Paolo Paruolo

Common features and common I(2) trends in VAR systems

2002/29

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

Printed in Italy in October 2002
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

Common trends and cycles in I(2) VAR systems*

Paolo Paruolo
Department of Economics, University of Insubria
Via Ravasi 2, 21100 Varese, Italy

July 31, 2003

Abstract

This paper discusses common cycles in I(2) vector autoregressive (VAR) systems. Both static and dynamic cofeatures are considered. We consider application of these notions to different choices of stationary variables extracted from a VAR, including deviations from equilibria. This extension is based on the equilibrium dynamics representation of the system, which is introduced in this paper.

Inference on the number of common features is addressed via reduced rank regression, as well as estimation of the cofeature relations and testing. An application to Australian prices illustrates the techniques presented in the paper. In the empirical application it is found that the deviation from one of the equilibria is an innovation process, whereas only trivial cases of cofeatures can be obtained for the equilibrium correction form.

Keywords: Common features, Cofeatures, Cointegration, Common trends, Common cycles, VAR, I(2), Reduced rank regression.
JEL code: C32, C51, C52.

*Invited paper at the conference 'Common features in Rio', Rio de Janeiro 28-31 July 2002. Partial financial support from Italian MIUR grants ex 60% and Cofin2002 is gratefully acknowledged. I wish to thank the participants of the conference for useful comments and in particular Alain Hecq for signaling the usefulness of a proof on the asymptotic distribution of the reduced rank regression test on cofeature rank, see Appendix B of this paper. I have also received useful comments on the first version of this paper (still available as working paper 2002/29 at `http://eco.uninsubria.it`) from Søren Johansen, the guest editors and two anonymous referees. The usual disclaimer applies.

Contents

1 Introduction

The notion of common factors is a classical idea in statistics. During the last two decades it has received new momentum in econometrics, thanks to the introduction of concepts like cointegration, see Engle and Granger (1987). The duality between cointegration and equilibrium correction models has proved a powerful tool for modelling, contributing significantly to the success of the concept.

The notion of common features, introduced by Vahid and Engle (1993) and Engle and Kozicki (1993), provided an even broader concept, which contains cointegration as a special case. Among the new areas of application, Vahid and Engle (1993) and Engle and Kozicki (1993) introduced non-innovation common features, i.e. common cycles. This notion is related to codependence, defined in Gourieroux and Peaucelle (1993).

The interplay between common trends and common cycles has been considered in Vahid and Engle (1993), who treated the case of processes integrated of order 1, I(1). I(1) systems have also been the focus of much of the ensuing literature on common features. Examples are Kugler and Neusser (1993), Vahid and Engle (1997), Vahid and Issler (2002), Hecq et al. (2000, 2002), Cubadda (1999, 2001), Cubadda and Hecq (2001) inter alia.

No contributions in the literature so far appear to investigate the interplay between common trends and cycles for systems integrated of order 2, I(2); this is focus of the present paper. I(2) systems have been analyzed in Johansen (1992a, 1995a, 1997), Stock and Watson (1993), Boswijk (2000) inter alia. These systems present a more complex, and inherently richer, structure. As for I(1) processes, I(2) systems have a dual error correction formulation, which however includes both integral and proportional control terms, see Haldrup (1998) for a survey and references.

In this paper it is shown how the corresponding equilibrium correction form is the basis for the analysis of common cycles in the second differences of the process. An additional representation is introduced, called the 'equilibrium dynamics' form, which is equivalent to the equilibrium correction one. This formulation is the basis for the analysis of common cycles in deviations from equilibria, i.e. in the cointegration relations.

Cofeature relations in the second differences of the variables are shown to represent second increments in the I(2) common trends. Cofeature relations in deviations from equilibria represent instead unpredictable cointegration relations. Unpredictable deviations from equilibria are expected under rational expectations; the associated adjustment can be interpreted as reactions to the unpredicted part of the equilibrium relation. Both applications of common features shed light on different features of the system.

The paper discusses both contemporaneous (static) and asynchronous (dynamic) common cycles. The lack of invariance of the notion of contemporaneous cofeatures can be overcome by allowing the cofeature relations to include lagged variables. This augmented notion is called 'dynamic cofeatures'; it is discussed both for the equilibrium correction and the equilibrium dynamics forms. Dynamic cofeatures correspond to polynomial cofeatures defined in Cubadda and Hecq (2001) when applied to the levels of the variables in the system.

As for I(1) systems, the notion of non-innovation common features is directly related to rank deficiency of some autoregressive coefficient matrices. This holds both for static and dynamic cofeatures, and provides a unified framework for inference.

When the cointegration parameters are known, this analysis can be based on reduced rank regression, see Anderson (1951).

In this paper it is shown that the same locally asymptotically normal (LAN) results apply once the integration indices (cointegration ranks) have been determined, and the cointegration parameters have been substituted with their maximum likelihood (ML) estimates or the two stage I(2) estimates (2SI2) of Johansen (1995a). This follows from the superconsistency of the cointegration parameters. Hence it is suggested to first test for cointegration and next for common features.

The possibility to fix the cointegration parameters at their estimated values permits to address inference both on the equilibrium correction form and the equilibrium dynamics form in a unified way. Other representations of the system which involve $T^{1/2}$-consistent parameters do not share this property.

We address inference on common features by likelihood-based techniques developed for reduced rank regression. These are also applied in nested reduced rank regression, see Ahn and Reinsel (1988) and in the scalar component models by Tiao and Tsay (1989). We show that the reduced rank regression model can be used to test for the number of cofeatures, as well as to test hypothesis on the specification of the cofeature vectors. This allows to develop a specification search for the cofeature vectors similar to the one for simultaneous systems of equations.

The techniques proposed in the paper are illustrated on the Australian prices data-set analyzed in Banerjee et al. (2001). For these data, the newly introduced equilibrium dynamics form supports the presence of a single cofeature vector, while only trivial cases of cofeatures can be obtained for the equilibrium correction form. This shows the empirical relevance of the proposed notions.

The rest of the paper is organized as follows: Section 2 reports notation and definitions. Section 3 reports the various representations of I(2) systems and introduces the equilibrium dynamics form. The application of common features to contemporaneous variables in treated in Section 4. Section 5 extends the notion to dynamic common features. Proofs of propositions in all these sections are reported in Appendix A.

Section 6 develops inference on common features in a unified way through reduced rank regression techniques. Proofs of this section are reported in Appendix B. Section 7 contains an application to Australian prices; Subsection 7.1 reports the cointegration analysis while Subsection 7.2 contains the common feature analysis. Section 8 concludes.

In the following $a := b$ and $b =: a$ indicate that a is defined by b; $(a : b)$ indicates the matrix obtained by horizontally concatenating a and b. e_i indicates the i-th column of the identity matrix. For any full column rank matrix H, $\text{col}(H)$ is the linear span of the columns of H, $\bar{H}$ indicates $H(H'H)^{-1}$ and $H_\perp$ indicates a basis of $\text{col}^\perp(H)$, the orthogonal complement of $\text{col}(H)$. $P_H = \bar{H}H' = H\bar{H}'$ is the orthogonal projector matrix onto $\text{col}(H)$. vec is the column stacking operator and $A \otimes B := [a_{ij}B]$ defines the Kronecker product. Finally $\xrightarrow{p}$ and $\xrightarrow{d}$ indicate convergence in probability and in distribution respectively.

All processes W_t are understood to be multivariate, i.e. of dimension $q \times 1$, $W_t = (W_{1t} : ... : W_{qt})'$. Individual time series, or linear combinations thereof, are called 'components' of the process. As usual, we reserve the word 'process' for the case where the probabilistic structure of W_t is known, and use the word 'model' to indicate a class of processes indexed by some parameter vector to be estimated.

2 Notation and definitions

In this section we introduce general notation and definitions. We follow Johansen (1996) Chapter 3, for the definition of (co)integration and Engle and Kozicki (1993), Vahid and Engle (1993) for the general definition of non-innovation common features. This section also introduces a few additional concepts that are needed in the present context, as the notion of I(0) rank.

We consider VAR(k), $k \geq 2$, systems of the type

$$X_t = \sum_{i=1}^{k} A_i X_{t-i} + \widetilde{\epsilon}_t, \tag{1}$$

where X_t and $\widetilde{\epsilon}_t = \mu_1 t + \mu_0 + \mu_d d_t + \epsilon_t$ are $p \times 1$. t, 1, d_t are the deterministic components, $d_t := (d_{1,t} : ..d_{r-1,t})'$ is a vector of seasonal dummies 'orthogonal' to the constant, i.e. of the form $d_{i,t} = 1(t \bmod r = i) - 1/r$, $1(\cdot)$ is the indicator function and r is the number of seasons. L and $\Delta := 1 - L$ are the lag and difference operators, where negative powers of Δ indicate summation. The innovations ϵ_t are assumed to i.i.d. $N(0, \Omega)$, where $E(\epsilon_t | F_{t-1}) = 0$, with F_{t-1} the sigma-field generated by X_{t-i}, $i \geq 1$.[1]

As it is well known, roots of the autoregressive polynomial $A(z) = I - \sum_{i=1}^{k} A_i z^i$ at $z = 1$ are responsible for the presence of common trends of the random-walk type. We assume, 'Assumption 1', that, apart from roots at $z = 1$, all other roots of $A(z)$ are outside the unit circle, i.e. of the stationary type. Hence (1) may generate random walk-type stochastic trends as well as stationary processes.

Stationary processes derived from $X_t - E(X_t)$ in (1) have a moving average representation $W_t = C_W(L)\epsilon_t$, where $C_W(z) := \sum_{i=0}^{\infty} C_{W,i} z^i$ is summable for $|z| < 1 + \varkappa$ and $\varkappa > 0$, i.e. they are linear processes. The first coefficient matrix $C_{W,0}$ is assumed to have full row-rank, but not necessarily to be equal to the identity matrix. When a linear process has sum of coefficients $C_W(1)$ different from the 0 matrix, then the linear process is called integrated of order 0, I(0); see Johansen (1996). In the following an I(0) process W_t is said to have rank q if $\text{rank}(C_W(1)) = q$. One can observe that the row-rank deficiency of $C_W(1)$ is associated with the presence of cointegration in $\Delta^{-1} W_t$, see Johansen (1996) Chapter 3.

Unit roots generate integrated processes of different order. A process X_t is said to be integrated of order d, I(d), if $\Delta^d X_t - E(\Delta^d X_t)$ is I(0), $d = \pm 1, \pm 2, ...$ In the case of an integrated system X_t of order 1, X_t is said to be cointegrated if there exist linearly independent vectors $\beta := (\beta_1 : ... : \beta_{p_0})$ such that the linear combinations $W_t = \beta' X_t - E(\beta' X_t)$ are stationary. The system X_t is said to be cointegrated with rank p_0, or equivalently, with $p - p_0$ common trends. For I(1) VAR processes (1), $W_t = \beta' X_t - E(\beta' X_t)$ can indeed be shown to be I(0); this is shown in Johansen's proof of Granger's representation theorem, see Johansen (1996), Theorem 4.2.

I(0) processes $W_t = C_W(L)\epsilon_t$ have j-th autocovariance

$$E(W_t W'_{t+j}) = \sum_{i=0}^{\infty} C_{W,i} \Omega C'_{W,i+j}, \tag{2}$$

[1] Other deterministic terms could also be incorporated. The innovations could be assumed to be a martingale difference process with respect to F_t with third moments. These generalizations are not pursued here for simplicity. The assumption $k \geq 2$ is common to the literature on I(2) systems.

which is in general different from 0. A special case of I(0) processes is an 'innovation process', which is defined as an I(0) process $W_t = C_W(L)\epsilon_t$ with $C_W(L) = C_{W,0}$, a contant matrix, assumed to be a full row-rank, and with positive definite covariance matrix $E(W_tW_t') = C_{W,0}\Omega C_{W,0}'$. As it is well known, innovation processes presents zero autocovariances because $C_{W,i} = 0$ for $i = 1, 2, ...$ in (2). In the following we abbreviate 'innovation process' with 'innovation'. With a slight abuse of language, we will also refer to any linear process that is not an innovation process as a 'cycle'.[2]

If a process W_t is I(0) with rank q but it is not an innovation process, it contains cycles. If there exist some non-zero vector b_i such that $b_i'W_t$ is an innovation process, then the system is said to present non-innovation common features, or common cycles, and b_i is called a cofeature vector. When there exist ℓ linearly independent cofeature vectors $b_1, ..., b_\ell$, then $b := (b_1 : ... : b_\ell)$ is called the cofeature matrix, and the systems is said to have cofeature rank ℓ. Equivalently, W_t is said to present $q - \ell$ common I(0) cycles. In the following non-innovation common features will be abbreviated into 'common features'.

Implicit in this notation is the notion that the maximum number of I(0) cycles is given by the rank of the I(0) process. We state this result as a proposition for later reference. This result is parallel to Theorem 1 in Vahid an Engle (1993) for I(1) systems, although it applies more generally; in particular it holds also for I(2) systems. The proofs is reported Appendix A.

Proposition 1 (upper bound on cofeature rank) *A $p \times 1$ I(0) process W_t with rank $q \le p$ presents at most q innovation processes; hence the cofeature rank ℓ is bounded by q, $\ell \le q$.*

When $q < p$, the remaining $p - q$ components of an I(0) process with rank q are integrated of negative order. Processes integrated of negative order are cyclic, and they cannot be innovation processes because they are not I(0) processes. Hence nothing can be said about the commonality in the remaining $p - q$ directions. This point is further discussed in Section 5 with respect to the special case of I(2) systems.

The notions of cointegration and common features describe the possibility that trends and cycles in economic systems may be shared by different component time series. Although the two notions refer to different aspects of the same system, i.e. the non-stationary and stationary ones, they have some interplay. In an I(1) system X_t, in fact, the stationary variables include not only the first differences ΔX_t, but also the cointegration relations $\beta' X_t$. For more on the interplay of the two notions in I(1) systems we refer to Vahid and Engle (1993) and to a companion paper, Paruolo (2003).

The following sections review the representation of common trends in I(2) systems and discuss several possible applications of common cycles. Before stating general results, we introduce two simple bivariate motivating examples that will be used to illustrate the various applications of common features.

Example 2 (real interest rates) *Consider a bivariate system $X_t := (X_{1t} : X_{2t})'$ run by i.i.d. innovations $\eta_t := (\eta_{1t} : \eta_{2t})'$, defined by the equations*

$$\begin{cases} X_{1t} & = & \Delta X_{2t} + \eta_{1t} \\ \Delta^2 X_{2t} & = & c_t \end{cases} \qquad \textit{where} \qquad c_t = \varrho c_{t-1} + \eta_{2t}, \qquad |\varrho| < 1.$$

[2] The word 'innovation' is preferred to 'contemporaneous white noise' as argued in Ericsson in his comments to Engle and Kozicki (1993).

Here c_t represents a cycle (for $\varrho \neq 0$), which we call the 'business cycle'. One could interpret X_{2t} as the log of the price level and X_{1t} as the level of the nominal interest rate. The second equation states that the growth of the inflation rate is proportional to the business cycle c_t, while the first equation states that the (ex-post) real interest rate $X_{1t} - \Delta X_{2t} = \eta_{1t}$ is stationary, and in particular an innovation process. Observe that the system is I(2) because X_{2t} needs second differences to become I(0); this is equivalent to saying that the inflation rate ΔX_{2t} is I(1).

From the first equation one sees that the nominal interest rate X_{1t} is I(1) and cointegrated with ΔX_{2t}. Hence also the increments of the interest rate $\Delta X_{1t} = \Delta^2 X_{2t} + \Delta\eta_{1t} = c_t + \Delta\eta_{1t}$ are affected by the business cycle c_t, which is common to the 2 variables in the system. One thus wishes to adopt a notion of common features that, when applied to this system, would indicate the presence of a common cycle.

Example 3 (profitability) *Consider a bivariate system $X_t := (X_{1t} : X_{2t})'$ run by i.i.d. innovations $\epsilon_t := (\epsilon_{1t} : \epsilon_{2t})'$, defined by the equations*

$$\left\{ \begin{array}{lcll} \Delta X_{1t} & = & & -\frac{1}{2}\Delta^2 X_{2t-1} + \epsilon_{1t} \\ \Delta^2 X_{2t} & = & \frac{1}{2}\Delta X_{1t-1} & +\frac{1}{4}\Delta^2 X_{2t-1} + \epsilon_{2t} \end{array} \right. . \tag{3}$$

The second variable X_{2t} can again be interpreted as the log of prices, while the first one X_{1t} could represent firms' profitability, which is negatively related to lagged inflation. Define $Y_t := (Y_{1t} : Y_{2t}) = (\Delta X_{1t} : \Delta^2 X_{2t})$, and observe that $Y_t = A_1^\circ Y_{t-1} + \epsilon_t$ is a VAR(1). The AR matrix

$$A_1^\circ = \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix} \tag{4}$$

has eigenvalues $\frac{1}{8} \mp \frac{1}{8} i \sqrt{15}$, where i is the imaginary unit, both with modulus equal to $1/2$. Hence the system for Y_t is stable. We deduce that Y_t is I(0), i.e. that X_{1t} is I(1) and X_{2t} is I(2). Both processes present some cyclic component. The cyclic component is associated with the A_1° matrix, where ΔX_{1t} is influenced by $\Delta^2 X_{2t-1}$ and $\Delta^2 X_{2t}$ is influenced by ΔX_{1t-1}, i.e. both variables receive some feedback from the other one. In this example it is not apparent if there are common cycles. In the following it will be argued that one common cycle can be defined, for an appropriate application of common features.

3 Common trends

In this section we review three representations for I(2) systems, the common trends representation, the equilibrium correction formulation and the equilibrium dynamics form. All these formulations are needed in order to discuss common features in I(2) systems.

In particular we summarize the I(2) Representation Theorem by Johansen (1992a), which describes the I(2) conditions and the common trends structure of these systems. The I(2) restrictions lead to equilibrium correction formulations; here we use the one introduced in Paruolo and Rahbek (1999). Finally this section introduces a novel representation of I(2) systems, which we call the equilibrium dynamics form.

It is convenient to rewrite $A(L)$ as $A(L) = -\Pi L - \Gamma\Delta L + \Upsilon(L)\Delta^2$, i.e. consider[3]

$$\begin{aligned}\Delta^2 X_t &= \Pi X_{t-1} + \Gamma\Delta X_{t-1} + \sum_{i=1}^{k-2}\Upsilon_i\Delta^2 X_{t-i} + \widetilde{\epsilon}_t \\ &= \Pi X_{t-1} + \Gamma\Delta X_{t-1} + \gamma V_t + \widetilde{\epsilon}_t\end{aligned} \tag{5}$$

for $k \geq 2$, where $\gamma := (\Upsilon_1 : ... : \Upsilon_{k-2})$, $V_t := (\Delta^2 X'_{t-1} : ... : \Delta^2 X'_{t-k+2})'$. The matrices Π and Γ are key elements in the characterization of the presence of I(2) variables in the system, as illustrated in the following assumptions. Let $\Upsilon := \sum_{i=1}^{k-2}\Upsilon_i$.

I(2) conditions

(a). Assumption 1 holds;

(b). $\Pi = \alpha\beta'$, where α and β are $p \times p_0$ matrices of full rank $p_0 < p$;

(c). $P_{\alpha_\perp}\Gamma P_{\beta_\perp} = \alpha_1\beta_1'$ where α_1 and β_1 are $p \times p_1$ matrices of full rank $p_1 < p - p_0$, or, equivalently, $\alpha'_\perp\Gamma\beta_\perp = \xi\eta'$ where $\xi = \alpha'_\perp\alpha_1$ and $\eta = \beta'_\perp\beta_1$ are $p - p_0 \times p_1$ matrices of full rank $p_1 < p - p_0$;

(d). $\alpha_2'\theta\beta_2$ has full rank $p - p_0 - p_1$, where $\alpha_2 = (\alpha : \alpha_1)_\perp$, $\beta_2 = (\beta : \beta_1)_\perp$ and θ is defined as

$$\theta := (\Gamma - \Pi)\bar{\beta}\bar{\alpha}'(\Gamma - \Pi) + I - \Upsilon; \tag{6}$$

(e). $\mu_1 = \alpha\beta_0'$, with β_0' a $p_0 \times 1$ vector;

(f). $\alpha'_\perp\mu_0 = \xi\eta_0' + \alpha'_\perp\Gamma\bar{\beta}\beta_0'$, with η_0' a $p_1 \times 1$ vector.

In the following 'I(2) assumptions' and 'I(2) conditions' are used as synonyms. Johansen's I(2) representation theorem, see Johansen (1992a) or (1996) Theorem 4.6, establishes under (a) that necessary and sufficient conditions for $\Delta^2 X_t$, $\beta' X_t + \delta\beta_2'\Delta X_t$, $\beta_1'\Delta X_t$ to be stationary, apart from deterministic components and initial values, are the conditions (b) to (d).

The dimensions p_0, p_1 and p_2 are called the integration indices of the system, $p = p_0 + p_1 + p_2$. They correspond to the ranks of β, β_1 and β_2; β are the linear combinations of the levels that cointegrate with the differences (polynomial cointegration) in $\beta' X_t + \delta\beta_2'\Delta X_t$. β_1 are the extra linear combinations that reduce the order of integration from 2 to 1, but that do not cointegrate with the differences, $\beta_1' X_t$. Finally β_2 are the remaining orthogonal directions, which are dominated by the I(2) component.

[3] One can use the decomposition, $B(z) = B(1) + B^*(z)(1-z)$ where $B(z)$ and $B^*(z)$ have the same radius of convergence. We use here the variant $B(z) = B(1)z + B^\circ(z)(1-z)$ where $B^\circ(z) = B(1) + B^*(z)$. Apply it once to $A(z) = A(1)z + A^\circ(z)(1-z)$ and once to $A^\circ(z) = A^\circ(1)z + A^{\circ\circ}(z)(1-z)$ to find

$$\begin{aligned}A(L) &= A(1)L + (A^\circ(1)L + A^{\circ\circ}(L)(1-L))(1-L) = \\ &= A(1)L + A^\circ(1)L(1-L) + A^{\circ\circ}(L)(1-L)^2.\end{aligned}$$

In presence of a constant and trend, it can be proved, see Rahbek (1997), that under (a), X_t is I(2) and presents linear trends in all directions iff the conditions (b) to (f) hold, see also Paruolo (2002b) who restated this result with the inclusion of dummies. Under the I(2) conditions $\Delta^2 X_t$, $Y_{0,t} := \beta' X_t + \delta\beta_2' \Delta X_t + \beta_0' t$, $Y_{1,t} := \beta_1' \Delta X_t$ are shown to be I(0), and the following common trends representation holds

$$X_t = C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + C_0(L)\epsilon_t + m_0 + m_1 t + A + Bt + m(L)d_t, \quad (7)$$

where the I(2) component is $C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} \epsilon_i$ and the I(1) component is $C_1 \sum_{i=1}^{t} \epsilon_i$. We observe that the I(1) and I(2) trends are built from cumulated ϵ_i, i.e. that they have i.i.d. first and second increments.

In (7) m_0, m_1 do not depend on initial values while A and B do, $m(z)$ is a polynomial of degree equal to the number of seasons and $C_0(L)\epsilon_t$ is a linear process with exponentially decreasing coefficients. The reduced rank matrix $C_2 = \beta_2(\alpha_2'\theta\beta_2)^{-1}\alpha_2'$ induces p_2 common I(2) trends in the system. In the remaining directions, $(\beta : \beta_1)' X_t$, there are I(1) trends, which are cancelled when considering the polynomial cointegrating relation $Y_{0,t} := \beta' X_t + \delta\beta_2' \Delta X_t + \beta_0' t$. For complete definitions of the matrix C_1 and of other expressions in (7) we refer e.g. to Paruolo (2002b).

The I(2) common trends representation implies that $\Delta^2 X_t$ is an I(0) process with rank p_2. Take in fact second differences in (7) and let $m_t^* := m(L)\Delta^2 d_t$; one obtains

$$\Delta^2 X_t =: C^*(L)\epsilon_t + m_t^* = C_2\epsilon_t + C_1\Delta\epsilon_t + C_0(L)\Delta^2\epsilon_t + m_t^*. \quad (8)$$

This equation shows that $C^*(1) = C_2$, which is of rank p_2. Hence $\Delta^2 X_t$ is an I(0) process with rank p_2, in the notation introduced in the previous section.

If the I(2) conditions hold, then system (5) can be rewritten in many equilibrium correction forms. The following one was introduced in Paruolo and Rahbek (1999), and will be employed in the following,

$$\Delta^2 X_t = \alpha[Y_{0,t-1}] + (\zeta_1 : \zeta_2)[(\beta : \beta_1)'\Delta X_{t-1}] + \gamma V_t + \mu D_t + \epsilon_t = \Psi U_t + \mu D_t + \epsilon_t, \quad (9)$$

where $\mu_1 = \alpha\beta_0'$ and $\mu := (\mu_0 : \mu_d)$, $D_t := (1 : d_t')'$, $\Psi := (\alpha : \zeta_1 : \zeta_2 : \gamma)$, $U_t := (Y_{0t-1}' : \Delta X_{t-1}'(\beta : \beta_1) : V_t')'$. The terms in square brackets in (9) are I(0) by Johansen's I(2) representation theorem.[4] In the following we let $\zeta := (\zeta_1 : \zeta_2)$. The equilibrium correction formulation (9) shows how the stationary cointegration relations in square brackets affect the acceleration rate of all variables $\Delta^2 X_t$ through the adjustment coefficients α, ζ. These equations emphasize the correction of the variables $\Delta^2 X_t$ towards equilibrium.

A final representation is the one that defines the dynamics of the stationary cointegration relations themselves. We consider $Y_{0t} := \beta' X_t + \delta\beta_2'\Delta X_t + \beta_0' t$, $Y_{1t} := \beta_1'\Delta X_t$, $Y_{2t} := \beta_2'\Delta^2 X_t$ as the stationary variables of interest, where $Y_t := (Y_{0t}' : Y_{1t}' : Y_{2t}')'$ is $p \times 1$.[5] The dynamic equations for Y_t can be obtained by rearranging (9) to reproduce the chosen stationary l.h.s. variables, as in

[4]In order to satisfy condition (f) in the I(2) conditions, the coefficient μ_0 should be constrained. However, for simplicity, in the statistical analysis of (9) μ_0 is estimated unrestrictedly.

[5]Other choices are possible. In particular one may wish to discuss the dynamics of $\beta'\Delta X_t$ instead of Y_{0t}, or $b_2'\Delta X_t$ in place of Y_{2t}, with $b_2'\beta_2$ of full rank. The present choice is representative of other ones, and illustrates the class of possible definitions of common features.

Theorem 4 (equilibrium dynamics representation) *Under the $I(2)$ conditions, the following equilibrium dynamics representation holds for $Y_t := (Y_{0,t}' : Y_{1,t}' : Y_{2,t}')'$ as defined above*

$$Y_t = \sum_{i=1}^{k} A_i^\circ Y_{t-i} + \mu^\dagger D_t + \epsilon_t^\circ \tag{10}$$

where $\mu^\dagger := (\mu_0^\dagger : \mu_d^\circ)$, $\mu_0^\dagger := D(\mu_0 - \zeta_1 \beta_0')$, $(\mu^\circ : \epsilon_t^\circ) := D(\mu : \epsilon_t)$, $D := (\beta + \beta_2\delta' : \beta_1 : \beta_2)'$. The AR polynomial $A^\circ(L) := I - \sum_{i=1}^{k} A_i^\circ L^i$ is stable, i.e. has all characteristic roots outside the unit circle, and can be inverted to give

$$Y_t = C^\circ(L)(\mu^\dagger D_t + \epsilon_t^\circ)$$

where $C^\circ(L)\epsilon_t^\circ = C_Y(L)\epsilon_t$ is a $I(0)$ linear process with rank p, where $C_{Y,0} = D := (\beta + \beta_2\delta' : \beta_1 : \beta_2)'$, a full rank matrix.

Let the AR matrices A_i° be partitioned column-wise conformably with Y_t, i.e. let $A_{i,j}^\circ$ be the $p \times p_j$ block that multiplies $Y_{j,t-i}$, $j = 0, 1, 2$. The last AR matrices A_i° in (10) *are constrained as follows*

$$A_{k,0}^\circ \delta = -A_{k-1,2}^\circ, \qquad (A_{k,1}^\circ : A_{k,2}^\circ) = 0. \tag{11}$$

(The specific definition of the A_i° matrices is given in the proof in Appendix A.) The constraints (11) *can be incorporated in* (10) *by substituting Y_{t-k}, $Y_{2,t-k+1}$ with $\Delta\beta' X_{t-k+1}$ on the r.h.s., obtaining*

$$Y_t = \Psi^\circ U_t^\circ + \mu^\ddagger D_t + \epsilon_t^\circ, \tag{12}$$

where $U_t^\circ := \left(Y_{t-1}' : \dots : Y_{0,t-k+1}' : Y_{1,t-k+1}' : \beta'\Delta X_{t-k+1}'\right)'$ contains the elements in the lags of Y_t with unrestricted coefficients, which are collected in Ψ°, $\mu^\ddagger := (\mu_0^\ddagger : \mu_d^\circ)$, $\mu_0^\ddagger := \mu^\dagger - D\Upsilon_{k-2}\bar{\beta}\beta_0'$. Alternatively Y_t also satisfies the following equations, which contain the same r.h.s. variables as the equilibrium correction form (9)*:*

$$Y_t = (\alpha^\circ : \zeta_1^\circ : \zeta_2^\circ) \begin{pmatrix} Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} + \gamma^\circ V_t + \mu^\circ D_t + \epsilon_t^\circ, \tag{13}$$

where $(\alpha^\circ : \zeta_1^\circ : \zeta_2^\circ : \gamma^\circ) := D(\alpha + \bar{\beta} : \zeta_1 + \bar{\beta} : \zeta_2 + \bar{\beta}_1 : \gamma)$. In the following this is called the 'mixed form'.

The equilibrium dynamics representation (12), (13) *and the equilibrium correction representation* (9) *are equivalent, in the sense that any pair can be derived from the other one. Note that when $k = 2$, the mixed form* (13) *and the constrained equilibrium dynamics representation* (12) *coincide.*

The equations (10) describe the dynamics of the equilibrium relations. In the following we call it the 'equilibrium dynamics form'. Note also that rank of the I(0) process Y_t is equal to the dimension p of the process X_t (and Y_t). Hence the transformation from $\Delta^2 X_t$ to Y_t allows to express all cycles as I(0) cycles. Observe also that Ψ in (9) and Ψ° in (12) are not similar and possibly have different ranks.

One may wonder if some cointegrating relation, i.e. some element of Y_{0t} or Y_{1t}, is an innovation process. Equilibria are often defined by rational expectations arguments, which imply that deviations from equilibria must be unanticipated, i.e. innovation processes. Given that the cointegrating relations represent deviations from equilibria, this would indeed be a test of rational expectations, which is of interest in its own right. Quite obviously, the equilibrium dynamics equation (10) is the representation best suited to address this type of question.

The equilibrium dynamics (10) or the equilibrium correction form (9) can be used to discuss serial correlation common features within I(2) systems. This issue is addressed in the following section.

4 Common features

This section discusses the application to I(2) systems of common features, as defined in Section 2. The notion is applied both to $\Delta^2 X_t$ and Y_t, where Y_t has been defined in Section 3, see (10). In this section the relative merits of these options are discussed theoretically and with respect to the two example introduced at the end of Section 2.

In the following we will indicate with W_t either $\Delta^2 X_t$, Y_t, or any other set of p linear combinations of X_t and its first $k-1$ lags.[6] One other possible choice would be $(\Delta X_t'\beta : \Delta X_t'\beta_1 : \Delta^2 X_t'\beta_2)'$. Following the proof of Theorem 4, it is possible to shown that $(\Delta X_t'\beta : \Delta X_t'\beta_1 : \Delta^2 X_t'\beta_2)'$ is an I(0) system with rank $p_1 + p_2$.

A matrix b, of dimension $p \times \ell$ and rank ℓ, is defined to be a cofeature matrix for W_t if $b'(W_t - E(W_t))$ is an innovation process, where $W_t - E(W_t)$ is a $p \times 1$ I(0) process of rank q. We say that b is a cofeature matrix for W_t with cofeature rank ℓ.

As stated in Proposition 1, the cofeature rank ℓ is bounded by the I(0) rank q, i.e. $\ell \leq q$. As noted in equation (8), $\Delta^2 X_t$ is an I(0) process of rank p_2, while Y_t is an I(0) process of rank p, see Theorem 4. Hence the upper bound $\ell \leq q$ is more restrictive for the choice $W_t = \Delta^2 X_t$ than $W_t = Y_t$ because $p_2 \leq p$. The option $W_t = (\Delta X_t'\beta : \Delta X_t'\beta_1 : \Delta^2 X_t'\beta_2)'$ is intermediate, because this process has I(0) rank $p_1 + p_2$, where $p_2 \leq p_1 + p_2 \leq p$. This intermediate option is not discussed further, and we concentrate on $\Delta^2 X_t$ and Y_t.

The existence of a cofeature matrix b for $\Delta^2 X_t$ or Y_t is linked to a rank reduction of the coefficient matrices in the equilibrium correction or equilibrium dynamics representations respectively.

Proposition 5 (cofeatures and reduced rank) *W_t presents common feature with cofeature rank ℓ if and only if $\Psi^{(\cdot)}$ is of reduced rank, where $\Psi^{(\cdot)} = \Psi$ in (9) for the choice $W_t = \Delta^2 X_t$, and $\Psi^{(\cdot)} = \Psi^{\circ}$ in (10) for the choice $W_t = Y_t$. The reduced rank condition rank$(\Psi^{(\cdot)}) = p - \ell$ can be written $\Psi^{(\cdot)} = \varphi\tau'$, with φ and τ of full column rank $s := p - \ell$. In this case the cofeature matrix b can be chosen equal to $\varphi_{\perp}$.*

We next discuss the various choices for W_t in more detail. Consider first the choice $W_t = \Delta^2 X_t$. The characteristics of the cofeature matrix b are described in the following representation theorem, which is parallel to Proposition 1 of Vahid and Engle (1993) for I(1) systems.

Theorem 6 (common cycles representation) *Under the I(2) conditions, there exist a cofeature matrix b such that $b'(\Delta^2 X_t - E(\Delta^2 X_t)) = b'\epsilon_t$ if and only if in (7)*

[6] Let $W_t = X_t - \sum_{i=1}^{k} A_i X_{t-i}$; equation (1) implies that $W_t - E(W_t)$ is an innovation process. We hence exclude this trivial case.

or (8) *one has*

$$\begin{aligned} b'C_{0,i} &= 0, \qquad i = 0,1,2,... && (14)\\ b'C_1 &= 0. && (15)\end{aligned}$$

When (14) *holds, the second condition* (15) *holds if and only if any of the following equations holds*

$$\begin{aligned} b'C_2 &= b' && (16)\\ b &= \alpha_2 c_\perp \qquad \alpha_2'(I-\theta)\beta_2 = cd', && (17)\end{aligned}$$

where c *and* d *are* $p_2 \times p_2 - \ell$ *matrices of rank* $p_2 - \ell$.

Theorem 6 shows that, when it exists, the cofeature matrix b such that $b'(\Delta^2 X_t - E(\Delta^2 X_t)) = b'\epsilon_t$ must be of the form $b = \alpha_2 u$ for an appropriate matrix u. Hence, the cofeature matrix isolates the second increments of ℓ common I(2) trends, because it is equal to $b'\epsilon_t$, and $\alpha_2'\epsilon_t$ are the second increments of the p_2 common trends, where $b \in \mathrm{col}(\alpha_2)$. Hence the interpretation of the cofeature linear combinations $b'(\Delta^2 X_t - E(\Delta^2 X_t))$ is that of *observable* second increments to the common I(2) trends, where in general the innovations ϵ_t and their cumulations are unobservable.

Summarizing, the main advantage of the choice $W_t = \Delta^2 X_t$ is to give cofeature relations that represent the second increments of common I(2) trends. The main disadvantage is that the number of cofeature vectors is restricted to be less or equal to p_2, the number of I(2) trends.

We next consider the choice $W_t = Y_t$. If the cofeature matrix b selects elements of Y_{0t} or Y_{1t}, then the cofeature relations imply that certain deviations from equilibria are innovation processes, as expected by several economic models with rational expectations. If the cofeature matrix b selects elements from Y_{2t}, the interpretation is similar to the one given for the choice $W_t = \Delta^2 X_t$. It would thus be useful in this context to test exclusion restrictions on $b'W_t$ similar to the ones of a system of structural equations. We refer to this possibility as specification-test on b.

A possible disadvantage of the choice $W_t = Y_t$ is that the components of Y_t are themselves linear combinations of X_t, ΔX_t, $\Delta^2 X_t$, so that the interpretation of the cofeature matrix is possible only after identification of the components of Y_t, and after specification-testing on b itself. This problem, however, is solved by a careful modelling of the cointegration properties of a system and by appropriate specification testing on b, see Section 6 below, where we address model specification.

We next apply the previous definitions to the two examples of Section 2. These examples show that both choices $W_t = \Delta^2 X_t$, Y_t are sensible. We hence suggest to use both choices when there is no a priori information on what type of common features may apply.

Example 7 (real interest rates - continued) *We first observe that* $\beta = (1:0)'$, $\delta = -1$, $\beta_2 = (0:1)'$, $Y_{0t} = X_{1t} - \Delta X_{2t}$, $Y_{2t} = \Delta^2 X_{2t}$, $Y_t = (Y_{0t} : Y_{2t})'$. *The equilibrium dynamics representation is easily obtained as*

$$\begin{pmatrix} X_{1t} - \Delta X_{2t} \\ \Delta^2 X_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \varrho \end{pmatrix} \begin{pmatrix} X_{1t-1} - \Delta X_{2t-1} \\ \Delta^2 X_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}.$$

Observe that the AR matrix A_1° *is of deficient rank, so that* $b = (1:0)'$ *is a cofeature vector. The cofeature relation is* $X_{1t} - \Delta X_{2t} = \eta_{1t}$, *which states that ex post real* *interest rates is an innovation process. Hence common features applied to Y_t correctly signal the presence of a common cycle.*

The equilibrium correction representation is obtained by substituting $X_{1t} = \Delta^2 X_{1t} + \Delta X_{1t-1} + X_{1t-1}$ and $\Delta X_{2t} = \Delta^2 X_{2t} + \Delta X_{2t-1}$ in the first equation. This gives $\Delta^2 X_{1t} = -(X_{1t-1} - \Delta X_{2t-1}) - \Delta X_{1t-1} + \Delta^2 X_{2t} + \eta_{1t}$. Substituting from the second equation one finds $\Delta^2 X_{1t} = -(X_{1t-1} - \Delta X_{2t-1}) - \Delta X_{1t-1} + \varrho \Delta^2 X_{2t-1} + (\eta_{1t} + \eta_{2t})$, i.e.

$$\begin{pmatrix} \Delta^2 X_{1t} \\ \Delta^2 X_{2t} \end{pmatrix} = \begin{pmatrix} -1 & -1 & \varrho \\ 0 & 0 & \varrho \end{pmatrix} \begin{pmatrix} X_{1t-1} - \Delta X_{2t-1} \\ \Delta X_{1t-1} \\ \Delta^2 X_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

where $\epsilon_{1t} := \eta_{1t} + \eta_{2t}$, $\epsilon_{2t} := \eta_{2t}$. Note that the regression matrix on the r.h.s. is of full rank for any $\varrho \neq 0$, so that there is no cofeature vector for $\Delta^2 X_t$.

In this case common features applied to $\Delta^2 X_t$ would not signal the presence of a common cycle. The reason is that when $W_t = \Delta^2 X_t$ the type of cofeature relation is the one of observable $I(2)$ trends, which is not the case here.

Example 8 (profitability - continued) *The system has already been described in terms of the equilibrium dynamics form for $Y_t := (Y_{1t} : Y_{2t}) = (\Delta X_{1t} : \Delta^2 X_{2t})$, where $Y_t = A_1^\circ Y_{t-1} + \epsilon_t$ is a stable VAR(1). The A_1° matrix (4) is of full rank, so that there are no common cycles when choosing $W_t = Y_t$. Consider next the equilibrium correction form. This is obtained by subtracting ΔX_{1t-1} from both sides of the first equation. This gives*

$$\begin{pmatrix} \Delta^2 X_{1t} \\ \Delta^2 X_{2t} \end{pmatrix} = \begin{pmatrix} -1 & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \Delta X_{1t-1} \\ \Delta^2 X_{2t-1} \end{pmatrix} + \epsilon_t = \begin{pmatrix} -1 \\ \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \Delta X_{1t-1} \\ \Delta^2 X_{2t-1} \end{pmatrix} + \epsilon_t.$$

We note that this representation admits one cofeature vector of the type $b = (2 : 1)'$, so that when choosing $W_t = \Delta^2 X_t$, one finds one common cycle.

In this case common features applied to Y_t would not signal the presence of a common cycle. In the case $W_t = Y_t$ the cofeature relations define unpredictable disequilibria, which is not the case here.

The conclusion is that for some systems there may exist cofeatures in the equilibrium correction formulation, $W_t = \Delta^2 X_t$, and for some other systems there may exist cofeatures in the equilibrium dynamics representation, $W_t = Y_t$. Both definitions may turn out to be important. Other representations in terms e.g. of $W_t = (\Delta X_t'\beta : \Delta X_t'\beta_1 : \Delta^2 X_t'\beta_2)'$ or other stationary transformation are also possible. Ultimately which option to choose remains an empirical question.

Before addressing the problem of inference we consider dynamic extensions of the concept of common features. These are considered in the following section.

5 Dynamic cofeatures

In this section we apply the notion of common features to W_t augmented with other lagged stationary variables taken from the r.h.s. of the equilibrium correction or the equilibrium dynamics forms. The main motivation for this extension is given by the lack of invariance of common features to timing of the variables.[7] This phenomenon

[7] In this respect, common features deviates from cointegration.

was first observed in Ericsson in his comments to Engle and Kozicki (1993). Again here we discuss several possible choices of additional lagged variables.

Like polynomial cointegration, see e.g. Haldrup (1998), this application of common features may be called 'polynomial cofeatures' or 'dynamic cofeatures', see Cubadda and Hecq (2001). In the following, we call the applications of common features in Section 4 'static' in order to contrast them with the 'dynamics cofeatures' notion defined below.

In the following we let W_t be an I(0) process with rank q, which is taken to be either $\Delta^2 X_t$ or Y_t as above. We will also indicate by R_t an $h \times 1$ vector of additional stationary variables, constructed from lags of X_t. Let $Z_t := (W_t' : R_t')'$. Again a matrix b, of dimension $(p+h) \times \ell$ and rank ℓ, is defined to be a cofeature matrix for Z_t if $b'(Z_t - E(Z_t))$ is an innovation process. We say that b is a cofeature matrix for Z_t with cofeature rank ℓ.

This definition nests the one of static cofeatures. In fact if $b := (b_1' : b_2')'$ is partitioned conformably with $Z_t := (W_t' : R_t')'$, choosing $b_2 = 0$ delivers the definition given in Section 4. The above definition is also a re-statement of the definition of 'polynomial serial correlation common features' given in Cubadda and Hecq (2001), Definition 1, when applied to the *levels of* X_t rather than to the differences. In fact let for instance $W_t = \Delta^2 X_t$, $R_t = v(L)X_{t-1}$; then the cofeature relations $b_1'W_t + b_2'R_t = (b_1'\Delta^2 + b_2'v(L)L)X_t =: b(L)X_t$ correspond to their Definition 1 for $b(L) := (b_1'\Delta^2 + b_2'v(L)L)$. Note that the levels are needed here to accommodate also the possibility that the cointegrating relations appear in R_t, and/or in W_t.

The interpretation of dynamic cofeatures is similar to the static case; they only differ for the list of variables to which the notion of common features is applied, W_t or $Z_t := (W_t' : R_t')'$. A consequence of the definition is that in dynamic cofeatures, the contemporaneous variables W_t are always involved, in the sense of the following proposition.

Proposition 9 *If $b := (b_1' : b_2')'$ is a $(p+h) \times \ell$ cofeature matrix for $Z_t := (W_t' : R_t')'$, where R_t depends on lagged X_ts and b is partitioned conformably with Z_t, then b_1 has full column rank ℓ.*

The inclusion of additional variables R_t is meant to be minimal. In this sense it would be interesting to investigate what set of additional variables R_t makes the choices $(\Delta^2 X_t' : R_t')'$ and $(Y_t' : R_t')'$ equivalent. This is reported in the following proposition.

Proposition 10 *The dynamic cofeature properties of $U_{1t} := (\Delta^2 X_t' : R_t')'$ and $U_{2t} := (Y_t' : R_t')'$ are identical for $R_t := (Y_{0,t-1}' : \Delta X_{t-1}'(\beta : \beta_1))'$.*

In the next proposition we state the necessary and sufficient conditions in order to have common features of dynamic type; this proposition extends Proposition 5. In the following we indicate W_t with Z_{0t}, and we let $Z_{2t} := (R_t' : d_t')'$, in order to simplify the notation of later statements. We define $\epsilon_t^* := C_{W,0}\epsilon_t$, where $C_{W,0} = I$ for $W_t = \Delta^2 X_t$ and $C_{W,0} = D$ for $W_t = Y_t$, see Theorem 4. The covariance matrix of ϵ_t^* is indicated by $\Omega^* := C_{W,0}\Omega C_{W,0}'$. Similarly we let μ_0^* indicate μ_0, $\mu_0^\dagger$, $\mu_0^\ddagger$ or μ_0°.

Proposition 11 *Let $Z_{2t} := (R_t' : d_t')'$, and assume that $Z_{0t} := W_t$, Z_{1t} and R_t be variables generated from a stationary VAR with i.i.d. innovations ϵ_t, where Z_{0t} satisfies*

$$Z_{0t} = \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \tag{18}$$

and Z_{1t}, R_t depend on lagged ϵ_ts. Partition also $\Phi := (\Phi_1 : \Phi_2)$ conformably with $Z_{2t} := (R_t' : d_t')'$. Then a necessary and sufficient condition for b to be a cofeature matrix for $(W_t' : R_t')' = (Z_{0t}' : R_t')'$ is that ς is of reduced rank, $\varsigma = \varphi\tau'$, with φ and τ of full column rank. In this case the cofeature matrix has representation

$$b' = (\varphi_\perp' : \varphi_\perp'\Phi_1) \qquad \text{and} \qquad b'(W_t' : R_t')' = \varphi_\perp' W_t + \varphi_\perp'\Phi_1 R_t = \varphi_\perp'(\Phi_2 d_t + \mu_0^*) + \epsilon_t^*.$$

We next illustrate possible choices for R_t using the equilibrium correction (9) formulation and the mixed form (13). Analogous remarks can be given for the equilibrium dynamics (10); these are not given here for conciseness. In the empirical application we use the characterization given in Proposition 11, simply stating the reduced rank restrictions implied by different choices of variables in Z_{0t}, Z_{1t}, Z_{2t}.

A list of different dynamic cofeatures cases is given in Table 1, using the format of equation (18). We observe that case d) for $W_t = \Delta^2 X_t$ corresponds to the conditions for $b_1' X_t$ to be weakly exogenous for the cointegrating parameters β, β_1, δ, see Paruolo and Rahbek (1999). In particular these conditions can be written as $b_1'(\alpha : \zeta) = 0$, which simply states that the equations of $b_1' X_t$ in the equilibrium correction formulation (9) have zero adjustment coefficients. This situation may be described as 'no levels- and difference-feedback' in the equations of $b_1'\Delta^2 X_t$. Cases b), c), e) are similar to the definition of 'weak form' of common features proposed in Hecq et al. (2000, 2002) for I(1) systems. The idea is that some elements in $\Delta^2 X_t$ inherit the cyclic part included in deviations from equilibria in $Y_{0,t-1}$ and/or $(\beta : \beta_1)'\Delta X_{t-1}$.

Several cases given in Table 1 are nested. This suggest the possibility to test down for cofeatures from the most general to the most specific model. This strategy is indicated as the 'testing down' procedure in the following. The sequence starts from models characterized by the less stringent restrictions, represented by case e). Rejection of the reduced rank restrictions in this model implies also rejection for any nested submodel. Hence, finding that model e) does not support the presence of cofeatures implies that no submodel (cases a), b), c)) presents cofeatures.

When the presence of cofeatures is supported in a model, like model e), one can continue testing more restricted submodels. Cases b) and c) are nested within model e), but mutually non-nested. Both submodels can be investigated. If both submodels do not support cofeatures, then one returns to the first nesting model that supports cofeatures. Eventually the sequence may reach the static cofeature model a). Obviously, the significance level of the individual tests in the testing-down procedure must be chosen in order to guarantee a given overall size, by use of Bonferroni-type inequalities. Hence a typically small nominal size is chosen for each component test.

One can also arrange the specification search starting from the most restricted model; we call this strategy the 'testing up' procedure. In this case the most restricted model is the static cofeatures model, case a). If this model does not support cofeatures, one considers less stringent models, like models b) and c). Eventually the specification search may reach the least stringent model e).

Note that, in all cases, the models with cofeatures are compared with a baseline reference model, which is the unrestricted equilibrium correction formulation (9) or the equilibrium dynamics mixed form (13). Hence also the 'testing up' procedure is in line with the general-to-specific framework, see Johansen (1992b) or Paruolo (2001). In this procedure, the sizes of the tests in the sequence are fixed at the overall nominal level; the overall procedure can be shown to have the asymptotic nominal size of each

case	cofeatures	$b_1'\alpha^{(\cdot)}$	$b_1'\zeta^{(\cdot)}$	$b_1'\gamma^{(\cdot)}$	Z_{1t}	$Z_{2t}, 1$
$a)$	static	0	0	0	$\begin{pmatrix} Y_{0,t-1} \\ (\beta : \beta_1)'\Delta X_{t-1} \\ V_t \end{pmatrix}$	D_t
$b)$	dynamic	*	0	0	$\begin{pmatrix} (\beta : \beta_1)'\Delta X_{t-1} \\ V_t \end{pmatrix}$	$\begin{pmatrix} Y_{0,t-1} \\ D_t \end{pmatrix}$
$c)$	dynamic	0	*	0	$\begin{pmatrix} Y_{0,t-1} \\ V_t \end{pmatrix}$	$\begin{pmatrix} (\beta : \beta_1)'\Delta X_{t-1} \\ D_t \end{pmatrix}$
$d)$	dynamic	0	0	*	$\begin{pmatrix} Y_{0,t-1} \\ (\beta : \beta_1)'\Delta X_{t-1} \end{pmatrix}$	$\begin{pmatrix} V_t \\ D_t \end{pmatrix}$
$e)$	dynamic	*	*	0	V_t	$\begin{pmatrix} Y_{0,t-1} \\ (\beta : \beta_1)'\Delta X_{t-1} \\ D_t \end{pmatrix}$

Table 1: Possible cofeature rank restrictions in the regression format of (19) using the notation $RRR(Z_{0t}, Z_{1t}|Z_{2t}, 1)$. The dependent variables Z_{0t} is either $\Delta^2 X_t$ for the equilibrium correction form (9) or Y_t for the equilibrium dynamics mixed form (13). $\alpha^{(\cdot)}$ indicates either α or α°; similarly for ζ, γ. * indicates unrestricted entries.

component test, if each test has probability of rejection that converges to 1 under a fixed alternative. For further details on this type of procedure we refer to Johansen (1992b) or Paruolo (2001) and reference therein.

6 Estimation and testing

This section describes inference on I(2) VAR systems with common trends and cycles. The cointegration analysis of I(2) systems has been extensively discussed in the literature, and it is not described here for space constraints. We refer to Johansen (1995a, 1997), Rahbek et al. (1999), Boswijk (2000), Paruolo (2000, 2002a), inter alia; see Haldrup (1998) for a review.

This section concentrates on the analysis of cofeatures after the cointegration analysis has been performed, fixing the cointegration parameters β, β_1, δ to their maximum likelihood estimates or the two stage I(2) estimates of Johansen (1995a), 2SI2, see Rahbek et al. (1999). These estimators of the cointegration parameters are superconsistent, and using the estimates in place of the parameters does not change the limit distributions of the common feature statistics described below, see Appendix B. In the rest of this section we simply do not distinguish β, β_1, δ and their estimated values.

For any given model, see Table 1, the analysis of common features may be organized by first determining the cofeature rank ℓ. The cofeature matrix b can then be estimated, for the selected cofeature rank ℓ, possibly testing restrictions on b. In some cases, economic theory may suggest the specific value of the cofeature matrix b; in this case it would be of interest to test that a certain vector is a cofeature vector. Finally one may analyze the cofeature relations $b'W_t = u_t$ or $b'Z_t = u_t$ as a system of simultaneous equations, where u_t are ℓ linear combinations of the innovations ϵ_t. All these hypotheses are considered in this section; we consider either likelihood ratio tests, labelled Q_i, or Wald-type tests, indicated by J_i. Proofs of the statements in this section are collected in Appendix B.

We first treat the case of unknown cofeature matrix b. Several cases of common features have been presented in Sections 4 and 5, see Table 1. As stated in Proposition 5 and Proposition 11, they can all be put in the regression format

$$Z_{0t} = \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \tag{19}$$

where the cofeature restriction is

$$H(s): \qquad \varsigma = \varphi\tau'. \tag{20}$$

and φ, τ, Φ, μ_0^* and $\Omega^* := E(\epsilon_t^*\epsilon_t^{*\prime})$ are unrestricted and s indicates the number of columns in φ, τ, where φ is $p \times s$ and τ is $j \times s$. Because when $j < p$, there always exist a cofeature matrix of rank $p - j$, we exclude these trivial cases by assuming $j \geq p$, i.e. there are more regressors than dependent variables.[8] We indicate as the '$H(s)$ model' the regression model (19) under the reduced rank restriction (20).

The $H(s)$ model is analyzed by reduced rank regression, indicated in the following with the shorthand $\mathrm{RRR}(Z_{0t}, Z_{1t}|Z_{2t}, 1)$. The Gaussian likelihood function is maximized by considering the following eigenvalue problem

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0 \tag{21}$$

with eigenvalues $\lambda_1 > ...\lambda_i > ... > \lambda_p$ and associated eigenvectors v_i, where $S_{ij} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $M_{ij} := T^{-1}\sum_{t=1}^{T}(Z_{it} - m_i)(Z_{jt} - m_j)'$, $m_i := T^{-1}\sum_{t=1}^{T} Z_{it}$, i, $j = 0, 1, 2$, see e.g. Johansen (1996).

The LR test statistic for hypothesis (20) of $H(s)$ versus $H(p)$ about the rank of ς is given by

$$Q_1(s) := -T\sum_{i=s+1}^{p}\ln(1 - \lambda_i).$$

This test is asymptotically $\chi^2((j - s)(p - s))$ under the null; moreover $Q_1(s - i) \to \infty$ for $i > 0$. These properties allow to adopt a testing-up sequence for the rank determination, see Johansen (1992b), Paruolo (2001).

Eq. (21) provides also the maximum likelihood estimates for given dimension s. In particular $\widehat{\tau} = (v_1 : ... : v_s)$ and

$$\begin{aligned}
\widehat{\varphi} &= S_{01}\widehat{\tau}(\widehat{\tau}'S_{11}\widehat{\tau})^{-1}, \qquad \widehat{\varsigma} = \widehat{\varphi}\widehat{\tau}' = S_{01}\widehat{\tau}(\widehat{\tau}'S_{11}\widehat{\tau})^{-1}\widehat{\tau}', \\
\widehat{\Phi} &= (M_{02} - \widehat{\varsigma}M_{12})M_{22}^{-1}, \qquad \widehat{\Omega}^* = S_{00} - S_{01}\widehat{\tau}(\widehat{\tau}'S_{11}\widehat{\tau})^{-1}\widehat{\tau}'S_{10},
\end{aligned} \tag{22}$$

where $\widehat{\tau}$ is normalized by $\widehat{\tau}'S_{11}\widehat{\tau} = I_s$, $\widehat{\tau}'S_{10}S_{00}^{-1}S_{01}\widehat{\tau} = diag(\lambda_1, ..., \lambda_s) =: \Lambda_1$.

In order to identify parameters, it is convenient to normalize $\widehat{\tau}$ by the just-identifying restrictions $\widehat{\tau}_c := \widehat{\tau}(c'\widehat{\tau})^{-1}$, where c is a known matrix of the same dimensions of τ, such that $c'\tau$ is a square nonsingular matrix, see Johansen (1996) Section 5.2 or Paruolo (1997). The choice of $\widehat{\varphi}$ obtained by substituting $\widehat{\tau}_c$ in place of $\widehat{\tau}$ in (22) is given by ${}_c\widehat{\varphi} := \widehat{\varsigma}c$, which satisfies ${}_c\widehat{\varphi}\widehat{\tau}_c' = \widehat{\varsigma}$.

In the following we use the just-identifying normalization $\widehat{\varphi}_{\perp a_\perp} := \widehat{\varphi}_\perp(a_\perp'\widehat{\varphi}_\perp)^{-1}$ also for the estimator of $\varphi_\perp$. We note that $\varphi_\perp$ is estimated unrestrictedly as the matrix of eigenvectors associated with the last $p - s$ eigenvalues of the dual problem to (21)

$$\left|\lambda S_{00} - S_{01}S_{11}^{-1}S_{10}\right| = 0, \tag{23}$$

[8] Most of the derivations are unaffected by this assumption.

which has the same λ_i eigenvalues of (21) and eigenvectors $u_1, ..., u_p$; one has $\widehat{\varphi}_\perp = (u_{s+1} : ... : u_p)$, see Johansen (1996) Theorem 8.5. Here $\widehat{\varphi}_\perp$ is normalized by $\widehat{\varphi}_\perp' S_{00} \widehat{\varphi}_\perp = I_{p-s}$, $\widehat{\varphi}_\perp' S_{01} S_{11}^{-1} S_{10} \widehat{\varphi}_\perp = diag(\lambda_{s+1}, ..., \lambda_p) =: \Lambda_2$. The corresponding just-identified estimator is $\widehat{\varphi}_{\perp a_\perp} := \widehat{\varphi}_\perp (a_\perp' \widehat{\varphi}_\perp)^{-1}$, where $a_\perp$ is a known, full column rank matrix of the same dimensions of $\varphi_\perp$, and it is assumed that $a_\perp' \varphi_\perp$ is of full rank.

By the just-identifying restriction, one has $a_\perp' \widehat{\varphi}_{\perp a_\perp} = a_\perp' \varphi_{\perp a_\perp} = I_\ell$, so that $a_\perp'(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp}) = 0$, and $(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp}) = P_a(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})$, see Paruolo (1997), and one only needs to report the limit distribution for $\bar{a}'(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})$. This is given in the following proposition.

Proposition 12 *One has*

$$T^{1/2} \left(vec(\bar{a}'(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})) \right) \xrightarrow{d} N \left(0, \left(\varphi_{\perp a_\perp}' \Omega^* \varphi_{\perp a_\perp} \otimes (a' \varsigma \Sigma_{11} \varsigma' a)^{-1} \right) \right). \tag{24}$$

A consistent estimator of the asymptotic covariance matrix is obtained substituting parameters with the corresponding ML estimator and Σ_{11} with S_{11}.

Hence one can consider generic linear hypothesis of the type $K' vec(\bar{a}' \widehat{\varphi}_{\perp a}) = j$, which nest (26), where K has h columns. The associated Wald test is given by

$$J_1 := T(K' vec(\bar{a}' \widehat{\varphi}_{\perp a_\perp}) - j)' \left(\left(\widehat{\varphi}_{\perp a_\perp}' \widehat{\Omega}^* \widehat{\varphi}_{\perp a_\perp} \right)^{-1} \otimes (a' \widehat{\varsigma} S_{11} \widehat{\varsigma}' a) \right) (K' vec(\bar{a}' \widehat{\varphi}_{\perp a_\perp}) - j). \tag{25}$$

Also this test is shown to be asymptotically $\chi^2(h)$ and to diverge under fixed alternatives in Appendix B.

In the analysis of the specification of the cofeature matrix, it may be of interest to consider restrictions of the type

$$H_0 : \qquad \varphi_\perp = H\phi, \tag{26}$$

which accommodate exclusion restrictions for all columns of $\varphi_\perp$ simultaneously. Here H is $p \times h$, $h \geq \ell$. Under the restriction (26), the likelihood function is maximized explicitly by solving

$$\left| \lambda^* H' S_{00} H - H' S_{01} S_{11}^{-1} S_{10} H \right| = 0,$$

with eigenvalues $\lambda_1^* > ... > \lambda_h^*$ and corresponding eigenvectors v_i^*, see e.g. Paruolo (1997), Appendix C, or Johansen (1996) Theorems 8.4 and 8.5. The corresponding LR test statistic of (26) in $H(s)$ is given by

$$Q_2 := T \left(\sum_{i=s+1}^{p} \ln(1 - \lambda_i) - \sum_{i=h-p+s+1}^{h} \ln(1 - \lambda_i^*) \right),$$

and the restricted estimate of $\varphi_\perp$ is $\widehat{\varphi}_\perp = H(v_{h-p+s+1}^* : ... : v_h^*)$. This test is asymptotically distributed as $\chi^2(df_{Q_2})$ and diverges under a fixed alternative, see Appendix B. The degrees of freedom correspond to the number of restrictions, $df_{Q_2} := 2ps - s^2 - 2p(p-h) - (p-s-h)(2h-p+s)$.

Consider now the case where b is (partly) known. Let K be a known $p \times h$ matrix of rank $h \leq \ell$, and consider the hypothesis that K is a submatrix of b, $b = (K, b_2^*)$, i.e.

$$H_0 : \qquad K' \varsigma = 0. \tag{27}$$

A Wald test of (27) can be based on the unrestricted maximum likelihood estimates $\widehat{\varsigma} := S_{01}S_{11}^{-1}$, $\widehat{\Omega}^* := S_{00.1} := S_{00} - S_{01}S_{11}^{-1}S_{10}$, and equals

$$J_2 := Ttr\left((K'S_{00.1}K)^{-1}K'\widehat{\varsigma}S_{11}\widehat{\varsigma}'K\right) = Ttr\left((K'S_{00.1}K)^{-1}K'S_{01}S_{11}^{-1}S_{10}K\right). \quad (28)$$

Also this test is shown to be asymptotically $\chi^2(hj)$ and to diverge under fixed alternatives in Appendix B. The corresponding LR test of (27) in $H(s)$, labelled Q_3, is found by solving the eigenvalue problem

$$\left|\lambda^\circ K_\perp' S_{00.K}K_\perp - K_\perp' S_{01.K}S_{11.K}^{-1}S_{10.K}K_\perp\right| = 0, \quad (29)$$

with eigenvalues $\lambda_1^\circ > ... > \lambda_h^\circ$ and corresponding eigenvectors v_i°, where $S_{ij.K} := S_{ij} - S_{i0}K(K'S_{00}K)^{-1}K'S_{0j}$, $i, j = 0, 1$, see Johansen (1996) Theorems 8.2 and 8.5. The test of (27) in $H(s)$ is given by

$$\begin{aligned} Q_3 := \quad & T\left(\sum_{i=1}^{s}\ln(1-\lambda_i^\circ) - \sum_{i=1}^{s}\ln(1-\lambda_i)\right) = \\ = \quad & T\left(\sum_{i=s+1}^{p}\ln(1-\lambda_i) - \sum_{i=s+1}^{h-p+s}\ln(1-\lambda_i^\circ) - \ln\frac{|K'S_{00.1}K|}{|K'S_{00}K|}\right), \end{aligned} \quad (30)$$

where $S_{00.1} := S_{00} - S_{01}S_{11}^{-1}S_{10}$. The restricted estimate under (27) is $\widehat{\varphi}_\perp = (K : K_\perp(v_{s+1}^\circ : ... : v_{h-p+s}^\circ))$, which again can be identified via $\widehat{\varphi}_{\perp a_\perp}$. The Q_3 test is asymptotically $\chi^2(df_{Q_3})$, with degrees of freedom equal to the number of constraints, $df_{Q_3} := sh$. The tests $Q_1(s)$ and Q_3 can be combined to obtain the LR test of (27) in $H(p)$, $Q_4 := Q_1(s) + Q_3$. Again it can be shown that $Q_4 \xrightarrow{d} \chi^2(df_{Q_4})$, with degrees of freedom equal to the number of constraints, $df_{Q_4} = df_{Q_1(s)} + df_{Q_3}$. Both Q_3 and Q_4 diverge under a fixed alternative.

Finally, observe that $\varphi_\perp' Z_{0t} = u_t$, where u_t are ℓ linear combinations of ϵ_t, defines a system of ℓ simultaneous equations. Homogeneous separable restrictions on each equation can be written in the form

$$\varphi_\perp = (H_1\phi_1 : ... : H_\ell\phi_\ell),$$

see Johansen (1995b) for the discussion of identification in this case. We just mention here that the algorithm for the maximization of the likelihood proposed there, see also Johansen (1996) Theorem 7.4, can be applied to the estimation of $\varphi_\perp$ in the dual problem (23), interchanging β and $\varphi_\perp$, the subscripts 0 and 1, and choosing the smallest eigenvalues instead of the largest ones.

7 An application to Australian prices

In this section we present an application to the Australian prices data-set analyzed by Banerjee et al. (2001).[9] The same data have also been analyzed in Omtzigt and Paruolo (2002) for common I(2) trends. We here summarize the common trends findings of Omtzigt and Paruolo (2002), and apply the common feature analysis proposed in the previous sections. The calculations of the I(2) cointegration analysis

[9]The data set is available at the data archive of the Journal of Applied Econometrics: http://qed.econ.queensu.ca/jae.

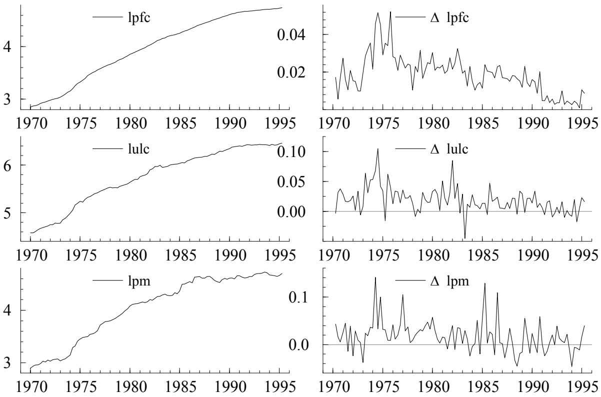

Figure 1: Data in levels and first differences.

are documented in Omtzigt and Paruolo (2002), to which we refer for full details. The remaining computations of the common feature analysis were performed in Gauss 3.6 and PcGive 10.0.

The data-set consists of three Australian macroeconomic time series: the consumer price deflator at factor cost (lpfc), unit labor costs in the non-farm sector (lulc) and import prices (lpm). All three variables are quarterly data, measured in natural logs, and run from 1970Q1 to 1995Q2 for a total of 102 observations. The variables are graphed in levels and first differences in Fig. 1. The levels of the variables appear non-stationary, and also the differences show signs of non-stationarity. No apparent break in the deterministic terms is visible in Fig. 1.

We included dummy variables to take account of a number of shocks to the economy, like the oil shocks. The dummies take value 1 in one quarter and zero otherwise; the quarters are 1974Q2, 1974Q3, 1975Q2, 1982Q1, 1983Q2, 1985Q2 and 1986Q3; these dummies are indicated by d_t^* in the following.[10] We fitted an unrestricted VAR in levels with $k = 2$ lags, dummies d_t and d_t^*, a constant and a trend; the model passed mis-specification tests for ARCH, normality and autocorrelation of the errors.

We next tested for the degree of integration of the system, allowing for the possibility of I(2). This analysis is reported in Subsection 7.1 below. The analysis of the common cycles was performed next, and it is reported in Subsection 7.2.

7.1 Common trends

The selection of the integration indices was based on the 2SI2 estimator (Johansen 1995, Rahbek et al. 1999); the test statistics for the specification $\mu_1 = \alpha\beta_0'$ are reported in Table 2. Below each entry we report the 95% quantiles of the asymptotic distribution, taken from Rahbek et al. (1999). The selected integration indices are $(p_0, p_1) = (1, 1)$, which corresponds to one I(1) direction and one I(2) trend. The restricted roots of the characteristics polynomial are 1, 1, 1, 0.38, -0.21 and 0.11; there is no evidence of additional non-stationary trends.[11] The same integration indices were selected by Banerjee et al. (2001).

We next tested the nominal-to-real transformation (see Kongsted, 2002), i.e. that

[10] Banerjee et al. (2001) conditioned on a number of stationary variables we do not consider here. Their selection of integration indices is the same as the one reached here; moreover we do not reject

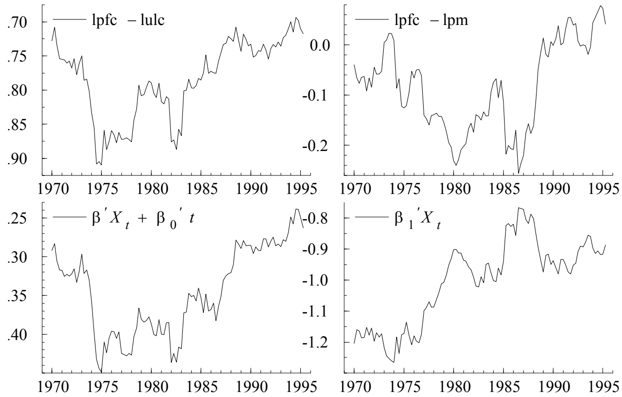

Figure 2: The top panels contain the log-ratios lpfc−lulc and lpfc−lpm, which are tested to be I(1). The bottom panels report linear combination of the top panels, $\beta' X_t + \beta_0' t$ and $\beta_1' X_t$, which are still both I(1), but the former cointegrates with the differences $\beta_2' \Delta X_t$.

$p_1{+}p_2$	p_0				
3	0	279.3	158.6	74.7	53.0
		(87.6)	(68.2)	(53.2)	(42.7)
2	1		117.8	**31.8**	23.7
			(47.6)	(34.4)	(25.4)
1	2			17.5	10.8
				(19.9)	(12.5)
p_2		3	2	1	0

Table 2: 2SI2 tests on the integration indices p_0, p_1, $p_2 := p - p_0 - p_1$. 5% asymptotic critical values are reported in parenthesis; they are taken from Rahbek et al. (1999). The sequence of tests is from the upper left corner to the lower right corner, proceeding row-wise from left to right. The first un-rejected model is shown in boldface.

lpfc−lulc (the markup of internal prices on unit labor cost) and lpfc−lpm (the markup of price over import prices) are at most I(1), see Fig. 2. We calculated the corresponding likelihood ratio (LR) statistic; under the null the test has an asymptotic $\chi^2(2)$-distribution, see Johansen (2002). The test statistic takes the value 0.935, with a p-value of 0.63, giving ample support to the transformation. This implies that $\beta = H\rho$, and $\beta_2 = H_\perp = (1:1:1)'$, where $H = (e_1 - e_2 : e_1 - e_3)$, and e_i is a 3×1 vector with all 0 and 1 in position i.

In other words, the common I(2) trend can be represented by the average of the three price series, $3^{-1}\beta_2' X_t = 3^{-1}\sum_{i=1}^{3} X_{it}$. The differences lpfc−lulc and lpfc−lpm, pictured in Fig. 2, are I(1), i.e. they are $CI(2,1)$ relations in the sense that they reduce by 1 the order of integration. They also cointegrate with the average inflation rate $\pi_t := 3^{-1}\beta_2' \Delta X_t$ because $p_0 = 1$.

The maximum likelihood estimates of the cointegration parameters are reported in Table 3. $\widehat{\beta} = b\widehat{\rho}$ is the linear combination of lpfc−lulc and lpfc−lpm which cointe-

the nominal-to-real transformation, as in their paper.

[11] The roots of the unrestricted polynomial are 1.00, $0.89 \pm 0.02i$, 0.41, −0.22 and 0.14.

			lpfc	lupc	lpm	t
ρ	0.7423	H'	1	-1	0	0.0013
	0.2577		1	0	-1	-0.0029
δ	2.6760	β_2'	1	1	1	

Table 3: ML estimates of the cointegration parameters under the nominal to real transformation; H is a basis of $\text{col}(\beta : \beta_1)$, $\beta = H\rho$.

grates with π_t:

$$\widehat{\beta}' X_t + \widehat{\beta}_0 t = \text{lpfc}_t - 0.74\text{lulc}_t - 0.26\text{lpm}_t + 0.0013t.$$

The remaining CI(2,1) relationship $\widehat{\beta}_1 = \bar{H}\widehat{\rho}_\perp$ is chosen orthogonal to $\widehat{\beta}$; this is also I(1), but does not cointegrate with π_t:

$$\widehat{\beta}_1' X_t = -0.28\text{lpfc}_t - 0.72\text{lulc}_t + \text{lpm}_t.$$

The fact that the combined mark-up, $\widehat{\beta}' X_t$, is still I(1) by itself is consistent with imperfect competition theories, which predict that a high mark-up is associated with low inflation.[12] The combined markup $\widehat{\beta}' X_t$ next cointegrates with the I(1) trend in the average inflation π_t to give the multicointegration relationship

$$\begin{aligned} Y_{0t} = \widehat{\beta}' X_t + \widehat{\delta}\widehat{\beta}_2' \Delta X_t + \widehat{\beta}_0' t = \text{lpfc}_t - 0.74\text{lulc}_t - 0.26\text{lpm}_t + \\ + 2.68\,(\Delta\text{lpfc}_t + \Delta\text{lulc}_t + \Delta\text{lpm}_t) + 0.0013t. \end{aligned} \quad (31)$$

This multicointegration relation represents a *compensated* markup relation, where the markup of internal prices over labor cost and imports depends negatively on the average inflation in the three series: high average inflation is associated with low markups and vice versa.

The other CI(2,1) cointegrating relation $\widehat{\beta}_1' X_t$ eliminates the I(2) trend but does not cointegrate further with π_t. Hence $\widehat{\beta}_1' X_t$ may be interpreted as the I(1) *autonomous* component in the system, in contrast with the I(1) linear combination $\widehat{\beta}' X_t$ which balances the average inflation rate π_t.

The equilibrium corrections shows how the original variables adjust to various disequilibria. The ML estimates of the adjustment coefficients α and ζ are reported in the upper panel of Table 4. They show a significant adjustment to the growth rate of the autonomous price component $\widehat{\beta}_1' \Delta X_{t-1}$, both for $\Delta^2\text{lulc}_t$ and $\Delta^2\text{lpm}_t$. We interpret this finding as evidence that the I(1) autonomous price component $\widehat{\beta}_1' X_t$ contains international trends, which also influence the labor market. Note also that the adjustment to the multicointegrating relation $Y_{0,t-1}$ is significant only in the equation for Δ^2lpfc, suggesting that Y_{0t} measures (deviations from) an internal price equilibrium.[13]

During the period under study, the Australian economy moved from a fixed to a floating exchange rate regime and from a national-award-based wage system to a localized system. In order to check for possible breaks in the model, we tested for

[12] For a full overview of the economic theory, we refer to Banerjee et al. (2001).

[13] Note that the effects of the main oil shocks is already modelled through the dummies d_t^*.

Unrestricted	$\widehat{\alpha}$	$\widehat{\zeta}_1$	$\widehat{\zeta}_2$
Δ^2lpfc	**−0.0655**	**−0.355**	0.0304
Δ^2lulc	0.0105	**1.310**	**0.189**
Δ^2lpm	−0.00391	**0.622**	**−0.839**
Restricted	$\widehat{\alpha}$	$\widehat{\zeta}_1$	$\widehat{\zeta}_2$
Δ^2lpfc	**−0.0596**	**−0.329**	0
Δ^2lulc	0	**1.256**	**0.179**
Δ^2lpm	0	**0.639**	**−0.866**

Table 4: Estimated adjustment coefficients α, ζ in the equilibrium correction form (9). Bold entries correspond to significant coefficients at the 5% level. Top panel gives the ML unrestricted estimates, the bottom panel gives the restricted estimates setting insignificant coefficients to 0, after fixing the cointegration coefficients. The LR test of these restrictions gave a test statistic of 2.5709 with a $\chi^2(3)$ p-value of 0.4626.

structural changes in the speed of adjustment to equilibrium. We calculated the estimated error correction terms Y_{0t}, Y_{1t}, $\widehat{\beta}'\Delta X_t$, using the estimates of the cointegration parameters; we then performed Andrews' (1993) stability test on the adjustment coefficients $(\alpha : \zeta)$. The unknown sample-fraction break-point was chosen in the range $[\pi_0, 1 - \pi_0] = [0.20, 0.80]$. The sup-LR test for breaks gives a test statistics of 23.36. The 5% critical value in Table 1 in Estrella (2003) for $\pi_0 = 0.2$ and dimension 9 is 25.16, which implies a non-rejection. We thus conclude that there is no evidence of breaks in the model.

7.2 Common cycles

This subsection presents the common cycle analysis, fixing the cointegration parameters β, β_1, δ at the estimates obtained in the previous subsection. We analyzed both the equilibrium correction form (9) and equilibrium dynamics form (10) for presence of cofeature vectors. Because the system has $k = 2$ lags, the mixed form (13) and the restricted equilibrium dynamics form (12) coincide. Moreover, given that there are no lagged terms in second differences $\Delta^2 X_{t-j}$, model *e*) in Table 1 is trivially satisfied.

We performed the analysis in the testing-up sequence, starting from model *a*) in Table 1 with static cofeatures, both for the equilibrium correction form and for the equilibrium dynamics form. We employed the Q_1 test statistic to investigate the presence of cofeature vectors, taking Z_{0t} either equal to $\Delta^2 X_t$ or Y_t, Z_{1t} equal to $(Y_{0,t-1} : Y_{1,t-1} : \beta'\Delta X_{t-1})'$ and $Z_{2t} = (d_t' : d_t^{*\prime})'$, where d_t^* indicate the intervention dummies introduced at the beginning of this section. Table 5 reports the result of the test statistics. It can be seen that the tests reject the presence of cofeatures in the equilibrium correction form (9) for $\Delta^2 X_t$, while they indicate the presence of a single cofeature vector for the equilibrium dynamics in mixed form (13) for Y_t.

The corresponding estimate of $\varphi_\perp$ for $s = 2$ i.e. $\ell = 1$, is reported in Table 6. We normalized the estimate on the coefficient to Y_{1t}, by choosing $a_\perp = e_2$ and $a = (e_1 : e_3)$ in (24), where e_i is a unit vector with all zeros and a 1 in position i. Table 6 reports also asymptotic standard errors based on Proposition 12, and the corresponding asymptotically normal t-ratios. These estimates suggest the vector

specification	s	0	1	2
(9)	$Q_1(s)$	250.28	116.33	33.00
	$df_{Q_1(s)}$	9	4	1
	p-value $\chi^2(df_{Q_1(s)})$	0.0000	0.0000	0.0000
(13)	$Q_1(s)$	562.08	43.00	0.3538
	$df_{Q_1(s)}$	9	4	1
	p-value $\chi^2(df_{Q_1(s)})$	0.0000	0.0000	0.5519

Table 5: Test statistics $Q_1(s)$ for the equilibrium correction form (9) for $\Delta^2 X_t$ and for the equilibrium dynamics in mixed form (13) for Y_t. *df* indicates the number of degrees of freedom.

equations	Y_{0t}	Y_{1t}	Y_{2t}
$\widehat{\varphi}'_{\perp a_\perp}$	-0.0092	1	0.1026
standard errors	0.0593	.	0.0756
t-ratios	-0.1551	.	1.3581

Table 6: Estimates of the cofeature vector $\varphi_\perp$ for the equilibrium dynamics in mixed form (13) for Y_t.

$(0:1:0)'$ as a candidate cofeature vector, i.e. that $Y_{1t} - E(Y_{1t})$ is an innovation process in this system.

In order to test the hypothesis $\varphi_\perp = (0:1:0)'$ we employed the LR test Q_2 of (26) in $H(s)$, specifying $H = (0:1:0)'$.[14] We obtained $Q_2 = 2.0358$, with a p-value of 0.3614 when compared with a $\chi^2(2)$. The corresponding Wald test J_1 in (25) was equal to 1.8453 with a p-value of 0.3975 when compared with a $\chi^2(2)$. Hence, both tests support the hypothesis $\varphi_\perp = (0:1:0)'$.

The same conclusion for the equilibrium dynamics can be derived by testing that single equations of the system $Y_t := (Y_{0t} : Y_{1t} : Y_{2t})'$ have all coefficients to stochastic regressors equal to 0 in $H(p)$. This hypothesis is of the type (27) with $K = e_i$; the associated Wald test statistic is J_2 in (28).

We report the results for test J_2 in Table 7. We also calculated the LR test of (27) in $H(s)$, i.e. the test Q_4.[15] Both tests confirm that there exists a static cofeature vector $b = (0:1:0)'$ in the system for Y_t, i.e. that all coefficients of lagged variables

[14] The same test can also be obtained as a special case of Q_3.

[15] These tests were calculated with PcGive 10.0.

specification	statistic	equation		
(9)		Δ^2 lpfc	Δ^2 lulc	Δ^2 lpm
	J_2	50.162 [0.0000]	165.74 [0.0000]	111.81 [0.0000]
	Q_4	45.949 [0.0000]	107.40 [0.0000]	83.297 [0.0000]
(13)		Y_{0t}	Y_{1t}	Y_{2t}
	J_2	31.243 [0.0000]	2.0798 [0.5560]	111.18 [0.0000]
	Q_4	30.990 [0.0000]	2.3896 [0.4956]	82.979 [0.0000]

Table 7: Test statistics J_2, Q_4 and $\chi^2(3)$ p-values in brackets of hypothesis (27) with $K = e_i$, $i = 1, 2, 3$, corresponding to the zero coefficients in the single equations indicated at the top of each column.

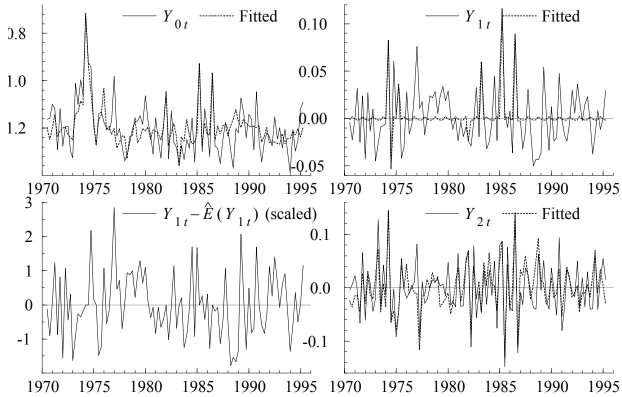

Figure 3: Fit of the equilibrium dynamics Y_t in mixed form, where the coefficients of $Y_{0,t-1}$, $Y_{1,t-1}$ and $\Delta\beta' X_{t-1}$ are constrained to 0 in the equation for Y_{1t}, i.e. (27) holds with $K = e_2$. The lower left panel reports $Y_{1t} - \widehat{E}(Y_{1t})$, which is an innovation process.

in the equation for Y_{1t} can be restricted to 0. Hence $Y_{1t} - E(Y_{1t})$ is an innovation process, i.e. the autonomous I(1) component is one of the I(1) trends in the system. The fit of the restricted equilibrium dynamics is graphed in Fig. 3, along with the estimated innovation process $Y_{1t} - \widehat{E}(Y_{1t})$.

This analysis also suggests that Y_t contains $3 - 1 = 2$ common I(0) cycles, which can be represented e.g. by the equations for Y_{0t} and Y_{2t}. Recall in fact that the equilibrium dynamics form has I(0) rank equal to $p = 3$. The cofeature restrictions provide a reduction in the number of parameters; the number of coefficients to stochastic regressors in the system are reduced from 9 to 6.

Moreover the finding that Y_{1t} is an innovation process allows to better interpret the adjustment coefficients both in the equilibrium correction and the equilibrium dynamics forms. In fact, the adjustment to $Y_{1,t-1}$ can be interpreted as reaction to the unpredictable autonomous I(1) component in the trend of inflation.

These results on the specification a) in Table 1 show that static cofeatures exist only for the equilibrium dynamics form. The testing-up sequence can further be applied to the equilibrium correction form, where no static cofeatures have been detected.

In this continuation of the analysis, one may wish to analyze submodels b) or c) which are less stringent than a). We observe, however, that in the case $k = 2$, V_t is void, and model c) will always present a cofeature matrix of dimension $p - p_0 = 2$. Similarly model b) will always present a cofeature matrix of dimension $p_2 = 1$. Hence these submodels appear not to be very interesting in the present case.

Despite their limited interest, we briefly comment on how the results in Table 4 already provide tests of common features of the type c). This comment illustrates possible further analysis, which can be of interest especially with more complicated structures. In this case $W_t = Z_{0t} = \Delta^2 X_t$, $R_t = (\beta' \Delta X_{t-1} : Y_{1,t-1})'$.

We observe that the α coefficients in the Δ^2lulc and Δ^2lpm equations are not significant; a joint Wald significance test J_2 in (28) gave a test statistic of 0.16712 with a $\chi^2(2)$ p-value of 0.9198. Hence $b_1 = (e_2 : e_3)$ is the contemporaneous part of the dynamic cofeature matrix of type c). The corresponding estimate of the cofeature matrix is

$$b'Z_t = b_1'W_t + b_2'R_t = \begin{pmatrix} 0 & 1 & 0 & 1.256 & 0.1793 \\ 0 & 0 & 1 & 0.6388 & -0.8655 \end{pmatrix} \begin{pmatrix} \Delta^2 \mathrm{lpfc}_t \\ \Delta^2 \mathrm{lulc}_t \\ \Delta^2 \mathrm{lpm}_t \\ \beta'\Delta X_{t-1} \\ Y_{1,t-1} \end{pmatrix},$$

see Proposition 11. Given that $\varphi_\perp = b_1 = (e_2 : e_3)$, the estimated coefficients $\varphi_\perp'\Phi_1$ are just the coefficients to $\beta'\Delta X_{t-1}$, $Y_{1,t-1}$ in the last 2 equations in the second part of Table 4. Because $\varphi_\perp$ is not estimated, the remaining estimated coefficients have the same standard errors as in the second part of Table 4, and they are all significant.

Summarizing, the application to Australian prices shows the relevance of the equilibrium dynamics form in cofeatures analysis. Only trivial cases of cofeatures could be obtained for the equilibrium correction form. The cofeature analysis attains a reduction in the number of parameters and improves the understanding of the equilibrium relations, and specifically of $Y_{1t} = \beta_1'\Delta X_t$ in this empirical application.

8 Conclusions

In this paper we have discussed various applications of the notion of common features that can possibly arise in I(2) systems. For each possibility we have discussed how to address inference both for known and unknown cofeature vectors, using reduced rank regression. As in the I(1) case, the cointegration analysis needs to precede the analysis of common features. After fixing the cointegration parameters, all subsequent inference is LAN.

The notions of cofeatures introduced in the paper have been found to have empirical relevance on the data-set on Australian inflation analyzed in Banerjee et al. (2001) inter alia. For these data, the equilibrium dynamics form supports the presence of a single cofeature vector, while only trivial cases of cofeatures can be obtained for the equilibrium correction form.

References

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

Anderson, T.W. (1971) *The statistical analysis of time series*, John Wiley and Sons, New York.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change points, *Econometrica* 61, 821-856.

Ahn S. K. and G. C. Reinsel (1998) Nested reduced rank autoregressive models for multiple time series, *Journal of the American Statistical Association* 83, 849-856.

Banerjee, A., L. Cockeress and B. Russell (2001) An I(2) analysis of inflation and the markup, *Journal of Applied Econometrics* 16, 221-240.

Boswijk P. (2000) Mixed normality and ancillarity in I(2) systems, *Econometric Theory*, 16, 878–904.

Cubadda G. (1999) Common cycles in seasonal non-stationary time series *Journal of Applied Econometrics*, 14, 273-291.

Cubadda G. (2001) Common features in time series with both deterministic and stochastic seasonality, *Econometric Reviews*, 20, 201-216.

Cubadda G. and A. Hecq (2001), On non-contemporaneous short-run comovements, *Economics Letters*, 73, 389-397.

Estrella A. (2003) Critical values and p values of Bessel Process Distributions: computation and application to structural break tests, manuscript, Federal Reserve Bank of New York, forthcoming in *Econometric Theory.*

Engle R. F. and C. W. J. Granger (1987). Co-integration and Error Correction: Representation. Estimation and Testing. *Econometrica* 55, 251-276.

Engle R.F. and S. Kozicki (1993) Testing for Common Features, *Journal of Business and Economic Statistics* 11, 369 - 380.

Gouriéroux C. and I. Peaucelle (1993) Séries codépendent: application à l'hypothèse de parité de pouvoir d'achat, in *Macroéconomie: dévelopment récents*, ed. Economica, 285-306.

Haldrup, N. (1998) An Econometric Analysis of I(2) Variables *Journal of Economic Surveys* 12 (5), 595-650.

Hecq, A., F. Palm and J.P. Urbain (2000), Testing for Common Cyclical Features in VAR models with cointegration, WP University of Maastricht, available at `http://www.personeel.unimaas.nl/A.Hecq/`.

Hecq, A., F. Palm and J.P. Urbain (2002), Separation, Weak Exogeneity and P-T Decompositions in Cointegrated VAR Systems with Common Features, *Econometric Reviews*, 21, 3, 273-307.

Johansen S. (1992a) A Representation of Vector Autoregressive Processes Integrated of Order 2. *Econometric Theory* 8, 188-202.

Johansen S. (1992b), Determination of cointegration rank in the presence of a linear trend, *Oxford Bullettin of Economics and Statistics* 54, 383–97.

Johansen S. (1995a) A Statistical Analysis of Cointegration for I(2) variables. *Econometric Theory* 11, 25-59.

Johansen S. (1995b) Identifying restrictions of linear equations with application to simultaneous equations and cointegration, *Journal of Econometrics* 69, 111-132

Johansen S. (1996) *Likelihood–based inference in cointegrated vector autoregressive models.* Second corrected printing, Oxford: Oxford University Press.

Johansen S. (1997) A Likelihood Analysis of the I(2) Model. *Scandinavian Journal of Statistics* 24, 433-462.

Johansen S. (2002) Statistical analysis of hypotheses on the cointegrating relations in the I(2) model, University of Copenhagen, Preprint 13/2002, available at `http://www.math.ku.dk/~sjo`

Kongsted, H.C. (2002), Testing the nominal-to-real transformation, Discussion papers 02/06, University of Copenhagen Institute of Economics, available at `http://www.econ.ku.dk/wpa/`

Kugler P. and K. Neusser (1993) International interest rates equalization: a multivariate time series approach, *Journal of Applied Econometrics* 8, 163-174.

Omtzigt P. and P. Paruolo (2002) Impact factors, WP Università dell'Insubria Varese, Italy, 2002/4.

Paruolo P. (1997) Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems. *Econometric Theory* 13, 79-118.

Paruolo P. (2000) Asymptotic efficiency of the two stage estimator in I(2) systems. *Econometric Theory* 16, 524-550.

Paruolo P. (2001) The power of lambda max, *Oxford Bulletin of Economics and Statistics* 63/3, 395-403.
Paruolo P. (2002a) Asymptotic inference on the moving average impact matrix in cointegrated I(2) VAR systems, *Econometric Theory*, 18, 673-690.
Paruolo P. (2002b) Testing for common trends in conditional I(2) VAR systems, Università dell'Insubria Varese, Italy, WP 2002/28; paper presented at LAMES 2002, São Paolo, Brasil, 24-27 July 2002.
Paruolo P. (2003) Common dynamics in I(1) systems, mimeo, to be presented at the conference in Maastricht, December 14-16, 2003.
Paruolo, P. and A. Rahbek (1999) Weak exogeneity in I(2) VAR systems. *Journal of Econometrics* 93, 281-308.
Rahbek A. (1997), Representation of cointegrated I(2) and I(1) processes with deterministic trends, manuscript, University of Copenhagen.
Rahbek A., Kongsted H.C., and C. Jørgensen (1999) Trend-Stationarity in the I(2) Cointegration Model. *Journal of Econometrics* 90, 265-89.
Srivastava M.S. and C.G. Khatri (1979) *An introduction to multivariate statistics.* New York: Elsevier North Holland.
Stock J. and Watson M. (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783-820.
Tiao G.C. and Tsay R. S. (1989) Model specification in multivariate time series, *Journal of the Royal Statistical Society*, series B, 51, 157-213.
Vahid F. and Engle R.F. (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics* 8, 341-360.
Vahid F. and Engle R.F. (1997), Codependent Cycles, *Journal of Econometrics* 80, 199-221.
Vahid F. and J. Issler (2002) The importance of common cyclical features in VAR analysis: A Monte Carlo study, *Journal of Econometrics* 109, 341-363.

Appendix A: representation

We here report proofs of the propositions in Sections 3 to 5.

Proof. of Proposition 1. Let $W_t = C_W(L)\epsilon_t$. From the definition of cofeature rank, $b'W_t = b'C_{W,0}\epsilon_t$ (i.e. $b'C_{W,i} = 0$, $i \geq 1$) and $V := b'C_{W,0}\Omega C'_{W,0}b$ is positive definite. Because $b'C_{W,i} = 0$, $i \geq 1$, one has $b'C_{W,0} = b'C_W(1)$, and $V = b'C_W(1)\Omega C'_W(1)b$. Because $C_{W,0}$ is assumed to be of full rank, $\ell = \text{rank}(b'C_{W,0}) = \text{rank}(b'C_W(1)) \leq \text{rank}(C_W(1))$. Alternatively, in order for V to be positive definite, $\ell := \text{rank}(b)$ must be less or equal to $\text{rank}(C_W(1)) =: q$, where Ω is of full rank by assumption. ■

Proof. of Theorem 4. Let $u_t := \mu_d d_t + \epsilon_t$, $\Upsilon(L) := I - \sum_{i=1}^{k-2} \Upsilon_i L^i$, $B := (\beta : \beta_1 : \beta_2)$ and

$$Y_t := \begin{pmatrix} Y_{0t} \\ Y_{1t} \\ Y_{2t} \end{pmatrix} := \begin{pmatrix} \beta' X_t + \delta\beta_2'\Delta X_t + \beta_0' t \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \end{pmatrix}.$$

Write the equilibrium correction form (9) as

$$\Upsilon(L)\Delta^2 X_t = \alpha Y_{0t-1} + \zeta_1\beta'\Delta X_{t-1} + \zeta_2\beta_1'\Delta X_{t-1} + \mu_0 + u_t \tag{32}$$

Insert $I = \bar{B}B'$ between $\Upsilon(L)$ and $\Delta^2 X_t$ in the l.h.s. of (32); one finds

$$\left(\Upsilon(L)\bar{\beta}\Delta : \Upsilon(L)\bar{\beta}_1\Delta : \Upsilon(L)\bar{\beta}_2\right)\begin{pmatrix} \beta'\Delta X_t \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \end{pmatrix} = (\alpha : \zeta_1 : \zeta_2)\begin{pmatrix} Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} + \mu_0 + u_t.$$

and rearranging with $\varrho_1(L) := \Upsilon(L)\bar{\beta}_1\Delta - \zeta_2 L$

$$\left(\Upsilon(L)\bar{\beta}\Delta - \zeta_1 L : \varrho_1(L) : \Upsilon(L)\bar{\beta}_2\right)\begin{pmatrix} \beta'\Delta X_t \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \end{pmatrix} = \alpha Y_{0t-1} + \mu_0 + u_t. \tag{33}$$

Let

$$A := \begin{pmatrix} I_{p_0} & & \delta \\ & I_{p_1} & \\ & & I_{p_2} \end{pmatrix} \qquad \text{with} \qquad A^{-1} = \begin{pmatrix} I_{p_0} & & -\delta \\ & I_{p_1} & \\ & & I_{p_2} \end{pmatrix},$$

and note that $D := AB' = (\beta + \beta_2\delta' : \beta_1 : \beta_2)'$, $D^{-1} = \bar{B}A^{-1} = (\bar{\beta} : \bar{\beta}_1 : \bar{\beta}_2 - \bar{\beta}\delta)$. Insert $I = A^{-1}A$ between the two factors in the l.h.s. of (33) and add $(\Upsilon(L)\bar{\beta}\Delta - \zeta_1 L)\beta_0' = -\zeta_1\beta_0'$ on both sides of the equations. Let also $\varrho_2(L) := \Upsilon(L)\bar{\beta}_2 - (\Upsilon(L)\bar{\beta}\Delta - \zeta_1 L)\delta$; one finds

$$\left(\Upsilon(L)\bar{\beta}\Delta - \zeta_1 L : \varrho_1(L) : \varrho_2(L)\right)\begin{pmatrix} \beta'\Delta X_t + \delta\beta_2'\Delta^2 X_t + \beta_0' \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \end{pmatrix} = \alpha Y_{0t-1} + (\mu_0 - \zeta_1\beta_0') + u_t.$$

Let $\varrho_0(L) := (\Upsilon(L)\bar{\beta}\Delta - \zeta_1 L)\Delta - \alpha L$, $\varrho(L) := (\varrho_0(L) : \varrho_1(L) : \varrho_2(L))$; rearranging one finds

$$(\varrho_0(L) : \varrho_1(L) : \varrho_2(L))\begin{pmatrix} Y_{0t} \\ Y_{1t} \\ Y_{2t} \end{pmatrix} = (\mu_0 - \zeta_1\beta_0') + u_t.$$

In order to normalize the zero-lag matrix of the VAR to be the identity, one needs to pre-multiply by D, so that the VAR equations read

$$D\,(\varrho_0(L) : \varrho_1(L) : \varrho_2(L))\begin{pmatrix} Y_{0t} \\ Y_{1t} \\ Y_{2t} \end{pmatrix} = D(\mu_0 - \zeta_1\beta_0') + Du_t.$$

Spelling out the coefficients of the lag polynomial for the first block of $\rho(L) := (\rho_0(L) : \rho_1(L) : \rho_2(L))$ one finds

$$\begin{aligned}
\rho_0(L) &= \bar{\beta}(1 - 2L + L^2) - \sum_{i=1}^{k-2}\Upsilon_i\bar{\beta}(L^i - 2L^{i+1} + L^{i+2}) - (\zeta_1 + \alpha)L + \zeta_1 L^2 = \\
&= \bar{\beta} - (\zeta_1 + \alpha + 2\bar{\beta} + \Upsilon_1\bar{\beta})L + (\zeta_1 + (I - \Upsilon_2 + 2\Upsilon_1)\bar{\beta})L^2 + \\
&\quad - \sum_{i=3}^{k-2}(\Upsilon_i - 2\Upsilon_{i-1} + \Upsilon_{i-2})\bar{\beta}L^i + (2\Upsilon_{k-2} - \Upsilon_{k-3})\bar{\beta}L^{k-1} - \Upsilon_{k-2}\bar{\beta}L^k
\end{aligned}$$

Similarly for the second and third blocks:

$$\begin{aligned}
\rho_1(L) &= \bar{\beta}_1(1 - L) - \sum_{i=1}^{k-2}\Upsilon_i\bar{\beta}_1(L^i - L^{i+1}) - \zeta_2 L \\
&= \bar{\beta}_1 - (\zeta_2 + \bar{\beta}_1 + \Upsilon_1\bar{\beta}_1)L - \sum_{i=3}^{k-2}(\Upsilon_i - \Upsilon_{i-1})\bar{\beta}_1 L^i + \Upsilon_{k-2}\bar{\beta}_1 L^{k-1}
\end{aligned}$$

$$
\begin{aligned}
\rho_2(L) &= \Upsilon(L)(\bar{\beta}_2 - \bar{\beta}\delta) + \Upsilon(L)\bar{\beta}\delta L + \zeta_1\delta L \\
&= \bar{\beta}_2 - \bar{\beta}\delta + (\bar{\beta} + \zeta_1)\delta L - \sum_{i=1}^{k-2}\Upsilon_i(\bar{\beta}_2 - \bar{\beta}\delta)L^i + \sum_{i=1}^{k-2}\Upsilon_i\bar{\beta}\delta L^{i+1} = \\
&= (\bar{\beta}_2 - \bar{\beta}\delta) + ((\bar{\beta} + \zeta_1)\delta - \Upsilon_1(\bar{\beta}_2 - \bar{\beta}\delta))L + \\
&\quad - \sum_{i=2}^{k-2}(\Upsilon_i(\bar{\beta}_2 - \bar{\beta}\delta) - \Upsilon_{i-1}\bar{\beta}\delta)L^i + \Upsilon_{k-2}\bar{\beta}\delta L^{k-1}.
\end{aligned}
$$

The AR matrices are thus

$$
\begin{aligned}
A_1^\circ &= D\left(\zeta_1 + \alpha + (2I + \Upsilon_1)\bar{\beta} : \zeta_2 + (I + \Upsilon_1)\bar{\beta}_1 : -(\bar{\beta} + \zeta_1)\delta + \Upsilon_1(\bar{\beta}_2 - \bar{\beta}\delta)\right), \\
A_2^\circ &= D\left(-\zeta_1 - (I - \Upsilon_2 + 2\Upsilon_1)\bar{\beta} : (\Upsilon_2 - \Upsilon_1)\bar{\beta}_1 : \Upsilon_2(\bar{\beta}_2 - \bar{\beta}\delta) - \Upsilon_1\bar{\beta}\delta)\right),
\end{aligned}
$$

$$
A_i^\circ = D\left((\Upsilon_i - 2\Upsilon_{i-1} + \Upsilon_{i-2})\bar{\beta} : (\Upsilon_i - \Upsilon_{i-1})\bar{\beta}_1 : \Upsilon_i(\bar{\beta}_2 - \bar{\beta}\delta) - \Upsilon_{i-1}\bar{\beta}\delta\right),
$$

$i = 3, ..., k-2$

$$
\begin{aligned}
A_{k-1}^\circ &= D\left((-2\Upsilon_{k-2} + \Upsilon_{k-3})\bar{\beta} : -\Upsilon_{k-2}\bar{\beta}_1 : -\Upsilon_{k-2}\bar{\beta}\delta\right), \\
A_k^\circ &= D\left(\Upsilon_{k-2}\bar{\beta} : 0 : 0\right).
\end{aligned}
$$

where $D := (\beta + \beta_2\delta' : \beta_1 : \beta_2)'$. These expressions imply the restrictions (11). In order to impose them, observe that $A_{k-1,0}^\circ Y_{0,t-k+1} + A_{k-1,2}^\circ Y_{2,t-k+1} + A_{k,0}^\circ Y_{0,t-k}$ equals

$$
\begin{aligned}
& D(-2\Upsilon_{k-2} + \Upsilon_{k-3})\bar{\beta}Y_{0,t-k+1} - D\Upsilon_{k-2}\bar{\beta}\delta Y_{2,t-k+1} + D\Upsilon_{k-2}\bar{\beta}Y_{0,t-k} = \\
= \; & -D\Upsilon_{k-2}\bar{\beta}(Y_{0,t-k+1} - Y_{0,t-k} + \delta Y_{2,t-k+1}) + D(-\Upsilon_{k-2} + \Upsilon_{k-3})\bar{\beta}Y_{0,t-k+1} = \\
= \; & -D\Upsilon_{k-2}\bar{\beta}(\Delta\beta' X_{t-k+1} + \beta_0') + D(-\Upsilon_{k-2} + \Upsilon_{k-3})\bar{\beta}Y_{0,t-k+1},
\end{aligned}
$$

so that one can simply substitute Y_{t-k}, $Y_{2,t-k+1}$ with $\Delta\beta' X_{t-k+1}$, changing the coefficients of the constant and of $Y_{0,t-k+1}$.

The stability of the roots of the AR polynomial $A^\circ(L)$ under the I(2) assumptions and that Y_t is an I(0) process of rank p are proved in Johansen's I(2) representation theorem, see Johansen (1992a, 1996) or Paruolo (2002b) Appendix 1. These references also describe how to transform Y_t back to the autoregressive form, and hence to the equilibrium correction form (10).

We finally show how the mixed form can be obtained. Let $u_t^* := \gamma V_t + \mu_d d_t + \epsilon_t$. One has

$$
\Delta^2 X_t = \alpha Y_{0t-1} + \zeta_1\beta'\Delta X_{t-1} + \zeta_2\beta_1'\Delta X_{t-1} + \mu_0 + u_t^*
$$

Insert $I = D^{-1}D = \bar{B}A^{-1}AB'$ before $\Delta^2 X_t$ in the l.h.s.; one finds

$$
(\bar{\beta}\Delta : \bar{\beta}_1\Delta : \bar{\beta}_2 - \bar{\beta}\delta\Delta)\begin{pmatrix} \beta'\Delta X_t + \delta\beta_2'\Delta^2 X_t \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \end{pmatrix} = (\alpha : \zeta_1 : \zeta_2)\begin{pmatrix} Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} + \mu_0 + u_t^*.
$$

Adding β_0 to the top block of variables on the l.h.s. and rearranging

$$
(\bar{\beta} : \bar{\beta}_1 : \bar{\beta}_2 - \bar{\beta}\delta)\begin{pmatrix} Y_{0t} \\ Y_{1t} \\ Y_{2t} \end{pmatrix} = \left(\alpha + \bar{\beta} : \zeta_1 + \bar{\beta} : \zeta_2 + \bar{\beta}_1\right)\begin{pmatrix} Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} + \mu_0 + u_t^*
$$

Pre-multiplication by $D := (\beta + \beta_2 \delta' : \beta_1 : \beta_2)'$ gives

$$Y_t = D\left(\alpha + \bar{\beta} : \zeta_1 + \bar{\beta} : \zeta_2 + \bar{\beta}_1\right) \begin{pmatrix} Y_{0,t-1} \\ \beta' \Delta X_{t-1} \\ \beta_1' \Delta X_{t-1} \end{pmatrix} + D\mu_0 + Du_t^*,$$

which is the stated result ■

Proof. of Theorem 6. Let $m_t^* := m(L)\Delta^2 d_t$. Taking second differences in (7) one finds that $\Delta^2 X_t - m_t^*$ equals

$$\begin{aligned} & C_2\epsilon_t + C_1\Delta\epsilon_t + C_0(L)\Delta^2\epsilon_t, \\ = \quad & C_2\epsilon_t + C_1(\epsilon_t - \epsilon_{t-1}) + \sum_{i=0}^{\infty} C_{0,i}L^i\epsilon_t - 2\sum_{i=0}^{\infty} C_{0,i}L^{i+1}\epsilon_t + \sum_{i=0}^{\infty} C_{0,i}L^{i+2}\epsilon_t \\ = \quad & (C_2 + C_1 + C_{0,0})\,\epsilon_t + (-C_1 + C_{0,1} - 2C_{0,0})\,\epsilon_t + \sum_{i=2}^{\infty} (C_{0,i} - 2C_{0,i-1} + C_{0,i-2})\,\epsilon_{t-i} \\ =: \quad & \epsilon_t + C_1^*\epsilon_{t-1} + \sum_{i=2}^{\infty} C_i^*\epsilon_{t-i} =: C^*(L)\epsilon_t, \end{aligned} \tag{34}$$

where in the last line we have used the normalization of the process $C^*(0) = I$, i.e.

$$C_2 + C_1 + C_{0,0} = I. \tag{35}$$

There exist a cofeature matrix b such that $b'(\Delta^2 X_t - m_t^*) = b'\epsilon_t$ if and only if all the coefficient matrices to the lagged ϵ_t in (34) cancel when pre-multiplied by b', i.e. iff $b'C_i^* = 0$, $i = 1, 2, \ldots$ Let $a_i := b'C_{0,i}$. The condition $b'C_i^* = 0$, for $i \geq 2$ is

$$a_j - 2a_{j+1} + a_{j+2} = 0, \qquad j = 0, 1, \ldots$$

This is a difference equation with solution $a_j = a_0 + j(a_1 - a_0)$. From the summability of $C_0(z)$ for $|z| < 1 + \varkappa$ and $\varkappa > 0$, it follows that $a_0 = a_1 - a_0 = 0$, i.e. $b'C_{0,i} = 0$ for all $i \geq 0$, which is condition (14).

The condition $b'C_1^* = 0$ gives $b'(-C_1 + C_{0,1} - 2C_{0,0}) = 0$, where $b'C_{0,1} = b'C_{0,0} = 0$ by (14). Hence one finds $b'C_1 = 0$, condition (15). From (35) one has $C_1 = I - C_2 - C_{0,0}$, so that $b'C_1 = b'(I - C_2 - C_{0,0}) = b'(I - C_2)$, where the last equality follows from (14). This proves the equivalence between (15) and (16).

Assume (16) holds, $b'C_2 = b'$. From the definition of C_2, see (7), it follows that $b \in \operatorname{col}(\alpha_2)$, i.e. that $b = \alpha_2 u$ for some u. Substituting into $b'C_2 = b'$ one finds $u'(\alpha_2'\beta_2(\alpha_2'\theta\beta_2)^{-1} - I_{p_2})\alpha_2' = 0$, which holds iff $u'(\alpha_2'\beta_2 - \alpha_2'\theta\beta_2) = 0$, i.e. if c belongs to $\mathcal{A} := \operatorname{col}^{\perp}(\alpha_2'(I - \theta)\beta_2)$. In order for $\mathcal{A}$ not to contain only the zero vector, $\alpha_2'(I - \theta)\beta_2$ must be of deficient rank, i.e. $\alpha_2'(I - \theta)\beta_2 = cd'$ for some full column rank $p_2 \times p_2 - \ell$ matrices c and d. Hence $u = c_{\perp}$. The converse statement is direct. This completes the proof. ■

Proof. of Proposition 5. If $\Psi^{(\cdot)} = \varphi\tau'$ then $\varphi_{\perp}'(W_t - E(W_t))$ is an innovation process. Conversely assume W_t has cofeature matrix b, i.e. $b'(W_t - E(W_t))$ is an innovation process. From (9) and (10) one finds that $b'(W_t - E(W_t))$ contains $b'\Psi U_t$ or $b'\Psi U_t^{\circ}$ in addition to an innovation process. Hence $b'\Psi = 0$, i.e. $b \in \operatorname{col}^{\perp}(\Psi)$. In order b to be different from the zero vector one must have $\operatorname{rank}(\Psi) = p - \ell$, i.e. $\Psi = \varphi\tau'$. This completes the proof. ■

Proof. of Proposition 9. Let $W_t = C_W(L)\epsilon_t$. Because b is a cofeature matrix, one has $(b_1' : b_2')(W_t' : R_t')' = b_1' C_{W,0}\epsilon_t$, given that R_t does not depend on ϵ_t. Moreover $V = var(b'Z_t) = b_1' C_{W,0}\Omega C_{W,0}' b_1$ is of full rank ℓ. This holds only if b_1 has full column rank ℓ. This completes the proof. ■

Proof. of Proposition 10. We wish to show U_{2t} can be obtained linearly from U_{1t} and vice versa. To this end simply observe that

$$\begin{pmatrix} Y_{0t} \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \\ Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} = \begin{pmatrix} \beta_0' \\ \\ \\ \\ \\ \\ \end{pmatrix} + \begin{pmatrix} \beta' + \delta\beta_2' & I_{p_0} & I_{p_0} & \\ \beta_1' & & & I_{p_1} \\ \beta_2' & & & \\ & I_{p_0} & & \\ & & I_{p_0} & \\ & & & I_{p_1} \end{pmatrix} \begin{pmatrix} \Delta^2 X_t \\ Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix},$$

where we have omitted zeros for readability. Conversely

$$\begin{pmatrix} \Delta^2 X_t \\ Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix} = \begin{pmatrix} -\bar\beta\beta_0' \\ \\ \\ \\ \end{pmatrix} + \begin{pmatrix} \bar\beta & \bar\beta_1 & \bar\beta_2 - \bar\beta\delta & -\bar\beta & -\bar\beta & -\bar\beta_1 \\ & & & I_{p_0} & & \\ & & & & I_{p_0} & \\ & & & & & I_{p_1} \end{pmatrix} \begin{pmatrix} Y_{0t} \\ \beta_1'\Delta X_t \\ \beta_2'\Delta^2 X_t \\ Y_{0,t-1} \\ \beta'\Delta X_{t-1} \\ \beta_1'\Delta X_{t-1} \end{pmatrix}.$$

This completes the proof. ■

Proof. of Proposition 11. Sufficiency is proved by substituting $\varsigma = \varphi\tau'$ in (18) and pre-multiplication by $\varphi_\perp'$. In order to prove necessity, assume $b := (b_1' : b_2')'$ is the cofeature matrix with $\ell > 0$ columns, and $b_1' Z_{0t} + b_2' Z_{2t} = b_1' u_t$ be the cofeature relations. Pre-multiplication of (18) by b_1' gives $b_1' Z_{0t} = b_1'\varsigma Z_{1t} + b_1'\Phi Z_{2t} + b_1' u_t$, which, substituted back implies

$$-b_1'\varsigma Z_{1t} + (b_2' - b_1'\Phi) Z_{2t} = 0.$$

In order for this to be zero for any t, one needs both coefficients of Z_{1t} and Z_{2t} to be zero. This shows that $b_1 \in \mathrm{col}^\perp(\varsigma)$ and that $b_2' = b_1'\Phi$. Since $\ell > 0$ was assumed, ς must be of deficient rank, $\varsigma = \varphi\tau'$, and $b_1 = \varphi_\perp$. ■

Appendix B: inference

In this appendix we report proofs that the tests J_i, Q_i are asymptotically χ^2. Similar arguments lead to the $T^{1/2}$ asymptotic normality of the maximum likelihood estimators. The appendix is organized in two parts. The first step is to prove results when the cointegration relations are known and fixed at their true value. This part is documented in various sources in the literature; it is reproduced here for completeness and for further reference. The second step is to show that the effect of estimation of the cointegration coefficients vanishes asymptotically, so that the limit distributions are the same as the ones for known cointegration coefficients.

The data generating process is taken to be

$$\begin{aligned} Z_{0t} &= \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \\ \varsigma &= \varphi\tau'. \end{aligned} \tag{36}$$

for an appropriate definition of Z_{0t}, Z_{1t}, Z_{2t}. The coefficient matrix ς is $p \times j$ of reduced rank s, and the matrices φ and τ are of dimension $p \times s$ and $j \times s$ and full column rank s. We assume $j \geq p$ in order to exclude trivial cases.

The stochastic variables in Z_{0t}, Z_{1t}, Z_{2t} are selected from a stationary process with companion form $Z_t = AZ_{t-1} + \epsilon_t^{\dagger}$, $\epsilon_t^{\dagger} := (\widetilde{\epsilon}_t' : 0')'$, where the eigenvalues of A are all inside the unit disk. For details of these companion forms in the I(1) and I(2) cases see Omtzigt and Paruolo (2002).

We use the notation $S_{ij} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $M_{ij} := T^{-1}\sum_{t=1}^{T}(Z_{it} - m_i)(Z_{jt} - m_j)'$, $m_i := T^{-1}\sum_{t=1}^{T} Z_{it}$, i, $j = 0, 1, 2$; we also write $S_{i\epsilon}$ when Z_{jt} is substituted by ϵ_t^*. Moreover $\widehat{Z}_{it}$ (resp. $\widehat{S}_{ij}$) indicates Z_{it} (resp. S_{ij}) calculated at the estimated values of the cointegration coefficients. Similarly for $\widehat{S}(\widehat{\lambda}) := \widehat{\lambda}\widehat{S}_{00} - \widehat{S}_{01}\widehat{S}_{11}^{-1}\widehat{S}_{10}$. We indicate by $\widehat{Q}_i$, $\widehat{J}_i$ the LR and Wald test statistic based on estimated cointegration coefficients. The maximum likelihood estimates (for the true dimension s) are indicated with a hat $\widehat{\ }$. We distinguish the estimators based on $\widehat{S}_{ij}$ in place of S_{ij} with a double hat, $\widehat{\widehat{\ }}$.

Let $\Sigma_{ij}^* := E((Z_{it} - E(Z_{it}))(Z_{jt} - E(Z_{jt}))')$, $\Sigma_{ij} := \Sigma_{ij}^* - \Sigma_{i2}^*\Sigma_{22}^{*-1}\Sigma_{2j}^*$. We collect the basic behavior of the various sample moment matrices in the following lemma, whose proof can be found e.g. in Anderson (1971) Chapter 5.

Lemma 13 *The following convergences hold*

$$\begin{aligned} S_{00} \xrightarrow{p} \Sigma_{00} &= \varphi\tau'\Sigma_{11}\tau\varphi' + \Omega^* \qquad S_{01} \xrightarrow{p} \varphi\tau'\Sigma_{11} \qquad S_{11} \xrightarrow{p} \Sigma_{11} \\ T^{1/2}vec(\varphi_\perp' S_{01}) &= T^{1/2}vec(\varphi_\perp' S_{\epsilon 1}) =: vec(C_T') \xrightarrow{d} N(0, \Sigma_{11} \otimes \varphi_\perp'\Omega^*\varphi_\perp). \end{aligned}$$

We next present results for known cointegration coefficients in Propositions 14, 15, 16 and 17.

Proposition 14 *For known cointegration coefficients $Q_1(s) \xrightarrow{d} \chi^2(df_{Q_1})$ under the null, where $df_{Q_1(s)} := (p-s)(j-s) = (j-s)\ell$. Under fixed alternatives, the test diverges.*

Proof. Let $B := (\bar{\varphi} : \varphi_\perp)$, of full rank, and consider $0 = |S(\lambda)| = |B'S(\lambda)B|$, since $|B| \neq 0$. One has, by the results in Lemma 13, that

$$|B'S(\lambda)B| = \left| \lambda B'\Sigma_{00}B - \begin{pmatrix} \tau'\Sigma_{11}\tau & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & O_p(T^{-1}) \end{pmatrix} \right| + o_p(1)$$

which shows that $(\lambda_1, ..., \lambda_s) = O_p(1)$ and $(\lambda_{s+1}, ..., \lambda_n) = O_p(T^{-1})$, and $\text{col}(\widehat{\varphi}_\perp) \xrightarrow{p} \text{col}(\varphi_\perp)$, i.e that $\widehat{\varphi}_\perp$ is consistent. This implies that $Q_1(s-i) \to \infty$ for any positive i, i.e. that the probability to select the number of columns in τ smaller than s goes to zero asymptotically.

Consider the last $n - s$ roots and let $\rho := T\lambda$, $B_T := (\bar{\varphi} : T^{1/2}\varphi_\perp)$; we also use the shorthand $S_\rho := S\left(\frac{\rho}{T}\right)$. One finds

$$|B_T'S_\rho B_T| = |\bar{\varphi}'S_\rho\bar{\varphi}| \left| T\varphi_\perp' \left(S_\rho - S_\rho\bar{\varphi}(\bar{\varphi}'S_\rho\bar{\varphi})^{-1}\bar{\varphi}'S_\rho\right)\varphi_\perp \right|,$$

where in the first factor $\bar{\varphi}'S_\rho\bar{\varphi} = -\tau'\Sigma_{11}\tau + o_p(1)$, and hence all the $n - s$ smallest roots come from the second factor in the limit.

Recall that $C_T' := T^{1/2}\varphi_\perp' S_{\epsilon 1}$; by the results in Lemma 13 one has

$$T^{1/2}\varphi_\perp' S_\rho \bar{\varphi} = -C_T'\tau + o_p(1) \qquad T\varphi_\perp' S_\rho \varphi_\perp = \rho\varphi_\perp'\Sigma_{00}\varphi_\perp - C_T'\Sigma_{11}^{-1}C_T + o_p(1)$$

and hence the second factor converges to

$$\begin{aligned} &\left|\rho\varphi_\perp'\Sigma_{00}\varphi_\perp - C_T'(\Sigma_{11}^{-1} - \tau(\tau'\Sigma_{11}\tau)^{-1}\tau')C_T\right| = \\ = \; &\left|\rho\varphi_\perp'\Sigma_{00}\varphi_\perp - C_T'\Sigma_{11}^{-1}\tau_\perp(\tau_\perp'\Sigma_{11}^{-1}\tau_\perp)^{-1}\tau_\perp'\Sigma_{11}^{-1}C_T\right| =: |\rho F - G| \end{aligned} \tag{37}$$

where we have used the non-orthogonal projection identity $I - \Sigma_{11}\tau(\tau'\Sigma_{11}\tau)^{-1}\tau' = \tau_\perp(\tau_\perp'\Sigma_{11}^{-1}\tau_\perp)^{-1}\tau_\perp{}'\Sigma_{11}^{-1}$, see Srivastava and Khatri (1979) p. 19. By a first order Taylor expansion

$$Q_1(s) := -T\sum_{i=s+1}^{n}\ln(1-\lambda_i) = T\sum_{i=s+1}^{n}\lambda_i + o_p(1) = \sum_{i=1}^{n-s}\rho_i + o_p(1) = tr(F^{-1}G) + o_p(1),$$

where F and G are defined in (37). Next note that by Lemma 13, $\varphi_\perp'\Sigma_{00}\varphi_\perp = \varphi_\perp'\Omega^*\varphi_\perp$ and that $(\varphi_\perp'\Omega^*\varphi_\perp)^{-1/2}C_T'\Sigma_{11}^{-1}\tau_\perp(\tau_\perp'\Sigma_{11}^{-1}\tau_\perp)^{-1/2} \xrightarrow{d} N(0, I_{(p-s)(j-s)})$ and hence

$$tr(F^{-1}G) = tr\left((\varphi_\perp'\Omega^*\varphi_\perp)^{-1}C_T'\Sigma_{11}^{-1/2}\mathcal{A}_1\Sigma_{11}^{-1/2}C_T\right) \xrightarrow{d} \chi^2((p-s)(j-s)). \tag{38}$$

where $\mathcal{A}_1 := \Sigma_{11}^{-1/2}\tau_\perp(\tau_\perp'\Sigma_{11}^{-1}\tau_\perp)^{-1}\tau_\perp'\Sigma_{11}^{-1/2}$. ■

Proposition 15 *For known cointegration coefficients $Q_2 \xrightarrow{d} \chi^2(df_{Q_2})$ under the null, where $df_{Q_2} := 2ps - s^2 - 2p(p-h) - (p-s-h)(2h-p+s)$. Under a fixed alternative, the test diverges.*

Proof. The null hypothesis (26) can be written as $\varphi = (H_\perp : \bar{H}\phi_\perp)$. Let $\varphi_1 := H_\perp$, $\varphi_2 := \bar{H}\phi_\perp$, $c := \varphi'\varphi_2$, $\tau_2 := \tau c$, $\tau_1 := \bar{\tau}c_\perp$, and note that $\text{col}(\tau) = \text{col}(\tau_1 : \tau_2)$. Next apply the format of the proof of Proposition 14 with $|H'S(\lambda^*)H| = 0$ in place of $|S(\lambda)| = 0$ and B replaced by $((H'H)^{-1}\phi_\perp : \phi)$. Similarly replace τ with τ_2 in (37); one finds $-T\sum_{i=h-p+s+1}^{h}\ln(1-\lambda_i^*) = tr((\varphi_\perp'\Sigma_{00}\varphi_\perp)^{-1}C_T'\Sigma_{11}^{-1/2}\mathcal{A}_2\Sigma_{11}^{-1/2}C_T) + o_p(1)$, where $\mathcal{A}_2 := \Sigma_{11}^{-1/2}\tau_{2\perp}(\tau_{2\perp}'\Sigma_{11}^{-1}\tau_{2\perp})^{-1}\tau_{2\perp}'\Sigma_{11}^{-1/2}$, where $\tau_{2\perp}$ can be chosen as $(\tau_\perp : \tau_1)$. Hence $Q_2 = tr((\varphi_\perp'\Sigma_{00}\varphi_\perp)^{-1}C_T'\Sigma_{11}^{-1/2}(\mathcal{A}_2 - \mathcal{A}_1)\Sigma_{11}^{-1/2}C_T) + o_p(1)$. The $\mathcal{A}_1$ and $\mathcal{A}_2$ matrices are orthogonal projectors onto $\text{col}(\Sigma_{11}^{-1/2}\tau_\perp)$ and $\text{col}(\Sigma_{11}^{-1/2}\tau_{2\perp})$ respectively, where $\text{col}(\Sigma_{11}^{-1/2}\tau_\perp) \subset \text{col}(\Sigma_{11}^{-1/2}\tau_{2\perp}) = \text{col}(\Sigma_{11}^{-1/2}(\tau_\perp : \tau_1))$. Because orthogonal projectors are invariant to the choice of bases of the linear spaces, we wish here to choose a basis for $\text{col}(\Sigma_{11}^{-1/2}\tau_{2\perp}) = \text{col}(\Sigma_{11}^{-1/2}(\tau_\perp : \tau_1))$ of the type $(f_1 : g_1)$, with $f_1 := \Sigma_{11}^{-1/2}\tau_\perp$ and g_1 orthogonal to f_1. This implies $P_{(f_1:g_1)} = P_{f_1} + P_{g_1}$, where $P_a := a(a'a)^{-1}a'$ for a of full column rank. It is simple to see that g_1 can e.g. be chosen as

$$g_1 := \Sigma_{11}^{1/2}\tau(\tau'\Sigma_{11}\tau)^{-1}\tau'\tau_1 =: q_1c_1,$$

where $q_1 = \Sigma_{11}^{1/2}\tau(\tau'\Sigma_{11}\tau)^{-1}$, $c_1 := \tau'\tau_1$. In fact $f_1'g_1 = \tau_\perp'\Sigma_{11}^{-1/2}\,\Sigma_{11}^{1/2}\tau(\tau'\Sigma_{11}\tau)^{-1}\tau'\tau_1 = 0$ i.e. g_1 is perpendicular to f_1. Moreover g_1 is perpendicular to the orthogonal complement of $\Sigma_{11}^{-1/2}\tau_{2\perp}$, where $\text{col}^\perp(\Sigma_{11}^{-1/2}\tau_{2\perp}) = \text{col}(\Sigma_{11}^{1/2}\tau_2)$. In fact

$$\tau_2'\Sigma_{11}^{1/2}\Sigma_{11}^{1/2}\tau(\tau'\Sigma_{11}\tau)^{-1}\tau'\tau_1 = c'\tau'\Sigma_{11}\tau(\tau'\Sigma_{11}\tau)^{-1}\tau'\bar{\tau}c_\perp = c'c_\perp = 0.$$

Hence we conclude that $(f_1 : g_1)$ is a basis of $\text{col}(\Sigma_{11}^{-1/2}\tau_{2\perp})$. Hence $\mathcal{A}_2 - \mathcal{A}_1 = P_{g_1}$ where

$$\Sigma_{11}^{-1/2} P_{g_1} \Sigma_{11}^{-1/2} = \tau(\tau'\Sigma_{11}\tau)^{-1} c_1 (c_1'(\tau'\Sigma_{11}\tau)^{-1} c_1)^{-1} c_1'(\tau'\Sigma_{11}\tau)^{-1}\tau'.$$

One sees that $L_T := (\varphi_\perp'\Omega^*\varphi_\perp)^{-1/2} C_T'\Sigma_{11}^{-1}\tau(\tau'\Sigma_{11}\tau)^{-1}c_1(c_1'(\tau'\Sigma_{11}\tau)^{-1}c_1)^{-1/2} \xrightarrow{d} N(0, I_{df_{Q_2}})$ so that

$$Q_2 = tr(L_T L_T') + o_p(1) \xrightarrow{d} \chi^2(df_{Q_2}).$$

Under a fixed alternative, at least one of the λ^* eigenvalues does not converge to zero, which implies that Q_2 diverges. This completes the proof. ■

Proposition 16 *For known cointegration coefficients $Q_3 \xrightarrow{d} \chi^2(df_{Q_3})$ under the null, where $df_{Q_3} := sh$. Moreover $Q_4 := Q_1(s) + Q_3 \xrightarrow{d} \chi^2(df_{Q_4})$ under the null, where $df_{Q_4} = df_{Q_1(s)} + df_{Q_3}$. Under a fixed alternative, Q_3 and Q_4 diverge.*

Proof. Observe that the null can be written $\varphi_\perp := (\varphi_{\perp 1} : \varphi_{\perp 2}) := (K : K_\perp\psi)$ or $\varphi = \bar{K}_\perp\psi_\perp$. Indicate the three terms on the r.h.s. in (30) as a, b, c, respectively, $Q_3 = a + b + c$. Using the results in Lemma 13, one finds that

$$c := -T\ln\frac{|K'S_{00.1}K|}{|K'S_{00}K|} = tr\left((\varphi_{\perp 1}'\Omega^*\varphi_{\perp 1})^{-1} C_{1T}'\Sigma_{11}^{-1}C_{1T}\right) + o_p(1)$$

where $C_{1T} := T^{1/2}S_{1\epsilon}K$, because $K \in \text{col}(\varphi_\perp)$ under the null. Under the alternative c is seen to diverge.

Consider now the second term $b := -T\sum_{i=s+1}^{h-p+s}\ln(1-\lambda_i^\circ)$, and apply the format of the proof of Proposition 14 with (29) in place of $|S(\lambda)| = 0$ and B replaced by $((K_\perp'K_\perp)^{-1}\psi_\perp(\varphi'\varphi)^{-1} : \psi)$.

One has $S_{11.K} = \Sigma_{11} + o_p(1)$ because $S_{10}K = O_p(T^{-1/2})$ given that $K \in \text{col}(\varphi_\perp)$ under the null. Similarly one finds $\psi'K_\perp'S_{01.K} = \psi'K_\perp'(I - S_{00}K(K'S_{00}K)^{-1}K')S_{01} = \omega'S_{\epsilon 1} + o_p(T^{-1/2})$, where $\omega' := \varphi_{\perp 2}'(I - \Omega^*\varphi_{\perp 1}(\varphi_{\perp 1}'\Omega^*\varphi_{\perp 1})^{-1}\varphi_{\perp 1}')$ and $(\varphi'\varphi)^{-1}\psi_\perp'\bar{K}_\perp'S_{01.K} = (\varphi'\varphi)^{-1}\psi_\perp'\bar{K}_\perp'(I - S_{00}K(K'S_{00}K)^{-1}K')S_{01} = \bar{\varphi}'S_{01} + o_p(1) = \tau'\Sigma_{11} + o_p(1)$.

Let $C_{2T} := T^{1/2}S_{1\epsilon}\omega'$, and recall $\mathcal{A}_1 := \Sigma_{11}^{-1/2}\tau_\perp(\tau_\perp'\Sigma_{11}^{-1}\tau_\perp)^{-1}\tau_\perp'\Sigma_{11}^{-1/2}$, see eq. (38). One finds for the last $h - p$ eigenvalues λ_i° in (29) that

$$b := -T\sum_{i=s+1}^{h-p+s}\ln(1-\lambda_i^\circ) = tr\left((\omega'\Omega^*\omega)^{-1} C_{2T}'\Sigma_{11}^{-1/2}\mathcal{A}_1\Sigma_{11}^{-1/2}C_{2T}\right) + o_p(1).$$

The asymptotic expansion of a is found in (38). Add and subtract d from $Q_3 := a+b+c$, where $d := tr\left((\varphi_{\perp 1}'\Omega^*\varphi_{\perp 1})^{-1} C_{1T}'\Sigma_{11}^{-1/2}\mathcal{A}_1\Sigma_{11}^{-1/2}C_{1T}\right)$; one finds that $b+d = a$, so that $Q_3 = c - d$, i.e.

$$\begin{aligned} Q_3 &= tr\left((\varphi_{\perp 1}'\Omega^*\varphi_{\perp 1})^{-1} C_{1T}'\Sigma_{11}^{-1/2}(I - \mathcal{A}_1)\Sigma_{11}^{-1/2}C_{1T}\right) + o_p(1) = \\ &= tr\left((\varphi_{\perp 1}'\Omega^*\varphi_{\perp 1})^{-1} C_{1T}'\tau(\tau'\Sigma_{11}\tau)^{-1}\tau'C_{1T}\right) + o_p(1) \xrightarrow{d} \chi^2(sh), \qquad (39) \end{aligned}$$

where we have used orthogonal projections. In order to prove that $Q_4 := Q_1(s) + Q_3$ is a $\chi^2(df_{Q_1(s)} + df_{Q_3})$ it is enough to note that $C_{1T}'\tau$ in (39) and $C_T'\Sigma_{11}^{-1}\tau_\perp$ in (38) are asymptotically jointly normal with 0 covariance, and hence independent. This completes the proof. ■

Proof. of Proposition 12. Recall that $\widehat{\varphi}_\perp = (u_{s+1} : ... : u_p)$ satisfies

$$S_{00}\widehat{\varphi}_\perp \Lambda_2 = S_{01}S_{11}^{-1}S_{10}\widehat{\varphi}_\perp. \tag{40}$$

Consistency of $\mathrm{col}(\widehat{\varphi}_\perp) \xrightarrow{p} \mathrm{col}(\varphi_\perp)$ has been shown in the proof of Proposition 14. By the assumption of $a'_\perp\varphi_\perp$ of full rank, $c := a'_\perp\widehat{\varphi}_\perp$ converges in probability to a full rank square matrix. Post-multiply (40) by c^{-1} one finds $S_{00}\widehat{\varphi}_\perp(\Lambda_2 c^{-1}) = S_{01}S_{11}^{-1}S_{10}\widehat{\varphi}_{\perp a_\perp}$, where $S_{00}\widehat{\varphi}_\perp = O_p(1)$, $c = O_p(1)$ and $\Lambda_2 := diag(\lambda_{s+1}, ..., \lambda_p) = O_p(T^{-1})$. This shows that

$$S_{01}S_{11}^{-1}S_{10}\widehat{\varphi}_{\perp a_\perp} = O_p(T^{-1}) \tag{41}$$

Insert next the non-orthogonal projection identity

$$I = \varphi_\perp(a'_\perp\varphi_\perp)^{-1}a'_\perp + a(\varphi' a)^{-1}\varphi' = (a : \varphi_{\perp a_\perp})(\varphi_a : a_\perp)'.$$

before $\widehat{\varphi}_{\perp a_\perp}$ to obtain $\widehat{\varphi}_{\perp a_\perp} = (a : \varphi_{\perp a_\perp})(\varphi_a : a_\perp)'\widehat{\varphi}_{\perp a_\perp} = a\varphi'_a\widehat{\varphi}_{\perp a_\perp} + \varphi_{\perp a_\perp} =: \varphi_{\perp a_\perp} + ah$ because $a'_\perp\widehat{\varphi}_{\perp a_\perp} = I_{p-s}$, where $h := \varphi'_a\widehat{\varphi}_{\perp a_\perp} = \bar{a}'(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})$.

Substituting into (41) one finds $S_{01}S_{11}^{-1}S_{10}(\varphi_{\perp a_\perp} + ah) = O_p(T^{-1})$. Pre-multiplying by a' and rearranging one finds

$$T^{1/2}h = -(a'S_{01}S_{11}^{-1}S_{10}a)^{-1}a'S_{01}S_{11}^{-1}(T^{1/2}S_{10}\varphi_{\perp a_\perp}) + O_p(T^{-1/2}).$$

By the results in Lemma 13, $(a'S_{01}S_{11}^{-1}S_{10}a)^{-1}a'S_{01}S_{11}^{-1} \xrightarrow{p} (a'\varsigma\Sigma_{11}\varsigma' a)^{-1}a'\varsigma$ and

$$vec(T^{1/2}S_{10}\varphi_{\perp a_\perp}) = vec(C_T(a'_\perp\varphi_\perp)^{-1}) \xrightarrow{d} N(0, \varphi'_{\perp a_\perp}\Omega^*\varphi_{\perp a_\perp} \otimes \Sigma_{11}).$$

Hence

$$T^{1/2}vec(h) := T^{1/2}vec(\bar{a}'(\widehat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})) \xrightarrow{d} N(0, \varphi'_{\perp a_\perp}\Omega^*\varphi_{\perp a_\perp} \otimes (a'\varsigma\Sigma_{11}\varsigma' a)^{-1}).$$

A different proof can be given exploiting the relation with $\widehat{\varphi}$ along the lines in Paruolo (1997), see also Paruolo (2003). ■

We next give results for the Wald tests.

Proposition 17 *For known cointegration coefficients* $J_1 \xrightarrow{d} \chi^2(df_{J_1})$ *and* $J_2 \xrightarrow{d} \chi^2(df_{J_2})$ *under the null, where* $df_{J_1} := h$ *and* $df_{J_2} := jh$. *Under fixed alternatives, both tests diverge.*

Proof. Consider J_1. The results is a simple consequence of Proposition 12. Under a fixed alternative, $K'vec(\bar{a}'\widehat{\varphi}_{\perp a}) - j \xrightarrow{p} c$, a non-zero vector, so that J_1 diverges.

Consider next J_2. Under the null $K \in \mathrm{col}(\varphi_\perp)$ and hence $K'S_{01} = o_p(1)$, $K'S_{00.1}K \xrightarrow{p} K'\Omega^*K$. Next observe that $C'_{1T} := T^{1/2}K'S_{01} = T^{1/2}K'S_{\epsilon 1}$ is a part of C'_T in Lemma 13, so that $vec(C'_{1T}) \xrightarrow{d} N(0, \Sigma_{11} \otimes K'\Omega^*K)$, and hence

$$N_T := vec((K'\Omega^*K)^{-1/2}C'_{1T}\Sigma_{11}^{-1/2}) \xrightarrow{d} N(0, I_{jh}).$$

Finally note that $J_2 = tr\,(N'_T N_T) + o_p(1) \to \chi^2(jh)$. Under the alternative, $K \notin \mathrm{col}(\varphi_\perp)$, one finds $K'S_{01}S_{11}^{-1} = K'\varphi\tau' + o_p(1)$, so that $J_2 \to \infty$. ■

The second part of this appendix shows that $Q_i - \widehat{Q}_i = o_p(1)$, $J_i - \widehat{J}_i = o_p(1)$, and that $T^{1/2}(\widehat{\widehat{\eta}} - \widehat{\eta}) = T^{1/2}(\widehat{\widehat{\eta}} - \eta) - T^{1/2}(\widehat{\eta} - \eta) = o_p(1)$, where η represents the parameter vector in the reduced rank regression model. We summarize this by saying that test

Q_i (respectively J_i) and $\widehat{Q}_i$ (respectively $\widehat{J}_i$) are equivalent, or that the estimators $\widehat{\widehat{\eta}}$ and $\widehat{\eta}$ are equivalent.

Take for instance Q_i; the above equivalence is proved by showing that $Q_i = Q_i^{\infty} + o_p(1)$ and that $\widehat{Q}_i = Q_i^{\infty} + o_p(1)$ for the same asymptotic term Q_i^{∞}. This proves that $Q_i - \widehat{Q}_i = Q_i^{\infty} + o_p(1) - Q_i^{\infty} + o_p(1) = o_p(1)$. The same format can be used for J_i and for $T^{1/2}(\widehat{\eta} - \eta)$. This proves that the same limit distributions apply.

We first state sufficient conditions on the sample moment matrices in order for the results in the first part to be still valid.

Lemma 18 *If*

$$\widehat{S}_{01} = S_{01} + o_p(T^{-1/2}) \qquad \widehat{S}_{ii} = S_{ii} + o_p(1) \qquad i = 0, 1 \tag{42}$$

then the tests Q_i (respectively J_i) and $\widehat{Q}_i$ (respectively $\widehat{J}_i$) are equivalent, and the estimators $\widehat{\widehat{\eta}}$ and $\widehat{\eta}$ are equivalent. The following are sufficient conditions to verify (42)*:*

$$\widehat{M}_{ij} = M_{ij} + o_p(T^{-1/2}),\ i = 0, 1, 2 \text{ and } j = 1, 2. \tag{43}$$

Proof. It is simple to see that under (42) Lemma 13 applies substituting $\widehat{S}_{00}$, $\widehat{S}_{01}$, $\widehat{S}_{11}$ in place of S_{00}, S_{01}, S_{11}, and the proofs of propositions in the first part of the appendix hold. This proves the first claim.

Let (43) hold; then for $i, j = 0, 1$

$$\widehat{S}_{ij} = \widehat{M}_{ij} - \widehat{M}_{i2}\widehat{M}_{22}^{-1}\widehat{M}_{2j} = S_{ij} + (\widehat{M}_{ij} - M_{ij}) - (\widehat{M}_{i2}\widehat{M}_{22}^{-1}\widehat{M}_{2j} - M_{i2}M_{22}^{-1}M_{2j}).$$

Let $a := M_{i2}$, $b := M_{22}^{-1}$, $c := M_{21}$. If is easy to see that

$$\begin{aligned} \widehat{a}\widehat{b}\widehat{c} - abc &= (\widehat{a} - a)bc + a(\widehat{b} - b)c + ab(\widehat{c} - c) + (\widehat{a} - a)b(\widehat{c} - c) + \\ &\quad + (\widehat{a} - a)(\widehat{b} - b)c + a(\widehat{b} - b)(\widehat{c} - c) + \\ &\quad + (\widehat{a} - a)(\widehat{b} - b)(\widehat{c} - c) \end{aligned} \tag{44}$$

such that when $(\widehat{a} - a)$, $(\widehat{b} - b)$, $(\widehat{c} - c)$ are $o_p(T^{-1/2})$, so is $\widehat{a}\widehat{b}\widehat{c} - abc = o_p(T^{-1/2})$. Finally $\widehat{M}_{ij} - M_{ij} = o_p(T^{-1/2})$. Thus the conditions (42) are verified. ■

We observe that $\widehat{S}_{01} - S_{01}$ needs to be of a smaller order than $T^{-1/2}$. Any $T^{1/2}$ estimator is hence not sufficient here. In the case of cointegration coefficients, superconsistency implies that $\widehat{M}_{ij} - M_{ij} = O_p(T^{-1})$, $i, j = 0, 1, 2$, so that Lemma 18 applies. In the following we assume that the cointegration coefficients have been estimated either by ML or by 2SI2.

Let $\mu_1 = \alpha\beta_0'$ and $\beta^* := (\beta' : \beta_0)'$, $X_{t-1}^* = (X_{t-1}' : t-1)'$. Let also $N_t := (X_{t-1}^{*\prime} : \Delta X_{t-1}')'$ and $\psi := (\beta^{*\prime} : \delta\beta_2')' = (\beta' : \beta_0' : \delta\beta_2')'$. The polynomial equilibrium correction Y_{0t} can be written as $\psi' N_t := \beta^{*\prime} X_{t-1}^* + \delta\beta_2'\Delta X_{t-1}$. The remaining equilibrium correction terms are $\kappa'\Delta X_t$ where $\kappa := (\beta : \beta_1)$; let also $\kappa_0 := (\beta_0 : \eta_0)$

Proposition 19 *Under the $I(2)$ assumptions $\widehat{M}_{ij} - M_{ij} = O_p(T^{-1})$, $i, j = 0, 1, 2$.*

Proof. We first want to show that $F_T(\widehat{\psi} - \psi) = O_p(T^{-1})$ where

$$F_T := \begin{pmatrix} \beta & \beta_1 & T\beta_2 & & & & \\ \beta_0 & \eta_0 & & T^{1/2} & & & \\ & & & & T\beta & \beta_1 & \beta_2 \end{pmatrix}' \text{ and}$$

$$F_T^{-1} = \begin{pmatrix} \bar{\beta} & \bar{\beta}_1 & T^{-1}\bar{\beta}_2 & -T^{-1/2}\bar{\kappa}\kappa_0' & & & \\ & & & T^{-1/2} & & & \\ & & & & T^{-1}\bar{\beta} & \bar{\beta}_1 & \bar{\beta}_2 \end{pmatrix}.$$

and zero entries are omitted for readability. The part regarding the first square submatrix on the main diagonal containing the first 2 block of rows and 4 blocks of columns in F_T is proved in Rahbek et al., eq. (B.8).

The second part follows from the expansion

$$\widehat{\delta\beta}_2' - \delta\beta_2' = (\widehat{\delta} - \delta)\beta_2' + \delta(\widehat{\beta}_2 - \beta_2)' + (\widehat{\delta} - \delta)(\widehat{\beta}_2 - \beta_2)',$$

and the fact that $(\widehat{\delta} - \delta) = O_p(T^{-1})$, $(\widehat{\beta}_2 - \beta_2)'\beta = O_p(T^{-2})$, $(\widehat{\beta}_2 - \beta_2)'(\beta_1 : \beta_2) = O_p(T^{-1})$ see Paruolo (2000) eq. (4.7), (4.5) and (4.6). Observe also that $F_T^{-1}K_t$ is normalized as an I(1) process.

Similarly consider $G_T(\widehat{\kappa} - \kappa)$ where $G_T := (\beta : \beta_1 : \beta_2)$, which is $O_p(T^{-1})$ by Theorem 4.1 in Paruolo (2000), and note that $G_T^{-1}\Delta X_t = \bar{G}_T'\Delta X_t$ is normalized as an I(1) process. Let U_t indicate any stationary process that appears in Z_{it}, and let subscript ΔX indicate ΔX_{t-1}.

We apply the same format of proof of Lemma 18, and observe that the non-zero entries in $\widehat{M}_{ij} - M_{ij}$ are given by the following type of terms:

$$\begin{aligned}
M_{UN}(\widehat{\psi} - \psi) &= M_{UN}F_T^{-1}F_T(\widehat{\psi} - \psi) = O_p(1)O_p(T^{-1}) = O_p(T^{-1}) \\
M_{U\Delta X}(\widehat{\kappa} - \kappa) &= M_{U\Delta X}G_T^{-1}G_T(\widehat{\kappa} - \kappa) = O_p(1)O_p(T^{-1}) = O_p(T^{-1}) \\
(\widehat{\kappa} - \kappa)'M_{\Delta X,N}(\widehat{\psi} - \psi) &= (\widehat{\kappa} - \kappa)'G_T'G_T^{-1\prime}M_{\Delta X,N}F_T^{-1}F_T(\widehat{\psi} - \psi) = \\
&= O_p(T^{-1})O_p(T)O_p(T^{-1}) = O_p(T^{-1}) \\
(\widehat{\psi} - \psi)'M_{N,N}(\widehat{\psi} - \psi) &= (\widehat{\psi} - \psi)'F_T'F_T^{-1\prime}M_{NN}F_T^{-1}F_T(\widehat{\psi} - \psi) = \\
&= O_p(T^{-1})O_p(T)O_p(T^{-1}) = O_p(T^{-1}) \\
(\widehat{\kappa} - \kappa)'M_{\Delta X,\Delta X}(\widehat{\kappa} - \kappa)' &= (\widehat{\kappa} - \kappa)'G_T'G_T^{-1\prime}M_{\Delta X,\Delta X}G_T^{-1}G_T(\widehat{\kappa} - \kappa) = \\
&= O_p(T^{-1})O_p(T)O_p(T^{-1}) = O_p(T^{-1})
\end{aligned}$$

Thus $\widehat{M}_{ij} - M_{ij} = O_p(T^{-1})$, $i, j = 0, 1, 2$; this completes the proof. ■