Paolo Paruolo

A LR rank test for a cointegration sub-matrix

2004/35

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

Printed in Italy in December 2004
Università degli Studi dell'Insubria
Via Monte Generoso 71, 21100 Varese, Italy

The likelihood ratio test for 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 proposes a likelihood ratio test for rank-deficiency of a submatrix of the cointegrating matrix. 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. The proposed test has χ^2 limit distribution and indicates the validity of the normalization with probability one in the limit, for valid normalizations. The asymptotic properties of several derived estimators of the rank are also discussed. It is found that a testing procedure that starts from the hypothesis of minimal rank is preferable.

JEL classification numbers: C12, C32, C52.
Word count: 7478

*Paper presented at the conference 'Frontiers in Time Series Analysis' Olbia, Italy, May 29-31, 2005. 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. I wish to thank Søren Johansen, Pentti Saikkonen, three anonymous referees and the conference participants for useful comments on a previous version of the paper.

I Introduction

This paper presents a likelihood ratio (LR) test for deficient rank of a submatrix of the cointegration (CI) matrix β in vector autoregressive models (VAR). A leading special case is given by tests for valid normalization of cointegrating vectors. Other special cases are feasibility of permanent-transitory decompositions of the type defined in Gonzalo and Granger (1995), tests of block long run Granger prior noncausality, see Bruneau and Jondeau (1999) and Yamamoto and Kurozumi (2006), validity of χ^2 asymptotics for the Wald test of Granger noncausality in Toda and Phillips (1993), feasibility of assigned long run changes of the type described in Johansen (2005) for the interpretation of cointegrating vectors.

The test is based on a representation result for the intersection of linear subspaces. As a by-product of this representation, the paper shows that the proposed LR test is also a rank test for a submatrix of a basis $\beta_\perp$, say, of the orthogonal complement of the cointegration space. Hence the proposed test can serve to test rank hypotheses both on submatrices of the cointegration matrix and of its orthogonal complement.

The normalization of cointegrating vectors forms the first step in the analysis of the cointegration space, once the dimension of the cointegration space has been fixed, see Johansen (1996). An instance of this normalization is the 'triangular form' used e.g. in Phillips (1991); this corresponds to the reduced form of the system of long run relations in terms of a subset of variables interpreted as dependent variables. Not all subsets of variables can be chosen as dependent variables, and the test presented here can serve to ascertain if any specified subset is a viable one.

Lukkonen, Ripatti and Saikkonen (1999) and Saikkonen (1999) were the first to consider this 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 $\beta_\perp$.

The tests proposed in this paper differ with respect to the tests in the literature reported above in the following aspects. First, they are LR tests; see Wald (1942), Andrews (1996) and reference therein for optimality properties of LR test in a classical context. Second, using the duality results mentioned above, it is shown that the same test can be used to test hypotheses on β and $\beta_\perp$. Third: the LR tests are a special case of a class of LR tests already considered in Johansen and Juselius (1992) for which computer software is already available. Fourth: the calculation of the test gives restricted maximum likelihood (ML) estimates of β under the null of incorrect normalization; this may help the econometrician to amend the specification of the normalization, in case the null is not rejected. Fifth: like the Kurozumi tests, the limit distribution of the LR test statistics is χ^2, so that asymptotic inference is standard, and no new tables need to be simulated. A final characteristic that the present LR test shares with the Kurozumi tests is the formulation of the null as a reduced rank hypothesis, and the property of the test of being consistent under the alternative, i.e. under the correct specification of normalizations. Thus a valid normalization is signaled with limit probability one by the tests.

The latter property has notable consequences concerning specification searches. If several (possibly dependent) specification tests are performed in a specification procedure, the econometrician may be compelled to bound the overall type-I error by employing some Bonferroni-type inequality, where the overall bound on size is computed as the sum of individual test sizes. Because the null is the one of incorrect normalization, which is rejected with probability one in the limit, the inclusion of this test does not give any contribution to the overall bound. In this sense, therefore, using this test is costless from the point of view of the overall asymptotic bound on size. This is not the case for the tests proposed by Saikkonen and co-authors, which consider the hypothesis of correct normalization as the null.

In the following we indicate by $\mathcal{A} := \text{col}(A)$ the linear space generated by a matrix A and by $A_\perp$ a basis of the orthogonal complement of $\text{col}(A)$, indicates as $\mathcal{A}^\perp = \text{col}^\perp(A)$. 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 $\text{col}(A)$. We let $\overset{w}{\to}$ and $\overset{p}{\to}$ indicate weak convergence and convergence in probability, respectively, for sample size $T \to \infty$.

The rest of the paper is organized as follows. Section II reports notation and introduces several motivating examples. Section III discusses the hypothesis of interest, presents the LR test and its limit distribution; this section also discusses the properties of rank-determination procedures based on the LR test. Section IV summarizes simulation results obtained in Paruolo (2006) on the finite sample performance of the tests and reports conclusions. Proofs are placed in the Appendix.

II Motivation

In this section we introduce notation and several motivating examples. In Subsection II.1 we introduce VAR processes integrated of order 1, I(1), and their equilibrium correction (EC) formulation. Subsection II.2 defines the relevant statistical model and presents the issue of valid normalization of the CI vectors. In Subsection II.3 we describe the possibility to use nonorthogonal permanent-transitory decompositions. Subsection II.4 relates the hypothesis of interest with the interpretation of cointegrating coefficients and with controllability of the system. Finally subsection II.5 discusses the relation with testing long run Granger noncausality.

II.1 Vector autoregressive processes

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,

[1]Several assumptions are made here for simplicity. For instance, similar arguments apply to other specification of the deterministic components, see Section III.4. The same asymptotic results can also be obtained under less restrictive conditions on ε_t and assuming that initial values are realizations of random variables whose distribution does not depend on μ_0, μ, Ω, A_i, $i = 1, ..., k$.

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

$$\operatorname{rk}(\alpha_{\perp}'\Gamma\beta_{\perp}) = p - r, \tag{1}$$

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, see Johansen (1996) and references therein:

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

II.2 Normalization of cointegrating vectors

In this subsection we describe the reference statistical model and discuss normalization of the CI matrix β.

The model 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. The parameter vector θ is not identified within $\mathcal{H}(r)$, see Johansen (1996), Boswijk (1996) and the discussion below.

The system of r equations

$$\beta' X_t + \phi' t = u_t \tag{3}$$

represents long run relations among the level of the variables. Here u_t is a $r \times 1$ vector of I(0) processes. The system (3) can be interpreted as a simultaneous system of equations, see Johansen (1995).

The matrix β is a basis of the cointegration space $\mathcal{B} := \operatorname{col}(\beta)$, where $\operatorname{col}(A)$ denotes the column space of the matrix A. We also use the notation $A_{\perp}$ for a basis of the orthogonal complement of $\operatorname{col}(A)$, indicated as $\operatorname{col}^{\perp}(A)$. It is well known that $\mathcal{B}$ is identified within model $\mathcal{H}(r)$, but β is not, see e.g. Johansen (1996, Chapter 5.2). In fact, let a be a square nonsingular matrix; any other basis $\beta^{\dagger} := \beta a$ can be used in the likelihood function in place of β obtaining the same likelihood value, provided also α, ϕ are replaced by $\alpha^{\dagger} := \alpha a'^{-1}$, $\phi^{\dagger} := \phi a$.

One possibility to obtain a just-identified β is to consider the 'triangular form' used in Phillips (1991). This corresponds to the reduced form of the system of long run relations (3) obtained by solving the equations for a block of r variables in X_t. Let $X_t := (X_{1t}' : X_{2t}')'$, where X_{1t} is $r \times 1$ and assume that one wishes to solve for X_{1t}. Note that it is always possible to reorder the variables such that this is the case. Let $\beta := (B : \Psi)'$ be the corresponding partition of β, where B is $r \times r$ and square. The triangular form is obtained by premultiplication of (3) by B^{-1}, when B is nonsingular, obtaining $\beta^{\circ} := (I_r : -\Pi)'$, $\Pi := -B^{-1}\Psi$, where I_r is the identity matrix of order r.

Noting that $B' = c'\beta$ for $c := (I_r : 0)'$, the triangular form β° is seen to be a special case of $\beta_c := \beta(c'\beta)^{-1}$, for the given choice of c, under the assumption of $c'\beta$ of full rank r. This paper considers hypotheses on $s := \operatorname{rk}(c'\beta)$. We call β_c the normalized CI matrix.

Many authors, including Phillips (1991) and Johansen (1996, p.78), have noted that β_c is identified within $\mathcal{H}(r)$. The next subsection describes possible permanent-transitory decompositions and their relation to $s := \mathrm{rk}(c'\beta)$.

II.3 Permanent-transitory decompositions

In this subsection we consider permanent-transitory decompositions similar to the one of Beveridge and Nelson. We wish to decompose the observed time series $\{X_t\}_{t=1}^T$ into two additive terms: a permanent I(1) component, indicated as X_t^P, and a transitory I(0) one, labeled X_t^T, where $X_t = X_t^P + X_t^T$. The X_t^P, X_t^T components are defined as linear combinations of the observed X_t.

From the CI properties of β, one can specify $X_t^T := \beta' X_t$, as in Gonzalo and Granger (1995) eq. (13), or as in Kasa (1992). We consider several possible choices for the linear combination X_t^P. Gonzalo and Granger (1995) propose $X_t^P := \alpha'_\perp X_t$, Kasa (1992) sets $X_t^P := \beta'_\perp X_t$. Another possible choice is $X_t^P := \alpha'_\perp \Gamma X_t$, as suggested in Johansen (1996, Corollary 4.4).

The above possibilities can be obtained through nonorthogonal projections, based on the condition that there exist a matrix $c_\perp$ of the same dimensions as $\beta_\perp$ such that $A := (\beta : c_\perp)'$ is a full rank $p \times p$ matrix. This allows to write

$$X_t = A^{-1}AX_t = c(\beta'c)^{-1}\beta'X_t + \beta_\perp(c'_\perp\beta_\perp)^{-1}c'_\perp X_t = c(\beta'c)^{-1}X_t^T + \beta_\perp(c'_\perp\beta_\perp)^{-1}X_t^P \tag{4}$$

for $X_t^T := \beta' X_t$, $X_t^P := c'_\perp X_t$.

The choice $c_\perp := \beta_\perp$ of Kasa (1992) satisfies the invertibility requirement for A because orthogonal rows are linearly independent, and $A := (\beta : \beta_\perp)'$ is hence invertible. The choice $c_\perp := \Gamma'\alpha_\perp$ is also seen to satisfy it because $c'_\perp\beta_\perp = \alpha'_\perp\Gamma\beta_\perp$ is of full rank, thanks to (1). Gonzalo and Granger (1995) assume that $c'_\perp\beta_\perp = \alpha'_\perp\beta_\perp$ is of full rank in order to obtain their decomposition; this assumption coincides with (1) if $k = 1$, but it is otherwise unwarranted, see Harbo et al. (1998) section 1.5 for a counterexample.

Also outside the realm of the above choices for $c_\perp$, an econometrician may have economic a priori rationale for a specific choice of $c_\perp$ in $X_t^P := c'_\perp X_t$. Then a valid permanent-transitory decomposition of the type (4) exists iff $(\beta : c_\perp)$ is of full rank. It is well known that, see e.g. Johansen (1996, Exercise 3.7) that this is equivalent to requiring $\mathrm{rk}\,(c'_\perp\beta_\perp) = p - r$, which is the same hypothesis as $s := \mathrm{rk}(c'\beta) = r$, as shown also below.

The next subsections describe different economic issues, that turn out to be also connected to $s := \mathrm{rk}(c'\beta)$.

II.4 Thought experiment

The interpretation of cointegrating coefficients can be associated with a thought experiment based on the long run forecast of the system, see Johansen (2005).

Let $X_{t+h|t}$ be the optimal linear predictor of X_{t+h} based on the complete past of the process $\{X_{t-s}, s \geq 0\}$, which is equivalent, due to the Markov structure of $\widetilde{X}_t := (X'_t : ... : X'_{t-k+1})'$, to the optimal linear predictor based on $\widetilde{X}_t$. Here the optimal predictor coincides with the condition expectation because of the Gaussianity of ε_t. When the normality assumption is dropped, the calculations are the same, but $X_{t+h|t}$ is interpreted as the best linear predictor in mean square sense.

For the VAR process in eq. (2) it can be shown, see e.g. Bruneau and Jondeau (1999), Johansen (2005), Omtzigt and Paruolo (2005), that when $h \to \infty$ one has

$$X_{\infty|t} = C\left(X_t - \sum_{i=1}^{k-1} \Gamma_i X_{t-i}\right),$$

where $C := \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp$ is the MA impact matrix, see e.g. Johansen (1996), where $\mathrm{rk}\,(C) = p - r$.

Johansen (2005) considers the following thought experiment: 'fix a $p \times 1$ vector q of changes in $X_{\infty|t}$, where $q = \beta_\perp \psi$; can one add some vector x to X_t in order to produce the given change q in $X_{\infty|t}$?'. This question is answered in the affirmative, and it turns out that there are many values of x that accomplish the given change, one being $x = \Gamma\beta_\perp\psi$.

Typically there is a subset of variables $Y_t := b'X_t$ that is interpreted as 'target' economic variables, where b is $p \times m$. One may ask if one can produce any given change in Y_t by the thought experiment given above. In other words: 'fix a $m \times 1$ vector v of changes in $Y_{\infty|t}$; can one find some vector $x = \Gamma\beta_\perp\psi$ to add to X_t in order to produce the given change v in $Y_{\infty|t}$? Is ψ unique?'.

Because of linearity, the change in $Y_{\infty|t}$ is b' times the change in $X_{\infty|t}$, i.e. $v = b'q = b'\beta_\perp\psi$. So the answer is affirmative, and ψ is unique iff $b'\beta_\perp$ is square and invertible, that is iff $m = p - r$ and $\mathrm{rk}\,(b'\beta_\perp) = p - r$. Again this is shown below to be related to $s := \mathrm{rk}(c'\beta)$.

This thought experiment is also related to the question of how to (long run) control a subset $b'X_t$ of variables in (2) using another subset of variables $a'X_t$ as instruments, see Johansen and Juselius (2001). Here a is $p \times n$ and such that $(a : b)$ is a full column rank matrix, and the notion of (long run) control is the one exposed in the thought experiment described above.

It turns out that a key condition for controllability when $n = m$ is $\mathrm{rk}(b'Ca) = m$. When $m = p - r$, this is equivalent to $\mathrm{rk}(b'\beta_\perp) = p - r$ and $\mathrm{rk}(a'\alpha_\perp) = p - r$. These examples show how hypotheses on $s := \mathrm{rk}(c'\beta)$ also arise in this context.

II.5 Long run neutrality and noncausality

In this subsection we describe the conditions for long run non causality as defined in Bruneau and Jondeau (1999), Yamamoto and Kurozumi (2006); also these conditions for long run noncausality turn out to be related to $s := \mathrm{rk}(c'\beta)$.

Long run Granger non causality is defined by the above authors in terms of $X_{\infty|t}$ as follows. Consider some target variables $b'X_t$ and some candidate causal variables $a'X_t$, where a, b are of dimension $p \times n$ and $p \times m$ respectively and $(a : b)$ is of full column rank, see the previous subsection. Bruneau and Jondeau (1999) take $n = m = 1$, Yamamoto and Kurozumi (2006) take $n + m \leq p$. Then $a'X_t$ is said not to Granger cause $b'X_t$ in the long run if $b'X_{\infty|t}$ does not depend on $(I_k \otimes a')\,\widetilde{X}_t = (X'_t a : X'_{t-1}a : \ldots : X'_{t-k+1}a)'$, which contains $a'X_t$ and its lags.

Bruneau and Jondeau (1999) show that this corresponds to the conditions[2]

$$b'Ca = 0, \tag{5}$$

$$b'C\Gamma_i a = 0, \qquad i = 1, \ldots, k-1. \tag{6}$$

[2] The conditions (5) and (6) can be phrased as restrictions on the first block of rows in the impact factors defined in Omtzigt and Paruolo (2005) Proposition 1, p. 40.

Condition (5) is also called long run neutrality by Yamamoto and Kurozumi (2006), see their definition 2. Tests of condition (5) are discussed in Paruolo (1997) when $n = m = 1$. Tests of (5), (6) are considered in Bruneau and Jondeau (1999) and Yamamoto and Kurozumi (2006), who explicitly recognize that the asymptotic covariance matrices used in their Wald tests are singular. The latter authors use a Kurozumi (2005) test for the possible singularity of the asymptotic covariance matrix.

We here observe that this is not the only approach to the problem, as shown in the rest of this subsection, which is based on Paruolo (1997). Consider (5); the singularity of the asymptotic covariance matrix used in the Wald tests mentioned above stems from the fact that (5) can be decomposed into 3 subhypotheses: the first one regards the possible rank reduction of $b'\beta_\perp$, the second one the possible rank reduction of $a'\alpha_\perp$. The third one considers (5) only for the subsets of a and b that have passed the full-rank assumption in the previous two tests. This is the strategy proposed in Paruolo (1997), and faces the third test only for the part of (5) that has a nonsingular asymptotic covariance matrix.

In order to illustrate, assume e.g. that $m = 2$ and consider the hypothesis $\mathrm{rk}\,(b'\beta_\perp) \leq i$, $i = 0, 1$. If one cannot reject $\mathrm{rk}\,(b'\beta_\perp) = 0$, then there is no need to test further, because (5) holds. On the contrary if one rejects both $\mathrm{rk}\,(b'\beta_\perp) = 0$ and $\mathrm{rk}\,(b'\beta_\perp) \leq 1$ one concludes $\mathrm{rk}\,(b'\beta_\perp) = 2$ and one can then use the knowledge $\mathrm{rk}\,(b'\beta_\perp) = 2$ in the further tests of the remaining subhypotheses corresponding to (5). If one rejects rank 0 but not rank ≤ 1, then one selects rank 1 for $b'\beta_\perp$ and estimates which linear combination of b belongs to the CI space, say b_1, and which other does not, say b_2. For b_1 there is no need of test (5) further, while one can continue with $b_2'Ca$, now assuming that $b_2'\beta_\perp$ has full rank.

All the examples of the last three subsections highlight the role of the hypothesis on $s := \mathrm{rk}(c'\beta)$ and $\mathrm{rk}\,(b'\beta_\perp)$, which correspond to the hypothesis of interest in this paper. We next present the statistical formulation of the problem.

III The likelihood ratio test

In this section we state the hypothesis of interest in Subsection III.1 and provide examples in Subsection III.2. A characterization of the null is presented in Subsection III.3. The statistical definition of the LR test is reported in Subsection III.4. The asymptotic distribution is stated in Subsection III.5. Finally Subsection III.6 discusses various possible procedures that combine such tests. Proofs are placed in the Appendix.

III.1 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

$$\mathcal{K} : s = r, \tag{7}$$

The complementary hypothesis to $\mathcal{K}$ in (7) is

$$\mathcal{H}_{0,j} : s \leq r - j, \qquad 0 < j \leq m := \min(r, p - r), \tag{8}$$

where j is the 'rank deficiency' index. This rank deficiency index cannot exceed r, because $0 \leq s \leq r$. On the other hand we show below 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.

We construct a test for $\mathcal{K}$ in (7) taking $\mathcal{H}_{0,j}$ in (8) 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

$$\mathcal{H}_{0,1} \supset \mathcal{H}_{0,2} \supset \ldots \supset \mathcal{H}_{0,m}. \tag{9}$$

Note that if $\mathcal{H}_{0,1}$ is false, one can conclude directly that $\mathcal{K}$ is valid. This suggests to use a single test of $\mathcal{H}_{0,1}$ to decide about $\mathcal{K}$. Various other options on how to combine the various tests of $\mathcal{H}_{0,j}$ to estimate s are also discussed in Subsection III.6.

III.2 Examples

In this subsection we present two examples that illustrate the main idea behind the formulation of the hypotheses underlying the LR tests in terms of $\mathcal{K}$ and $\mathcal{H}_{0,j}$.

Consider first the bivariate case, $p = 2$, with CI rank $r = 1$ and specification $c = (1 : 0)'$; β is 2×1, $\beta = (\kappa_1 : \kappa_2)'$ and c selects the first element of β, i.e. κ_1. β_c normalizes the CI vector $\beta = (\kappa_1 : \kappa_2)'$ by dividing by $\kappa_1 = c'\beta$, provided $\kappa_1 \neq 0$, i.e. $\mathrm{rk}(c'\beta) = 1$. This is hypothesis $\mathcal{K}$ in (7) above.

In order to ascertain if $\kappa_1 \neq 0$, we test $\kappa_1 = 0$ instead, which can be written $\mathrm{rk}(c'\beta) = \mathrm{rk}(\kappa_1) = 0$. This is hypothesis $\mathcal{H}_{0,1}$ in (8) above with $r = j = 1$. Thus in order to decide if $\mathcal{K}$ holds, we test $\mathcal{H}_{0,1}$ instead. If ones rejects $\mathcal{H}_{0,1}$ then one decides in favor of $\mathcal{K}$; conversely if $\mathcal{H}_{0,1}$ is not rejected then one rejects $\mathcal{K}$.

Consider next a more complicated example with $p = 7$, $r = 3$ and the specification $c = (I_3 : 0_{3\times 4})'$. This corresponds to a system of 3 equations involving 7 variables. This specification of c attempts to solve the system in triangular form taking the first 3 variables as the dependent variables X_{1t} and the remaining 4 as independent variables X_{2t}. The hypothesis $\mathcal{K}$ is that $c'\beta$ has rank 3, so that one can premultiply $\beta' X_t$ by $(\beta' c)^{-1}$ in order to find the reduced form of the system.

In order to decide about hypothesis $\mathcal{K}$ we consider instead hypothesis $\mathcal{H}_{0,1}$: $\mathrm{rk}(c'\beta) \leq 2$. If $\mathcal{H}_{0,1}$ is rejected, then one decides in favor of $\mathcal{K}$; conversely if $\mathcal{H}_{0,1}$ is not rejected then one concludes that $\mathcal{K}$ does not hold, i.e. that one cannot solve $\beta' X_t + \phi' t = u_t$ for X_{1t}.

Both these examples illustrate how decisions about $\mathcal{K}$ can be based on a single test of $\mathcal{H}_{0,1}$ in $\mathcal{H}(r)$. If interest is not centered on $\mathcal{K}$ but on the estimation of $s := \mathrm{rk}(c'\beta)$, then the LR tests of $\mathcal{H}_{0,j}$ within $\mathcal{H}(r)$ can be arranged to give estimators of s. The definition of estimators of s and their properties are postponed to Subsection III.6 below.

In the following subsection we consider an equivalent formulation of $\mathcal{H}_{0,j}$ which is later used to define the test.

III.3 Characterization of the hypothesis

As argued in Lukkonen, Ripatti and Saikkonen (1999) and Saikkonen (1999), the validity of the chosen normalization (7) can be characterized in terms of the column spaces $\mathcal{B} := \mathrm{col}(\beta)$, $\mathcal{B}^{\perp} := \mathrm{col}(\beta_{\perp})$, $\mathcal{C} := \mathrm{col}(c)$, $\mathcal{C}^{\perp} := \mathrm{col}(c_{\perp})$. The following result connects (7) and (8) with the dimension of the intersection of $\mathcal{B}$ and $\mathcal{C}^{\perp}$.

This characterization is subsequently used as the basis of the formulation of the null hypothesis $\mathcal{H}_{0,j}$.

It is well known that the intersection of two linear subspaces is itself a linear subspace. Moreover it is simple to see that $\dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)\leq m:=\min(r,p-r)$, where $r=\dim(\mathcal{B})$, $p-r=\dim(\mathcal{C}^{\perp})$. The following theorem applies.

Theorem 1 *One has*

$$s:=\operatorname{rk}(c'\beta)=r-j \qquad\Longleftrightarrow\qquad \dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)=j. \tag{10}$$

In particular $\mathcal{K}$ holds if and only if $\dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)=0$ i.e. $\mathcal{B}\cap\mathcal{C}^{\perp}=\{0\}$. Analogously hypothesis $\mathcal{H}_{0,j}$ in (8) holds if and only if $\dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)>0$.

We next note that the choice of $\mathcal{B}$ and $\mathcal{C}^{\perp}$ can be interchanged with the one of $\mathcal{C}$ and $\mathcal{B}^{\perp}$, given that the argument is symmetric with respect to β and c. This proves the following corollary.

Corollary 2 *Eq. (7) holds if and only if $\mathcal{B}^{\perp}\cap\mathcal{C}=\{0\}$, i.e. $\dim\left(\mathcal{B}^{\perp}\cap\mathcal{C}\right)=0$. Analogously hypothesis $\mathcal{H}_{0,j}$ in (8) holds if and only if $\dim\left(\mathcal{B}^{\perp}\cap\mathcal{C}\right)>0$.*

One can also prove that $\dim\left(\mathcal{B}^{\perp}\cap\mathcal{C}\right)=\dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)$ by showing that there is a linear relation between $\operatorname{rk}(c'\beta)$ and $\operatorname{rk}(c_{\perp}'\beta_{\perp})$. This is stated in the following theorem.

Theorem 3 *In the previous notation*

$$\operatorname{rk}\left(c_{\perp}'\beta_{\perp}\right)=p-2r+\operatorname{rk}\left(c'\beta\right).$$

This implies also that $\dim\left(\mathcal{B}^{\perp}\cap\mathcal{C}\right)=\dim\left(\mathcal{B}\cap\mathcal{C}^{\perp}\right)$. When in particular $\operatorname{rk}\left(c'\beta\right)=r$, one has $\operatorname{rk}\left(c_{\perp}'\beta_{\perp}\right)=p-r$.

Corollary 2 and Theorem 3 can be summarized by saying that hypotheses on $\operatorname{rk}\left(c_{\perp}'\beta_{\perp}\right)$ and $\operatorname{rk}\left(c'\beta\right)$ are not separate hypotheses, but they are the same one. In particular one can reformulate hypotheses on $\operatorname{rk}\left(c_{\perp}'\beta_{\perp}\right)$ as hypotheses on $\operatorname{rk}\left(c'\beta\right)$, and there is no need to devise separate LR test for $\operatorname{rk}\left(c'\beta\right)$ and $\operatorname{rk}\left(c_{\perp}'\beta_{\perp}\right)$.

We next recall how hypotheses on $\mathcal{B}\cap\mathcal{C}^{\perp}$ can be formulated, see Johansen and Juselius (1992, pp. 225-226). This gives an explicit formulation of hypothesis $\mathcal{H}_{0,j}$ in terms of β.

Corollary 4 *$\mathcal{H}_{0,j}$ in (8) holds if and only if*

$$\beta:=(\beta_1:\beta_2)=(H_1\varphi_1:H_2\varphi_2) \tag{11}$$

with $H_1=c_{\perp}$, $H_2=I_p$, φ_1 of dimension $(p-r)\times j$, and φ_2 of dimension $p\times(r-j)$, (both of full column rank).

We hence consider (11) for unrestricted φ_i, $i=1,2$, as the formulation of hypothesis $\mathcal{H}_{0,j}$.

We note that the ability to use the format (11) 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 (11) can be stated in terms of the extended matrix $\beta^* := (\beta' : \phi')'$ as follows:

$$\beta^* = (H_1^*\varphi_1^* : H_2^*\varphi_2^*), \qquad H_i^* := \begin{pmatrix} H_i & 0 \\ 0 & 1 \end{pmatrix}, \qquad i = 1, 2 \tag{12}$$

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. (11), 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 (11). Statistical calculations are discussed in the following subsection.

III.4 The likelihood ratio test

In this subsection we describe how one calculates the LR test of (11) within the $\mathcal{H}(r)$ model. ML estimation 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 (11) 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). Any other numerical maximization algorithm may be used in its place to produce the restricted ML estimator of β^*, indicated here as $\widehat{\beta}^*$. The formulas below would then hold inserting $\widehat{\beta}^*$ as 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

$$\ell_T(\beta^*) := -\frac{T}{2}\left(p + \ln|S_{00}| + \ln\frac{|\beta^{*\prime}S_{11.0}\beta^*|}{|\beta^{*\prime}S_{11}\beta^*|}\right), \tag{13}$$

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 $\{\widetilde{\beta}^*, \widetilde{\lambda}_{1:r}\} := \mathrm{RRR}\left(Z_{0t}, Z_{1t}; Z_{2t}, r\right)$. 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 (13).

Under the null hypothesis $\mathcal{H}_{0,j}$ the maximum of $\ell_T(\theta)$ 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. Paruolo (2006) discusses its properties and the choice of starting values.

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:

$$\begin{aligned}
\{\varphi_1^{*(h)}, \rho_{1:j}^{(h)}\} &:= \mathrm{RRR}\left(Z_{0t}, H_1^{*\prime} Z_{1t}; (Z_{1t}'\beta_2^{*(h-1)} : Z_{2t}')', j\right), \\
\{\varphi_2^{*(h)}, \zeta_{1:r-j}^{(h)}\} &:= \mathrm{RRR}\left(Z_{0t}, H_2^{*\prime} Z_{1t}; (Z_{1t}'\beta_1^{*(h)} : Z_{2t}')', r-j\right).
\end{aligned} \tag{14}$$

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. $\ell_T\left(\beta^{*(h)}\right) - \ell_T\left(\beta^{*(h-1)}\right) < \xi$. 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 $\widehat{\beta}^* = \left(\widehat{\beta}_1^* : \widehat{\beta}_2^*\right)$, the maximum $\ell_{T,j}$ of $\ell_T(\theta)$ under $\mathcal{H}_{0,j}$ can be calculated as

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

see (13), where for the alternating algorithm considered here $\widehat{\beta}_i^* = \beta_i^{*(n_{\mathrm{it}})}$, $i = 1, 2$, see (14). The LR test of $\mathcal{H}_{0,j}$ within $\mathcal{H}(r)$ is thus given by[3]

$$
\begin{aligned}
Q_j &:= -2\left(\ell_{T,j} - \ell_{T,A}\right) = T\left(\ln \frac{\left|\widehat{\beta}_1^{*\prime} S_{11.0}^{*} \widehat{\beta}_1^{*}\right|}{\left|\widehat{\beta}_1^{*\prime} S_{11}^{*} \widehat{\beta}_1^{*}\right|} + \sum_{i=1}^{r-j} \ln\left(1-\widehat{\zeta}_i\right) - \sum_{i=1}^{r} \ln\left(1-\widetilde{\lambda}_i\right)\right) \\
&= T\left(\sum_{i=1}^{j} \ln\left(1-\widehat{\rho}_i\right) + \ln \frac{\left|\widehat{\beta}_2^{*\prime} S_{11.0}^{*} \widehat{\beta}_2^{*}\right|}{\left|\widehat{\beta}_2^{*\prime} S_{11}^{*} \widehat{\beta}_2^{*}\right|} - \sum_{i=1}^{r} \ln\left(1-\widetilde{\lambda}_i\right)\right) \qquad (15) \\
&= T\left(\ln \frac{\left|\widehat{\beta}^{*\prime} S_{11.0}^{*} \widehat{\beta}^{*}\right|}{\left|\widehat{\beta}^{*\prime} S_{11}^{*} \widehat{\beta}^{*}\right|} - \sum_{i=1}^{r} \ln\left(1-\widetilde{\lambda}_i\right)\right).
\end{aligned}
$$

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 (11) in place of (12) and omitting t from Z_{1t}. The case of restricted constant, $\mu_0 = \alpha\phi_0'$ is similar to the present case with (11) in place of (12), 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 (11) in place of (12). 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 (15) 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 (15) above setting $j = r$ and omitting the term with the ratio of two determinants. This corresponds to a special case of hypothesis $\mathcal{H}_5$ in eq. (15) or of hypothesis $\mathcal{H}_4$ in eq. (14) of Johansen and Juselius (1992), with H_4 of dimension $p \times r$ (their notation).

III.5 Asymptotics

In this subsection we report the asymptotic distribution for $T \to \infty$ of the LR test Q_j defined in (15). The limit distribution of the class of tests for hypotheses of this form has been discussed in Boswijk (1996), see also Johansen (1991, Appendix C) and Saikkonen (1999, Section 3.3). We here re-state asymptotic results in the following Theorem 5; the proof of this theorem can be found in Boswijk (1996), except for the proof that Q_j diverges when $j > r - s$, which is given in the Appendix.

We use the notation $\preceq$ to signify stochastic dominance, defined as follows. Let X and Y be two nonnegative random variables; $X \preceq Y$ indicates that $\Pr(X > c) \leq \Pr(Y > c)$ for all $c > 0$. Finally let $\overset{w}{\to}$ denote weak convergence.

[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).

Theorem 5 *Let $Y_j \sim \chi^2(j^2)$, a chi square distribution with j^2 degrees of freedom, let Q_j be the LR test in* (15) *and let $s := \mathrm{rk}(c'\beta)$. Then:*

1. *$Q_j \xrightarrow{w} Y_j$ for $j = r - s$;*

2. *$Q_j \xrightarrow{w} Y \preceq Y_j$ for $1 \leq j < r - s$;*

3. *Q_j diverges for $r - s < j \leq m := \min(r, p - r)$.*

The χ^2 asymptotics in Theorem 5 is a consequence of the fact that the LR test is performed after fixing the CI rank. Inference for known CI rank is in fact locally asymptotically mixed normal (LAMN), and χ^2 asymptotics applies after conditioning on the stochastic limit observed information matrix, see e.g. Boswijk (1996) and references therein. The number j^2 of degrees of freedom equals the difference in the number of identified parameters under the null and the alternative.

The present χ^2 asymptotics for the LR test does not require to simulate asymptotic critical values. Note that the tests for correct normalization in Saikkonen (1999) and Lukkonen, Ripatti and Saikkonen (1999) all have limit distributions of the unit root-type, instead. This difference is due to the fact that their tests are not formulated as of submodels of model $\mathcal{H}(r)$.

The implications of Theorem 5 for estimators of s based on sequences of Q_j tests are discussed in the following subsection.

III.6 Sequences of tests

In this subsection we discuss three procedures which combine the LR tests Q_j. The first procedure considers the single test Q_1 of $\mathcal{H}_{0,1}$; we call this the ST procedure, short for 'single test'. The second procedure considers the sequence of tests Q_j, from $j = 1$ to $m := \min(r, p - r)$; this procedure is called TD, short for 'testing down' the sequence of nested hypothesis (9). Finally the same tests Q_j may be performed in the reverse order, i.e. from $j = m$ to 1, yielding a third procedure, called TU, short for 'testing up'.

Both the TD and TU procedures can be viewed as applications of a general-to-specific principle, because all tests in the sequence compare the null $\mathcal{H}_{0,j}$ with the unrestricted model $\mathcal{H}(r)$, see e.g. Paruolo (2001) and reference therein. The asymptotic properties of these different procedures are discussed in Theorem 7 below, which gives guidance on how to choose significance levels appropriately.

We next describe the three procedures in more detail. The ST procedure considers hypothesis $\mathcal{H}_{0,1}$ and test Q_1, with critical values k_{1,υ_1} from a $\chi^2(1)$ distribution where υ_1 is the significance level. The ST procedure can be summarized by means of an indicator $\widehat{\iota}_{\mathrm{ST}}$ that signals if $s = r$ (i.e. $\mathcal{K}$ holds) or not. Specifically $\widehat{\iota}_{\mathrm{ST}}$ takes on the value 1 if Q_1 rejects $\mathcal{H}_{0,1}$ and $\widehat{\iota}_{\mathrm{ST}} = 0$ otherwise. By Theorem 5, as $T \to \infty$ one finds $\Pr(\widehat{\iota}_{\mathrm{ST}} = 1) \to 1$ if $s = r$ and $\Pr(\widehat{\iota}_{\mathrm{ST}} = 1) \to \gamma_1 \leq \upsilon_1$ if $s < r$, where the inequality $\gamma_1 \leq \upsilon_1$ is an equality for $s = r - 1$, see Theorem 7 below.

The ST procedure can hence control the size γ_1 of type I errors effectively, even though γ_1 is just a bound on the asymptotic size if $s < r - 1$. One may expect that, when $s < r - 1$, using exact asymptotic critical values for Q_1 in place of k_{1,υ_1} may increase the power of the procedure in finite samples. However, there are many different limit distributions corresponding for Q_1, each one corresponding to a different value of $s < r - 1$; choosing the appropriate distribution entails estimating

s, which is the goal of the TD and TU procedures described below. Instead of doing so, the ST procedure uses conservative critical values to bound type I errors.

Note also that if one finds $\widehat{\imath}_{\mathrm{ST}} = 0$ for a certain choice of the normalization c, the restricted ML estimator delivers an estimate $\beta_1 = c_\perp \varphi_1$ of the CI relations that are the most likely source of failure of the condition $\mathrm{rk}(c'\beta) = r$. This estimate can help the econometrician amend the specification of c, in a way as to avoid the rank reduction. The new choice of c can then be re-submitted to test, again via Q_1.

Example 6 *In order to illustrate how restricted ML estimates may help in selecting a new c, assume that a system with $p = 3$ is analyzed, with CI rank $r = 1$ and with the specification $c = (1, 0, 0)'$. Assume that Q_1 does not reject, so that $\widehat{\imath}_{ST} = 0$. The restricted ML estimation under $\mathcal{H}_{0,1}$ then delivers an estimate of $\beta = \beta_1$ of the form $\beta = (0, \varphi_{11}, \varphi_{12})'$. If the restricted numerical estimates φ_{11}, φ_{12} are sufficiently far from the restriction $\varphi_{11} = -\varphi_{12}$, one may wish to consider $c = (0, 1, 1)'$ as the new specification; this choice of c normalizes $\beta_c = (\kappa_1, \kappa_2, \kappa_3)'$ such that $\kappa_2 + \kappa_3 = 1$. In case the numerical estimates φ_{11}, φ_{12} are close to the restriction $\varphi_{11} = -\varphi_{12}$, then it would be better to use $c = (0, 1, -1)'$ as the new specification instead.*

We now describe the TD and TU procedure in more detail. We describe the TD and TU procedure by defining estimators $\widehat{s}_{\mathrm{TD}}$ and $\widehat{s}_{\mathrm{TU}}$ of s. The TD procedure considers tests Q_j for $j = 1$ to $m := \min(r, p - r)$; the first test in the TD sequence is Q_1, as in the ST procedure. This test is compared with the critical value k_{1,υ_1}; if Q_1 rejects $\mathcal{H}_{0,1}$, then $\widehat{s}_{\mathrm{TD}} = r$, and the TD procedure concludes that $\mathrm{rk}(c'\beta) = r$, just as the ST procedure.

When Q_1 does not reject, the TD procedure considers test Q_2 of hypothesis $\mathcal{H}_{02}$, using critical value k_{2,υ_2} from a $\chi^2(4)$ distribution; again here υ_2 is the significance level. If Q_2 rejects, then $\widehat{s}_{\mathrm{TD}} = r - 1$, otherwise test Q_3 is considered, etc. Test Q_j is hence performed only if all preceding tests Q_1, ..., Q_{j-1} do not reject; test Q_j uses the critical value k_{j,υ_j} from a $\chi^2(j^2)$ distribution. If Q_j is the first test in the sequence to reject, then $\widehat{s}_{\mathrm{TD}} = r - j + 1$. If all the tests Q_1, ..., Q_m do not reject, then $\widehat{s}_{\mathrm{TD}} = r - m$. This completes the description of the $\widehat{s}_{\mathrm{TD}}$ estimator of s, i.e. of the TD procedure.

Note that in the TD procedure the econometrician must set the significance levels υ_j of each test Q_j, $j = 1$, ..., m, in order to achieve a high probability of selecting the correct value of s. A solution for this problem is given below, using the results reported in the following Theorem 7.

We next describe the TU procedure, which considers the Q_j tests in the reverse order, i.e. from $j = m$ to 1. The first test is Q_m, which corresponds to most restricted hypothesis $\mathcal{H}_{0,m}$ on β; the test Q_m uses critical value k_{m,υ_m} from a $\chi^2(m^2)$ distribution. If $\mathcal{H}_{0,m}$ is not rejected, then $\widehat{s}_{\mathrm{TU}} = r - m$, otherwise the procedure proceeds to test $\mathcal{H}_{0,m-1}$. A generic test Q_j in the sequence employs critical values k_{j,υ_j} from a $\chi^2(j^2)$ distribution, and it is performed if all the preceding tests Q_m, ..., Q_{j+1} reject. If Q_j is the first test in the sequence not to yield a rejection, then $\widehat{s}_{\mathrm{TU}} = r - j$. If all Q_m, ..., Q_1 tests reject, then $\widehat{s}_{\mathrm{TU}} = r$. This completes the description of the $\widehat{s}_{\mathrm{TU}}$ estimator of s, i.e. of the TU procedure.

As in the TD procedure, also for the TU procedure one must choose significance levels υ_j of each test Q_j, $j = 1$, ..., m in order to guarantee a given limit probability of selection for the correct value of s. Note that also for the TD and TU procedures when $\widehat{s}_{\mathrm{TD}}$ or $\widehat{s}_{\mathrm{TU}}$ selects a value $< r$, the restricted ML estimate of β may suggests how to modify the choice of c, as described above for the ST procedure.

The following theorem collects asymptotic properties of the three procedures ST, TD, TU; γ_i indicate probabilities.

Theorem 7 *Assume* $s := \mathrm{rk}(c'\beta) < r$*; then as* $T \to \infty$ *one has:*

(i). $\Pr(\widehat{\iota}_{ST} = 1) \to \gamma_1 \leq \upsilon_1$*; in particular when* $s = r - 1$ *one has* $\gamma_1 = \upsilon_1$*;*

(ii). $\Pr(\widehat{s}_{TD} = s) \to \gamma_2 \geq 1 - \sum_{j=1}^{r-s} \upsilon_j$*;*

(iii). $\Pr(\widehat{s}_{TU} = s) \to 1 - \upsilon_{r-s}$, $\Pr(\widehat{s}_{TU} > s) \to \upsilon_{r-s}$ *and* $\Pr(\widehat{s}_{TU} < s) \to 0$.

If instead $s = r$*, then*

(iv). $\Pr(\widehat{\iota}_{ST} = 1) \to 1$, $\Pr(\widehat{s}_{TD} = r) \to 1$, $\Pr(\widehat{s}_{TU} = r) \to 1$.

Note that the asymptotic properties stated in Theorem 7 allow to choose the significance levels υ_j of the tests Q_j in such a way as to control the limits of $\Pr(\widehat{s}_{\mathrm{TD}} = s)$ and $\Pr(\widehat{s}_{\mathrm{TU}} = s)$. Consider, in fact, the TD procedure. Let υ be the desired asymptotic significance level and consider the choice $\upsilon_j = \upsilon/m$, which divides the overall significance level into equal fractions for each test Q_j in the sequence.

This choice ensures that $\Pr(\widehat{s}_{\mathrm{TD}} = s)$ converges to a constant probability γ_2 greater or equal to

$$1 - \sum_{j=1}^{r-s} \upsilon_j = 1 - \frac{r-s}{m}\upsilon \geq 1 - \upsilon,$$

because $(r - s)/m \leq 1$. Hence the choice $\upsilon_j = \upsilon/m$ ensures that the correct value of s is selected with limit probability at least equal to $1 - \upsilon$.

A different choice of υ_j is suggested by Theorem 7 for the TU procedure. In fact one can choose $\upsilon_j = \upsilon$, the nominal level, for all tests Q_j in the sequence, and still obtain that $\Pr(\widehat{s}_{\mathrm{TU}} = s) \to 1 - \upsilon_{r-s} = 1 - \upsilon$. Hence in the TU sequence the econometrician does not need to divide the significance level by the number of tests, in order to control the limit probability of correct selection. In this sense the TU sequence appears preferable to TD. The TU procedure is of the same type considered for instance in CI rank determination, see e.g. Paruolo (2001) and reference therein.

Note that the choice of individual significance levels υ_j does not affect the properties of the TD and TU procedures when $s = r$. In this case all $\widehat{\iota}_{\mathrm{ST}}$, $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ signal with limit probability 1 that $s = r$ is supported by the data.

If several (possibly dependent) specification tests are performed along with one of the procedures $\widehat{\iota}_{\mathrm{ST}}$, $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ in a specification stage, the econometrician may be compelled to bound the overall type-I error by employing some Bonferroni-type inequality across the various tests, where the overall bound on size is computed as the sum of individual test sizes.

Because $\widehat{\iota}_{\mathrm{ST}}$, $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ all select $s = r$ with limit probability 1 under the null of correct specification $\mathcal{K} : s = r$, the inclusion of any of these procedures does not give any contribution to the overall bound. In this sense the inclusion of any of $\widehat{\iota}_{\mathrm{ST}}$, $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ is costless in the specification analysis.[4]

[4]The same asymptotic properties of the $\widehat{\iota}_{\mathrm{ST}}$, $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ procedures can be obtained substituting the LR test with tests with similar asymptotic properties. For instance, a TU procedure based on the Kurozumi (2005) tests would have asymptotic properties similar to the ones stated in Theorem 7.

IV Conclusions

In this paper we have proposed LR tests on the rank of a submatrix of the matrix of CI relations. These tests have χ^2 limit distribution, like the test of Kurozumi (2005) and unlike the tests in Lukkonen, Ripatti and Saikkonen (1999) and Saikkonen (1999). This result is due to the formulation of the test as a test of over-identifying restrictions in a CI model with known CI rank.

The LR tests can be organized into selection procedures that infer the rank of $c'\beta$. The asymptotic properties of these procedures have been discusses. The procedure that starts testing from the most restricted model is proposed as the one with better asymptotic behavior.

In Paruolo (2006) a MC simulation study was performed in order to analyze the small sample properties of the LR tests, and to compare them with the performance of alternative tests. It was found that the proposed LR test had a better finite sample performance than the tests in Lukkonen, Ripatti and Saikkonen (1999) and Saikkonen (1999). The performance of the Wald test of Kurozumi (2005) was similar to the one of the LR test, although for smaller sample sizes the LR test performed marginally better.

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.

Andrews D.W.K. (1996). 'Admissibility of the likelihood ratio test when the parameter space is restricted under the alternative', *Econometrica* Vol. 64, pp. 705-718.

Boswijk, H. P. (1996). 'Testing identifiability of cointegrating vectors', *Journal of Business and Economic Statistics*, Vol. 14, pp. 153-160.

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

Bruneau C. and E. Jondeau (1999). 'Long-run causality. with application to international links between long-term interest rates', *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 545-568.

Gonzalo J. and Granger C. W. J (1995). 'Estimation of Common Long-Memory Components in Cointegrated Systems', *Journal of Business and Economic Statistics*, Vol. 13, pp. 27-36.

Harbo I., S. Johansen, B. Nielsen and A. Rahbek (1998). 'Asymptotic inference on cointegrating rank in partial systems', *Journal of Business and Economic Statistics*, Vol. 16, pp. 388-399.

Johansen S. (1991). 'Estimation and Hypothesis Testing of Cointegration vectors in Gaussian Vector Autoregressive Models', *Econometrica* Vol. 59, pp. 1551-1580.

Johansen S. (1995). 'Identifying restrictions of linear equations with application to simultaneous equations and cointegration.' *Journal of econometrics* Vol. 69, pp. 111-132.

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

Johansen S. (2005). 'Interpretation of cointegrating coefficients in the cointegrated vector autoregressive model', *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 93-104.

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, pp. 211-244.

Johansen S. and K. Juselius (2001). 'Controlling a non-stationary variable in the CI VAR model'. EUI Working Paper ECO No. 2001/2, available at `http://www.math.ku.dk/~sjo`

Kasa K. (1992). 'Common stochastic trends in international stock markets', *Journal of Monetary Economics*, Vol. 29. pp. 95-124.

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.

Omtzigt P. and P. Paruolo (2005). 'Impact factors', *Journal of Econometrics*, Vol. 128, pp. 31-68.

Paruolo P. (1997). 'Asymptotic inference on the moving average impact matrix in cointegratared I(1) VAR systems', *Econometric Theory*, Vol. 13, pp. 79-118.

Paruolo, P. (2001). 'The power of Lambda Max', *Oxford Bulletin of Economics and Statistics*, Vol. 63, pp. 395–403.

Paruolo P. (2006). 'Finite sample comparison of alternative tests on the rank of a cointegration submatrix', Quaderno della Facoltà di Economia 2006/6, University of Insubria, Varese, Italy; available at `http://eco.uninsubria.it/dipeco/Quaderni`

Phillips P. C. B. (1991). 'Optimal inference in cointegrated systems', *Econometrica* Vol. 59, pp. 283-306.

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

Srivastava M.S. and C.G. Khatri (1979). *An introduction to multivariate statistics.* Elsevier North Holland, New York.

Toda H. Y. and P. C. B. Phillips (1993). 'Vector autoregressions and causality', *Econometrica* Vol. 61, pp. 1367-1393.

Yamamoto T. and E. Kurozumi (2006). 'Tests for long run Granger non-causality in cointegrated systems', *Journal of Time Series Analysis*, Vol. 27, pp. 703-723.

Wald A. (1942). 'On the power function of the analysis of variance test', *The Annals of Mathematical Statistics* Vol. 13, pp. 434-439.

A Appendix

This appendix collects proofs. We use the notation $O_p(T^a)$ in the strict sense, i.e. for quantities that are not $o_p(T^a)$.

Proof. of Theorem 1. Let $s := \text{rk}(c'\beta)$. One has $s = r - j$ iff there exists an $r \times j$ full column rank matrix v such that

$$c'\beta v = 0, \tag{16}$$

and there is no full column rank matrix v with more columns that satisfies (16). Let $\beta_1 := \beta v$, $\beta_2 := \bar{\beta} v_\perp$, and observe that $(\beta_1 : \beta_2) := \beta(v : (\beta'\beta)^{-1} v_\perp)$ is an alternative basis of $\mathcal{B} := \text{col}(\beta)$, because $(v : (\beta'\beta)^{-1}v_\perp)$ is a full rank square matrix, see Srivastava and Khatri (1979), p. 19. The columns of β_1 are linearly independent and belong both to $\mathcal{B} := \text{col}(\beta)$ and to $\mathcal{C}^\perp := \text{col}^\perp(c)$ because of (16). They hence form a basis of $\mathcal{B} \cap \mathcal{C}^\perp$, and $\dim(\mathcal{B} \cap \mathcal{C}^\perp)$ equals the number of columns in β_1. This proves that $s = r - j$ implies $\dim(\mathcal{B} \cap \mathcal{C}^\perp) = j$.

The converse statement is proved as follows. If $\dim(\mathcal{B} \cap \mathcal{C}^\perp) = j$ then let β_1 be a basis of $\mathcal{B} \cap \mathcal{C}^\perp$, which can be represented in terms of β as βv, because the columns of β_1 belong to $\mathcal{B}$, with v of full column rank. Because the columns of β_1 belong also to $\mathcal{C}^\perp$, they satisfy $c'\beta_1 = 0$, i.e. (16). This proves (10). The remaining implications on $\mathcal{K}$ and $\mathcal{H}_{0,j}$ follow directly from (10). ■

Proof. of Corollary 2. Immediate. ■

Proof. of Theorem 3. Recall that $s := \text{rk}(c'\beta)$ and define $A := (c : \beta_\perp)$; we wish to show that $\text{rk}\,(A) = p - r + s$. In fact, let $B := \left(\beta : \overline{\beta}_\perp\right)$, which is by construction a full rank square matrix. The product

$$B'A = \begin{pmatrix} \beta' c & 0 \\ \overline{\beta}'_\perp c & I_{p-r} \end{pmatrix}$$

has rank $p-r+s$, because of the lower block-diagonal structure of the right hand side, where we have used the definition $s := \text{rk}(c'\beta)$. This proves $\text{rk}\,(A) = \text{rk}\,(c : \beta_\perp) = p - r + s$.

Next let $K := (\overline{c} : c_\perp)$, a full rank square matrix, and consider the product

$$K'A = \begin{pmatrix} I_r & \overline{c}'\beta_\perp \\ 0 & c'_\perp \beta_\perp \end{pmatrix}$$

which, by the same argument, has the same rank of A, i.e. $\text{rk}\,(K'A) = p - r + s$. By the upper block-diagonal structure, moreover, $\text{rk}\,(K'A) = r + \text{rk}\,(c'_\perp\beta_\perp)$. Hence, equating the two expressions, $\text{rk}\,(c'_\perp\beta_\perp) = p - 2r + s = p - 2r + \text{rk}(c'\beta)$.

Next observe that j is the rank deficiency of $c'\beta$, which equals $\dim\left(\mathcal{B} \cap \mathcal{C}^\perp\right)$ by Theorem 1. From the above equality, substituting $s = r - j$, one finds $\text{rk}\,(c'_\perp\beta_\perp) = p - 2r + r - j = p - r - j$. Hence j is also the rank deficiency of $c'_\perp\beta_\perp$, which equals $\dim\left(\mathcal{B}^\perp \cap \mathcal{C}\right)$. This shows $\dim\left(\mathcal{B} \cap \mathcal{C}^\perp\right) = \dim\left(\mathcal{B}^\perp \cap \mathcal{C}\right)$. ■

Proof. of Corollary 4. Immediate. ■

In order to prove Theorem 5, we first prove the following Lemma, which was used also e.g. in Paruolo (2001).

Lemma 8 *Consider two generic linear subspaces* $\mathcal{G} := \text{col}(\gamma)$, $\mathcal{P} := \text{col}(\pi)$, *with bases* γ *and* π, *both of dimension* $p \times n$, *and let* $\dim(\mathcal{G} \cap \mathcal{P}^\perp) = n - s$, $0 < s < \min(n, p - n)$. *Then one can choose* $\gamma := (\gamma_1 : \gamma_2)$ *as basis of* $\mathcal{G}$, *and* $\pi_\perp := (\gamma_1 : \pi_{\perp 2})$ *as basis of* $\mathcal{P}^\perp$ *with the following properties:*

(i). γ_1 *is a basis of* $\mathcal{G} \cap \mathcal{P}^\perp$ *of dimension* $p \times (n - s)$;

γ_2 *is of dimension* $p \times s$ *and of full column rank;*

$\pi_{\perp 2}$ *is of dimension* $p \times (p - 2n + s)$ *and of full column rank;*

(ii). $\gamma'_1\gamma_2 = 0$, $\gamma'_1\pi_{\perp 2} = 0$;

(iii). $\gamma'_\perp\pi_{\perp 2}$ *has full column rank, i.e.* $\text{rk}(\gamma'_\perp\pi_{\perp 2}) = p - 2n + s$.

Proof. Because $\dim(\mathcal{G} \cap \mathcal{P}^{\perp}) = n - s$, by definition there exist $n - s$ linearly independent vectors that belong both to $\mathcal{G}$ and $\mathcal{P}^{\perp}$; these are collected in γ_1, which has thus full column rank. γ_2 are then chosen as any set of $n-s$ linearly independent vectors in $\mathcal{G}$ that are perpendicular to the columns in γ_1. Because $\gamma_1'\gamma_2 = 0$, $\gamma := (\gamma_1 : \gamma_2)$ has full column rank equal to $\dim(\mathcal{G}) = n$. Similarly one chooses $\pi_{\perp 2}$, such that $\pi = (\gamma_1 : \pi_{\perp 2})$ is a basis of $\mathcal{P}^{\perp}$ and $\gamma_1'\pi_{\perp 2} = 0$. This proves (i) and (ii).

Result (iii) is proved by contradiction. $\gamma_{\perp}'\pi_{\perp 2}$ has dimension $(p-n) \times (p-2n+s)$ and hence $\mathrm{rk}(\gamma_{\perp}'\pi_{\perp 2}) \leq p - 2n + s$. Assume $\mathrm{rk}(\gamma_{\perp}'\pi_{\perp 2}) < p - 2n + s$; this holds iff $\gamma_{\perp}'\pi_{\perp 2}u = 0$ for a nonzero full column rank matrix u. We show that this implies that there exists at least one vector v in $\mathrm{col}(\pi_{\perp 2})$ that belongs to $\mathcal{G}$, which can be added to γ_1 to define a enlarged basis $(\gamma_1 : v)$ of $\mathcal{G} \cap \mathcal{P}^{\perp}$; this implies that $\dim(\mathcal{G} \cap \mathcal{P}^{\perp}) > n - s$, yielding a contradiction.

In fact one has by orthogonal projections

$$\pi_{\perp 2} = \left(P_{\gamma_1} + P_{\gamma_2} + P_{\gamma_{\perp}}\right)\pi_{\perp 2} = \gamma_2\left(\bar{\gamma}_2'\pi_{\perp 2}\right) + \bar{\gamma}_{\perp}\left(\gamma_{\perp}'\pi_{\perp 2}\right).$$

Post multiplying by u one finds $v := \pi_{\perp 2}u = \gamma_2\left(\bar{\gamma}_2'\pi_{\perp 2}\right)u$, where $v \in \mathrm{col}(\gamma_2)$. This proves (iii). ∎

Proof. of Theorem 5. The proof of the limit distribution of Q_j when $j \leq r - s$ is given in Boswijk (1996). We here report the proof that Q_j diverges when $j > r - s$. For simplicity of exposition we consider the case without restricted deterministic components, i.e. $\beta = (\beta_1 : \beta_2) = (H_1\varphi_1 : H_2\varphi_2)$, $H_1 = c_{\perp}$, $H_2 = I$. We indicate by $\Sigma_{ab} := \mathbb{E}(a_t b_t' | Z_{2t})$, a_t, $b_t = Z_{0t}$, $\beta_i' Z_{1t}$. Here the subscript 0 refers to Z_{0t} and the subscript β_i to $\beta_i' Z_{1t}$, where $i = 1, 2, 21$ and β_{21} is defined below. We also let $\Sigma_{aa.b} := \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$, and we use the notation $\Sigma_{aa.bc}$ to indicate $\Sigma_{aa.d}$ with $d_t := (b_t' : c_t')'$.

The plan of the proof is to show that $T^{-1}Q_j$ converges in probability to a positive constant when $j > r - s$. This shows that Q_j is of the order of T in probability, $Q_j = O_p(T)$. Specifically let $T^{-1}Q_j =: c_{T,j} + c_{T,A}$ where

$$c_{T,j,1} := \ln \frac{\left|\widehat{\beta}_1' S_{11.0} \widehat{\beta}_1\right|}{\left|\widehat{\beta}_1' S_{11} \widehat{\beta}_1\right|}, \qquad c_{T,j,2} := \sum_{i=1}^{r-j} \ln\left(1 - \widehat{\zeta}_i\right),$$

$$c_{T,j} := c_{T,j,1} + c_{T,j,2}, \qquad c_{T,A} := -\sum_{i=1}^{r} \ln\left(1 - \widetilde{\lambda}_i\right).$$

The proof is divided into three steps. The first step defines various subspaces and their respective bases. The second one deals with the term $c_{T,A}$. The third one deals with $c_{T,j}$.

Step 1. Subspaces and bases

We use Lemma 8 on $\mathcal{G} = \mathcal{B}$ and $\mathcal{P} = \mathcal{C}$ to define $\beta = (\beta_1 : \beta_2)$, $c_{\perp} = (\beta_1 : c_{\perp 2})$ as bases of $\mathcal{B}$ and $\mathcal{C}^{\perp}$ respectively, where β_1 is a basis of $\mathcal{B} \cap \mathcal{C}^{\perp}$, with $r - s$ columns. Applying again Lemma 8 to $\mathcal{G} = \mathcal{B}^{\perp}$ and $\mathcal{P} = \mathcal{C}^{\perp}$ we define bases $c = (\beta_{\perp 1} : c_2)$, $\beta_{\perp} = (\beta_{\perp 1} : \beta_{\perp 2})$ of $\mathcal{C}$ and $\mathcal{B}^{\perp}$ respectively, where $\beta_{\perp 1}$ is a basis of $\mathcal{B}^{\perp} \cap \mathcal{C}$. Note that by Theorem 3 $\beta_{\perp 1}$ has $r - s$ columns, β_2 and c_2 have s columns, $\beta_{\perp 2}$ and $c_{\perp 2}$ have $p - 2r + s$ columns.

The express $\widehat{\beta}_1 = c_{\perp}\widehat{\varphi}_1$ in terms of the present choice of $c_{\perp}$ as $\widehat{\beta}_1 = (\beta_1 : c_{\perp 2})\widehat{\varphi}_1$, with

$$\widehat{\varphi}_1 := \begin{array}{c} \begin{array}{cc} r-s & j-r+s \end{array} \\ (\widehat{\varphi}_{11} : \widehat{\varphi}_{12}) \end{array} := \begin{array}{cc} \begin{array}{cc} r-s & j-r+s \end{array} & \\ \begin{pmatrix} \widehat{a}_{11} & \widehat{a}_{12} \\ \widehat{a}_{21} & \widehat{a}_{22} \end{pmatrix} & \begin{array}{c} r-s \\ p-2r+s \end{array} \end{array}$$

where we have also reported dimensions of blocks. Recall that $j > r-s$, so that $j-r+s \geq 1$, i.e. $\widehat{\varphi}_{12}$ has at least one column.

Step 2. Term $c_{T,A}$

We first consider $c_{T,A} := -\sum_{i=1}^{r} \ln\left(1-\widetilde{\lambda}_i\right)$. It is well known, see Johansen (1996) eq. (11.17), that under the present assumptions $\widetilde{\lambda}_{1:r}$ converge in probability to $\lambda_{1:r} := (\lambda_1 : \cdots : \lambda_r)'$ eigenvalues of $\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}$, $\lambda_1 \geq \ldots \geq \lambda_r$. Hence $c_{T,A} \xrightarrow{p} c_A := -\sum_{i=1}^{r} \ln\left(1-\lambda_i\right)$.

Using properties of eigenvalues, one can also express c_A as $c_A = -\ln\left(|\Sigma_{\beta\beta.0}| \,/\, |\Sigma_{\beta\beta}|\right)$. Using properties of determinants, one can write $|\Sigma_{\beta\beta.0}| = |\Sigma_{\beta_2\beta_2.\beta_1 0}|\,|\Sigma_{\beta_1\beta_1.0}|$. Similarly $|\Sigma_{\beta\beta}| = |\Sigma_{\beta_2\beta_2.\beta_1}|\,|\Sigma_{\beta_1\beta_1}|$. Thus $c_A = -\ln\left(|\Sigma_{\beta_1\beta_1.0}| \,/\, |\Sigma_{\beta_1\beta_1}|\right) - \ln\left(|\Sigma_{\beta_2\beta_2.\beta_1 0}| \,/\, |\Sigma_{\beta_2\beta_2.\beta_1}|\right)$.

Again using properties of eigenvalues one has $\ln\left(|\Sigma_{\beta_1\beta_1.0}| \,/\, |\Sigma_{\beta_1\beta_1}|\right) = \sum_{i=1}^{r-s} \ln(1-\omega_i)$ where $\omega_1 \geq \cdots \geq \omega_{r-s}$ are the eigenvalues of $\Sigma_{\beta_1\beta_1}^{-1}\Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1}$, i.e. the solutions to

$$\left|\omega\Sigma_{\beta_1\beta_1} - \Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1}\right| = 0. \tag{17}$$

Similarly $\ln\left(|\Sigma_{\beta_2\beta_2.\beta_1 0}| \,/\, |\Sigma_{\beta_2\beta_2.\beta_1}|\right) = \sum_{i=1}^{s} \ln(1-\zeta_i)$, where $\zeta_1 \geq \cdots \geq \zeta_s$ are the eigenvalues of $\Sigma_{\beta_2\beta_2.\beta_1}^{-1}\Sigma_{\beta_2 0.\beta_1}\Sigma_{00.\beta_1}^{-1}\Sigma_{0\beta_2.\beta_1}$, i.e. the solutions to

$$\left|\zeta\Sigma_{\beta_2\beta_2.\beta_1} - \Sigma_{\beta_2 0.\beta_1}\Sigma_{00.\beta_1}^{-1}\Sigma_{0\beta_2.\beta_1}\right| = 0. \tag{18}$$

Summarizing one has the following expression of c_A

$$c_{T,A} \xrightarrow{p} c_A = -\sum_{i=1}^{r-s} \ln(1-\omega_i) - \sum_{i=1}^{s} \ln(1-\zeta_i). \tag{19}$$

Step 3. Term $c_{T,j}$

We next consider $c_{T,j}$. We aim to show that $\mathrm{col}(\widehat{\beta}_1)$, of dimension $j > r-s$, converges to $\mathrm{col}(\beta_1 : c_{\perp 2}a_{22})$, i.e. $\mathrm{col}(\widehat{\beta}_1)$ converges to a subspace spanned by β_1, with $r-s$ columns, and by $c_{\perp 2}a_{22}$, with $j-r+s$ columns, where $c_{\perp 2}$ has a nonsingular projection on $\beta_\perp$ by Lemma 8. We next show that $\mathrm{col}(\widehat{\beta}_2)$ converges to $\mathrm{col}(\beta_2 v)$, where v is $s \times r-j$ and of full column rank, with $r-j < s$. These results are then used to find the probability limit c_j of $c_{T,j}$.

Because $\widehat{\beta} := (\widehat{\beta}_1 : \widehat{\beta}_2)$ is the restricted ML estimator, it must satisfy the first order conditions implicit in (14) regardless of which algorithm is used in the likelihood maximization. Hence we use the first order conditions implicit in (14) in order to derive the asymptotic behavior of $\widehat{\beta}$.

We first consider $\widehat{\beta}_1 = c_\perp\widehat{\varphi}_1$ which is found selecting $\widehat{\varphi}_1$ as the first j eigenvectors associated with the largest j eigenvalues $\widehat{\rho}_{1:j}$ of the problem

$$\left|\widehat{\rho}c_\perp' S_{11.\widehat{\beta}_2}c_\perp - c_\perp' S_{10.\widehat{\beta}_2}S_{00.\widehat{\beta}_2}^{-1}S_{01.\widehat{\beta}_2}c_\perp\right| = 0.$$

Pre and post multiply the equations by $D_{1T} := \text{diag}(I_{r-s}, T^{-1/2}I_{j-r+s})$, one finds, using results and notation as the ones in Lemma 10.3 of Johansen (1996),

$$
\begin{aligned}
D_{1T}c_{\perp}'S_{11}c_{\perp}D_{1T} &= \begin{pmatrix} \beta_1'S_{11}\beta_1 & T^{-1/2}\beta_1'S_{11}c_{\perp 2} \\ T^{-1/2}c_{\perp 2}'S_{11}\beta_1 & T^{-1}c_{\perp 2}'S_{11}c_{\perp 2} \end{pmatrix} \\
&\xrightarrow{w} \begin{pmatrix} \Sigma_{\beta_1\beta_1} & 0 \\ 0 & q_1'\left(\int_0^1 G_u G_u' du\right) q_1 \end{pmatrix}
\end{aligned}
\tag{20}
$$

Moreover $D_{1T}c_{\perp}'S_{11}\widehat{\beta}_2(\widehat{\beta}_2'S_{11}\widehat{\beta}_2)^{-1}\widehat{\beta}_2'S_{11}c_{\perp}D_{1T} = O_p(1)$ for any choice of $\widehat{\beta}_2$ and

$$
D_{1T}S_{10.\widehat{\beta}_2}S_{00.\widehat{\beta}_2}^{-1}S_{01.\widehat{\beta}_2}D_{1T} = \begin{pmatrix} O_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{pmatrix}.
$$

This shows that $\widehat{\rho}_{1:r-s} = O_p(1)$ while $\widehat{\rho}_{r-s+1:j} = o_p(1)$. Hence $\widehat{\varphi}_{11}$, the first block or $r-s$ columns in $\widehat{\varphi}_1$, converges in probability to $(I_{r-s} : 0)'$. Because we can take $\widehat{\varphi}_{11}'\widehat{\varphi}_{12} = 0$, this also shows that $\widehat{a}_{12} \xrightarrow{p} 0$ in $\widehat{\varphi}_{12}$, i.e. that $\widehat{\beta}_1 = (\beta_1 : c_{\perp 2}\widehat{a}_{22}) + o_p(1)$. Before completing the description of the limit of $\widehat{\beta}_1$ by discussing $\widehat{a}_{22}$, we study the limit behavior of $\widehat{\beta}_2$.

$\widehat{\beta}_2$ is given by the eigenvectors $\widehat{u}_{1:r-j} := (\widehat{u}_1 : \cdots : \widehat{u}_{r-j})$ associated with the largest $\widehat{\zeta}_{1:r-j}$ eigenvalues of the problem

$$
\left|\widehat{\zeta}S_{11.\widehat{\beta}_1} - S_{10.\widehat{\beta}_1}S_{00.\widehat{\beta}_1}^{-1}S_{01.\widehat{\beta}_1}\right| = 0.
$$

Pre and post multiplying by $B_T' := \left(\beta_1 : \beta_2 : T^{-1/2}\beta_{\perp}\right)'$ and B_T, one finds that

$$
\begin{aligned}
B_T'S_{11.\widehat{\beta}_1}B_T &\xrightarrow{w} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Sigma_{\beta_2\beta_2.\beta_1} & 0 \\ 0 & 0 & K_1 \end{pmatrix} \\
B_T'S_{10.\widehat{\beta}_1}S_{00.\widehat{\beta}_1}^{-1}S_{01.\widehat{\beta}_1}B_T &\xrightarrow{p} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Sigma_{\beta_2 0.\beta_1}\Sigma_{00.\beta_1}^{-1}\Sigma_{0\beta_2.\beta_1} & 0 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

where K_1 is the weak limit of $T^{-1}\beta_{\perp}'S_{11.\widehat{\beta}_1}\beta_{\perp}$. Hence $\widehat{\zeta}_{1:s}$ converge to $\zeta_{1:s}$, the eigenvalues of $\Sigma_{\beta_2\beta_2.\beta_1}^{-1}\Sigma_{\beta_2 0.\beta_1}\Sigma_{00.\beta_1}^{-1}\Sigma_{0\beta_2.\beta_1}$, where $r-j < s$, i.e. the solution of (18). $\widehat{\beta}_2$, which has $r-j$ columns, converges to $\beta_{21} := \beta_2 v$, where $v = u_{1:r-j}$, the first $r-j$ eigenvectors of (18). $\widehat{\zeta}_{1:r-j}$ hence converge to $\zeta_{1:r-j}$ and

$$
c_{T,j,2} := \sum_{i=1}^{r-j} \ln\left(1 - \widehat{\zeta}_i\right) \xrightarrow{p} c_{j,2} := \sum_{i=1}^{r-j} \ln\left(1 - \zeta_i\right)
$$

We next return to $\widehat{a}_{22}$ in $\widehat{\beta}_1$, in order to show that it converges weakly to a random matrix a_{22}. We consider the smaller set of eigenvalues $\widehat{\rho}_{r-s+1:j} = o_p(1)$ and let $\widehat{\sigma} := T\hat{\rho}$; one finds that the relevant eigenvalue problem is

$$
\left|\widehat{\sigma}(T^{-1}c_{\perp}'S_{11.\beta_{21}}c_{\perp}) - c_{\perp}'S_{10.\beta_{21}}S_{00.\beta_{21}}^{-1}S_{01.\beta_{21}}c_{\perp}\right| = 0.
$$

Now

$$
T^{-1}c_{\perp}'S_{11.\beta_{21}}c_{\perp} \xrightarrow{w} \begin{pmatrix} 0 & 0 \\ 0 & q_1'\left(\int_0^1 G_u G_u' du\right) q_1 \end{pmatrix}
$$

so that, by an argument similar to Johansen (1996), p.159, $\widehat{\sigma}_{1:p-2r+s}$ converge to the eigenvalues of

$$\left|\sigma q_1' \left(\int_0^1 G_u G_u' du\right) q_1 - K_2\right| = 0, \tag{21}$$

where K_2 is the weak limit of $c_{\perp 2}' S_{10.\beta_{21}} N S_{01.\beta_{21}} c_{\perp 2}$ and

$$N := \Sigma_{00.\beta_{21}}^{-1} - \Sigma_{00.\beta_{21}}^{-1} \Sigma_{0\beta_1.\beta_{21}} (\Sigma_{\beta_1 0.\beta_{21}} \Sigma_{00.\beta_{21}}^{-1} \Sigma_{0\beta_1.\beta_{21}})^{-1} \Sigma_{\beta_1 0.\beta_{21}} \Sigma_{00.\beta_{21}}^{-1}.$$

Let $\sigma_1 \geq \cdots \geq \sigma_{p-2r+s}$ be the ordered eigenvalues of (21) with associated eigenvectors w_1, ..., w_{p-2r+s}, and let $w_{1:n} := (w_1 : \cdots : w_n)$. Then $\widehat{a}_{22} \xrightarrow{w} a_{22} := w_{1:r-j}$. This shows that $\text{col}(\widehat{\beta}_1)$ converges to $\text{col}(\beta_1 : c_{\perp 2} a_{22})$ and $\text{col}(\widehat{\beta}_2)$ converges to $\text{col}(\beta_2 v)$.

We finally consider

$$\begin{aligned} c_{T,j,1} &:= \ln \frac{\left|\widehat{\beta}_1' S_{11.0} \widehat{\beta}_1\right|}{\left|\widehat{\beta}_1' S_{11} \widehat{\beta}_1\right|} = \ln \left| I_j - \left(\widehat{\beta}_1' S_{11} \widehat{\beta}_1\right)^{-1} \widehat{\beta}_1' S_{10} S_{00}^{-1} S_{01} \widehat{\beta}_1 \right| \\ &= \ln \left| I_j - \left(D' \widehat{\beta}_1' S_{11} \widehat{\beta}_1 D\right)^{-1} D' \widehat{\beta}_1' S_{10} S_{00}^{-1} S_{01} \widehat{\beta}_1 D \right| \end{aligned}$$

where D is square and nonsingular. We choose D equal to

$$D_{2T} := \begin{pmatrix} \widehat{a}_{11}^{-1} & -\widehat{a}_{11}^{-1} \widehat{a}_{12} \\ 0 & I_{j-r+s} \end{pmatrix} D_{1T}$$

where the leading block on the main diagonal $\widehat{a}_{11} = I_{r-s} + o_p(1)$. This gives

$$\widehat{\beta}_1 D_{2T} = (\beta_1 : c_{\perp 2}) \begin{pmatrix} I & 0 \\ \widehat{a}_{21} \widehat{a}_{11}^{-1} & T^{-1/2} I_{j-r+s} \end{pmatrix},$$

and

$$\begin{aligned} D_{2T}' \widehat{\beta}_1' S_{11} \widehat{\beta}_1 D_{2T} &\xrightarrow{w} \begin{pmatrix} \Sigma_{\beta_1 \beta_1} & 0 \\ 0 & q_1' \left(\int_0^1 G_u G_u' du\right) q_1 \end{pmatrix}, \\ D_{2T}' \widehat{\beta}_1' S_{10} S_{00}^{-1} S_{01} \widehat{\beta}_1 D_{2T} &\xrightarrow{w} \begin{pmatrix} \Sigma_{\beta_1 0} \Sigma_{00}^{-1} \Sigma_{0\beta_1} & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Hence $c_{T,j,1} \xrightarrow{p} \sum_{i=1}^{r-s} \ln(1-\omega_i)$, where $\omega_1 \geq \cdots \geq \omega_{r-s}$ are the eigenvalues $\Sigma_{\beta_1\beta_1}^{-1} \Sigma_{\beta_1 0} \Sigma_{00}^{-1} \Sigma_{0\beta_1}$, i.e. the solutions of (17). Summarizing

$$\begin{aligned} c_{T,j} &\xrightarrow{p} c_j = \sum_{i=1}^{r-s} \ln(1-\omega_i) + \sum_{i=1}^{r-j} \ln\left(1-\zeta_i\right), \\ c_{T,A} &\xrightarrow{p} c_A = -\sum_{i=1}^{r-s} \ln(1-\omega_i) - \sum_{i=1}^{s} \ln(1-\zeta_i), \\ T^{-1} Q_j &\xrightarrow{p} \delta := c_j + c_A = -\sum_{i=r-j+1}^{s} \ln(1-\zeta_i) > 0, \end{aligned}$$

which proves that Q_j diverges. The term δ can be interpreted as a non-centrality parameter. ∎

Proof. of Theorem 7. Let $\mathcal{A}_j := \{Q_j \leq k_{j,\upsilon_j}\}$, $\mathcal{R}_j := \{Q_j > k_{j,\upsilon_j}\}$ be the acceptance and rejection regions of test Q_j. Let also $0 < \gamma_{ij} < 1$ indicate generic probabilities. The estimators $\widehat{s}_{\mathrm{TD}}$, $\widehat{s}_{\mathrm{TU}}$ have the following representations in terms of $\mathcal{A}_j$, $\mathcal{R}_j$ with $0 < h < r$:

$$\{\widehat{s}_{\mathrm{TD}} = r\} = \mathcal{R}_1, \quad \{\widehat{s}_{\mathrm{TD}} = r-h\} = \bigcap_{j=1}^{h} \mathcal{A}_j \cap \mathcal{R}_{h+1}, \quad \{\widehat{s}_{\mathrm{TD}} = r-m\} = \bigcap_{j=1}^{m} \mathcal{A}_j;$$
$$\{\widehat{s}_{\mathrm{TU}} = r-m\} = \mathcal{A}_1, \quad \{\widehat{s}_{\mathrm{TU}} = r-h\} = \bigcap_{j=h+1}^{m} \mathcal{R}_j \cap \mathcal{A}_h, \quad \{\widehat{s}_{\mathrm{TU}} = r\} = \bigcap_{j=1}^{m} \mathcal{R}_j.$$

If $s < r$ then $\Pr(\widehat{\iota}_{\mathrm{ST}} = 1) = \Pr(\mathcal{R}_1) \to \gamma_1 \leq \upsilon_1$ by Theorem 5, and $\gamma_1 = \upsilon_1$ if $s = r-1$. This proves (i). Moreover, applying Bonferroni's inequality

$$\Pr(\widehat{s}_{\mathrm{TD}} = s) = \Pr\left(\bigcap_{j=1}^{r-s} \mathcal{A}_j \cap \mathcal{R}_{r-s+1}\right) \geq \sum_{j=1}^{r-s} \Pr(\mathcal{A}_j) + \Pr(\mathcal{R}_{r-s+1}) - (r-s).$$

By Theorem 5, $\Pr(\mathcal{A}_j) \to \gamma_{3j} \geq 1 - \upsilon_j$, for $j \leq r-s$ and $\Pr(\mathcal{R}_{r-s+1}) \to 1$. Hence $\Pr(\widehat{s}_{\mathrm{TD}} = s) \to \gamma_2 \geq 1 - \sum_{j=1}^{r-s} \upsilon_j$. This proves (ii). Similarly consider

$$\Pr(\widehat{s}_{\mathrm{TU}} = s) = \Pr\left(\bigcap_{j=r-s+1}^{m} \mathcal{R}_j \cap \mathcal{A}_{r-s}\right).$$

By Theorem 5, $\Pr(\mathcal{R}_j) \to 1$ for $j > r-s$, and hence $\Pr(\widehat{s}_{\mathrm{TU}} = s) \to \lim \Pr(\mathcal{A}_{r-s}) = 1 - \upsilon_{r-s}$. Moreover for $a < s$ (i.e. $r-a > r-s$) one has

$$\Pr(\widehat{s}_{\mathrm{TU}} = a) = \Pr\left(\bigcap_{j=r-a+1}^{m} \mathcal{R}_j \cap \mathcal{A}_{r-a}\right).$$

As above, $\Pr(\mathcal{R}_j) \to 1$ for $j \geq r-a > r-s$ and hence $\Pr(\mathcal{A}_{r-a}) \to 0$. This proves that $\Pr(\widehat{s}_{\mathrm{TU}} = a) \to 0$ for $a < s$. Summing these probabilities over a for $m \leq a < s$ one finds $\Pr(\widehat{s}_{\mathrm{TU}} < s) \to 0$. This completes the proof for (iii).

Assume now that $s = r$. Then by Theorem 5, $\Pr(\mathcal{R}_j) \to 1$ for $1 \leq j \leq m$. Result (iv), follows directly. ∎