

Article

Maximum Likelihood Estimation of the I(2) Model under Linear Restrictions

Jurgen A. Doornik

Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, OX1 3UQ, UK; jurgen.doornik@nuffield.ox.ac.uk

Academic Editor: Katarina Juselius

Received: 27 February 2017; Accepted: 8 May 2017; Published: 15 May 2017

Abstract: Estimation of the I(2) cointegrated vector autoregressive (CVAR) model is considered. Without further restrictions, estimation of the I(1) model is by reduced-rank regression (Anderson (1951)). Maximum likelihood estimation of I(2) models, on the other hand, always requires iteration. This paper presents a new triangular representation of the I(2) model. This is the basis for a new estimation procedure of the unrestricted I(2) model, as well as the I(2) model with linear restrictions imposed.

Keywords: cointegration; I(2); vector autoregression; representation; maximum likelihood estimation; reduced rank regression; generalized least squares

JEL Classification: C32; C51; C61

1. Introduction

The I(1) model or cointegrated vector autoregression (CVAR) is now well established. The model is developed in a series of papers and books (see, e.g., Johansen (1988), Johansen (1991), Johansen (1995a), Juselius (2006)) and generally available in econometric software. The I(1) model is formulated as a rank reduction of the matrix of ‘long-run’ coefficients. The Gaussian log-likelihood is estimated by reduced-rank regression (RRR; see Anderson (1951), Anderson (2002)).

Determining the cointegrating rank only finds the cointegrating vectors up to a rank-preserving linear transformation. Therefore, the next step of an empirical study usually identifies the cointegrating vectors. This may be followed by imposing over-identifying restrictions. Common restrictions, i.e., the same restrictions on each cointegrating vector, can still be solved by adjusting the RRR estimation; see Johansen and Juselius (1990) and Johansen and Juselius (1992). Estimation with separate linear restrictions on the cointegrating vectors, or more general non-linear restrictions, requires iterative maximization. The usual approach is based on so-called switching algorithms; see Johansen (1995b) and Boswijk and Doornik (2004). The former proposes an algorithm that alternates between cointegrating vectors, estimating one while keeping the others fixed. The latter consider algorithms that alternate between the cointegrating vectors and their loadings: when one is kept fixed, the other is identified. The drawback is that these algorithms can be very slow and occasionally terminate prematurely. Doornik (2017) proposes improvements that can be applied to all switching algorithms.

Johansen (1995c) and Johansen (1997) extend the CVAR to allow for I(2) stochastic trends. These tend to be smoother than I(1) stochastic trends. The I(2) model implies a second reduced rank restriction, but this is now more complicated, and estimation under Gaussian errors can no longer be performed by RRR. The basis of an algorithm for maximum likelihood estimation is presented in Johansen (1997), with an implementation in Dennis and Juselius (2004).

The general approach to handling the I(2) model is to create representations that introduce parameters that vary freely without changing the nature of the model. This facilitates both the statistical analysis and the estimation.

The contributions of the current paper are two-fold. First, we present the triangular representation of the I(2) model. This is a new trilinear formulation with a block-triangular matrix structure at its core. The triangular representation provides a convenient framework for imposing linear restrictions on the model parameters. Next, we introduce several improved estimation algorithms for the I(2) model. A simulation experiment is used to study the behaviour of the algorithms.

Notation

Let α ($p \times r$) be a matrix with full column rank $r, r \leq p$. The perpendicular matrix α_{\perp} ($p \times p-r$) has $\alpha'_{\perp}\alpha = 0$. The orthogonal complement α_{\perp} has $\alpha'_{\perp}\alpha = 0$ with the additional property that $\alpha'_{\perp}\alpha_{\perp} = I_{p-r}$. Define $\tilde{\alpha} = \alpha(\alpha'\alpha)^{-1/2}$ and $\bar{\alpha} = \alpha(\alpha'\alpha)^{-1}$. Then, $(\tilde{\alpha} : \alpha_{\perp})$ is a $p \times p$ orthogonal matrix, so $I_p = \tilde{\alpha}\tilde{\alpha}' + \alpha_{\perp}\alpha'_{\perp} = \bar{\alpha}\alpha' + \alpha_{\perp}\alpha'_{\perp} = \alpha\bar{\alpha}' + \alpha_{\perp}\alpha'_{\perp}$.

The (thin) singular value decomposition (SVD) of α is $\alpha = UWV'$, where $U(p \times r), V(r \times r)$ are orthogonal: $U'U = V'V = VV' = I_r$, and W is a diagonal matrix with the ordered positive singular values on the diagonal. If $\text{rank}(\alpha) = s < r$, then the last $r - s$ singular values are zero. We can find $\alpha_{\perp} = U_2$ from the SVD of the square matrix $(\alpha : 0) = (U_1 : U_2)WV' = (U_1W_1V'_1 : 0)$.

The (thin) QR factorization of α with pivoting is $\alpha P = QR$, with $Q(p \times r)$ orthogonal and R upper triangular. This pivoting is the reordering of columns of α to better handle poor conditioning and singularity, and is captured in P , as discussed [Golub and Van Loan \(2013, §5.4.2\)](#).

The QL decomposition of A can be derived from the QR decomposition of JAJ : $JAJ = QR$, so $A = JAJJ = JQJJZJ = Q'L$. J is the exchange matrix, which is the identity matrix with columns in reverse order: premultiplication reverses rows; postmultiplication reverses columns; and $JJ = I$.

Let $\bar{\alpha} = \Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}$, then $\alpha'_{\perp}\Omega\bar{\alpha} = 0$.

Finally, $a \leftarrow b$ assigns the value of b to a .

2. The I(2) Model

The vector autoregression (VAR) with p dependent variables and $m \geq 1$ lags:

$$y_t = A_1y_{t-1} + \dots + A_my_{t-m} + \Phi x_t^U + \epsilon_t, \quad \epsilon_t \sim \text{IIN}_p[0_p, \Omega],$$

for $t = 1, \dots, T$, and with $y_j, j = -m + 1, \dots, 0$ fixed and given, can be written in equilibrium correction form as:

$$\Delta y_t = \Pi_y y_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{m-1} \Delta y_{t-m+1} + \Phi x_t^U + \epsilon_t,$$

without imposing any restrictions. The I(1) cointegrated VAR (CVAR) imposes a reduced rank restriction on $\Pi_y(p \times p)$: $\text{rank}\Pi_y = r$; see, e.g., [Johansen and Juselius \(1990\)](#), [Johansen \(1995a\)](#).

With $m \geq 2$, the model can be written in second-differenced equilibrium correction form as:

$$\Delta^2 y_t = \Pi_y y_{t-1} - \Gamma_y \Delta y_{t-1} + \Psi_1 \Delta^2 y_{t-1} + \dots + \Psi_{m-2} \Delta^2 y_{t-m+2} + \Phi x_t^U + \epsilon_t. \tag{1}$$

The I(2) CVAR involves an additional reduced rank restriction:

$$\text{rank}(\alpha'_{\perp} \Gamma_y \beta_{y,\perp}) = s,$$

where $\alpha'_{\perp}\alpha = 0$. The two rank restrictions can be expressed more conveniently in terms of products of matrices with reduced dimensions:

$$\Pi_y = \alpha\beta'_y, \tag{2}$$

$$\alpha'_{\perp} \Gamma_y \beta_{y,\perp} = \zeta\eta'_y, \tag{3}$$

where α and β_y are $p \times r$ matrices. The second restriction needs rank s , so ξ and η_y are a $(p - r) \times s$. This requires that the matrices on the right-hand side of (2) and (3) have full column rank. The number of I(2) trends is $s_2 = p - r - s$.

The most relevant model in terms of deterministic allows for linearly trending behaviour: $\Phi x_t^U = \mu_0 + \mu_1 t$. Using the representation theorem of Johansen (1992) and assuming $E[y_t] = a + bt$ imply:

$$\mu_1 = -\alpha\beta'_y b, \tag{4}$$

$$\mu_0 = -\alpha\beta'_y a + \Gamma_y b, \tag{5}$$

which restricts and links μ_0 and μ_1 ; we see that $\alpha'_\perp \mu_1 = 0$ and $\alpha'_\perp \mu_0 = \alpha'_\perp \Gamma_y b$.

2.1. The I(2) Model with a Linear Trend

The model (1) subject to the I(1) and I(2) rank restrictions (2) and (3) with $\Phi x_t^U = \mu_0 + \mu_1 t$, subject to (4) and (5) can be written as:

$$\Delta^2 y_t = \alpha\beta' \begin{pmatrix} y_{t-1} \\ t \end{pmatrix} - \Gamma \begin{pmatrix} \Delta y_{t-1} \\ 1 \end{pmatrix} + \Psi_1 \Delta^2 y_{t-1} + \dots + \Psi_{m-2} \Delta^2 y_{t-m+2} + \epsilon_t, \tag{6}$$

subject to:

$$\alpha'_\perp \Gamma \beta_\perp = \xi \eta', \tag{7}$$

where β is $p_1 \times r$, Γ is $p \times p_1$ and η is $(p_1 - r) \times s$. In this case, $p_1 = p + 1$. Because α is the leading term in (4), we can extend β_y by introducing $\beta'_c = -\beta'_y b$, so $\beta' = (\beta'_y : \beta'_c)$. Furthermore, Γ has been extended to $\Gamma = (\Gamma_y : \Gamma_c) = (\Gamma_y : -\mu_0)$.

To see that (6) and (7) remains the same I(2) model, consider $\alpha'_\perp \Gamma_c$ and insert $I_p = \beta_y \bar{\beta}'_y + \beta_{y\perp} \beta'_{y\perp}$:

$$\alpha'_\perp \Gamma_c = -\alpha'_\perp \Gamma_y \bar{\beta}_y \beta'_y b - \alpha'_\perp \Gamma_y \beta_{y\perp} \beta'_{y\perp} b = \alpha'_\perp \Gamma_y \bar{\beta}_y \beta'_c - \xi \eta'_y \beta'_{y\perp} b = \alpha'_\perp \Gamma_y \bar{\beta}_y \beta'_c + \xi \eta'_c.$$

Using the perpendicular matrix:

$$\beta_\perp = \begin{pmatrix} \beta_{y\perp} & -\bar{\beta}_y \beta'_c \\ 0 & 1 \end{pmatrix}$$

we see that the rank condition is unaffected:

$$\alpha'_\perp \Gamma \beta_\perp = (\alpha'_\perp \Gamma_y \beta_{y\perp} : \xi \eta'_c) = (\alpha'_\perp \Gamma_y \beta_{y\perp} : \alpha'_\perp [-\Gamma_y \bar{\beta}_y \beta'_c + \Gamma_c]) = \xi (\eta'_y : \eta'_c).$$

A more general formulation allows for restricted deterministic and weakly exogenous variables x_t^R and unrestricted variables x_t^U :

$$\begin{aligned} \Delta^2 y_t &= \Pi \begin{pmatrix} y_{t-1} \\ x_{t-1}^R \end{pmatrix} - \Gamma \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1}^R \end{pmatrix} + \Psi_1 \Delta^2 y_{t-1} + \dots + \Psi_{m-2} \Delta^2 y_{t-m+2} + \Phi x_t^U + \epsilon_t, \\ &= \Pi w_{2t} - \Gamma w_{1t} + \Psi w_{3t} + \epsilon_t, \end{aligned}$$

where $\Delta^2 x_t^R$, and its lags are contained in x_t^U ; this in turn, is subsumed under $w_{3t} = (\Delta^2 y'_{t-1}, \dots, x_t^{U'})'$. The number of variables in x_t^R is $p_1 - p$, so Π and Γ always have the same dimensions. Ψ is unrestricted, which allows it to be concentrated out by regressing all other variables on w_{3t} :

$$z_{0t} = \alpha\beta' z_{2t} - \Gamma z_{1t} + \epsilon_t, \quad \epsilon_t \sim \text{IIN}_p[0_p, \Omega]. \tag{8}$$

To implement likelihood-ratio tests, it is necessary to count the number of restrictions:

$$\begin{aligned} \text{restrictions on } \Pi &: && (p-r)(p_1-r) \text{ restrictions,} \\ \text{restrictions on } \Gamma &: && (p-r-s)(p_1-r-s) = s_2 s_2^* \text{ restrictions,} \end{aligned}$$

defining $s_2^* = p_1 - r - s$. The restrictions on Π follow from the representation. Several representations of the I(2) model have been introduced in the literature to translate the implicit non-linear restriction (3) on Γ into an explicit part of the model. These representations reveal the number of restrictions imposed on Γ , as is shown below.

First, we introduce the new triangular representation.

2.2. The Triangular Representation

Theorem 1. Consider the model:

$$z_{0t} = \Pi z_{2t} - \Gamma z_{1t} + \epsilon_t,$$

with rank restrictions $\Pi = \alpha\beta'$ and $\alpha'_\perp \Gamma \beta_\perp = \zeta\eta'$ where α is a $p \times r$ matrix, β is $p_1 \times r$, ζ is $(p-r) \times s$, η is $(p_1-r) \times s$. This can be written as:

$$z_{0t} = AWB'z_{2t} - AVB'z_{1t} + \epsilon_t, \tag{9}$$

where:

$$W = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ W_{11} & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} V_{31} & 0 & 0 \\ V_{21} & V_{22} & 0 \\ V_{11} & V_{12} & V_{13} \end{pmatrix}. \tag{10}$$

A, B, W_{11}, V_{22} are full rank matrices. A is $p \times p$, and B is $p_1 \times p_1$; moreover, A, B and the nonzero blocks in W and V are freely varying. A and B are partitioned as:

$$A = (A_2 : A_1 : A_0), \quad B = (B_0 : B_1 : B_2),$$

where the blocks in A have s_2, s, r columns respectively; for B , this is: r, s, s_2^* ; $p_1 = r + s + s_2^*$. W and V are partitioned accordingly.

Proof. Write $\tilde{\alpha} = \alpha(\alpha'\alpha)^{-1/2}$, such that $\tilde{\alpha}'\tilde{\alpha} = I_r$. Construct A and B as:

$$A = (\alpha_\perp \zeta_\perp : \alpha_\perp \tilde{\zeta} : \tilde{\alpha}), \quad B = (\tilde{\beta} : \beta_\perp \tilde{\eta} : \beta_\perp \eta_\perp).$$

Now, $A'A = I$ and $B'B = I$. $A(p \times p)$ and $B(p_1 \times p_1)$ are full rank by design. Define $V = A'GB$:

$$V = \begin{pmatrix} \zeta'_\perp \alpha'_\perp \Gamma \tilde{\beta} & \zeta'_\perp \alpha'_\perp \Gamma \beta_\perp \tilde{\eta} & \zeta'_\perp \alpha'_\perp \Gamma \beta_\perp \eta_\perp \\ \tilde{\zeta}' \alpha'_\perp \Gamma \tilde{\beta} & \tilde{\zeta}' \alpha'_\perp \Gamma \beta_\perp \tilde{\eta} & \tilde{\zeta}' \alpha'_\perp \Gamma \beta_\perp \eta_\perp \\ \tilde{\alpha}' \Gamma \tilde{\beta} & \tilde{\alpha}' \Gamma \beta_\perp \tilde{\eta} & \tilde{\alpha}' \Gamma \beta_\perp \eta_\perp \end{pmatrix} = \begin{pmatrix} V_{31} & 0 & 0 \\ V_{21} & V_{22} & 0 \\ V_{11} & V_{12} & V_{13} \end{pmatrix}.$$

$V_{22} = (\zeta'_\perp \zeta_\perp)^{1/2} (\eta'_\perp \eta_\perp)^{1/2}$ is a full rank $s \times s$ matrix. The zero blocks in V arise because, e.g., $\zeta'_\perp \alpha'_\perp \Gamma \beta_\perp = \zeta'_\perp \zeta_\perp \eta'_\perp = 0$. Trivially:

$$\Pi = \alpha\beta' = A \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ W_{11} & 0 & 0 \end{pmatrix} B' = AWB'.$$

$W_{11} = (\alpha'\alpha)^{1/2} (\beta'\beta)^{1/2}$ is a full rank $r \times r$ matrix. Both W and V are $p \times p_1$ matrices. Because A and B are each orthogonal: $\Gamma = AA'\Gamma BB' = AVB'$.

The QR decomposition shows that a full rank square matrix can be written as the product of an orthogonal matrix and a triangular matrix. Therefore, $AVB' = AL_a L_a^{-1} V L_b L_b^{-1} B' = A_* V_* B'_*$ preserves

the structure in V_* when L_a, L_b are lower triangular, as well as that in W_* . This shows that (9) holds for any full rank A and B , and the orthogonality can be relaxed.

Therefore, any model with full rank matrices A and B , together with any W, V that have the zeros as described above, satisfies the I(2) rank restrictions. We obtain the same model by restricting A and B to be orthogonal. □

When Γ is restricted only by the I(2) condition: $\text{rank}\Gamma = r + s + \min(r, s_2)$. Then, V varies freely, except for the zero blocks, and the I(2) restrictions are imposed through the trilinear form of (9). $\Gamma = 0$ implies $V = 0$. Another way to have $s = 0$ is $\Gamma = (\alpha : 0)G$; in that case, $V \neq 0$.

The s_2 restrictions on the intercept (5) can be expressed as $A'_2(\mu_0 - \mu_c) = 0$, using $\mu_c = \Gamma_y \bar{\beta}_y \beta'_c$, or $\mu_0 = (A_1 : A_0)v + \mu_c$, for a vector v of length $r + s$.

2.3. Obtaining the Triangular Representation

The triangular representation shows that the I(2) model can be written in trilinear form:

$$z_{0t} = AWB'z_{2t} - AVB'z_{1t} + \epsilon_t,$$

where A and B are freely varying, provided W and V have the appropriate structure.

Consider that we are given α, β, Γ of an I(2) CVAR with rank indices r, s and wish to obtain the parameters of the triangular representation. First compute $\alpha'_\perp \Gamma \beta_\perp = \zeta \eta'$, which can be done with the SVD, assuming rank s . From this, compute A and B :

$$A = (A_2 : A_1 : A_0) = (\alpha_\perp \zeta_\perp : \alpha_\perp \tilde{\zeta} : \alpha), \quad B = (B_0 : B_1 : B_2) = (\beta : \beta_\perp \tilde{\eta} : \beta_\perp \eta_\perp).$$

Then, $V = A^{-1} \Gamma B^{-1'}$. Because Γ satisfies the I(2) rank restriction, V will have the corresponding block-triangular structure.

It may be of interest to consider which part of the structure can be retrieved in the case where $\text{rank}(\Pi) = r$, but $\text{rank}(\alpha'_\perp \Gamma \beta_\perp) = p - r$, while it should be s . This would happen when using I(1) starting values for I(2) estimation. The off anti-diagonal blocks of zeros:

$$V^* = \begin{pmatrix} V_{31}^* & V_{32}^* & V_{33}^* \\ V_{21} & V_{22} & V_{23}^* \\ V_{11} & V_{12} & V_{13}^* \end{pmatrix} \rightarrow \begin{pmatrix} V_{31} & 0 & V_{33} \\ V_{21} & V_{22} & 0 \\ V_{11} & V_{12} & V_{13} \end{pmatrix} = V \tag{11}$$

can be implemented with two sweep operations:

$$\begin{pmatrix} I_{s_2} & -V_{32}^* V_{22}^{-1} & 0 \\ 0 & I_s & 0 \\ 0 & 0 & I_r \end{pmatrix} V^* \begin{pmatrix} I_r & 0 & 0 \\ 0 & I_s & -V_{22}^{-1} V_{23}^* \\ 0 & 0 & I_{s_2}^* \end{pmatrix}.$$

The offsetting operations affect A_1 and B_1 only, so Π and Γ are unchanged. However, we cannot achieve $V_{33} = 0$ in a similar way, because it would remove the zeros just obtained. The V_{33} block has dimension $s_2 s_2^*$ and represents the number of restrictions imposed on Γ in the I(2) model. Similarly, the anti-diagonal block of zeros in W captures the restrictions on Π .

Note that the $r \times s_2^*$ block V_{13} can be made lower triangular. Write the column partition of V as $(V_{\cdot 1} : V_{\cdot 2} : V_{\cdot 3})$, and use $V_{13} = LQ$ to replace $V_{\cdot 3}$ by $V_{\cdot 3} Q'$ and B_2 by $B_2 Q'$. When $r < s_2^*$, the rightmost $s_2^* - r$ columns of L will be zero, and the corresponding columns of B_2 are not needed to compute Γ . This part can then be omitted from the likelihood evaluation. This is an issue when we propose an estimation procedure in §4.2.1.

2.4. Restoring Orthogonality

Although A and B are freely varying, interpretation may require orthogonality between column blocks. The column blocks of A are in reverse order from B to make V and W block lower triangular. As a consequence, multiplication of V or W from either side by a lower triangular matrix preserves their structure. This allows for the relaxation of the orthogonality of A and B , but also enables us to restore it again.

To restore orthogonality, let $\Gamma = A_*V_*B_*'$, where A_*, B_* are not orthogonal, but with V_* block-triangular. Now, use the QL decomposition to get $A_* = AL$, with A orthogonal and L lower triangular. Use the QR decomposition to get $B_* = BR$, with B orthogonal and R upper triangular. Then, $A_*V_*B_*' = ALV_*R'B' = AVB'$ with the blocks of zeros in V preserved. $A_*W_*B_*'$ must be adjusted accordingly. When β is restricted, B_0 cannot be modified like this. However, we can still adjust $(A_2 : A_1) = A_*$ to get $A_*'A_* = I_{p-r}$ and $A_0'A_* = 0$; with similar adjustments to $(B_1 : B_2)$.

The orthogonal version is convenient mathematically, but for estimation, it is preferable to use the unrestricted version. We do not distinguish through notation, but the context will state when the orthogonal version is used.

2.5. Identification in the Triangular Representation

The matrices A and B are not identified without further restrictions. For example, rescaling α and β as in $\alpha W_{11}\beta' = \alpha'c^{-1}cW_{11}dd^{-1}\beta = \alpha^*cW_{11}d\beta^{*'}$ can be absorbed in V :

$$\begin{pmatrix} V_{31}d & 0 & 0 \\ V_{21}d & V_{22} & 0 \\ cV_{11}d & cV_{12} & cV_{13} \end{pmatrix}.$$

When β is identified, W_{11} remains freely varying, and we can, e.g., set $c = W_{11}^{-1}$. However, it is convenient to transform to $W_{11} = I$, so that A_0 and B_0 correspond to α and β . This prevents part of the orthogonality, in the sense that $A_0'A_0 \neq I$ and $B_0'B_0 \neq I$.

The following scheme identifies A and B , under the assumption that B_0 is already identified through prior restrictions.

1. Orthogonalize to obtain $A_0'A_1 = 0, A_0'A_2 = 0, A_1'A_2 = 0$.
2. Choose s full rank rows from B_1 , denoted M_{B_1} , and set $B_1 \leftarrow B_1M_{B_1}^{-1}$. Adjust V accordingly.
3. Do the same for $B_2 \leftarrow B_2M_{B_2}^{-1}$.
4. Set $A_1 \leftarrow A_1V_{22}, V_{21} \leftarrow V_{22}^{-1}V_{21}$ and $V_{22} \leftarrow I$.
5. $A_2 \leftarrow A_2M_{A_2}^{-1}$.

The ordering of columns inside A_i, B_i is not unique.

3. Relation to Other Representations

Two other formulations of the I(2) model that are in use are the so-called τ and δ representations. All representations implement the same model and make the rank restrictions explicit. However, they differ in their definitions of freely-varying parameters, so may facilitate different forms of analysis, e.g., asymptotic analysis, estimation or the imposition of restrictions. The different parametrizations may also affect economic interpretations.

3.1. τ Representation

Johansen (1997) transforms (8) into the τ -representation:

$$z_{0t} = \alpha (q'\tau'z_{2t} + \psi'z_{1t}) + w\kappa'\tau'z_{1t} + \epsilon_t, \tag{12}$$

where $\varrho(p_1 \times r+s)$ is used to recover β : $\beta = \tau\varrho$. The parameters $(\alpha, \varrho, \tau, \psi, \kappa)$ vary freely. If we normalize on $\varrho' = (I_r : 0)$ and adjust κ, τ accordingly, then $\tau = (\beta : \beta_1)$, and:

$$z_{0t} = \alpha (\beta' z_{2t} + \psi' z_{1t}) + w\kappa' \tau' z_{1t} + \epsilon_t.$$

We shall derive the τ representation. The first step is to define a transformation of $\epsilon_t \sim N[0, \Omega]$:

$$\begin{pmatrix} \alpha_{\perp} & \bar{\alpha} \end{pmatrix}' \epsilon_t \sim N \left[0, \begin{pmatrix} \alpha'_{\perp} \Omega \alpha_{\perp} & 0 \\ 0 & (\alpha' \Omega^{-1} \alpha)^{-1} \end{pmatrix} \right]. \tag{13}$$

This splits the p -variate systems into two independent parts. The first has any terms with leading α knocked out, while the second has all leading α 's cancelled. The inverse transformation is given by: $(\alpha_{\perp} : \bar{\alpha})^{-1} = (\alpha_{\perp} - \alpha \bar{\alpha}' \alpha_{\perp} : \alpha)' = (w : \alpha)'$.

The next step is to apply (13) to (8) to create two independent systems and insert $I_p = \bar{\beta}\beta' + \beta_{\perp}\beta'_{\perp}$ in the 'marginal' equation:

$$\begin{pmatrix} \alpha_{\perp} & \bar{\alpha} \end{pmatrix}' z_{0t} = \begin{cases} -\alpha'_{\perp} \Gamma (\bar{\beta}\beta' + \beta_{\perp}\beta'_{\perp}) z_{1t} + \epsilon_{1t} & = \kappa' (\beta : \beta_{\perp} \eta)' z_{1t} + \epsilon_{1t}, \\ \beta' z_{2t} - \bar{\alpha}' \Gamma z_{1t} + \epsilon_{2t} & = \beta' z_{2t} + \psi' z_{1t} + \epsilon_{2t}. \end{cases} \tag{14}$$

where $\psi' = -\bar{\alpha}' \Gamma$ and $\kappa' = -(\alpha'_{\perp} \Gamma \bar{\beta} : \zeta)$ are freely varying. Removing the transformation:

$$z_{0t} = w' \kappa' (\beta : \beta_{\perp} \eta)' z_{1t} + \alpha' (\beta' z_{2t} + \psi' z_{1t}) + \epsilon_t$$

and introducing the additional parameters $\tau = (\beta : \beta_{\perp} \eta)$ and ϱ completes the τ -representation (12). Table 1 provides definitions of the parameters that are used (cf. Johansen (1997, Tables 1 and 2)).

Table 1. Definitions of the symbols used in the τ and δ representations of the I(2) model.

Definition	Dimension
$\tau = (\beta : \beta_{\perp} \eta)$ when $\varrho' = (I : 0)$	$p_1 \times (r + s)$
$\tau_{\perp} = \beta_{\perp} \eta_{\perp}$	$p_1 \times s_2^*$
$\psi = -(\bar{\alpha}' \Gamma)'$	$p_1 \times r$
$\kappa' = -\alpha'_{\perp} \Gamma \bar{\tau} = -(\alpha'_{\perp} \Gamma \bar{\beta} : \zeta) = (\kappa_1 : \kappa_2)'$	$(p - r) \times (r + s)$
$\delta = -\bar{\alpha}' \Gamma \tau_{\perp}$	$r \times s_2^*$
$\zeta = -\Gamma \bar{\tau} = (\zeta_1 : \zeta_2)$	$p \times (r + s)$
$w = \alpha_{\perp} - \alpha \bar{\alpha}' \alpha_{\perp} = \Omega \alpha_{\perp} (\alpha'_{\perp} \Omega \alpha_{\perp})^{-1} = \bar{\alpha}_{\perp}$	$p \times (p - r)$
$d = \tau_{\perp} \delta'$	$p_1 \times r$
$e = \tau \zeta'$	$p_1 \times p$

Corollary 1. Triangular representation (9) is equivalent to the τ -representation (12) when $A'_0(A_2 : A_1) = 0$.

Proof. Write $A_* = (A_2 : A_1)$, so $A = (A_* : A_0)$. First, the system (9) is premultiplied by $A^{-1} = \bar{A}'$ and subsequently with a lower triangular matrix L to create two independent subsystems. The matrix L and its inverse are given by:

$$L = \begin{pmatrix} I_{p-r} & 0 \\ -\bar{A}'_0 w_* & I_r \end{pmatrix}, \quad L^{-1} = \begin{pmatrix} I_{p-r} & 0 \\ \bar{A}'_0 w_* & I_r \end{pmatrix},$$

where $w_* = \Omega \bar{A}_* (\bar{A}'_* \Omega \bar{A}_*)^{-1}$, cf. (14). Because $A'_0 A_* = 0$, we have that $A_* + A_0 \bar{A}'_0 w_* = A_* + (I - A_* \bar{A}'_*)' w_* = w_*$, so $AL^{-1} = (w_* : A_0)$. Furthermore: $LW = W$. The identity matrix $L^{-1}L$ can also be inserted directly in (9):

$$\begin{aligned} z_{0t} &= A_0 W_{11} B'_0 z_{2t} - (w_* : A_0) L V B' z_{1t} + \epsilon_t \\ &= A_0 W_{11} B'_0 z_{2t} - w_* \begin{pmatrix} V_{31} & 0 \\ V_{21} & V_{22} \end{pmatrix} (B_0 : B_1)' z_{1t} + A_0 \bar{A}'_0 w_* (V_{11} : V_{12} : V_{13}) B' z_{1t} + \epsilon_t \\ &= \alpha [\beta' z_{2t} + \psi' z_{1t}] + w \kappa' \tau' z_{1t} + \epsilon_t, \end{aligned}$$

where $\psi' = -\bar{A}'_0 w_* (V_{11} : V_{12} : V_{13}) B'$ and $w \kappa' = -w_* \begin{pmatrix} V_{31} & 0 \\ V_{21} & V_{22} \end{pmatrix}$. \square

3.2. δ Representation

Paruolo and Rahbek (1999) and Paruolo (2000a) use the δ representation:

$$z_{0t} = \alpha (\beta' z_{2t} + \delta \tau'_{\perp} z_{1t}) + \zeta \tau' z_{1t} + \epsilon_t. \tag{15}$$

Here, $(\alpha, \delta, \zeta, \tau = [\beta : \beta_1])$ vary freely. To derive the δ representation, use $\bar{\tau}' + \tau_{\perp} \tau'_{\perp} = I_{r+s}$:

$$\begin{aligned} -\Gamma \bar{\tau}' &= -(\Gamma \bar{\beta} : \Gamma \beta_{\perp} \bar{\eta}) \tau' = (\zeta_1 : \zeta_2) \tau' = \zeta \tau', \\ -\Gamma \tau_{\perp} \tau'_{\perp} &= -\alpha \bar{\alpha}' \Gamma \tau_{\perp} \tau'_{\perp} - \alpha_{\perp} \alpha'_{\perp} \Gamma \tau_{\perp} \tau'_{\perp} = \alpha \delta \tau'_{\perp}, \end{aligned}$$

and insert in (8). The term with $\alpha'_{\perp} \Gamma \tau_{\perp}$ disappears because $\tau_{\perp} = \beta_{\perp} \eta_{\perp}$, so $\alpha'_{\perp} \Gamma \tau_{\perp} = \zeta \eta' \eta_{\perp} = 0$.

When β is identified both $\delta \tau'_{\perp}$ and $\zeta \tau'$ are unique, but not yet ζ or δ . In the τ representation, the variable ψ is also unique with ρ chosen as $(I : 0)'$ and β identified. Table 2 relates the τ , δ and triangular representations.

Table 2. Links between symbols used in the representations of the I(2) model, assuming $W_{11} = I_r$ and $a'_{\perp} a_{\perp} = I$.

$-\Gamma = \alpha \psi' + w \kappa' \tau' = \alpha \delta \tau'_{\perp} + \zeta \tau' = \alpha d' + e'$		
ζ	$= \alpha \psi' \bar{\tau} + w \kappa'$	(from $\Gamma \bar{\tau}$)
d'	$= \psi' \tau_{\perp} \tau'_{\perp}$	(from $\Gamma \tau_{\perp}$)
κ'	$= \alpha'_{\perp} \zeta$	(from $\alpha'_{\perp} \Gamma$)
ψ'	$= d' + \bar{\alpha}' \zeta \tau'$	(from $\bar{\alpha}' \Gamma$)
α	$= A_0$	
β	$= B_0$	
d'	$= -V_{13} B'_2$	
e'	$= -A(V_{.1} : V_{.2})(B_0 : B_1)'$	
τ	$= (B_0 : B_1)$	

Corollary 2. Triangular representation (9) is equivalent to the δ -representation (15) when $B'_2(B_0 : B_1) = 0$.

Proof. Write $B_2 = \tau_{\perp}$ and $(B_0 : B_1) = \tau$. Using the column partitioning if $V = (V_{.1} : V_{.2} : V_{.3})$: $\Gamma \bar{\tau} = AVB' \bar{\tau} = A(V_{.1} : V_{.2})$. From (9):

$$\begin{aligned} z_{0t} &= A_0 [W_{11} B'_0 z_{2t} - V_{13} \tau'_{\perp} z_{1t}] - A(V_{.1} : V_{.2}) \tau' z_{1t} + \epsilon_t \\ &= A_0 [W_{11} B'_0 z_{2t} - V_{13} \tau'_{\perp} z_{1t}] - \Gamma \bar{\tau}' \tau' z_{1t} + \epsilon_t = \alpha [\beta' z_{2t} + \delta \tau'_{\perp} z_{1t}] + \zeta \tau' z_{1t} + \epsilon_t. \end{aligned}$$

\square

4. Algorithms for Gaussian Maximum Likelihood Estimation

Algorithms to estimate the Gaussian CVAR are usually alternating over sets of variables. In the cointegration literature, these are called switching algorithms, following [Johansen and Juselius \(1994\)](#).

The advantage of switching is that each step is easy to implement, and no derivatives are required. Furthermore, the partitioning circumvents the lack of identification that can occur in these models and which makes it harder to use Newton-type methods. The drawback is that progress is often slow, taking many iterations to converge. Occasionally, this will lead to premature convergence. Although the steps can generally be shown to be in a non-downward direction, this is not enough to show convergence to a stationary point. The work in [Doornik \(2017\)](#) documents the framework for the switching algorithms and also considers acceleration of these algorithms; both results are used here.

[Johansen \(1997, §8\)](#) proposes an algorithm based on the τ -representation, called τ -switching here. This is presented in detail in Appendix B. Two new algorithms are given next, the first based on the δ -representation, the second on the triangular representation. Some formality is required to describe the algorithms with sufficient detail.

4.1. δ -Switching Algorithm

The free parameters in the δ -representation (15) are $(\alpha, \delta, \zeta, \tau)$ with symmetric positive definite Ω . The algorithm alternates between estimating τ given the rest and fixing τ . The model for τ given the other parameters is linear:

1. To estimate $\tau = [\beta : \beta_1]$, rewrite (15) as:

$$z_{0t} = \alpha\beta'z_{2t} + \zeta_1\beta'z_{1t} + \zeta_2\beta'_1z_{1t} + \alpha dz_{1t} + \epsilon_t,$$

where d replaces $\delta\tau'_\perp$. Then, vectorize, using $\alpha\beta'z_{2t} = \text{vec}(z'_{2t}\beta\alpha') = (\alpha \otimes z'_{2t})\text{vec}\beta$:

$$z_{0t} = (\alpha \otimes z'_{2t} + \zeta_1 \otimes z'_{1t})\text{vec}\beta + (\zeta_2 \otimes z'_{1t})\text{vec}\beta_1 + (\alpha \otimes z'_{1t})\text{vec}(d') + \epsilon_t. \quad (16)$$

Given $\alpha, \zeta_1, \zeta_2, \Omega$, we can treat β, β_1 and d as free parameters to be estimated by generalized least squares (GLS). This will give a new estimate of τ .

We can treat d as a free parameter in (16). First, when $r \geq s_2^*$, δ has more parameters than τ_\perp . Second, when $r < s_2^*$, then Γ is reduced rank, and $s_2^* - r$ columns of τ_\perp are redundant. Orthogonality is recovered in the next step.

2. Given τ and derived τ_\perp , we can estimate α and δ by RRR after concentrating out $\tau'z_{1t}$. Introducing ρ with dimension $(r+s_2^*) \times r$ allows us to write (15) as:

$$z_{0t} = \alpha^*\rho' \begin{pmatrix} \beta'z_{2t} \\ \tau'_\perp z_{1t} \end{pmatrix} + \zeta\tau'z_{1t} + \epsilon_t. \quad (17)$$

RRR provides estimates of α^* and ρ' . Next, $\alpha^*\rho'$ is transformed to $\alpha(I_r : \delta)$, giving new estimates of α and δ . Finally, ζ can be obtained by OLS from (17) given α, δ, τ , and hence, Ω .

The RRR step is the same as used in [Dennis and Juselius \(2004\)](#) and [Paruolo \(2000b\)](#). However, the GLS step for τ is different from both. We have found that the specification of the GLS step can have a substantial impact on the performance of the algorithm.

For numerical reasons (see, e.g. [Golub and Van Loan \(2013, Ch.5\)](#)), we prefer to use the QR decomposition to implement OLS and RRR estimation rather than moment matrices. However, in iterative estimation, there are very many regressions, which would be much faster using precomputed moment matrices. As a compromise, we use precomputed 'data' matrices that are transformed by a QR decomposition. This reduces the effective sample size from T to $2p_1$. The regressions (16) and (17) can then be implemented in terms of the transformed data matrices; see Appendix A.

Usually, starting values of α and β are available from I(1) estimation. The initial τ is then obtained from the marginal equation of the τ -representation, (14a), written as:

$$\alpha'_{\perp} z_{0t} = \kappa' \tau' z_{1t} + \varepsilon_{1t} = \kappa'_1 \beta' z_{1t} + \xi \eta' \beta'_{\perp} z_{1t} + \varepsilon_{1t}. \tag{18}$$

RRR of $\alpha'_{\perp} z_{0t}$ on $\beta'_{\perp} z_{1t}$ corrected for $\beta' z_{1t}$ gives estimates of η , and so, τ .

δ -switching algorithm:

To start, set $k = 1$, and choose starting values $\alpha^{(0)}, \beta^{(0)}$, tolerance ε_1 and the maximum number of iterations. Compute $\tau_c^{(0)}$ from (18) and $\alpha^{(0)}, \delta^{(0)}, \zeta^{(0)}, \Omega^{(0)}$ from (17). Furthermore, compute $f^{(0)} = -\log |\Omega^{(0)}|$.

1. Get $\tau_c^{(k)}$ from (16); get the remaining parameters from (17).
2. Compute $f_c^{(k)} = -\log |\Omega_c^{(k)}|$.
3. Enter a line search for τ .

The change in τ is $\nabla = \tau_c^{(k)} - \tau^{(k-1)}$ and the line search find a step length λ with $\tau^{(k)} = \tau^{(k-1)} + \lambda \nabla$. Because only τ is varied, a GLS step is needed to evaluate the log-likelihood for each trial τ .

The line search gives new parameters with corresponding $f^{(k)}$.

- T. Compute the relative change from the previous iteration:

$$c^{(k)} = \frac{f^{(k)} - f^{(k-1)}}{1 + |f^{(k-1)}|}.$$

Terminate if:

$$|c^{(k)}| \leq \varepsilon_1 \text{ and } \max_{ij} \frac{|\Pi_{ij}^{(k)} - \Pi_{ij}^{(k-1)}|}{1 + |\Pi_{ij}^{(k-1)}|} \leq \varepsilon_1^{1/2}. \tag{19}$$

Else increment k , and return to Step 1. □

The subscript c indicates that these are candidate values that may be improved upon by the line search. The line search is the concentrated version, so the I(2) equivalent to the LBeta line search documented in Doornik (2017). This means that the function evaluation inside the line search needs to re-evaluate all of the other parameters as τ changes. Therefore, within the line search, we effectively concentrate out all other parameters.

Normalization of τ prevents the scale from growing excessively, and it was found to be beneficial to normalize in the first iteration every hundredth or when the norm of τ gets large. Continuous normalization had a negative impact in our experiments. Care is required when normalizing: if an iteration uses a different normalization from the previous one, then the line search will only be effective if the previous coefficients are adjusted accordingly.

The algorithm is incomplete without starting values, and it is obvious that a better start will lead to faster and more reliable convergence. Experimentation also showed that this and other algorithms struggled more in cases with $s = 0$. To improve this, we generate two initial values, follow three iterations of the τ -switching algorithms, then select the best for continuation. The details are in Appendix C.

4.2. MLE with the Triangular Representation

We set $W_{11} = I$. This tends to lead to slower convergence, but is required when both α and β are restricted. V_{22} is kept unrestricted: fewer restrictions seem to lead to faster convergence. All regressions use the data matrices that are pre-transformed by an orthogonal matrix as described in Appendix A. In the next section, we describe the estimation steps that can be repeated until convergence.

4.2.1. Estimation Steps

Equation (9) provides a convenient structure for an alternating variables algorithm. We can solve three separate steps by ordinary or generalized least squares for the case with orthogonal A :

1. B -step: estimate B , and fix A, V, W, Ω at A_c, V_c, W_c, Ω_c . The resulting model is linear in B :

$$z_{0t} = A_c W_c B' z_{2t} - A_c V_c B' z_{1t} + \varepsilon_t. \quad (20)$$

Estimation by GLS can be conveniently done as follows. Start with the Cholesky decomposition $\Omega_c = PP'$, and premultiply (20) by P^{-1} . Next take the QL decomposition of $P^{-1}A$ as $P^{-1}A = HL$ with L lower diagonal and H orthogonal. Now, premultiply the transformed system by H' :

$$H'P^{-1}z_{0t} = LW_c B' z_{2t} - LV_c B' z_{1t} + u_t = \tilde{W}_c B' z_{2t} - \tilde{V}_c B' z_{1t} + u_t,$$

which has the unit variance matrix. Because the structures of W and V are preserved, this approach can also be used in the next step.

2. V -step: estimate W, V , and fix A, B, Ω . This is a linear model in (W, V) , which can be solved by GLS as in the B step.
3. A -step: estimate A, Ω and fix W, V, B at W_c, V_c, B_c :

$$z_{0t} = A (W_c B_c' z_{2t} - V_c B_c' z_{1t}) + \varepsilon_t$$

This is the linear regression of z_{0t} on $W_c B_c' z_{2t} - V_c B_c' z_{1t}$.

The likelihood will not go down when making one update that consists of the three steps given above, provided V is full rank. If that does not hold, as noted at the end of §2.3, some part of B_2 or A_2 is not identified from the above expressions. To handle this, we make the following adjustments to steps 1 and 3:

- 1a. B -step: Remove the last $s_2^* - \min\{r, s_2^*\}$ columns from B, V and W , as they do not affect the log-likelihood. When iteration is finished, we can add columns of zeros back to W and V and the orthogonal complement of the reduced B to get a rectangular B .
- 3a. A -step: we wish to keep A invertible and, so, square during iteration. The missing part of A_2 is filled in with the orthogonal complement of the remainder of A after each regression. This requires re-estimation of V_1 by OLS.

4.2.2. Triangular-Switching Algorithm

The steps described in the previous section form the basis of an alternating variables algorithm:

Triangular-switching algorithm:

To start, set $k = 1$, and choose $\alpha^{(0)}, \beta^{(0)}$ and the maximum number of iterations. Compute $A^{(0)}, B^{(0)}, V^{(0)}, W^{(0)}$ and $\Omega^{(0)}$.

- 1.1 B -step: obtain $B^{(k)}$ from $A^{(k-1)}, V^{(k-1)}, W^{(k-1)}, \Omega^{(k-1)}$.
- 1.2 V step: obtain $W^{(k)}, V^{(k)}$ from $A^{(k-1)}, B^{(k)}, \Omega^{(k-1)}$.
- 1.3 A step: obtain $A^{(k)}, \Omega^{(k)}$ from $B^{(k)}, V^{(k)}, W^{(k)}$.
- 1.4 V_1 step: if necessary, obtain new $V_1^{(k)}$ from $A^{(k)}, B^{(k)}, V_2^{(k)}, V_3^{(k)}, W^{(k)}$.
- 2... As steps 2,3,T from the δ -switching algorithm. In this case, the line search is over all of the parameters in A, B, V . □

The starting values are taken as for the δ -switching algorithm; see Appendix C. This means that two iterations of δ -switching are taken first, using only restrictions on β .

4.3. Linear Restrictions

4.3.1. Delta Switching

Estimation under linear restrictions on β or τ of the form:

$$\beta = (H_1\phi_1 : \dots : H_r\phi_r) \text{ or } \tau = (H_1\phi_1 : \dots : H_{r+s}\phi_{r+s})$$

can be done by adjusting the GLS step in §4.1. However, estimation of α is by RRR, which is not so easily adjusted for linear restrictions. Restricting δ requires replacing the RRR step by regression conditional on δ , which makes the algorithm much slower. Estimation under $\delta = 0$, which implies $d = 0$, is straightforward.

4.3.2. Triangular Switching

Triangular switching avoids RRR, and restrictions on $\beta = B_0$ or $\tau = (B_0 : B_1)$ can be implemented by adjusting the B -step. In general, we can test restrictions of the form:

$$B = (H_1\phi_1 : \dots : H_{p_1}\phi_{p_1}) \text{ and } A = (G_1\theta_1 : \dots : G_p\theta_p).$$

Such linear restrictions on the columns of A and B are a straightforward extension of the GLS steps described above.

Estimation without multi-cointegration is also feasible. Setting $\delta = 0$ corresponds to $V_{13} = 0$ in the triangular representation. This amounts to removing the last s_2^* columns from B, V, W . Boswijk (2010) shows that the test for $\delta = 0$ has an asymptotic $\chi^2(rs_2^*)$ distribution.

Paruolo and Rahbek (1999) derives conditions for weak exogeneity in (15). They decompose this into three sub-hypotheses: $\mathcal{H}_0: b'\alpha = 0$, $\mathcal{H}_1: b'(\alpha \perp \xi) = 0$, $\mathcal{H}_2: b'\zeta_1 = 0$. These restrictions, taking $b = e_{p,i}$, where $e_{p,i}$ is the i -th column of I_p , correspond to a zero right-hand side in a particular equation in the triangular representation. First is $e'_{p,i}A_0 = 0$ creating a row of zeros in AWB' . Next is $e'_{p,i}A_1 = 0$, which extends the row of zeros. However, A must be full rank, so the final restriction must be imposed on V as $e'_{p,i}AV = (e'_{p,i}A_2 : 0 : 0)V = 0$, expressed as $e'_{p,i}A_2V_{31} = 0$. Paruolo and Rahbek (1999) shows that the combined test for a single variable has an asymptotic $\chi^2(2r + s)$ distribution.

5. Comparing Algorithms

We have three algorithms that can be compared:

1. The δ -switching algorithm, §4.1, which can handle linear restrictions on β or τ .
2. The triangular-switching algorithm proposed in §4.2.2. This can optionally have linear restrictions on the columns of A or B .
3. The improved τ -switching algorithm, Appendix B, implemented to allow for common restrictions on τ .

These algorithms, as well as two pre-existing ones, have been implemented in Ox 7 Doornik (2013).

The comparisons are based on a model for the Danish data (five variables: $m_3 = \log$ real money, $y = \log$ real GDP, $\Delta p = \log$ GDP deflator, and r_m, r_b , two interest rates); see Juselius (2006, §4.1.1). This has two lags in the VAR, with an unrestricted constant and restricted trend for the deterministic terms, i.e., specification H_I . The sample period is 1973(3) to 2003(1). First computed is the I(2) rank test table.

Table 3 records the number of iterations used by each of the algorithms; this is closely related to the actual computational time required (but less machine specific). All three algorithms converge rapidly to the same likelihood value. Although τ switching takes somewhat fewer iterations, it tends to take a bit more time to run than the other two algorithms. The new triangular I(2) switching procedure is largely competitive with the new δ -switching algorithm.

Table 3. Estimation of all I(2) models by τ, δ and triangular switching; all using the same starting value procedure. Number of iterations to convergence for $\epsilon_1 = 10^{-14}$.

$r \setminus s_2$	τ Switching				δ Switching				Triangular Switching			
	4	3	2	1	4	3	2	1	4	3	2	1
1	19	25	36	34	15	24	37	30	31	31	39	32
2		18	32	25		18	32	34		22	27	50
3			37	23			42	38			50	59
4				29				28				85

To illustrate the advances made with the new algorithms, we report in Table 4 how the original τ -switching, as well as the CATS2 version of δ -switching performed. CATS2, [Dennis and Juselius \(2004\)](#), is a RATS package for the estimation of I(1) and I(2) models, which uses a somewhat different implementation of an I(2) algorithm that is also called δ -switching. The number of iterations of that CATS 2 algorithm is up to 200-times higher than that of the new algorithms, which are therefore much faster, as well as more robust and reliable.

Table 4. Estimation of all I(2) models by old versions of τ, δ switching. Number of iterations to convergence for $\epsilon_1 = 10^{-14}$.

$r \setminus s_2$	Old τ Switching				CATS2 Switching			
	4	3	2	1	4	3	2	1
1	126	198	338	201	5229	8329	8516	5371
2		79	211	229		7234	709	861
3			483	237			550	432
4				4851				5771

6. A More Detailed Comparison

A Monte Carlo experiment is used to show the difference between algorithms in more detail. The first data generation process is the model for the Danish data, estimated with the I(1) and I(2) restrictions r, s imposed. $M = 1000$ random samples are drawn from this, using, for each case, the estimated parameters and estimated residual variance assuming normality. The number of iterations and the progress of the algorithm is recorded for each sample. The maximum number of iterations was set to 10 000, $\epsilon_1 = 10^{-11}$, and all replications are included in the results.

Figure 1 shows the histograms of the number of iterations required to achieve convergence (or 10,000). Each graph has the number of iterations (on a log 10 scale) on the horizontal axis and the count (out of 1000 experiments) represented by the bars and the vertical axis. Ideally, all of the mass is to the left, reflecting very quick convergence. The top row of histograms is for δ switching, the bottom row for triangular switching. In each histogram, the data generation process (DGP) uses the stated r, s values, and estimation is using the correct values of r, s .

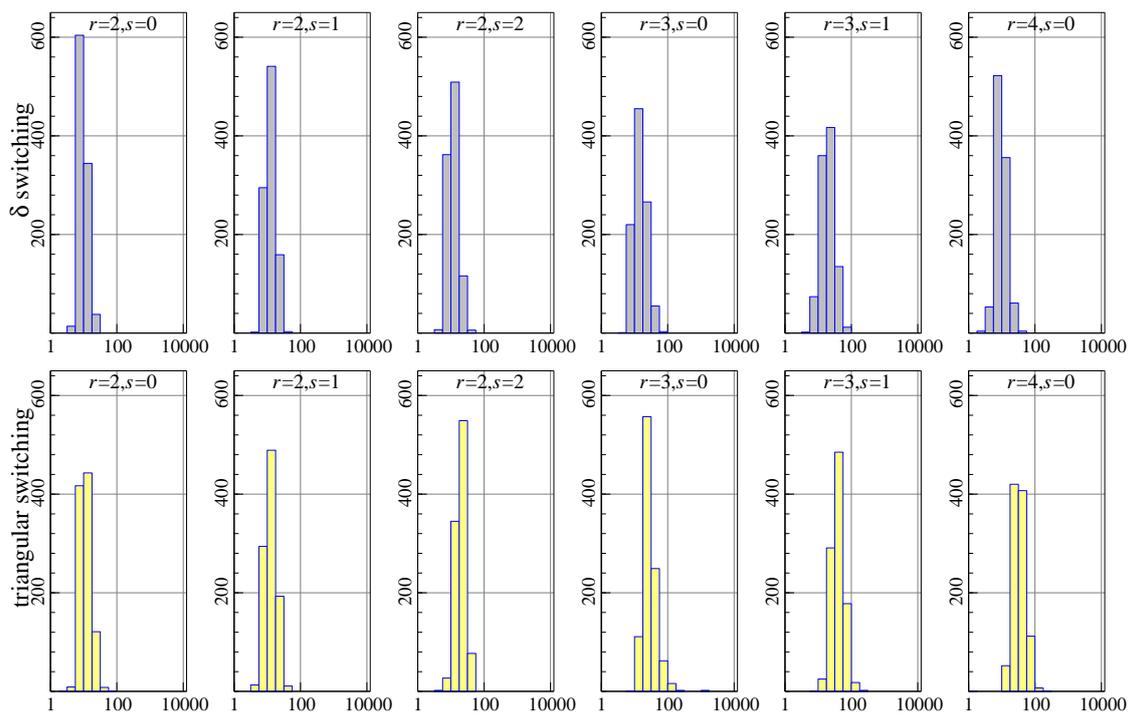


Figure 1. Comparison of algorithms: δ -switching (top row) and triangular-switching (bottom row). Simulating a range of r, s . Number of iterations on the horizontal axis, count (out of 1000) on the vertical.

The histograms show that triangular switching (bottom row) uses more iterations than δ switching (top row), in particular when $s = 0$. Nonetheless, the experiment using triangular switching runs slightly faster as measured by the total time taken (and τ switching is the slowest).

An important question is whether the algorithms converge to the same maximum. The function value that is maximized is:

$$f(\theta) = -\log |\Omega(\theta)|.$$

Out of 10,000 experiments, counted over all r, s combinations that we consider, there is only a single experiment with a noticeable difference in $f(\hat{\theta})$. This happens for $r = 3, s = 0$, and δ -switching finds a higher function value by almost 0.05. Because $T = 119$, the 0.05 translates to a difference of three in the log-likelihoods.

A second issue of interest is how the algorithms perform when restrictions are imposed. The following restrictions are imposed on the three columns of β with $r = 3$:

	m_3	y	Δp	r_m	r_b	t
β'_1	a	$-a$	0	1	-1	*
β'_2	0	*	1	$-a$	a	*
β'_3	0	0	1	*	0	*

This specification identifies the cointegrating vectors and imposes two over-identifying restrictions. For $r = 3, s = 0$ this is accepted with a p -value of 0.4, while for $r = 3, s = 1$, the p -value is 0.5 using the model on the actual Danish data. Simulation is from the estimated restricted model.

In terms of the number of iterations, as Figure 2 shows, δ -switching converges more rapidly in most cases. This makes triangular switching slower, but only by about 10%–20%.

Figure 3 shows $f(\hat{\theta}_\delta) - f(\hat{\theta}_{\text{triangular}})$, so a positive value means that triangular switching obtained a lower log-likelihood. There are many small differences, mostly to the advantage of δ -switching when

$s = 1$ (right-hand plot), but to the advantage of triangular switching on the left, when $s = 0$. The latter case is also much more noisy.

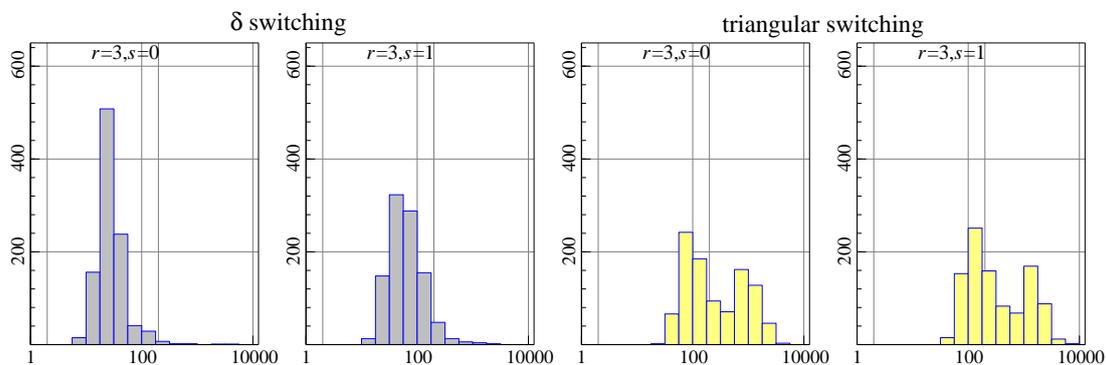


Figure 2. Comparison of algorithms: δ -switching (left two) and triangular-switching (right two). Simulating a range of r, s . Number of iterations on the horizontal axis, count (out of 1000) on the vertical.

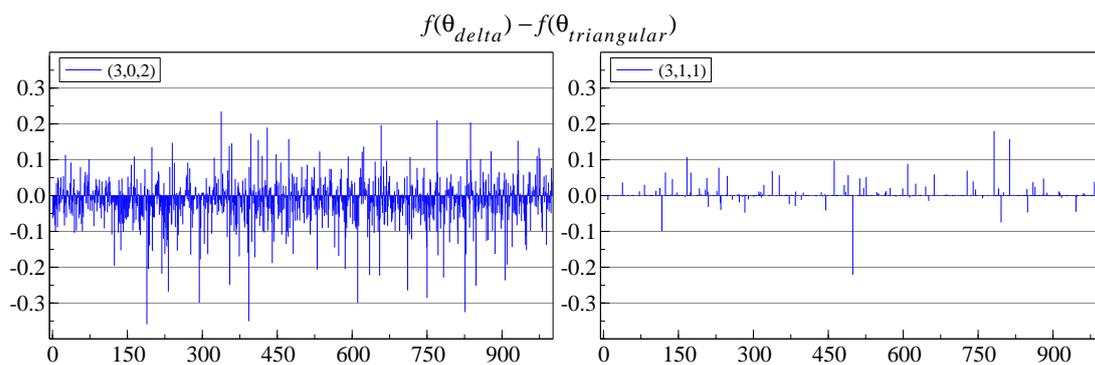


Figure 3. δ -switching function value minus the triangular switching function value (vertical axis) for each replication (horizontal axis). Both starting from their default starting values. The labels are the cointegration indices (r, s, s_2) .

6.1. Hybrid Estimation

To increase the robustness of the triangular procedure, we also consider a hybrid procedure, which combines algorithms as follows:

1. standard starting values, as well as twenty randomized starting values, then
2. triangular switching, followed by
3. BFGS optimization (the Broyden-Fletcher, Goldfarb, and Shanno quasi-Newton method) for a maximum of 200 iterations, followed by
4. triangular switching.

This offers some protection against false convergence, because BFGS is based on first derivatives combined with an approximation to the inverse Hessian.

More importantly, we add a randomized search for better starting values as perturbations of the default starting values. Twenty versions of starting values are created this way, and each is followed for ten iterations. Then, we discard half, merge (almost) identical ones and run another ten iterations. This is repeated until a single one is left.

Figure 4 shows that this hybrid approach is an improvement: now, it is almost never beaten by δ switching. Of course, the hybrid approach is a bit slower again. The starting value procedure for δ switching could be improved in the same way.

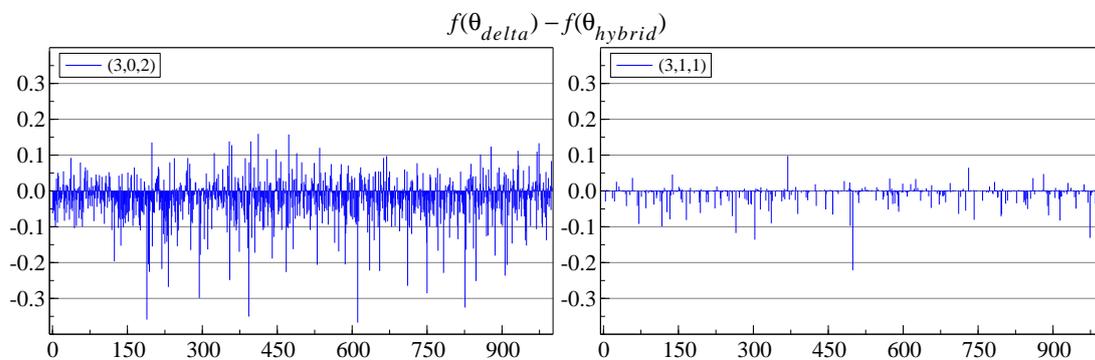


Figure 4. δ -switching function value minus the hybrid triangular-switching function value (vertical axis) for each replication (horizontal axis).

7. Conclusions

We introduced the triangular representation of the I(2) model and showed how it can be used for estimation. The trilinear form of the triangular representation has the advantage that estimation can be implemented as alternating least squares, without using reduced-rank regression. This structure allows us to impose restrictions on (parts of) the A and B matrices, which gives more flexibility than is available in the δ and τ representations.

We also presented an algorithm based on the δ -representation and compared the performance to triangular switching in an application based on Danish data, as well as a parametric bootstrap using that data. Combined with the acceleration of Doornik (2017), both algorithms are fast and give mostly the same result. This will improve empirical applications of the I(2) model and facilitate recursive estimation and Monte Carlo analysis. Expressions for the computation of t -values of coefficients will be reported in a separate paper.

Because they are considerably faster than the previous generation, bootstrapping the I(2) model can now be considered, as Cavaliere et al. (2012) did for the I(1) model.

Acknowledgments: I am grateful to Peter Boswijk, Søren Johansen and Katarina Juselius for providing detailed comments as this paper was developing. Their suggestions have greatly improved the results presented here. Financial support from the Robertson Foundation (Award 9907422) and the Institute for New Economic Thinking (Grant 20029822) is gratefully acknowledged. All computations and graphs use OxMetrics and Ox, Doornik (2013).

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Estimation Using the QR Decomposition

The data matrices in the I(2) model (8) are $Z'_i = (z_{i1} : \dots : z_{iT})$ for $i = 0, 1, 2$.

Take the QR decomposition of $(Z_2 : Z_1)$ as $(Z_2 : Z_1)P = QR = Q_z(R_2 : R_1)$ where Q is a $T \times T$ orthogonal matrix and R a $T \times 2p_1$ upper triangular matrix, while Q_z are the $T \times 2p_1$ leading columns of Q and $(R_2 : R_1)$ a $2p_1 \times 2p_1$ upper triangular matrix. P is the orthogonal matrix that captures the column reordering. Then:

$$Q'_z(Z_2 : Z_1) = (R_2 : R_1)P' = (X_2 : X_1),$$

where $(X_2 : X_1)$ is no longer triangular. Introduce:

$$\begin{pmatrix} X_0 \\ X_0^* \end{pmatrix} = Q'Z_0,$$

where $X_0 = Q_z Z_0$ is $2p_1 \times 2p_1$, then:

$$Z'_i Z_j = Z'_i Q Q' Z_j = \begin{cases} X'_0 X_0 + X_0^{*'} X_0^* & \text{if } i = j = 0, \\ X'_i X_j & \text{otherwise.} \end{cases}$$

Now, e.g., a regression of $A' z_{0t}$ on $B' z_{1t}$ for known A, B :

$$A' z_{0t} = \gamma B' z_{1t} + \epsilon_t, \quad t = 1, \dots, T, \quad (\text{A1})$$

has:

$$\hat{\gamma} = (B' Z'_1 Z_1 B)^{-1} B' Z'_1 Z_0 A = (B' X'_1 X_1 B)^{-1} B' X'_1 X_0 A.$$

This is a regression of $X_0 A$ on $X_1 B$. If such regressions need to be done often for the same Z 's, it is more efficient to do them in terms of the X_i :

$$A' x_{0i} = \gamma B' x_{1i} + e_i, \quad i = 1, \dots, 2p_1,$$

with estimated residual variance:

$$\hat{\Omega}_e = T^{-1} \left(A' X_0^{*'} X_0^* A + \sum_{i=1}^{2p_1} \hat{e}_i \right).$$

This regression has fewer 'observations', while at the same time avoiding the creation of moment matrices. Precomputed moment matrices would be faster, but not as good numerically. For recursive estimation, it is useful to be able to handle singular regressions because dummy variables can be zero over a subsample; this happens naturally in the QR approach. This approach needs to be adjusted when (A1) also has z_{0t} on the right-hand side, as happens for τ -switching in (A3).

Reduced Rank Regression

Let $\text{RRR}(Z_0, Z_1 | Z_x)$ denote reduced rank regression of z_{0t} on z_{1t} corrected for z_{xt} . Assume that (Z_0, Z_1, Z_x) have been transformed into (X_0, X_1, X_x) using the QR decomposition described above. Concentrating X_x out can be done by regression of (X_0, X_1) on X_x , with residuals (Y, X) . Form $S_{00} = Y'Y + X_0^{*'} X_0^*$, and decompose using the Cholesky decomposition: $S_{00} = LL'$.

We need to solve the matrix pencil:

$$X'Y S_{00}^{-1} Y' X x = \lambda X' X x.$$

Start by using the QR decomposition $X = Q R P'$, $y = P' x$:

$$\begin{aligned} R' Q' Y L^{-1} L^{-1} Y' Q R y &= \lambda R' R y, \\ R' W' W R y &= \lambda R' R y, \\ W' W z &= \lambda z, \\ U \Sigma^2 U' z &= \lambda z. \end{aligned}$$

The second line introduces $W = L^{-1} Y' Q$; the next line removes R ; and the final line takes the SVD of W' . The eigenvalues are the squared singular values that are on the diagonal of Σ^2 , and the eigenvectors are $P R^{-1} U$.

When X is singular, as may be the case in recursive estimation, the upper triangular matrix R will have rows and columns that are zero at the bottom and end. These are the same on the left and right of the pencil, so they can be dropped. The resulting reduced dimension R is full rank, and we can set the

corresponding rows in the eigenvectors to zero. When the regressors are singular, their corresponding coefficients in β will be set to zero, just as in our regressions.

This approach differs somewhat from Doornik and O'Brien (2002) because of the different structure of S_{00} as a consequence of the prior QR transformation.

Appendix B. Tau-Switching Algorithm

The algorithm of Johansen (1997, §8) is based on the τ -representation and involves three stages:

1. The estimate of τ is obtained by GLS given all other parameters except ψ . Johansen (1997, p. 451) shows the GLS expressions using second moment matrices. Define the orthogonal matrix $A = (\alpha_{\perp} : \bar{\alpha})$, then using $\kappa' \tau' z_{1t} = \text{vec}(z'_{1t} \tau \kappa) = (\kappa' \otimes z_{1t}) \text{vec} \tau$:

$$\begin{aligned} A' z_{0t} &= \begin{pmatrix} \kappa' \otimes z_{1t} \\ \rho' \otimes z_{2t} \end{pmatrix} \text{vec} \tau + \begin{pmatrix} 0 \\ I_r \otimes z_{1t} \end{pmatrix} \text{vec} \psi + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\ &= \left\{ \begin{pmatrix} \kappa' \\ 0 \end{pmatrix} \otimes z_{1t} + \begin{pmatrix} 0 \\ \rho' \end{pmatrix} \otimes z_{2t} \right\} \text{vec} \tau + \left\{ \begin{pmatrix} 0 \\ I_r \end{pmatrix} \otimes z_{1t} \right\} \text{vec} \psi + u_t. \end{aligned} \tag{A2}$$

The error term u_t has variance $A' \Omega A$, which is block diagonal. Given $\alpha, \kappa, \rho, \Omega$, (A2) is linear in τ and ψ . The estimates of the latter are discarded.

2. Given just τ , reduced-rank regression of z_{0t} corrected for $\tau' z_{1t}$ on z_{0t} corrected for $z_{1t}, \tau' z_{2t}$ is used to estimate α . Details are in Johansen (1997, p. 450).
3. Given τ and α , the remaining parameters can be obtained by GLS. The equivalence $\bar{\alpha}' = \alpha' - \alpha' w \alpha'_{\perp}$ is used to write the conditional equation as:

$$\bar{\alpha}' z_{0t} = \gamma' \alpha'_{\perp} z_{0t} + \rho' \tau' z_{2t} + \psi' z_{1t} + \varepsilon_{2t}, \tag{A3}$$

from which ρ and ψ are estimated by regression. Then, κ is estimated from the marginal equation:

$$\alpha'_{\perp} z_{0t} = \kappa' \tau' z_{1t} + \varepsilon_{1t}. \tag{A4}$$

Together, they give Ω and w . We always transform to set $\rho' = (I : 0)$, adjusting κ and τ accordingly.

τ -switching algorithm:

To start, set $k = 1$, and choose starting values $\alpha^{(0)}, \beta^{(0)}$, tolerance ε_1 and the maximum number of iterations. Compute $\tau_c^{(0)}$ from (18) and $\kappa^{(0)}, \psi^{(0)}, \Omega^{(0)}$ from (A3) and (A4). Furthermore, compute $f^{(0)} = -\log |\Omega^{(0)}|$.

1. Get $\tau_c^{(k)}$ from (A2). Identify this as follows. Select the non-singular $(r + s) \times (r + s)$ submatrix from τ with the largest volume, say M . We find M by using the first $r + s$ column pivots that are chosen by the QR decomposition of τ (Golub and Van Loan (2013, Algorithm 5.4.1)). Set $\tau_c^{(k)} \leftarrow \tau_c^{(k)} M$. Get $\alpha_c^{(k)}$ by RRR; finally, get the remaining parameters from (A3) and (A4).
- 2... As steps 2,3,T from the δ -switching algorithm. □

The line search is only for the $p_1 s_2^*$ parameters in τ as part of it is set to a unit matrix every time. The function evaluation inside the line search needs to obtain all of the other parameters as τ changes.

This is the algorithm of Johansen (1997) except for the normalization of τ and the line search. The former protects the parameter values from exploding, while the latter improves convergence speed and makes it more robust. Removing ρ is largely for convenience: it has little impact on convergence. The τ -switching algorithm is easily adjusted for common restrictions on τ in the form of $\tau = H\tilde{\tau}$. However, ρ gets in the way of more general restrictions.

Appendix C. Starting Values

The first starting value procedure is:

1. Set $\alpha^{(-1)}, \beta^{(-1)}$ to their I(1) values (i.e., with full rank Γ).
2. Get $\tau^{(-1)}$ from (A4), then $\Omega^{(-1)}$ from (A3), ignoring restrictions.
3. Take two iterations with the relevant switching algorithm subject to restrictions.

The second starting value procedure is:

1. Get $\alpha^{(-2)}, \beta^{(-2)}$ by RRR from the τ -representation using $\kappa = 0$:

$$z_{0t} = \alpha(\beta' z_{2t} + \psi' z_{1t}) + \epsilon_t.$$

2. Get $\kappa^{(-2)}, w^{(-2)}$ from (A3), (A4).
3. Get $\alpha^{(-1)}, \beta^{(-1)}$ by RRR from the τ -representation:

$$z_{0t} - w\kappa' \beta' z_{1t} = \alpha(\beta' z_{2t} + \psi' z_{1t}) + \epsilon_t.$$

4. Get $\tau^{(-1)}$ from (A4), then $\Omega^{(-1)}$ from (A3), ignoring restrictions.
5. Take two iterations with the relevant switching algorithm subject to restrictions.

Finally, choose the final starting values as those that have the highest function value.

References

- Anderson, Theodore W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22: 327–51. (Erratum in *Annals of Statistics* 8, 1980).
- Anderson, Theodore W. 2002. Reduced rank regression in cointegrated models. *Journal of Econometrics* 106: 203–16.
- Boswijk, H. Peter. 2010. Mixed Normal Inference on Multicointegration. *Econometric Theory* 26: 1565–76.
- Boswijk, H. Peter, and Jurgen A. Doornik. 2004. Identifying, Estimating and Testing Restricted Cointegrated Systems: An Overview. *Statistica Neerlandica* 58: 440–65.
- Cavaliere, Giuseppe, Anders Rahbek, and R.A.M. Taylor. 2012. Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models. *Econometrica* 80: 1721–40.
- Dennis, Jonathan G., and Katarina Juselius. 2004. *CATS in RATS: Cointegration Analysis of Time Series Version 2*. Technical Report. Evanston: Estima.
- Doornik, Jurgen A. 2013. *Object-Oriented Matrix Programming using Ox*, 7th ed. London: Timberlake Consultants Press.
- Doornik, Jurgen A. 2017. *Accelerated Estimation of Switching Algorithms: The Cointegrated VAR Model and Other Applications*. Oxford: Department of Economics, University of Oxford.
- Doornik, Jurgen A., and R. J. O'Brien. 2002. Numerically Stable Cointegration Analysis. *Computational Statistics & Data Analysis* 41: 185–93.
- Golub, Gen H., and Charles F. Van Loan. 2013. *Matrix Computations*, 4th ed. Baltimore: The Johns Hopkins University Press.
- Johansen, Søren. 1988. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12: 231–54. Reprinted in R. F. Engle, and C. W. J. Granger, eds. 1991. *Long-Run Economic Relationships*. Oxford: Oxford University Press, pp. 131–52.
- 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. 1995a. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Johansen, Søren. 1995b. Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration. *Journal of Econometrics* 69: 111–32.

- Johansen, Søren. 1995c. A Statistical Analysis of Cointegration for I(2) Variables. *Econometric Theory* 11: 25–59.
- Johansen, Søren. 1997. Likelihood Analysis of the I(2) Model. *Scandinavian Journal of Statistics* 24: 433–62.
- Johansen, Søren, and Katarina Juselius. 1990. Maximum Likelihood Estimation and Inference on Cointegration—With Application to the Demand for Money. *Oxford Bulletin of Economics and Statistics* 52: 169–210.
- Johansen, Søren, and Katarina Juselius. 1992. Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK. *Journal of Econometrics* 53: 211–44.
- Johansen, Søren, and Katarina Juselius. 1994. Identification of the Long-run and the Short-run Structure. An Application to the ISLM Model. *Journal of Econometrics* 63: 7–36.
- Juselius, Katarina. 2006. *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.
- 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. Varese: Università dell’Insubria.
- Paruolo, Paolo, and Anders Rahbek. 1999. Weak exogeneity in I(2) VAR Systems. *Journal of Econometrics* 93: 281–308.



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).