R}^{p \times kp}$ for which $\operatorname{vec}(X) = x$. The parameter space for φ is the open set

$$\mathcal{P}_n := \{\operatorname{vec}[(\mathbf{\Phi} - \mathbf{\Phi}_n)\mathbf{R}_n] \mid \mathbf{\Phi} \in \mathscr{P}\}, \tag{C.2}$$

and the true parameters correspond to $\varphi = 0$. Although $\mathcal{P}_n$ depends on n, since $\mathbf{\Phi}_n \to \mathbf{\Phi}_0 \in \mathscr{P}$ and $\mathscr{P}$ is open (Lemma B.1), there is an $\epsilon > 0$ such that $\mathcal{P}_n$ contains a ball of radius ϵ centred at the origin, for all n sufficiently large. Let

$$\ell_n^*(\varphi) := \ell_n[\mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}, \hat{\Sigma}_n].$$

Define $D_n := \operatorname{diag}\{nI_{\#\mathrm{LU}}, n^{1/2}I_{\#\mathrm{ST}}\}$, where $\#\mathrm{LU} := pq$ and $\#\mathrm{ST} := p(kp - q)$ correspond to the dimensions of the vectors φ_{LU} and φ_{ST} respectively.

Lemma C.2. *There exist S_n and H_n such that for all $\varphi \in \mathcal{P}_n$,*

$$\ell_n^*(\varphi) - \ell^*(0) = S_n^{\mathsf{T}}(D_n\varphi) - \tfrac{1}{2}(D_n\varphi)^{\mathsf{T}}H_n(D_n\varphi)$$

where

$$S_n \rightsquigarrow \begin{bmatrix} \int_0^1 [\bar{Z}_C(r) \otimes \Sigma^{-1}\mathrm{d}E(r)] \\ \xi \end{bmatrix} =: \begin{bmatrix} S_{\mathrm{LU}} \\ S_{\mathrm{ST}} \end{bmatrix} =: S$$
$$H_n \rightsquigarrow \begin{bmatrix} \int \bar{Z}_C\bar{Z}_C^{\mathsf{T}} & 0 \\ 0 & \Omega \end{bmatrix} \otimes \Sigma^{-1} =: \begin{bmatrix} H_{\mathrm{LU}} & 0 \\ 0 & H_{\mathrm{ST}} \end{bmatrix} =: H,$$

for ξ as in Lemma C.1.

Define the constraint maps

$$\begin{aligned} \theta_n(\varphi) &:= \operatorname{vec}\{\Lambda_{\mathrm{LU}}[\mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}] - (I_q + C/n)\} \\ \gamma_n(\varphi) &:= a_{ij}[\mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}] - a_{ij}(\mathbf{\Phi}_n), \end{aligned} \tag{C.3}$$

and the associated restricted parameter spaces

$$\begin{aligned} \mathcal{P}_{n|\theta} &:= \{\varphi \in \mathcal{P}_n \mid \theta_n(\varphi) = 0\} \\ \mathcal{P}_{n|\theta,\gamma} &:= \{\varphi \in \mathcal{P}_n \mid \theta_n(\varphi) = 0 \text{ and } \gamma_n(\varphi) = 0\}. \end{aligned}$$

Let $\hat{\varphi}_n$, $\hat{\varphi}_{n|\theta}$ and $\hat{\varphi}_{n|\theta,\gamma}$ denote exact maximisers of $\ell_n^*(\varphi)$ over the sets $\mathcal{P}_n$, $\mathcal{P}_{n|\theta}$ and $\mathcal{P}_{n|\theta,\gamma}$ respectively: which may be shown to exist at least with with probability approaching one (w.p.a.1), and may be arbitrarily defined otherwise.

Lemma C.3. *Each of $D_n\hat{\varphi}_n$, $D_n\hat{\varphi}_{n|\theta}$ and $D_n\hat{\varphi}_{n|\theta,\gamma}$ are $O_p(1)$.*

Let $\nabla_\varphi g(\varphi_0)$ denote the gradient of $g : \mathcal{P} \to \mathbb{R}^{d_g}$ at $\varphi = \varphi_0$. The derivatives of the maps θ_n and γ_n can be inferred from Lemma B.2. Part (ii) of that result gives the derivatives with respect to φ_{LU}, and part (i) implies that when $\varphi_{\text{LU}} = 0$, the first (and higher order) derivatives with respect to φ_{ST} are identically zero. Now letting $e_{d,i} \in \mathbb{R}^d$ denote a vector with zero everywhere except for a 1 in the ith position, define

$$\Pi \coloneqq [\Theta; \quad \Gamma] \coloneqq [I_q \otimes L_{\text{LU}}; \quad e_{q,j} \otimes L_{\text{ST}}(I_{kp-q} - \Lambda_{\text{ST}}^{\mathsf{T}})^{-1} R_{\text{ST}}^{\mathsf{T}} \beta e_{r,i}],$$

which by Lemma B.3 has full column rank, and

$$\mathbf{\Theta} \coloneqq \begin{bmatrix} \Theta \\ 0_{\#\text{ST} \times q^2} \end{bmatrix} \qquad\qquad \mathbf{\Pi} \coloneqq \begin{bmatrix} \Pi \\ 0_{\#\text{ST} \times (q^2+1)} \end{bmatrix}.$$

Lemma C.4.

(i) *Let $\{\tilde{\varphi}_n\}$ denote a random sequence in $\mathcal{P}_n$ with $\tilde{\varphi}_n \xrightarrow{p} 0$. Then*

$$\nabla_\varphi \theta_n(\tilde{\varphi}_n) \xrightarrow{p} \mathbf{\Theta} \qquad\qquad \nabla_\varphi \gamma_n(\tilde{\varphi}_n) \xrightarrow{p} \mathbf{\Gamma}.$$

(ii) *Let $\mathcal{Q}_{\mathbf{\Theta},\perp}$ and $\mathcal{Q}_{\mathbf{\Pi},\perp}$ denote orthogonal projections from $\mathbb{R}^{kp^2}$ onto the subspaces orthogonal to the the columns of $\mathbf{\Theta}$ and $\mathbf{\Pi}$. Then*

$$\begin{aligned} D_n \hat{\varphi}_{n|\theta} &= \mathcal{Q}_{\mathbf{\Theta},\perp} D_n \hat{\varphi}_{n|\theta} + o_p(1) \\ D_n \hat{\varphi}_{n|\theta,\gamma} &= \mathcal{Q}_{\mathbf{\Pi},\perp} D_n \hat{\varphi}_{n|\theta,\gamma} + o_p(1). \end{aligned}$$

Let $\Theta_\perp \in \mathbb{R}^{pq \times qr}$ and $\Pi_\perp \in \mathbb{R}^{pq \times (qr-1)}$ denote matrices having full column rank, such that $\Theta_\perp^{\mathsf{T}} \Theta = 0$ and $\Pi_\perp^{\mathsf{T}} \Pi = 0$. We may take $\Theta_\perp = I_q \otimes L_{\text{LU},\perp}$, for $L_{\text{LU},\perp}$ a $p \times r$ matrix having rank r and for which $L_{\text{LU},\perp}^{\mathsf{T}} L_{\text{LU}} = 0$. Since $\Pi = [\Theta, \Gamma]$ there exists a full column rank matrix $\Xi \in \mathbb{R}^{qr \times (qr-1)}$ for which $\Pi_\perp \coloneqq \Theta_\perp \Xi$.

Proposition C.1.

(i) $D_n \hat{\varphi}_n = \begin{bmatrix} n\hat{\varphi}_{n,\text{LU}} \\ n^{1/2}\hat{\varphi}_{n,\text{ST}} \end{bmatrix} \rightsquigarrow \begin{bmatrix} H_{\text{LU}}^{-1} S_{\text{LU}} \\ H_{\text{ST}}^{-1} S_{\text{ST}} \end{bmatrix}$,

(ii) $D_n \hat{\varphi}_{n|\theta} = \begin{bmatrix} n\hat{\varphi}_{n,\text{LU}|\theta} \\ n^{1/2}\hat{\varphi}_{n,\text{ST}|\theta} \end{bmatrix} \rightsquigarrow \begin{bmatrix} \Theta_\perp(\Theta_\perp^{\mathsf{T}} H_{\text{LU}} \Theta_\perp)^{-1} \Theta_\perp^{\mathsf{T}} S_{\text{LU}} \\ H_{\text{ST}}^{-1} S_{\text{ST}} \end{bmatrix}$,

(iii) $2[\ell_n^*(\hat{\varphi}_n) - \ell_n^*(\hat{\varphi}_{n|\theta})] \rightsquigarrow S_{\text{LU}}^{\mathsf{T}} H_{\text{LU}}^{-1} \Theta(\Theta^{\mathsf{T}} H_{\text{LU}}^{-1} \Theta)^{-1} \Theta^{\mathsf{T}} H_{\text{LU}}^{-1} S_{\text{LU}}$.

Let $H_{\Theta,\perp} \coloneqq \Theta_\perp^{\mathsf{T}} H_{\text{LU}} \Theta_\perp$, and $\mathcal{Q} \in \mathbb{R}^{qr \times qr}$ denote the orthogonal projection onto $\operatorname{sp} H_{\Theta,\perp}^{1/2} \Xi$. Then

(iv) $2[\ell_n^*(\hat{\varphi}_{n|\theta}) - \ell_n^*(\hat{\varphi}_{n|\theta,\gamma})] \rightsquigarrow (H_{\Theta,\perp}^{-1/2} \Theta_\perp^{\mathsf{T}} S_{\text{LU}})^{\mathsf{T}} [I_{qr} - \mathcal{Q}] (H_{\Theta,\perp}^{-1/2} \Theta_\perp^{\mathsf{T}} S_{\text{LU}})$.

The preceding gives the limiting distribution of $\hat{\mathbf{\Phi}}_n$ under the reparametrisation (C.1); the limiting distributions of estimators of A and Λ_{LU} will then follow by an application of the delta method, as per

Proposition C.2. *Let $\{\mathbf{\Phi}_n\}$ be as in* LOC, $\mathbf{\Phi}_0 \coloneqq \lim_{n\to\infty}\mathbf{\Phi}_n \in \mathscr{P}$, *and $\{\tilde{\mathbf{\Phi}}_n\}$ a random sequence in $\mathscr{P}$ with $\tilde{\mathbf{\Phi}}_n = \mathbf{\Phi}_n + o_p(1)$. Then*

$$\begin{bmatrix} \operatorname{vec}\{A(\tilde{\mathbf{\Phi}}_n) - A(\mathbf{\Phi}_n)\} \\ \operatorname{vec}\{\Lambda_{\text{LU}}(\tilde{\mathbf{\Phi}}_n) - \Lambda_{\text{LU}}(\mathbf{\Phi}_n)\} \end{bmatrix} = \left(\begin{bmatrix} J_A(\mathbf{\Phi}_0) \\ J_\Lambda(\mathbf{\Phi}_0) \end{bmatrix} + o_p(1) \right) \operatorname{vec}\{(\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n)\mathbf{R}_{n,\text{LU}}\} \tag{C.4}$$

where $\mathbf{R}_{n,\text{LU}} \coloneqq \mathbf{R}_{\text{LU}}(\mathbf{\Phi}_n)$.

Proof of Lemma C.1. (i)–(iv) follow by Donsker's theorem for partial sums, Lemma 3.1 in Phillips (1988) and the continuous mapping theorem; (v) by the martingale central limit theorem (Hall and Heyde, 1980, Thm. 3.2); and (vi) by arguments similar to those given in Section 3.2.2 of Lütkepohl (2007). □

Proof of Lemma C.2. Let $\Phi_i \coloneqq \mathbf{\Phi}\mathbf{R}_{n,i}$ and $\Phi_{n,i} \coloneqq \mathbf{\Phi}_n\mathbf{R}_{n,i}$ for $i \in \{\text{LU}, \text{ST}\}$. Then

$$\begin{aligned} \ell_n(\mathbf{\Phi}, \Sigma) &= -\frac{n}{2}\log\det\Sigma - \min_{m,d}\frac{1}{2}\sum_{t=1}^n \|y_t - m - dt - \mathbf{\Phi}\mathbf{y}_{t-1}\|^2_{\Sigma^{-1}} \\ &= -\frac{n}{2}\log\det\Sigma - \min_{m,d}\frac{1}{2}\sum_{t=1}^n \|x_t - m - dt - \mathbf{\Phi}\mathbf{x}_{t-1}\|^2_{\Sigma^{-1}} \\ &= -\frac{n}{2}\log\det\Sigma - \min_{m,d}\frac{1}{2}\sum_{t=1}^n \|x_t - m - dt - \Phi_{\text{LU}}z_{\text{LU},t-1} - \Phi_{\text{ST}}z_{\text{ST},t-1}\|^2_{\Sigma^{-1}} \\ &= -\frac{n}{2}\log\det\Sigma - \frac{1}{2}\sum_{t=1}^n \|\bar{x}_t - \Phi_{\text{LU}}\bar{z}_{\text{LU},t-1} - \Phi_{\text{ST}}\bar{z}_{\text{ST},t-1}\|^2_{\Sigma^{-1}} \end{aligned}$$

Twice differentiating the r.h.s. (as in Lütkepohl 2007, Sec. 3.4) with respect to Φ_{LU} and Φ_{ST}, and noting that $\varphi_i = \operatorname{vec}(\Phi_i - \Phi_{n,i})$, we thus have

$$\ell_n^*(\varphi) - \ell_n^*(0) = \ell_n(\mathbf{\Phi}, \hat{\Sigma}_n) - \ell_n(\mathbf{\Phi}_n, \hat{\Sigma}_n) = S_n^\mathsf{T}(D_n\varphi) - \tfrac{1}{2}(D_n\varphi)^\mathsf{T}H_n(D_n\varphi)$$

where

$$\begin{aligned} S_n &\coloneqq \begin{bmatrix} n^{-1}\sum_{t=1}^n(\bar{z}_{\text{LU},t-1}\otimes\hat{\Sigma}_n^{-1}\bar{\varepsilon}_t) \\ n^{-1/2}\sum_{t=1}^n(\bar{z}_{\text{ST},t-1}\otimes\hat{\Sigma}_n^{-1}\bar{\varepsilon}_t) \end{bmatrix} =_{(1)} \begin{bmatrix} \frac{1}{n}\sum_{t=1}^n(\bar{z}_{\text{LU},t-1}\otimes\hat{\Sigma}_n^{-1}\varepsilon_t) \\ \frac{1}{n^{1/2}}\sum_{t=1}^n(\bar{z}_{\text{ST},t-1}\otimes\hat{\Sigma}_n^{-1}\varepsilon_t) \end{bmatrix} \\ H_n &\coloneqq \begin{bmatrix} n^{-2}\sum_{t=1}^n \bar{z}_{\text{LU},t-1}\bar{z}_{\text{LU},t-1}^\mathsf{T} & n^{-3/2}\sum_{t=1}^n \bar{z}_{\text{LU},t-1}\bar{z}_{\text{ST},t-1}^\mathsf{T} \\ n^{-3/2}\sum_{t=1}^n \bar{z}_{\text{ST},t-1}\bar{z}_{\text{LU},t-1}^\mathsf{T} & n^{-1}\sum_{t=1}^n \bar{z}_{\text{ST},t-1}\bar{z}_{\text{ST},t-1}^\mathsf{T} \end{bmatrix} \otimes \hat{\Sigma}_n^{-1}, \end{aligned}$$

and $\bar{\varepsilon}_t$ denotes the residual from an OLS regression of $\{\varepsilon_t\}_{t=1}^n$ on a constant and a linear trend; $=_{(1)}$ holds because each element of $\bar{z}_{\text{LU},t-1}$ and $\bar{z}_{\text{ST},t-1}$ is orthogonal to a constant and linear trend. The stated convergences of S_n and H_n then follow by Lemma C.1 and the continuous mapping theorem. □

Proof of Lemma C.3. By Lemma C.2, we have

$$\ell_n^*(\varphi) - \ell_n^*(0) \le \|D_n\varphi\|[\|S_n\| - \tfrac{1}{2}\lambda_{\min}(H_n)\|D_n\varphi\|].$$

Let $M < \infty$ and $\epsilon > 0$. Since $D_n = \operatorname{diag}\{nI_{\#\text{LU}}, n^{1/2}I_{\#\text{ST}}\}$, $S_n = O_p(1)$ and $H_n \rightsquigarrow H$ is positive definite w.p.a.1, it is evident that

$$\mathbb{P}\left\{\sup_{\{\varphi\in\mathcal{P}_n \mid \|D_n\varphi\|\geq M\}}[\ell_n^*(\varphi)-\ell_n^*(0)]<-\epsilon\right\}\geq\mathbb{P}\left\{M[\|S_n\|-\tfrac{1}{2}\lambda_{\min}(H_n)M]<-\epsilon\right\}$$

and

$$\limsup_{n\to\infty}\mathbb{P}\left\{M[\|S_n\|-\tfrac{1}{2}\lambda_{\min}(H_n)M]<-\epsilon\right\}\geq\mathbb{P}\left\{M[\|S\|-\tfrac{1}{2}\lambda_{\min}(H)M]<-\epsilon\right\}\to 1$$

as $M\to\infty$. Deduce that $D_n\hat{\varphi}_n=O_p(1)$. Since $\mathcal{P}_{n|\theta,\gamma}\subset\mathcal{P}_{n|\theta}\subset\mathcal{P}_n$ and $0\in\mathcal{P}_{n|\theta,\gamma}$, that $D_n\hat{\varphi}_{n|\theta}$ and $D_n\hat{\varphi}_{n|\theta,\gamma}$ are stochastically bounded follows by the same argument. □

Proof of Lemma C.4. **(i).** Since $\mathbf{\Phi}_n\to\mathbf{\Phi}_0$, $\mathbf{L}_n\to\mathbf{L}_0$ and $\Lambda_{\text{LU}}(\cdot)$ is continuously differentiable (Lemma B.1),

$$\nabla_\varphi\theta_n(\tilde{\varphi}_n)\xrightarrow{p}\nabla_\varphi\operatorname{vec}\{\Lambda_{\text{LU}}[\mathbf{\Phi}_0+\operatorname{vec}^{-1}(\varphi)\mathbf{L}_0^{\mathsf{T}}]\}|_{\varphi=0}=_{(1)}\begin{bmatrix}I_q\otimes L_{\text{LU}}\\0_{\#\text{ST}\times q^2}\end{bmatrix}=\mathbf{\Theta}$$

where $=_{(1)}$ follows by Lemmas B.2 and B.3. The probability limit of $\nabla_\varphi\gamma_n(\tilde{\varphi}_n)$ follows similarly.

(ii). By Lemma C.3 and the remarks following (C.2), there exists a ball $B(0,\epsilon)$ of radius $\epsilon>0$, centred on the origin, such that $B(0,\epsilon)\subset\mathcal{P}_n$ for all n sufficiently large, and $\mathbb{P}\{\hat{\varphi}_{n|\theta}\in B(0,\epsilon)\}\to 1$. We may take ϵ sufficiently small that $\mathbf{\Phi}_\varphi:=\mathbf{\Phi}_n+\operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}$ has $|\lambda_{q+1}(\mathbf{\Phi}_\varphi)|<|\lambda_q(\mathbf{\Phi}_n)|$ for all n sufficiently large, for all $\varphi\in B(0,\epsilon)$. In particular, suppose $\varphi_{\text{LU}}=0$; then $(\mathbf{\Phi}_\varphi-\mathbf{\Phi}_n)\mathbf{R}_{n,\text{LU}}=0$ and we have by Lemma B.2(i) that $\Lambda_{\text{LU}}(\mathbf{\Phi}_\varphi)=\Lambda_{\text{LU}}(\mathbf{\Phi}_n)=C/n$. It follows that $\theta_n(0,\hat{\varphi}_{n,\text{ST}|\theta})=0$ w.p.a.1., whence

$$\begin{aligned}0=\theta_n(\hat{\varphi}_{n,\text{LU}|\theta},\hat{\varphi}_{n,\text{ST}|\theta})&=\theta_n(\hat{\varphi}_{n,\text{LU}|\theta},\hat{\varphi}_{n,\text{ST}|\theta})-\theta_n(0,\hat{\varphi}_{n,\text{ST}|\theta})\\&=[\Theta+o_p(1)]^{\mathsf{T}}\hat{\varphi}_{n,\text{LU}|\theta}=\Theta^{\mathsf{T}}\hat{\varphi}_{n,\text{LU}|\theta}+o_p(\|\hat{\varphi}_{n,\text{LU}|\theta}\|)\end{aligned}$$

by part (i) of the lemma and a mean value expansion. Hence, letting $\mathcal{Q}_\Theta$ and $\mathcal{Q}_{\Theta,\perp}$ denote the matrices that orthogonally project from $\mathbb{R}^{\#\text{LU}}$ onto $\operatorname{sp}\Theta$ and $(\operatorname{sp}\Theta)^\perp$ respectively, we have

$$\begin{aligned}D_n\hat{\varphi}_{n|\theta}&=\begin{bmatrix}nI_{\#\text{LU}}&0\\0&n^{1/2}I_{\#\text{ST}}\end{bmatrix}\begin{bmatrix}\mathcal{Q}_\Theta+\mathcal{Q}_{\Theta,\perp}&0\\0&I_{\#\text{ST}}\end{bmatrix}\begin{bmatrix}\hat{\varphi}_{n,\text{LU}|\theta}\\\hat{\varphi}_{n,\text{ST}|\theta}\end{bmatrix}\\&=\begin{bmatrix}\mathcal{Q}_{\Theta,\perp}&0\\0&I_{\#\text{ST}}\end{bmatrix}\begin{bmatrix}n\hat{\varphi}_{n,\text{LU}|\theta}\\n^{1/2}\hat{\varphi}_{n,\text{ST}|\theta}\end{bmatrix}+o_p(n\|\hat{\varphi}_{n,\text{LU}|\theta}\|)=\mathcal{Q}_{\mathbf{\Theta},\perp}D_n\hat{\varphi}_{n|\theta}+o_p(\|D_n\hat{\varphi}_{n|\theta}\|).\end{aligned}$$ □

Proof of Proposition C.1. **(i).** Immediate from Lemma C.2.

(ii). As in the proof of Lemma C.4(ii), we may take $\epsilon>0$ such that $B(0,\epsilon)\subset\mathcal{P}_n$ for all n sufficiently large, and $\mathbb{P}\{\hat{\varphi}_{n|\theta}\in B(0,\epsilon)\}\to 1$. Hence w.p.a.1., $\hat{\varphi}_{n|\theta}$ satisfies the first-order conditions for a constrained interior maximum,

$$\nabla_\varphi\ell_n^*(\hat{\varphi}_{n|\theta})=D_nS_n-D_nH_n(D_n\hat{\varphi}_{n|\theta})=\nabla_\varphi\theta_n(\hat{\varphi}_{n|\theta})\mu_n,$$

where $\mu_n \in \mathbb{R}^{q^2}$ is a vector of Lagrange multipliers; whence

$$S_n - H_n(D_n\hat{\varphi}_{n|\theta}) = (nD_n^{-1})\nabla_\varphi \theta_n(\hat{\varphi}_{n|\theta})(n^{-1}\mu_n) =: \mathbf{\Theta}_n(n^{-1}\mu_n) \tag{C.5}$$

w.p.a.1. By a similar argument as given in the proof of Lemma C.4(ii), it follows from Lemma B.2(i) that $\nabla_{\varphi_{\mathrm{ST}}}\theta_n(0, \hat{\varphi}_{n,\mathrm{ST}|\theta}) = 0$ w.p.a.1, and so by a a mean value expansion and Lemma C.3,

$$\nabla_{\varphi_{\mathrm{ST}}}\theta_n(\hat{\varphi}_{n|\theta}) = \nabla_{\varphi_{\mathrm{ST}}}\theta_n(\hat{\varphi}_{n,\mathrm{LU}|\theta}, \hat{\varphi}_{n,\mathrm{ST}|\theta}) - \nabla_{\varphi_{\mathrm{ST}}}\theta_n(0, \hat{\varphi}_{n,\mathrm{ST}|\theta}) = O_p(\|\hat{\varphi}_{n,\mathrm{LU}|\theta}\|) = O_p(n^{-1}).$$

Deduce from the preceding and Lemma C.4(i) that

$$\mathbf{\Theta}_n = (nD_n^{-1})\nabla_\varphi \theta_n(\hat{\varphi}_{n|\theta}) = \begin{bmatrix} \nabla_{\varphi_{\mathrm{LU}}}\theta_n(\hat{\varphi}_{n|\theta}) \\ n^{1/2}\nabla_{\varphi_{\mathrm{ST}}}\theta_n(\hat{\varphi}_{n|\theta}) \end{bmatrix} \xrightarrow{p} \mathbf{\Theta},$$

which has full column rank. Let $\mathbf{\Theta}_\perp := \operatorname{diag}\{\Theta_\perp, I_{\#\mathrm{ST}}\}$, a full column rank matrix for which $\mathbf{\Theta}_\perp^\mathsf{T}\mathbf{\Theta} = 0$; then $\mathbf{\Theta}_{n,\perp} := [I_{kp^2} - \mathbf{\Theta}_n(\mathbf{\Theta}_n^\mathsf{T}\mathbf{\Theta}_n)^{-1}\mathbf{\Theta}_n^\mathsf{T}]\mathbf{\Theta}_\perp \xrightarrow{p} \mathbf{\Theta}_\perp$ and $\mathbf{\Theta}_{n,\perp}^\mathsf{T}\mathbf{\Theta}_n = 0$ for all n. Hence w.p.a.1

$$\begin{aligned} 0 &=_{(1)} \mathbf{\Theta}_{n,\perp}^\mathsf{T} S_n - \mathbf{\Theta}_{n,\perp}^\mathsf{T} H_n(D_n\hat{\varphi}_{n|\theta}) \\ &=_{(2)} \mathbf{\Theta}_{n,\perp}^\mathsf{T} S_n - \mathbf{\Theta}_{n,\perp}^\mathsf{T} H_n[\mathbf{\Theta}_\perp(\mathbf{\Theta}_\perp^\mathsf{T}\mathbf{\Theta}_\perp)^{-1}\mathbf{\Theta}_\perp^\mathsf{T}(D_n\hat{\varphi}_{n|\theta}) + o_p(\|D_n\hat{\varphi}_{n|\theta}\|)] \end{aligned}$$

where $=_{(1)}$ follows from premultiplying (C.5) by $\mathbf{\Theta}_{n,\perp}^\mathsf{T}$, and $=_{(2)}$ from Lemma C.4(ii). A further appeal to that result and rearranging the preceding yields

$$D_n\hat{\varphi}_{n|\theta} = \mathcal{Q}_{\mathbf{\Theta},\perp}D_n\hat{\varphi}_{n|\theta} + o_p(\|D_n\hat{\varphi}_{n|\theta}\|) = \mathbf{\Theta}_\perp(\mathbf{\Theta}_{n,\perp}^\mathsf{T} H_n\mathbf{\Theta}_\perp)^{-1}\mathbf{\Theta}_{n,\perp}^\mathsf{T} S_n + o_p(1 + \|D_n\hat{\varphi}_{n|\theta}\|).$$

The result then follows by Lemmas C.2 and C.3.

(iii). From parts (i) and (ii) and Lemma C.2 we have

$$2[\ell_n^*(\hat{\varphi}_n) - \ell_n^*(0)] \rightsquigarrow S_{\mathrm{LU}}^\mathsf{T} H_{\mathrm{LU}}^{-1} S_{\mathrm{LU}} + S_{\mathrm{ST}}^\mathsf{T} H_{\mathrm{ST}}^{-1} S_{\mathrm{ST}} \tag{C.6}$$

$$2[\ell_n^*(\hat{\varphi}_{n|\theta}) - \ell_n^*(0)] \rightsquigarrow S_{\mathrm{LU}}^\mathsf{T}\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}} + S_{\mathrm{ST}}^\mathsf{T} H_{\mathrm{ST}}^{-1} S_{\mathrm{ST}} \tag{C.7}$$

whence the result follows by subtracting (C.7) from (C.6) and noting that

$$H_{\mathrm{LU}}^{-1/2}\Theta(\Theta^\mathsf{T} H_{\mathrm{LU}}^{-1}\Theta)^{-1}\Theta^\mathsf{T} H_{\mathrm{LU}}^{-1/2} + H_{\mathrm{LU}}^{1/2}\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1}\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}^{1/2} = I_{pq}$$

since the columns of $H_{\mathrm{LU}}^{-1/2}\Theta$ and $H_{\mathrm{LU}}^{1/2}\Theta_\perp$ are mutually orthogonal, and collectively span the whole of $\mathbb{R}^{pq}$.

(iv). The same argument as which yielded (C.7) also gives

$$2[\ell_n^*(\hat{\varphi}_{n|\theta,\gamma}) - \ell_n^*(0)] \rightsquigarrow S_{\mathrm{LU}}^\mathsf{T}\Pi_\perp(\Pi_\perp^\mathsf{T} H_{\mathrm{LU}}\Pi_\perp)^{-1}\Pi_\perp^\mathsf{T} S_{\mathrm{LU}} + S_{\mathrm{ST}}^\mathsf{T} H_{\mathrm{ST}}^{-1} S_{\mathrm{ST}} \tag{C.8}$$

so that subtracting (C.8) from (C.7), and recalling $\Pi_\perp = \Theta_\perp\Xi$, yields

$$2[\ell_n^*(\hat{\varphi}_{n|\theta}) - \ell_n^*(\hat{\varphi}_{n|\theta,\gamma})] \rightsquigarrow S_{\mathrm{LU}}^\mathsf{T}\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}}^\mathsf{T} - S_{\mathrm{LU}}^\mathsf{T}\Pi_\perp(\Pi_\perp^\mathsf{T} H_{\mathrm{LU}}\Pi_\perp)^{-1}\Pi_\perp^\mathsf{T})S_{\mathrm{LU}}$$

$$
\begin{aligned}
&= (\Theta_\perp^\mathsf{T} S_{\mathrm{LU}})^\mathsf{T}[H_{\Theta,\perp}^{-1} - \Xi(\Xi^\mathsf{T} H_{\Theta,\perp}\Xi)^{-1}\Xi^\mathsf{T}](\Theta_\perp^\mathsf{T} S_{\mathrm{LU}}) \\
&= (H_{\Theta,\perp}^{-1/2}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}})^\mathsf{T}[I_{qr} - H_{\Theta,\perp}^{1/2}\Xi(\Xi^\mathsf{T} H_{\Theta,\perp}\Xi)^{-1}\Xi^\mathsf{T} H_{\Theta,\perp}^{1/2}](H_{\Theta,\perp}^{-1/2}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}}).
\end{aligned}
$$

□

Proof of Proposition C.2. Recall the definitions of $\mathbf{R}_n = [\mathbf{R}_{n,\mathrm{LU}}, \mathbf{R}_{n,\mathrm{ST}}]$ and $\mathbf{L}_n = [\mathbf{L}_{n,\mathrm{LU}}, \mathbf{L}_{n,\mathrm{ST}}]$ given at the beginning of this appendix. Since $I_{kp} = \mathbf{R}_{n,\mathrm{LU}}\mathbf{L}_{n,\mathrm{LU}}^\mathsf{T} + \mathbf{R}_{n,\mathrm{ST}}\mathbf{L}_{n,\mathrm{ST}}^\mathsf{T}$, we may write

$$
\tilde{\mathbf{\Phi}}_n = \mathbf{\Phi}_n + [(\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}]\mathbf{L}_{n,\mathrm{LU}}^\mathsf{T} + [(\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{ST}}]\mathbf{L}_{n,\mathrm{ST}}^\mathsf{T} =: \mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{LU}} + \tilde{\Delta}_{n,\mathrm{ST}}.
$$

Since $\tilde{\Delta}_{n,\mathrm{LU}} = o_p(1)$ and $\mathbf{\Phi}_n \to \mathbf{\Phi}_0$, we have $|\lambda_{q+1}(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}})| < |\lambda_q(\mathbf{\Phi}_n)|$ w.p.a.1, and so by Lemma B.2(i)

$$
A(\tilde{\mathbf{\Phi}}_n) - A(\mathbf{\Phi}_n) = A(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}} + \tilde{\Delta}_{n,\mathrm{LU}}) - A(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}}) \tag{C.9}
$$

w.p.a.1. Since $A(\cdot)$ is smooth, a second-order Taylor series expansion and Lemma B.2(ii) yield

$$
\begin{aligned}
&\operatorname{vec}\{A(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}} + \tilde{\Delta}_{n,\mathrm{LU}}) - A(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}})\} \\
&\qquad = [J_A(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}}) + o_p(1)]\operatorname{vec}\{\tilde{\Delta}_{n,\mathrm{LU}}\mathbf{R}_{\mathrm{LU}}(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}})\} \\
&\qquad = [J_A(\mathbf{\Phi}_0) + o_p(1)]\operatorname{vec}\{\tilde{\Delta}_{n,\mathrm{LU}}\mathbf{R}_{n,\mathrm{LU}}\}
\end{aligned} \tag{C.10}
$$

where the second equality holds w.p.a.1, and follows from the continuity of J_A (Lemma B.2(iii)), $\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}} = \mathbf{\Phi}_0 + o_p(1)$, and $\mathbf{R}_{\mathrm{LU}}(\mathbf{\Phi}_n + \tilde{\Delta}_{n,\mathrm{ST}}) = \mathbf{R}_{\mathrm{LU}}(\mathbf{\Phi}_n) = \mathbf{R}_{n,\mathrm{LU}}$ (w.p.a.1, as implied by Lemma B.2(i)). Finally, since

$$
\tilde{\Delta}_{n,\mathrm{LU}}\mathbf{R}_{n,\mathrm{LU}} = [(\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}]\mathbf{L}_{n,\mathrm{LU}}^\mathsf{T}\mathbf{R}_{n,\mathrm{LU}} = (\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}, \tag{C.11}
$$

the first part of (C.4) follows from (C.9)–(C.11). The proof of the second part is analogous. □

D Limiting experiments

The assumptions DGP and LOC are maintained throughout this appendix. Recall the re-parametrisation given in (3.9) above, which in view of (C.1) we can equivalently write as

$$
\boldsymbol{\pi} := n\operatorname{vec}\begin{bmatrix} A[\mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^\mathsf{T}] - A(\mathbf{\Phi}_n) \\ \Lambda_{\mathrm{LU}}[\mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^\mathsf{T}] - \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_n) \end{bmatrix} \tag{D.1a}
$$

$$
f := n^{1/2}\varphi_{\mathrm{ST}}. \tag{D.1b}
$$

Under LOC, $R_{n,\mathrm{ST}}$ and $\Lambda_{n,\mathrm{ST}}$ associated with $\{\mathbf{\Phi}_n\} \subset \mathscr{P}$ are constant (see LOC3), so $\mathbf{R}_{n,\mathrm{ST}} = \mathbf{R}_{0,\mathrm{ST}}$ for all $n \in \mathbb{N}$, so that in particular $\varphi_{\mathrm{ST}} = \operatorname{vec}\{(\mathbf{\Phi} - \mathbf{\Phi}_n)\mathbf{R}_{n,\mathrm{ST}}\}$. Let $\psi_n(\varphi)$ denote the smooth mapping $\varphi \mapsto (\boldsymbol{\pi}, f)$ implied by (D.1), which has domain $\mathcal{P}_n$ (defined in (C.2) above) and $\psi_n(0) = 0$ for all $n \in \mathbb{N}$.

Lemma D.1.

(i) *There exists an* $n_0 \in \mathbb{N}$ *and an open neighbourhood* $N \subset \mathbb{R}^{kp^2}$ *of the origin, such that* ψ_n *is a smooth diffeomorphism on* N*, for all* $n \geq n_0$

(ii) *Let* $\mathcal{K} \subset \mathbb{R}^{kp^2}$ *be any compact neighbourhood of zero. Then there exists an* $n_1 \geq n_0$ *such that* ψ_n^{-1} *is well defined (and smooth) on* $\mathcal{K}$*, for all* $n \geq n_1$*. Moreover, for any* $(\boldsymbol{\pi}, f) \in \mathcal{K}$*,* $\varphi_n := \psi_n^{-1}(\boldsymbol{\pi}, f)$ *is such that* $D_n \varphi_n = O(1)$*.*

Thus so long as we restrict attention to $\varphi \in N$, we may equivalently parametrise the model in terms of $(\boldsymbol{\pi}, f)$. For a given $(\boldsymbol{\pi}, f) \in \mathbb{R}^{kp^2}$, ψ_n^{-1} is well-defined (and smooth) at $(\boldsymbol{\pi}, f)$ for all n sufficiently large, in which case we shall define (with a slight abuse of notation) $\ell_n(\boldsymbol{\pi}, f) := \ell_n(\varphi, \Sigma)$, where $\varphi = \psi_n^{-1}(\boldsymbol{\pi}, f)$; and set $\ell_n(\boldsymbol{\pi}, f) := -\infty$ otherwise (to simplify arguments, we treat Σ as known here.) To state our next result, recall the definitions of $S_{\boldsymbol{\pi}}$ and $H_{\boldsymbol{\pi}}$ given in (3.13).

Lemma D.2. *Jointly over any finite collection of* $(\boldsymbol{\pi}, f) \in \mathbb{R}^{kp^2}$*,*

$$\ell_n(\boldsymbol{\pi}, f) - \ell_n(0,0) \rightsquigarrow [S_{\boldsymbol{\pi}}^{\mathsf{T}} \boldsymbol{\pi} - \tfrac{1}{2} \boldsymbol{\pi}^{\mathsf{T}} H_{\boldsymbol{\pi}} \boldsymbol{\pi}] + [S_{\mathrm{ST}}^{\mathsf{T}} f - \tfrac{1}{2} f^{\mathsf{T}} H_{\mathrm{ST}} f].$$

We next show that, up to the term depending on f, the preceding is also the limit of the loglikelihood ratio process in a multivariate predictive regression with a known covariance matrix; recall PR given in Section 3.2.

Lemma D.3. *Suppose that* $\{y_{\mathrm{PR},t}\}$ *and* $\{z_{\mathrm{PR}t}\}$ *are generated under* PR*, and that* $\xi_t = [\begin{smallmatrix} \xi_{yt} \\ \xi_{zt} \end{smallmatrix}] \sim_{\text{i.i.d.}} N[0, \Omega]$ *with* $\Omega = K \Sigma K^{\mathsf{T}}$*. Then for*

$$\boldsymbol{\pi} = n \operatorname{vec} \begin{bmatrix} A - A(\boldsymbol{\Phi}_n) \\ \Lambda - \Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}_n) \end{bmatrix},$$

we have, jointly over any finite collection of $\boldsymbol{\pi} \in \mathbb{R}^{pq}$

$$\ell_{n,\mathrm{PR}}(\boldsymbol{\pi}) - \ell_{n,\mathrm{PR}}(0) \rightsquigarrow S_{\boldsymbol{\pi}}^{\mathsf{T}} \boldsymbol{\pi} - \tfrac{1}{2} \boldsymbol{\pi}^{\mathsf{T}} H_{\boldsymbol{\pi}} \boldsymbol{\pi},$$

where $\ell_{n,\mathrm{PR}}(\boldsymbol{\pi})$ *is the loglikelihood defined in Theorem 3.1.*

Finally, we show that when (the entirety of) $\boldsymbol{\Phi}$ is in unknown, and the model is estimated subject to the constraint (D.1a), then the limit of the *concentrated* loglikelihood ratio process is asymptotically identical to that of the predictive regression, up to (random) terms that do not depend on $\boldsymbol{\pi}$. Let $\hat{f}_{n|\boldsymbol{\pi}} := n^{1/2} \hat{\varphi}_{n,\mathrm{ST}|\boldsymbol{\pi}}$, where $\hat{\varphi}_{n|\boldsymbol{\pi}}$ denotes the maximiser of $\ell_n^*(\varphi)$ subject to φ satisfying (D.1a).

Lemma D.4. *Jointly over every finite collection of* $\boldsymbol{\pi} \in \mathbb{R}^{pq}$*,*

$$\ell_n(\boldsymbol{\pi}, \hat{f}_{n|\boldsymbol{\pi}}) - \ell_n(0) \rightsquigarrow S_{\boldsymbol{\pi}}^{\mathsf{T}} \boldsymbol{\pi} - \tfrac{1}{2} \boldsymbol{\pi}^{\mathsf{T}} H_{\boldsymbol{\pi}} \boldsymbol{\pi} + S_{\mathrm{ST}}^{\mathsf{T}} H_{\mathrm{ST}}^{-1} S_{\mathrm{ST}}$$

Proof of Lemma D.1. Consider the mapping Ψ and the permutation matrix $M \in \mathbb{R}^{pq \times pq}$ such that

$$\Psi(\mathbf{\Phi}) \coloneqq \begin{bmatrix} \operatorname{vec} A(\mathbf{\Phi}) \\ \operatorname{vec} \Lambda_{\mathrm{LU}}(\mathbf{\Phi}) \\ \operatorname{vec} \mathbf{\Phi}\mathbf{R}_{0,\mathrm{ST}} \end{bmatrix} \qquad \mathcal{M}\Psi(\mathbf{\Phi}) \coloneqq \begin{bmatrix} M & 0 \\ 0 & I_{\#\mathrm{ST}} \end{bmatrix} \Psi(\mathbf{\Phi}) = \begin{bmatrix} \operatorname{vec} \begin{bmatrix} A(\mathbf{\Phi}) \\ \Lambda_{\mathrm{LU}}(\mathbf{\Phi}) \end{bmatrix} \\ \operatorname{vec} \mathbf{\Phi}\mathbf{R}_{0,\mathrm{ST}} \end{bmatrix} \tag{D.2}$$

By Lemmas B.1 and B.2, Ψ is smooth and at $\mathbf{\Phi} = \mathbf{\Phi}_0$ has first differential

$$\mathrm{d}\Psi = \begin{bmatrix} J_A(\mathbf{\Phi}_0) & 0 \\ J_\Lambda(\mathbf{\Phi}_0) & 0 \\ 0 & I_{\#\mathrm{ST}} \end{bmatrix} \left(\begin{bmatrix} \mathbf{R}_{0,\mathrm{LU}}^{\mathsf{T}} \\ \mathbf{R}_{0,\mathrm{ST}}^{\mathsf{T}} \end{bmatrix} \otimes I_p \right) \operatorname{vec}(\mathrm{d}\mathbf{\Phi}). \tag{D.3}$$

The Jacobian on the r.h.s. is invertible by Lemma A.1(iii) and Lemma B.3 (for the latter, since $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}_0) = I_q$). Thus by the inverse mapping theorem, there is an open neighbourhood $N_{\mathscr{P}} \subset \mathscr{P}$ of $\mathbf{\Phi}_0$ on which Ψ has a smooth inverse.

Now let $\tau_n(\varphi) \coloneqq \mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi)\mathbf{L}_n^{\mathsf{T}}$, which converges (uniformly on compacta) to a linear and invertible mapping $\tau_0(\varphi)$ for which $\tau_0(0) = \mathbf{\Phi}_0$. Hence there exists an $n_0 \in \mathbb{N}$ and a (fixed) open neighbourhood $N \subset \mathcal{P}_n$ of zero such that $\tau_n(N) \subset N_{\mathscr{P}}$, for all $n \geq n_0$, with τ_n being invertible on N. By composition, the sequence of maps defined by

$$D_n^{-1}\psi_n(\varphi) = \mathcal{M}\{\Psi[\tau_n(\varphi)] - \Psi(\mathbf{\Phi}_n)\} \tag{D.4}$$

is smooth and invertible on N, for all $n \geq 0$, and has a smooth inverse there; hence part (i) holds. Finally, since the image of N under the r.h.s. must itself be an open neighbourhood of zero, and

$$\begin{bmatrix} \boldsymbol{\pi} \\ f \end{bmatrix} = \psi_n(\varphi) = D_n\mathcal{M}\{\Psi[\tau_n(\varphi)] - \Psi(\mathbf{\Phi}_n)\}, \tag{D.5}$$

we may deduce that for any compact neighbourhood $\mathcal{K}$ of zero, there is an $n_1 \geq n_0$ such that the inverse ψ_n^{-1} is well-defined and smooth for all $(\boldsymbol{\pi}, f) \in \mathcal{K}$, for all $n \geq n_1$. Finally, to show that the $\varphi_n \coloneqq \psi_n^{-1}(\boldsymbol{\pi}, f)$ has $D_n\varphi_n = O(1)$, we note that since the r.h.s. of (D.4) is (locally to zero) a diffeomorphism, which itself equals zero at $\varphi = 0$, the fact that

$$\mathcal{M}\{\Psi[\tau_n(\varphi_n)] - \Psi(\mathbf{\Phi}_n)\} = D_n^{-1} \begin{bmatrix} \boldsymbol{\pi} \\ f \end{bmatrix} \to 0$$

must imply that $\varphi_n \to 0$. Hence follows from (D.3) and (D.5) that, by a Taylor expansion of (D.5) around $\varphi = 0$,

$$\begin{bmatrix} MJ + o_p(1) & 0 \\ 0 & I_{\#\mathrm{ST}} \end{bmatrix} \begin{bmatrix} \varphi_{n,\mathrm{LU}} \\ \varphi_{n,\mathrm{ST}} \end{bmatrix} = D_n^{-1} \begin{bmatrix} \boldsymbol{\pi} \\ f \end{bmatrix}$$

whence $D_n\varphi_n = O(1)$ as claimed. Thus part (ii) holds. □

Proof of Lemma D.2. In view of Lemma D.1, we may take n sufficiently large such that ψ_n^{-1} is well defined at $(\boldsymbol{\pi}, f)$. Let $\varphi_n^* \coloneqq \psi_n^{-1}(\boldsymbol{\pi}, f) = o(1)$, $\mathbf{\Phi}_n^* \coloneqq \mathbf{\Phi}_n + \operatorname{vec}^{-1}(\varphi_n^*)\mathbf{L}_n^{\mathsf{T}}$, and $M \in \mathbb{R}^{pq\times pq}$

be as in (D.2). Then by Proposition C.2, for $J \coloneqq J(\mathbf{\Phi}_0)$

$$\boldsymbol{\pi} = n \operatorname{vec} \begin{bmatrix} A(\mathbf{\Phi}_n^*) - A(\mathbf{\Phi}_n) \\ \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_n^*) - \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_n) \end{bmatrix} = [MJ + o(1)] n \varphi_{n,\mathrm{LU}}^* \tag{D.6}$$

where by Lemma B.3 and the definition of M,

$$MJ = M \begin{bmatrix} I_q \otimes \mathcal{J} \\ I_q \otimes L_{\mathrm{LU}}^{\mathsf{T}} \end{bmatrix} = I_q \otimes \begin{bmatrix} \mathcal{J} \\ L_{\mathrm{LU}}^{\mathsf{T}} \end{bmatrix} =: I_q \otimes K$$

for $\mathcal{J}$ and K as defined in (3.10). Noting also that $n^{1/2}\varphi_{n,\mathrm{ST}}^* = f$, it follows from Lemma C.2 and (D.6) that

$$\begin{aligned} \ell_n(\boldsymbol{\pi}, f) - \ell_n(0,0) &= \ell_n(\varphi_{n,\mathrm{LU}}^*, \varphi_{n,\mathrm{ST}}^*) - \ell_n(0,0) \\ &= S_n^{\mathsf{T}}(D_n\varphi_n^*) - \tfrac{1}{2}(D_n\varphi_n^*)^{\mathsf{T}} H_n (D_n\varphi_n^*) \\ &\rightsquigarrow [S_{\mathrm{LU}}^{\mathsf{T}}(MJ)^{-1}\boldsymbol{\pi} - \tfrac{1}{2}\boldsymbol{\pi}^{\mathsf{T}}[(MJ)^{-1}]^{\mathsf{T}} H_{\mathrm{LU}}(MJ)^{-1}\boldsymbol{\pi}] + [S_{\mathrm{ST}}^{\mathsf{T}} f - \tfrac{1}{2} f^{\mathsf{T}} H_{\mathrm{ST}} f]. \end{aligned}$$

To complete the proof, we note that

$$\begin{aligned} [(MJ)^{-1}]^{\mathsf{T}} S_{\mathrm{LU}} &= (I_q \otimes K^{-1})^{\mathsf{T}} \int_0^1 [\bar{Z}_C(r) \otimes \Sigma^{-1} \mathrm{d}E(r)] \\ &= \int_0^1 [\bar{Z}_C(r) \otimes (K\Sigma K^{\mathsf{T}})^{-1/2} \mathrm{d}W(r)] = S_{\boldsymbol{\pi}} \end{aligned}$$

where we have used that $E(s) = \Sigma^{-1/2} W(s)$ and,

$$\begin{aligned} [(MJ)^{-1}]^{\mathsf{T}} H_{\mathrm{LU}} (MJ)^{-1} &= (I_q \otimes K^{-1})^{\mathsf{T}} \left(\int \bar{Z}_C \bar{Z}_C^{\mathsf{T}} \otimes \Sigma^{-1} \right) (I_q \otimes K^{-1}) \\ &= \int \bar{Z}_C \bar{Z}_C^{\mathsf{T}} \otimes (K\Sigma K^{\mathsf{T}})^{-1} = H_{\boldsymbol{\pi}}. \end{aligned}$$

□

Proof of Lemma D.3. Letting $\mathbf{\Pi} = [\begin{smallmatrix} A_{\mathrm{PR}} \\ \Lambda_{\mathrm{PR}} \end{smallmatrix}]$ and noting that $\boldsymbol{\pi} = n \operatorname{vec}(\mathbf{\Pi} - [\begin{smallmatrix} A(\mathbf{\Phi}_n) \\ \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_n) \end{smallmatrix}])$, we have

$$\ell_n^{\mathrm{PR}}(\boldsymbol{\pi}) = K_n - \frac{1}{2} \sum_{t=1}^n \| x_t - \mathbf{\Pi} z_{t-1} \|_{\Omega^{-1}},$$

where $K_n \coloneqq -\frac{n}{2} \log(2\pi \log\det\Omega)$. It then follows by exactly the same arguments as were used in the proof of Lemma C.2 that

$$\ell_{n,\mathrm{PR}}(\boldsymbol{\pi}) - \ell_{n,\mathrm{PR}}(0) = S_{n,\mathrm{PR}}^{\mathsf{T}} \boldsymbol{\pi} - \tfrac{1}{2} \boldsymbol{\pi}^{\mathsf{T}} H_{n,\mathrm{PR}} \boldsymbol{\pi}$$

where

$$S_{n,\mathrm{PR}} = \frac{1}{n} \sum_{t=1}^n (\bar{z}_{t-1} \otimes \Omega^{-1/2} \eta_t) \qquad H_{n,\mathrm{PR}} = \frac{1}{n} \sum_{t=1}^n (\bar{z}_{t-1} \bar{z}_{t-1}^{\mathsf{T}} \otimes \Omega^{-1}).$$

Under PR, it follows by Lemma C.1(iii) that $n^{-1/2} \bar{z}_{\lfloor nr \rfloor} \rightsquigarrow \bar{Z}_{C,\mathrm{PR}}(r)$ on $D[0,1]$, where $\bar{Z}_{C,\mathrm{PR}}(r)$

denotes the residual from the projection of

$$Z_{C,\mathrm{PR}}(r) := \int_0^r \mathrm{e}^{C(r-s)}\Omega_{zz}^{1/2}\mathrm{d}W(s) \tag{D.7}$$

on a constant and a linear trend, and we have partitioned $\Omega = \left[\begin{smallmatrix}\Omega_{yy} & \Omega_{yz}\\ \Omega_{zy} & \Omega_{zz}\end{smallmatrix}\right]$ conformably with $\xi_t = \left[\begin{smallmatrix}\xi_{yt}\\ \xi_{zt}\end{smallmatrix}\right]$. Then by the continuous mapping theorem and the same arguments as used in the proof of Lemma C.1(iv),

$$S_{n,\mathrm{PR}} \rightsquigarrow \int_0^1 [\bar{Z}_{C,\mathrm{PR}}(r) \otimes \Omega^{-1/2}W(r)]\,\mathrm{d}r \qquad H_{n,\mathrm{PR}} \rightsquigarrow \int \bar{Z}_{C,\mathrm{PR}}\bar{Z}_{C,\mathrm{PR}}^{\mathsf{T}} \otimes \Omega^{-1}. \tag{D.8}$$

Thus we can bring (D.7) into agreement with (3.6), and the limits on the r.h.s. of (D.8) with (3.13), by setting

$$\Omega = \begin{bmatrix}\Omega_{yy} & \Omega_{yz}\\ \Omega_{zy} & \Omega_{zz}\end{bmatrix} = \begin{bmatrix}\mathcal{J}\Sigma\mathcal{J}^{\mathsf{T}} & \mathcal{J}\Sigma L_{\mathrm{LU}}\\ L_{\mathrm{LU}}^{\mathsf{T}}\Sigma\mathcal{J}^{\mathsf{T}} & L_{\mathrm{LU}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}}\end{bmatrix} = K\Sigma K^{\mathsf{T}}. \qquad \square$$

Proof of Lemma D.4. We first show that $D_n\hat{\varphi}_{n|\boldsymbol{\pi}} = O_p(1)$. By Lemma D.1, for all n sufficiently large, there exists a (deterministic) sequence $\varphi_{n|\boldsymbol{\pi}} \in \mathcal{P}_n$ with $D_n\varphi_{n|\boldsymbol{\pi}} = O(1)$, such that (D.1a) holds at $\varphi = \varphi_{n|\boldsymbol{\pi}}$. It follows from Lemma C.2 that for each $\epsilon > 0$, there exists an $N < \infty$ such that

$$\limsup_{n\to\infty} \mathbb{P}\{\ell_n^*(\varphi_{n|\boldsymbol{\pi}}) - \ell_n^*(0) < -N\} < \epsilon/2.$$

On the other hand, adapting the argument given in the proof of Lemma C.3, we may also choose $M < \infty$ sufficiently large such that

$$\mathbb{P}\left\{\sup_{\{\varphi\in\mathcal{P}_n \mid \|D_n\varphi\|\geq M\}} [\ell_n^*(\varphi) - \ell_n^*(0)] < -2N\right\} \geq \mathbb{P}\left\{M[\|S_n\| - \tfrac{1}{2}\lambda_{\min}(H_n)M] \leq -2N\right\}$$
$$> 1 - \epsilon/2$$

for all n sufficiently large. Deduce that with probability at least $1-\epsilon$, $\ell_n^*(\varphi_{n|\boldsymbol{\pi}})$ must strictly exceed $\ell_n(\varphi)$ over all $\varphi \in \mathcal{P}_n$ with $\|D_n\varphi\| \geq M$; it follows that the constrained maximiser $\hat{\varphi}_{n|\boldsymbol{\pi}}$ must have $\|D_n\hat{\varphi}_{n|\boldsymbol{\pi}}\| < M$. Deduce that $D_n\hat{\varphi}_{n|\boldsymbol{\pi}} = O_p(1)$ as claimed.

Now it follows from (D.3) and (D.5) that, at $\varphi = \hat{\varphi}_{n|\boldsymbol{\pi}}$,

$$\begin{bmatrix}\mathrm{d}\boldsymbol{\pi}\\ \mathrm{d}f\end{bmatrix} = \begin{bmatrix}MJ + o_p(1) & 0\\ 0 & I_{\#\mathrm{ST}}\end{bmatrix} D_n\mathrm{d}\varphi$$

and hence

$$D_n\mathrm{d}\varphi = \begin{bmatrix}(MJ)^{-1} + o_p(1) & 0\\ 0 & I_{\#\mathrm{ST}}\end{bmatrix}\begin{bmatrix}\mathrm{d}\boldsymbol{\pi}\\ \mathrm{d}f\end{bmatrix}$$

at $(\boldsymbol{\pi}, \hat{f}_{n|\boldsymbol{\pi}})$. Thus

$$n\hat{\varphi}_{n,\mathrm{LU}} = [(MJ)^{-1} + o_p(1)]\boldsymbol{\pi}, \tag{D.9}$$

and since $\hat{f}_{n|\boldsymbol{\pi}}$ must satisfy the first-order conditions for a maximum, we have from Lemma C.2

that

$$\nabla_f \ell_n(\boldsymbol{\pi}, \hat{f}_{n|\boldsymbol{\pi}}) = \nabla_f [S_n^{\mathsf{T}}(D_n\varphi) - \tfrac{1}{2}(D_n\varphi)^{\mathsf{T}} H_n(D_n\varphi)]_{f=\hat{f}_{n|\boldsymbol{\pi}}}$$
$$= S_{n,\mathrm{ST}} - (n\hat{\varphi}_{n,\mathrm{LU}|\boldsymbol{\pi}})^{\mathsf{T}} H_{n,\mathrm{LS}} - H_{n,\mathrm{ST}} n^{1/2} \hat{\varphi}_{n,\mathrm{ST}|\boldsymbol{\pi}}$$

whence

$$n^{1/2}\hat{\varphi}_{n,\mathrm{ST}|\boldsymbol{\pi}} = H_{n,\mathrm{ST}}^{-1} S_{n,\mathrm{ST}} + o_p(1) \rightsquigarrow H_{\mathrm{ST}}^{-1} S_{\mathrm{ST}}. \tag{D.10}$$

Thus, in view of (D.9) and (D.10), the weak limit of

$$\ell_n(\boldsymbol{\pi}, \hat{f}_{n|\boldsymbol{\pi}}) - \ell_n(0) = \ell_n^*(\hat{\varphi}_{n|\boldsymbol{\pi}}) - \ell_n^*(0) = S_n^{\mathsf{T}}(D_n\hat{\varphi}_{n|\pi}) - \tfrac{1}{2}(D_n\hat{\varphi}_{n|\pi})^{\mathsf{T}} H_n(D_n\hat{\varphi}_{n|\pi})$$

is as claimed. □

E Proofs of theorems

Proof of Theorem 3.1. This follows directly from Lemmas D.2–D.4, noting in particular that $\ell_n^*(\boldsymbol{\pi}) = \ell_n(\boldsymbol{\pi}, \hat{f}_{n|\boldsymbol{\pi}})$, where the latter is as appears in Lemma D.4. □

Proof of Theorem 3.2. **(i).** In the notation of Appendix C, $\hat{\varphi}_{n,\mathrm{LU}} = \operatorname{vec}\{(\hat{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}\}$. By Proposition C.1(i)

$$n \operatorname{vec}\{(\hat{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}\} \rightsquigarrow \left[\left(\int \bar{Z}_C \bar{Z}_C^{\mathsf{T}}\right)^{-1} \otimes I_p\right] \int_0^1 [\bar{Z}_C(r) \otimes \mathrm{d}E(r)]$$
$$= \operatorname{vec}\left\{\int (\mathrm{d}E)\bar{Z}_C^{\mathsf{T}} \left(\int \bar{Z}_C \bar{Z}_C^{\mathsf{T}}\right)^{-1}\right\},$$

and so by Proposition C.2

$$\begin{bmatrix} \operatorname{vec}\{A(\hat{\boldsymbol{\Phi}}_n) - A(\boldsymbol{\Phi}_n)\} \\ \operatorname{vec}\{\Lambda_{\mathrm{LU}}(\hat{\boldsymbol{\Phi}}_n) - \Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}_n)\} \end{bmatrix} \rightsquigarrow \begin{bmatrix} J_A(\boldsymbol{\Phi}_0) \\ J_\Lambda(\boldsymbol{\Phi}_0) \end{bmatrix} \operatorname{vec}\left\{\int (\mathrm{d}E)\bar{Z}_C^{\mathsf{T}} \left(\int \bar{Z}_C \bar{Z}_C^{\mathsf{T}}\right)^{-1}\right\}. \tag{E.1}$$

Since $\boldsymbol{\Phi}_n \to \boldsymbol{\Phi}_0$ with $\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}_0) = I_q$ under LOC, we have by Lemma B.3 that

$$\begin{bmatrix} J_A(\boldsymbol{\Phi}_0) \\ J_\Lambda(\boldsymbol{\Phi}_0) \end{bmatrix} = \begin{bmatrix} I_q \otimes \beta^{\mathsf{T}} R_{\mathrm{ST}}(I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1} L_{\mathrm{ST}}^{\mathsf{T}} \\ I_q \otimes L_{\mathrm{LU}}^{\mathsf{T}} \end{bmatrix}. \tag{E.2}$$

The result then follows from (E.1) and (E.2), by reversing the vectorisation.

(ii). In the notation of Appendix C, maximising $\ell_n^*(\boldsymbol{\Phi})$ subject to $\Lambda_{\mathrm{LU}}(\boldsymbol{\Phi}) = \Lambda_{n,\mathrm{LU}} = I_q + C/n$ corresponds to maximising $\ell_n(\varphi)$ subject to $\theta_n(\varphi) = 0$. Thus $\hat{\varphi}_{n,\mathrm{LU}|\theta} = \operatorname{vec}\{(\hat{\boldsymbol{\Phi}}_{n|\Lambda_{n,\mathrm{LU}}} - \boldsymbol{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}\}$, and so by Proposition C.1(ii)

$$n \operatorname{vec}\{(\hat{\boldsymbol{\Phi}}_{n|\Lambda_{n,\mathrm{LU}}} - \boldsymbol{\Phi}_n)\mathbf{R}_{n,\mathrm{LU}}\} \rightsquigarrow \Theta_\perp(\Theta_\perp^{\mathsf{T}} H_{\mathrm{LU}} \Theta_\perp)^{-1} \Theta_\perp^{\mathsf{T}} S_{\mathrm{LU}}$$

where $\Theta_\perp = I_q \otimes L_{\mathrm{LU},\perp}$. Hence by Proposition C.2,

$$\operatorname{vec}\{A(\hat{\mathbf{\Phi}}_{n|\Lambda_{n,\mathrm{LU}}}) - A(\mathbf{\Phi}_n)\} \rightsquigarrow J_A(\mathbf{\Phi}_0)\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}}.$$

To determine the distribution of the r.h.s., we note that

$$\Theta_\perp^\mathsf{T} S_{\mathrm{LU}} = \int_0^1 [\bar{Z}_C(r) \otimes L_{\mathrm{LU},\perp}^\mathsf{T}\Sigma^{-1}\mathrm{d}E(r)] =: \int_0^1 [\bar{Z}_C(r) \otimes \mathrm{d}U(r)]. \tag{E.3}$$

Recall that $\bar{Z}_C$ is a function only of Z_C, which from (3.6) is given by

$$Z_C(r) = \int_0^r \mathrm{e}^{C(r-s)} L_{\mathrm{LU}}^\mathsf{T}\mathrm{d}E(s) =: \int_0^r \mathrm{e}^{C(r-s)}\mathrm{d}V(s). \tag{E.4}$$

$(U, V) = (L_{\mathrm{LU},\perp}^\mathsf{T}\Sigma^{-1}E, L_{\mathrm{LU}}^\mathsf{T}E)$ is a pair of vector Brownian motions, with covariance

$$\mathbb{E}U(1)V(1)^\mathsf{T} = L_{\mathrm{LU},\perp}^\mathsf{T}\Sigma^{-1}\mathbb{E}[E(1)E(1)^\mathsf{T}]L_{\mathrm{LU}} = L_{\mathrm{LU},\perp}^\mathsf{T} L_{\mathrm{LU}} = 0;$$

whence U and V are independent. In particular, we have from (E.4) that U is independent of $\bar{Z}_C$. This, combined with the fact that

$$J_A(\mathbf{\Phi}_0)\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1} = \left(\int \bar{Z}_C\bar{Z}_C^\mathsf{T}\right)^{-1} \otimes \mathcal{J} L_{\mathrm{LU},\perp}(L_{\mathrm{LU},\perp}^\mathsf{T}\Sigma^{-1}L_{\mathrm{LU},\perp})^{-1}$$

depends only on $\bar{Z}_C$, implies $J_A(\mathbf{\Phi}_0)\Theta_\perp(\Theta_\perp^\mathsf{T} H_{\mathrm{LU}}\Theta_\perp)^{-1}\Theta_\perp^\mathsf{T} S_{\mathrm{LU}}$ is mixed normal with variance

$$\left(\int \bar{Z}_C\bar{Z}_C^\mathsf{T}\right)^{-1} \otimes \mathcal{J} L_{\mathrm{LU},\perp}(L_{\mathrm{LU},\perp}^\mathsf{T}\Sigma^{-1}L_{\mathrm{LU},\perp})^{-1}L_{\mathrm{LU},\perp}^\mathsf{T}\mathcal{J}^\mathsf{T},$$

which proves (3.14).

Finally, note that the preceding holds for any choice of $L_{\mathrm{LU},\perp} \in \mathbb{R}^{p\times r}$ having full column rank and $L_{\mathrm{LU},\perp}^\mathsf{T} L_{\mathrm{LU}} = 0$. Let $\alpha := \Phi_0(1)\beta(\beta^\mathsf{T}\beta)^{-1} \in \mathbb{R}^{p\times r}$, where $\Phi_0(1) := \lim_{n\to\infty}\Phi_n(1)$; then

$$L_{\mathrm{LU}}^\mathsf{T}\alpha = L_{\mathrm{LU}}^\mathsf{T}\Phi_0(1)\beta(\beta^\mathsf{T}\beta)^{-1} = 0$$

by (2.6) with $\Lambda_{\mathrm{LU}} = \Lambda_{\mathrm{LU}}(\mathbf{\Phi}_0) = I_q$. Further, $\operatorname{rk}\alpha = r$ since $\operatorname{sp}\Phi_0(1) = \operatorname{sp}\beta$, and thus we may indeed choose $L_{\mathrm{LU},\perp} = \alpha$. In this case,

$$\mathcal{J} L_{\mathrm{LU},\perp} = \beta^\mathsf{T} R_{\mathrm{ST}}(I_{kp-q} - \Lambda_{\mathrm{ST}})^{-1}L_{\mathrm{ST}}^\mathsf{T}\Phi_0(1)\beta(\beta^\mathsf{T}\beta)^{-1} =_{(1)} \beta^\mathsf{T}\beta(\beta^\mathsf{T}\beta)^{-1} = I_r,$$

where $=_{(1)}$ follows from (B.16) above. Thus (3.15) is proved. □

Proof of Theorem 3.3. We first prove (3.17). In the notation of Appendix C, $\mathcal{LR}_n(\Lambda_{n,\mathrm{LU}}) = 2[\ell_n^*(\hat{\varphi}_n) - \ell_n^*(\hat{\varphi}_{n|\theta})]$. By Proposition C.1(iii),

$$\mathcal{LR}_n(\Lambda_{n,\mathrm{LU}}) \rightsquigarrow S_{\mathrm{LU}}^\mathsf{T} H_{\mathrm{LU}}^{-1}\Theta(\Theta^\mathsf{T} H_{\mathrm{LU}}^{-1}\Theta)^{-1}\Theta^\mathsf{T} H_{\mathrm{LU}}^{-1}S_{\mathrm{LU}} =: \mathcal{LR},$$

where $\Theta = I_q \otimes L_{\mathrm{LU}}$, $S_{\mathrm{LU}} = \int[\bar{Z}_C(r)\otimes\Sigma^{-1}\mathrm{d}E]$, and $H_{\mathrm{LU}} = \int\bar{Z}_C\bar{Z}_C^\mathsf{T}\otimes\Sigma^{-1}$. To obtain the claimed expression for $\mathcal{LR}$, note that

$$S_{\mathrm{LU}} = \int [\bar{Z}_C(r) \otimes \Sigma^{-1}\mathrm{d}E] = \mathrm{vec}\left\{\Sigma^{-1}\int (\mathrm{d}E)\bar{Z}_C^{\mathsf{T}}\right\}$$

and

$$H_{\mathrm{LU}}^{-1}\Theta(\Theta^{\mathsf{T}}H_{\mathrm{LU}}^{-1}\Theta)^{-1}\Theta^{\mathsf{T}}H_{\mathrm{LU}}^{-1} = \left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\right)^{-1} \otimes \Sigma L_{\mathrm{LU}}(L_{\mathrm{LU}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}})^{-1}L_{\mathrm{LU}}^{\mathsf{T}}\Sigma$$

whence, using $\mathrm{vec}(A)^{\mathsf{T}}\,\mathrm{vec}(B) = \mathrm{tr}(A^{\mathsf{T}}B)$,

$$\mathcal{LR} = \mathrm{tr}\left\{\Delta^{-1/2}L_{\mathrm{LU}}^{\mathsf{T}}\int(\mathrm{d}E)\bar{Z}_C^{\mathsf{T}}\left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}}\right)^{-1}\int \bar{Z}_C(\mathrm{d}E)^{\mathsf{T}}L_{\mathrm{LU}}\Delta^{-1/2}\right\} \tag{E.5}$$

where $\Delta \coloneqq L_{\mathrm{LU}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}}$. To simplify this further, note that $L_{\mathrm{LU}}^{\mathsf{T}}E$ is a q-dimensional Brownian motion with variance Δ, and so for $W_*(r) \coloneqq \Delta^{-1/2}L_{\mathrm{LU}}^{\mathsf{T}}E(r) \sim \mathrm{BM}(I_q)$, we have

$$\begin{aligned} Z_C(r) = \int_0^r \mathrm{e}^{C(r-s)}L_{\mathrm{LU}}^{\mathsf{T}}\mathrm{d}E(s) = \int_0^r \mathrm{e}^{C(r-s)}\Delta^{1/2}\mathrm{d}W_*(s) \\ =_{(1)} \Delta^{1/2}\int_0^r \mathrm{e}^{C_*(r-s)}\mathrm{d}W_*(s) =: \Delta^{1/2}Z_{C_*}(r) \end{aligned}$$

where $C_* \coloneqq \Delta^{-1/2}C\Delta^{1/2}$ is as in the statement of the theorem, and $=_{(1)}$ follows from $\mathrm{e}^{C}D = D\mathrm{e}^{D^{-1}CD}$ for any nonsingular D. Hence $\bar{Z}_C(r) = \Delta^{1/2}\bar{Z}_{C_*}(r)$, whereupon (3.17) follows from (E.5) and the definition of W_*.

We next prove (3.18). Maximisation of $\ell_n^*(\mathbf{\Phi})$ subject to $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = I_q + C/n$ and $a_{ij}(\mathbf{\Phi}) = a_0$ corresponds, in the notation of Appendix C, to maximisation of $\ell_n(\varphi)$ subject to $\theta_n(\varphi) = 0$ and $\gamma_n(\varphi) = 0$. Therefore by Proposition C.1(iv),

$$\mathcal{LR}_n[a_{ij}(\mathbf{\Phi}_n); \Lambda_{n,\mathrm{LU}}] = 2[\ell_n(\hat{\varphi}_{n|\theta}) - \ell_n(\hat{\varphi}_{n|\theta,\gamma})] \rightsquigarrow (H_{\Theta,\perp}^{-1/2}\Theta_\perp^{\mathsf{T}}S_{\mathrm{LU}})^{\mathsf{T}}[I_{qr} - \mathcal{Q}](H_{\Theta,\perp}^{-1/2}\Theta_\perp^{\mathsf{T}}S_{\mathrm{LU}}).$$

Recall from (E.3) and the subsequent arguments that

$$\mathrm{vec}\{\Theta_\perp^{\mathsf{T}}S_{\mathrm{LU}}\} =_d \left(\int \bar{Z}_C\bar{Z}_C^{\mathsf{T}} \otimes L_{\mathrm{LU},\perp}^{\mathsf{T}}\Sigma^{-1}L_{\mathrm{LU},\perp}\right)^{1/2}\eta$$

for $\eta \sim \mathrm{N}[0, I_{qr}]$ independent of $\bar{Z}_C$, and therefore also of

$$H_{\Theta,\perp} = \Theta_\perp^{\mathsf{T}}H_{\mathrm{LU}}\Theta_\perp = \int \bar{Z}_C\bar{Z}_C^{\mathsf{T}} \otimes L_{\mathrm{LU},\perp}^{\mathsf{T}}\Sigma^{-1}L_{\mathrm{LU},\perp}.$$

Thus $\mathrm{vec}\{H_{\Theta,\perp}^{-1/2}\Theta_\perp^{\mathsf{T}}S_{\mathrm{LU}}\} \sim \mathrm{N}[0, I_{qr}]$ is independent of H_{LU}, and therefore also of $\mathcal{Q}$. The result follows by noting that $H_{\Theta,\perp}^{1/2}\Xi$ has rank $qr-1$ a.s., whence $I_{qr} - \mathcal{Q}$ projects orthogonally onto a subspace of dimension 1, a.s. $\square$

F Computational appendix

F.1 Test statistics

Computation of $\mathcal{C}_{a_{ij}|\Lambda_0}$ and $\mathcal{C}_{\mathrm{NP}}$ involves maximising $\ell_n^*(\mathbf{\Phi})$ subject to the restrictions that $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0$ for some specified $\Lambda_0 \in \mathscr{L}$, and possibly also that for and $a_{ij}(\mathbf{\Phi}) = a_0$ for some $a_0 \in \mathbb{R}$. To implement this estimator numerically, we introduce the constraint

$$\begin{bmatrix} A \\ I_q \end{bmatrix} \Lambda_0^k - \sum_{i=1}^{k} \Phi_i \begin{bmatrix} A \\ I_q \end{bmatrix} \Lambda_0^{k-i} = 0, \tag{F.1}$$

which incorporates (2.6) and (3.3) above: it forces $\Phi(\lambda)$ to have roots at the eigenvalues of Λ_0, and the associated R_{LU} matrix to respect the normalisation (3.3). We then proceed as follows:

(i) Given $A \in \mathbb{R}^{r\times q}$ and $\Lambda_0 \in \mathscr{L}$, maximise $\ell_n^*(\mathbf{\Phi})$ over $\mathbf{\Phi} \in \mathbb{R}^{p\times kp}$, subject to (F.1), to obtain the maximum likelihood estimate $\hat{\mathbf{\Phi}}_{n|A,\Lambda_0}$ using two-step, restricted least-squares estimation. This is straightforward, since (F.1) is a linear restriction on $\mathbf{\Phi}$ (see Lütkepohl, 2007, Ch. 7).

(ii) Using a general purpose optimiser, compute

$$\max_{A\in\mathbb{R}^{r\times q}} \ell_n^*(\hat{\mathbf{\Phi}}_{n|A,\Lambda_0}). \tag{F.2}$$

The maximum of $\ell_n^*(\mathbf{\Phi})$ subject to $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \Lambda_0$ and $a_{ij}(\mathbf{\Phi}) = a_0$ obtains by holding restricting a_{ij} when computing the maximum in (F.2). Point estimates of $\mathbf{\Phi}$ that merely impose the requirement that $\Lambda_0 \in \mathscr{L}$ (as appears e.g. in (3.16a)) obtain by maximising $\ell_n^*(\hat{\mathbf{\Phi}}_{n|A,\Lambda_0})$ over both A and $\Lambda_0 \in \mathscr{L}$.

When $q = 1$, and in the special case where $\Lambda_0 = \lambda_0 I_q$, (F.2) simplifies so that model (3.1) becomes

$$\Delta_{\lambda_0} y_t = m + dt - \lambda_0^{-k+1}\Phi(\lambda_0)y_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta_{\lambda_0} y_{t-i} \tag{F.3}$$

where $\Delta_{\lambda_0} y_t \coloneqq y_t - \lambda_0 y_{t-1}$ denotes a quasi-difference (see Lemma A.5). Since $\Phi(\lambda_0)$ has rank $p - q = r$, $\ell_n^*(\mathbf{\Phi})$ can then be efficiently maximised, subject to $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \lambda_0 I_q$, via a reduced rank regression, exactly as in Johansen (1995, Ch. 6).

When $q \geq 2$, some care needs to be taken with the parametrisation of $\mathscr{L}$. If we take this to be either the set of real normal ($\mathscr{L}_{\mathrm{n}}$) or symmetric ($\mathscr{L}_{\mathrm{s}}$) matrices, then each $\Lambda_{\mathrm{LU}} \in \mathscr{L}$ can be expressed as $\Lambda_{\mathrm{LU}} = QD_{\mathrm{LU}}Q^{\mathsf{T}}$, where $Q \in \mathbb{R}^{q\times q}$ is an orthogonal matrix ($Q^{\mathsf{T}}Q = I_q$) and D_{LU} is a block diagonal, with blocks that are either: 1×1 and equal to each of the real eigenvalues of Λ_{LU}, or (2×2) and of the form $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$, if Λ_{LU} has a pair of complex eigenvalues at $\lambda = a \pm \mathrm{i}b$ (Horn and Johnson, 2013, Thm. 2.5.6 and 2.5.8). Since Q can be constructed from $q(q-1)/2$ plane rotations (Horn and Johnson, 2013, Prob. 2.1.P29), both $\mathscr{L}_{\mathrm{n}}$ and $\mathscr{L}_{\mathrm{s}}$ can thus be expressed in terms of of $q(q+1)/2$ free parameters lying in a compact set.

F.2 Nearly optimal tests

For simplicity of exposition, suppose that $p = 2$ and $q = r = 1$, so that $A = a$ and $\Lambda_{\text{LU}} = \lambda_{\text{LU}}$. We want to test

$$H_0 : a = a_0,\ \lambda_{\text{LU}} \in \mathscr{L} \qquad \text{against} \qquad H_1 : a \neq a_0,\ \lambda_{\text{LU}} \in \mathscr{L},$$

where $\mathscr{L} = [\rho, 1]$ for some user-chosen $\rho < 1$. Let F_0 and F_1 denote discrete distributions on $\mathbb{R} \times \mathscr{L}$, that respectively concentrate on those subsets of the parameter space consistent with the null and the alternative. Consider a test of the form

$$\mathcal{NP}_n(a) \coloneqq \mathbf{1}\left\{\int_{\mathbb{R}\times\mathscr{L}} \mathrm{e}^{\ell_n^*(a,\lambda)} F_1(\mathrm{d}a, \mathrm{d}\lambda) > \mathrm{cv}_\alpha \int_{\mathscr{L}} \mathrm{e}^{\ell_n^*(a_0,\lambda)} F_0(a_0, \mathrm{d}\lambda)\right\}$$

To implement the procedure of Elliott et al. (2015), we require estimates of the following probabilities:

(i) the null rejection probability $\mathbb{P}_{(a_0,\lambda)}\{\mathcal{NP}_n(a_0) = 1\}$ for λ in (a discretisation of) $\mathscr{L}$;

(ii) the weighted null rejection probability $\int \mathbb{P}_{(a_0,\lambda)}\{\mathcal{NP}_n(a_0) = 1\} F_0(a_0, \mathrm{d}\lambda)$; and

(iii) the power under the weighted alternative, $\int_{\mathbb{R}\times\mathscr{L}} \mathbb{P}_{(a,\lambda)}\{\mathcal{NP}_n(a_0) = 1\} F_1(\mathrm{d}a, \mathrm{d}\lambda)$.

In view of Theorem 3.1, aside from (a, λ) the only parameter that these probabilities depend on, in large samples, is the long-run covariance matrix $\Omega \coloneqq K\Sigma K^{\mathsf{T}}$.

Suppose for the moment that $\Omega = \Omega_0$ is known. To estimate the probabilities in (i)–(iii), we only need to simulate data from a VAR with the same implied values of a, λ and Ω_0. Fixing (a, λ), consider the bivariate VAR(1) with autoregressive coefficient matrix $\Phi(a, \lambda) \coloneqq R(a)\Lambda(\lambda)L(a)^{\mathsf{T}}$, where

$$R(a) = \begin{bmatrix} R_{\text{LU}}(a) & R_{\text{ST}} \end{bmatrix} = \begin{bmatrix} a & 1 \\ 1 & 0 \end{bmatrix} \qquad \Lambda(\lambda) = \operatorname{diag}\{\lambda, \lambda_{\text{ST}}\} \qquad L(a)^{\mathsf{T}} = \begin{bmatrix} L_{\text{LU}}(a)^{\mathsf{T}} \\ L_{\text{ST}}(a)^{\mathsf{T}} \end{bmatrix} \coloneqq R(a)^{-1},$$

where we may take $\lambda_{\text{ST}} = 0$. The implied quasi-cointegrating relation is $\beta(a)^{\mathsf{T}} = (1, -a)$. Then taking

$$K(a) \coloneqq \begin{bmatrix} \beta(a)^{\mathsf{T}} R_{\text{ST}}(1 - \lambda_{\text{ST}})^{-1} L_{\text{ST}}(a)^{\mathsf{T}} \\ L_{\text{LU}}(a)^{\mathsf{T}} \end{bmatrix}$$

we can ensure that the implied Ω in this model agrees with Ω_0, at the null value of a, by setting the variance matrix Σ of the reduced-form errors to

$$\Sigma_0 \coloneqq K(a_0)^{-1}\Omega_0[K(a_0)^{\mathsf{T}}]^{-1}.$$

Now for $(a, \lambda) \in \mathbb{R} \times \mathscr{L}$, let $Y_n^{(b)}(a, \lambda) \coloneqq \{y_t^{(b)}\}_{t=1}^n$ denote a sample of length n (i.e. of the same length as the observed sample) generated as

$$y_t^{(b)} = \Phi(a, \lambda) y_{t-1}^{(b)} + \Sigma_0^{1/2} w_t$$

where $w_t \sim_{\text{i.i.d.}} N[0, I_p]$, with $y_0^{(b)} = 0$. For each (a, λ) lying on a discrete grid that contains the supports of F_0 and F_1, we generate a total of $B = 20,000$ such samples. Let $\ell_n^*[\mathbf{\Phi}; Y_n^{(b)}(a,\lambda)]$ denote the model loglikelihood (with the deterministic terms concentrated out), computed on the basis of the data $Y_n^{(b)}(a,\lambda)$. For another (or possibly the same) $(a',\lambda') \in \mathbb{R} \times \mathscr{L}$, define

$$\hat{\ell}_n^{(b)}(a',\lambda' \mid a,\lambda) \coloneqq \max_{\{\mathbf{\Phi}\in\mathscr{P} \mid \Lambda_{\mathrm{LU}}(\mathbf{\Phi})=\lambda', A(\mathbf{\Phi})=a'\}} \ell_n^*[\mathbf{\Phi}; Y_{n_0}^{(b)}(a,\lambda)]$$

to be the concentrated loglikelihood at $A(\mathbf{\Phi}) = a'$ and $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) = \lambda'$. Then we can compute

$$\begin{aligned}\mathcal{NP}_n^{(b)}(a_0 \mid a,\lambda) \coloneqq \mathbf{1}\biggl\{\int_{\mathbb{R}\times\mathscr{L}} &\exp\{\ell_n^*(a',\lambda' \mid a,\lambda)\}F_1(\mathrm{d}a',\mathrm{d}\lambda') \\ &> \mathrm{cv}_\alpha \int_{\mathscr{L}} \exp\{\ell_n^*(a_0,\lambda' \mid a,\lambda)\}F_0(a_0,\mathrm{d}\lambda')\biggr\}\end{aligned}$$

as the realisation of the nearly optimal test on the dataset $Y_n^{(b)}(a,\lambda)$. Hence we can estimate the probabilities in (i)–(iii) above by replacing each instance of $\mathbb{P}_{(a,\lambda)}\{\mathcal{NP}_n(a_0) = 1\}$ with

$$\frac{1}{B}\sum_{b=1}^{B} \mathcal{NP}_n^{(b)}(a_0 \mid a,\lambda)$$

for each value of $(a,\lambda) \in \mathbb{R}\times\mathscr{L}$ (in practice, for a discrete subset thereof).

Finally, since Ω_0 is unknown, it needs to be consistently estimated. To that end, we recognise that for $\varepsilon_{\mathrm{LU},t} = L_{\mathrm{LU}}^{\mathsf{T}}\varepsilon_t$

$$\Omega = K\Sigma K^{\mathsf{T}} = \mathrm{lrvar}\left(\begin{bmatrix}\beta^{\mathsf{T}}x_t \\ \varepsilon_{\mathrm{LU},t}\end{bmatrix}\right),$$

where the final equality follows in particular by observing that

$$\begin{aligned}\sum_{l=-\infty}^{\infty}\mathbb{E}(\beta^{\mathsf{T}}x_t)\varepsilon_{\mathrm{LU},t-l}^{\mathsf{T}} &= \beta^{\mathsf{T}}\varepsilon_t\varepsilon_{\mathrm{LU},t} + \beta^{\mathsf{T}}R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^k\sum_{l=1}^{\infty}\mathbb{E}z_{\mathrm{ST},t-1}\varepsilon_{\mathrm{LU},t-l}^{\mathsf{T}} \\ &= \beta^{\mathsf{T}}\left[I + R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^k\sum_{l=0}^{\infty}\Lambda_{\mathrm{ST}}^l L_{\mathrm{ST}}^{\mathsf{T}}\right]\Sigma L_{\mathrm{LU}} \\ &=_{(1)} \beta^{\mathsf{T}}R_{\mathrm{ST}}\left[\Lambda_{\mathrm{ST}}^{k-1} + \Lambda_{\mathrm{ST}}^k(I-\Lambda_{\mathrm{ST}})^{-1}\right]L_{\mathrm{ST}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}} \\ &= \beta^{\mathsf{T}}R_{\mathrm{ST}}\Lambda_{\mathrm{ST}}^{k-1}(I-\Lambda_{\mathrm{ST}})^{-1}L_{\mathrm{ST}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}} \\ &=_{(2)} \beta^{\mathsf{T}}R_{\mathrm{ST}}(I-\Lambda_{\mathrm{ST}})^{-1}L_{\mathrm{ST}}^{\mathsf{T}}\Sigma L_{\mathrm{LU}},\end{aligned}$$

where $=_{(1)}$ follows by $R\Lambda^{k-1}L^{\mathsf{T}} = I_p$ and $\beta^{\mathsf{T}}R_{\mathrm{LU}} = 0$, and $=_{(2)}$ by $R\Lambda^i L^{\mathsf{T}} = 0$ for $i \in \{0,\ldots,k-2\}$, which are themselves implied by $\mathbf{R}\mathbf{L}^{\mathsf{T}} = I_{kp}$ (see Lemma A.1). Hence we can estimate Ω by computing ML estimates of β and L_{LU} (under only the restriction that $\Lambda_{\mathrm{LU}}(\mathbf{\Phi}) \in \mathscr{L}$), and then computed an estimate (after demeaning and detrending) of the long-run covariance matrix of $\hat{\beta}^{\mathsf{T}}y_t$ and $\hat{L}_{\mathrm{LU}}^{\mathsf{T}}\hat{\varepsilon}_t$, where $\hat{\varepsilon}_t$ denote the VAR residuals.