Common trends and cycles in I(2) VAR systems

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

1 May 2004

Abstract

This paper analyzes common cycles in I(2) vector autoregressive (VAR) systems. We consider 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, 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 is presented; it is found that the deviation from one equilibrium relation is an innovation process, whereas no common cycles can be obtained for the acceleration rates.

Keywords: Common features, Cointegration, Common cycles, I(2), Reduced rank regression.

JEL code: C32, C51, C52.
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), and of common features, see Engle and Koizicki (1993) and Vahid and Engle (1993). As for the case of cointegration, common cycles (CC) – defined as non-innovation common features – imply restrictions on the dynamic representation of the system.

The interplay between common trends and common cycles in processes integrated of order 1, I(1), has been considered in Vahid and Engle (1993). I(1) systems have also been the focus of much of the ensuing literature; see Kugler and Neusser (1993), Vahid and Engle (1997), Cubadda and Hecq (2001) inter alia. No contributions in the literature has investigated 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 (1992, 1995a, 1997), Stock and Watson (1993), and Boswijk (2000) inter alia; see Haldrup (1998) for a survey and references prior to 1998.

In this paper it is shown how the I(2) 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 the basis for the analysis of common cycles in deviations from equilibrium. The application of CC both to second differences and to equilibrium relations sheds light on different features of a system. Extensions of the notion of CC to unpredictable linear combinations of several lags of the process are also considered.

As for I(1) systems, the notion of CC is directly related to rank deficiency of some autoregressive coefficient matrices; this provides a unified framework for inference. When the cointegration parameters are known, the Gaussian likelihood analysis consists of reduced rank regression (RRR), see Anderson (1951). The same locally asymptotically normal (LAN) results apply once the cointegration parameters have been substituted with their maximum likelihood (ML) estimates or the two stage I(2) estimates of Johansen (1995a), thanks to the superconsistency of estimates of cointegration parameters. This property does not extend to other permanent-transitory transformations involving estimated parameters that converge at the standard $T^{1/2}$ rate, as in the I(2) analog of the transformation of Gonzalo and Granger (1995).

The RRR framework provides a common format for testing, estimation and specification search on the cofeature vectors. These techniques are illustrated on the Australian prices data-set analyzed in Banerjee et al. (2001); it is found that the newly introduced equilibrium dynamics form supports the presence of a single cofeature vector, while none can be obtained for the equilibrium correction form.

The rest of the paper is organized as follows: Section 2 presents notation and definitions. Section 3 reports various representations of I(2) systems. CC are treated in Section 4. Section 5 discusses unpredictable combinations involving variables at different lags. Section 6 addresses inference through RRR techniques. Section 7 contains the application to Australian prices. Section 8 concludes. Proofs are reported in the Appendix.

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
they are linear processes. The first coefficient matrix powers of ∆ indicate summation.

Stationary processes derived from of the stationary type; if all roots are outside the unit circle then be a martingale difference sequence w.r.t. \( X \) and the presence of cycles. Innovation processes are a special case of I(0) processes

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.

When a linear process has sum of coefficients \( \sum_{i=1}^{\infty} C_i \) different from the 0 matrix, then the linear process is called integrated of order \( q \), I(0); see Johansen (1996). In the following an I(0) process \( W_t \) is said to have I(0) rank \( q \) if rank(\( C_W(1) \)) = \( q \). 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. A process \( X_t \) is said to be integrated of order \( d \), I(d), if \( \Delta^dX_t - E(\Delta^dX_t) \) is I(0), \( d = \pm 1, \pm 2, \ldots \).

I(0) processes \( W_t - E(W_t) = C_W(L)\epsilon_t \) are in general autocorrelated, with \( j \)-th autocovariance \( \gamma_j := \sum_{i=0}^{\infty} C_{W,i}\Omega C_{W,i+j} \), a feature associated with predictability and the presence of cycles. Innovation processes are a special case of I(0) processes without this feature; they correspond to \( C_W(L) = C_W(0) \), with 0 autocovariances because \( C_{W,i} = 0 \) for \( i = 1, 2, \ldots \) in \( \gamma_j \). With a slight abuse of language, we refer to any linear process that is not an innovation process as a cycle.

Consider an I(0) process \( W_t \) with rank \( q > 0 \); if there exist some non-zero vector \( b_i \) such that \( b_i'(W_t - E(W_t)) \) is an innovation process, then the system is said to present non-innovation common features, or common cycles, CC, and \( b_i \) is called a cofeature vector. Let \( \ell \) be the maximum number of linearly independent cofeature vectors \( b_1, \ldots, b_{\ell} \).

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.

We consider VAR(\( k \)), \( k \geq 2 \), systems of the type \( X_t = \sum_{i=1}^{k} A_iX_{t-i} + \varepsilon_t \), where \( X_t \) and \( \varepsilon_t = \mu_1t + \mu_0 + \mu_d\delta_t + \varepsilon_t \) are \( p \times 1 \). \( t, 1, \delta_t \) are the deterministic components, \( d_t := (d_{1,t} : \ldots : d_{r,t})' \) is a vector of demeaned seasonal dummies, i.e. of the form \( d_{i,t} = 1(t \text{mod}r = i) - 1/r, 1(\cdot) \) is the indicator function and \( r \) is the number of seasons. \( L \) and \( \Delta := I - L \) are the lag and difference operators, where negative powers of \( \Delta \) indicate summation. \( \varepsilon_t \) is assumed to be an innovation process w.r.t. \( F_{t-1} \), the sigma-field generated by \( X_{t-1}, i \geq 1 \); more specifically \( \varepsilon_t \) is assumed to be a martingale difference sequence w.r.t. \( F_t \), \( E(\varepsilon_t|F_{t-1}) = 0 \), with second moments \( E(\varepsilon_t^2|F_{t-1}) = \Omega \), where \( 0 < x'\Omega x < \infty \) for all \( x \in \mathbb{R}^p \). All innovation processes in this paper are understood to be linear combination of \( \varepsilon_t \).

We assume, ‘Assumption 1’, that, apart from roots at \( z = 1 \), all other characteristic roots of the AR polynomial \( A(z) := I - \sum_{i=1}^{k} A_i z^i \) are outside the unit circle, i.e. of the stationary type; if all roots are outside the unit circle then \( A(z) \) is called stable. Stationary processes derived from \( X_t - E(X_t) \) have a moving average representation \( C_W(L)\epsilon_t \), where \( C_W(z) := \sum_{i=0}^{\infty} C_{W,i}z^i \) is summable for \( |z| < 1 + \kappa \) and \( \kappa > 0 \), 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 I(0) rank \( q \) if rank(\( C_W(1) \)) = \( q \). 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. A process \( X_t \) is said to be integrated of order \( d \), I(d), if \( \Delta^dX_t - E(\Delta^dX_t) \) is I(0), \( d = \pm 1, \pm 2, \ldots \).

I(0) processes \( W_t - E(W_t) = C_W(L)\epsilon_t \) are in general autocorrelated, with \( j \)-th autocovariance \( \gamma_j := \sum_{i=0}^{\infty} C_{W,i}\Omega C_{W,i+j} \), a feature associated with predictability and the presence of cycles. Innovation processes are a special case of I(0) processes without this feature; they correspond to \( C_W(L) = C_W(0) \), with 0 autocovariances because \( C_{W,i} = 0 \) for \( i = 1, 2, \ldots \) in \( \gamma_j \). With a slight abuse of language, we refer to any linear process that is not an innovation process as a cycle.

Consider an I(0) process \( W_t \) with rank \( q > 0 \); if there exist some non-zero vector \( b_i \) such that \( b_i'(W_t - E(W_t)) \) is an innovation process, then the system is said to present non-innovation common features, or common cycles, CC, and \( b_i \) is called a cofeature vector. Let \( \ell \) be the maximum number of linearly independent cofeature vectors \( b_1, \ldots, b_{\ell} \).

\(^1\)In the statistical analysis \( \epsilon_t \) is assumed i.i.d. \( N(0,\Omega) \). The assumption \( k \geq 2 \) is common to the literature on I(2) systems. Other deterministic terms could also be incorporated.
..., \( b; \) the matrix \( \mathbf{b} := (b_1 : ... : b_\ell) \) is called the CC cofeature matrix, and the systems is said to have cofeature rank \( \ell \). Equivalently, \( W_t \) is said to present \( q - \ell \) common I(0) cycles.

Implicit in this notation is the fact that the maximum number of I(0) cycles is given by the I(0) rank. This result is parallel to Theorem 1 in Vahid an Engle (1993) for I(1) systems, although it applies more generally also to I(2) systems.

**Proposition 1 (upper bound on cofeature rank)**

A \( p \times 1 \) I(0) process \( W_t \) with I(0) rank \( q \leq p \) presents at most \( q \) linearly-independent innovation processes; hence the cofeature rank \( \ell \) is bounded by \( q \), \( \ell \leq q \).

When \( q < p \), the remaining \( p - q \) components of an I(0) process with rank \( q \) are integrated of negative order, which are cyclic by definition. Hence nothing can be said about the commonality in the remaining \( p - q \) directions. This point is further discussed in Section 5. Before closing this section, we introduce a motivating example.

**Example 2** 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

\[
\begin{align*}
X_{1t} &= \Delta X_{2t} + \eta_{1t} \\
\Delta^2 X_{2t} &= c_t
\end{align*}
\]

where \( c_t = \varphi c_{t-1} + \eta_{2t}, \quad |\varphi| < 1 \).

Here \( c_t \) represents a cycle for \( \varphi \neq 0 \). Observe that the system is I(2) because \( X_{2t} \) needs second differences to become I(0). From the first equation one sees that \( \Delta X_{1t} = \Delta^2 X_{2t} + \Delta \eta_{1t} = c_t + \Delta \eta_{1t} \) is affected by the 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.

### 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. The connection between the first two representations is treated in Johansen’s representation theorem, see Johansen (1992) or (1996, Theorem 4.6). The last representation can be found in the proof of Johansen’s representation theorem, and it is used here in order to discuss common features in deviations from equilibrium.

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

\[
\Delta^2 X_t = \Pi X_{t-1} + \Gamma \Delta X_{t-1} + \gamma V_t + \tilde{\epsilon}_t
\]

for \( k \geq 2 \), where notation is collected in Table 1. We next list the conditions for (1) to be I(2).

**I(2) conditions**

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

(b). \( P_{\alpha_1} \Gamma P_{\beta_1} = \alpha_1 \beta_1' \) where \( \alpha_1 \) and \( \beta_1 \) are \( p \times p_1 \) matrices of full rank \( p_1 < p - p_0 \), or, equivalently, \( \alpha_1' \Gamma \beta_1 = \xi \eta' \) where \( \xi = \alpha_1' \alpha_1 \) and \( \eta = \beta_1' \beta_1 \) are \( p - p_0 \times p_1 \) matrices of full rank \( p_1 < p - p_0 \);
(c). $\alpha_2^t \theta_2$ has full rank $p - p_0 - p_1$, where $\alpha_2 := (\alpha : \alpha_1) \perp$, $\beta_2 := (\beta : \beta_1) \perp$.

(d). $\mu_1 = \alpha \beta_0^t$, with $\beta_0^t$ a $p_0 \times 1$ vector;

(e). $\alpha^t \mu_0 = \xi \eta_0^t + \alpha^t \Gamma \beta \beta_0^t$, with $\eta_0^t$ a $p_1 \times 1$ vector.

Under Assumption 1, Johansen’s representation theorem, see Rahbek et al. (1999), establishes that necessary and sufficient conditions for $\Delta^2 X_t$, $Y_{0,t} := \beta' X_t + \delta \Delta X_t + \beta_0^t t$, $Y_{1,t} := \beta_1^t \Delta X_t$ to be stationary, apart from deterministic components and initial values, are the conditions (a) to (e). Under the I(2) conditions the following common trends representation holds

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

where the I(2) component is $C_2 \sum_{s=1}^t \sum_{i=1}^s \epsilon_i$, with $C_2 = \beta_2 (\alpha_2' \theta_2) \perp \alpha_2'$ of rank $p_2$; hence there are $p_2$ common I(2) trends. The I(1) component is $C_1 \sum_{i=1}^t \epsilon_i$; for remaining definitions see Rahbek et al. (1999).

The I(2) common trends representation implies that $\Delta^2 X_t$ is an I(0) process with rank $p_2$. Taking in fact second differences in (2) one find

$$\Delta^2 X_t := C^* (L) \epsilon_t + m^*_t = C_2 \epsilon_t + C_1 \Delta \epsilon_t + C_0(L) \Delta^2 \epsilon_t + m^*_t,$$

where $m^*_t := m(L) \Delta^2 d_t$. This equation shows that $C^*(1) = C_2$, which is of rank $p_2$, and hence $\Delta^2 X_t$ is an I(0) process with rank $p_2$.

If the I(2) conditions hold, then system (1) can be rewritten in many equilibrium correction forms. The following one is taken from Paruolo and Rahbek (1999):

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

where $\mu_1 = \alpha \beta_0^t$, see also Table 1. The terms in square brackets in (4) are I(0) by Johansen’s representation theorem; 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 $Y_{0,t} := \beta' X_t + \delta \Delta X_t + \beta_0^t t$, $Y_{1,t} := \beta_1^t \Delta X_t$, $Y_{2,t} := \beta_2^t \Delta^2 X_t$ as the stationary variables of interest, where $Y_t := (Y_{0,t} : Y_{1,t} : Y_{2,t})'$ is $p \times 1$. Other choices are possible, see Corollary 4.

**Theorem 3 (equilibrium dynamics representation)** Let the I(2) conditions hold and let $Y_t := (Y_{0,t} : Y_{1,t} : Y_{2,t})'$; then $Y_t$ follows a VAR($k$),

$$A^o(L) Y_t = \mu^1 D_t + \epsilon^o_t$$

where the AR polynomial $A^o(L) := I - \sum_{i=1}^k A^o_i L^i$ is stable, and can be inverted to give $Y_t = C^o(L)[\mu^1 D_t + \epsilon^o_t]$, with $C^o(L) \epsilon^o_t = C_Y(L) \epsilon_t$ is a I(0) process with rank $p$, where $C_Y 0 = D$, a full rank matrix.

Let the AR matrices $A^o_i$ be partitioned column-wise conformably with $Y_t$, i.e. let $A^o_{i,j}$ be the $p \times p_j$ block that multiplies $Y_{j,t-1}$. The last AR matrices $A^o_k$ in (5) are constrained as follows

$$A^o_{k,0} \delta = -A^o_{k-1,2}, \quad (A^o_{k,1} : A^o_{k,2}) = 0. \quad (6)$$
The constraints (6) can be incorporated in (5) by substituting $Y_{t-k}$, $Y_{2,t-k+1}$ with $\Delta^2 X_{t-k+1}$ in the r.h.s. of (5), or, alternatively using the same r.h.s. variables as (4), obtaining respectively

\[ Y_t = \Psi^t U_t^i + \mu^t D_t + \epsilon_t^i, \quad (7) \]

\[ = \Psi^o U_t^o + \mu^o D_t + \epsilon_t^o. \quad (8) \]

Moreover $\Psi^o = \Psi^t A$, for $A$ square and nonsingular, and hence $\text{rank}(\Psi^o) = \text{rank}(\Psi^t)$ and $\text{col}(\Psi^o) = \text{col}(\Psi^t)$. (7) is called the ‘equilibrium dynamics form’ and (8) is called the ‘mixed form’.  

The equations (5) or (7) describe the dynamics of the equilibrium relations. Note also that rank of the I(0) process $Y_t$ is equal to the dimension $p$ of the process $X_t$ (and $Y_t$). Thus the transformation from $\Delta^2 X_t$ to $Y_t$ allows to express all cycles as I(0) cycles. Observe also that $\Psi$ in (4) and $\Psi^o$ (respectively $\Psi^t$) in (7) (resp. (8)) are not similar and possibly have different ranks.

In the following corollary we show that the properties of the equilibrium dynamics for $Y_t$ discussed in Theorem 3 are carried over to a different choice of the $Y_{1t}$ component. Let $\tau_1$ be any $p \times p_1$ matrix such that $\text{col}(\beta : \beta_1) = \text{col}(\beta : \tau_1)$ and define $G_t := (Y_{0t} : \Delta X_t^i : Y_{0t}^o)^o$.

**Corollary 4** Under the same assumptions of Theorem 3, $G_t$ follows a VAR($k$) of the type $B^o(L)Y_t = K\mu^t D_t + K\epsilon_t^o$ where $B^o(L)$ is stable, $K$ is a square matrix of full rank defined in the Appendix and the AR coefficients of $B^o(L)$ satisfy the following constraints, equivalent to (6):

\[ B^o_{k,0} \delta + B^o_{k-1,2} - 2B^o_{k-1,1} \tau_1^t \beta \delta = 0, \quad (B^o_{k,1} : B^o_{k,2}) = 0. \]

The choice $G_t$ is hence equivalent to $Y_t$. For ease of exposition, in the following we will only discuss the choice $Y_t$, simply noting that $G_t$ enjoys the same properties.

## 4 Common cycles

This section discusses the application of common features to I(2) systems, 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 (5). In the following we will indicate with $W_t$ either $\Delta^2 X_t$ or $Y_t$.

A matrix $b$, of dimension $p \times \ell$ and rank $\ell$, defines unpredictable linear combinations for $W_t$ if $b'(W_t - E(W_t)) = g' \epsilon_t$ 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 CC cofeature matrix for $W_t$ with cofeature rank $\ell$ when $\ell$ is chosen to be maximal.

The above definition of CC allows to decompose the time series $W_t$ into cyclical and idiosyncratic components, analogously to the permanent-transitory decompositions discussed in Gonzalo and Granger (1995). Let $b$ be the cofeature matrix of $W_t$. Using the orthogonal projection identity $I_p = P_b + P_{b\perp}$, one can define a first decomposition

\[ W_t = \bar{b} \eta_t + \bar{b}_{\perp} \epsilon_t, \quad (9) \]

\[ \text{Note that when } k = 2, \text{ the mixed form } (8) \text{ and the equilibrium dynamics representation } (7) \text{ coincide.} \]
where $c_i := b_i' W_i$ is the common cyclical component of dimension $p - \ell$ and I(0) rank $q - \ell$, while $\eta_t := b' W_t = b' E(W_t) + g' \epsilon_t$ is the idiosyncratic noise, i.e. a white noise component, of dimension and rank equal to $\ell$.

A second decomposition can be obtained using non-orthogonal projections; let $a$ be of the same dimension of $b$ with $b' a$ of full rank. One has $I_n = a(b'a)^{-1} b' + b_1 (a_1' b_1)^{-1} a_1' =: a b' + b_1 a_1 a_1'$, and hence

$$W_t = a t_0 + b_1 a_1 c_i t_0$$

where $c_i := a_i' W_i$, $a_b := a(b'a)^{-1}$, $b_1 a_1 := b_1 (a_1' b_1)^{-1}$. Note that both decompositions (9) and (10) contain the same idiosyncratic component $\eta_t$ with different loading matrices, which correspond to a different definition of the cyclical component, $c_i$ and $c_i^2$. The cyclical parts are autocorrelated and present no CC.

The cyclical ($c_i$ or $c_i^2$) and idiosyncratic components ($\eta_t$) are in general correlated, except for the choice $a_1 := \Omega^{-1} g_1$. This particular decomposition has the property that $Cov(\eta_t, c_i^2 g_1) = g' \Omega^2 g_1 = 0$, i.e. the cyclic and idiosyncratic parts are uncorrelated.

As stated in Proposition 1, the cofeature rank $\ell$ is bounded by the I(0) rank $q$, i.e. $\ell \leq q$. As noted in equation (3), $\Delta^2 X_t$ is an I(0) process of rank $p_2$, while $Y_t$ is an I(0) processes of rank $p$, see Theorem 3. 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 existence of a cofeature matrix $b$ for $\Delta^2 X_t$ or $Y_t$ is associated to a rank reduction of the coefficient matrices in the equilibrium correction or equilibrium dynamics representations respectively, as shown in the following proposition.

**Proposition 5 (cofeatures and reduced rank)** $W_t$ presents common feature with cofeature rank $\ell$ if and only if $\Psi^{(1)}$ is of reduced rank, where $\Psi^{(1)} = \Psi$ in (4) for the choice $W_t = \Delta^2 X_t$, and $\Psi^{(1)} = \Psi^o$ in (7) for the choice $W_t = Y_t$. The reduced rank condition $rank(\Psi^{(1)}) = p - \ell$ can be written $\Psi^{(1)} = \varphi \tau$, with $\varphi$ and $\tau$ 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 two choices for $W_t$ in more detail. Consider first the choice $W_t = \Delta^2 X_t$. The following representation theorem for cofeatures in I(2) systems parallels 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 (2) or (3) one has

$$b' C_{0,i} = 0, \quad i = 0, 1, 2, ...$$

$$b' C_1 = 0.$$  \hspace{1cm} (11) \hspace{1cm} (12)

When (11) holds, the second condition (12) holds if and only if any of the following equations holds

$$b' C_2 \quad b'$$

$$b' = \alpha_2 c_\perp \quad \alpha_2' (I - \theta) b_2 = c d',$$

where $c$ and $d$ are $p_2 \times p_2 - \ell$ matrices of rank $p_2 - \ell$.  \hspace{1cm} (13) \hspace{1cm} (14)
Theorem 6 shows that, when it exists, the cofeature matrix \( b \) for \( \Delta^2 X_t \) must be of the form \( b = \alpha_2 u \) for an appropriate matrix \( u \), \( b'\epsilon_t = u'\alpha_2'\epsilon_t \). Recall also from (2) that \( \alpha_2'\epsilon_t \) are the second increments of the \( p_2 \) common trends. Hence the interpretation of the CC 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.

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 equilibrium are innovation processes, as expected by some 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 \( X_t \), \( \Delta X_t \), \( \Delta^2 X_t \), 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 careful modelling of the cointegration properties of a system and by appropriate specification testing on \( b \), see Section 6 below, where we discuss model specification.

We next apply the previous definitions to the example of Section 2.

**Example 7 (Example 2 continued)** 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

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

Observe that the AR matrix \( A_2^* \) is of deficient rank, so that \( b = (1 : 0)' \) is a cofeature vector. The cofeature relation is \( X_{1t} - \Delta X_{2t} = \eta_{1t} \); hence common features applied to \( Y_t \) correctly signal the presence of a common cycle.

The equilibrium correction representation is

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

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.

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 hence turn out to be important. Ultimately which option is relevant remains an empirical question.

In the following section we briefly consider some more general notions of unpredictable linear combinations.
5 Unpredictable combinations

In this section we briefly discuss unpredictable linear combinations of possibly larger sets of variables, which may include other lags not included in \( X_t \); we call this case UP, Unpredictable Polynomial combinations.\(^3\) This notion offers some relief to the lack of invariance of CC to changes in the timing of the variables. This phenomenon was first observed in Ericsson in his comments to Engle and Kozicki (1993), and it applies as well to the results in previous sections. Unlike CC, however, there is no direct decomposition in idiosyncratic and cyclic components associated with UP, of the type discussed in Section 4.

Unpredictable combinations of \( X_t \) and its lags are motivated e.g. by present value models. Consider in fact a present value model of the form \( X_{1t} = \varphi_0 (1 - \varphi_1) \sum_{i=0}^{\infty} \varphi_i^t E_t (X_{2t+i}) \) for the bivariate time series \( X_t := (X_{1t} : X_{2t})' \), see Campbell and Shiller (1987).\(^4\) Special cases of this model are obtained in the expectation theory of interest rates, where \( X_{1t} \) represents the long term yield and \( X_{2t} \) the one-period interest rate, the present value model for stock prices, where \( X_{1t} \) is the stock price and \( X_{2t} \) the dividend, and (with modifications) to the permanent income theory of consumption.

This model implies that the linear combination \( X_{1t} - \varphi_1^{-1} (X_{1t-1} - \varphi_0 (1 - \varphi_1) X_{2t-1}) \) is an innovation process with respect to past history \( F_t \) of \( X_t \). Note that this innovation process

\[
\Delta X_{1t} - \frac{1 - \varphi_1}{\varphi_1} (1 - \varphi_0) \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix}
\]

involves \( \Delta X_t \) and \( X_{t-1} \) or \( X_t \) and \( X_{t-1} \), i.e. \( X_t \) at different lags. The relation (15) is compatible with \( X_t \) being I(1) or I(2); in the latter case (15) represents a multicointegrating relation. The unpredictable combination (15) is similar to a CC cofeature relation defined in Section 4, although applied to two lags of \( X_t \).

We hence consider a polynomial matrix \( b(L) \) in the lag operator; a set of \( \ell \) linear combinations \( b(L)'X_t \) is defined to be an Unpredictable Polynomial linear combinations (UP) if it is an innovation process with respect to \( F_t \) and \( \ell \) is maximal. Observe that CC discussed in Section 4 corresponds to the special case \( b(L)'X_t = b'W_t \).\(^5\)

The notion of UP combinations provides a partial answer to the lack of invariance of CC with respect to timing of the variables, as illustrated by the following example.

Example 8 Consider the stationary bivariate system \( W_t := (W_{1t} : W_{2t})' \) with innovations \( \eta_t \)

\[
\begin{align*}
W_{1t} & = \varrho_1 W_{2t-j} + \eta_{1t} \\
W_{2t} & = \varrho_2 W_{2t-2} + \eta_{2t}
\end{align*}
\]

where \( |\varrho_2| < 1 \), \( \varrho_1 \neq 0 \), \( \varrho_2 \neq 0 \), and the lag \( j \) in the first equation can be either 1 or 2. When \( j = 2 \), the system \( W_t \) presents a common cycle in the definition of Section 4, because \( \varrho_2 W_{1t} - \varrho_1 W_{2t} = \varrho_2 \eta_{1t} - \varrho_1 \eta_{2t} \) is an innovation process. Changing the lag to \( j = 1 \) yields no common

\(^{3}\)The transformation \( Y_t \) already involves lags of \( X_t \). The unpredictable linear combinations may involve other lags not included in \( Y_t \).

\(^{4}\)This reference was suggested by one of the referees; it assumes \( X_t \) is I(1) and discusses the cointegration implications of (15) in this case.

\(^{5}\)The UP notion contains also the special case of 'polynomial serial correlation common feature' defined as any polynomial linear combination of \( \Delta X_t \) that is an innovation process, see Cubadda and Hecq (2001).
cycles for \( W_t \) in the definition of Section 4. However, if one considers UP in the extended vector \((W'_t : W'_{t-1})'\), one finds that \( W_{1t} - \varphi_1 W_{2t-1} = \eta_{1t} \) is an innovation process. Hence, looking for UP in the extended vector \((W'_t : W'_{t-1})'\) for \( j = 1 \) yields the same number of unpredictable linear combinations as when considering CC for \( j = 2 \), although the weights of the linear combination are different in the two cases. This shows that UP goes some way in preserving the common cycle properties of a system when changing the timing of the variables.

The previous example shows that employing the definition of UP relations, the system in the example contains a common cycle both in the case \( j = 1 \) and \( j = 2 \), although the linear combination involving \( W_{1t} \) and \( W_{2t-h} \) has different weights in the two cases.

Observe that not all UP combination can be interpreted as cofeature combination in the CC sense. Some UP combinations may for instance involve the same variable at different lags; for instance \( W_{1t} - \varphi W_{1t-1} \) could be an innovation. This occurrence would signal that \( W_{1t} \) has autonomous dynamics, i.e. it is not Granger-caused by the other variables in the system. Obviously the interpretation in this case would be very different from the case of CC.

Finding some UP relations in a system does not tell whether these conform to the predictions of an economic model as in (15), or they correspond to asynchronous cycles in different variables as in Example 8, or to autonomous dynamics of the above type, or even some other case. Just as in cointegration analysis, one can investigate these issue e.g. by hypothesis testing on the UP relations; this specification analysis is discussed in Section 6. In the rest of this section we analyze some representation properties of UP combinations.

In the following we let \( W_t \) be an I(0) process with rank \( q \), which is taken to be either \( \Delta X_t \) 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 \( Z_t := (W'_t : R'_t)' \). A matrix \( b \), of dimension \((p+h) \times \ell \) and rank \( \ell \), is defined to be an UP matrix for \( Z_t \) if \( b'(Z_t - E(Z_t)) \) is an innovation process. We say that \( b \) is a UP matrix for \( Z_t \) with UP rank \( \ell \). If \( b := (b'_1 : b'_2)' \) is partitioned conformably with \( Z_t := (W'_t : R'_t)' \), choosing \( b_2 = 0 \) delivers the CC definition given in Section 4.

A consequence of the definition is that in UP combinations 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 \((p+h) \times \ell \) UP matrix for \( Z_t := (W'_t : R'_t)' \), where \( R_t \) depends on lagged \( X_t \)'s and \( b \) is partitioned conformably with \( Z_t \), then \( b_1 \) has full column rank \( \ell \).

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 for UP purposes. This is reported in the following proposition.

**Proposition 10** The UP properties of \( U_{1t} := (\Delta^2 X_t' : R'_t)' \) and \( U_{2t} := (Y'_t : R'_t)' \) are identical for \( \Delta X_{t-1} = \Delta X_{t-1}(\beta : \beta_1) \).
of later statements. We define $\epsilon_t^* := C_{W,0} \delta_t$, where $C_{W,0} = I$ for $W_t = \Delta^2 X_t$ and $C_{W,0} = D$ for $W_t = Y_t$, see Theorem 3. The covariance matrix of $\epsilon_t^*$ is indicated by $\Omega^* := C_{W,0} \Omega C_{W,0}$. Similarly we let $\mu_0^*$ indicate $\mu_0$, $\mu_0^\dagger$, $\mu_0^\ddagger$ or $\mu_0^\star$.

**Proposition 11** Let $Z_{2t} := (R_t^\prime : d_t^\prime)^\prime$, and assume that $Z_{0t} := W_t$, $Z_{1t}$ and $R_t$ be variables generated from a stationary VAR with innovations $\epsilon_t$, where $Z_{0t}$ satisfies

$$Z_{0t} = \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \quad (16)$$

and $Z_{1t}$, $R_t$ depend on lagged $\epsilon_t$s. Partition also $\Phi := (\Phi_1 : \Phi_2)$ conformably with $Z_{2t} := (R_t^\prime : d_t^\prime)^\prime$. Then a necessary and sufficient condition for $b$ to be an UP matrix for $(W_t^\prime : R_t)^\prime = (Z_{0t} : R_t)^\prime$ is that $\varsigma$ is of reduced rank, $\varsigma = \varphi \tau^\prime$, with $\varphi$ and $\tau$ of full column rank. In this case the UP matrix has representation

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

In the following we use the characterization given in Proposition 11, simply stating the reduced rank restrictions implied by different choices of variables in $Z_{0t}$, $Z_{1t}$, $Z_{2t}.$

## 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; hence it is not described here for space constraints. We refer to Johansen (1995a, 1997), Rahbek et al. (1999), Boswijk (2000), Paruolo (2000).

This section concentrates on the analysis of cofeatures after the cointegration analysis has been performed, fixing the cointegration parameters $\beta$, $\beta_1$, $\delta$ to their maximum likelihood estimates or the two stage I(2) estimates of Johansen (1995a) 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. In the rest of this section we simply do not distinguish $\beta$, $\beta_1$, $\delta$ and their estimated values, and we assume that $\epsilon_t$ i.i.d. Gaussian.

In Subsection 6.1 we review the statistical analysis of CC and UP through RRR; in Subsection 6.2 we discuss specification analysis.

### 6.1 RRR analysis

For any given model, see Table 2, the analysis of common features may be organized by first determining the cofeature rank $\ell$. The cofeature matrix $b$ can then be estimated, for the selected cofeature rank $\ell$, 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 $\ell$ linear combinations of the innovations $\epsilon_t$. All these hypotheses are considered in this subsection; we consider either likelihood ratio tests, labelled $Q_\ell$, or Wald-type tests, indicated by $J_\ell$. All test statistics are asymptotically $\chi^2$ distributed with degrees of freedom equal to the number of constraints; moreover all tests are consistent under fixed alternatives. Proofs of these statements
can be found in the working paper version of this paper, posted on the author’s web page.

We first treat the case of unknown cofeature matrix $b$. Several cases of CC and UP have been presented in Sections 4 and 5, see Table 2. As stated in Proposition 5 and Proposition 11, they can all be put in the regression format

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

where the cofeature restriction is

$$H(s) : \varsigma = \varphi \tau'.$$

and $\varphi$, $\tau$, $\Phi$, $\mu_0^*$ and $\Omega^* := E(\epsilon_t^* \epsilon_t'^*)$ are unrestricted. $s$ indicates the number of columns in $\varphi$, $\tau$, where $\varphi$ is $p \times s$ and $\tau$ 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.\(^6\) We indicate as the \(^6\)‘$H(s)$ model’ the regression model (17) under the reduced rank restriction (18).

The $H(s)$ model is analyzed by reduced rank regression, indicated in the following with the shorthand $\text{RRR}(Z_{0t}, Z_{1t}|Z_{2t}, 1)$. The Gaussian likelihood function is proportional to $-T(\ln |\Omega^*| + \text{tr}(\Omega^*^{-1} T^{-1} \sum_{t=1}^T \epsilon_t^* \epsilon_t'^*))$, which is maximized by considering the following eigenvalue problem

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

(19)

with eigenvalues $\lambda_1 > ... > \lambda_p > 0$ and associated eigenvectors $v_t$, where $S_{ij} := M_{ij} - M_{21} M_{22}^{-1} M_{1j}$, $M_{ij} := T^{-1} \sum_{t=1}^T (Z_{it} - m_i) (Z_{jt} - m_j)'$, $m_i := T^{-1} \sum_{t=1}^T Z_{it}$, $i, j = 0, 1, 2$, see e.g. Johansen (1996).

The LR test statistic for hypothesis (18) of $H(s)$ versus $H(p)$ about the rank of $\varsigma$ is given by

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

This test is asymptotically $\chi^2$ with degrees of freedom $df_{Q_1(s)} = (j-s)(p-s)$. Eq. (19) provides also the maximum likelihood estimates for given dimension $s$. In particular $\hat{\tau} = (v_1 : ... : v_s)$ and

$$\hat{\varphi} = S_{01}\hat{\tau}'(\hat{\tau}' S_{11} \hat{\tau})^{-1}, \quad \hat{\varsigma} = \hat{\varphi}' \hat{\tau}' = S_{01}\hat{\tau}'(\hat{\tau}' S_{11} \hat{\tau})^{-1} \hat{\tau}', \quad \hat{\Phi} = (M_{02} - \hat{\varsigma} M_{12}) M_{22}^{-1}, \quad \hat{\Omega}^* = S_{00} - S_{01}(\hat{\tau}' S_{11} \hat{\tau})^{-1} \hat{\tau}' S_{10},$$

(20)

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

In order to identify parameters, it is convenient to normalize $\hat{\tau}$ by the just-identifying restrictions $\hat{\tau}_c := \hat{\tau}(c' \hat{\tau})^{-1}$, where $c$ is a known matrix of the same dimensions of $\tau$, such that $c' \tau$ is a square nonsingular matrix, see Johansen (1996) Section 5.2 or Paruolo (1997). The choice of $\hat{\tau}$ obtained by substituting $\hat{\tau}_c$ in place of $\hat{\tau}$ in (20) is given by $c\hat{\tau} := \hat{\varsigma} c$, which satisfies $c\hat{\tau}' \hat{\tau}_c = \hat{\varsigma}$.

In the following we use the just-identifying normalization $\hat{\tau}_c := \hat{\tau}_c (a_{11}' \hat{\tau}_c)^{-1}$ 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 (19)

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

(21)

\(^6\)Most of the derivations are unaffected by this assumption.
which has the same \( \lambda_i \) eigenvalues of (19) and eigenvectors \( u_1, \ldots, u_p \); one has \( \hat{\varphi}_\perp = (u_{s+1} : \ldots : u_p) \), see Johansen (1996) Theorem 8.5. Here \( \hat{\varphi}_\perp \) is normalized by \( \hat{\varphi}_\perp S_{00} \hat{\varphi}_\perp = I_p \). \( \hat{\varphi}_\perp S_{01} S_{11}^{-1} S_{10} \hat{\varphi}_\perp = \text{diag}(\lambda_{s+1}, \ldots, \lambda_p) =: \Lambda_2 \). The corresponding just-identified estimator is \( \hat{\varphi}_{\perp a} := \hat{\varphi}_\perp (a'_\perp \hat{\varphi}_\perp)^{-1} \), 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.

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

**Proposition 12** Let the \( I(2) \) assumption hold. Under (18) one has

\[
T^{1/2} \left( \text{vec}(\hat{a}'(\hat{\varphi}_{\perp a} - \varphi_{\perp a})) \right) \overset{d}{\to} N \left( 0, \left( \varphi'_{\perp a} \Omega \varphi_{\perp a} \otimes (a' \Sigma_{11} \varsigma a)^{-1} \right) \right). \tag{22}
\]

A consistent estimator of the asymptotic covariance matrix is obtained substituting the unrestricted maximum likelihood estimator and \( \Sigma_{11} \) with \( S_{11} \).

Consider exclusion restrictions for all columns of \( \varphi_\perp \) simultaneously of the type

\[
H_0 : \quad \varphi_\perp = H \phi, \tag{23}
\]

where \( H \) is \( p \times h \), \( h \geq \ell' \). Under the restriction (23), the likelihood function is maximized by solving

\[
|\lambda^* H' S_{00} H - H' S_{01} S_{11}^{-1} S_{10} H | = 0,
\]

with eigenvalues \( \lambda^*_1 > \ldots > \lambda^*_h \) and corresponding eigenvectors \( v^*_1 \), see e.g. Paruolo (1997), Appendix C, or Johansen (1996) Theorems 8.4 and 8.5. The corresponding LR test statistic of (23) 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 \( \hat{\varphi}_\perp = H (v^*_{h-p+s+1} : \ldots : v^*_h) \). This test is asymptotically distributed \( \chi^2 \) with degrees of freedom \( df_{Q_2} := 2ps - s^2 - 2p(h - p) - (p - s - h)(2h - p + s) \).

Hypothesis (23) is a special case of \( K' \text{vec}(\hat{a}' \varphi_{\perp a}) = j \), where \( K \) has \( h \) columns. The Wald test associated with the latter is given by

\[
J_1 := T (K' \text{vec}(\hat{a}' \varphi_{\perp a}) - j)' \left( \left( \varphi_{\perp a} \hat{\Omega}' \varphi_{\perp a} \right)^{-1} \otimes (a' \varsigma S_{11} \varsigma a) \right) (K' \text{vec}(\hat{a}' \varphi_{\perp a}) - j), \tag{24}
\]

which is asymptotically \( \chi^2(h) \).

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.

\[
H_0 : \quad K' \varsigma = 0. \tag{25}
\]

A Wald test of (25) can be based on the unrestricted maximum likelihood estimates \( \hat{\varsigma} := S_{01} S_{11}^{-1}, \hat{\Omega}' := S_{00.1} := S_{00} - S_{01} S_{11}^{-1} S_{10} \) and equals

\[
J_2 := T \text{tr} \left( (K' S_{00.1} K)^{-1} K' \hat{\varsigma} S_{11} \hat{\varsigma} K \right) = T \text{tr} \left( (K' S_{00.1} K)^{-1} K' S_{01} S_{11}^{-1} S_{10} K \right). \tag{26}
\]
which is asymptotically \( \chi^2(hj) \). The corresponding LR test of (25) in \( H(s) \), labelled \( Q_3 \), is found by solving the eigenvalue problem

\[
\lambda^v K'_\perp S_{00,K} K_\perp - K'_\perp S_{01,K} S_{11,K}^{-1} S_{10,K} K_\perp = 0, \tag{27}
\]

with eigenvalues \( \lambda^1 > ... > \lambda^n \) and corresponding eigenvectors \( v_i^v \), where \( S_{ij,K} := S_{ij} - S_{00} K_\perp (K' S_{00} K_\perp)^{-1} K' S_{0ij}, i, j = 0, 1 \), see Johansen (1996) Theorems 8.2 and 8.5. The test of (25) in \( H(s) \) is given by

\[
Q_3 := T \left( \sum_{i=1}^s \ln(1 - \lambda^v_i) - \sum_{i=1}^s \ln(1 - \lambda_i) \right) = T \left( \sum_{i=s+1}^p \ln(1 - \lambda_i) - \sum_{i=s+1}^{h-p+s} \ln(1 - \lambda^v_i) - \ln \left| \frac{K'_\perp S_{00,K} K_\perp}{K'_\perp S_{00} K_\perp} \right| \right), \tag{28}
\]

where \( S_{00,1} := S_{00} - S_{01} S_{11}^{-1} S_{10} \). The restricted estimate under (25) is \( \hat{\varphi}_\perp = (K : K_\perp (v_0^{v+1} : ... : v_{h-p+s}^{v+1})) \), which again can be identified via \( \hat{\varphi}_{\perp,\perp} \). The \( Q_3 \) test is asymptotically \( \chi^2 \), with degrees of freedom \( df_{Q_3} := sh \). The tests \( Q_1(s) \) and \( Q_3 \) can be combined to obtain the LR test of (25) in \( H(p) \), \( Q_4 := Q_1(s) + Q_3 \). Again \( Q_4 \overset{d}{\rightarrow} \chi^2(df_{Q_4}) \), with \( df_{Q_4} = df_{Q_1(s)} + df_{Q_3} \).

Finally, observe that \( \varphi'_\perp Z_{00} = u_\ell \), where \( u_\ell \) contains \( \ell \) linear combinations of \( \epsilon_t \), defines a system of \( \ell \) 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 (21), interchanging \( \beta \) and \( \varphi_\perp \), the subscripts 0 and 1, and choosing the smallest eigenvalues instead of the largest ones.

### 6.2 Specification analysis

A list of different UP and CC cases is given in Table 2, using the format of equation (16). 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 \( \beta, \beta_1, \delta \), 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 (4) 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. (2004) for I(1) systems. The idea is that some elements in \( \Delta^2 X_t \) inherit the cyclic part included in deviations from equilibrium in \( Y_{0,t-1} \) and/or \( (\beta : \beta_1)' \Delta X_{t-1} \).

Several models given in Table 2 are nested. This suggest the possibility to ‘test down’ for cofeatures from the most general to the most specific model. 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 UP 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 CC 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, e.g. by use of Bonferroni-type inequalities. Hence a typically small nominal size must be 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 restrictions are compared with a baseline reference model, which is the unrestricted equilibrium correction formulation (4) or the equilibrium dynamics mixed form (8). Hence also the ‘testing up’ procedure is in line with the general-to-specific framework, see Paruolo (2001) and reference therein. 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 component test, if each test has probability of rejection that converges to 1 under a fixed alternative, which is the case for the tests in Subsection 6.1. For further details on this type of procedure we refer to Paruolo (2001) and reference therein.

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) and Omtzigt and Paruolo (2004), inter alia. We refer to the latter paper for the analysis of the initial specification and the cointegration analysis. The remaining computations of the CC analysis reported here 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 time series are graphed in Fig. 1. The common trend analysis selected \( p_0 = 1, p_1 = 1; \) the ML estimates of the cointegration parameters turned out to be

\[
\hat{\hat{\beta}}_1 X_t = -0.28lpfc_t - 0.72lulc_t + lpm_t,
\]

\[
Y_{0t} = \hat{\beta} X_t + \hat{\delta} \Delta X_t + \hat{\beta}_0 t = lpfc_t - 0.74lulc_t - 0.26lpm_t + 2.68 (\Delta lpfc_t + \Delta lulc_t + \Delta lpm_t) + 0.0013 t. \tag{29}
\]

We next fixed the cointegration parameters \( \hat{\beta}, \hat{\beta}_1, \hat{\delta} \) at these estimated values. The equilibrium corrections shows how the original variables adjust to various disequilibria. The ML estimates of the adjustment coefficients \( \alpha \) and \( \zeta \) are reported in

\[\text{http://qed.econ.queensu.ca/dae.}\]
Table 3, after deleting insignificant coefficients. They show a significant adjustment to the growth rate of the autonomous price component $\beta_1^t \Delta X_{t-1}$, both for $\Delta^2 \text{lulc}_t$ and $\Delta^2 \text{lpfc}_t$. We interpret this finding as evidence that the I(1) autonomous price component $\beta_1^t X_t$ 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 $\Delta^2 \text{lpfc}$, suggesting that $Y_{0t}$ measures (deviations from) an internal price equilibrium. Using this specification we also were not able to reject the null of absence of structural breaks.

We performed the CC analysis in model a) in Table 2, 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 $(Y_{0,t-1} : Y_{1,t-1} : \beta^t_1 \Delta X_{t-1})'$ and $Z_{2t} = (d_{1t}^t : d_{2t}^t)'$, where $d_{it}$ indicate intervention dummies defined in Omtzigt and Paruolo (2004). Table 4 reports the result of the test statistics. It can be seen that the tests reject the hypothesis $H_{s,t}^\prime = 0 \; \forall \; s$, which corresponds to the specification in model 3. Table 5 reports also asymptotic standard errors based on Proposition 12, and the corresponding asymptotically normal $t$-ratios. These estimates suggest the vector $(0 : 1 : 0)'$ 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 $H_{s,t}^\prime = 0 \; \forall \; s$ we employed the LR test $Q_2$ of (23) in $H(s)$, specifying $H = (0 : 1 : 0)'$. 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 (24) 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 $H_{s,t}^\prime = 0 \; \forall \; s$.

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 (25) with $K = e_i$; the associated Wald test statistic $J_2$ in (26) is reported in Table 6. We also calculated the LR test of (25) in $H(s)$, i.e. the test $Q_4$. Both tests confirm that there exists a CC cofeature vector $b = (0 : 1 : 0)'$ in the system for $Y_t$, i.e. that all coefficients of lagged variables 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. 2, along with the estimated innovation process $Y_{1t} - \hat{E}(Y_{1t})$. 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 $p = 3$.

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 inflation.

Summarizing, the application to Australian prices shows the relevance of the

---

8The same test can also be obtained as a special case of $Q_3$. 

equilibrium dynamics form in cofeature analysis. Only trivial cases of cofeatures could be obtained for the equilibrium correction form.

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 case we have discussed how to address inference both for known and unknown cofeature vectors, using RRR. 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 in the data-set of 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


Appendix

Proof. of Proposition 1. Let \( W_t = C_W(L)\epsilon_t \). From the definition of cofeature rank, \( bW_t = b'C_W0e_t \) (i.e. \( b'C_Wi = 0, i \geq 1 \)) and \( V := b'C_W0\Omega C_W'b \) is positive definite. Because \( b'C_Wi = 0, i \geq 1 \), one has \( b'C_W0 = b'C_W(1) \), and \( V = b'C_W(1)\Omega C_W'(1)b \). Because \( C_W0 \) is assumed to be of full rank, \( \ell = \text{rank}(b'C_W0) = \text{rank}(b'C_W(1)) \leq \text{rank}(C_W(1)) \). 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 \( \Omega \) is of full rank by assumption.

Proof. of Theorem 3. Let \( u_t := \mu_d\delta_t + \epsilon_t, Y(L) := I - \sum_{k=1}^N Y_kL^k, D := (\beta + \beta_2\delta') : (\beta_1 : \beta_2)' \). By the proof of Johansen’s representation theorem, see Johansen (1992, 1996), the process \( Y_t \) follows an I(0) VAR(\( k \)) process, with stable AR polynomial \( A^o(L) \) under the I(2) assumptions. These references also describe how to transform \( Y_t \) back to the autoregressive form, and hence to the equilibrium correction form (5).

Tedious calculations show that the AR matrices are given by

\[
A^o_1 = D \left( \beta_1 + \alpha + (I + \tau Y_1)\beta : \tau \right) \epsilon_2 + (I + \tau Y_1)\beta : \tau, \\
A^o_2 = D \left( \beta_1 - (I - \tau Y_2 + 2\tau Y_1)\beta : (\tau Y_2 - \tau Y_1)\beta : \tau, \\
A^o_i = D \left( (\tau Y_i - 2\tau Y_{i-1} + \tau Y_{i-2})\beta : (\tau Y_i - \tau Y_{i-1})\beta : \tau \right) \epsilon_i, \quad i = 3, ..., k - 2
\]

These expressions imply the restrictions (6). In order to impose them, observe that

\[
A^o_{k-1} Y_{0,t-k+1} + A^o_{k-2} Y_{2,t-k+1} + A^o_{k-0} Y_{0,t-k} = D(2Y_{k-2} + Y_{k-3})\beta Y_{0,t-k+1} - DY_{k-2}\delta Y_{2,t-k+1} + DY_{k-2}\delta Y_{0,t-k} = \\
- DY_{k-2}\beta (Y_{0,t-k+1} - Y_{0,t-k} + \delta Y_{2,t-k+1}) + D(-\beta Y_{k-2} + Y_{k-3})\beta Y_{0,t-k+1} = \\
- DY_{k-2}\beta (\Delta \beta'X_{t-k+1} + \beta') + D(-\beta Y_{k-2} + Y_{k-3})\beta Y_{0,t-k+1},
\]

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 \( Y_{0,t-k+1} \).

We finally show how the mixed form can be obtained. Let \( u^*_t := \gamma \nu_t + \mu_d \delta_t + \epsilon_t \). Insert \( I = D^{-1}D \) before \( \Delta^2X_t \) in the l.h.s. of (4); one finds

\[
(\beta \Delta : \beta_1 \Delta : \beta_2 - \beta \delta \Delta) \left( \begin{array}{c}
\beta' \Delta X_t + \delta \beta_2' \Delta^2 X_t \\
\beta_1' \Delta X_t \\
\beta_2' \Delta^2 X_t
\end{array} \right) = (\alpha : \xi_1 : \xi_2) \left( \begin{array}{c}
Y_{0,t-1} \\
\beta' \Delta X_{t-1} \\
\beta_1' \Delta X_{t-1}
\end{array} \right) + \mu_0 + u^*_t.
\]

Adding \( \beta_0 \) to the top block of variables on the l.h.s. and rearranging

\[
(\beta : \beta_1 : \beta_2 - \beta \delta) \left( \begin{array}{c}
Y_{0t} \\
Y_{1t} \\
Y_{2t}
\end{array} \right) = (\alpha + \beta : \xi_1 + \beta : \xi_2 + \beta_1) \left( \begin{array}{c}
Y_{0,t-1} \\
\beta' \Delta X_{t-1} \\
\beta_1' \Delta X_{t-1}
\end{array} \right) + \mu_0 + u^*_t
\]

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

\[
Y_t = D(\alpha + \beta : \xi_1 + \beta : \xi_2 + \beta_1) \left( \begin{array}{c}
Y_{0,t-1} \\
\beta' \Delta X_{t-1} \\
\beta_1' \Delta X_{t-1}
\end{array} \right) + D\mu_0 + Du^*_t,
\]

19
which gives (8).

The proof that \( \Psi^* = \Psi^\dagger A \), for \( A \) square and nonsingular is similar to the proof of Proposition 10 below and it is therefore omitted.

**Proof.** of Corollary 4. Given that \( \text{col}(\beta : \beta_1) = \text{col}(\beta : \tau_1) \), one has \( \tau_1 = \beta f' + \beta_1 h' \), where \( f' = \beta' \tau_1 \), \( h' = \beta_1 \tau_1 \) and \( h \) must be invertible. Let \( a := h^{-1} f \) and define

\[
D(L) := \begin{pmatrix} I_{p_0} & -a \alpha \\ a \Delta & I_{p_1} \end{pmatrix} \quad D^{-1}(L) = \begin{pmatrix} I_{p_0} & a \alpha \\ -a \Delta & I_{p_1} \end{pmatrix}.
\]

Let \( H := \text{diag}(I_{p_0}, h, I_{p_2}) \) and note that \( G_t = HD(L)Y_t \). Let also \( u_t := \mu^2 D_t + \epsilon_t^2 \); inserting \( D^{-1}(L)H^{-1}HD(L) \) between \( A^0(L) \) and \( Y_t \) in \( A^0(L)Y_t = u_t \), one obtains \( B^0(L)G_t = u_t \) where \( B^0(L) := KA^0(L)D^{-1}(L)H^{-1} \), \( u_t := K v_t \), \( K := HD(0) \). The AR polynomial \( B^0(L) := KA^0(L)D^{-1}(L)H^{-1} \) is of the same degree \( k \) as \( A^0(L) \) and has the same roots, because \( D(L) \) is a unimodular matrix. The constraints on the coefficients of \( B^0(L) \) are obtained from (6) substituting \( A^0(L) \) into the definition of \( B^0(L) \).

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

\[
C_2 \epsilon_t + C_1 \Delta \epsilon_t + C_0(L) \Delta^2 \epsilon_t,
\]

\[
= C_2 \epsilon_t + C_1 (\epsilon_t - \epsilon_{t-1}) + \sum_{i=0}^{\infty} C_{0,i} L^i \epsilon_t - 2 \sum_{i=0}^{\infty} C_{0,i} L^{i+1} \epsilon_t + \sum_{i=0}^{\infty} C_{0,i} L^{i+2} \epsilon_t
\]

\[
= (C_2 + C_1 + C_{0,0}) \epsilon_t + (-C_1 + C_{0,1} - 2C_{0,0}) \epsilon_t + \sum_{i=2}^{\infty} (C_{0,i} - 2C_{0,i-1} + C_{0,i-2}) \epsilon_{t-i}
\]

\[
=: \epsilon_t + C_t^* \epsilon_{t-1} + \sum_{i=2}^{\infty} C_t^* \epsilon_{t-i} =: C_t^*(L) \epsilon_t,
\]

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

\[
C_2 + C_1 + C_{0,0} = I.
\]

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 \( \epsilon_t \) in (30) cancel when pre-multiplied by \( b' \), i.e. iff \( b'C_t^* = 0 \), \( i = 1, 2, \ldots \). Let \( a_t := b'C_{0,i} \). The condition \( b'C_t^* = 0 \), for \( i \geq 2 \) is

\[
a_{j} - 2a_{j+1} + a_{j+2} = 0, \quad j = 0, 1, \ldots
\]

This is a difference equation with solution \( a_j = a_0 + j(a_1 - a_0) \). From the summissibility of \( C_0(z) \) for \( |z| < 1 + \varepsilon \) and \( \varepsilon > 0 \), it follows that \( a_0 = a_1 = a_0 = 0 \), i.e. \( b'C_{0,i} = 0 \) for all \( i \geq 1 \), which is condition (11).

The condition \( b'C_t^* = 0 \) gives \( b'(-C_1 + C_{0,1} - 2C_{0,0}) = 0 \), where \( b'C_{0,1} = b'C_{0,0} = 0 \) by (11). Hence one finds \( b'C_1 = 0 \), condition (12). From (31) 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 (11). This proves the equivalence between (12) and (13).

Assume (13) holds, \( b'C_2 = b' \). From the definition of \( C_2 \), see (2), it follows that \( b \in \text{col}(\alpha_2) \), i.e. that \( b = \alpha_2 u \) for some \( u \). Substituting into \( b'C_2 = b' \) one finds \( u'(\alpha'_2 \beta_2 (\alpha'_2 \beta_2)^{-1} - I_{p_0}) \alpha'_2 = 0 \), which holds iff \( u'(\alpha'_2 \beta_2 - \alpha'_2 \beta_2) = 0 \), i.e. if \( c \) belongs
to $A := \text{col}^{+}(\alpha_2'(I - \theta)\beta_2)$. In order for $A$ 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 $\Psi^{(i)} = \varphi^{\tau'}$ then $\varphi^{(i)}_\perp(W_t - E(W_t))$ 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 (4) and (5) one finds that $b'(W_t - E(W_t))$ contains $b'\Psi U_t$ or $b'\Psi U_t^p$ in addition to an innovation process. Hence $b'\Psi = 0$, i.e. $b \in \text{col}^{+}(\Psi)$. In order for $b$ to be different from the zero vector one must have 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 $(b'_1 : b'_2)(W'_t : R'_t)' = b'_1C_{W,0}\epsilon_t$, given that $R_t$ does not depend on $\epsilon_t$. Moreover $V = \text{Var}(b'_1Z_t) = b'_1C_{W,0}\Omega C_{W,0}b_1$ is of full rank $\ell$. This holds only if $b_1$ has full column rank $\ell$. 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

$$
\begin{pmatrix}
Y_{0t} \\
\beta'_1 \Delta X_t \\
\beta'_2 \Delta^2 X_t \\
Y_{0,t-1} \\
\beta'_1 \Delta X_{t-1} \\
\beta'_2 \Delta^2 X_{t-1}
\end{pmatrix} =
\begin{pmatrix}
\beta'_0 \\
\beta'_1 \\
\beta'_2 \\
\beta'_0 \\
\beta'_1 \\
\beta'_2
\end{pmatrix} +
\begin{pmatrix}
I_{p_0} & I_{p_0} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1}
\end{pmatrix}
\begin{pmatrix}
\Delta^2 X_t \\
Y_{0,t-1} \\
\beta'_1 \Delta X_{t-1} \\
\beta'_2 \Delta^2 X_{t-1}
\end{pmatrix},
$$

where we have omitted zeros for readability. Conversely

$$
\begin{pmatrix}
\Delta^2 X_t \\
Y_{0,t-1} \\
\beta'_1 \Delta X_{t-1} \\
\beta'_2 \Delta^2 X_{t-1}
\end{pmatrix} =
\begin{pmatrix}
\bar{\beta}_0 \\
\bar{\beta}_1 \\
\bar{\beta}_2 - \bar{\beta} \delta \\
\bar{\beta} - \bar{\beta} - \bar{\beta}_1
\end{pmatrix} +
\begin{pmatrix}
I_{p_0} & I_{p_0} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1} \\
I_{p_0} & I_{p_1}
\end{pmatrix}
\begin{pmatrix}
Y_{0t} \\
Y_{0,t-1} \\
\beta'_1 \Delta X_t \\
\beta'_2 \Delta^2 X_t \\
\beta'_1 \Delta X_{t-1} \\
\beta'_2 \Delta^2 X_{t-1}
\end{pmatrix}.
$$

This completes the proof. ■

**Proof.** of Proposition 11. Sufficiency is proved by substituting $\varsigma = \varphi^{\tau'}$ in (16) and pre-multiplication by $\varphi^{\tau'}$. In order to prove necessity, assume $b := (b'_1 : b'_2)'$ is the cofeature matrix with $\ell > 0$ columns, and $b'_1Z_{0t} + b'_2Z_{2t} = b'_1u_t$ be the cofeature relations. Pre-multiplication of (16) by $b'_1$ gives $b'_1Z_{0t} = b'_1\varsigma Z_{1t} + b'_1\Phi Z_{2t} + b'_1u_t$, which, substituted back implies

$$-b'_1\varsigma Z_{1t} + (b'_2 - b'_1\Phi)Z_{2t} = 0.$$ 

In order for this to be identically zero, both coefficients of $Z_{1t}$ and $Z_{2t}$ must be zero. This shows that $b_1 \in \text{col}^{+}(\varsigma)$ and that $b'_2 = b'_1\Phi$. Since $\ell > 0$ was assumed, $\varsigma$ must be of deficient rank, $\varsigma = \varphi^{\tau'}$, and $b_1 = \varphi^{\tau'}$. ■
Figures

Captions

Figure 1: Data in levels and first differences.

Figure 2: 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. (25) holds with $K = e_2$. The lower left panel reports $Y_{1t} - \hat{E}(Y_{1t})$, which is an innovation process.
Tables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1t}$</td>
<td>$\beta_1 \Delta X_t$</td>
</tr>
<tr>
<td>$Y_{2t}$</td>
<td>$\beta_2 \Delta^2 X_t$</td>
</tr>
<tr>
<td>$Y_{0,t}$</td>
<td>$\beta' X_t + \delta \beta_2 \Delta X_t + \beta_0 t$</td>
</tr>
<tr>
<td>$Y_t$</td>
<td>$(Y_{0,t}': Y_{1t}' : Y_{2t}')'</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$(\gamma_1 : \ldots : \gamma_{k-2})$</td>
</tr>
<tr>
<td>$V_t$</td>
<td>$(\Delta^2 X_{t-1}' : \ldots : \Delta^2 X_{t-k+2})'$</td>
</tr>
<tr>
<td>$\Upsilon$</td>
<td>$\sum_{i=1}^{k-2} \Upsilon_i$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$(\Gamma - \Pi)\beta' \alpha'(\Gamma - \Pi) + I - \Upsilon$</td>
</tr>
<tr>
<td>$D_t$</td>
<td>$(1 : d_t)'$</td>
</tr>
<tr>
<td>$D$</td>
<td>$(\beta + \beta_2 \delta': \beta_1 : \beta_2)'$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$(\mu_0 : \mu_d)$</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>$(\zeta_1 : \zeta_2)$</td>
</tr>
<tr>
<td>$\mu^1$</td>
<td>$(\mu_1^1 : \mu_2^1)$</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>$(\alpha : \zeta_1 : \zeta_2 : \gamma)$</td>
</tr>
<tr>
<td>$\mu^0$</td>
<td>$D(\mu_0 - \zeta_1 \beta_0)$</td>
</tr>
<tr>
<td>$\mu^o$</td>
<td>$\Delta \mu$</td>
</tr>
<tr>
<td>$\epsilon_t$</td>
<td>$U_t \mu_t$ $(\alpha + \beta : \zeta_1 + \beta : \zeta_2 + \beta_1 : \gamma)$</td>
</tr>
<tr>
<td>$\mu^+$</td>
<td>$(\mu_1^+ : \mu_2^+)$</td>
</tr>
<tr>
<td>$U_t^+$</td>
<td>$(Y_{0,t-1}' : \Delta X_{t-1}'(\beta : \beta_1) : V_{t}'))$</td>
</tr>
</tbody>
</table>

Table 1: Notation.

<table>
<thead>
<tr>
<th>Case</th>
<th>$b_1^{\alpha(\cdot)}$</th>
<th>$b_1^{\zeta(\cdot)}$</th>
<th>$b_1^{\gamma(\cdot)}$</th>
<th>$Z_{1t}$</th>
<th>$Z_{2t, 1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>CC</td>
<td></td>
<td></td>
<td>$Y_{0,t-1}'$</td>
<td>$D_t$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(\beta : \beta_1)' \Delta X_{t-1}$</td>
<td>$V_{t}$</td>
<td>$(Y_{0,t-1}'$</td>
</tr>
<tr>
<td>(b)</td>
<td>UP *</td>
<td>0</td>
<td>0</td>
<td>$(\beta : \beta_1)' \Delta X_{t-1}$</td>
<td>$V_{t}$</td>
</tr>
<tr>
<td>(c)</td>
<td>UP 0 *</td>
<td>0</td>
<td>0</td>
<td>$(Y_{0,t-1}'$</td>
<td>$V_{t}$</td>
</tr>
<tr>
<td>(d)</td>
<td>UP 0 0 *</td>
<td></td>
<td></td>
<td>$(Y_{0,t-1}'$</td>
<td>$V_{t}$</td>
</tr>
<tr>
<td>(e)</td>
<td>UP * *</td>
<td>0</td>
<td></td>
<td>$V_{t}$</td>
<td>$(\beta : \beta_1)' \Delta X_{t-1}$</td>
</tr>
</tbody>
</table>

Table 2: Rank restrictions in the regression format of (17) 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 (4) or $Y_t$ for the equilibrium dynamics mixed form (8). $\alpha^{(\cdot)}$ indicates either $\alpha$ or $\alpha^o$; similarly for $\zeta, \gamma$. * indicates unrestricted entries.

Table 3: Restricted estimates of the adjustment coefficients $\alpha$, $\zeta$ in the equilibrium correction form (4). The zero restrictions, for fixed cointegration coefficients, gave a LR test of 2.5709 with a $\chi^2(3)$ p-value of 0.4626. All remaining coefficients are significant at conventional confidence levels.

<table>
<thead>
<tr>
<th>$\Delta^2$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\zeta}_1$</th>
<th>$\hat{\zeta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta^2$</td>
<td>-0.0596</td>
<td>-0.329</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta^2$</td>
<td>0</td>
<td>1.256</td>
<td>0.179</td>
</tr>
<tr>
<td>$\Delta^2$</td>
<td>0</td>
<td>0.639</td>
<td>-0.866</td>
</tr>
</tbody>
</table>
Table 4: Test statistics $Q_1(s)$ for the equilibrium correction form (4) for $\Delta^2 X_t$ and for the equilibrium dynamics in mixed form (8) for $Y_t$. df indicates the number of degrees of freedom.

<table>
<thead>
<tr>
<th>specification</th>
<th>$s$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4) $Q_1(s)$</td>
<td></td>
<td>250.28</td>
<td>116.33</td>
<td>33.00</td>
</tr>
<tr>
<td>df</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>p-value $\chi^2(df)$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>(8) $Q_1(s)$</td>
<td></td>
<td>562.08</td>
<td>43.00</td>
<td>0.3538</td>
</tr>
<tr>
<td>df</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>p-value $\chi^2(df)$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.5519</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Estimates of the cofeature vector $\varphi_\perp$ for the equilibrium dynamics in mixed form (8) for $Y_t$.

<table>
<thead>
<tr>
<th>equations</th>
<th>$\varphi_\perp$</th>
<th>$Y_{0t}$</th>
<th>$Y_{1t}$</th>
<th>$Y_{2t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{0t}$</td>
<td>$-0.0092$</td>
<td>1</td>
<td>0.1026</td>
<td></td>
</tr>
<tr>
<td>$Y_{1t}$</td>
<td>$0.0593$</td>
<td>.</td>
<td>0.0756</td>
<td></td>
</tr>
<tr>
<td>$Y_{2t}$</td>
<td>$-0.1551$</td>
<td>.</td>
<td>1.3581</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Test statistics $J_2$, $Q_4$ and $\chi^2(3)$ p-values in brackets of hypothesis (25) with $K = e_i$, $i = 1, 2, 3$, corresponding to the zero coefficients in the equations indicated at the top of each column.

<table>
<thead>
<tr>
<th>eq. test statistic</th>
<th>equation</th>
<th>$\Delta^2$ lpc</th>
<th>$\Delta^2$ lulc</th>
<th>$\Delta^2$ lpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4) $J_2$</td>
<td>$\Delta^2$ lpc</td>
<td>$50.162 [0.0000]$</td>
<td>$165.74 [0.0000]$</td>
<td>$111.81 [0.0000]$</td>
</tr>
<tr>
<td>$Q_4$</td>
<td>$\Delta^2$ lulc</td>
<td>$45.949 [0.0000]$</td>
<td>$107.40 [0.0000]$</td>
<td>$83.297 [0.0000]$</td>
</tr>
<tr>
<td>(8) $J_2$</td>
<td>$Y_{0t}$</td>
<td>$31.243 [0.0000]$</td>
<td>$2.0798 [0.5560]$</td>
<td>$111.18 [0.0000]$</td>
</tr>
<tr>
<td>$Q_4$</td>
<td>$Y_{1t}$</td>
<td>$30.990 [0.0000]$</td>
<td>$2.3896 [0.4956]$</td>
<td>$82.979 [0.0000]$</td>
</tr>
</tbody>
</table>