

Paolo Paruolo

Common dynamics in I (1)  
VAR systems

2003/35



UNIVERSITÀ DELL'INSUBRIA  
FACOLTÀ DI ECONOMIA

<http://eco.uninsubria.it>

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright Paolo Paruolo  
Printed in Italy in December 2003  
Università degli Studi dell'Insubria  
Via Ravasi 2, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

# Common dynamics in I(1) VAR systems\*

Paolo Paruolo

Department of Economics, University of Insubria

Via Ravasi 2, 21100 Varese, Italy

December 2003

## Abstract

This paper discusses common cycles in I(1) vector autoregressive (VAR) systems, both for the first differences of the process and for deviations from equilibrium. This extension is based on the equilibrium dynamics representation of I(1) systems, which is presented 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 specification testing. An empirical application on five US monthly macro and financial time series illustrates the techniques presented in the paper. We find one cointegrating relation and one cofeature vector in the equilibrium dynamics formulation, implying four common trends and four common cycles in the system.

**Keywords:** Common features, Cofeatures, Cointegration, Common trends, Common cycles, Common dynamics, Vector autoregressions, I(1), Reduced rank regression.

**JEL code:** C32, C51, C52.

---

\*Invited paper at ‘Common features in Maastricht’, December 14-16, 2003. Partial financial support from Italian MIUR grants ex 60% and Cofin2002 is gratefully acknowledged. I wish to thank Andrea Savoia for his help in recovering the US monthly dataset and the participants of ‘Common features in Maastricht’ for helpful comments.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Notation and definitions</b>	<b>4</b>
<b>3</b>	<b>Equilibrium dynamics form</b>	<b>6</b>
<b>4</b>	<b>Common cycles</b>	<b>8</b>
4.1	Decomposition . . . . .	8
4.2	CC in growth rates . . . . .	9
4.3	CC in cointegrating relations . . . . .	9
<b>5</b>	<b>UP relations</b>	<b>11</b>
<b>6</b>	<b>Estimation and testing</b>	<b>13</b>
6.1	Test on cofeature rank . . . . .	14
6.2	Estimation of the cofeature matrix . . . . .	14
6.3	Specification tests . . . . .	15
<b>7</b>	<b>An application to monthly US data</b>	<b>16</b>
7.1	Initial specification . . . . .	17
7.2	Common trends . . . . .	20
7.3	Common cycles . . . . .	21
<b>8</b>	<b>Conclusions</b>	<b>22</b>
	<b>References</b>	<b>22</b>
	<b>Appendix A: representation</b>	<b>24</b>
	<b>Appendix B: inference</b>	<b>26</b>

# 1 Introduction

Common cycles in the first differences of I(1) systems have been one of the prominent instances of common features, see Vahid and Engle (1993), Kugler and Neusser (1993), Vahid and Engle (1997), Vahid and Issler (2002), Hecq et al. (2000, 2002), Cubadda and Hecq (2001) *inter alia*. This paper extends these analyses to common cycles, CC, in deviations from equilibrium for I(1) VAR systems. By definition, in fact, cointegration relations are I(0), and they represent additional candidates for innovation processes along with the first differences of I(1) systems. The present analysis thus complements the analysis of growth rates in the search for common cycles.

The interpretation of CC cofeature relations is shown here to be different in the case of the first differences and the one of the cointegrating relations. Cofeatures in the first differences of the variables represent increments in the I(1) common trends; cofeature relations in cointegration relations represent unpredictable deviations from equilibrium. Some simple models with no-arbitrage and rational expectation (RE) imply non-predictable deviations from equilibrium, see Example 1 below. For these models tests of common cycles in cointegration relations are tests of the underlying economic model, see Campbell and Shiller (1987). Common cycles of the two types have thus very different economic meaning.

The analysis of CC in the cointegration relations is based on the dynamics of the stationary variables in I(1) systems, which is called here the ‘equilibrium dynamics’ form, ED. This form has been used implicitly in the proof of Granger’s representation theorem, see Johansen (1996, Theorem 4.2), and it is very similar to the transformation used in Campbell and Shiller (1987). The ED form is equivalent to the Equilibrium correction (EC) form. It is shown that tests of CC in growth rates or in cointegrating relations correspond to adjustment matrices of reduced rank in the EC or ED form, respectively.

We also discuss the implications of present value models, which imply the unpredictability of some linear combination involving variables at different lags. This implication does not correspond to a CC cofeature relation, but to a generalization which we call Unpredictable Polynomial linear combination, UP. It is well known that cofeatures relations are not invariant with respect to timing of the variables. In case cycles are asynchronous, it is shown that asynchronous cofeature relations become UP relations.

UP relations may include both the growth rates and the equilibrium relations. We show that when considering a particular set of lagged variables, the UP extension of CC relations in the growth rates and in the equilibrium relations coincide, in the sense that there is a 1-1 correspondence between the two. In a general-to-specific approach one could thus start from testing UP relations in this common specification, and then perform specification tests. If CC involving only the growth rates exist (or only in deviations from equilibrium), they can be hopefully reached in the specification search. Appropriate specification tests are thus also introduced and discussed.

As already noted in the literature, the notion of CC cofeatures is directly related to rank deficiency of some function of the autoregressive coefficient matrices. This holds both for CF and UP relations, and provides a unified framework for inference, although applied to different functions of the autoregressive parameters, i.e. to different representations. When the cointegration parameters are known, this analysis can be based on reduced rank regression (RRR), see Anderson (1951).

Not surprisingly, the same locally asymptotically normal (LAN) results apply once the cointegration parameters have been substituted with their maximum likelihood (ML) estimates or any superconsistent estimate. This follows from the supercon-

sistency of the cointegration parameters. The possibility to fix the cointegration parameters at their estimated values permits to address inference both on the EC and the ED forms in a unified way. Other transformations of the system, like the one in Gonzalo and Granger (1995), involve parameters that are not estimated superconsistently; they do not share this property. The econometrician should thus first test for common trends, and then, for fixed cointegrating relation, should inquire about CC.

We address inference on common features by likelihood-based techniques developed for RRR. 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 for specification testing on the CC (or UP) vectors. This allows to develop a specification search similar to the one for simultaneous systems of equations. A similar approach is presented in Paruolo (2004) for the analysis of common cycles in I(2) systems.

We finally report an application on a US monthly dataset analyzed in Kim (2003), which illustrates the techniques proposed in the paper. The dataset covers the period 1974.1 to 1998.12; it includes a stock price index, the industrial production index, the effective US exchange rate, the long term yield on corporate bonds and CPI inflation. These macroeconomic and financial time series represent the US stock market and its macro fundamentals.

As in Kim's paper, we find one cointegrating relation. Moreover we don't find any common cycles in the EC representation. However, we find one cofeature vector in the ED form. This implies that the given 5 time series share 4 common trends and 4 common cycles.

The rest of the paper is organized as follows: Section 2 reports notation and definitions. Section 3 introduces the ED form. The characterization of CC is treated in Section 4. Section 5 defines the notion of UP relations. Proofs of propositions in these sections are reported in Appendix A. Section 6 addresses 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 the application to US monthly data. 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  $\text{col}(H)$ .  $\text{vec}$  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. Unless when explicitly stated, all processes  $W_t$  are understood to be multivariate, i.e. of dimension  $q \times 1$ ,  $W_t = (W_{1t} : \dots : W_{qt})'$ . Individual time series, or linear combinations thereof, are called components of the process.

## 2 Notation and definitions

In this section we introduce general notation and definitions. We consider a VAR( $k$ ),  $k \geq 1$ , systems of the type  $A(L)X_t = \tilde{\epsilon}_t$ ,  $A(L) := I - \sum_{i=1}^k A_i L^i$ , where  $X_t$  and  $\tilde{\epsilon}_t = \mu_1 t + \mu_0 + \mu_d d_t + \epsilon_t$  are  $p \times 1$ . For simplicity we assume, Assumption 1, that the roots of  $|A(z)| = 0$  at either at  $z = 1$  or outside the unit circle.  $t, 1, d_t$  are the

deterministic components.  $d_t$  contains a vector of  $r - 1$  de-measured seasonal dummies, i.e. of the form  $d_{i,t} = 1(t \bmod r = i) - 1/r$ , where  $1(\cdot)$  is the indicator function and  $r$  is the number of seasons.  $d_t$  may also contain additional single-period dummies, if needed.<sup>1</sup>  $L$  and  $\Delta := 1 - L$  are the lag and difference operators, where negative powers of  $\Delta$  indicate summation.

$\epsilon_t$  is an innovation process with respect to  $F_t$ , the sigma-field generated by  $X_{t-i}$ ,  $i \geq 0$ . We define an innovation process  $W_t$  as a measurable process with respect to  $F_t$ , with  $E_{t-1}(W_t) = 0$ ,  $E_{t-1}(|W_{it}|^{2+\delta}) < \infty$ ,  $i = 1, \dots, m$ ,  $\delta > 0$ , and  $E_{t-1}(W_t W_t')$  positive definite and finite, where  $E_{t-1}(\cdot) := E(\cdot | F_{t-1})$ . All innovations processes in the paper are linear combinations of  $\epsilon_t$ .

Recall the equivalent formulation of the VAR

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \tilde{\epsilon}_t = \Pi X_{t-1} + \gamma V_t + \tilde{\epsilon}_t \quad (1)$$

for  $k \geq 1$ , where the parameter matrices are defined in Table 1. The VAR system  $X_t$  is I(1) and cointegrated under the following conditions, see Johansen (1996), Theorem 4.2:

**I(1) conditions**

- (a). Assumption 1 holds;
- (b).  $\Pi = \alpha \beta'$ , where  $\alpha$  and  $\beta$  are  $p \times p_0$  matrices of full rank  $p_0 < p$ ;
- (c).  $\alpha'_\perp \Gamma \beta_\perp$  has full rank  $p_1 = p - p_0$
- (d).  $\mu_1 = \alpha \beta'_0$ , with  $\beta'_0$  a  $p_0 \times 1$  vector.

Under the I(1) conditions,  $\Delta X_t$  and  $\beta' X_t$  are I(0) and  $\Delta X_t$  has the following moving average representation

$$\Delta X_t = C \epsilon_t + C_0(L) \Delta \epsilon_t + m_t^*, \quad (2)$$

where  $C := \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$  is of rank  $p_1 := p - p_0$ ; here  $m_t^* := m(L) \Delta d_t$ .  $\Delta X_t$  is thus an I(0) process because it has a MA representation  $\Delta X_t - E(\Delta X_t) = C^*(L) \epsilon_t$  with  $C^*(1) = C$  different from the zero matrix. We say that  $\Delta X_t$  is I(0) with I(0) rank  $p_1 = \text{rank}(C^*(1))$ .

I(0) processes  $W_t = C_W(L) \epsilon_t$  are in general autocorrelated with  $j$ -th autocovariance  $\gamma_j := E(W_t W_{t+j}') = \sum_{i=0}^{\infty} C_{W,i} \Omega C_{W,i+j}'$ . A special case is the one of an innovation processes with  $W_t = C_{W,0} \epsilon_t$  with  $C_{W,0}$ , a constant matrix, of full row-rank. Innovation processes presents zero autocovariances because  $C_{W,i} = 0$  for  $i = 1, 2, \dots$  in the formula for  $\gamma_j$ . In the following, with a slight abuse of language, we will refer to any linear process that is not an innovation process as a ‘cycle’.

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 CC, and  $b_i$  is called a CC cofeature vector. When there exist at most  $\ell$  linearly independent cofeature vectors  $b_1, \dots, b_\ell$ , then  $b := (b_1 : \dots : b_\ell)$  is called the cofeature matrix, and the system is said to have cofeature rank  $\ell$ . Equivalently,  $W_t$  is said to present  $q - \ell$  common I(0) cycles.  $B = \text{col}(b)$  is called the cofeature space.

---

<sup>1</sup>Other deterministic terms could also be incorporated. The asymptotic analysis is performed assuming that the single-period dummies take the value 1 at most a finite number of times as  $T$  goes to infinity.

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. More specifically a  $p \times 1$   $I(0)$  process  $W_t$  with rank  $q \leq p$  presents at most  $q$  innovation processes; hence the cofeature rank  $\ell$  is bounded by  $q$ ,  $\ell \leq q$ . This result is given in Theorem 1 in Vahid and Engle (1993) for  $I(1)$  systems, although it applies more generally; in particular it holds also for  $I(2)$  systems, see Paruolo (2004) and Section 4.2 below.

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. 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 simple bivariate example that will be used below to motivate the ED representation.

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

$$\begin{cases} X_{1t} &= a + X_{2t} + \epsilon_{1t}^* \\ \Delta X_{2t} &= c_t \end{cases} \quad \text{where} \quad c_t = \varrho c_{t-1} + \epsilon_{2t}^*, \quad |\varrho| < 1, \varrho \neq 0$$

Here  $c_t$  represents a cycle. Taking differences in the first equation one sees that  $\Delta X_{1t} = \Delta X_{2t} + \Delta \epsilon_{1t}^* = c_t + \Delta \epsilon_{1t}^*$  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.

Example 1 can be phrased as a simple economic model where  $X_{1t}$  and  $X_{2t}$  are yields on different assets, with the same risk, so that an arbitrage condition holds of the type  $\mathcal{E}_{t-1}(X_{1t} - X_{2t}) = a$ , where  $\mathcal{E}_{t-1}$  indicates agents' conditional expectations. Assuming moreover that agents' conditional expectations  $\mathcal{E}_{t-1}$  are rational,  $\mathcal{E}_{t-1} = E_{t-1}$ , one obtains the CC relation  $E_{t-1}(X_{1t} - X_{2t}) = a$ . Hence Example 1 is consistent with an economic model with no arbitrage opportunities and RE.

### 3 Equilibrium dynamics form

In this section we introduce the ED form and discuss its relation to the EC form. As it is well known, see Johansen (1996), if the  $I(1)$  conditions hold, then system (1) can be rewritten in many EC forms, like

$$\Delta X_t = \alpha Y_{0,t-1} + \gamma V_t + \mu D_t + \epsilon_t = \Psi U_t + \mu D_t + \epsilon_t, \quad (3)$$

where the coefficient matrices are defined in Table 1.

Other EC formulations differ by the choice of lag  $j$  at which  $Y_{0,t-j}$  is measured. The EC formulation (3) shows how the stationary cointegration relations affect the growth rate of all variables  $\Delta X_t$  through the adjustment coefficients  $\alpha$ . These equations emphasize the correction of the variables  $\Delta X_t$  towards equilibrium.

We now turn to the ED form. Consider the  $p \times 1$  vector  $Y_t := (Y_{0t}' : Y_{1t}')'$  where  $Y_{0t} := \beta' X_t + \beta_0' t$ ,  $Y_{1t} := \beta_\perp' \Delta X_t$ , as the stationary variables of interest. The ED form is defined in Theorem 2 below, where the AR matrices  $A_i^\circ$  are partitioned column-wise conformably with  $Y_t$ . In other words the product  $A_i^\circ Y_{t-i}$  decomposes into  $A_i^\circ Y_{t-i} = A_{i,0}^\circ Y_{0,t-i} + A_{i,1}^\circ Y_{1,t-i}$  where  $A_{i,j}^\circ$  is  $p \times p_j$ . The definition of all coefficient matrices is reported in Table 1 except for  $A_i^\circ$ , defined in the proof reported in Appendix A.

symbol	definition	symbol	definition
$\gamma :=$	$(\Gamma_1 : \dots : \Gamma_{k-2})$	$V_t :=$	$(\Delta X'_{t-1} : \dots : \Delta X'_{t-k+1})'$
$\Pi :=$	$-A(1)$	$Y_t :=$	$(Y'_{0t} : Y'_{1t})'$
$\mu_1 =$	$\alpha\beta'_0$	$Y_{0,t} :=$	$\beta'X_t + \beta'_0t$
$\mu :=$	$(\mu_0 - \mu_1 : \mu_d)$	$Y_{1,t} :=$	$\beta'_\perp \Delta X_t$
$\mu^\dagger :=$	$(\mu_0^\dagger : \mu_d^\dagger)$	$D_t :=$	$(1 : d'_t)'$
$\mu_0^\dagger :=$	$B'(\mu_0 - \mu_1 + \Gamma\bar{\beta}\beta'_0)$	$U_t :=$	$(Y'_{0,t-1} : V'_t)'$
$B :=$	$(\beta : \beta_\perp)$	$U_t^\dagger :=$	$(Y'_{t-1} : \dots : Y'_{t-k+1} : Y'_{0,t-k})'$
$\mu^\circ :=$	$(\mu_0^\circ : \mu_d^\circ)$	$G_t :=$	$(Y'_{0t} : \Delta X'_t c_\perp)'$
$(\mu_d^\circ : \epsilon_t^\circ) :=$	$B'(\mu_d : \epsilon_t)$	$\Psi :=$	$(\alpha : \gamma)$
$\mu_0^\circ :=$	$B'(\mu_0 - \mu_1 + \bar{\beta}\beta'_0)$	$\Psi^\dagger :=$	$(A_1^\circ : A_2^\circ : \dots : A_{k-1}^\circ : A_{k,0}^\circ)$
		$\Psi^\circ :=$	$(\alpha^\circ : \gamma^\circ) := B'(\alpha + \bar{\beta} : \gamma)$

Table 1: Notation.

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

$$Y_t = \sum_{i=1}^k A_i^\circ Y_{t-i} + \mu^\dagger D_t + \epsilon_t^\circ \quad (4)$$

where the last AR matrix  $A_k^\circ$  in (4) satisfies  $A_{k,1}^\circ = 0$ . 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 the moving average representation

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

where  $C^\circ(L)\epsilon_t^\circ = C_Y(L)\epsilon_t$  is a I(0) linear process with rank  $p$ , where  $C_{Y,0} = B' := (\beta : \beta_\perp)'$ , a full rank  $p \times p$  matrix. Incorporating  $A_{k,1}^\circ = 0$  in (4) one obtains  $U_t^\dagger := (Y'_{t-1} : \dots : Y'_{t-k+1} : Y'_{0,t-k})'$  as regressors, see (5) below; alternatively the regressors  $U_t^\dagger$  can be rotated into the same r.h.s. variables as the equilibrium correction form (3) obtaining the ‘mixed form’ (6):

$$Y_t = \Psi^\dagger U_t^\dagger + \mu^\dagger D_t + \epsilon_t^\circ \quad (5)$$

$$= \Psi^\circ U_t + \mu^\circ D_t + \epsilon_t^\circ, \quad (6)$$

The coefficients matrices  $\Psi^\dagger$  and  $\Psi^\circ$  in the ED and the mixed form are linked by  $\Psi^\circ = \Psi^\dagger A$ , where  $A$  is a square non-singular matrix; hence  $\text{col}(\Psi^\circ) = \text{col}(\Psi^\dagger)$  and  $\text{rank}(\Psi^\circ) = \text{rank}(\Psi^\dagger)$ .

**Remark i)** The transformation from  $X_t$  to  $Y_t$  maps the non-stationary VAR for  $X_t$  into a stable I(0) VAR for  $Y_t$ , which has full I(0) rank  $p$ .

**Remark ii)** The ED form (5) or (6) and the EC form (3) are equivalent, in the sense that any pair can be derived from the other one. However,  $\Psi$  in (3) may have a different rank than  $\Psi^\circ$  and  $\Psi^\dagger$ .

**Remark iii)** When  $k = 1$ ,  $V_t$  is void, the mixed form (6) coincides with the ED representation (5).

In the following corollary we show that the properties of the equilibrium dynamics for  $Y_t$  discussed in Theorem 2 are carried over to a different choice of the  $Y_{1t}$  component. Let  $c_\perp$  be any  $p \times p_1$  matrix such that  $\text{col}(\beta) \cap \text{col}(c_\perp) = \{0\}$  and define  $G_t := (Y'_{0t} : \Delta X'_t c_\perp)'$ .

**Corollary 3** *Under the same assumptions of Theorem 2,  $G_t$  follows a VAR( $k$ ) of the type  $B^\circ(L)Y_t = K\mu^\dagger D_t + K\epsilon_t^\circ$  where  $B^\circ(L)$  is stable,  $K$  is a square matrix of full rank defined in Appendix A and the AR coefficients of  $B^\circ(L)$  satisfy the constraint  $B_{k,1}^\circ = 0$ , equivalent to  $A_{k,1}^\circ = 0$ .*

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. The ED (5) or the EC form (3) can be used to discuss non-innovation common features. This issue is addressed in the following section.

## 4 Common cycles

This section discusses CC for a generic vector time series  $W_t$  extracted from  $X_t$ , where  $W_t$  is applied both to  $\Delta X_t$  and  $Y_t$ , and  $Y_t$  has been defined in Section 3. In this section the relative merits of these options are discussed also with respect to the example introduced at the end of Section 2.

A matrix  $b$ , of dimension  $p \times \ell$  and rank  $\ell$ , is defined to be a CC 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  $\ell$  when  $\ell$  is chosen to be maximal.

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

**Theorem 4**  *$W_t$  presents common feature with cofeature rank  $\ell$  if and only if  $\Psi^{(\cdot)}$  is of reduced rank, where  $\Psi^{(\cdot)} = \Psi$  in (3) for the choice  $W_t = \Delta X_t$ , and  $\Psi^{(\cdot)} = \Psi^\circ$  in (6) for the choice  $W_t = Y_t$ . The reduced rank condition  $\text{rank}(\Psi^{(\cdot)}) = p - \ell$  can be written  $\Psi^{(\cdot)} = \varphi\tau'$ , with  $\varphi$  and  $\tau$  of full column rank  $s := p - \ell$ . In this case the cofeature matrix  $b$  can be chosen equal to  $\varphi_\perp$ .*

In the rest of this section we discuss how the definition of CC allows to define a decomposition of the observed time series of  $W_t$ , see Subsection 4.1. We next discuss the two choices  $W_t = \Delta X_t$  and  $W_t = Y_t$  in more detail in Subsections 4.2 and 4.3. Since the CC cofeature rank  $\ell$  is bounded by the I(0) rank  $q$ , the upper bound  $\ell \leq q$  is seen to be more restrictive for the choice  $W_t = \Delta X_t$  than for  $W_t = Y_t$  because the I(0) rank of  $\Delta X_t$  is  $p_1 := p - p_0 \leq p$ , which is the I(0) rank of  $Y_t$ . However it is wise to investigate both choices  $W_t = \Delta X_t$ ,  $Y_t$  when there is no a priori information on what type of common features may apply.

### 4.1 Decomposition

The 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

$$W_t = \bar{b}\eta_t + \bar{b}_\perp c_t, \quad (7)$$

where  $c_t := b'_\perp W_t$  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  $I_p = a(b'a)^{-1}b' + b_{\perp}(a'_{\perp}b_{\perp})^{-1}a'_{\perp} =: a_b b' + b_{\perp a_{\perp}} a'_{\perp}$ , and hence

$$W_t = a_b \eta_t + b_{\perp a_{\perp}} c_t^{a_{\perp}}, \quad (8)$$

where  $c_t^{a_{\perp}} := a'_{\perp} W_t$ ,  $a_b := a(b'a)^{-1}$ ,  $b_{\perp a_{\perp}} := b_{\perp}(a'_{\perp}b_{\perp})^{-1}$ . Note that both decompositions (7) and (8) contain the same idiosyncratic component  $\eta_t$  with different loading matrices, which correspond to a different definition of the cyclical component,  $c_t$  and  $c_t^{a_{\perp}}$ . The cyclical parts are autocorrelated and present no CC.

The cyclical ( $c_t$  or  $c_t^{a_{\perp}}$ ) and idiosyncratic components ( $\eta_t$ ) are in general correlated, except for the choice  $a_{\perp} := \Omega^{-1}g_{\perp}$ . This particular decomposition has the property that  $Cov(\eta_t, c_t^{\Omega^{-1}g_{\perp}}) = g'\Omega\Omega^{-1}g_{\perp} = 0$ , i.e. the cyclic and idiosyncratic parts are uncorrelated. Both decompositions may be applied either to  $W_t = \Delta X_t$  or to  $W_t = Y_t$ .

## 4.2 CC in growth rates

Consider first the choice  $W_t = \Delta X_t$ . The properties of the CC cofeature matrix  $b$  are described in the following representation theorem, which is a restatement of Proposition 1 of Vahid and Engle (1993), with the addition of the new characterization (11).

**Theorem 5 (common cycles representation)** *Under the I(1) conditions, there exist a cofeature matrix  $b$  such that  $b'(\Delta X_t - E(\Delta X_t)) = b'\epsilon_t$  if and only if in (??) or (2) one has*

$$b' C_{0,i} = 0, \quad i = 0, 1, 2, \dots \quad (9)$$

$$b' C = b'. \quad (10)$$

When (9) holds, the second condition (10) holds if and only if

$$b = \alpha_{\perp} \zeta_{\perp} \quad \text{and} \quad \alpha'_{\perp} (I - \Gamma) \beta_{\perp} = \zeta \rho', \quad (11)$$

where  $c$  and  $d$  are  $p_1 \times p_1 - \ell$  matrices of rank  $p_1 - \ell$ .

Theorem 5 shows that, when it exists, the cofeature matrix  $b$  such that  $b'(\Delta X_t - E(\Delta X_t)) = b'\epsilon_t$  must be of the form  $b = \alpha_{\perp} u$  for an appropriate matrix  $u = \zeta_{\perp}$ . Hence, the CC cofeature matrix isolates the increments of  $\ell \leq p_1$  common I(1) trends, because it is equal to  $b'\epsilon_t = u'\alpha'_{\perp} \epsilon_t$ , and  $\alpha'_{\perp} \epsilon_t$  are the increments of the  $p_1$  common I(1) trends. The interpretation of the CC cofeature linear combinations  $b'(\Delta X_t - E(\Delta X_t))$  is that of increments of the common I(1) trends, so that  $b'X_t$  may be interpreted as driving stochastic trends in the system.<sup>2</sup>

## 4.3 CC in cointegrating relations

We next consider the choice  $W_t = Y_t$ . If the cofeature matrix  $b$  selects only elements of  $Y_{0t}$ , then the cofeature relations imply that certain deviations from equilibrium are innovation processes. If the cofeature matrix  $b$  selects elements from  $Y_{1t}$ , the interpretation is similar to the one given for the choice  $W_t = \Delta X_t$ . It would thus be useful in this context to test exclusion restrictions on  $b'Y_t$  similar to the ones of a

<sup>2</sup>We observe that the characterization (11) relates a necessary condition for CC in  $W_t = \Delta X_t$ , i.e. condition (10), to the same matrix  $\alpha'_{\perp} \Gamma \beta_{\perp}$  that appears in the I(1) conditions sub (c).

system of structural equations. We refer to this possibility as specification-test on  $b$ , which is treated in Section 6.3 below.

A consequence of Theorems 2 and 4 is that the CC properties of  $W_t = Y_t$  can be based on the ED (5) or (6), which share the same cofeature rank and cofeature space.

A possible disadvantage of this choice  $W_t = Y_t$  is that the components of  $Y_t$  are themselves linear combinations of  $X_t$  and  $\Delta X_t$ ; the economic interpretation of the components of  $Y_t$  is thus needed in order to interpret the cofeature matrix  $b$ . This problem, however, is solved by a careful modelling of the cointegration properties of a system and by appropriate specification testing on  $b$ , see again Section Section 6.3.

Moreover note that  $b'Y_t$  can always be expressed in terms of  $X_t$  and  $\Delta X_t$ . Partition in fact  $b = (b'_0 : b'_1)'$  conformably with  $Y_t := (Y'_{0t} : Y'_{1t})'$ , and note that  $b'Y_t = b'_0\beta'X_t + b'_1\beta'_\perp\Delta X_t =: c'_0X_t + c'_1\Delta X_t$ , say. This re-formulation is illustrated in Section 7.

We next apply the previous definitions to the Example 1 of Section 2; this example shows that the choice  $W_t = Y_t$  may be a sensible one.

**Example 6 (Ex. 1 continued)** *We first observe that  $\beta = (1 : -1)'$ ,  $\beta_\perp = (1 : 1)'$ ,  $Y_{0t} = X_{1t} - X_{2t}$ ,  $Y_{1t} = \Delta X_{1t} + \Delta X_{2t}$ ,  $Y_t = (Y_{0t} : Y_{1t})'$ . After tedious algebra, see the proof of Theorem 2, one can write the system in ED form (5) as follows*

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

Observe that the coefficient matrix  $\Psi^\dagger$  is of deficient rank, and that  $b = (1 : 0)'$  is a cofeature vector. The cofeature relation is  $X_{1t} - X_{2t} = a + \epsilon_{1t}^*$ , which states that no arbitrage opportunities exist between the two types of investments. Hence common features applied to  $Y_t$  correctly signal the presence of a common cycle. The ED mixed form (6) is

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

also for this representation one finds a common cycle with cofeature vector  $b = (1 : 0)'$ . Finally the EC representation is

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

where  $\epsilon_{1t} := \epsilon_{1t}^* + \epsilon_{2t}^*$ ,  $\epsilon_{2t} := \epsilon_{2t}^*$ . Note that the coefficient matrix  $\Psi$  on the r.h.s. is of full rank for any  $\varrho \neq 0$ , so that there is no cofeature vector for  $\Delta X_t$ . Note that in this case CC applied to  $\Delta X_t$  would not signal the presence of a common cycle, because there is no observable increment of the  $I(1)$  trend.

Both options  $W_t = \Delta X_t$ ,  $Y_t$  may turn out to be important. Ultimately which case applies remains an empirical question. Before addressing the problem of inference we consider extensions of the concept of CC. These are considered in the following section.

## 5 UP relations

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 EC or ED forms. This extension goes some way in providing an answer to the lack of invariance of CC to timing of the variables.<sup>3</sup> This phenomenon was first observed in Ericsson in his comments to Engle and Kozicki (1993). Several possible choices of additional lagged variables are considered.

Let  $R_t$  indicate 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 Unpredictable polynomial linear combination, UP, 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 cofeature rank  $\ell$  if  $\ell$  is chosen to be maximal.

This definition nests the one of CC. 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  $W_t = \Delta X_t$ ,  $R_t = v(L)X_{t-1}$ ; then the cofeature relations  $b_1'W_t + b_2'R_t = (b_1'\Delta + b_2'v(L)L)X_t =: b(L)X_t$  correspond to their Definition 1 for  $b(L) := (b_1'\Delta + b_2'v(L)L)$ . 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 UP relations is similar to CC; 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 UP relations always load on the contemporaneous variables  $W_t$ , in the sense of the following proposition.

**Proposition 7** *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  $X_t$ s and  $b$  is partitioned conformably with  $Z_t$ , then  $b_1$  has full column rank  $\ell$ .*

In the next proposition we state the necessary and sufficient conditions on the coefficient matrices in order to have common features of dynamic type; this proposition extends Theorem 4. 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  $W_t = \Delta X_t$  and  $C_{W,0} = B'$  for  $W_t = Y_t$ , see Theorem 2. The covariance matrix of  $\epsilon_t^*$  is indicated by  $\Omega^* := C_{W,0}\Omega C_{W,0}'$ . Similarly we let  $\mu_0^*$  indicate the appropriate coefficient of the constant.

**Theorem 8** *Let  $Z_{2t} := (R_t' : d_t)'$ ,  $Z_{0t} := W_t$ , and assume that  $Z_{0t}$ ,  $Z_{1t}$  and  $R_t$  are variables generated from a stationary VAR with innovations  $\epsilon_t$ , where  $Z_{0t}$  satisfies*

$$Z_{0t} = \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \quad (12)$$

and  $Z_{1t}$ ,  $R_t$  depend on lagged  $\epsilon_t^*$ s. 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  $\varsigma$  is of reduced rank,  $\varsigma = \varphi\tau'$ , with  $\varphi$  and  $\tau$  of full column rank. In this case the cofeature matrix has representation

$$b' := (b_1' : b_2) = (\varphi'_\perp : \varphi'_\perp \Phi_1) \quad \text{and} \\ b'(W_t' : R_t)' = \varphi'_\perp W_t + \varphi'_\perp \Phi_1 R_t = \varphi'_\perp (\Phi_2 d_t + \mu_0^* + \epsilon_t^*).$$

<sup>3</sup>In this respect, CC deviates from cointegration.

case	cofeatures	$b'_1\alpha^{(\cdot)}$	$b'_1\gamma^{(\cdot)}$	$Z_{1t}$	$Z_{2t}, 1$
(a)	CC	0	0	$\begin{pmatrix} Y_{0,t-1} \\ V_t \end{pmatrix}$	$D_t$
(b)	UP	0	unrestricted	$Y_{0,t-1}$	$\begin{pmatrix} V_t \\ D_t \end{pmatrix}$
(c)	UP	unrestricted	0	$V_t$	$\begin{pmatrix} Y_{0,t-1} \\ D_t \end{pmatrix}$

Table 2: Possible cofeature rank restrictions in the regression format of (13) using the notation  $RRR(Z_{0t}, Z_{1t}|Z_{2t}, 1)$ . The dependent variables  $Z_{0t}$  is either  $\Delta X_t$  for the EC form (3) or  $Y_t$  for the ED mixed form (6).  $\alpha^{(\cdot)}$  indicates either  $\alpha$  or  $\alpha^\circ$ ; similarly for  $\gamma$ .

We next illustrate possible choices for  $R_t$  using the equilibrium correction form (3) and the equilibrium dynamics form (6). In the empirical application we use the characterization given in Proposition 8, simply stating the reduced rank restrictions implied by different choices of variables in  $Z_{0t}$ ,  $Z_{1t}$ ,  $Z_{2t}$ . The corresponding technique of reduced rank regression RRR, see Anderson (1951) and the following section, is indicated with the shorthand  $RRR(Z_{0t}, Z_{1t}|Z_{2t}, 1)$ .

A list of different dynamic cofeatures cases is given in Table 2, using the format of equation (12). We observe that case (b) for  $W_t = \Delta X_t$  corresponds to the conditions for  $b'_1 X_t$  to be weakly exogenous for the cointegrating parameters  $\beta$ , see Johansen (1992a). These conditions simply state that the equations of  $b'_1 \Delta X_t$  in the equilibrium correction formulation (3) have zero adjustment coefficients, which may be described as ‘no feedback’ in the equations of  $b'_1 \Delta X_t$ . Observe also that  $1 \leq p_0 < p$  in the presence of cointegration, so that there always exist a cofeature matrix of this type, which corresponds to a basis of  $\text{col}^\perp(\alpha)$ , i.e.  $b_1 = \alpha_\perp$ . This case is thus trivially true for all cointegrated systems, and it is not discussed further.

Case (c) corresponds 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  $\Delta X_t$  inherit the cyclic part included in deviations from equilibrium in  $Y_{0,t-1}$ . Case (a) is the CC definition. Note that all these three cases equally apply to the choice  $W_t = Y_t$ , this enlarges the list of possible occurrences of common features. In the following proposition, we show that in one of the UP case both choices  $W_t = Y_t$  and  $W_t = \Delta X_t$  lead to the same system.

In the following we say that the cofeature properties of two sets of variables are equivalent if they have the same cofeature rank and the cofeature matrices are linearly related.

**Theorem 9** *The cofeature properties of  $U_{1t} := (\Delta X'_t : Y'_{0,t-1})'$  and  $U_{2t} := (Y'_t : Y'_{0,t-1})'$  are equivalent. Moreover let  $Z_{2t} := (Y'_{0,t-1} : d'_t)'$  and  $Z_{1t} := V_t$ ; then the canonical correlations involved in  $RRR(\Delta X_t, Z_{1t}|Z_{2t}, 1)$  and  $RRR(Y_t, Z_{1t}|Z_{2t}, 1)$  are the same, so that all statistics based on them are invariant with respect to the choice  $Z_{0t} = \Delta X_t$  or  $Y_t$  in these two systems.*

All submodels in Table 2 are compared with a baseline reference model, which is the unrestricted equilibrium correction formulation (3) or the equilibrium dynamics mixed form (6). Hence all possible testing procedures can be thought in line with the general-to-specific framework, see Johansen (1992b) or Paruolo (2001), to which we refer for details.

## 6 Estimation and testing

This section reviews inference on I(1) VAR systems with common trends and cycles based on reduced rank regression, RRR, see Anderson (1951). For the purpose of likelihood inference and in the application we take  $\epsilon_t$  to be i.i.d.  $N(0, \Omega)$ , where  $\Omega$  is positive definite. The cointegration analysis of I(1) systems has been extensively discussed in the literature; we refer to Johansen (1996) and reference therein.

In this section we concentrate on the CC and UP analysis after the cointegration analysis has been performed, fixing the cointegration parameters  $\beta$  to their maximum likelihood (ML) estimates, or to any superconsistent estimates. Due to superconsistency, using the estimates in place of the parameters does not change the limit distributions of the cofeature statistics described below, see Appendix B. In the rest of this section we simply do not distinguish  $\beta$  and its estimated values. Proofs of other statements in this section are also collected in Appendix B.

The RRR technique has been employed extensively in the literature of common features, see the comment of Hansen to Engle and Kozicki (1993), Vahid and Engle (1993), Hecq et al. (2000, 2002), Cubadda (1999, 2001). The RRR approach has been mainly employed to test for the cofeature rank. We here show that the same approach delivers various Wald tests and likelihood ratio (LR) tests that can be usefully employed in the specification of the CC and UP cofeature relations. The RRR approach also delivers maximum likelihood estimates under the various restrictions for fixed cointegration coefficients.

All instances of common features described in Section 4 and 5 correspond to regression models with reduced rank restrictions, see Theorems 4 and 8. The regression format is

$$Z_{0t} = \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \quad (13)$$

where the cofeature restriction is

$$H(s) : \quad \varsigma = \varphi \tau'. \quad (14)$$

$\varphi$ ,  $\tau$ ,  $\Phi$ ,  $\mu_0^*$  and  $\Omega^* := E(\epsilon_t^* \epsilon_t^{*\prime})$  are unrestricted and  $s$  indicates the number of columns in  $\varphi$ ,  $\tau$ , where  $\varphi$  is  $p \times s$  and  $\tau$  is  $j \times s$ .

For any given model, see Table 2, the analysis of cofeatures 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  $\epsilon_t$ .

Because when  $j < p$ , there always exist a cofeature matrix of rank  $p - j$ , we exclude this trivial case by assuming  $j \geq p$ , i.e. that there are at least as many regressors as dependent variables.<sup>4</sup> We indicate as the ‘ $H(s)$  model’ the regression model (13) under the reduced rank restriction (14), see Subsection 6.1.

We review four LR tests,  $Q_i$   $i = 1, \dots, 4$  and two Wald tests  $J_1$  and  $J_2$  that can be used in the specification analysis in Subsection 6.3. These are calculated explicitly in terms of eigenvalues and sample moment matrices, and hence do not require iterations. The proofs of the asymptotic distributions of the tests statistics when the cointegration parameters are known and fixed are summarized in Paruolo (2004) and reference therein. In Appendix B we show that the same proofs can be used here, due to the superconsistency of the cointegration parameters estimators.

---

<sup>4</sup>Most of the derivations are unaffected by this assumption.

In Subsection 6.2 we also report the asymptotic distribution of the dynamic cofeature matrix, which nests the results for the static cofeature matrix reported in Paruolo (2004). These form the basis of some of the Wald tests considered in Subsection 6.3.

## 6.1 Test on cofeature rank

The  $H(s)$  model is analyzed by the reduced rank regression  $\text{RRR}(Z_{0t}, Z_{1t}|Z_{2t}, 1)$ . The Gaussian log likelihood function is proportional to  $-T(\ln |\Omega^*| + \text{tr}(\Omega^{*-1} T^{-1} \sum_{t=1}^T \epsilon_t^* \epsilon_t^{*\prime}))$  and it is maximized by considering the following eigenvalue problem

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \quad (15)$$

with eigenvalues  $\lambda_1 > \dots > \lambda_i > \dots > \lambda_p$  and associated eigenvectors  $v_i$ , where  $S_{ij} := M_{ij} - M_{I1} M_{22}^{-1} M_{2j}$ ,  $M_{ij} := T^{-1} \sum_{t=1}^T (Z_{it} - m_i)(Z_{jt} - m_j)'$ ,  $m_i := T^{-1} \sum_{t=1}^T Z_{it}$ ,  $i, j = 0, 1, 2$ , see e.g. Johansen (1996). The LR test statistic for hypothesis (14) of  $H(s)$  versus  $H(p)$  about the rank of  $\zeta$  is given by

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

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

## 6.2 Estimation of the cofeature matrix

Eq. (15) provides also the ML estimates for given dimension  $s$ . In particular  $\hat{\tau} = (v_1 : \dots : v_s)$  and

$$\begin{aligned} \hat{\varphi} &= S_{01} \hat{\tau} (\hat{\tau}' S_{11} \hat{\tau})^{-1}, & \hat{\zeta} &= \hat{\varphi} \hat{\tau}' = S_{01} \hat{\tau} (\hat{\tau}' S_{11} \hat{\tau})^{-1} \hat{\tau}', \\ \hat{\Phi} &= (M_{02} - \hat{\zeta} M_{12}) M_{22}^{-1}, & \hat{\Omega}^* &= S_{00} - S_{01} \hat{\tau} (\hat{\tau}' S_{11} \hat{\tau})^{-1} \hat{\tau}' S_{10}, \end{aligned} \quad (16)$$

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

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  $\hat{\varphi}$  obtained by substituting  $\hat{\tau}_c$  in place of  $\hat{\tau}$  in (16) is given by  ${}_c \hat{\varphi} := \hat{\zeta} c$ , which satisfies  ${}_c \hat{\varphi} \hat{\tau}_c' = \hat{\zeta}$ . In the following we use the just-identifying normalization  $\hat{\varphi}_{\perp a_{\perp}} := \hat{\varphi}_{\perp} (a'_{\perp} \hat{\varphi}_{\perp})^{-1}$  also for the estimator of  $\varphi_{\perp}$ .

We consider the limit distribution of the estimator of the cofeature relations  $b' = (\varphi'_{\perp} : \varphi'_{\perp} \Phi_1)$  based on the results in Proposition 8. In order to state this limit distribution, we introduce the following notation. Let  $x_t := (Z'_{1t} : Z'_{2t})'$ ,  $y_t := (Z'_{0t} : x'_t)'$ ,  $\Sigma^* := E((y_t - E(y_t))(y_t - E(y_t))')$ , and indicate blocks of  $\Sigma^*$  with the subscripts 0, 1, 2,  $x$ . Recall also that  $Z_{2t} := (R'_t : d'_t)'$ , and let  $\Sigma_{ij,h}^* := \Sigma_{ij}^* - \Sigma_{ih}^* \Sigma_{hh}^{*-1} \Sigma_{hj}^*$  for  $i, j, h = 0, 1, 2, x, R, d$  and in particular  $\Sigma_{ij} := \Sigma_{ij,2}^*$  for  $i, j = 0, 1, x$ .

**Theorem 10** *Under the  $I(1)$  assumption and the hypothesis  $H(s)$  in eq. (14), the limit distribution of the dynamic cofeature matrix  $\hat{b}' = (\hat{\varphi}'_{\perp a_{\perp}} : \hat{\varphi}'_{\perp a_{\perp}} \hat{\Phi}_1)$ , where  $\varphi_{\perp}$  is identified via  $\varphi_{\perp a_{\perp}}$ , is given by*

$$T^{1/2} \text{vec}(\hat{b} - b) \xrightarrow{d} N(0, \mathcal{V}), \quad (17)$$

where

$$\begin{aligned}\mathcal{V} &:= \varphi'_{\perp a} \Omega^* \varphi_{\perp a} \otimes \mathcal{H}' \Sigma_{xx}^* \mathcal{H}, & \mathcal{H} &:= (-\mathcal{A}_1 : \mathcal{A}_2 - \mathcal{A}_1 \Phi_1 + \mathcal{A}_3), \\ \mathcal{A}_1 &:= \mathcal{B}_1 \zeta' a (a' \zeta \Sigma_{11} \zeta' a)^{-1} a', & \mathcal{A}_2 &:= \mathcal{B}_1 \tau (\tau' \Sigma_{11} \tau)^{-1} \tau' \Sigma_{1R.d}^* \Sigma_{RR.d}^{*-1}, \\ \mathcal{A}_3 &:= (0 : I : -\Sigma_{dd}^{*-1} \Sigma_{dR}^*)', & \mathcal{B}_1 &:= (I : -\Sigma_{22}^{*-1} \Sigma_{21}^*)'.\end{aligned}$$

The asymptotic variance  $\mathcal{V}$  can be consistently estimated by  $\widehat{\mathcal{V}}$  where parameters are substituted with their ML estimates and the moment matrices  $\Sigma_{ij}^*$  with sample moments  $M_{ij}$ .

In the special case of contemporaneous cofeatures  $\widehat{b} = \widehat{\varphi}_{\perp a_{\perp}}$ , the limit distribution given in (17) simplifies to

$$T^{1/2} \left( \text{vec}(\widehat{a}'(\widehat{\varphi}_{\perp a_{\perp}} - \varphi_{\perp a_{\perp}})) \right) \xrightarrow{d} N \left( 0, (\varphi'_{\perp a_{\perp}} \Omega^* \varphi_{\perp a_{\perp}} \otimes (a' \zeta \Sigma_{11} \zeta' a)^{-1}) \right). \quad (18)$$

Again a consistent estimator of the asymptotic covariance matrix is obtained substituting parameters with the corresponding ML estimator and  $\Sigma_{11}$  with  $S_{11}$ .

### 6.3 Specification tests

Theorem 10 allows to derive specification tests on the cofeature matrix  $b$ . Hypothesis of the form  $K' \text{vec}(b) = j$ , where  $K$  has rank  $h$ , can be tested via the Wald statistic

$$J_1 := T(K' \text{vec}(\widehat{b}) - j)' (K' \widehat{\mathcal{V}} K)^{-1} (K' \text{vec}(\widehat{b}) - j).$$

The test  $J_1$  has an asymptotic  $\chi^2(h)$  limit distribution, provided the hypothesis  $K' \text{vec}(b) = j$  regards the un-normalized parameters in  $b$ , i.e. if  $K' \mathcal{V} K$  is positive definite.<sup>5</sup>

We next deal with the special case of a generic linear hypothesis on  $\varphi_{\perp}$ ; this is relevant e.g. in the contemporaneous cofeature case. Consider hypotheses of the type  $K' \text{vec}(\widehat{a}' \varphi_{\perp a_{\perp}}) = j$ , where  $K$  has  $h$  columns. Note that  $\widehat{a}' \varphi_{\perp a_{\perp}}$  are the un-normalized coefficients in  $\varphi_{\perp}$  when it is identified via  $\varphi_{\perp a_{\perp}}$ . The associated Wald test is given by

$$J_1 := T(K' \text{vec}(\widehat{a}' \widehat{\varphi}_{\perp a_{\perp}}) - j)' \left( \left( \widehat{\varphi}'_{\perp a_{\perp}} \widehat{\Omega}^* \widehat{\varphi}_{\perp a_{\perp}} \right)^{-1} \otimes (a' \widehat{\zeta} S_{11} \widehat{\zeta}' a) \right) (K' \text{vec}(\widehat{a}' \widehat{\varphi}_{\perp a_{\perp}}) - j). \quad (19)$$

Also this test is shown to be asymptotically  $\chi^2(h)$  on the basis of Theorem 10 and to diverge under fixed alternatives.

A further special case of the hypothesis  $K' \text{vec}(\widehat{a}' \varphi_{\perp a_{\perp}}) = j$  is given by restrictions of the type

$$H_0 : \quad \varphi_{\perp} = H \phi, \quad (20)$$

which accommodate exclusion restrictions for all columns of  $\varphi_{\perp}$  simultaneously. Here  $H$  is  $p \times h$ ,  $h \geq \ell$ . Under the restriction (20), the likelihood function is maximized by solving

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

with eigenvalues  $\lambda_1^* > \dots > \lambda_h^*$  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 (20) in  $H(s)$  is given by

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

<sup>5</sup>We hence exclude hypothesis concerning the normalized coefficients  $b'_{1a_{\perp}} = \widehat{b}'_{1a_{\perp}} = I$ , see e.g. the discussion in Paruolo (1997).

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

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.

$$H_0 : \quad K'\zeta = 0. \quad (21)$$

A Wald test of (21) can be based on the unrestricted maximum likelihood estimates  $\widehat{\zeta} := S_{01}S_{11}^{-1}$ ,  $\widehat{\Omega}^* := S_{00.1} := S_{00} - S_{01}S_{11}^{-1}S_{10}$ , and equals

$$J_2 := Ttr \left( (K'S_{00.1}K)^{-1}K'\widehat{\zeta}S_{11}\widehat{\zeta}'K \right) = Ttr \left( (K'S_{00.1}K)^{-1}K'S_{01}S_{11}^{-1}S_{10}K \right). \quad (22)$$

Also this test is asymptotically  $\chi^2(hj)$  and diverges under fixed alternatives. The corresponding LR test of (21) in  $H(s)$ , labelled  $Q_3$ , is found by solving the eigenvalue problem

$$\left| \lambda^{\circ} K'_{\perp} S_{00.K} K_{\perp} - K'_{\perp} S_{01.K} S_{11.K}^{-1} S_{10.K} K_{\perp} \right| = 0, \quad (23)$$

with eigenvalues  $\lambda_1^{\circ} > \dots > \lambda_{p-h}^{\circ}$  and corresponding eigenvectors  $v_i^{\circ}$ , 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 (21) in  $H(s)$  is given by

$$Q_3 := T \left( \sum_{i=1}^s \ln(1 - \lambda_i^{\circ}) - \sum_{i=1}^s \ln(1 - \lambda_i) \right) \quad (24)$$

where  $S_{00.1} := S_{00} - S_{01}S_{11}^{-1}S_{10}$ . The restricted estimate under (21) is  $\widehat{\varphi}_{\perp} = (K : K_{\perp}(v_{s+1}^{\circ} : \dots : v_{p-s-h}^{\circ}))$ , 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 (21) in  $H(p)$ ,  $Q_4 := Q_1(s) + Q_3$ . Again  $Q_4 \xrightarrow{d} \chi^2(df_{Q_4})$ , 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.

We also observe that  $\varphi'_{\perp} Z_{0t} = u_t$ , where  $u_t$  are  $\ell$  linear combinations of  $\epsilon_t$ , defines a system of  $\ell$  simultaneous equations. Homogeneous separable restrictions on each equation can be written in the form

$$\varphi_{\perp} = (H_1\phi_1 : \dots : H_{\ell}\phi_{\ell}),$$

see Johansen (1995) 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 to (15), interchanging  $\beta$  and  $\varphi_{\perp}$ , the subscripts 0 and 1, and choosing the smallest eigenvalues instead of the largest ones.

## 7 An application to monthly US data

In this section we illustrate the techniques proposed in the paper, by reconsidering the US monthly time series data analyzed in Kim (2003). The data consists of the log of S&P 500 composite stock price index, indicated by  $lp$ , the log of index of industrial production,  $lip$ , the log of the US dollar exchange rate,  $lxr$ , the log of Moody's Aaa corporate bond yield,  $lr$ , and the inflation rate  $if$ , calculated as the first difference of

the log of the CPI index times 100. The exchange rate series  $lax$  is trade-weighted and price adjusted. For a complete description of the data series and of the statistical sources we refer to Kim's original paper. All data series are monthly, and the available sample period is 1974:1 to 1998:12, for a total of 300 periods.

These time series were selected by Kim to investigate the relation between the stock price index and its fundamentals, with special reference to exchange rates. The choice of 1974 and of 1998 as the starting and ending years of the period is motivated by the beginning of floating exchange rates of the major currencies and the introduction of the Euro. A review of the applied econometric literature on the relation between exchange rates and stock market is reported in Kim's paper, which innovates on the existing literature by considering all the above five variables simultaneously within a VAR framework, and by allowing for cointegration.

The econometric analysis of this system was performed in 3 steps. In the first step we investigated the initial specification of the system, i.e. the choice of lag length, and we performed a preliminary specification analysis. This analysis is reported in Subsection 7.1; the results suggest  $k = 13$  and the absence of autoregressive conditionally heteroskedasticity (ARCH) effects. Given the large number of regressors, we performed a preliminary reduction of relevant variables in  $V_t$  that brought the dimension of  $V_t$  from 60 to 15.

We next tested for common trends; this analysis is reported in Subsection 7.2 below. As in Kim's paper, we found one cointegrating vector. The analysis of the common cycles was performed next, and it is reported in Subsection 7.3. One cofeature vector could be detected in the equilibrium dynamics form but not in the equilibrium correction form.

In all the analysis we employed a nominal size of 1% in each test, given the moderately large sample size. Calculations were performed in Gauss 6 and PcGive 10.

## 7.1 Initial specification

We reconstructed the dataset by accessing the original statistical sources. We managed to replicate the sample correlations among the five time series as reported in Kim's (2003) Table 1 to the second (and last reported) digit, except for the correlations involving  $if$ , which we only managed to match to the first digit. This may be due to recent revisions in the CPI index.

The time series in levels and first differences are pictured in Fig. 1. All time series appear non-stationary in levels, but not in first differences. Un-surprisingly, the time series of the industrial production series shows signs of seasonal variation. We thus included seasonal dummies in  $d_t$ , expanding Kim's original specification.

We also included five dummy variables that take the value 1 only for a single time period, labelled  $d_{yy:mm}$ , where  $yy$  stands for year and  $mm$  for month. These dummies take into account the stock market crash of October 1987,  $d_{87:10}$ , and the Russian crisis of August 1998,  $d_{98:8}$ . Several other large changes in the variables occurred in 1978:4 (second oil shock), 1980:7, 1991:3 (first gulf war); we thus created the dummies  $d_{78:4}$ ,  $d_{80:7}$ ,  $d_{91:3}$ . In the final specification  $D_t$  thus contained the constant, 11 seasonal dummies, and the 5 specific dummies listed above, for a total of 17 deterministic regressors.

We next estimated a VAR in the form (1), excluding the linear trend as in Kim.<sup>6</sup> Kim adopted a VAR(12) specification; in order to include the 12-th lag of  $\Delta X_{t-i}$  in

---

<sup>6</sup>Inclusion of a trend did not change the results, so we do not report results for this extension.

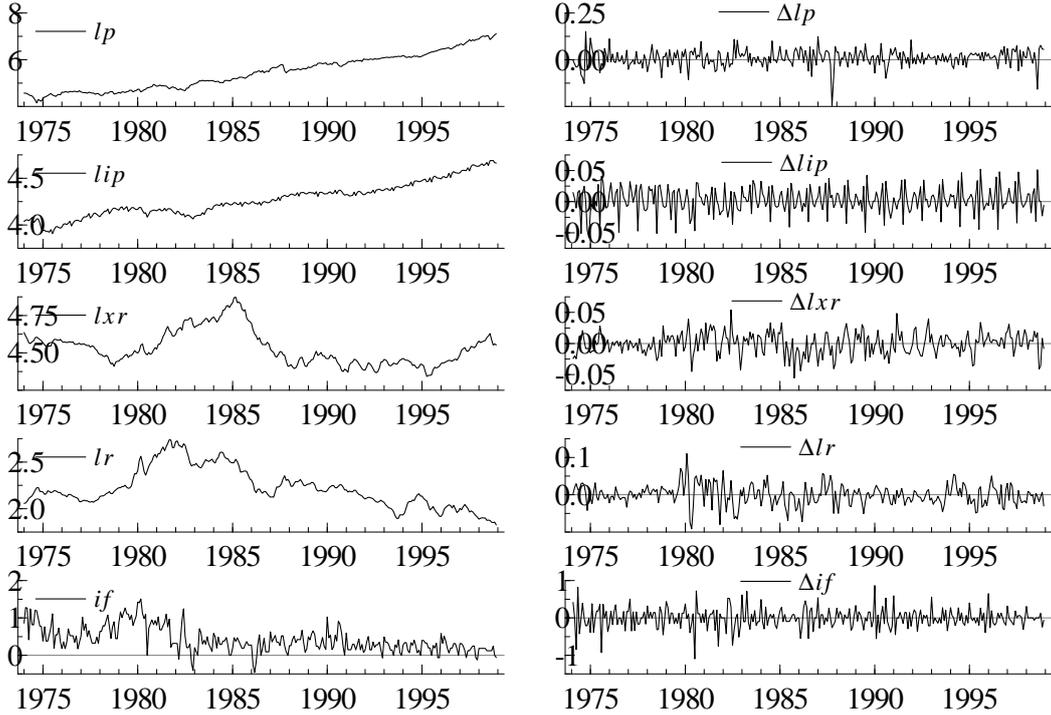


Figure 1: Levels and first differences of the monthly time series of  $lp$ ,  $lip$ ,  $lxr$ ,  $lr$ ,  $if$ , 1974:1 to 1998:12.

$V_t$  in (1), we started from a VAR(13) specification. This choice leads to  $5 \cdot 12 = 60$  stochastic regressors in  $V_t$ . The present specification encompasses Kim's specification because of the additional lag and for the inclusion of dummies. We estimated all parameters in (1) unrestrainedly, including  $\Pi$ . The effective sample period was 1975:2 to 1998:12 for a total of 287 time periods.

The residuals were tested for possible mis-specification, using Pc-Give 10.0. The summary of these tests are reported in Table 3. Both the Pc-Give individual and system mis-specification statistics indicated no departure from normality and no serial correlation. Univariate mis-specification statistics did not indicate any departure from homoskedasticity against ARCH. We thus concluded that the unrestricted system was correctly specified.

While correctly specified, the initial VAR(13) specification could well be overparameterized. Each equation presents  $60 + 17 + 5 = 82$  regressors. The total number of parameters in the unrestricted initial specification is equal to  $kp^2 + k(k+1)/2 + 17 = 13 \cdot 25 + 13 \cdot 7 + 17 = 433$ , a rather large number. This number should be contrasted with a total number of observations of  $Tp = 287 \cdot 5 = 1435$ . The ratio of number of parameters to observations is  $433/1435 \simeq 30\%$ , which is usually considered acceptable. Before performing common trends and cointegration analyses, we tried to reduce the dimension of the model in order not to dilute sample information.

Note that if all stochastic regressors in  $V_t$  were insignificant, then common features would trivially hold; this gives another motivation for attempting this reduction. The preliminary selection of relevant stochastic regressors in  $V_t$  is thus very important to substantiate the common features analysis.

We hence considered various submodels in the format of eq. (1) obtained by reducing the number of variables included in  $V_t$ . The choice was based on the summary regression statistics of Pc-Give 10.0 under the heading "F-tests on retained regres-

	Portm.(24)	AR 1-6	Normality	ARCH 1-12	heterosk.
<i>Stat<sub>i</sub></i> for eq. <i>i</i>					
$\max_{1 \leq i \leq 5} Stat_i$	29.8596	1.9604	8.1531	1.0756	0.70825
ref. distribution		$F(6,199)$	$\chi^2(2)$	$F(12,181)$	$F(130,74)$
<i>p</i> -value		[0.0730]	[0.0170]	[0.3832]	[0.9566]
system <i>Stat</i>					
<i>Stat</i>	419.58	1.2136	14.415		0.51996
ref. distrib.		$F(150,850)$	$\chi^2(10)$		$F(1950,978)$
<i>p</i> -value		[0.0538]	[0.1549]		[1.0000]

Table 3: Mis-specification tests for the unrestricted VAR(13) specification. *Stat<sub>i</sub>* indicates test statistics associated with the equation of  $\Delta X_i$ . *Stat* indicates system test statistics. We report the maximal statistics across equations,  $\max_{1 \leq i \leq 5} Stat_i$ . *p*-values in square brackets [ ]. Output of Pc-Give 10.0.

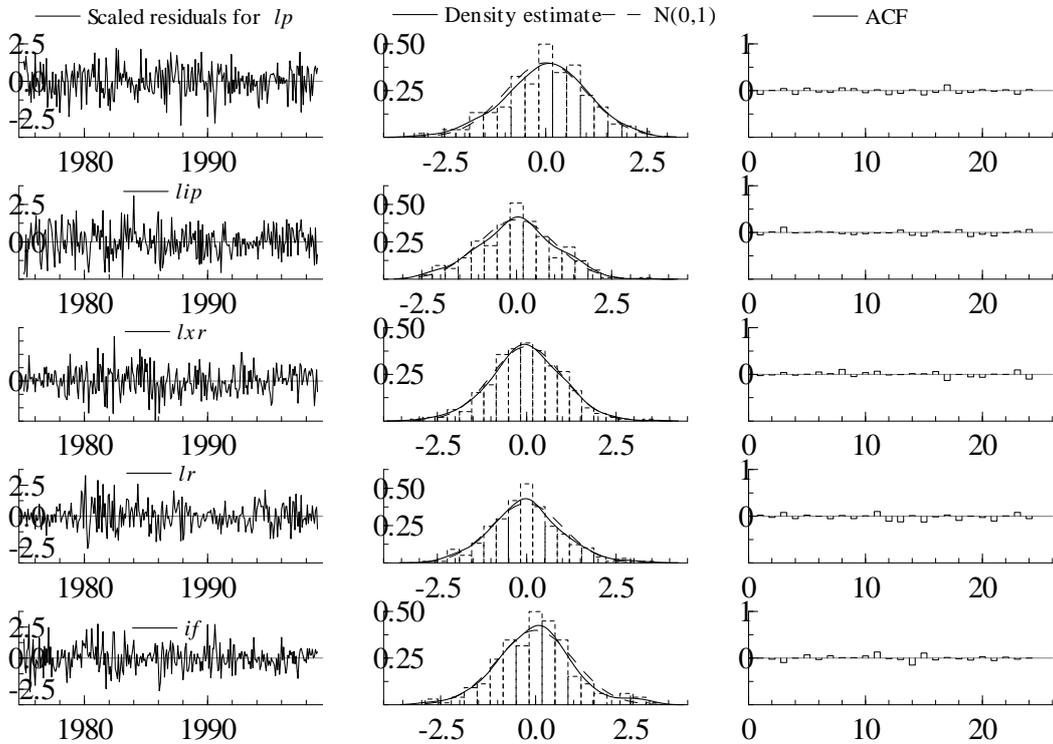


Figure 2: Residuals for the specification with  $V_t$  described in Table 4, with their histogram, estimated densities and autocorrelation functions (ACF); 1975:2 to 1998:12.

	$\Delta lp_{t-i}$	$\Delta lip_{t-i}$	$\Delta lxr_{t-i}$	$\Delta lr_{t-i}$	$\Delta if_{t-i}$
$i$	1, 6, 10	1, 10, 12	1, 2	1, 2, 6, 9	1, 7, 8

Table 4: Selected variables in  $V_t$  for the restricted model.

$H_0$	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r \leq 4$
LR trace test	126.86	52.82	23.62	7.90	2.28
[1% critical values]	[76.37]	[53.91]	[34.87]	[19.69]	[6.64]

Table 5: Cointegration trace tests. The 1 % critical values are taken from Johansen (1996) Table 15.3.

sors". The regressors included in  $V_t$  in the resulting restricted model are listed in Table 4. The  $F(225, 1005)$  statistics for the reduction from the unrestricted to the restricted model was equal to 1.0258 with a  $p$ -value of 0.3949, giving ample support to the reduction. All regressors were found to be significant except for  $\Delta lip_{t-1}$  and  $\Delta if_{t-1}$ , which were included anyway in order to include all first differences at lag 1. The total number of retained regressors was thus reduced to 37, down from 82.

Summarizing, the final breakdown of the 37 regressors was the following:  $p = 5$  variables in levels in  $X_{t-1}$ , 17 deterministic regressors in  $D_t$  (the constant, 11 seasonal and 5 single-period dummies), and 15 stochastic regressors included in  $V_t$ , listed in Table 4. The residuals from the restricted specification still preserved the general features of the unrestricted residuals detailed above. These residuals are graphed in Fig. 2, along with their histograms, estimated densities and autocorrelation functions (ACF). The rest of the analysis was thus performed with this restricted specification for  $V_t$ .

## 7.2 Common trends

We next performed the cointegration analysis. Kim found 1 cointegrating vector, i.e. 4 common trends. The trace tests for the present specification are reported in Table 5; they differ from Kim's because of the difference in the specification of  $V_t$  and  $D_t$ . Also the present results suggest  $p_0 = 1$  cointegrating vector, as in Kim's paper.

The estimated  $\beta$  vector was

$$ci_t := \hat{\beta}' X_t = lp_t - 2.36lip_t + 0.88lxr_t + 0.29lr_t + 1.87if_t,$$

which has the same signs as in Kim's paper, although there are some differences in the numerical estimates. We next tested the following specification

$$lp_t - lip_t = lxr_t + a(lip_t - if_t) + ci_t$$

which is an hypothesis of the type (20). If this specification holds,  $lr$  can be excluded from the cointegrating vector. A possible interpretation of this relation is that the log of the ratio of stock prices to industrial production (similar to an aggregate price/earning ratio) is homogeneous of degree 1 to the effective exchange rate and of degree  $a$  to real productivity as measured by  $lip_t - if_t$ .

This gave a LR test of  $Q_2 = 1.1677$  with a  $\chi^2(3)$   $p$ -value of 0.7608. The further restriction of  $a = 1.5$  gave an overall  $Q_2$  test of 3.7368 with a  $\chi^2(4)$   $p$ -value of 0.4428. We thus concluded that

$$\beta = \beta_{CI}, \quad \text{where } \beta_{CI} := \left( 1 \quad -5/2 \quad 1 \quad 0 \quad 3/2 \right)'. \quad (25)$$

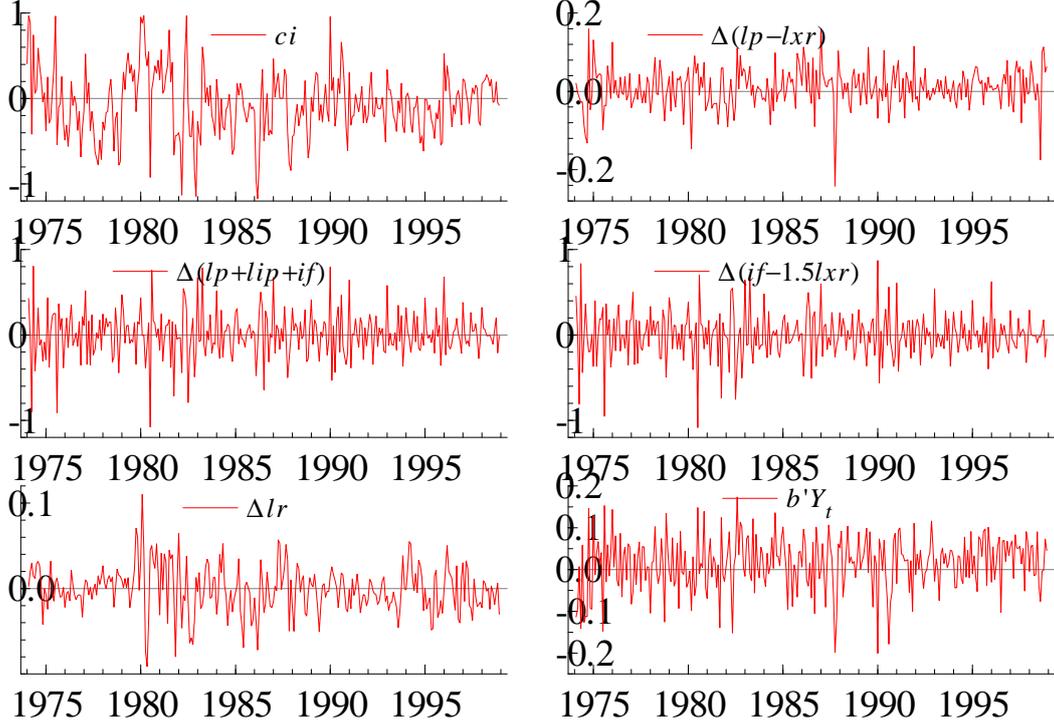


Figure 3: Transformed system  $Y_t$  and CC relation  $b'Y_t$ .

We hence fixed  $\beta$  at  $\beta_{CI}$  in (25) and indicate  $ci_t := \beta'_{CI}X_t$ . The basis of the orthogonal complement of  $\beta_{CI}$  can be chosen as follows

$$\beta_{\perp} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & -1.5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

The transformed system  $Y_t := (Y'_{0,t} : Y'_{1,t})' := (X'_t\beta : \Delta X'_t\beta_{\perp})'$  is pictured in Fig. 3.

### 7.3 Common cycles

We finally performed the static cofeature analysis, both for the equilibrium dynamics form,  $Z_{0t} := Y_t$ , and for the equilibrium correction form,  $Z_{0t} := \Delta X_t$ . In the case of the equilibrium dynamics, we chose to work with the mixed form, in order to have the same right-hand-side variables.  $Z_{1t}$  was hence set equal to  $Z_{1t} := (ci_{t-1} : V'_t)'$  (16 regressors) while  $Z_{2t} := D_t$  (17 regressors). The results of the LR test on the cofeature dimension are reported in Table 6 both for  $Z_{0t} = \Delta X_t$  and  $Z_{0t} = Y_t$ . It can be seen that, at the nominal 1% level, one must conclude that no cofeature vectors exist for the equilibrium correction form, while  $\ell = 1$  cofeature vector exists for the equilibrium dynamics mixed form.

The estimated cofeature vector  $b$  is reported in Table 7, along with standard errors based on eq. (18) and the  $t$ -version of the corresponding  $J_1$  statistic. It can be seen that all variables load significantly on  $b$ .

Given that  $Y_t$  is a transformed system, one may wish to express the cofeature relations in terms of the original variables in  $X_t$ . Recall that, partitioning  $b = (b'_0 : b'_1)'$  conformably with  $Y_t := (Y'_{0,t} : Y'_{1,t})'$ , one has  $b'Y_t = b'_0\beta'X_t + b'_1\beta'_{\perp}\Delta X_t =: c'_0X_t + c'_1\Delta X_t$ ,

$H(s)$		$Z_{0t} = \Delta X_t$			$Z_{0t} = Y_t$		
$\ell$	$s$	$Q_1(s)$	$df$	$100 \cdot p\text{-value}$	$Q_1(s)$	$df$	$100 \cdot p\text{-value}$
$\ell = 5$	$s = 0$	419.5565	80	0.00	1651.1291	80	0.00
$\ell \geq 4$	$s \leq 1$	248.0962	60	0.00	250.7385	60	0.00
$\ell \geq 3$	$s \leq 2$	144.5994	42	0.00	158.1744	42	0.00
$\ell \geq 2$	$s \leq 3$	67.0296	26	0.00	81.9780	26	0.00
$\ell \geq 1$	$s \leq 4$	26.7864	12	0.83	23.8917	12	2.10

Table 6: Cofeature dimension LR tests  $Q_1(s)$  for the equilibrium correction form ( $Z_{0t} = \Delta X_t$ ) and the equilibrium dynamics mixed form ( $Z_{0t} = Y_t$ ).  $100 \cdot p$  values are calculated on the basis of the  $\chi^2$  asymptotic distribution.

	$ci$	$\Delta(lp - lxr)$	$\Delta(lp + lip + if)$	$\Delta(if - 1.5lxr)$	$\Delta lr$
$b'$	-0.045	0.55	1	-1.10	0.29
SE	0.022	0.31	.	0.035	0.33
$t$ statistics	-2.09	1.73	.	-31.49	0.86

Table 7: Estimated cofeature vector;  $a_{\perp} = (0 : 0 : 0 : 0 : 1)'$ . SE indicates the standard errors based on eq. (18).

say, see Section 4. The implied coefficients  $c_0$  and  $c_1$  for the levels and differences are reported in the Table 8. These coefficients reflect the impact of the initial values that affect the levels but not the differences of the process.

## 8 Conclusions

In this paper we have extended the areas of application of the notion of common features in I(1) systems, allowing for innovation processes that depend also on the cointegrating relations. We have discussed how to address inference both for known and unknown cofeature vectors, using reduced rank regression.

Similarly to the case of cofeature analysis for the first differences of the process, the asymptotics suggests to perform the cointegration analysis first and then to proceed to the analysis of common features. After fixing the cointegration parameters, all subsequent inference is LAN.

Applying some of the proposed techniques to the monthly US data analyzed in Kim (2003), we found one cointegrating relation and one cofeature vector which involves the cointegrating relation. This system thus presents 4 common cycles in the equilibrium dynamics form and one innovation process. No common cycles were instead detected for the equilibrium correction form.

## References

$i$		$\Delta^i lp$	$\Delta^i lip$	$\Delta^i lxr$	$\Delta^i lr$	$\Delta^i if$
0	$c'_0$	1	-2.5	1	0	1.5
1	$c'_1$	-34.01	-22.02	-24.42	-6.35	2.25

Table 8: Coefficients of the cofeature relation for the levels ( $c_0$ ) and the first differences ( $c_1$ ) of  $X_t$ , rescaled such that the coefficient for the level of  $lp$  is equal to 1.

- 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.
- 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.
- Campbell, J. Y. and R. J. Shiller (1987), Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062–1087.
- Cubadda G. and A. Hecq (2001), On non-contemporaneous short-run comovements, *Economics Letters*, 73, 389-397.
- 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.
- Gonzalo J. and Granger C. W. J (1995) Estimation of Common Long-Memory Components in Cointegrated Systems, *Journal of Business and Economic Statistics*, 13, 27-36.
- Granger C. W. J, (2003) Methodological Questions from Using Very Large Data Sets, in Giles D. E. A. (ed.) *Computer-Aided Econometrics*, Marcel Dekker e-book, Statistics: Textbooks and Monographs series Volume169.
- Hecq, A., F. C. Palm and J. P. Urbain, 2004, Common cyclical features analysis in VAR models with cointegration, *Journal of Econometrics*, forthcoming.
- 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), Cointegration analysis in partial systems and the efficiency of single-equation analysis, *Journal of Econometrics* 52, 389-402.
- Johansen S. (1992b), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics* 54, 383–97.
- Johansen S. (1995), 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.
- Kim K. (2003), Dollar exchange rate and stock price: evidence from a multivariate cointegration and error correction model, *Review of Financial Economics* 12, 301-313
- 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, Università dell’Insubria Varese, Italy, Department of economics WP 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. (2001) The power of lambda max, *Oxford Bulletin of Economics and Statistics* 63/3, 395-403.
- Paruolo P. (2004) Common trend and cycles in I(2) VAR systems, *Journal of Econometrics*, forthcoming.
- 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 Theorem 2. Let  $u_t := \mu_d d_t + \epsilon_t$ ,  $\mu_0^\dagger := \mu_0 - \mu_1$ ,  $\Gamma(L) := I - \sum_{i=1}^{k-1} \Gamma_i L^i$ ,  $\Gamma = \Gamma(1)$ ,  $B := (\beta : \beta_\perp)$  and

$$Y_t := \begin{pmatrix} Y_{0t} \\ Y_{1t} \end{pmatrix} := \begin{pmatrix} \beta' X_t + \beta_0' t \\ \beta_\perp' \Delta X_t \end{pmatrix}.$$

Write the equilibrium correction form (3) as

$$\Gamma(L) \Delta X_t = \alpha Y_{0,t-1} + \mu_0^\dagger + u_t$$

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

$$(\Gamma(L) \bar{\beta} \Delta : \Gamma(L) \bar{\beta}_\perp) \begin{pmatrix} \beta' X_t + \beta_0' t \\ \beta_\perp' \Delta X_t \end{pmatrix} = \alpha Y_{0,t-1} + \mu_0^\dagger + u_t.$$

where  $\mu_0^\ddagger := \mu_0^\dagger + \Gamma \bar{\beta} \beta_0'$ . Rearranging

$$(\rho_0(L) : \rho_1(L)) Y_t := (\Gamma(L) \bar{\beta} \Delta - \alpha L : \Gamma(L) \bar{\beta}_\perp) \begin{pmatrix} Y_{0t} \\ Y_{1t} \end{pmatrix} = \mu_0^\ddagger + u_t. \quad (26)$$

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

$$B' (\rho_0(L) : \rho_1(L)) \begin{pmatrix} Y_{0t} \\ Y_{1t} \end{pmatrix} = \mu_0^\circ + B' u_t.$$

where  $\mu_0^\circ := B' \mu_0^\ddagger$ . Spelling out the coefficients of the lag polynomial for the first block of  $\rho(L) := (\rho_0(L) : \rho_1(L))$  one finds

$$\begin{aligned} \rho_0(L) &= \Gamma(L) \bar{\beta} \Delta - \alpha L = \bar{\beta} (1 - L) - \sum_{i=1}^{k-1} \Gamma_i \bar{\beta} (L^i - L^{i+1}) - \alpha L = \\ &= \bar{\beta} - (\alpha + \bar{\beta} + \Gamma_1 \bar{\beta}) L - \sum_{i=2}^{k-1} (\Gamma_i - \Gamma_{i-1}) \bar{\beta} L^i + \Gamma_{k-1} \bar{\beta} L^k \end{aligned}$$

Similarly for the second block  $\rho_1(L) = \Gamma(L) \bar{\beta}_\perp = \bar{\beta}_\perp - \sum_{i=1}^{k-1} \Gamma_i \bar{\beta}_\perp L^i$ . Note that this polynomial is of degree  $k - 1$ . The AR matrices are thus

$$\begin{aligned} A_1^\circ &= B' (\alpha + (I + \Gamma_1) \bar{\beta} : \Gamma_1 \bar{\beta}_\perp), & A_i^\circ &= B' ((\Gamma_i - \Gamma_{i-1}) \bar{\beta} : \Gamma_i \bar{\beta}_\perp), & i &= 2, \dots, k-1 \\ A_k^\circ &= -B' (\Gamma_{k-1} \bar{\beta} : 0). \end{aligned}$$

The last expression implies the restrictions (??) and that  $Y_t = \Psi^\dagger U_t^\dagger + \mu^\dagger D_t + \epsilon_t^\circ$ .

The stability of the roots of the AR polynomial  $A^\circ(L)$  under the I(1) assumptions and that  $Y_t$  is an I(0) process of rank  $p$  are proved in Johansen's proof of Granger's representation theorem, see Johansen (1996), who also describes how to transform  $Y_t$  back to the autoregressive form, and hence to the equilibrium correction form (??).

In order to derive the mixed form, simply note that  $U_t := (Y'_{0,t-1} : V'_t)' = QU_t^\dagger + a_0$ , where  $U_t^\dagger := (Y'_{t-1} : \dots : Y'_{t-k+1} : Y'_{0,t-k})'$  and  $Q$  is a full rank matrix. In fact

$$\begin{pmatrix} Y_{0,t-1} \\ \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-k+1} \end{pmatrix} = \begin{pmatrix} 0 \\ -\beta'_0 \\ \vdots \\ 0 \\ -\beta'_0 \end{pmatrix} + \begin{pmatrix} I & & & & & & \\ \bar{\beta} & \bar{\beta}_\perp & -\bar{\beta} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \bar{\beta} & \bar{\beta}_\perp & -\bar{\beta} \end{pmatrix} \begin{pmatrix} Y_{0,t-1} \\ \beta'_\perp \Delta X_{t-1} \\ \vdots \\ Y_{0,t-k+1} \\ \beta'_\perp \Delta X_{t-k+1} \\ Y_{0,t-k} \end{pmatrix}$$

and  $U_t^\dagger = AU_t + m := Q^{-1}(U_t - a_0)$  i.e.

$$\begin{pmatrix} Y_{0,t-1} \\ \beta'_\perp \Delta X_{t-1} \\ Y_{0,t-2} \\ \vdots \\ Y_{0,t-k+1} \\ \beta'_\perp \Delta X_{t-k+1} \\ Y_{0,t-k} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\beta'_0 \\ \vdots \\ -\beta'_0(k-2) \\ 0 \\ -\beta'_0(k-1) \end{pmatrix} + \begin{pmatrix} I & & & & & & \\ & \beta'_\perp & & & & & \\ I & -\beta' & & & & & \\ & & \ddots & & & & \\ I & -\beta' & \cdots & -\beta' & & & \\ & & \ddots & & & \beta'_\perp & \\ I & -\beta' & \cdots & -\beta' & -\beta' & & \end{pmatrix} \begin{pmatrix} Y_{0,t-1} \\ \Delta X_{t-1} \\ \vdots \\ \vdots \\ \Delta X_{t-k+1} \end{pmatrix}.$$

Substitute in (5); one finds  $Y_t = \Psi^\dagger(AU_t + m) + \mu^\dagger D_t + \epsilon_t^\circ = (\Psi^\dagger A)U_t + (\Psi^\dagger m + \mu_0^\dagger) + \mu_t^\circ d_t + \epsilon_t^\circ$ , which is the mixed form. ■

**Proof.** of Theorem 4. 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 (3) and (5) one finds that  $b'(W_t - E(W_t))$  contains  $b'\Psi^\circ U_t$  or  $b'\Psi^\dagger U_t^\dagger$  in addition to an innovation process. Hence  $b'\Psi^{(\cdot)} = 0$ , i.e.  $b \in \text{col}^\perp(\Psi^{(\cdot)})$ . In order  $b$  to be different from the zero vector one must have  $\text{rank}(\Psi^{(\cdot)}) = p - \ell$ , i.e.  $\Psi^{(\cdot)} = \varphi\tau'$ . This completes the proof. ■

**Proof.** of Theorem 5. From (2) one has

$$\begin{aligned} \Delta X_t - m_t^* &= C\epsilon_t + C_0(L)\Delta\epsilon_t = C\epsilon_t + \sum_{i=0}^{\infty} C_{0,i}L^i\epsilon_t - \sum_{i=0}^{\infty} C_{0,i}L^{i+1}\epsilon_t = \\ &= (C + C_{0,0})\epsilon_t + \sum_{i=1}^{\infty} (C_{0,i} - C_{0,i-1})\epsilon_{t-i} \\ &=: \epsilon_t + \sum_{i=1}^{\infty} C_i^*\epsilon_{t-i} =: C^*(L)\epsilon_t, \end{aligned} \tag{27}$$

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

$$C + C_{0,0} = I. \tag{28}$$

There exist a cofeature matrix  $b$  such that  $b'(\Delta X_t - m_t^*) = b'\epsilon_t$  if and only if all the coefficient matrices to the lagged  $\epsilon_t$  in (27) cancel when pre-multiplied by  $b'$ , i.e. iff  $b'C_i^* = 0$ ,  $i = 1, 2, \dots$ . Let  $a_i := b'C_{0,i}$ . The condition  $b'C_i^* = 0$ , for  $i \geq 1$  is  $a_j - a_{j+1} = 0$ ,  $j = 0, 1, \dots$ , with solution  $a_j = a_0$ . From the summability of  $C_0(z)$  for

$|z| < 1 + \varkappa$  and  $\varkappa > 0$ , it follows that  $a_0 = 0$ , i.e.  $b'C_{0,i} = 0$  for all  $i \geq 0$ , which is condition (9). From (28) one has  $C_{0,0} = I - C$ , so that  $0 = b'C_{0,0} = b'(I - C)$ ; this proves (10).

Let (9) hold and assume (10),  $b'C = b'$ . From the definition of  $C$ , see (??), it follows that  $b \in \text{col}(\alpha_\perp)$ , i.e. that  $b = \alpha_\perp u$  for some  $u$ . Substituting into  $b'C = b'$  one finds  $u'(\alpha'_\perp \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} - I_{p_1})\alpha'_\perp = 0$ , which holds iff  $u'(\alpha'_\perp \beta_\perp - \alpha'_\perp \Gamma \beta_\perp) = 0$ , i.e. if  $c$  belongs to  $\mathcal{A} := \text{col}^\perp(\alpha'_\perp (I - \Gamma)\beta_\perp)$ . In order for  $\mathcal{A}$  not to contain only the zero vector,  $\alpha'_\perp (I - \Gamma)\beta_\perp$  must be of deficient rank, i.e.  $\alpha'_\perp (I - \Gamma)\beta_\perp = cd'$  for some full column rank  $p_1 \times p_1 - \ell$  matrices  $c$  and  $d$ . Hence  $u = c_\perp$ . The converse statement is direct. This completes the proof. ■

**Proof.** of Proposition 7. 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  $\epsilon_t$ . Moreover  $V = \text{var}(b'Z_t) = b'_1 C_{W,0} \Omega C'_{W,0} b_1$  is of full rank  $\ell$ . This holds only if  $b_1$  has full column rank  $\ell$ . This completes the proof. ■

**Proof.** of Theorem 8. Sufficiency is proved by substituting  $\varsigma = \varphi\tau'$  in (12) 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  $b'_1 Z_{0t} + b'_2 Z_{2t} = b'_1 u_t$  are the cofeature relations. Pre-multiplication of (12) 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  $b_1 \in \text{col}^\perp(\varsigma)$  and that  $b'_2 = b'_1 \Phi$ . Since  $\ell > 0$  was assumed,  $\varsigma$  must be of deficient rank,  $\varsigma = \varphi\tau'$ , and  $b_1 = \varphi_\perp$ . ■

**Proof.** of Theorem 9. We wish to show  $U_{2t}$  can be obtained linearly from  $U_{1t}$  and vice versa. In fact

$$\begin{pmatrix} Y_{0t} \\ \beta'_\perp \Delta X_t \\ Y_{0,t-1} \end{pmatrix} = \begin{pmatrix} \beta'_0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \beta' & I_{p_0} \\ \beta'_\perp & I_{p_0} \end{pmatrix} \begin{pmatrix} \Delta X_t \\ Y_{0,t-1} \end{pmatrix},$$

and conversely

$$\begin{pmatrix} \Delta X_t \\ Y_{0,t-1} \end{pmatrix} = \begin{pmatrix} -\bar{\beta}\beta'_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\beta} & \bar{\beta}_\perp & -\bar{\beta} \\ & & I_{p_0} \end{pmatrix} \begin{pmatrix} Y_{0t} \\ \beta'_\perp \Delta X_t \\ Y_{0,t-1} \end{pmatrix}.$$

Hence  $b'(U_{2t} - E(U_{2t})) = b'A(U_{1t} - E(U_{1t}))$ , and if  $b$  is the cofeature matrix for  $U_{2t}$  then  $A'b$  is the corresponding cofeature matrix for  $U_{1t}$  and vice versa. Hence the cofeature rank is the same and the cofeature matrices are equivalent, being linked by the same linear transformation associated with the transformation from  $U_{1t}$  to  $U_{2t}$ . This shows that the cofeature properties of  $U_{1t}$  and  $U_{2t}$  are equivalent.

We next show that the eigenvalues of  $\text{RRR}(\Delta X_t, Z_{1t}|Z_{2t}, 1)$  and of  $\text{RRR}(Y_t, Z_{1t}|Z_{2t}, 1)$  are identical. It is well known that the eigenvalues of (15) are invariant to nonsingular linear transformations of  $Z_{0t}$ . Hence one can substitute  $\Delta X_t$  with  $B'\Delta X_t$  without affecting the eigenvalues, where we choose  $B := (\beta : \beta_\perp)$ , which is of full rank. Next note that adding  $GZ_{2t} + g$  to  $Z_{0t}$  does not affect the eigenvalues either, because of the least squares correction by  $Z_{2t}$  and 1. Hence one can modify  $B'\Delta X_t$  by adding  $(Y'_{0,t-1} + \beta_0 : 0)'$  without changing the eigenvalues, because  $Z_{2t}$  contains  $Y_{0,t-1}$ . Note that by this choice one obtains  $B'\Delta X_t + (Y'_{0,t-1} + \beta_0 : 0)' = Y_t$  as dependent variables. This completes the proof. ■

## Appendix B: inference

In the first part of this appendix we assume that the cointegration parameters  $(\beta' : \beta'_0)$  are known. In the second part of the appendix we 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

$$\begin{aligned} Z_{0t} &= \varsigma Z_{1t} + \Phi Z_{2t} + \mu_0^* + \epsilon_t^*, \\ \varsigma &= \varphi \tau'. \end{aligned} \quad (29)$$

for an appropriate definition of  $Z_{0t}$ ,  $Z_{1t}$ ,  $Z_{2t}$ . The coefficient matrix  $\varsigma$  is  $p \times j$  of reduced rank  $s$ , and the matrices  $\varphi$  and  $\tau$  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  $Z_t = AZ_{t-1} + \epsilon_t^\dagger$ ,  $\epsilon_t^\dagger := (\tilde{\epsilon}_t' : 0)'$ , where the eigenvalues of  $A$  are all inside the unit disk. For details of these companion forms in the I(1) case, see e.g. Omtzigt and Paruolo (2002).

Let  $x_t := (Z'_{1t} : Z'_{2t})'$ ,  $y_t := (Z'_{0t} : x'_t)'$ ,  $\Sigma^* := E((y_t - E(y_t))(y_t - E(y_t))')$ , and indicate blocks of  $\Sigma^*$  with the subscripts 0, 1, 2,  $x$ . Recall also that  $Z_{2t} := (R'_t : d'_t)'$ , and let  $\Sigma_{ij}^* := \Sigma_{ij}^* - \Sigma_{ih}^* \Sigma_{hh}^{*-1} \Sigma_{hj}^*$  for  $i, j, h = 0, 1, 2, x, R, d$  and in particular  $\Sigma_{ij} := \Sigma_{ij,2}^*$  for  $i, j = 0, 1, x$ . Similarly we use the notation  $S_{ij} := M_{ij} - M_{i2} M_{22}^{-1} M_{2j}$ ,  $M_{ij} := T^{-1} \sum_{t=1}^T (Z_{it} - m_i)(Z_{jt} - m_j)'$ ,  $m_i := T^{-1} \sum_{t=1}^T Z_{it}$ ,  $i, j = 0, 1, 2, x, R, d, \epsilon$ , where  $Z_{jt}$  is substituted by  $R_t, d_t, \epsilon_t^*$  in the last three cases.

Let  $\hat{Z}_{it}$  (respectively  $\hat{S}_{ij}$ ) indicates  $Z_{it}$  (respectively  $S_{ij}$ ) calculated at the estimated values of the cointegration coefficients. Similarly  $\hat{S}(\hat{\lambda}) := \hat{\lambda} \hat{S}_{00} - \hat{S}_{01} \hat{S}_{11}^{-1} \hat{S}_{10}$ . We indicate by  $\hat{Q}_i, \hat{J}_i$  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  $\hat{\cdot}$ . We distinguish the estimators based on  $\hat{S}_{ij}$  in place of  $S_{ij}$  with a double hat,  $\hat{\hat{\cdot}}$ .

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 11** *The following convergences hold for  $T \rightarrow \infty$ :*

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

We next give the proof of Theorem 10.

**Proof.** of Theorem 10. Note that

$$(\hat{b} - b)' = (\hat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})' P_a(I : \Phi_1) + \varphi'_{\perp a_\perp} (0 : (\hat{\Phi}_1 - \Phi_1)) + (\hat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})' (0 : (\hat{\Phi}_1 - \Phi_1)). \quad (30)$$

where  $P_a$  appears because  $(\hat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})' P_{a_\perp} = 0$  by normalization. From proposition 12 in Paruolo (2003) one has that  $(\hat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})$  is  $T^{1/2}$  consistent with asymptotic representation

$$(\hat{\varphi}_{\perp a_\perp} - \varphi_{\perp a_\perp})' \bar{a} = -\varphi'_{\perp a_\perp} S_{\epsilon 1} \varsigma' a (a' \varsigma \Sigma_{11}^{-1} \varsigma' a)^{-1} + o_p(T^{-1/2}).$$

From the definition of  $\widehat{\Phi}_1$  in (16) it is simple to show that  $\widehat{\Phi}_1 - \Phi_1 = O_p(T^{-1/2})$ , so that the other leading term in (30) is  $\varphi'_{\perp a_{\perp}}(0 : (\widehat{\Phi}_1 - \Phi_1))$ , where one finds

$$\begin{aligned}\varphi'_{\perp a_{\perp}}(\widehat{\Phi}_1 - \Phi_1) &= \varphi'_{\perp a_{\perp}} M_{\epsilon R.d} M_{RR.d}^{-1} + \varphi'_{\perp a_{\perp}} (\widehat{\varphi} - \varphi) \tau' M_{1R.d} M_{RR.d}^{-1} + o_p(T^{-1/2}) \\ &= \varphi'_{\perp a_{\perp}} M_{\epsilon R.d} M_{RR.d}^{-1} - (\widehat{\varphi}_{\perp a_{\perp}} - \varphi_{\perp a_{\perp}})' P_a \varphi \tau' M_{1R.d} M_{RR.d}^{-1} + o_p(T^{-1/2})\end{aligned}$$

where we have applied the identities (4.2) (4.3) in Paruolo (1997) in the second line. Collecting terms in (30) one finds

$$\begin{aligned}(\widehat{b} - b)' &= (\widehat{\varphi}_{\perp a_{\perp}} - \varphi_{\perp a_{\perp}})' P_a ((I : \Phi_1 - \varphi \tau' M_{1R.d} M_{RR.d}^{-1}) + \\ &\quad + \varphi'_{\perp a_{\perp}} (0 : M_{\epsilon R.d} M_{RR.d}^{-1}) + o_p(T^{-1/2})) \\ &= \varphi'_{\perp a_{\perp}} S_{\epsilon 1} \zeta' a (a' \zeta \Sigma_{11}^{-1} \zeta' a)^{-1} a' (-I : -\Phi_1 + \varphi \tau' \Sigma_{1R.d}^* \Sigma_{RR.d}^{*-1}) + \\ &\quad + \varphi'_{\perp a_{\perp}} M_{\epsilon R.d} (0 : M_{RR.d}^{-1}) + o_p(T^{-1/2})\end{aligned}$$

Finally note that  $S_{\epsilon 1} = M_{\epsilon 1.2} = M_{\epsilon x} \mathcal{B}_1 + o_p(T^{-1/2})$  and  $M_{\epsilon R.d} = M_{\epsilon x} \mathcal{A}_3 + o_p(T^{-1/2})$ . The result then follows from Lemma 11. ■

The proof that  $Q_i \xrightarrow{d} \chi^2(df_{Q_i})$  is given in Paruolo (2003) for  $i = 1, 2, 3, 4$ . The proof that  $J_i \xrightarrow{d} \chi^2(df_{J_i})$  follows from Theorem 10 for  $i = 1$ , while for  $i = 2$  it follows from the standard limit distribution of the OLS estimator.

We next indicate by  $\eta$  the parameter vector in model  $H(s)$ . We next want to show that  $Q_i - \widehat{Q}_i = o_p(1)$ ,  $J_i - \widehat{J}_i = o_p(1)$ ,  $T^{1/2}(\widehat{\eta} - \eta) = o_p(1)$ . 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  $Q_i^{\infty}$ . 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 12** *If*

$$\widehat{S}_{01} = S_{01} + o_p(T^{-1/2}) \quad \widehat{S}_{ii} = S_{ii} + o_p(1) \quad i = 0, 1 \quad (31)$$

*then the tests  $Q_i$  (respectively  $J_i$ ) and  $\widehat{Q}_i$  (respectively  $\widehat{J}_i$ ) are equivalent, and the estimators  $\widehat{\eta}$  and  $\widehat{\eta}$  are equivalent. The following are sufficient conditions to verify (31):*

$$\widehat{M}_{ij} = M_{ij} + o_p(T^{-1/2}), \quad i = 0, 1, 2 \text{ and } j = 1, 2. \quad (32)$$

**Proof.** It is simple to see that under (31) Lemma 11 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 (32) 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  $a := M_{i2}$ ,  $b := M_{22}^{-1}$ ,  $c := M_{21}$ . It is easy to see that

$$\begin{aligned}\widehat{a}\widehat{b}\widehat{c} - abc &= (\widehat{a} - a)bc + a(\widehat{b} - b)c + ab(\widehat{c} - c) + (\widehat{a} - a)b(\widehat{c} - c) + \\ &\quad + (\widehat{a} - a)(\widehat{b} - b)c + a(\widehat{b} - b)(\widehat{c} - c) + \\ &\quad + (\widehat{a} - a)(\widehat{b} - b)(\widehat{c} - c)\end{aligned} \quad (33)$$

such that when  $(\widehat{a} - a)$ ,  $(\widehat{b} - b)$ ,  $(\widehat{c} - c)$  are  $o_p(T^{-1/2})$ , so is  $\widehat{abc} - abc = o_p(T^{-1/2})$ . Finally  $\widehat{M}_{ij} - M_{ij} = o_p(T^{-1/2})$ . Thus the conditions (31) 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 12 applies.

**Proposition 13** *For the I(1) case  $\widehat{M}_{ij} - M_{ij} = O_p(T^{-1})$ ,  $i, j = 0, 1, 2$ .*

**Proof.** Let  $\beta^* := (\beta' : \beta_0)'$  and  $X_{t-1}^* = (X'_{t-1} : t)'$ , so that the ECM term is  $\beta^{*'} X_{t-1}^*$ , which is stationary by Granger's representation theorem. Let  $\widehat{Z}_{jt} := ((\widehat{\beta}^{*'} X_{t-1}^*)' : v_t)$ , where  $v_t$  indicates other variables. For  $i \neq j$  one has  $\widehat{M}_{ij} - M_{ij} = (M_{iX^*}(\widehat{\beta}^* - \beta^*) : 0)$ , where a subscript  $X^*$  indicates  $X_{t-1}^*$ . Let

$$H_T := \begin{pmatrix} \beta & \beta_{\perp} & 0 \\ \beta_0 & 0 & T^{1/2} \end{pmatrix}' \quad H_T^{-1} := \begin{pmatrix} \bar{\beta} & \bar{\beta}_{\perp} & T^{-1/2} \bar{\beta} \beta_0' \\ 0 & 0 & T^{-1/2} \end{pmatrix}'$$

and recall that  $H_T(\widehat{\beta}^* - \beta^*) = O_p(T^{-1})$  while  $M_{iX^*} H_T^{-1} = O_p(1)$  because  $Z_{it}$  is I(0) and  $H_T^{-1'} X_{t-1}^*$  is normalized as an I(1) process. Hence, inserting  $H_T^{-1} H_T = I_{p+1}$  one finds  $M_{iX^*} H_T^{-1} H_T(\widehat{\beta}^* - \beta^*) = O_p(1) O_p(T^{-1})$  and hence  $\widehat{M}_{ij} - M_{ij} = O_p(T^{-1})$ .

Consider next  $\widehat{M}_{ii} - M_{ii} = a + a' + b$  where  $a := (\beta^{*'} M_{X^* X^*} (\widehat{\beta}^* - \beta^*) : 0)$  and  $b = \text{diag}((\widehat{\beta}^* - \beta^*)' M_{X^* X^*} (\widehat{\beta}^* - \beta^*), 0)$ . As before  $a = \beta^{*'} M_{X^* X^*} H_T^{-1} H_T(\widehat{\beta}^* - \beta^*) = O_p(1) O_p(T^{-1}) = O_p(T^{-1})$ . Note that  $H_T^{-1'} M_{X^* X^*} H_T^{-1} = O_p(T)$ , and hence  $b = (\widehat{\beta}^* - \beta^*)' H_T' H_T^{-1'} M_{X^* X^*} H_T^{-1} H_T(\widehat{\beta}^* - \beta^*) = O_p(T^{-1}) O_p(T) O_p(T^{-1}) = O_p(T^{-1})$  and hence  $\widehat{M}_{ii} - M_{ii} = O_p(T^{-1})$ . This completes the proof. ■