econometrics

Article

Likelihood Ratio Tests of Restrictions on Common Trends Loading Matrices in I(2) VAR Systems

H. Peter Boswijk [1] and Paolo Paruolo [2,*]

1 Amsterdam School of Economics and Tinbergen Institute, University of Amsterdam, 1001 NJ Amsterdam, The Netherlands; h.p.boswijk@uva.nl
2 Joint Research Centre, European Commission, 21027 Ispra (VA), Italy
* Correspondence: paolo.paruolo@ec.europa.eu

Academic Editor: Katarina Juselius
Received: 28 February 2017; Accepted: 21 June 2017; Published: 29 June 2017

Abstract: Likelihood ratio tests of over-identifying restrictions on the common trends loading matrices in I(2) VAR systems are discussed. It is shown how hypotheses on the common trends loading matrices can be translated into hypotheses on the cointegration parameters. Algorithms for (constrained) maximum likelihood estimation are presented, and asymptotic properties sketched. The techniques are illustrated using the analysis of the PPP and UIP between Switzerland and the US.

Keywords: cointegration; common trends; identification; VAR; I(2)

JEL Classification: C32

1. Introduction

The duality between the common trends representation and the vector equilibrium-correction model-form (VECM) in cointegrated systems allows researchers to formulate hypotheses of economic interest on any of the two. The VECM is centered on the adjustment with respect to disequilibria in the system; in this way it facilitates the interpretation of cointegrating relations as (deviations from) equilibria.

The common trends representation instead highlights how variables in the system as pushed around by common stochastic trends, which are often interpreted as the main persistent economic factors influencing the long-term. Both representations provide economic insights on the economic system under scrutiny. Examples of both perspectives are given in Juselius (2017a, 2017b)

The common trends and VECM representations are connected through representation results such as the Granger Representation Theorem, in the case of I(1) systems, see Engle and Granger (1987) and Johansen (1991), and the Johansen Representation Theorem, for the case of I(2) systems, see Johansen (1992). In particular, both representation theorems show that the loading matrix of the common stochastic trends of highest order is a basis of the orthogonal complement of the matrix of cointegrating relations. Because of this property, these two matrices are linked, and any one of them can be written as a function of the other one.

This paper focuses on I(2) vector autoregressive (VAR) systems, and it considers the situation where (possibly over-identifying) economic hypotheses are entertained for the factor loading matrix of the I(2) trends. It is shown how they can then be translated into hypotheses on the cointegrating relations, which appear in the VECM representation; the latter forms the basis for maximum likelihood (ML) estimation of I(2) VAR models. In this way, constrained ML estimators are obtained and the associated likelihood ratio (LR) tests of these hypotheses can be defined. These tests are discussed in the present paper; Wald tests on just-identified loading matrices of the I(1) and I(2) common trends have already been proposed by Paruolo (1997, 2002).

The running example of the paper is taken from Juselius and Assenmacher (2015), which is the working paper version of Juselius and Assenmacher (2017). The following notation is used: for a full column-rank matrix a, $\operatorname{col} a$ denotes the space spanned by the columns of a and $a_\perp$ indicates a basis of the orthogonal complement of $\operatorname{col} a$. For a matrix b of the same dimensions of a, and for which $b'a$ is full rank, let $b_a := b(a'b)^{-1}$; a special case is when $a = b$, for which $\bar{a} := a_a = a(a'a)^{-1}$. Let also $P_a := a(a'a)^{-1}a'$ indicate the orthogonal projection matrix onto $\operatorname{col} a$, and let the matrix $P_{a_\perp} = I - P_a$ denote the orthogonal projection matrix on its orthogonal complement. Finally e_j is used to indicate the j-th column of an identity matrix of appropriate dimension.

The rest of this paper is organized as follows: Section 2 contains the motivation and the definition of the problem considered in the paper. The identification of the I(2) common trends loading matrix under linear restrictions is analysed in Section 3. The relationship between the identified parametrization of I(2) common trends loading matrix and an identified version of the cointegration matrix is also discussed. Section 4 considers a parametrization of the VECM, and discusses its identification. ML estimation of this model is discussed in Section 5; the asymptotic distributions of the resulting ML estimator of the I(2) loading matrix and the LR statistic of the over-identifying restrictions are sketched in Section 6. Section 7 reports an illustration of the techniques developed in the paper on a system of US and Swiss prices, interest rates and exchange rate. Section 8 concludes, while two appendices report additional technical material.

2. Common Trends Representation for I(2) Systems

This section introduces quantities of interest and presents the motivation of the paper. Consider a p-variate VAR(k) process X_t:

$$X_t = A_1 X_{t-1} + \ldots + A_k X_{t-k} + \mu_0 + \mu_1 t + \varepsilon_t, \tag{1}$$

where $A_i, i = 1, \ldots, k$ are $p \times p$ matrices, μ_0 and μ_1 are $p \times 1$ vectors, and ε_t is a $p \times 1$ i.i.d. $N(0, \Omega)$ vector, with Ω positive definite. Under the conditions of the Johansen Representation Theorem, see Appendix A, called *the* I(2) *conditions*, X_t admits a common trends I(2) representation of the form

$$X_t = C_2 S_{2t} + C_1 S_{1t} + Y_t + v_0 + v_1 t, \tag{2}$$

where $S_{2t} := \sum_{i=1}^{t} \sum_{s=1}^{i} \varepsilon_s$ are the I(2) stochastic trends (cumulated random walks), $S_{1t} := \Delta S_{2t} = \sum_{i=1}^{t} \varepsilon_i$ is a random walk component, and Y_t is an I(0) linear process.

Cointegration occurs when the matrix C_2 has reduced rank $r_2 < p$, such that $C_2 = ab'$, where a and b are $p \times r_2$ and of full column rank. This observation lends itself to the following interpretation: $b'S_{2t}$ defines the r_2 common I(2) trends, while a acts as the loading matrix of X_t on the I(2) trends. The reduced rank of C_2 implies that there exist $m := p - r_2$ linearly independent cointegrating vectors, collected in a $p \times m$ matrix τ, satisfying $\tau' C_2 = 0$; hence $\tau' X_t$ is I(1). Combining this with $C_2 = ab'$, it is clear that $a = \tau_\perp$, i.e., the columns of the loading matrix span the orthogonal complement of the cointegration space $\operatorname{col} \tau$. Interest in this paper is on hypotheses on $a = \tau_\perp$[1].

Observe that $C_2 = ab'$ is invariant to the choice of basis of either $\operatorname{col} a$ and $\operatorname{col} b$. In fact, (a, b) can be replaced by (aQ, bQ'^{-1}) with Q square and nonsingular without affecting C_2. One way to resolve this identification problem is to impose restrictions on the entries of $a = \tau_\perp$; enough restrictions of this kind would make the choice of $\tau_\perp$ unique. Such an approach to identification is common in confirmatory factor analysis in the statistics literature, see Jöreskog et al. (2016).

If more restrictions are imposed than needed for identification, they are over-identifying. Such over-identifying restrictions on $\tau_\perp$ usually correspond to (similarly over-identifying) restrictions

[1] In the I(2) cointegration literature, $\tau_\perp$ is also referred to as β_2, see the Johansen Representation Theorem in Appendix A.

on τ, see Section 3 below. Although economic hypotheses may directly imply restrictions on the cointegrating vectors in τ, in some cases it is more natural to formulate restrictions on the I(2) loading matrix $\tau_\perp$. This is illustrated by the two following examples.

2.1. Example 1

Kongsted (2005) considers a model for $X_t = (m_t : y_t^n : p_t)'$, where m_t, y_t^n and p_t denote the nominal money stock, nominal income and the price level, respectively (all variables in logs); here ':' indicates horizontal concatenation. He assumes that the system is I(2), with $r_2 = 1$. Given the definition of the variables, Kongsted (2005) considers the natural question of whether real money $m_t - p_t$ and real income $y_t^n - p_t$ are at most I(1). This corresponds to an (over-identified) cointegrating matrix τ and loading vector $\tau_\perp$ of the form

$$\tau = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \qquad \tau_\perp = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The form of τ corresponds to the fact that the I(1) linear combinations $\tau' X_t$ are (linear combinations of) $((m_t - p_t) : (y_t^n - p_t))'$, as required. On the other hand, the restriction on $\tau_\perp$ says that each of the three series have exactly the same I(2) trend, with the same scale factor. Both formulations are easily interpretable.

Note that the hypothesis on $\tau_\perp$ involves two over-identifying restrictions (the second and third component are equal to the first component), in addition to a normalization (the first component equals 1). Similarly, the restriction that the matrix consisting of the first two rows of τ equals I_2 is a normalization; the two over-identifying restrictions are that the entries in both columns sum to 0.

As this first example shows, knowing τ is the same as knowing $\tau_\perp$ and vice versa[2].

2.2. Example 2

Juselius and Assenmacher (2015) consider a 7-dimensional VAR with $X_t = (p_{1t} : p_{2t} : e_{12t} : b_{1t} : b_{2t} : s_{1t} : s_{2t})'$ with $r_2 = 2$, where p_{it}, b_{it}, s_{it} are the (log of) the price index, the long and the short interest rate of country i at time t respectively, and e_{12t} is the log of the exchange rate between country 1 (Switzerland) and 2 (the US) at time t. They expect the common trends representation to have a loading matrix $\tau_\perp$ of the form:

$$\tau_\perp = \begin{pmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \\ \phi_{31} & \phi_{32} \\ 0 & \phi_{42} \\ 0 & \phi_{52} \\ 0 & \phi_{62} \\ 0 & \phi_{72} \end{pmatrix}. \tag{3}$$

where ϕ_{ij} indicates an entry not restricted to 0.

The second I(2) trend is loaded on the interest rates b_{1t}, b_{2t}, s_{1t}, s_{2t}, as well as on US prices p_{2t} and the exchange rate e_{12t}; this can be interpreted as a financial (or 'speculative') trend affecting world prices. The first I(2) trend, instead, is only loaded on p_{1t}, p_{2t}, e_{12t} and embodies a 'relative price' I(2) trend; it can be interpreted as the Swiss contribution to the trend in prices.

The cointegrating matrix τ in this example is of dimension 7×5. It is not obvious what type of restrictions on τ correspond to the structure in (3). However, it is τ rather than $\tau_\perp$ that enters the likelihood function (as will be analyzed in Section 4). The rest of the paper shows that the restrictions

---

2 Up to normalizations, see below.

in (3) are over-identifying, how they can be translated into hypotheses on τ, and how they can be tested via LR tests.

3. Hypothesis on the Common Trends Loadings

This section discusses linear hypotheses on $\tau_\perp$ and their relation to τ. First, attention is focused on the case of linear hypotheses on the normalized version $\tau_{\perp c_\perp} := \tau_\perp \left(c'_\perp \tau_\perp\right)^{-1}$ of $\tau_\perp$. Here $c_\perp$ is a full-column-rank matrix of the same dimension of $\tau_\perp$ such that $c'_\perp \tau_\perp$ is square and nonsingular[3]. This normalization was introduced by Johansen (1991) in the context of the I(1) model in order to isolate the (just-) identified parameters in the cointegration matrix.

Later, linear hypotheses formulated directly on $\tau_\perp$ are discussed. The main result of this section is the fact that the parameters of interest appears linearly both in $\tau_{\perp c_\perp}$ *and* in τ_c in the first case; this is not necessarily true in the second case.

The central relation employed in this section (for both cases), is the following identity:

$$\tau_c := \tau \left(c'\tau\right)^{-1} = (I - c_\perp \left(\tau'_\perp c_\perp\right)^{-1} \tau'_\perp)\bar{c} = (I - c_\perp \tau'_{\perp c_\perp})\bar{c}, \tag{4}$$

where $\bar{c} := c(c'c)^{-1}$. This identity readily follows from the oblique projections identity

$$I = \tau \left(c'\tau\right)^{-1} c' + c_\perp \left(\tau'_\perp c_\perp\right)^{-1} \tau'_\perp,$$

see e.g. Srivastava and Kathri (1979, p. 19), by post-multiplication by $\bar{c}$.

3.1. Linear hypotheses on $\tau_{\perp c_\perp}$

Johansen (1991) noted that the function $a_b := a(b'a)^{-1}$ is invariant with respect to the choice of basis of the space spanned by a. in fact, consider in the present context any alternative basis $\tau^\star_\perp$ of the space spanned by $\tau_\perp$; this has representation $\tau^\star_\perp = \tau_\perp Q$ for Q square and full rank. Inserting $\tau^\star_\perp$ in place of $\tau_\perp$ in the definition of $\tau_{\perp c_\perp} := \tau_\perp \left(c'_\perp \tau_\perp\right)^{-1}$, one finds

$$\tau^\star_{\perp c_\perp} = \tau^\star_\perp \left(c'_\perp \tau^\star_\perp\right)^{-1} = \tau_\perp Q \left(c'_\perp \tau_\perp Q\right)^{-1} = \tau_{\perp c_\perp}.$$

Hence $\tau_{\perp c_\perp}$, similarly to the cointegration matrix in the I(1) model in Johansen (1991), is (just-)identified.

To facilitate stating hypotheses on the unconstrained elements of $\tau_{\perp c_\perp}$, the following representation of $\tau_{\perp c_\perp}$ appears useful:

$$\tau_{\perp c_\perp} = \bar{c}_\perp + c\,\vartheta \tag{5}$$

where ϑ is an $m \times r_2$ matrix of free coefficients in $\tau_\perp$[4]. For example, one may have

$$c_\perp = \begin{pmatrix} 0_{3\times 2} \\ I_2 \end{pmatrix}, \qquad c = \begin{pmatrix} I_3 \\ 0_{2\times 3} \end{pmatrix}, \qquad \tau_{\perp c_\perp} = \bar{c}_\perp + c \begin{pmatrix} \vartheta_{11} & \vartheta_{12} \\ \vartheta_{21} & \vartheta_{22} \\ \vartheta_{31} & \vartheta_{32} \end{pmatrix} = \begin{pmatrix} \vartheta_{11} & \vartheta_{12} \\ \vartheta_{21} & \vartheta_{22} \\ \vartheta_{31} & \vartheta_{32} \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6}$$

with $p = 5$, $m = 3$, $r_2 = 2$.

---

3 When $c'_\perp \tau_\perp$ is square and nonsingular, then one can prove that also $c'\tau$ is square and nonsingular, see e.g., Johansen (1996, Exercise 3.7).

4 This equation is obtained by using orthogonal projection of $\tau_{\perp c_\perp}$ on the columns spaces of c and $c_\perp$, and applying the equality $c'_\perp \tau_{\perp c_\perp} = I_{r_2}$ which follows by definition.

Consider over-identifying linear restrictions on the columns of ϑ in (5). Typically, such restrictions will come in the form of zero (exclusion) restrictions or unit restrictions, where the latter would indicate equal loadings of a specific variable and the variable on which the column of $\tau_{\perp c_\perp}$ has been normalized. The general formulation of such restrictions is

$$\vartheta_i = k_i + K_i \phi_i, \quad i = 1, \ldots, r_2, \tag{7}$$

where ϑ_i is the i-th column vector of ϑ, k_i and K_i are conformable vectors and matrices, and ϕ_i contains the remaining unknown parameters in ϑ_i. If only zero restrictions are imposed, then $k_i = 0_m$.

The formulation in (7) includes several notable special cases. For instance, if all $K_i = K$ and $k_i = 0_m$, one obtains the hypothesis that ϑ is contained in a given linear space, $\vartheta = K\phi$. Another example is given by the case where one column ϑ_1 is known, $\vartheta = (k_1 : \phi)$; this corresponds to the choice $\vartheta_1 = k_1$ with K_1 and ϕ_1 void and $k_2 = \ldots = k_{r_2} = 0$, $K_2 = \ldots = K_{r_2} = I$.

The restrictions in (7) may be summarized as

$$\text{vec } \vartheta = k + K\phi, \tag{8}$$

where $k = (k_1' : \ldots : k_{r_2}')'$, $K = \text{blkdiag}(K_1, \ldots, K_{r_2})$ and $\phi = (\phi_1' : \ldots : \phi_{r_2}')'$. Here $\text{blkdiag}(B_1, B_2, \ldots, B_n)$ indicates a matrix with the (not necessarily square) blocks $B_1, B_2, \ldots, B_n$ along the main diagonal. Formulation (8) generalises (7).

The main result of this section is stated in the next theorem.

Theorem 1 (Hypotheses on $\tau_{\perp c_\perp}$)**.** *Assume that ϑ satisfies linear restrictions of the type (8); then these restrictions are translated into a linear hypothesis on* vec τ_c *via*

$$\text{vec } \tau_c = (\text{vec } \bar{c} - (I_m \otimes c_\perp)\mathcal{K}_{m,r_2} k) - (I_m \otimes c_\perp)\mathcal{K}_{m,r_2} K\phi, \tag{9}$$

where $\mathcal{K}_{m,n}$ is the commutation matrix satisfying $\mathcal{K}_{m,n}$ vec A = vec A', with A of dimensions $m \times n$, see Magnus and Neudecker (2007).

Proof. Substitute (8) into (4) and vectorize using standard properties of the vec operator, see Magnus and Neudecker (2007). □

The previous theorem shows that, when one can express a linear hypothesis on the coefficients in ϑ that are unrestricted in $\tau_{\perp c_\perp}$, then the same linear hypothesis is translated into a restriction on vec τ_c. Note that the proof simply exploits (4).

Identification of the restricted coefficients ϕ under these hypothesis can be addressed in a straightforward way. In fact, the parameters in ϑ are identified; hence ϕ is identified provided that the matrix K is of full column rank, which in turn will imply that the Jacobian matrix $\partial \text{vec } \tau_c / \partial \phi' = -(I_m \otimes c_\perp)\mathcal{K}_{m,r_2} K$ in (9) has full column rank.

Because, in practice, econometricians may explore the form of $\tau_\perp$ via unrestricted estimates of $\tau_{\perp c_\perp}$, see Paruolo (2002), before formulating restrictions on $\tau_\perp$, using hypothesis on the unrestricted coefficients in $\tau_{\perp c_\perp}$ appears a natural sequential step.

The next subsection discusses the alternative approach of specifying hypotheses directly on $\tau_\perp$.

3.2. Linear Hypotheses on $\tau_\perp$

In case placing restrictions on the unrestricted coefficients in $\tau_{\perp c_\perp}$ is not what the econometrician wants, this subsection considers linear hypothesis on $\tau_\perp$ directly. It is shown that sometimes it is possible to translate linear hypothesis on $\tau_\perp$ into linear hypothesis on $\tau_{\perp c_\perp}$ for some $c_\perp$. It is also shown that this is always possible for $r_2 = 2$, for which a constructive proof is provided.

Analogously to (7), consider linear hypotheses on the columns of $\tau_\perp$, of the following type:

$$\tau_{\perp,i} = h_i + H_i\phi_i, \quad i = 1, \dots, r_2, \tag{10}$$

summarized as

$$\text{vec } \tau_\perp = h + H\phi. \tag{11}$$

In this case, non-zero vectors h_i represent normalizations of the columns of the loading matrix, and as before, ϕ_i collects the unknown parameters in $\tau_{\perp,i}$.

Theorem 2 (Hypotheses on $\tau_\perp$). *Assume that $\tau_\perp = \tau_\perp(\phi)$ satisfies linear restrictions of the type (11), then these restrictions are translated in general into a* non-linear *hypothesis on* vec τ_c *via*

$$\tau_c = (I - c_\perp \left(\tau_\perp(\phi)' c_\perp\right)^{-1} \tau_\perp(\phi)')\bar{c} \tag{12}$$

and the Jacobian of the transformation from ϕ to vec τ_c *is*

$$\mathcal{J}(\cdot) := \frac{\partial \text{ vec } \tau_c(\cdot)}{\partial \phi'} = -(\tau_c(\cdot)' \otimes c_\perp(\tau_\perp(\cdot)' c_\perp)^{-1})\mathcal{K}_{p,r_2} H. \tag{13}$$

This parametrization is smooth on an open set in the parameter space Φ of ϕ where $c_\perp' \tau_\perp$ is of full rank.

Proof. Equation (12) is a re-statement of (4). Differentiation of (12) delivers (13). □

One can note that the Jacobian matrix in (13) can be used to check local identification using the results in Rothenberg (1971).

The result of Theorem 2 is in contrast with the result of Theorem 1, because the latter delivers a linear hypothesis for τ_c while Theorem 2 gives in general non-linear restrictions on τ_c. One may hence ask the following question: when is it possible to reduce the more general linear hypothesis on $\tau_\perp$ given in (11) to the simpler linear hypothesis on ϑ given in (8)?

In the special case of $r_2 = 2$, the following theorem states that this can be always obtained. This applies for instance to the motivating example (3), where one can choose some $c_\perp$ so that $\tau_\perp' c_\perp$ is equal to the identity, as shown below. Consider the formulation (10) with $r_2 = 2$, and assume that no normalizations have been imposed yet, such that $h_1 = h_2 = 0$. It is assumed that $\tau_\perp$, under the equation-by-equation restrictions, satisfies the usual rank conditions for identification, see Johansen (1995, Theorem 1) :

$$\text{rank } R_i' \tau_\perp = 1 \qquad \text{for} \qquad i = 1, 2, \tag{14}$$

where $R_i = H_{i,\perp}$.

Theorem 3 (Case $r_2 = 2$). *Let $\tau_\perp$ obey the restrictions $\tau_\perp = (H_1\phi_1 : H_2\phi_2)$ satisfying the rank conditions (14); then one can choose normalization conditions on ϕ_1 and ϕ_2 so that there exists a matrix $c_\perp$ such that $c_\perp' \tau_\perp = I$. This implies that a hypotheses on $\tau_\perp$ can be stated in terms of ϑ in (5), and, by Theorem 1, a linear hypotheses on* vec ϑ *corresponds to linear hypothesis on* vec τ_c.

Proof. Because $R_1' \tau_\perp = (0 : R_1' H_2 \phi_2)$ has rank 1, one can select (at least) one linear combination of R_1, $R_1 a_1$ say, so that ϕ_2 is normalized to be one in the direction $b_2' := a_1' R_1' H_2$, i.e., $b_2' \phi_2 = 1$. Similarly, $R_2' \tau_\perp = (R_2' H_1 \phi_1 : 0)$ has rank 1, and one can select (at least) one linear combination of R_2, $R_2 a_2$ say, so that ϕ_1 is normalized to be one in the direction $b_1' := a_2' R_2' H_1$, i.e., $b_1' \phi_1 = 1$. Next define $c_\perp = (R_2 a_2 : R_1 a_1)$ which by construction satisfies $c_\perp' \tau_\perp = I_2$. □

The proof of the previous theorem provides a way to construct $c_\perp$ when $r_2 = 2$ and the usual rank condition for identification (14) holds. The rest of the paper focuses attention on the case of linear restrictions on ϑ in (8), which can be translated linearly into restrictions on τ_c as shown in Theorem 1.

3.3. Example 2 Continued

Consider (3); this hypothesis is of type $\tau_\perp = (H_1\phi_1 : H_2\phi_2)$ with

$$H_1 = \begin{pmatrix} I_3 \\ 0_{4\times 3} \end{pmatrix}, \qquad H_2 = \begin{pmatrix} 0_{1\times 6} \\ I_6 \end{pmatrix},$$

and hence $R_1' = (I_4 : 0_{4\times 3})$ and $R_2' = (I_6 : 0_{6\times 1})$. In this case one can define $c = (e_2 : e_3 : e_5 : e_6 : e_7)$ and $c_\perp = (e_1 : e_4)$ where e_j is the j-th column of I_7.

It is simple to verify that, under the additional normalization restrictions $\phi_{11} = 1$ and $\phi_{42} = 1$, $\tau_\perp$ in (3) satisfies $c_\perp' \tau_\perp = I_2$. Therefore, define $\tau_{\perp c_\perp}$ as (3) under these normalization restrictions. Using formula (4) one can see that

$$\tau_c = (I - c_\perp \tau_{\perp c_\perp}')\bar{c} = \begin{pmatrix} -\phi_{21} & -\phi_{31} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\phi_{22} & -\phi_{32} & -\phi_{52} & -\phi_{62} & -\phi_{72} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{15}$$

so that vec τ_c is linear in ϕ, as predicted by Theorem 3.

4. The VECM Parametrization

This section describes the I(2) parametrization employed in the statistical analysis of the paper. Consider the following τ-parametrization (τ-par) of the VECM for I(2) VAR systems[5]. See Mosconi and Paruolo (2017):

$$\Delta^2 X_t = \alpha \left(\rho'\tau' X_{t-1} + \psi' \Delta X_{t-1}\right) + \lambda\tau'\Delta X_{t-1} + \Upsilon \Delta^2 \mathbb{X}_{t-1} + \varepsilon_t, \tag{16}$$

with $\Upsilon\Delta^2\mathbb{X}_{t-1} = \sum_{j=1}^{k-2} \Upsilon_j \Delta^2 X_{t-j}$. Recall that $m = p - r_2$ is the total number of cointegrating relations, i.e., the number of I(1) linear combinations $\tau' X_t$. The number of linear combinations of $\tau' X_t$ that cointegrate with ΔX_t to I(0), i.e., the number of I(0) linear combinations $\rho'\tau' X_t + \psi'\Delta X_t$, is indicated[6] by $r \leq m$. Here α is $p \times r$, τ is $p \times m$ and the other parameter matrices are conformable; the parameters are α, ρ, τ, ψ, λ, Υ, Ω, all freely varying, and Ω is assumed to be positive definite. When λ is restricted as $\lambda = \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'$ with κ' a $(p - r) \times m$ matrix of freely varying parameters, the τ-par reduces to the parametrization of Johansen (1997); this restriction on λ is not imposed here.

4.1. Identification of τ

The parameters in the τ-par (16) are not identified; in particular τ' can be replaced by $B\tau'$ with B square and nonsingular, provided ρ and λ are simultaneously replaced by $B^{-1\prime}\rho$ and λB^{-1}. This is because τ enters the likelihood only via (16) in the products $\rho'\tau' = \rho' B^{-1} B\tau'$ and $\lambda\tau' = (\lambda B^{-1})(B\tau')$. The transformation that generates observationally equivalent parameters, i.e., the post multiplication of τ by a square and invertible matrix B', is the same type of transformation that induces observational

---

[5] In the general VAR(k) model (1), ε_t in (16) is replaced by $\mu_0 + \mu_1 t + \varepsilon_t$; see Section 4.3 below.
[6] The difference $m - r = p - r - r_2$ is referred to as either s or r_1 in the I(2) cointegration literature, see Appendix A.

equivalence in the classical system of simultaneous equations, see Sargan (1988), or to the set of cointegrating equations in I(1) systems, see Johansen (1995). This leads to the following result.

Theorem 4 (Identification of τ in the τ-par)**.** *Assume that τ_c is specified as the restricted τ_c in (9), which is implied by the general linear hypothesis (8) on $\tau_{\perp c_\perp}$; then the restricted τ_c is identified within the τ-par if and only if*

$$\operatorname{rank}\left(R_\tau'(I_m \otimes \tau)\right) = m^2, \qquad \underset{mp\times m_\tau}{R_\tau} = G_\perp, \qquad G := -(I_m \otimes c_\perp)\mathcal{K}_{m,r_2}K \tag{17}$$

(rank condition), where $m_\tau = mp - \dim \phi$. The corresponding order condition is $m_\tau \geq m^2$, or equivalently $mr_2 \geq \dim \phi$.

Alternatively, consider the general linear hypothesis (11) on $\tau_\perp$; then the constrained τ_c in (12) is identified in a neighborhood of the point $\phi = \phi^\star$ provided the Jacobian $\mathcal{J}(\phi^\star) := \partial \operatorname{vec} \tau_c(\phi^\star)/\partial\phi'$ in (13) is of full rank.

Proof. The rank condition follows from Sargan (1988), given that the class of transformation that induce observational equivalence is the same as the classical one for systems of simultaneous equations. The local identification condition follows from Rothenberg (1971). □

4.2. The Identification of Remaining Parameters

This subsection discusses conditions for remaining parameters of the τ-par to be identified, when τ is identified as in Theorem 4. These additional conditions are used in the discussion of the ML algorithms of the next section.

The VECM can be rewritten as

$$\Delta^2 X_t = \nu\varsigma' \begin{pmatrix} \tau' X_{t-1} \\ \Delta X_{t-1} \end{pmatrix} + \Upsilon \Delta^2 \mathbb{X}_{t-1} + \varepsilon_t, \quad \text{with} \quad \varsigma' := \begin{pmatrix} \rho' & \psi' \\ 0 & \tau' \end{pmatrix}, \qquad \nu := (\alpha : \lambda)\,.$$

One can see that the equilibrium correction terms $\nu\varsigma'\left((\tau' X_{t-1})' : \Delta X_{t-1}'\right)'$ may be replaced by $\nu^\circ\varsigma'^\circ\left((\tau^{\circ\prime} X_{t-1})' : \Delta X_{t-1}'\right)'$ without changing the likelihood, where $\nu^\circ := \nu Q^{-1} = (\alpha A^{-1} : \lambda B^{-1} - \alpha A^{-1} C)$, $\varsigma'^\circ := Q\varsigma' W^{-1}$ and

$$Q := \begin{pmatrix} A & CB \\ 0 & B \end{pmatrix}, \qquad W := \begin{pmatrix} B & 0 \\ 0 & I_p \end{pmatrix}, \qquad \varsigma'^\circ := Q\varsigma' W^{-1} = \begin{pmatrix} A\rho' B^{-1} & A\psi' + CB\tau' \\ 0 & B\tau' \end{pmatrix};$$

here A and B are square nonsingular matrices, and C is a generic matrix. Hence one observes that $(\alpha, \rho, \tau, \psi, \lambda, \Upsilon, \Omega)$ is observationally equivalent to $(\alpha^\circ, \rho^\circ, \tau^\circ, \psi^\circ, \lambda^\circ, \Upsilon, \Omega)$. A, B and C define the class of observationally equivalent transformations in the τ-par for all parameters, including τ. When τ is identified one has $B = I_m$ in the above formulae.

Consider additional restrictions on φ of the type:

$$\underset{m_\varphi \times f_\varphi}{R_\varphi'} \operatorname{vec} \varphi' = q_\varphi, \qquad \varphi' := (\rho' : \psi')\,. \tag{18}$$

where $f_\varphi = r(p+m)$. The next theorem states rank conditions for these restrictions to identify the remaining parameters.

Theorem 5 (Identification of other parameters in the τ-par)**.** *Assume that τ is identified as in Theorem 4; the restrictions (18) identify φ and all other parameters in the τ-par if and only if (rank condition)*

$$\operatorname{rank} R_\varphi' (\varsigma \otimes I_r) = r(r+m). \tag{19}$$

A necessary but not sufficient condition (order condition) for this is that

$$m_\varphi \geq r(r+m). \tag{20}$$

Proof. Because τ is identified, one has $B = I_m$ in Q. For the identification of φ, observe that $\varsigma - \varsigma^\circ = \varsigma(I - Q')$. One finds $\varphi - \varphi^\circ = (\varsigma - \varsigma^\circ)(I_r : 0)' = \varsigma(I_{m+r} - Q')(I_r : 0)'$. Because both φ and φ° satisfy (18), one has $0 = R'_\varphi \operatorname{vec}(\varphi' - \varphi^{\circ\prime}) = R'_\varphi(\varsigma \otimes I_r)\operatorname{vec}((I_r : 0)(I_{r+m} - Q))$. This implies that $(I_r : 0)(I_{m+r} - Q) = 0$, i.e., that both $A = I_r$ and $C = 0_{r\times m}$, and that φ is identified, if and only if rank $R'_\varphi(\varsigma \otimes I_r) = r(r+m)$. This completes the proof. □

Observe that the identification properties of the τ-par differ from the ones of the parametrization of Johansen (1997), where $\lambda = \Omega\alpha_\perp(\alpha'_\perp\Omega\alpha_\perp)^{-1}\kappa'$ is restricted, and hence the adding-and-subtracting associated with C above is not permitted.

4.3. Deterministic Terms

The τ-par in (16) does not involve deterministic terms. Allowing a constant and a trend to enter the VAR Equation (1) in a way that rules out quadratic trends, one obtains the following equilibrium correction I(2) model—for simplicity still called the τ-par below:

$$\Delta^2 X_t = \alpha\left(\rho'\tau^{\star\prime}X^\star_{t-1} + \psi^{\star\prime}\Delta X^\star_{t-1}\right) + \lambda\tau^{\star\prime}\Delta X^\star_{t-1} + \Upsilon\Delta^2\mathbb{X}_{t-1} + \varepsilon_t. \tag{21}$$

Here $X^\star_{t-1} = (X'_{t-1} : t)'$ so that $\Delta X^\star_{t-1} = (\Delta X'_{t-1} : 1)'$; and $\tau^\star = (\tau' : \tau_1)$ and $\psi^\star = (\psi' : \psi_0)'$.

This parametrization satisfies the conditions of the Johansen Representation Theorem and it generates deterministic trends up to first order, as shown in Appendix A. This is the I(2) model used in the application, with the addition of unrestricted dummy variables.

5. Likelihood Maximization

This section discusses likelihood maximization of the τ-par of the I(2) model (16) under linear, possibly over-identifying, restrictions on $\tau_{\perp c_\perp}$, i.e., on ϑ in (5). The same treatment applies to (21) replacing $(X_{t-1}, \Delta X_{t-1})$ with $(X^\star_{t-1}, \Delta X^\star_{t-1})$, and (τ, ψ), with $(\tau^\star, \psi^\star)$. The formulation (16) is preferred here for simplicity in exposition.

The alternating maximization procedure proposed here is closely related, but not identical, to the algorithms proposed by Doornik (2017b); related algorithms were discussed in Paruolo (2000b). Restricted ML estimation in the I(1) model was discussed in Boswijk and Doornik (2004).

5.1. Normalizations

Consider restrictions (8), which are translated into linear hypotheses on τ_c in (9) as follows

$$\operatorname{vec}\,\tau_c = (\operatorname{vec}\bar{c} - (I_m \otimes c_\perp)\mathcal{K}_{m,r_2}k) - (I_m \otimes c_\perp)\mathcal{K}_{m,r_2}K\phi =: g + G\phi,$$

where by construction g and G satisfy $(I_m \otimes c')g = \operatorname{vec} I_{rm}$ and $(I_m \otimes c')G = 0$ such that $c'\tau_c = I_m$.

Next, consider just-identifying restrictions on the remaining parameters. For ψ, the linear combinations of first differences entering the multicointegration relations, one can consider

$$c'\psi = 0 \qquad \Longleftrightarrow \qquad \psi = c_\perp\delta', \tag{22}$$

where δ is the $r \times r_2$ matrix of multicointegration parameters. This restriction differs from the restriction $\psi = \tau_\perp\delta'$ which is considered e.g., in Juselius (2017a, 2017b), and it was proposed and analysed by Boswijk (2000).

Furthermore, the $m \times r$ matrix ρ can be normalized as follows

$$d'\rho = I_r \qquad \Longleftrightarrow \qquad \rho = \bar{d} + d_\perp\varrho, \tag{23}$$

where d is some known $m \times r$ matrix, and where ϱ, of dimension $(m-r) \times r$, contains freely varying parameters.

It can be shown that restrictions (22) and (23) identify the remaining parameters using Theorem 5. In fact, (22) and (23) can be written as $\varphi' V = v$ where $V := \text{blkdiag}(d, c)$ and $v := (I_r : 0_{r\times m})$. Vectorizing, one obtains an equation $R'_\varphi \operatorname{vec} \varphi' = q_\varphi$ of the form (18) with $R_\varphi = (V \otimes I_r)$ and $q_\varphi = \operatorname{vec} v$. The rank condition (19) is satisfied, since $R'_\varphi (\varsigma \otimes I_r) = (V'\varsigma \otimes I_r) = I_{r(m+r)}$ because

$$V'\varsigma = \begin{pmatrix} d'\rho & 0 \\ c'\psi & c'\tau \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & I_m \end{pmatrix},$$

where the last equality follows from (22) and (23) and $\tau = \tau_c$.

5.2. The Concentrated Likelihood Function

The model (16), after concentrating out the unrestricted parameter matrix Y, can be represented by the equations

$$Z_{0t} = \alpha(\rho'\tau' Z_{2t} + \psi' Z_{1t}) + \lambda\tau' Z_{1t} + \varepsilon_t(\xi), \tag{24}$$

where ξ indicates the vector of free parameters in $(\alpha, \varrho, \phi, \delta, \lambda)$, Z_{0t}, Z_{1t} and Z_{2t} are residual vectors of regressions of $\Delta^2 X_t$, ΔX_{t-1} and X_{t-1}, respectively, on $\mathbb{X}_{t-1}$;[7] this derivation follows similarly to Chapter 6.1 in Johansen (1996). The associated log-likelihood function, concentrated with respect to Y, is given by

$$\ell(\xi, \Omega) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t(\xi)' \,\Omega^{-1} \varepsilon_t(\xi),$$

In the rest of this section, ε_t is used as shorthand for $\varepsilon_t(\xi)$.

Algorithms for the maximization of the concentrated log-likelihood function $\ell(\xi, \Omega)$ are proposed below. The first one, called AL1, considers the alternative maximization of $\ell(\xi, \Omega)$ over $(\alpha, \varrho, \delta, \lambda, \Omega)$ for a fixed value of ϕ (called the α-step), and over (ϕ, δ) for a given value of $(\alpha, \varrho, \lambda, \Omega)$ (called the τ-step).

A variant of this algorithm, called AL2, can be defined fixing δ in the τ-step to the value of δ obtained in the α-step. It can be shown that the increase in $\ell(\xi, \Omega)$ obtained in one combination of α-step and τ-step of AL1 is greater or equal to the one obtained by AL2. The proof of this result is reported in Proposition A1 in Appendix B. Because of this property, and because AL2 may display very slow convergence properties in practice, AL1 is implemented in the illustration below.

The rest of this section presents algorithms AL1 and AL2, defining first the τ-step, then the α-step and finally discussing the starting values, a line search and normalizations.

5.2.1. τ Step

Taking differentials, one has $\mathrm{d}\ell = -\sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \mathrm{d}\varepsilon_t$. Keeping $(\alpha, \varrho, \lambda)$ fixed, one finds

$$\begin{aligned} -\mathrm{d}\varepsilon_t &= \mathrm{d}\left(\alpha\rho'\tau' Z_{2t} + \alpha\psi' Z_{1t} + \lambda\tau' Z_{1t}\right) \\ &= \left((Z'_{2t} \otimes \alpha\rho') + (Z'_{1t} \otimes \lambda)\right) \mathrm{d}\operatorname{vec}\tau' + (Z'_{1t} \otimes \alpha)\mathrm{d}\operatorname{vec}\psi' \\ &= \left((Z'_{2t} \otimes \alpha\rho') + (Z'_{1t} \otimes \lambda)\right) \mathcal{K}_{m,r_1} G\mathrm{d}\phi + (Z'_{1t} c_\perp \otimes \alpha)\mathrm{d}\operatorname{vec}\delta. \end{aligned}$$

Writing ε_t in terms of ϕ and $\operatorname{vec}\delta$, i.e., $\varepsilon_t = Z_{0t} - \left((Z'_{2t} \otimes \alpha\rho') + (Z'_{1t} \otimes \lambda)\right) \mathcal{K}_{m,r_1}(G\phi + g) - (Z'_{1t} c_\perp \otimes \alpha) \operatorname{vec}\delta$, the first-order conditions $\partial\ell/\partial\phi = 0$ and $\partial\ell/\partial \operatorname{vec}\delta = 0$ are solved by

---

[7] If a restricted constant and linear trend are included in the model, as in (21), then Z_{1t} and Z_{2t} are defined as the residual vectors of regressions of $\Delta X^\star_{t-1}$ and $X^\star_{t-1}$, respectively, on $\mathbb{X}_{t-1}$.

$$\begin{pmatrix} \widehat{\varphi} \\ \text{vec } \widehat{\delta}' \end{pmatrix} = \begin{pmatrix} G'U_1'\left(\Omega^{-1} \otimes I_T\right) U_1 G & G'U_1'\left(\Omega^{-1} \otimes I_T\right) U_2 \\ U_2'\left(\Omega^{-1} \otimes I_T\right) U_1 G & U_2'\left(\Omega^{-1} \otimes I_T\right) U_2 \end{pmatrix}^{-1} \cdot \\ \cdot \begin{pmatrix} G'U_1'\left(\Omega^{-1} \otimes I_T\right) \\ U_2'\left(\Omega^{-1} \otimes I_T\right) \end{pmatrix} \left(\text{vec } Z_0 - U_1 g\right), \tag{25}$$

where $Z_j = \left(Z_{j1} : \ldots : Z_{jT}\right)'$, $j = 0, 1, 2$, and where $U_1 = (\alpha\rho' \otimes Z_2) + (\lambda \otimes Z_1)$, and $U_2 = (\alpha \otimes Z_1 c_\perp)$. Note that (25) is the GLS estimator in a regression of vec $Z_0 - U_1 g$ on $(U_1 G : U_2)$. This defines the τ-step for AL1.

The τ-step for AL2 is defined similarly, but keeping δ fixed. In this case it is simple to see that

$$\widehat{\varphi} = \left(G'U_1'\left(\Omega^{-1} \otimes I_T\right) U_1 G\right)^{-1} G'U_1'\left(\Omega^{-1} \otimes I_T\right) \left(\text{vec } Z_0 - U_1 g - \text{vec } \left(Z_1 \psi \alpha'\right)\right).$$

5.2.2. α Step

When φ is fixed (and hence τ is fixed), one can construct $Z_{3t} = \tau' Z_{1t}$ and

$$Z_{4t} = \begin{pmatrix} \bar{d}'\tau' Z_{2t} \\ d_\perp' \tau' Z_{2t} \\ c_\perp' Z_{1t} \end{pmatrix}, \qquad \gamma = \begin{pmatrix} I_r \\ \varrho \\ \delta' \end{pmatrix}.$$

The concentrated model (24) can then be written as a reduced rank regression:

$$Z_{0t} = \alpha\gamma' Z_{4t} + \lambda Z_{3t} + \varepsilon_t,$$

for which the Guassian ML estimator for α, γ, λ has a closed-form solution, see Johansen (1996). Specifically, let $M_{ij} := T^{-1} \sum_{t=1}^{T} Z_{it} Z_{jt}'$, $i, j = 0, 3, 4$ and $S_{ij} := M_{ij} - M_{i3} M_{33}^{-1} M_{3j}$, $i, j = 0, 4$. If v_i, $i = 1, \ldots, r$, are the eigenvectors corresponding to the largest r eigenvalues of the problem

$$(\mu S_{44} - S_{40} S_{00}^{-1} S_{04}) v = 0,$$

and $v = (v_i, \ldots, v_r)$ is the matrix of the corresponding eigenvectors, then the optimal solutions for ϱ, δ, α, λ is given by

$$\widehat{\gamma} = \begin{pmatrix} I_{r_0} \\ \widehat{\varrho} \\ \widehat{\delta}' \end{pmatrix} = v(e'v)^{-1}, \qquad \widehat{\alpha} = S_{04}\widehat{\gamma}(\widehat{\gamma}' S_{44} \widehat{\gamma})^{-1}, \qquad \widehat{\lambda} = (M_{03} - \widehat{\alpha}\widehat{\gamma}' M_{43}) M_{33}^{-1},$$

where $e' = (I_r : 0)$. Optimization with respect to $\widehat{\Omega}$ is performed using $\Omega(\xi) = T^{-1} \sum_{t=1}^{T} \varepsilon_t(\xi) \varepsilon_t(\xi)'$ replacing ξ with $\widehat{\xi}$ formed from the previous expressions, namely taking $(\alpha, \varrho, \delta, \lambda)$ equal to $(\widehat{\alpha}, \widehat{\varrho}, \widehat{\delta}, \widehat{\lambda})$ in the above display and $\varphi = \widehat{\varphi}$ from the τ-step. Using the S_{ij} matrices, one can also compute $\widehat{\Omega}$ directly as $\widehat{\Omega} = S_{00} - S_{04}\widehat{\gamma}(\widehat{\gamma}' S_{44} \widehat{\gamma})^{-1} \widehat{\gamma}' S_{40}$. This completes the definition of the α-step.

5.2.3. Starting Values and Line Search

If the system is just-identified, consistent starting values for all parameters can be obtained by imposing the identifying restrictions on the two-stage estimator for the I(2) model (2SI2), see Johansen (1995) and Paruolo (2000a). In case of over-identification, this method may be used to produce starting values for $(\alpha, \varrho, \lambda)$, which may then be used as input for the first τ-step to obtain starting values for φ and δ.

Let η be the vector containing all free parameters in $(\alpha, \varrho, \delta, \lambda)$, and let $\xi := (\varphi' : \eta')'$. Denote by $\xi_{j-1} = (\varphi_{j-1}' : \eta_{j-1}')'$ the value of ξ in iteration $(j-1)$ of algorithms. Denote as $\widehat{\xi}_j = (\widehat{\varphi}_j' : \widehat{\eta}_j')'$ the value of ξ obtained by the application of a τ-step and α-step of algorithms AL1 and AL2 at iteration j starting from ξ_{j-1}. In an I(1) context, Doornik (2017a) found that better convergence properties can be obtained if a line search is added. For this purpose, define the final value of the j-th iteration as

$$\xi_j(\omega) = \xi_{j-1} + \omega(\widehat{\xi}_j - \xi_{j-1})$$

where ω is chosen in $\mathbb{R}_+ = (0, \infty)$ using a line search; note that values of ω greater than 1 are admissible. A simple (albeit admittedly sub-optimal) implementation of the line search is employed in Doornik (2017a); it consists of evaluating the log-likelihood function $\ell(\xi, \Omega(\xi))$ with $\Omega(\xi) = T^{-1}\sum_{t=1}^{T} \varepsilon_t(\xi)\varepsilon_t(\xi)'$ setting ξ equal to $\xi_j(\omega)$ for $\omega \in \{1.2, 2, 4, 8\}$, and in choosing the value of ω with the highest loglikelihood ℓ. This simple choice of line search is used in the empirical illustration.

5.3. Standard Errors

The asymptotic variance matrix of the ML estimators may be obtained from the inverse observed (concentrated) information matrix as usual. Writing (24) as $Z_{0t} = \Pi Z_{2t} + \Gamma Z_{1t} + \varepsilon_t$, and letting $\theta = (\text{vec}\,(\Pi')' : \text{vec}\,(\Gamma')')'$, the observed concentrated information matrix for the reduced-form parameter vector θ is obtained from

$$\mathcal{I}_\theta = -\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \Omega^{-1} \otimes Z_2'Z_2 & \Omega^{-1} \otimes Z_2'Z_1 \\ \Omega^{-1} \otimes Z_1'Z_2 & \Omega^{-1} \otimes Z_2'Z_2 \end{pmatrix}.$$

This leads to the following information matrix in terms of the parameters (ϕ, η):

$$\mathcal{I}_{\phi,\eta} = \begin{pmatrix} J_\phi' \\ J_\eta' \end{pmatrix} \mathcal{I}_\theta \begin{pmatrix} J_\phi & J_\eta \end{pmatrix},$$

where $J_\phi = \partial\theta/\partial\phi'$ and $J_\eta = \partial\theta/\partial\eta'$. From $\Pi = \alpha\rho'\tau'$ and $\Gamma = \alpha\psi' + \lambda\tau'$, one obtains

$$J_\phi = \begin{pmatrix} \alpha\rho' \otimes I_p \\ \lambda \otimes I_p \end{pmatrix} G.$$

Define $\eta = (\text{vec}\,(\alpha')' : \text{vec}\,(\varrho)' : \text{vec}\,(\delta')' : \text{vec}\,(\lambda')')'$, so that $J_\eta = [J_\alpha : J_\varrho : J_\delta : J_\lambda]$, with

$$J_\alpha = \begin{pmatrix} I_p \otimes \tau\rho \\ I_p \otimes \psi \end{pmatrix}, \qquad J_\varrho = \begin{pmatrix} \alpha \otimes \tau d_\perp \\ 0 \end{pmatrix}, \qquad J_\delta = \begin{pmatrix} 0 \\ \alpha \otimes c_\perp \end{pmatrix}, \qquad J_\lambda = \begin{pmatrix} 0 \\ I_p \otimes \tau \end{pmatrix}.$$

With these ingredients, one finds

$$\widehat{\text{var}}(\widehat{\phi}) = \left(\widehat{J}_\phi'\widehat{\mathcal{I}}_\theta\widehat{J}_\phi - \widehat{J}_\phi'\widehat{\mathcal{I}}_\theta\widehat{J}_\eta(\widehat{J}_\eta'\widehat{\mathcal{I}}_\theta\widehat{J}_\eta)^{-1}\widehat{J}_\eta'\widehat{\mathcal{I}}_\theta\widehat{J}_\phi\right)^{-1},$$

where $\widehat{\mathcal{I}}_\theta$, $\widehat{J}_\phi$ and $\widehat{J}_\eta$ are the expressions given above, evaluated at the ML estimators. Standard errors of individual parameters estimates are obtained as the square root of the diagonal elements of $\widehat{\text{var}}(\widehat{\phi})$. Asymptotic normality of resulting t-statistics (under the null hypothesis), and χ^2 asymptotic null distributions of likelihood ratio test statistics for the over-identifying restrictions, depend on conditions for asymptotic mixed normality being satisfied; this is discussed next.

6. Asymptotics

The asymptotic distribution of the ML estimator in the I(2) model has been discussed in Johansen (1997, 2006). As shown there and discussed in Boswijk (2000), the limit distribution of the ML estimator is not jointly mixed normal as in the I(1) case. As a consequence, the limit distribution of LR test statistics of generic hypotheses need not be χ^2 under the null hypothesis.

In some special cases, the asymptotic distribution of the just-identified ML estimator of the cointegration parameters can be shown to be asymptotically mixed normal. Consider the case $r_1 = 0$ (i.e., $r = m$), and assume as before that no deterministic terms are included in the model. In this case, the limit distribution of the cointegration parameters in Theorem 4 in Johansen (2006), J06 hereafter, can be described in terms of the estimated parameters $\widehat{B}_0 := \bar{\tau}_\perp'(\widehat{\psi} - \psi)$ and $\widehat{B}_2 := \bar{\tau}_\perp'(\widehat{\tau} - \tau)$, where $\widehat{\tau}$ is identified as τ_c with $c = \tau$. Note that the components C and B_1 in the above theorem do not appear here, because $r_1 = 0$. One has

$$\begin{pmatrix} T\widehat{B}_0 \\ T^2\widehat{B}_2 \end{pmatrix} \xrightarrow{w} B^\infty := \left(\int_0^1 H_*(s)H_*(s)'\mathrm{d}s\right)^{-1} \int_0^1 H_*(s)\mathrm{d}W_1(s)$$

with $H_*(u) := (H_0(u)' : H_2(u)')'$,

$$H_{2u} := \int_0^u H_0(s)\mathrm{d}s, \qquad H_0(u) := \tau_\perp' C_2 W(u), \qquad W_1(u) := \left(\alpha'\Omega^{-1}\alpha\right)^{-1} \alpha'\Omega^{-1}W(u),$$

and where $T^{-\frac{1}{2}} \sum_{i=1}^{\lfloor Tu \rfloor} \varepsilon_i \xrightarrow{w} W(u)$, a vector Brownian motion with covariance matrix Ω[8].

As noted in J06, B^∞ has a mixed normal distribution with mean 0, because $H_*(u)$ is a function of $\alpha_\perp' W(u)$, which is independent of $W_1(u)$. Moreover in the case $r_1 = 0$, the C^∞ component of the ML limit distribution does not appear, so that the whole limit distribution of the cointegration parameters is *jointly* mixed normal, unlike in the case $r_1 > 0$.

One can see that hypothesis (8) defines a smooth restriction of the B_2 parameters[9]. More precisely B_2 depends smoothly only on ϕ_2, $B_2 = B_2(\phi_2)$, where ϕ_2 contains the ϕ parameters in (8). Note also that B_0 depends on the parameters in ψ, which are unrestricted by (8); hence B_0 depends only on ϕ_1, $B_0 = B_0(\phi_1)$, where ϕ_1 contains the parameters in δ in (22).

The conditions of Theorem 5 in J06 are next shown to be verified, and hence the LR test of the hypothesis (8) is asymptotically χ^2 with degrees of freedom equal to the number of constraints, in case $r_1 = 0$. In fact, $B_0(\phi_1)$, $B_2(\phi_2)$ are smoothly parametrizated by the continuously identified parameters ϕ_1 and ϕ_2. Because B_2 does not depend on ϕ_1, one easily deduces $\partial B_2/\partial\phi_1 = \partial^2 B_2/\partial\phi_1^2 = 0$ in (37) of J06. Similarly, one has $\phi_1 = \phi_{1B}$ with $\partial B_0/\partial\phi_1$ and $\partial B_2/\partial\phi_2$ of full rank; hence (38) of J06 is satisfied. This shows that the LR statistic is asymptotically χ^2 under the null, for $r_1 = 0$.

In case $r_1 = (m - r) > 0$, the asymptotic distribution of $\widehat{\tau}$ is defined in terms of (B^∞, C^∞) in J06 p. 92, which is not jointly mixed normal. In such cases, Boswijk (2000) showed that inference is mixed normal if the restrictions on $\widehat{\tau}_c$ can be asymptotically linearized in (B^∞, C^∞), and separated into two sets of restrictions, the first group involving B^∞ only, and the second group involving C^∞ only. Because the conditions of Theorem 5 in J06 cannot be easily verified for general linear hypotheses of the form (8) in this case, they will need to be checked case by case. The authors intend to develop more readily verifiable conditions for χ^2 inference on τ in their future research.

7. Illustration

Following Juselius and Assenmacher (2015), consider a 7-dimensional VAR with

$$X_t = (p_{1t} : p_{2t} : e_{12t} : b_{1t} : b_{2t} : s_{1t} : s_{2t})',$$

where p_{it}, b_{it}, s_{it} are the (log of) the price index, the long and the short interest rate of country i at time t respectively, and e_{12t} is the log of the exchange rate between country 1 (Switzerland) and 2 (the US)

8 Here $\xrightarrow{w}$ indicates weak convergence and $\lfloor \cdot \rfloor$ denotes the greatest integer part.
9 In the rest of this section the notation ϕ_1, ϕ_2 and $\partial B_i/\partial\phi_j$ are used in accordance to the notation in J06.

at time t. The results are based on quarterly data over the period 1975:1–2013:3. The model has two lags, a restricted linear trend as in (21), which appears in the equilibrium correction only appended to the vector of lagged levels, and a number of dummy variables; see Juselius and Assenmacher (2017), which is an updated version of Juselius and Assenmacher (2015), for further details on the empirical model. The data set used here is taken from Juselius and Assenmacher (2017).

Specification (3) is based on the prediction that $r_2 = 2$. Based on I(2) cointegration tests, Juselius and Assenmacher (2017) choose a model with $r = m = 5$, which indeed implies $r_2 = 2$, but also $r_1 = m - r = 0$; arguably, however, the test results in Table 1 of their paper also support the hypothesis $(r, r_1) = (4, 1)$, which has the same number $r_2 = 2$ of common $I(2)$ trends. The latter model would be selected applying the sequential procedure in Nielsen and Rahbek (2007) using a 5% or 10% significance level in each test in the sequence.

Consider the case $(r, r_1) = (5, 0)$. The over-identifying restrictions on $\tau_\perp$ implied by (3) are incorporated in the parametrization (3), with normalizations $\phi_{11} = \phi_{42} = 1$, which in turn leads to the over-identified structure for τ_c in (15), to be estimated by ML. The restricted ML estimate of $\tau_{\perp c_\perp}$ is (standard errors in parentheses):

$$\widehat{\tau}_{\perp c_\perp} = \begin{pmatrix} 1 & 0 \\ \underset{(0.11)}{1.49} & \underset{(5.23)}{-25.14} \\ \underset{(0.72)}{-1.88} & \underset{(29.81)}{-35.70} \\ 0 & 1 \\ 0 & \underset{(0.53)}{-1.91} \\ 0 & \underset{(0.29)}{1.23} \\ 0 & \underset{(0.95)}{-3.02} \end{pmatrix}.$$

The LR statistics for the 3 over-identifying restrictions equals 16.11. Using the $\chi^2(3)$ asymptotic limit distribution, one finds an asymptotic p-value of 0.001, and hence a rejection of the null hypothesis. This indicates that the hypothesized structure on $\tau_\perp$ is rejected.

For comparison, consider also the case $(r, r_1) = (4, 1)$, for which the LR test for cointegration ranks has a p-value of 0.13. The resulting restricted estimate of $\tau_{\perp c_\perp}$ is:

$$\widehat{\tau}_{\perp c_\perp} = \begin{pmatrix} 1 & 0 \\ \underset{(0.09)}{1.38} & \underset{(5.22)}{-24.67} \\ \underset{(0.56)}{-1.07} & \underset{(22.42)}{-30.10} \\ 0 & 1 \\ 0 & \underset{(0.52)}{-1.75} \\ 0 & \underset{(0.28)}{1.20} \\ 0 & \underset{(1.02)}{-2.97} \end{pmatrix}.$$

The estimates and standard errors are similar to those obtained under the hypothesis $(r, r_1) = (5, 0)$. The LR statistic for the over-identifying restrictions now equals 10.08. If one conjectured that the limit distribution of the LR test is also $\chi^2(3)$ in this case, one would obtain an asymptotic p-value of 0.018, so the evidence against the hypothesized structure of τ appears slightly weaker in this model.

The results for both model $(r, r_1) = (5, 0)$ and for model $(r, r_1) = (4, 1)$ are in line with the preferred specification of Juselius and Assenmacher (2017), who select an over-identified structure for τ, which is not nested in (15), and therefore implies a different impact of the common I(2) trends.

8. Conclusions

Hypotheses on the loading matrix of I(2) common trends are of economic interest. They are shown to be related to the cointegration relations. This link is explicitly discussed in this paper, also for hypotheses that are over-identifying. Likelihood maximization algorithms are proposed and discussed, along with LR tests of the hypotheses.

The application of these LR tests to a system of prices, exchange rates and interest rates for Switzerland and the US shows support for the existence of two I(2) common trends. These may represent a 'speculative' trend and a 'relative prices' trend, but there is little empirical support for the corresponding exclusion restrictions in the loading matrix.

Acknowledgments: Helpful comments and suggestions from two anonymous referees and the Academic Editor, Katarina Juselius, are gratefully acknowledged.

Author Contributions: Both authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Theorem A1 (Johansen Representation Theorem)**.** *Let the vector process X_t satisfy $A(L)X_t = \mu_0 + \mu_1 t + \varepsilon_t$, where $A(L) := I_p - \sum_{i=1}^{k} A_i L^i$, a matrix lag polynomial of degree k, and where ε_t is an i.i.d. $(0, \Omega)$ sequence. Assume that $A(z)$ is of full rank for all $|z| < 1 + c$, $c > 0$, with the exception of $z = 1$. Let A, $\dot{A}$ and $\ddot{A}$ denote $A(1)$, the first and second derivative of $A(z)$ with respect to z, evaluated at $z = 1$; finally define $\Gamma = \dot{A} - A$. Then X_t is I(2) if and only if the following conditions hold:*

- *(i)* $A = -\alpha\beta'$ *where* α, β *are* $p \times r$ *matrices of full column rank* $r < p$,
- *(ii)* $P_{\alpha_\perp} \Gamma P_{\beta_\perp} = \alpha_1 \beta_1'$ *where* α_1, β_1 *are* $p \times r_1$ *matrices of full column rank* $r_1 < p - r$,
- *(iii)* $\alpha_2' \Theta \beta_2$ *is of full rank* $r_2 := p - r - r_1$, *where* $\Theta := \frac{1}{2}\ddot{A} + \dot{A}\bar{\beta}\bar{\alpha}'\dot{A}$, $\alpha_2 := (\alpha, \alpha_1)_\perp$ *and* $\beta_2 := (\beta, \beta_1)_\perp$,
- *(iv)* $\mu_1 = \alpha\beta_D$ *for some* β_D,
- *(v)* $\alpha_2'\mu_0 = \alpha_2'\Gamma\bar{\beta}\beta_D$.

Under these conditions, X_t admits a common trends I(2) representation of the form

$$X_t = C_2 \sum_{i=1}^{t}\sum_{s=1}^{i} \varepsilon_s + C_1 \sum_{i=1}^{t} \varepsilon_i + C^\star(L)\varepsilon_t + v_0 + v_1 t, \tag{A1}$$

where

$$C_2 = \beta_2(\alpha_2'\Theta\beta_2)^{-1}\alpha_2', \tag{A2}$$

$C^\star(L)\varepsilon_t$ is an I(0) linear process, and v_0 and v_1 depend on the VAR coefficients and on the initial values of the process.

Proof. See Johansen (1992), Johansen (2009) and Rahbek et al. (1999), which also contain expressions for C_1, $C^*(L)$ and (v_0, v_1). □

It is next shown that conditions (*iv*) and (*v*) are satisfied by the τ-par (21). In fact, condition (*iv*) holds for $\beta_D = \rho'\tau_1$. Note that $\Gamma = \alpha\psi' + \lambda\tau'$, $\beta = \tau\rho$ and $P_{\alpha_\perp}\Gamma P_{\beta_\perp} = P_{\alpha_\perp}\lambda\tau' P_{\beta_\perp} = \alpha_1\beta_1'$. The l.h.s. of (*v*) is

$$\alpha_2'\mu_0 = \alpha_2'\lambda\tau_1. \tag{A3}$$

Next write the r.h.s. of (*v*) using $\tau'\tau\rho(\rho'\tau'\tau\rho)^{-1}\rho' = I - \rho_\perp(\rho_\perp'(\tau'\tau)^{-1}\rho_\perp)^{-1}\rho_\perp'(\tau'\tau)^{-1}$ by oblique projections; one finds

$$\begin{aligned}\alpha_2'\Gamma\bar{\beta}\beta_D &= \alpha_2'\lambda\tau'\tau\rho(\rho'\tau'\tau\rho)^{-1}\rho'\tau_1 \\ &= \alpha_2'\lambda\tau_1 - \alpha_2'\lambda\rho_\perp(\rho_\perp'(\tau'\tau)^{-1}\rho_\perp)^{-1}\rho_\perp'(\tau'\tau)^{-1}\tau_1 = \alpha_2'\lambda\tau_1\end{aligned} \tag{A4}$$

where the last equality holds because $\alpha_2'\lambda\rho_\perp = 0$, as shown below. Note in fact that $\beta_1 = \bar{\tau}\rho_\perp$ lies in $\operatorname{col}\beta_\perp$ and α_2 lies in $\operatorname{col}\alpha_\perp$; hence one can write

$$\alpha_2'\lambda\rho_\perp = \alpha_2'\lambda\tau'\bar{\tau}\rho_\perp = \alpha_2'P_{\alpha_\perp}\lambda\tau'P_{\beta_\perp}\beta_1 = \alpha_2'P_{\alpha_\perp}\Gamma P_{\beta_\perp}\beta_1 = \alpha_2'\alpha_1\beta_1'\beta_1 = 0.$$

Hence, because (A3) equals (A4), condition (v) is satisfied.

Appendix B

This Appendix contains a proof that the increase in ℓ in one combination of α-step and τ-step of AL1 is greater or equal to the one obtained by AL2. In order to state the argument in somewhat greater generality, define a parameter vector θ partitioned in 3 components, denoted $(\theta_1, \theta_2, \theta_3)$, where each θ_j represents a subvector of parameters, respectively of dimensions n_1, n_2, n_3. Let $\ell(\theta)$ be the log-likelihood function. Define also the following switching algorithms, both starting at the same initial value $(\theta_1^{(j-1)}, \theta_2^{(j-1)}, \theta_3^{(j-1)})$:

Definition A1. ALGO1 *(3 way switching)*

Step 1: *for fixed θ_1, maximize ℓ with respect to (θ_2, θ_3);*
Step 2: *for fixed θ_2, maximize ℓ with respect to (θ_1, θ_3).*

Let $\ell(\theta^{(1,j)})$ be the value of ℓ corresponding to the application of step 1 and 2 of ALGO1.

Definition A2. ALGO2 *(Pure switching)*

Step 1: *for fixed θ_1, maximize ℓ with respect to (θ_2, θ_3);*
Step 2: *for fixed (θ_2, θ_3), maximize ℓ with respect to θ_1.*

Let $\ell(\theta^{(2,j)})$ be the value of ℓ corresponding to the application of step 1 and 2 of ALGO2.

Proposition A1 (Pure versus 3-way switching)**.** *One has $\ell(\theta^{(1,j)}) \geq \ell(\theta^{(2,j)})$.*

Proof. In order to see this, let

$$(\theta_2^\star, \theta_3^\star) = \arg\max_{\theta_2,\theta_3} \ell(\theta_1^{(j-1)}, \theta_2, \theta_3).$$

Step 1 is the same for ALGO1 and ALGO2. In the second step of ALGO1 one considers

$$\ell(\theta^{(1,j)}) = \max_{\theta_1,\theta_3} \ell(\theta_1, \theta_2^\star, \theta_3), \tag{A5}$$

while for ALGO2 one considers

$$\ell(\theta^{(2,j)}) = \max_{\theta_1} \ell(\theta_1, \theta_2^\star, \theta_3^\star). \tag{A6}$$

The conclusion that $\ell(\theta^{(1,j)}) \geq \ell(\theta^{(2,j)})$ follows from the fact that the maximization problem (A6) is a constrained version of (A5) under $\theta_3 = \theta_3^\star$. □

It is simple to observe that the argument of the proof implies that the larger the dimension of n_3, the better.

References

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

Boswijk, H. Peter, and Jurgen A. Doornik. 2004. Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica* 58: 440–65.

Doornik, Jurgen A. 2017a. Accelerated estimation of switching algorithms: The cointegrated VAR model and other applications. Working paper, University of Oxford, Oxford, UK.

Doornik, Jurgen A. 2017b. Maximum likelihood estimation of the I(2) model under linear restrictions. *Econometrics* 5: 19.

Engle, Robert F., and Clive W. J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55: 251–76.

Johansen, Søren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–80.

Johansen, Søren. 1992. A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8: 188–202.

Johansen, Søren. 1995. Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics* 69: 111–32.

Johansen, Søren. 1996. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd printing. Oxford: Oxford University Press.

Johansen, Søren. 1997. A likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24: 433–62.

Johansen, Søren. 2006. Statistical analysis of hypotheses on the cointegrating relations in the I(2) model. *Journal of Econometrics* 132: 81–115.

Johansen, Søren. 2009. Representation of cointegrated autoregressive processes with application to fractional processes. *Econometric Reviews* 28: 121–45.

Jöreskog, Karl G., Ulf H. Olsson, and Fan Y. Wallentin. 2016. *Multivariate Analysis with LISREL*. Basel: Springer International Publishing.

Juselius, Katarina. 2017a. A theory consistent CVAR scenario for a standard monetary model using data-generated expectations. Working paper, University of Copenhagen, Copenhagen, Denmark.

Juselius, Katarina. 2017b. Using a theory-consistent CVAR scenario to test an exchange rate model based on imperfect knowledge. Working paper, University of Copenhagen, Copenhagen, Denmark.

Juselius, Katarina, and Katrin Assenmacher. 2015. Real exchange rate persistence: The case of the Swiss franc-US dollar rate. SNB Working paper 2015-03, Swiss National Bank, Zürich, Switzerland.

Juselius, Katarina, and Katrin Assenmacher. 2017. Real Exchange Rate Persistence and the Excess Return Puzzle: The Case of Switzerland Versus the US. *Journal of Applied Econometrics*, forthcoming. doi:10.1002/jae.2562.

Kongsted, Hans Christian. 2005. Testing the nominal-to-real transformation. *Journal of Econometrics* 124: 205–25.

Magnus, Jan R., and Heinz Neudecker. 2007. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. New York: Wiley.

Mosconi, Rocco, and Paolo Paruolo. 2017. Cointegration and error correction in I(2) vector autoregressive models: Identification, estimation and testing. Mimeo, Polictenico di Milano, Milano, Italy.

Nielsen, Heino Bohn, and Anders Rahbek. 2007. The likelihood ratio test for cointegration ranks in the I(2) model. *Econometric Theory* 23: 615–37.

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

Paruolo, Paolo. 2000a. Asymptotic efficiency of the two stage estimator in I(2) systems. *Econometric Theory* 16: 524–50.

Paruolo, Paolo. 2000b. On likelihood-maximizing algorithms for I(2) VAR models. Mimeo, University of Insubria, Varese, Italy.

Paruolo, Paolo. 2002. Asymptotic inference on the moving average impact matrix in cointegrated I(2) VAR systems. *Econometric Theory* 18: 673–90.

Rahbek, Anders, Hans Christian Kongsted, and Clara Jørgensen. 1999. Trend-stationarity in the I(2) cointegration model. *Journal of Econometrics* 90: 265–89.

Rothenberg, Thomas J. 1971. Identification in parametric models. *Econometrica* 39: 577–91.

Sargan, J. Denis. 1988. *Lectures on Advanced Econometric Theory* (edited by Meghnad Desai). Oxford: Basil Blackwell.

Srivastava, Muni S., and C. G. Kathri. 1979. *An Introduction to Multivariate Statistics*. New York: North Holland.