Paolo Paruolo

Finite sample comparison of alternative tests on the rank of a cointegration submatrix

2006/6

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

The views expressed in the working papers reflect the opinions of the authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

Printed in Italy in September 2006
Università degli Studi dell'Insubria
Via Monte Generoso 71, 21100 Varese, Italy

Finite sample comparison of alternative tests on the rank of a cointegration submatrix*

Paolo Paruolo
Department of Economics, University of Insubria,
Via Monte Generoso 71, 21100 Varese, Italy
(e-mail: paolo.paruolo@uninsubria.it)

June 30, 2006

Abstract

This paper compares the finite sample performance of alternative tests for rank-deficiency of a submatrix of the cointegrating matrix. The paper focuses on the (implementation of the) likelihood ratio test proposed in Paruolo (2007, *Oxford Bulletin of Economics and Statistics*), and compares its finite sample performance with the ones of alternative tests proposed in Saikkonen (1999, *Econometric Reviews*) and Kurozumi (2005, *Econometric Theory*). All the tests have well-documented limit distributions; their finite sample performance is analyzed in this paper through a Monte Carlo simulation study. We use the Monte Carlo design used in Lukkonen, Ripatti and Saikkonen (1999, *Journal of Business and Economic Statistics*). It is found that the LR and the Kurozumi test perform remarkably better than the alternatives, with a marginal advantage of the LR test. The paper also investigates the properties and the numerical performance of the alternating maximization algorithm that is employed to maximize the likelihood under the null. Alternative ways to choose its starting values are also discussed. In the simulations it is found that the algorithm requires a few iterations when the null is correctly specified and a rather limited number of iteration in 90% of the other cases. The choice of starting values is found to have a significant effect on the number of iteration required by the algorithm.

JEL classification numbers: C15, C32, C63.
Keywords: Vector autoregressive processes, cointegration, likelihood ratio test, numerical optimization, alternating algorithm, starting values.

*Paper submitted to "the 8th workshop of the ERCIM working group on Matrix computation and Statistics" to be held in Salerno September 2-3, 2006. The support of the ESF-EMM network is gratefully acknowledged, as well as partial financial support from grants MIUR Cofin2004 and University of Insubria FAR 2004-2006.

Contents

1 Introduction

This paper investigates the finite sample properties of tests for deficient rank of a submatrix of the cointegration (CI) matrix β in vector autoregressive models (VAR). Special cases of the test include the one of invalid normalization in systems of cointegrating equations, the feasibility of permanent-transitory decompositions and of subhypotheses related to neutrality and long run Granger noncausality; see the discussion in Paruolo (2007), where the likelihood ratio (LR) test for this hypothesis is defined and its large-sample properties are discussed.

Lukkonen, Ripatti and Saikkonen (1999) and Saikkonen (1999) were the first to consider this hypothesis-testing problem. They proposed several tests for the validity of normalization restrictions based on an auxiliary regression formed from unrestricted estimates of the CI VAR model. The null hypothesis of these tests is one of correct normalization. These tests all have a limit distribution of the (multivariate) unit root type.

More recently Kurozumi (2005) has proposed Wald tests for the rank of a submatrix of the cointegrating matrix. The tests are based on properly normalized unrestricted estimates of β obtained from the VAR, with fixed cointegration rank r. The null is one of reduced rank $f < r$. The test statistics against the alternative of full rank have, under regularity conditions, a limit χ^2 distribution. Kurozumi (2005) also proposed similar tests for a submatrix of $\beta_\perp$, a basis of the orthogonal complement of the cointegration space.

This paper focuses on the numerical aspects involved in the implementation of the LR test. We compare the performance of the LR test with the ones of the alternative tests listed above through a Monte Carlo simulation study. We use the Monte Carlo design used in Lukkonen, Ripatti and Saikkonen (1999). It is found that the LR and the Kurozumi tests perform remarkably better than the alternatives, with a marginal advantage for the LR test.

This paper also discusses the numerical properties and performance of an alternating maximization algorithm, which is needed in order to maximize the likelihood function under the null hypothesis. The choice of starting values is also analyzed; it is found to have a significant effect on the number of iteration of the algorithm. The alternating algorithm appears to be fast and perform rather well in 90% of the cases.

In the following we indicate by $\mathcal{A} := \mathrm{col}(A)$ the linear space generated by a matrix A and by $A_\perp$ a basis of the orthogonal complement of $\mathrm{col}(A)$, indicates as $\mathcal{A}^\perp = \mathrm{col}^\perp(A)$. For a full column rank matrix A, we define $\bar{A} := A(A'A)^{-1}$ and let $P_A := \bar{A}A' = A\bar{A}'$ indicate the orthogonal projection matrix onto $\mathrm{col}(A)$.

The rest of the paper is organized as follows. Section 2 reports notation and defines the LR test. Section 3 discusses the choice of starting values and the properties of the alternating maximization algorithm. Section 4 reports results on the Monte Carlo simulation study. Section 5 concludes. Proofs are placed in the Appendix.

2 Model and test

In this section we present the model in Subsection 2.1 and define the hypothesis of interest in Subsection 2.2. We next specify the LR test in Subsection 2.3. We follow Johansen (1996) in most notation conventions.

2.1 VAR processes and statistical model

Consider the VAR(k) process $A(L)X_t = \mu_0 + \mu t + \varepsilon_t$, $t = -k+1, ..., 0, 1, 2, ...$, where $A(L) := -\sum_{i=0}^{k} A_i L^i$ is a k-order matrix polynomial in the lag operator L, $A_0 := -I_p$ and X_t, ϵ_t, μ_0, μ are $p \times 1$ vectors. ε_t is assumed i.i.d. $N(0, \Omega)$, Ω positive definite. Initial values are indicated as X_{-k+1}, ..., X_0, and are assumed to be fixed. [1]

We assume that the roots of $|A(z)| = 0$ are outside the unit disc or at $z = 1$. Under this assumption Granger's representation theorem (see e.g. Johansen 1996, Theorem 4.2) shows that X_t is I(1) and presents at most a linear trend if and only if $-A(1) = \alpha\beta'$, $\mu = \alpha\phi'$, where α and β have full column rank r, and

$$\mathrm{rk}\,(\alpha'_\perp \Gamma \beta_\perp) = p - r, \tag{1}$$

where $\Gamma := I - \sum_{i=1}^{k-1} \Gamma_i$, $\Gamma_i := -\sum_{j=i+1}^{k} A_j$. When these conditions hold, both $\beta' X_t + \phi' t$ and ΔX_t are I(0); these variables appear in the following equilibrium correction (EC) form:

$$\Delta X_t = \mu_0 + \alpha\,(\beta' : \phi') \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t. \tag{2}$$

We now recall the definition of the relevant statistical model. This is expressed in terms of the EC form in eq. (2) by taking all the matrices μ_0, α, β, ϕ, Γ_i, Ω to be unrestricted parameter matrices, with Ω positive definite. Note that α and β have dimensions $p \times r$, $r \leq p$, but not necessarily of full column rank; ϕ has dimensions $r \times 1$, and $\mu = \alpha\phi'$. The resulting model under normality is called the 'I(1) submodel' $\mathcal{H}(r)$ of the VAR. In the rest of the paper we assume that r is known and fixed, and we indicate by θ the column vector of parameters in μ_0, α, β, ϕ, Γ_i, Ω.

2.2 Hypothesis of interest

In this subsection we describe the hypothesis of interest. Let $s := \mathrm{rk}(c'\beta)$, where we take c to be a known, user-defined $p \times r$ matrix of full column rank r, where r is the CI rank, assumed known and fixed. Consider the hypothesis

$$\mathcal{K} : s = r, \tag{3}$$

The complementary hypothesis to $\mathcal{K}$ in (3) is

$$\mathcal{H}_{0,j} : s \leq r - j, \qquad 0 < j \leq m := \min(r, p - r), \tag{4}$$

where j is the 'rank deficiency' index. This rank deficiency index cannot exceed r, which is the dimension of $c'\beta$, $0 \leq s \leq r$. On the other hand in Paruolo (2007) it is shown that the rank deficiency j is associated to how many columns in β can be selected from the orthogonal complement $\mathcal{C}^\perp := \mathrm{col}(c_\perp)$, which has dimension $p - r$. This implies that $m := \min(r, p - r)$ is the maximum possible value for j.

In order to test $\mathcal{K}$ in (3), we take the complementary hypothesis $\mathcal{H}_{0,j}$ in (4) as the null hypothesis. If $\mathcal{H}_{0,j}$ is rejected for any $j > 0$, then one concludes that $\mathcal{K}$ is valid. We here observe that the $\mathcal{H}_{0,j}$ hypotheses are nested as follows

$$\mathcal{H}_{0,1} \supset \mathcal{H}_{0,2} \supset ... \supset \mathcal{H}_{0,m}. \tag{5}$$

[1] Initial conditions can alternatively be assumed to the realizations of random variables whose distribution does not depend on μ_0, μ, Ω, A_i, $i = 1, ..., k$. This includes the case of fixed initial values as a special case.

Note that if the most general hypothesis $\mathcal{H}_{0,1}$ is false, one can conclude directly that $\mathcal{K}$ is valid. In Paruolo (2007) it is shown that the null hypothesis $\mathcal{H}_{0,j}$ in eq. (4), (5) can be expressed as follows:

$$\beta := (\beta_1 : \beta_2) = (H_1\varphi_1 : H_2\varphi_2), \tag{6}$$

with $H_1 = c_\perp$, $H_2 = I_p$, φ_1 of dimension $(p-r)\times j$, and φ_2 of dimension $p\times(r-j)$. We hence consider (6) as the formulation of hypothesis $\mathcal{H}_{0,j}$, for unrestricted φ_i, $i = 1, 2$.

We note that the ability to use the format (6) is not affected by the introduction of restricted or unrestricted deterministic terms in the VAR. Consider in fact model (2) with restricted linear trend; restriction (6) can be stated in terms of the extended matrix $\beta^* := (\beta' : \phi')'$ as follows:

$$\beta^* = (H_1^*\varphi_1^* : H_2^*\varphi_2^*), \qquad H_i^* := \begin{pmatrix} H_i & 0 \\ 0 & 1 \end{pmatrix}, \qquad i = 1, 2 \tag{7}$$

where φ_i^* have one additional row with respect to their φ_i counterparts.

We close this subsection by describing the most restricted hypothesis $\mathcal{H}_{0,m}$. If $m = r \leq p - r$, then $\mathcal{H}_{0,m}$ states that $\beta \in \mathrm{col}\,(c_\perp)$, which can be written as $\beta = H_1\varphi_1$ for $H_1 = c_\perp$. On the other hand if $m = p - r < r$, then $\mathcal{H}_{0,m}$ corresponds to $\beta = (c_\perp : \varphi_2)$. Both cases correspond to the format of eq. (6), where either β_2 is absent or $\beta_1 = H_1$. The statistical calculation of the restricted ML estimator involved in these two special cases are considerably simpler than for the general case in (6). Statistical calculations are discussed in the following subsection.

2.3 The likelihood ratio test

In this subsection we describe how one calculates the LR test of (6) within the $\mathcal{H}(r)$ model. Maximum likelihood estimation (ML) of $\mathcal{H}(r)$ consists of Reduced Rank Regression (RRR), see e.g. Anderson (1951) and Johansen (1996). ML estimation of the model under $\mathcal{H}_{0,j}$ as specified in (6) requires in general numerical optimization, except in the special case $\mathcal{H}_{0,m}$. The present subsection reviews the statistical calculations needed to obtain the LR test.

Many algorithms for the solution of the numerical optimization problem under restrictions on β have been proposed, see Boswijk and Doornik (2004) for a recent review. A full discussion of the relative merits of these algorithms is beyond the scope of the present paper. We here choose to present and discuss the format of the alternating optimizing procedure proposed by Johansen and Juselius (1992) and summarized in Johansen (1996, Chapter 7.2.3, Theorem 7.4). If other numerical maximization algorithms are used in its place to produce the restricted ML estimator of β^*, indicated here as $\widehat{\beta}^*$, the expression below for the LR would then hold inserting $\widehat{\beta}^*$ as the the restricted ML estimator.

The Gaussian log-likelihood function conditional on the initial values X_{-k+1}, ..., X_0 is

$$\ell_T(\theta) := -\frac{T}{2}\left(\ln|\Omega| + T^{-1}\sum_{i=1}^{T}\epsilon_t'\Omega^{-1}\epsilon_t\right),$$

where the constant $-2^{-1}Tp\ln(2\pi)$ is omitted for simplicity. The log-likelihood function can be concentrated with respect to all parameters in θ except β^*, obtaining

$$\ell_T(\beta^*) := -\frac{T}{2}\left(p + \ln|S_{00}| + \ln\frac{|\beta^{*\prime}S_{11.0}\beta^*|}{|\beta^{*\prime}S_{11}\beta^*|}\right), \tag{8}$$

where $S_{11.0} := S_{11} - S_{10}S_{00}^{-1}S_{01}$; $\ell_T(\beta^*)$ denotes the concentrated log-likelihood function. The maximum of ℓ_T in the model $\mathcal{H}(r)$ is found solving the generalized eigenvalue problem

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

with eigenvalues $\widetilde{\lambda}_1 \geq ... \geq \widetilde{\lambda}_p > 0$ where $S_{ij} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $M_{ij} := T^{-1}\sum_{t=1}^{T} Z_{it}Z_{jt}'$, $Z_{0t} := \Delta X_t$, $Z_{1t} := (X_{t-1}' : t)'$, $Z_{2t} := (\Delta X_{t-1}' : ... : \Delta X_{t-k+1}' : 1)'$. In the following we use the notation $\widetilde{\lambda}_{1:m} := (\lambda_1 : \cdots : \lambda_m)'$ to indicate a $m \times 1$ vector containing the largest m eigenvalues, and let $V_{1:m} := (V_1 : \cdots : V_m)$ indicate the corresponding matrix of eigenvectors.

The estimates of the cointegration matrix $\beta^* := (\beta' : \phi')'$ correspond to the r eigenvectors $V_{1:r}$ associated with the eigenvalues $\widetilde{\lambda}_{1:r}$. We indicate this calculation by $\{\widetilde{\beta}^*, \widetilde{\lambda}_{1:r}\} := \mathrm{RRR}(Z_{0t}, Z_{1t}; Z_{2t}, r)$. The maximized value of $\ell_T(\theta)$ in $\mathcal{H}(r)$ is

$$\ell_{T,A} := \ell_T\left(\widetilde{\beta}^*\right) = -\frac{T}{2}\left(p + \ln|S_{00}| + \sum_{i=1}^{r}\ln\left(1 - \widetilde{\lambda}_i\right)\right),$$

where $\ell_{T,A}$ indicates the maximized value under the alternative hypothesis $\mathcal{H}(r)$, see (8).

Under the null hypothesis $\mathcal{H}_{0,j}$, $j \neq m$, the maximum of the $\ell_T(\theta)$ has to be found by numerical optimization; in the following we indicate the corresponding maximized value by $\ell_{T,j}$. Johansen and Juselius (1992) and Johansen (1996, Theorem 7.4) have proposed an alternating algorithm that maximizes the concentrated likelihood function $\ell_T(\beta^*)$ with respect to β_1 for fixed β_2 and vice versa. This algorithm is reviewed below; its properties are reported in the following Subsection 3.1; the choice of starting values is discussed in Subsection 3.2.

Let $\beta^{*(0)} := (\beta_1^{*(0)} : \beta_2^{*(0)})$ be a starting value for the iterative maximization of $\ell_T(\beta^*)$ under $\mathcal{H}_{0,j}$. Let $\beta_i^{*(h)} = H_i^*\varphi_i^{*(h)}$, $i = 1, 2$, be the values of β_1^* and β_2^* in iteration $h = 1, 2, \ldots$; the algorithm is defined as follows:

$$\begin{aligned}
\{\varphi_1^{*(h)}, \rho_{1:j}^{(h)}\} &:= \mathrm{RRR}\left(Z_{0t}, H_1^{*\prime}Z_{1t}; (Z_{1t}'\beta_2^{*(h-1)} : Z_{2t}')', j\right), \\
\{\varphi_2^{*(h)}, \zeta_{1:r-j}^{(h)}\} &:= \mathrm{RRR}\left(Z_{0t}, H_2^{*\prime}Z_{1t}; (Z_{1t}'\beta_1^{*(h)} : Z_{2t}')', r-j\right).
\end{aligned} \tag{9}$$

At each step h, ℓ_T is increased, i.e. $\ell_T\left(\beta^{*(h)}\right) - \ell_T\left(\beta^{*(h-1)}\right) > 0$. The algorithm is terminated when the increment in ℓ_T is inferior to a given tolerance level ξ, i.e. $\ell_T\left(\beta^{*(h)}\right) - \ell_T\left(\beta^{*(h-1)}\right) < \xi$.[2] The last value of h corresponds to the number of iterations of the algorithm, and it is indicated as n_{it} in the following.

After convergence of this (or any other) algorithm to the restricted ML estimate

[2] Other termination criteria can be used; the given criterion is the one used in the simulations reported in the paper.

$\widehat{\beta}^* = \left(\widehat{\beta}_1^* : \widehat{\beta}_2^*\right)$, the maximum $\ell_{T,j}$ of $\ell_T(\theta)$ under $\mathcal{H}_{0,j}$ can be calculated as

$$\begin{aligned}
\ell_{T,j} := \ell_T\left(\widehat{\beta}^*\right) &= -\frac{T}{2}\left(p + \ln|S_{00}| + \sum_{i=1}^{j}\ln\left(1-\widehat{\rho}_i\right) + \ln\frac{\left|\widehat{\beta}_2^{*\prime} S_{11.0}\widehat{\beta}_2^*\right|}{\left|\widehat{\beta}_2^{*\prime} S_{11}\widehat{\beta}_2^*\right|}\right) \\
&= -\frac{T}{2}\left(p + \ln|S_{00}| + \sum_{i=1}^{r-j}\ln\left(1-\widehat{\zeta}_i\right) + \ln\frac{\left|\widehat{\beta}_1^{*\prime} S_{11.0}\widehat{\beta}_1^*\right|}{\left|\widehat{\beta}_1^{*\prime} S_{11}\widehat{\beta}_1^*\right|}\right) \\
&= -\frac{T}{2}\left(p + \ln|S_{00}| + \ln\frac{\left|\widehat{\beta}^{*\prime} S_{11.0}\widehat{\beta}^*\right|}{\left|\widehat{\beta}^{*\prime} S_{11}\widehat{\beta}^*\right|}\right),
\end{aligned}$$

see (8), where for the alternating algorithm considered here $\widehat{\beta}_i^* = \beta_i^{*(n_{\mathrm{it}})}$, $i = 1, 2$, see (9).

The likelihood ratio test of $\mathcal{H}_{0,j}$ within $\mathcal{H}(r)$ is thus given by[3]

$$\begin{aligned}
Q_j := -2\left(\ell_{T,j} - \ell_{T,A}\right) &= T\left(\ln\frac{\left|\widehat{\beta}_1^{*\prime} S_{11.0}^*\widehat{\beta}_1^*\right|}{\left|\widehat{\beta}_1^{*\prime} S_{11}^*\widehat{\beta}_1^*\right|} + \sum_{i=1}^{r-j}\ln\left(1-\widehat{\zeta}_i\right) - \sum_{i=1}^{r}\ln\left(1-\widetilde{\lambda}_i\right)\right) \\
&= T\left(\sum_{i=1}^{j}\ln\left(1-\widehat{\rho}_i\right) + \ln\frac{\left|\widehat{\beta}_2^{*\prime} S_{11.0}^*\widehat{\beta}_2^*\right|}{\left|\widehat{\beta}_2^{*\prime} S_{11}^*\widehat{\beta}_2^*\right|} - \sum_{i=1}^{r}\ln\left(1-\widetilde{\lambda}_i\right)\right) \qquad (10) \\
&= T\left(\ln\frac{\left|\widehat{\beta}^{*\prime} S_{11.0}^*\widehat{\beta}^*\right|}{\left|\widehat{\beta}^{*\prime} S_{11}^*\widehat{\beta}^*\right|} - \sum_{i=1}^{r}\ln\left(1-\widetilde{\lambda}_i\right)\right).
\end{aligned}$$

Other specifications of the deterministic components can be treated in a similar way, as explained here. The case of no linear trend in the equations, $\phi = 0$, is treated using (6) in place of (7) and omitting t from Z_{1t}. The case of restricted constant, $\mu_0 = \alpha\phi_0'$ is similar to the present case with (6) in place of (7), a constant 1 in place of the trend t in Z_{1t} and no constant in Z_{2t}. The case of no deterministic is obtained by deleting 1 and t from Z_{1t} and Z_{2t}, and using (6) in place of (7). Hence in all submodel with restricted deterministic components one can apply the previous calculations, simply changing the definitions of the variables.

We now discuss the special case of hypothesis $\mathcal{H}_{0,m}$. The calculations for the restricted ML can be described in terms of a single step of the alternating algorithm described above. Specifically when $m = p - r < r$, then $\beta = (c_\perp : \varphi_2)$ one can simply set $\widehat{\beta}_1^* = c_\perp$ and $\{\widehat{\varphi}_2^*, \widehat{\zeta}_{1:r-j}\} := \mathrm{RRR}\left(Z_{0t}, H_2^{*\prime} Z_{1t}; (Z_{1t}' c_\perp : Z_{2t}')', r - m\right)$, $\widehat{\beta}_2^* = H_2^*\widehat{\varphi}_2^*$. This explicit solution is documented as hypothesis $\mathcal{H}_5$ in eq. (15) of Johansen and Juselius (1992). The $Q_m = Q_{p-r}$ statistics can be computed using the first or the third expressions in (10) above.

When $m = r < p - r$, one has $\beta = c_\perp$. This can be seen as a special case of the previous specification $\beta = (c_\perp : \varphi_2)$ when φ_2 (i.e. β_2) is absent. The $Q_m = Q_r$ statistics can be computed using the second expressions in (10) above setting $j = r$

[3]The expressions given here for $\ell_{T,j}$ and Q_j correct minor errors in eq. (34) in Johansen and Juselius (1992) and in the expression for $L_{\max}^{-2/T}$ in Theorem 7.4 in Johansen (1996).

and omitting the term with the ratio of two determinants. Using the notation of Johansen and Juselius (1992), this corresponds to a special case of hypothesis $\mathcal{H}_5$ in their eq. (15) or of hypothesis $\mathcal{H}_4$ in their eq. (14), with H_4 of dimension $p \times r$.

3 Numerical algorithm

In this section we discuss properties of the alternating algorithm in Subsection 3.1, while Subsection 3.2 discusses the choice of starting values for the restricted maximization of the likelihood. Proofs are placed in the Appendix.

3.1 Properties of the alternating algorithm

This subsection discusses some properties of the alternating algorithm, concerning convergence of the algorithm, the number of iterations n_{it} and the type of accumulation points of the algorithm.

Alternating algorithms as an optimization tool have a long tradition in econometrics; they characterize classical econometric procedures such as iterative generalized least squares, iterative seemingly unrelated least squares, the iterative Cochran-Orcutt procedure, iterative three stage least squares. Properties of the alternating algorithm presented in Subsection 2.3 are derived in the same manner as for the above classical econometric estimators, and date back at least to Sargan (1964).

We first introduce some additional notation. We let $a_i := \varphi_i^*(\varphi_i^{*\prime}\varphi_i^*)^{-1/2}$, $i = 1, 2$. Define the sets $\mathcal{U}_1 := \{a_1 \in \mathbb{R}^{(p-r+1)\times j} : a_1'a_1 = I_j\}$, $\mathcal{U}_2 := \{a_2 \in \mathbb{R}^{(p+1)\times(r-j)} : a_2'a_2 = I_{r-j}\}$. Let a be the ordered pair $a := (a_1, a_2)$, where $a \in \mathcal{U} := \mathcal{U}_1 \times \mathcal{U}_2$. Observe that $\ell_T(\beta_1^*, \beta_2^*)$ can be written as a function of a, $\ell_T(\beta_1^*, \beta_2^*) = \ell_T(a_1, a_2) = \ell_T(a)$. Let $a^{(0)}$, $a^{(h)}$ be the values of a corresponding respectively to the starting values $\beta^{*(0)}$ and to the values $\beta^{*(h)}$ at each iteration h. Finally let S be the level set $S := \{a \in \mathcal{U} : \ell_T(a) \geq \ell_T(a^{(0)})\}$. The following proposition reviews properties of the algorithm.

Proposition 1 *(i) The sequence $\{a^{(h)}\}$ has at least one accumulation point $a^\dagger$ in S. (ii) If $a^\dagger$ and $a^\ddagger$ are two accumulation points of the sequence, then $\ell_T(a^\dagger) = \ell_T(a^\ddagger)$. (iii) For every accumulation point $a^\dagger := (a_1^\dagger, a_2^\dagger)$, one has*

$$\max_{a_1 \in \mathcal{U}_1} \ell_T(a_1, a_2^\dagger) = \max_{a_2 \in \mathcal{U}_2} \ell_T(a_1^\dagger, a_2) = \ell_T(a_1^\dagger, a_2^\dagger) = \ell_T(a^\dagger).$$

(iv) If ℓ_T is differentiable at an accumulation point $a^\dagger$, then

$$\partial \ell_T(a^\dagger)/\partial a_1 = 0, \partial \ell_T(a^\dagger)/\partial a_2 = 0.$$

(v) Finally $n_{it} \leq n_{it}^{\max} := (\ell_T(\widetilde{\beta}^*) - \ell_T(a^{(0)}))/\xi < \infty$.

Several remarks are in order. First of all, not all accumulation points coincide with the global maxima, but can well be local maxima and saddle points. Secondly, as shown in (ii), the value of ℓ_T is the same at all accumulation points of the sequence; this implies that the value of Q_j is not affected by the presence of several accumulation points.

However, the choice of starting values for the algorithm does influence which accumulation point is reached by the algorithm and the number of iterations. This calls for a careful selection of starting values, which is discussed in the next subsection.

From a practical point of view, Proposition 1.(v) shows that the number of iterations is bounded by a finite number $n_{\text{it}}^{\max}$. This is the consequence of the fact that the algorithm increases the log-likelihood function ℓ_T at each iteration, and that ℓ_T is bounded by the value $\ell_T(\widetilde{\beta}^*)$ obtained in the unrestricted model $\mathcal{H}(r)$. This means that the algorithm always stops in a finite number of iterations, less or equal to $n_{\text{it}}^{\max}$. Note that $n_{\text{it}}^{\max}$ depends on the choice of starting values and on the tolerance value ξ used for convergence.

An advantage of this alternating algorithm over quasi-Newton methods, say, is that it avoids potential problems associated with non-identification of the parameters α_1, φ_1, α_2, φ_2, see Boswijk and Doornik (2004); here $\alpha = (\alpha_1 : \alpha_2)$ is partitioned conformably with $\beta = (\beta_1 : \beta_2) = (H_1\varphi_1 : H_2\varphi_2)$. On the other hand, other algorithms may require fewer iterations than the alternating algorithm. While a complete discussion of this point is beyond the scope of the present paper, in Subsection 4.2 below we document the number of iterations needed to obtain convergence of the alternating algorithm in the Monte Carlo study.

3.2 Choice of starting values

This subsection discusses the choice of starting values $\beta^{*(0)}$ for the maximization of ℓ_T under $\mathcal{H}_{0,j}$. In a strictly related model, Johansen (1996, Chapter 7.2.3) proposed to choose the starting values close to the unrestricted ML estimates $\widetilde{\beta}^*$. His proposal of starting values in indicated by $\beta^*_{(0)}$ and it is reviewed below. We also propose a modification of it, indicated as $\beta^*_{\langle 0\rangle}$, which ensures that the starting values of β^* satisfy the restrictions $\mathcal{H}_{0,j}$ and presents perpendicular blocks of columns. We finally consider another choice of starting values, indicated as $\beta^*_{\diamondsuit}$.

3.2.1 Starting values for identified systems of equations

We first summarize Johansen's choice of starting values, indicated as $\beta^*_{(0)} := (\beta^*_{1(0)} : \beta^*_{2(0)})$. This choice of starting values was proposed for identified systems; we show below that it must be modified in the present setting, where β_1^* and β_2^* are not identified.

Consider hypothesis $\mathcal{H}_{0,j}$ in (7) of the form $\beta^* = (\beta_1^* : \beta_2^*) = (H_1^*\varphi_1^* : H_2^*\varphi_2^*)$ with β_i^* of dimension $(p+1) \times r_i$; here $r_1 = j$, $r_2 = r - j$. Choose starting values for β_i^* as the linear combinations of $\widetilde{\beta}^*$ that are closest to $\text{col}(H_i^*)$. This is accomplished by solving the eigenvalue problem

$$\left|\varsigma\widetilde{\beta}^{*\prime}\widetilde{\beta}^* - \widetilde{\beta}^{*\prime}P_{H_i^*}\widetilde{\beta}^*\right| = 0, \tag{11}$$

with eigenvalues $\varsigma_{i,1} \geq ... \geq \varsigma_{i,r} > 0$ and associated eigenvectors v_1, ..., v_r. Let $U_{i,1} := (v_1 : ... : v_{r_i})$, $U_{i,2} := (v_{r_i+1} : ... : v_r)$, $U_i := (U_{i,1} : U_{i,2})$. With this notation $\beta^*_{i(0)} := \widetilde{\beta}^* U_{i,1}$, $i = 1, 2$.

The unrestricted ML estimator $\widetilde{\beta}^*$ is expected to be close to the restricted ML estimator $\widehat{\beta}^*$ when $\mathcal{H}_{0,j}$ is true. Hence one expects starting values derived from the unrestricted ML estimator $\widetilde{\beta}^*$, like $\beta^*_{i(0)}$, to be good choices when $\mathcal{H}_{0,j}$ is true. Note that this is not necessarily the case if $\mathcal{H}_{0,j}$ is false.

One may note that $\beta^*_{i(0)}$ does not satisfy the restriction $\beta_i^* = H_i^*\varphi_i^*$, $i = 1, 2$ of hypothesis $\mathcal{H}_{0,j}$. Hence one may question the idea of starting the iterations for the

restricted ML estimation at the point $\beta^* = \beta^*_{(0)}$, which is outside the parameter space for the restricted model $\mathcal{H}_{0,j}$.

This shortcoming can be fixed e.g. by orthogonal projection of $\beta^*_{i(0)}$ on $\text{col}(H^*_i)$. We indicate this modification of the starting values as $\beta^*_{i[0]} := P_{H^*_i}\beta^*_{i(0)}$. This modified choice of starting values can also be obtained by the following argument. Instead of choosing $\beta^*_{i(0)}$ as the linear combination of $\widetilde{\beta}^*$ that is closest to $\text{col}(H^*_i)$, one can find the linear combination of H^*_i that is closest to $\text{col}(\widetilde{\beta}^*)$. This corresponds to interchanging $\widetilde{\beta}^*$ and H^*_i in (11) i.e. use the dual eigenvalue problem

$$\left|\varsigma H^{*\prime}_i H^*_i - H^{*\prime}_i P_{\widetilde{\beta}^*} H^*_i\right| = 0.$$

The dual problem has the same eigenvalues $\varsigma_{i,1} \geq ..., \geq \varsigma_{i,r} > 0$ as (11) and eigenvectors V_i equal to $V_i = \left(H^{*\prime}_i H^*_i\right)^{-1} H^{*\prime}_i \widetilde{\beta} U \,\text{diag}(\varsigma_{i,1}^{-1/2}, ..., \varsigma_{i,r}^{-1/2})$, see e.g. Johansen (1996, Lemma A.9, p. 226). The choice of starting values from the dual problem is $H^*_i V_{i,1}$, where V_i is partitioned conformably with U_i. One sees that $\beta^*_{i[0]} = H^*_i \left(H^{*\prime}_i H^*_i\right)^{-1} H^{*\prime}_i \widetilde{\beta} U_{i,1} = H^*_i V_{i,1} \,\text{diag}(\varsigma_{i,1}^{1/2}, ..., \varsigma_{i,r_1}^{1/2})$. In other words, the choice of starting values from the dual problem differs from $\beta^*_{i[0]}$ just by the scaling factor $\text{diag}(\varsigma_{i,1}^{1/2}, ..., \varsigma_{i,r_1}^{1/2})$. Given that the log-likelihood ℓ_T is invariant to scaling of β^*_i, $i = 1$, 2, the dual eigenvalue problem gives in effect the same choice of starting values as $\beta^*_{i[0]}$.

A limitation which applies to both $\beta^*_{2(0)}$ and $\beta^*_{2[0]}$ is that the present choice $H^*_2 = I_{p+1}$ in $\mathcal{H}_{0,j}$ implies that $\varsigma_{2,1} = ... = \varsigma_{2,r} = 1$, $\varsigma_{2,r+1} = ... = \varsigma_{2,p+1} = 0$, and one can choose any subset of columns in $\widetilde{\beta}^*$ as $\beta^*_{2(0)}$ or $\beta^*_{2[0]}$. This poses the question of which subset of columns of $\widetilde{\beta}^*$, or combination of columns of $\widetilde{\beta}^*$ one should choose as $\beta^*_{2(0)}$ or $\beta^*_{2[0]}$.

3.2.2 First choice of starting values

In this subsection we define the first choice of starting values, indicated as $\beta^*_{\langle 0\rangle} := \left(\beta^*_{1\langle 0\rangle} : \beta^*_{2,\langle 0\rangle}\right)$; they are derived from $\beta^*_{[0]}$ described above.

We set $\beta^*_{1\langle 0\rangle} := \beta^*_{1[0]}$. Define the linear combinations of $\widetilde{\beta}^*$, indicated as $\beta^*_{2\{0\}}$, that are perpendicular to $\beta^*_{1(0)}$ and such that $\text{col}(\beta^*_{1(0)} : \beta^*_{2\{0\}}) = \text{col}(\widetilde{\beta}^*)$. One has $\beta^*_{2\{0\}} = \widetilde{\beta}^* U_{i,2}$ in the notation above. In order to make the starting values of $\beta^*_{2,\langle 0\rangle}$ perpendicular to $\beta^*_{1,\langle 0\rangle} = \beta^*_{1[0]}$, one can moreover project $\beta^*_{2\{0\}}$ on $\text{col}^{\perp}(\beta^*_{1[0]})$ obtaining $\beta^*_{2\langle 0\rangle} := \left(I - P_{\beta^*_{1[0]}}\right)\beta^*_{2\{0\}}$. This gives the following choice of starting values $\beta^*_{\langle 0\rangle} := (\beta^*_{1\langle 0\rangle} : \beta^*_{2\langle 0\rangle})$.

$\beta^*_{\langle 0\rangle}$ satisfies the restrictions of hypothesis $\mathcal{H}_{0,j}$, as well as guaranteeing that the starting values for β^*_1 and β^*_2 are perpendicular, $\beta^{*\prime}_{1\langle 0\rangle}\beta^*_{2\langle 0\rangle} = 0$.

3.2.3 Second choice of starting values

We also define a second choice of starting values, indicated as $\beta^*_{\diamondsuit} := (\beta^*_{1\diamondsuit} : \beta^*_{2\diamondsuit})$. This choice selects as starting values $\beta^*_{1\diamondsuit}$ for β^*_1 the first j columns of the restricted ML estimates under the hypothesis $\beta^* = H^*_1\varphi^*$ obtained by applying the restrictions in H^*_1 to all column of β^*. This choice is calculated as $\beta^*_{1\diamondsuit} = H^*_1\varphi^*_{1\diamondsuit}$, where $\{\varphi^*_{1\diamondsuit}, \rho^{\diamondsuit}_{1:j}\} := \text{RRR}\left(Z_{0t}, H^{*\prime}_1 Z_{1t}; Z_{2t}, j\right)$. For fixed $\beta^*_{1\diamondsuit}$, the starting value for β^*_2 is obtained as in

(9), i.e. $\beta^*_{2\diamondsuit} = H^*_2 \varphi^*_{2\diamondsuit}$, $\{\varphi^*_{2\diamondsuit}, \lambda^{\diamondsuit}_{1:r-j}\} := \text{RRR}\left(Z_{0t}, H^{*\prime}_2 Z_{1t}; (Z'_{1t}\beta^*_{1\diamondsuit} : Z'_{2t})', r-j\right)$. One expects the present choice of starting values $\beta^*_{\diamondsuit}$ to be less sensitive than the starting values $\beta^*_{\langle 0\rangle}$ to wether hypothesis $\mathcal{H}_{0,j}$ is true or not.

The two choices of starting values $\beta^*_{\langle 0\rangle}$ and $\beta^*_{\diamondsuit}$ were implemented in the Monte Carlo simulations in Subsection 4.2 below.

4 Simulation study

This section reports the Monte Carlo (MC) simulation study. The Q_j tests are compared with the ones proposed by Lukkonen, Ripatti and Saikkonen (1999) and Kurozumi (2005). The MC design is taken from Lukkonen, Ripatti and Saikkonen (1999), and it is described in Subsection 4.1. Subsection 4.2 documents the performance of the alternating algorithm described in Subsection 2.3 for the choices of starting values described in subsection 3.2. Results on the performance of the Q_j tests and the one of alternative tests are presented in Subsection 4.3.

4.1 Monte Carlo design

We employ the MC design of Lukkonen, Ripatti and Saikkonen (1999). We selected the same number of replications, equal to 10^4, and the same nominal size of 5%.

The simulated process is a VAR process $X_t := (X'_{1t} : X'_{2t})'$, where X_{1t} and X_{2t} are 2×1 vectors, $p = 4$. We selected $X_0 = 0$.[4] The process is defined as follows:

$$\begin{pmatrix} I_2 & -\gamma \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} b_1 I_2 & -b_1\gamma \\ b_2 I_2 & I_2 - b_2\gamma \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

where $u_t := (u'_{1t} : u'_{2t})' \sim N(0, I_4)$, and

$$\gamma = \begin{pmatrix} -a & 1 \\ a-1 & -1 \end{pmatrix}.$$

Inverting the matrix that multiplies X_t on the left hand side, one finds the VAR(1) representation:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} b_1 I_2 + b_2\gamma & -\left((b_1-1) I_2 + b_2\gamma\right)\gamma \\ b_2 I_2 & I_2 - b_2\gamma \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

where ε_t i.i.d. $N(0, \Omega)$ with

$$\Omega = \begin{pmatrix} I_2 & \gamma \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ \gamma' & I_2 \end{pmatrix} = \begin{pmatrix} \gamma\gamma' + I_2 & \gamma \\ \gamma' & I_2 \end{pmatrix}.$$

One has

$$-A(1) = -I_4 + A_1 = \begin{pmatrix} (b_1 - 1) I_2 + b_2\gamma \\ b_2 I_2 \end{pmatrix} \begin{pmatrix} I_2 & -\gamma \end{pmatrix} =: \alpha\beta'$$

[4] Unlike in Lukkonen, Ripatti and Saikkonen (1999), we did not discard the first 100 data points of each time series because the likelihood function is conditional on the initial observations.

DGP_i	a	b_1	R^2		a	b_1	R^2
$i=1$	1	0.4	0.4	$i=13$	0.2	0.4	0.4
2			0.8	14			0.8
3		0.8	0.4	15		0.8	0.4
4			0.8	16			0.8
5	0.5	0.4	0.4	17	0.1	0.4	0.4
6			0.8	18			0.8
7		0.8	0.4	19		0.8	0.4
8			0.8	20			0.8
9	0.3	0.4	0.4	21	0	0.4	0.4
10			0.8	22			0.8
11		0.8	0.4	23		0.8	0.4
12			0.8	24			0.8

Table 1: Description and numbering of the DGPs used in the Monte Carlo simulations.

with

$$\alpha := \begin{pmatrix} (b_1-1)\, I_2 + b_2\gamma \\ b_2 I_2 \end{pmatrix}, \qquad \beta := \begin{pmatrix} I_2 \\ -\gamma' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a & 1-a \\ -1 & 1 \end{pmatrix}.$$

We first check condition (1), where $\Gamma = I$ because there is only one lag. Moreover it can be shown, see e.g. Theorem 3 in Paruolo (2007), that $\text{rk}\,(\alpha'_\perp \beta_\perp) = p - r$ iff $\text{rk}(\beta'\alpha) = r$. In this design $\beta'\alpha = (b_1 - 1)\, I_2$ which has rank equal to 2 for all values of $b_1 \neq 1$. Hence condition (1) is satisfied.

Let $eig\,(\cdot)$ indicate a generic eigenvalue of the argument matrix, and observe that $eig\,(\beta'\alpha) = (b_1 - 1)$ with multiplicity 2. Because $eig(A_1) = eig(A_1 - I_4) + 1 = eig(\alpha\beta') + 1$ and $eig(\alpha\beta')$ is either 0 (twice) or equal to $eig(\beta'\alpha)$ (2 eigenvalues), one finds that $|eig(\beta'\alpha) + 1| < 1$, i.e. $-2 < eig(\beta'\alpha) < 0$ when b_1 is chosen in $\mathcal{E} := \{0.4, 0.8\}$. Hence all the characteristic roots of the autoregressive polynomial are either equal to 1 or are outside the unit circle. The conditions of Granger's representation theorem are thus satisfied, with CI rank $r = 2$, and $p - r = 2$ common trends.

The normalization matrix c is selected as follows:

$$c' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \text{so that} \qquad c'\beta = \begin{pmatrix} 0 & 1 \\ a & 1-a \end{pmatrix}.$$

This makes the normalization $c'\beta$ a valid one for $a \neq 0$, and an invalid one for $a = 0$.

The MC design sets $(a,\, b_1,\, R^2) \in \mathcal{D} \times \mathcal{E} \times \mathcal{F}$ with $\mathcal{D} := \{0, 0.1, 0.2, 0.3, 0.5, 1\}$, $\mathcal{E} = \mathcal{F} := \{0.4, 0.8\}$, where b_2 is chosen as the positive square root of $R^2\,(1 - b_1^2)\,/\,(1 - R^2)$. Moreover T is selected in the set $\mathcal{T} := \{100, 200, 500\}$.

We simulated the sample paths of the process $\{X_t\}_{t=1}^{T}$ for $T = 500$, and selected the first 100 data as the sample for $T = 100$, and the first 200 data points as the sample for $T = 200$. No other parts of the sample were then reused. This implies that simulations for various sample sizes T are dependent; this gives a basic form of variance reduction technique, see e.g. Hendry (1984, Section 4.1) and also Paruolo (2002).

Each combination of a, b_1, R^2 results in a different data generating process, DGP. Table 1 lists the various DGPs, using the notation DGP_i for DGP number i. Note that DGP_1-DGP_{20} correspond to $s := \text{rk}(c'\beta) = 2$, while DGP_{21}-DGP_{24} to $s = 1$. Hence the procedures which test hypothesis $\mathcal{K} : s = 2$ should indicate that $\mathcal{K}$ is supported by the data in DGP_1-DGP_{20}, and it is not supported in DGP_{21}-DGP_{24}.

This design of experiments is limited, in the sense that a few DGPs are simulated for a single choice of $p = 4$, $r = 2$ and $k = 1$. This design is chosen here to allow us to compare results with the Monte Carlo simulations in Lukkonen, Ripatti and Saikkonen (1999), to which we refer for motivations of the choice of design.

4.2 Performance of the alternating algorithm

In this subsection we document the performance of the alternating algorithm described in Subsection 2.3, using as starting values $\beta^{(0)}$ the choices $\beta_{\langle 0 \rangle}$ and $\beta_{\diamondsuit}$ described in Subsection 3.2. We used the convergence criterion with $\xi = 10^{-3}$. Calculations were performed with Gauss 6.0 on a personal computer with Pentium 4 processor, running at 3 Ghz. Each iteration of the alternating algorithm took 0.01067 seconds on average.[5]

Table 2 reports the number of iterations n_{it} for the alternating algorithm using $\beta_{\langle 0 \rangle}$ as starting values $\beta^{(0)}$. The table reports the minimum and maximum n_{it}, along with quantiles q_η of the MC distribution of n_{it} where η indicates the probability in the left tail of the distribution. In particular the table reports q_η for $\eta = 0.1, 0.5, 0.9$. The left panel concerns $T = 100$ and the right panel $T = 500$. Results for $T = 200$ are similar and they not reported for brevity.

For DGP_{21}-DGP_{24} the number of iterations n_{it} is always very small, with maximum n_{it} equal to 8 for $T = 100$ and to 3 for $T = 500$. This shows that the performance of the algorithm is very good for the case where $\mathcal{H}_{0,1}$ is valid; in these cases one expects the unconstrained estimator $\widetilde{\beta}$ to be close to the constrained estimator and the choice of starting values $\beta_{\langle 0 \rangle}$ described in Subsection 3.2 to be a very good starting point. This is indeed the case.

This behavior is not expected when $\mathcal{H}_{0,1}$ fails, i.e. in DGP_1-DGP_{20}. For these DGPs there is no reason for the unconstrained ML estimator $\widetilde{\beta}$ to be close to the constrained estimator $\widehat{\beta}$. Moreover, one may expect a misspecified likelihood, as the one for DGP_1-DGP_{20} to be numerically more difficult to maximize than a correctly specified one. For DGP_1-DGP_{20} the median n_{it} is inferior to 20 across DGPs. The 90% quantile is inferior to 800, with associated computing time approximately equal to 8.54 seconds. This means that in 90% of the simulations for DGP_1-DGP_{20}, the number of iterations and the associated computing times were reasonably good.

The last 10% of the simulations has quite large number of iterations, with maximum equal to 26650. This unsatisfactory performance of the alternating algorithm is however met in a very small percentage of cases.[6]The maximum number of iterations n_{it} in DGP_1-DGP_{20} appears to increase with T: it is 12204 for $T = 100$, and 26650 for $T = 500$. $n_{\text{it}} = 26650$ corresponds to 4 minutes 45 seconds of computing time. The fact that the maximum n_{it} increases with time may be associated with the fact

[5]The description of the random number generator and of other properties of the Gauss programming enviroment are available at `http://www.aptech.com/whitepapers.html`.

[6]From a pratical point of view, if computing time exceeds a certain threshold, the econometrician is lead to stop iterations and restart them with a different choice of initial values and/or with a different algorithm. Hence this maximum computing time is arguably never met in practice.

		$T =$	100				$T =$	500		
DGP_i	min	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$	max	min	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$	max
$i = 1$	2	2	2	105.5	2197	2	2	2	2	2735
2	2	2	2	93	3833	2	2	2	2	1107
3	2	2	4	157.5	2024	2	2	2	136	9287
4	1	2	2	569	12204	2	2	2	2	26650
5	1	2	7	105.5	2497	2	2	2	137	3972
6	2	2	2	311	5064	2	2	2	2	5579
7	2	2	8	113	1371	2	2	2	358	5782
8	2	2	3	609.5	7825	2	2	2	3	17256
9	2	2	5	50	1073	2	2	8	158	4201
10	1	2	11	285	6911	2	2	2	172	7418
11	2	2	6	63	1335	2	2	11	286.5	4652
12	2	2	16	431	5838	2	2	2	511	18229
13	1	2	4	15	949	2	2	9	72	4023
14	2	2	12	179	2959	2	2	2	316	5884
15	1	2	4	26	1145	2	2	10	162.5	4158
16	2	2	14	273	4618	2	2	2	771	23025
17	1	2	3	5	732	2	2	5	13	2545
18	2	2	6	26	1915	2	2	16	147	3841
19	1	2	3	6	368	2	2	5	18	3122
20	2	2	6	70	3899	2	2	19	403	8725
21	1	2	2	2	4	1	2	2	2	2
22	1	2	2	2	4	1	2	2	2	2
23	1	2	2	2	8	1	2	2	2	3
24	1	2	2	2	5	1	2	2	2	2

Table 2: Summary statistics of the number of iterations n_{it} of the alternating algorithm in 10^4 replication, for the choice of starting values $\beta_{\langle 0 \rangle}$. $T = 100$ left panel, $T = 500$ right panel. q_η is the η-quantile of the Monte Carlo distribution of n_{it}.

		$T =$	100				$T =$	500		
DGP_i	min	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$	max	min	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$	max
1	1	1	1	78	1663	1	1	1	3	4774
2	1	1	1	53.5	3173	1	1	1	1	1673
3	1	1	4.5	134	1869	1	1	1	100	5501
4	1	1	1	374	6282	1	1	1	1	6903
5	1	1	3	92	1640	1	1	1	75	3353
6	1	1	1	169	2955	1	1	1	1	6227
7	1	1	6	102	1718	1	1	1	224	3986
8	1	1	1	379.5	7213	1	1	1	18	14545
9	1	1	3	37	1353	1	1	1	91	4480
10	1	1	4	188	2345	1	1	1	59.5	6395
11	1	1	4	56	1069	1	1	3	210	3676
12	1	1	9	326	6198	1	1	1	241.5	9277
13	1	1	2	10	903	1	1	2	32	3150
14	1	1	4	125	2502	1	1	1	127	4599
15	1	1	3	23	1133	1	1	4	114	4950
16	1	1	8	213	3330	1	1	1	380	11992
17	1	1	2	4	850	1	1	2	5	2433
18	1	1	3	15	1904	1	1	3	56	5043
19	1	1	2	5	361	1	1	2	10	2236
20	1	1	4	52	2310	1	1	5	248	4891
21	1	1	2	2	8	1	1	1	2	3
22	1	1	2	2	5	1	1	1	2	3
23	1	1	2	3	12	1	1	1	2	4
24	1	1	2	3	7	1	1	1	2	4

Table 3: Summary statistics of the number of iterations n_{it} of the alternating algorithm in 10^4 replication, for the choice of starting values $\beta_\diamondsuit$. $T = 100$ left panel, $T = 500$ right panel. q_η is the η-quantile of the Monte Carlo distribution of n_{it}.

that the convergence criterion uses the same tolerance $\xi = 10^{-3}$, and ℓ_T increases linearly in T.

We also performed the same set of simulations using $\beta_\diamondsuit$ as starting values. The complete set of simulations for all DGPs and sample sizes took 2 hours 16 minutes for the choice $\beta_{\langle 0 \rangle}$ of starting values and 1 hour 33 minutes for the choice $\beta_\diamondsuit$. This implies approximately a 48% decrease in average computing time when using $\beta_\diamondsuit$ instead of $\beta_{\langle 0 \rangle}$ as starting values.

Table 3 reports summary statistics of n_{it} for the choice $\beta_\diamondsuit$ of starting values, for $T = 100$ and 500. We first compare results in Tables 2 and 3 for DGP_1-DGP_{20}. The median of n_{it} is at most 9 across DGPs, and the 90% quantile is inferior to 400 both for $T = 100$ and 500. The sample size T does not appear to influence n_{it} in a systematic way, except for the maximum of n_{it}, which is 6282 for $T = 100$ and 14546 for $T = 500$.

For these DGPs, the choice $\beta_\diamondsuit$ of starting values provides a sizable improvement over $\beta_{\langle 0 \rangle}$ at the right tail of the n_{it} distribution. In fact the 90% quantile and the maximum of n_{it} for $\beta_\diamondsuit$ are half the size than for the corresponding DGP using $\beta_{\langle 0 \rangle}$ as $\beta^{(0)}$.

The computing time required to perform the maximum number of iterations across DGPs for $\beta_\diamondsuit$, 14546, is 2 minutes 35 seconds. This is arguably not such a large computing time across $24{\cdot}10^4$ simulations. It thus appears that $\beta_\diamondsuit$ provides a better starting point than $\beta_{\langle 0 \rangle}$ for the alternating algorithm, and a reasonably good one.

For DGP_{21}-DGP_{24} results for $\beta_\diamondsuit$ and $\beta_{\langle 0 \rangle}$ are similar and very good, with fewer iterations for $T = 500$ than for $T = 100$. The maximum n_{it} is 12 for $T = 100$ and 3 for $T = 500$. Hence both $\beta_\diamondsuit$ and $\beta_{\langle 0 \rangle}$ provide remarkably good starting points for these DGPs.

Overall, the present results on n_{it} indicate that the alternating algorithm delivers a sensible performance. $\beta_\diamondsuit$ provides a comparably better starting point than $\beta_{\langle 0 \rangle}$, and it is hence the choice of starting values recommended in practice.

We next present result concerning the performance of the tests.

4.3 Test performance

In this subsection we report the performance of the Q_j tests. The test outcomes using $\beta_\diamondsuit$ or $\beta_{\langle 0 \rangle}$ as starting values $\beta^{(0)}$ were identical, as expected from Proposition 1.(ii). Because $m = r = p - r = 2$, two Q_j tests can be performed, i.e. Q_1 and Q_2. The test Q_2 rejected in all samples for all DGPs. Hence we confine attention to the performance of the Q_1 test.

The Q_1 test is compared with a representative tests of the class proposed in Lukkonen, Ripatti and Saikkonen (1999); we chose $\widetilde{\lambda}_{LM}$ in their notation because this is described as the test with best Monte Carlo performance. This test takes $s = 2$ as the null and $s < 2$ as the alternative.

Another alternative test is the Wald test $\mathcal{L}$ proposed in Kurozumi (2005, p. 303). The null hypothesis is $s = f$ against an alternative $s > f$; we indicate these tests as $\mathcal{L}_f$. In the present situation one can consider $\mathcal{L}_0$ and $\mathcal{L}_1$, which correspond to Q_2 and Q_1 respectively. Under the null $\mathcal{L}_0$ is asymptotically $\chi^2(4)$ and $\mathcal{L}_1$ is asymptotically $\chi^2(1)$. We performed both tests $\mathcal{L}_0$ and $\mathcal{L}_1$. As for the test Q_2, $\mathcal{L}_0$ always rejected in all samples for all DGPs. Therefore, also in this case, results are presented only for the $\mathcal{L}_1$ test.

	$T=$	100		$T=$	200		$T=$	500	
DGP_i	$\widetilde{\lambda}_{LM}$	$\mathcal{L}_1$	Q_1	$\widetilde{\lambda}_{LM}$	$\mathcal{L}_1$	Q_1	$\widetilde{\lambda}_{LM}$	$\mathcal{L}_1$	Q_1
$i=1$	4.3	0	0	4.8	0	0		0	0
2	3.3	0	0	4.0	0	0		0	0
3	8.5	0	0.04	8.0	0	0	6.4	0	0
4	6.3	0	0	6.1	0	0		0	0
5	6.6	0	0	6.5	0	0		0	0
6	3.7	0	0	4.2	0	0		0	0
7	11.6	0.02	0.2	10.1	0	0	7.3	0	0
8	6.7	0	0	6.3	0	0		0	0
9	11.9	0	0.06	10.4	0	0		0	0
10	4.2	0	0	4.6	0	0		0	0
11	18.4	0.14	0.97	14.9	0	0	9.1	0	0
12	7.6	0	0	6.7	0	0		0	0
13	20.6	0.18	0.59	17.3	0	0		0	0
14	5.5	0	0	5.5	0	0		0	0
15	28.5	1.38	3.5	22.3	0	0.01	12.6	0	0
16	9.1	0	0	7.4	0	0		0	0
17	46.8	6.25	10.22	41.8	0.03	0.07		0	0
18	11.8	0	0	9.7	0	0		0	0
19	54.1	14.4	23.4	46.6	0.49	0.98	28.6	0	0
20	16.4	0	0	11.5	0	0		0	0
21	96.4	89.63	93.23	99.2	92.93	94.47		93.93	94.44
22	98.7	91.83	94.14	99.6	93.71	94.85		94.30	94.66
23	93.8	85.30	91.52	98.9	91.53	93.85	99.8	93.46	94.39
24	98.4	89.73	93.18	99.6	92.49	94.41		94.06	94.62

Table 4: Monte Carlo percentage frequency of rejections of $\mathcal{K}: s=2$ for the tests $\widetilde{\lambda}_{LM}$, $\mathcal{L}_1$, Q_1. $\mathcal{K}: s=2$ holds for DGP_1-DGP_{20} and does not hold for DGP_{21}-DGP_{24}. The columns labelled $\widetilde{\lambda}_{LM}$ are taken from Lukkonen, Ripatti, Saikkonen (1999), Tables 1, 2, 3.

The performance of the tests $\widetilde{\lambda}_{LM}$, $\mathcal{L}_1$, Q_1 are reported in Table 4. The table reports 100 times the frequency with which the tests reject $\mathcal{K} : s = 2$. Recall that the correct decision is to reject $\mathcal{K}$ in DGP_{21}-DGP_{24} and decide in favor of $\mathcal{K}$ in DGP_1-DGP_{20}. Entries for $\widetilde{\lambda}_{LM}$ are taken from Tables 1, 2, 3 of Lukkonen, Ripatti and Saikkonen (1999); they report results only for a subset of the designs for $T = 500$, and this is reflected in some missing entries in the corresponding column.

Consider DGP_1-DGP_{20} for $T = 100$. It is seen that the power of $\mathcal{L}_1$ and Q_1 is very close to 1 also in cases of small a, close to the case of incorrect normalization, thus giving the correct decision in favor of $\mathcal{K}$ with probability close to 1. The $\widetilde{\lambda}_{LM}$ test, instead, should give the incorrect decision approximately 5% of the time, because of the choice of null hypothesis. In practice $\widetilde{\lambda}_{LM}$ decides against $\mathcal{K}$ much more frequently than in 5% of the times; this frequency is higher than 10% in DGP_i for i =7, 9, 11, 13, 15, 17, 18, 19. For these DGPs the $\mathcal{L}_1$ and Q_1 tests deliver a better performance.

We next compare the $\mathcal{L}_1$ and Q_1 tests for DGP_{21}-DGP_{24} and $T = 100$. Q_1 does not reject the null between 92% and 94% of the time, close to 1 minus the nominal size of the test. The $\mathcal{L}_1$ test has a similar behavior, even though the frequency of acceptance ranges between 85% and 92%, sightly less than to 1 minus the nominal level. Hence the Q_1 test appears to give the best overall performance among the three tests for $T = 100$.

For sample sizes $T = 200$ and $T = 500$, the decisions based on $\mathcal{L}_1$ and Q_1 progressively converge; their performance is similar to the one described above for the Q_1 test at $T = 100$. One observes that in DGP_{21}-DGP_{24} where $\mathcal{K}$ should be rejected, the power of $\widetilde{\lambda}_{LM}$ converges to 1 as T increases, giving the correct decision with asymptotic probability 1. In the same situation Q_1 and $\mathcal{L}_1$ give the correct decision with probability 95%, due to the choice of the null hypothesis. Overall the Q_1 test appears to deliver the most accurate decisions.

We observe that the significance level of all the tests could be reduced to zero when T increases. This would give a limit probability 1 of delivering the correct decision; see e.g. Poetscher (1983).

5 Conclusions

In this paper we have compared the LR tests on the rank of a submatrix of the matrix of CI relations with alternatives. The MC simulation study shows that the LR test has a better finite sample performance than the tests proposed in Saikkonen (1999) for the Monte Carlo design used in Lukkonen, Ripatti and Saikkonen (1999). The performance of the Wald test of Kurozumi (2005) is similar to the one of the LR test, although for sample size equal to 100 the LR test performs marginally better.

The alternating maximization algorithm always terminates, usually with very few and fast iterations. When the null hypothesis is correctly specified, the number of iteration was always less than a dozen. For the DGPs where the null hypothesis was incorrect, in 90% of the simulations the number of iteration was reasonably low. In the worst 10% of the simulations, the number of iterations became sometimes very large, but still with a maximum computing time below 5 minutes.

The choice of starting values is also investigated in this paper, and it is found to affect the performance of the alternating maximization algorithm, where a careful choice of starting values can reduce the number of iteration by a factor of 1/2.

References

Anderson, T.W. (1951). 'Estimating linear restrictions on regression coefficients for multivariate normal distributions', *The Annals of Mathematical Statistics*, Vol. 22, pp. 327–351. Correction, *The Annals of Statistics* Vol. 8, 1980, pp. 1400.

Boswijk, H. P. and J. A. Doornik (2004). 'Identifying, estimating and testing restricted cointegrated systems: An overview', *Statistica Neerlandica*, Vol. 58, pp. 440-465.

Hendry D. (1984). 'Monte Carlo experimentation in econometrics', Chapter 16 in Z. Grilitches and M.D. Intriligator (eds) *Handbook of econometrics*, Vol. 2, North Holland, Amsterdam.

Johansen S. (1996). *Likelihood–based inference in cointegrated vector auto-regressive models*. Oxford: Oxford University Press, second printing.

Johansen, S. and K. Juselius (1992). 'Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK', *Journal of Econometrics*, Vol. 53, 211-244.

Kurozumi, E. (2005). 'The Rank of a Sub-Matrix of Cointegration', *Econometric theory* Vol. 21, pp. 299-325.

Lukkonen R., A. Ripatti and P. Saikkonen (1999). 'Testing for a valid normalization of cointegrating vectors in vector autoregressive processes', *Journal of Business and Economic Statistics*, Vol. 17, pp. 195-204.

Oberhofer W. and J. Kmenta (1974). 'A general procedure for obtaining maximum likelihood estimates in generalized regression models', *Econometrica*, Vol. 42, pp. 579-590.

Paruolo P. (2002). 'On Monte Carlo Estimation of Relative Power', *Econometrics Journal*, Vol. 5, pp. 65-75.

Paruolo P. (2007). 'The likelihood ratio test for the rank of a cointegration submatrix', *Oxford Bulletin of Economics and Statistics*, forthcoming.

Saikkonen, P. (1999). 'Testing normalization and overidentification of cointegrating vectors in vector autoregressive processes', *Econometric Reviews* Vol. 18, pp. 235-257.

Sargan J. D. (1964) 'Wages and prices in the United Kingdom: a study in econometric methodology', in Hart P. E., G. Mills and J. K. Whitaker (eds) *Econometric analysis for national economic planning*, London: Butterworths.

A Appendix

Proof. of Proposition 1. We wish to show that the hypothesis of Lemma 1 in Oberhofer and Kmenta (1974) are satisfied in this case. The level set S is non-empty because it contains the starting values $a^{(0)}$, and it is bounded because each column in a_i lies on an unit sphere of appropriate dimension, $i = 1, 2$. Moreover ℓ_T is continuous in a on S. Finally the parameter space $\mathcal{U}$ is closed. Thus the hypothesis of Lemma 1 in Oberhofer and Kmenta (1974) are satisfied and (i) (ii) (iii) follow. (iv) is a consequence of (iii). Finally we prove (v) by contradiction. Assume that $n_{\text{it}} > n_{\text{it}}^{\max}$. At each iteration $h \leq n_{\text{it}}$, the increment $\ell_T\left(a^{(h)}\right) - \ell_T\left(a^{(h-1)}\right) > \xi$, because otherwise the algorithm stops. Summing over h,

$$\sum_{h=1}^{n_{\text{it}}^{\max}} \left(\ell_T\left(a^{(h)}\right) - \ell_T\left(a^{(h-1)}\right)\right) = \ell_T\left(a^{(n_{\text{it}}^{\max})}\right) - \ell_T\left(a^{(0)}\right) > \xi n_{\text{it}}^{\max} = \ell_T(\widetilde{\beta}^*) - \ell_T(a^{(0)})$$

which implies $\ell_T\left(a^{(n_{\text{it}}^{\max})}\right) > \ell_T(\widetilde{\beta}^*)$, giving a contradiction. ■