ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Journal of Econometrics ▌(▌▌▌▌) ▌▌▌–▌▌▌

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Generalized reduced rank tests using the singular value decomposition

Frank Kleibergen[a,b,*], Richard Paap[c]

[a]Department of Economics, Box B, Brown University, Providence RI 02912, USA
[b]Department of Quantitative Economics, University of Amsterdam, The Netherlands
[c]Econometric Institute (H11-25), Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands

Accepted 9 February 2005

Abstract

We propose a novel statistic to test the rank of a matrix. The rank statistic overcomes deficiencies of existing rank statistics, like: a Kronecker covariance matrix for the canonical correlation rank statistic of Anderson [Annals of Mathematical Statistics (1951), 22, 327–351] sensitivity to the ordering of the variables for the LDU rank statistic of Cragg and Donald [Journal of the American Statistical Association (1996), 91, 1301–1309] and Gill and Lewbel [Journal of the American Statistical Association (1992), 87, 766–776] a limiting distribution that is not a standard chi-squared distribution for the rank statistic of Robin and Smith [Econometric Theory (2000), 16, 151–175] usage of numerical optimization for the objective function statistic of Cragg and Donald [Journal of Econometrics (1997), 76, 223–250] and ignoring the non-negativity restriction on the singular values in Ratsimalahelo [2002, Rank test based on matrix perturbation theory. Unpublished working paper, U.F.R. Science Economique, University de Franche-Comté]. In the non-stationary cointegration case, the limiting distribution of the new rank statistic is identical to that of the Johansen trace statistic.

JEL classification: C12; C13; C30

Keywords: Stochastic discount factor model; Cointegration; GMM

*Corresponding author. Department of Economics, Box B, Brown University, Providence RI 02912, USA.

E-mail addresses: Frank_Kleibergen@brown.edu (F. Kleibergen), paap@few.eur.nl (R. Paap).

URL: http://www.econ.brown.edu/fac/Frank_Kleibergen, http://www.few.eur.nl/few/people/paap.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.02.011

1. Introduction

Tests of many economic hypotheses boil down to testing the rank of a matrix, see, for example, Lewbel (1991). Also the identification of parameters in econometric models often depends on the rank of a matrix. Therefore, since the early stages of the development of econometric methodology a literature focusing on rank tests has developed. An early contribution to this line of work is Anderson (1951). Anderson (1951) proposes a likelihood ratio rank statistic in a multivariate regression model based on canonical correlations. This canonical correlation rank statistic has a χ^2 limiting distribution when the covariance matrix of the unrestricted matrix estimator has a Kronecker structure. A Kronecker covariance matrix is however not always valid. For example, it does not apply if one opts for the use of heteroscedasticity autocorrelation consistent covariance (HACC) matrix estimators, see, for example, Andrews (1991), Newey and West (1987), West (1997) and White (1980). Alternative approaches to test the rank of a matrix that overcome this deficit have therefore been proposed.

Gill and Lewbel (1992) and Cragg and Donald (1996) use a LDU decomposition (see, for example, Golub and van Loan, 1989) of the unrestricted matrix to construct a rank statistic. They use a (root-T) consistent estimator of the unrestricted matrix, which does not need to have a Kronecker covariance matrix. A disadvantage of their approach is that it can be sensitive to the ordering of the variables. To overcome this sensitivity, the estimator of the unrestricted matrix is permuted using Gaussian elimination. A high level assumption is however needed to ensure that the Gaussian elimination performs adequately.

Cragg and Donald (1997) propose a rank statistic based on a minimum χ^2 criterion. They assume that the rank of the covariance matrix of the unrestricted matrix is maximal and minimize the objective function numerically. Numerical optimization is often difficult when the dimension of the optimization space is large. Hence, numerical optimization of the criterion function does not always perform satisfactorily when the number of columns or rows of the matrix of interest is large.

Recently, Robin and Smith (2000) construct a rank statistic on the basis of a (root-T) consistent estimator for the unrestricted matrix. The covariance matrix of this estimator does not need to have the Kronecker structure and its rank does not have to be maximal. The limiting distribution of the rank statistic is, however, a weighted average of χ^2 distributions. The asymptotic critical values of the rank statistic are therefore not tabulated. The limiting distribution shows that this rank statistic is not always appropriately normalized so the range of values of the weights affects the power of the statistic.

To overcome the deficiencies of the rank statistics, we propose a novel rank statistic. The rank statistic involves a (root-T) consistent estimator of the unrestricted matrix, which does not need to have a Kronecker covariance matrix. We decompose the unrestricted matrix estimator using a singular value decomposition. If the limiting distribution of the unrestricted matrix estimator is normal, an orthogonal transformation of the smallest singular values has a normal limiting distribution. Our rank statistic is equal to the quadratic form of this orthogonal transformation with respect to the inverse of its covariance matrix, and hence it has a standard χ^2 limiting distribution. The rank statistic simplifies to the canonical correlation rank statistic of Anderson (1951) in case of a Kronecker covariance matrix. The rank statistic can also be applied in case of non-stationary variables, for example, for cointegration testing. The asymptotic distribution is in that case a functional of Brownian motions and equal to the asymptotic distribution of the Johansen (1991) trace statistic.

The novel rank statistic is related to the rank statistic proposed by Ratsimalahelo (2001, 2002). The limiting distribution of his rank statistic is however based on asymptotic normality of the smallest singular values. Since singular values are non-negative, they cannot have a mean zero normal limiting distribution. We therefore consider an orthogonal transformation of the smallest singular values which removes the non-negativity restriction on the singular values.

The outline of the paper is as follows. In Section 2 we discuss the singular value decomposition of a matrix. We show how this decomposition can be used to obtain a parameter that reflects rank reduction. In Section 3 we construct a rank statistic based on this parameter. We derive its limiting distribution and show that the statistic is a generalization of the canonical correlation rank statistic. We limit ourselves in this section to stationary variables and compare the rank statistic with existing rank statistics. The non-stationary cointegration case is discussed in Section 4. In this case, the limit distribution of our rank statistic is identical to that of the Johansen trace statistic. In Section 5, we apply the rank statistic to test the rank of a matrix that governs the identification of the parameters in the stochastic discount factor model of Jagannathan and Wang (1996). Furthermore, we use a factor model which stems from this application to illustrate the validity of the limiting distribution and to conduct a power comparison of different rank statistics. We end with some concluding remarks and topics for further research in Section 6.

Throughout the paper we use the notation: $a = \mathrm{vec}(A)$ for the column vectorization of the $k \times m$ matrix A such that for $A = (A_1 \cdots A_m)$, $\mathrm{vec}(A) = (A_1' \cdots A_m')'$, I_m is the $m \times m$ identity matrix and $\mathrm{tr}(A)$ stands for the trace of a square matrix A. Furthermore, "$\underset{p}{\rightarrow}$" stands for convergence in probability and "$\underset{d}{\rightarrow}$" for convergence in distribution.

2. Singular value decomposition and rank reduction

To construct our rank test, we use the fact that a $k \times m$ real matrix C can be decomposed as

$$C = A_q B_q + A_{q,\perp} \Lambda_q B_{q,\perp}, \tag{1}$$

with A_q a $k \times q$ matrix, B_q a $q \times m$ matrix, Λ_q a $(k-q) \times (m-q)$ matrix, $A_{q,\perp}$ a $k \times (k-q)$ matrix, $B_{q,\perp}$ a $(m-q) \times m$ matrix and where $A_q' A_{q,\perp} \equiv 0$, $B_{q,\perp} B_q' \equiv 0$, $A_{q,\perp}' A_{q,\perp} \equiv I_{k-q}$ and $B_{q,\perp} B_{q,\perp}' \equiv I_{m-q}$ with $q < \min(k, m)$, see Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002). If $\Lambda_q = 0$, the rank of the matrix C is determined by the rank of $A_q B_q$. If both A_q and B_q have full rank, the rank of C is then equal to q. Our test for rank reduction will be based on a test for $\Lambda_q = 0$. To obtain a value of Λ_q which reflects the distance to rank reduction, we define Λ_q as a transformation of the smallest singular values of the matrix C since these values determine the rank of the matrix C in a unambiguous way. In fact, to identify the matrices in decomposition (1) we link the decomposition to a singular value decomposition of C.

The singular value decomposition (SVD) of a real $k \times m$ matrix C is given by

$$C = USV', \tag{2}$$

where U is a $k \times k$ orthonormal matrix ($U'U = I_k$), V is a $m \times m$ orthonormal matrix ($V'V = I_m$), and S is a $k \times m$ matrix that contains the singular values of C on its main diagonal and is equal to zero elsewhere. If $k = m$, S is a $k \times k$ diagonal matrix with the singular values in decreasing order on its main diagonal. If $k<m$, S consists of a $k \times k$ diagonal matrix with the k singular values on its main diagonal (in decreasing order) extended on the right-hand side with a $k \times (m-k)$ matrix of zeros. If $k>m$, S consists of a $m \times m$ diagonal matrix with the m singular values on its main diagonal (in decreasing order) on top of a $(k-m) \times m$ matrix of zeros, see, for example, Golub and van Loan (1989) and Horn and Johnson (1991) for details. Restricting the smallest singular values in Λ_q to zero leads to the least squares approximation in dimension q of the matrix C, see, for example, Reinsel and Velu (1998).

The expressions for A_q, B_q, and Λ_q in terms of the smallest singular values of C follow from the relation

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V'_{11} & V'_{21} \\ V'_{12} & V'_{22} \end{pmatrix} = A_q B_q + A_{q,\perp} \Lambda_q B_{q,\perp}, \tag{3}$$

where U_{11}, S_1 and V_{11} are $q \times q$ matrices, U_{12} and U'_{21} are $q \times (k-q)$ matrices, V_{12} and V'_{21} are $q \times (m-q)$ matrices, U_{22} is a $(k-q) \times (k-q)$ matrix, V_{22} a $(m-q) \times (m-q)$ matrix, and S_2 a $(k-q) \times (m-q)$ matrix. Eq. (3) implies that

$$A_q B_q = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 \begin{pmatrix} V'_{11} & V'_{21} \end{pmatrix} \text{ and } A_{q,\perp} \Lambda_q B_{q,\perp} = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} S_2 \begin{pmatrix} V'_{12} & V'_{22} \end{pmatrix}. \tag{4}$$

The exact relation between A_q, B_q, Λ_q, $A_{q,\perp}$, $B_{q,\perp}$ and U, S, V depends on the specification of A_q and B_q. If we do not normalize $A_q B_q$, the number of free elements in A_q and B_q $(= kq + qm)$ is larger than the number of free elements of a $k \times m$ matrix with rank q $(= km - q^2)$. Hence, if we want to solve for A_q and B_q we need to impose a normalization for $A_q B_q$. Several normalizations are possible and the normalization usually depends on the model at hand. As an illustration we show the normalization involved in three econometric models that imply rank reduction, that is (a) the linear instrumental variables regression model, (b) the error correction cointegration model and (c) a factor model.

(a) The linear instrumental variables regression model for the q-dimensional endogenous variable $y_{1,i}$ and $(k-q)$-dimensional endogenous variable $y_{2,i}$ with the m-dimensional vector of instruments z_i

$$y_{1,i} = B_q z_i + \varepsilon_{1,i}$$
$$y_{2,i} = A_{q,2} y_{1,i} + \varepsilon_{2,i}, \text{ for } i = 1, \ldots, T \tag{5}$$

can be written as

$$\begin{pmatrix} y_{1,i} \\ y_{2,i} \end{pmatrix} = C z_i + \begin{pmatrix} u_{1,i} \\ u_{2,i} \end{pmatrix}, \tag{6}$$

where $u_{1,i} = \varepsilon_{1,i}$, $u_{2,i} = \varepsilon_{2,i} + A_{q,2}\varepsilon_{1,i}$, and $C = A_q B_q$ with normalization $A_q = [I_q \vdots A'_{q,2}]'$.

(b) In a vector autoregression of order 1 for the k-dimensional time series $\{y_i\}_{i=0}^T$ in error correction form

$$\Delta y_i = C y_{i-1} + \varepsilon_i, \tag{7}$$

rank reduction of the $k \times k$ matrix C implies cointegration, see, for example, Engle and Granger (1987) and Johansen (1991). If the cointegration rank is q, we can write (7) as

$$\Delta y_i = A_q B_q y_{i-1} + \varepsilon_i. \tag{8}$$

For economic interpretation of the cointegration relation $B_q y_i$, it is convenient to normalize B_q as $[I_q \vdots B_{q,2}]$, where $B_{q,2}$ is a $q \times (k-q)$ matrix.

(c) In a factor model, where one relates a k-dimensional vector of variables r_i to an m-dimensional vector of explanatory variables f_i,

$$r_i = C f_i + \varepsilon_i, \text{ for } i = 1, \ldots, T, \tag{9}$$

one typically imposes rank reduction on C to reduce the number of parameters. If the rank of C is q, we can write $C = A_q B_q$. Since there is often no clear interpretation of $B_q f_i$, one may impose that $B_q\left(\frac{1}{T}\sum_{i=1}^T f_i f'_i\right)B'_q = I_q$ as a normalization.

Given an appropriate normalization, we can solve for A_q and B_q in (4). For example, if we impose that $B_q = [I_q \vdots B_{q,2}]$ with $B_{q,2}$ a $q \times (m-q)$ matrix, as in the cointegration model, the first equation in (4) implies that

$$A_q = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V'_{11} \text{ and } B_{q,2} = (V'_{11})^{-1} V_{21}. \tag{10}$$

Note that the A_q and B_q matrix are constructed with respect to the q largest singular values of C contained in S_1. Although Λ_q is initially only identified up to an orthonormal transformation, since $A^*_{q,\perp}\Lambda^*_q B^*_{q,\perp} = A_{q,\perp}\Lambda_q B_{q,\perp}$ with $A^*_{q,\perp} = A_{q,\perp}R_q$, $\Lambda^*_q = R'_q \Lambda_q Q'_q$, $B^*_{q,\perp} = Q_q B_{q,\perp}$ and R_q and Q_q are $(k-q)\times(k-q)$ and $(m-q)\times(m-q)$ orthonormal matrices, a unique specification of Λ_q results when we use an appropriate normalization of A_q and B_q. This normalization allows us to express $A_{q,\perp}$ and $B_{q,\perp}$ as functions of the unrestricted elements of A_q and B_q and hence in terms of U_{11}, U_{21}, S_1, V_{11}, and V_{21}. We show below that this leads to the expression of Λ_q[1]

$$\Lambda_q = (U_{22}U_{22}')^{-1/2}U_{22}S_2V_{22}'(V_{22}V_{22}')^{-1/2}, \tag{11}$$

which implies, given the second equation in (4), that

$$A_{q,\perp} = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1}(U_{22}U_{22}')^{1/2} \text{ and } B_{q,\perp} = (V_{22}V_{22}')^{1/2}(V_{22}')^{-1}[V_{12}' \vdots V_{22}']. \tag{12}$$

For (11) and (12), we can express $A_{q,\perp}$ in terms of $A_q = [A_{q,1}' \vdots A_{q,2}']'$ with $A_{q,1}$ a $q \times q$ matrix and $A_{q,2}$ a $(k-q) \times q$ matrix as

$$\begin{aligned} A_{q,\perp} &= \begin{pmatrix} U_{12}U_{22}^{-1} \\ I_{k-q} \end{pmatrix} ((U_{22}')^{-1}U_{22}^{-1})^{-1/2} \\ &= \begin{pmatrix} -(U_{11}')^{-1}U_{21}' \\ I_{k-q} \end{pmatrix} (I_{k-q} + U_{21}U_{11}^{-1}(U_{11}')^{-1}U_{21}')^{-1/2} \\ &= \begin{pmatrix} -(A_{q,1}')^{-1}A_{q,2} \\ I_{k-q} \end{pmatrix} (I_{k-q} + A_{q,2}A_{q,1}^{-1}(A_{q,1}')^{-1}A_{q,2}')^{-1/2}, \end{aligned} \tag{13}$$

where we use the orthonormality properties of U and V. Likewise, we can write $B_{q,\perp}$ in terms of $B_{q,2}$

$$\begin{aligned} B_{q,\perp} &= ((V_{22}')^{-1}V_{22}^{-1})^{-1/2}[(V_{22}')^{-1}V_{12}' \vdots I_{m-q}] \\ &= (I_{m-q} + V_{21}'V_{11}^{-1}(V_{11}')^{-1}V_{21})^{-1/2}[-V_{21}'V_{11}^{-1} \vdots I_{m-q}] \\ &= (I_{m-q} + B_{q,2}'B_{q,2})^{-1/2}[-B_{q,2}' \vdots I_{m-q}]. \end{aligned} \tag{14}$$

We note that every normalization of A_q and B_q leads to the same expressions of Λ_q, $A_{q,\perp}$ and $B_{q,\perp}$ in (11) and (12). The matrices $(U_{22}U_{22}')^{-1/2}U_{22}$ and $V_{22}'(V_{22}V_{22}')^{-1/2}$ in the expression of Λ_q (11) are both orthonormal matrices so Λ_q remains well-defined when U_{22} or V_{22} are of reduced rank. Hence, the matrix Λ_q is equal to the matrix S_2, that contains the smallest singular values, pre- and postmultiplied by orthonormal matrices or stated differently, vec(Λ_q) is just a rotation of the vectorized smallest singular values around the origin. The singular values are non-negative and therefore all elements of S_2 are also non-negative. If we pre- and post-multiply S_2 by orthonormal matrices, the elements of the resulting matrix Λ_q are no longer restrained to be non-negative. Hence, when C is a matrix of random variables, the elements of Λ_q can be asymptotically normal distributed while the elements of S_2

[1]If D is a positive definite real symmetric matrix, then $D^{1/2} = ELE'$, where L is a diagonal matrix containing the square roots of the eigenvalues of D, E contains the orthonormal eigenvectors of D, and $D^{-1/2} = EL^{-1}E'$, see, for example, Johansen (1995, p. 222).

cannot be asymptotically normal distributed. This property allows us to base tests for rank reduction with χ^2 limiting distributions on Λ_q as we show in the next section.

3. Rank test: the standard (stationary) case

We are concerned with testing the rank of the $k \times m$ matrix of parameters Π. We assume that the matrix is non-symmetric in case $k = m$ and hence Π is not allowed to be a covariance matrix. Under our null hypothesis, the rank of the matrix Π is equal to q with $q < \min(k, m)$, that is, $\mathrm{H}_0 : \mathrm{rank}(\Pi) = q$. To derive the limiting distributions of our test statistics and estimators, we make an assumption about the limiting behavior of an estimator of the unrestricted value of Π denoted by $\hat{\Pi}$, where we use the notation $\hat{\pi} = \mathrm{vec}(\hat{\Pi})$ and $\pi = \mathrm{vec}(\Pi)$.

Assumption 1. The limiting behavior of the estimator of the matrix Π is characterized by

$$\sqrt{T}(\hat{\pi} - \pi) \underset{\mathrm{d}}{\rightarrow} \mathrm{N}(0, \mathscr{V}), \tag{15}$$

where T is the sample size and $\mathscr{V}$ is a $km \times km$ covariance matrix.

If we apply the SVD given in (2) directly on $\hat{\Pi}$ to test for the rank of Π, the resulting procedure can be sensitive to scaling of $\hat{\Pi}$, that is, pre- or post-multiplying $\hat{\Pi}$ by some matrix can affect the outcome of the rank test. We therefore allow $\hat{\Pi}$ to be normalized before we conduct a SVD of it. We can normalize $\hat{\Pi}$ by means of pre-multiplication by a $k \times k$ finite non-singular matrix G and post-multiplication by a $m \times m$ finite non-singular matrix F. We are free to choose the specification of F and G. The matrices F and G adapt the scaling of $\hat{\Pi}$

$$\hat{\Theta} = G\hat{\Pi}F', \tag{16}$$

such that $\hat{\theta} = \mathrm{vec}(\hat{\Theta}) = (F \otimes G)\hat{\pi}$. Using an appropriate specification of F and G, the estimator $\hat{\Theta}$ becomes invariant to invertible transformations of the data that are identical over all observations. For example, $\hat{\Pi}$ is often a least squares estimator, $\hat{\Pi} = (\sum_{i=1}^{T} y_i x_i')^{-1}(\sum_{i=1}^{T} x_i x_i')^{-1}$, and appropriate specifications of F and G then result from $FF' = T^{-1}(T^{-1}\sum_{i=1}^{T} x_i x_i')\mathrm{covar}(x_i)^{-1}\ (T^{-1}\sum_{i=1}^{T} x_i x_i')$ and $G'G = \mathrm{covar}(y_i)^{-1}$. It is therefore best to specify the matrices F and G such that $(F \otimes G)\mathscr{V}(F \otimes G)'$ is close to the identity matrix which improves the power properties of the proposed test.

The limiting behavior of $\hat{\theta}$ is characterized by

$$\sqrt{T}(\hat{\theta} - \theta) \underset{\mathrm{d}}{\rightarrow} \mathrm{N}(0, \mathscr{W}), \tag{17}$$

where $\mathscr{W} = (F \otimes G)\mathscr{V}(F \otimes G)'$ and $\theta = (F \otimes G)\pi$.

To test $\mathrm{H}_0 : \mathrm{rank}(\Pi) = q$, which is equivalent to $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$, we use decomposition (1) for Θ

$$\Theta = A_q B_q + A_{q,\perp}\Lambda_q B_{q,\perp}, \tag{18}$$

where the components follow from the SVD (2) as described in Section 2. Under H_0, the $(k-q)\times(m-q)$ matrix Λ_q is identical to zero. The null hypothesis H_0 : $\text{rank}(\Theta)=q$ is identical to $H_0:\Lambda_q=0$. To obtain Λ_q from (18), we need to pre- and post-multiply Θ by $A'_{q,\perp}$ and $B'_{q,\perp}$, respectively. To test $H_0:\Lambda_q=0$, we therefore have to make an assumption concerning the covariance matrix $\mathscr{W}$.

Assumption 2. The $(k-q)(m-q)\times(k-q)(m-q)$ covariance matrix

$$\Omega_q=(B_{q,\perp}\otimes A'_{q,\perp})\mathscr{W}(B_{q,\perp}\otimes A'_{q,\perp})' \tag{19}$$

is non-singular.

Decomposition (18) is also applied to the estimator $\hat{\Theta}$,

$$\hat{\Theta}=\hat{A}_q\hat{B}_q+\hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}. \tag{20}$$

The limiting behavior of the different elements of $\hat{\Theta}$ in (20) is stated in Theorem 1.

Theorem 1. *Under* H_0 *and Assumptions* 1–2, *the limiting behavior of the elements of* $\hat{\Theta}$ *in* (20) *is such that* $\hat{A}_q\hat{B}_q$ *resulting from the SVD of* $\hat{\Theta}$ *is a root-T consistent estimator of* A_qB_q *and*

$$\sqrt{T}\hat{\lambda}_q\underset{d}{\to}\text{N}(0,\Omega_q), \tag{21}$$

where $\hat{\lambda}_q=\text{vec}(\hat{\Lambda}_q)$ *and* $\hat{\Lambda}_q=\hat{A}'_{q,\perp}\hat{\Theta}\hat{B}'_{q,\perp}$.

Proof. See the appendix. □

We use $\hat{\lambda}_q$ to define the statistic to test $H_0:\text{rank}(\Theta)=q$.

Corollary 1. *Under Assumptions* 1–2, *the statistic*

$$\text{rk}(q)=T\hat{\lambda}'_q\Omega_q^{-1}\hat{\lambda}_q, \tag{22}$$

converges under $H_0:\text{rank}(\Theta)=q$ *in distribution to a* $\chi^2((k-q)(m-q))$ *random variable*.

The statistic (22), to which we refer as the rk-statistic, becomes an operational statistic for testing the rank of Θ when we have specified a consistent estimator for the covariance matrix Ω_q, which corresponds with a consistent estimator for the covariance matrix $\mathscr{V}$. The rk-statistic therefore allows for the use of both parametric as well as non-parametric covariance matrix estimators, like, for example, HACC matrix estimators, see Andrews (1991), Newey and West (1987), West (1997) and White (1980).

The rk-statistic is related to the rank tests discussed in the literature. To illustrate this, we separately discuss the relationship between the rk-statistic and the rank statistics proposed by Anderson (1951), Robin and Smith (2000), Cragg and Donald (1997), Gill and Lewbel (1992)–Cragg and Donald (1996) and Ratsimalahelo (2001, 2002).

Anderson (1951) *canonical correlation rank statistic*. When the covariance matrix $\mathscr{V}$ has a Kronecker form and $\hat{\Pi}$ is a least squares estimator, the rk-statistic is identical to the canonical correlation rank statistic of Anderson (1951) when F and G are appropriately specified.

Proposition 1. *When* $\mathscr{V} = ((F'F)^{-1} \otimes (G'G)^{-1})$ *and* $\hat{\Pi}$ *in* (16) *is a least squares estimator, the rank statistic* rk(q) (22) *is identical to the canonical correlation rank statistic of Anderson* (1951). *When* $k > m$, *it is equal to the sum of the* $m - q$ *smallest eigenvalues of* $\hat{\Theta}'\hat{\Theta}$ *divided by* T, *and when* $m > k$, *it is equal to the sum of the* $k - q$ *smallest eigenvalues of* $\hat{\Theta}\hat{\Theta}'$ *divided by* T. *The smallest eigenvalues of* $\hat{\Theta}'\hat{\Theta}$ *and* $\hat{\Theta}\hat{\Theta}'$ *represent the smallest canonical correlations when* $\hat{\Pi}$ *is a least squares estimator.*

Proof. See the appendix. □

Proposition 1 shows that the rk-statistic is a generalization of the canonical correlation rank statistic of Anderson (1951). It generalizes the canonical correlation rank statistic to a more general specification of the covariance matrix $\mathscr{V}$ with the same limiting distribution. A non-Kronecker structure of the covariance matrix $\mathscr{V}$ occurs when we use an estimator $\hat{\Pi}$ that does not give the same weight to every observation or when the disturbances of the underlying model are heteroscedastic or correlated.

Robin and Smith (2000). Propose a rank test which results from the limiting distribution of the roots of the polynomial

$$|\hat{\Pi}'G'G\hat{\Pi} - \mu(F'F)^{-1}| = 0, \quad k \geqslant m, \tag{23}$$

or

$$|\hat{\Pi}F'F\hat{\Pi}' - \mu(G'G)^{-1}| = 0, \quad m < k, \tag{24}$$

which are identical to the eigenvalues of $\hat{\Theta}'\hat{\Theta}$ and $\hat{\Theta}\hat{\Theta}'$. Robin and Smith (2000) show that, under $\mathrm{H}_0 : \mathrm{rank}(\Pi) = q$ and Assumption 1, the limiting distribution of the statistic

$$\mathrm{CRT}(q) = T \sum_{i=q+1}^{\min(k,m)} h(\hat{\mu}_i), \tag{25}$$

where $h(z)$ is a continuous differentiable non-negative scalar function, $h(0) = 0$, $(\mathrm{d}h(z)/\mathrm{d}z)|_{z=0} = 1$, and $\hat{\mu}_1 > \hat{\mu}_2 > \cdots > \hat{\mu}_{\min(k,m)}$ are the roots of the polynomial (23) or (24), is characterized by

$$\sum_{i=1}^{(k-q)(m-q)} \tau_i \psi_i, \tag{26}$$

where ψ_i, $i = 1, \ldots, (k-q)(m-q)$, are independent $\chi^2(1)$ random variables and τ_i, $i = 1, \ldots, (k-q)(m-q)$, are the $(k-q)(m-q)$ eigenvalues of Ω_q.[2] An example of a function h is $h(z) = z$, for other examples we refer to Robin and Smith (2000).

[2] Robin and Smith (2000) express the eigenvectors of the roots of the polynomials (23) and (24) normalized with respect to either $(F'F)^{-1}$ and $(G'G)^{-1}$ so $D'_{k-q}(G'G)^{-1}D_{k-q} = I_{k-q}$ and $E'_{m-q}(F'F)^{-1}E_{m-q} = I_{m-q}$ where the $k \times (k-q)$ and $m \times (m-q)$ matrices D_{k-q} and E_{m-q} contain the eigenvectors that belong to the smallest roots of the polynomials (24) and (23), respectively. The matrix $(E_{m-q} \otimes D_{k-q})'V(E_{m-q} \otimes D_{k-q})$ that results from Theorem 3.2 of Robin and Smith (2000) is therefore

The relation between the two tests follows from the fact that the roots of polynomials (23) and (24) are equal to the squared singular values of $\hat{\Theta}$, since they are equal to the eigenvalues of $\hat{\Theta}'\hat{\Theta}$ and $\hat{\Theta}\hat{\Theta}'$. The smallest $\min(k-q, m-q)$ singular values are contained in S_2. Hence, Robin and Smith (2000) construct the limiting distribution of $\mathrm{tr}(S_2'S_2)$ $(=\mathrm{tr}(S_2S_2'))$. To construct our rk-statistic, we first conduct an orthogonal transformation from S_2 to Λ_q. Because Λ_q has a normal limiting distribution we can obtain its asymptotic covariance matrix. Since S_2 has only non-negative elements, it does not have a normal limiting distribution and we cannot construct the asymptotic covariance matrix of S_2. We use the asymptotic covariance matrix of Λ_q to normalize Λ_q which leads to a statistic with a standard χ^2 limiting distribution. Since S_2 is not appropriately normalized in $\mathrm{tr}(S_2'S_2)$ and $\mathrm{tr}(S_2S_2')$, the limiting distribution of the CRT-statistic is not a standard χ^2 one. It consists of a weighted sum of $\chi^2(1)$ random variables. The weights result from the asymptotic covariance matrix of Λ_q since $\mathrm{tr}(S_2'S_2) = \mathrm{tr}(\Lambda_q'\Lambda_q)$. The critical values of the limiting distribution of the CRT-statistic are therefore not readily available, and given the value of the weights we have to simulate from the limiting distribution to obtain them. Because S_2 is not appropriately normalized in the CRT-statistic, the τ_i, $i = 1, \ldots, (k-q)(m-q)$ are of importance for the power of the rank test. When there is a large difference between the different τ_i, $i = 1, \ldots, (k-q)(m-q)$, the CRT-statistic essentially only focusses on the large τ_i since the small τ_i will not contribute much to the critical value. The rk-statistic normalizes Λ_q using its asymptotic covariance matrix and hence it focusses on all elements of S_2 in a more evenly distributed manner.

When $\mathscr{V} = ((F'F)^{-1} \otimes (G'G)^{-1})$, the values of τ_i are equal to one and the limiting distribution of the CRT-statistic is a $\chi^2((k-q)(m-q))$ distribution. In this case, the CRT-statistic is also identical to the canonical correlation rank statistic of Anderson (1951).

Cragg and Donald (1997). Construct a rank statistic by means of the minimal value of a criterion function,

$$\mathrm{CD}(q) = \min_{\Pi_0 \in \Gamma(q)} \; T(\hat{\pi} - \pi_0)'\mathscr{V}^{-1}(\hat{\pi} - \pi_0), \tag{27}$$

where $\pi_0 = \mathrm{vec}(\Pi_0)$ and $\Gamma(q)$ is the space of $k \times m$ matrices with rank less than or equal to q. Cragg and Donald (1997) show that the CD-statistic (27) has a $\chi^2((k-q)(m-q))$ limiting distribution. Unless $\mathscr{V}$ has a Kronecker structure, in which case the CD-statistic corresponds with the canonical correlation rank statistic, Π_0 cannot be obtained analytically and a numerical optimization procedure has to be used. To test for the rank of Π, we compute $\mathrm{CD}(q)$ sequentially for values of q equal to 0, 1, $\ldots, \min(k, m)$. For large values of k or m, the number of elements of Π_0 becomes large so we have to optimize over a large number of parameters. Numerical optimization does not always perform satisfactory in such cases and it is then

(*footnote continued*)
identical to Ω_q in Assumption 2 since $E_{m-q} = (B_{q,\perp}F)'$ and $D_{k-q} = GA_{q,\perp}$. Note that the full rank assumption of Ω_q is not required in their approach.

important to use an appropriate normalization, see Cragg and Donald (1993). Numerical optimization is therefore a deficiency of the CD-statistic. Another difference with the rk-statistic concerns the covariance matrix $\mathscr{V}$ which has to be non-singular while the rk-statistic only assumes a non-singular covariance matrix Ω_q. In case of a singular covariance matrix $\mathscr{V}$, we have to use a generalized inverse to compute the CD-statistic which can alter the degrees of freedom parameter of the χ^2 limiting distribution, see e.g. Andrews (1987).

Gill and Lewbel (1992)–*Cragg and Donald* (1996). Alongside the singular value decomposition, another well-known matrix decomposition is the LDU-decomposition, see, for example, Golub and van Loan (1989). Gill and Lewbel (1992) and Cragg and Donald (1996) use the LDU-decomposition to construct statistics to test $\mathrm{H}_0 : \mathrm{rank}(\Pi) = q$. To construct these LDU-statistics, permutations are conducted on Π which transform it into Θ,

$$\Theta = G\Pi F', \tag{28}$$

where G and F are $k \times k$ and $m \times m$ permutation matrices that, in case we want to test $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$, are such that the first q columns and rows of Θ perform a Gaussian elimination of Π. For the LDU-decomposition, Θ is specified as

$$\Theta = LDU, \tag{29}$$

with $L : k \times k$, $D : k \times m$ and $U : m \times m$ matrices that are specified by

$$L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & I_{k-q} \end{pmatrix}, \quad D = \begin{pmatrix} D_{11} & 0 \\ 0 & D_{22} \end{pmatrix}, \quad U = \begin{pmatrix} U_{11} & U_{12} \\ 0 & I_{m-q} \end{pmatrix}, \tag{30}$$

where L_{11}, D_{11}, $U_{11} : q \times q$; $L_{21} : (k-q) \times q$; $D_{22} : (k-q) \times (m-q)$; $U_{12} : q \times (m-q)$; and L_{11} and U'_{11} are lower triangular matrices with ones on the diagonal and D_{11} is a diagonal matrix.[3] The Gaussian elimination accomplishes that the largest elements of Θ are grouped in D_{11}, L_{11}, L_{21}, U_{11} and U_{12} and the smallest in D_{22}. Under $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$, D_{22} is therefore equal to zero.

To construct the LDU-statistic to test $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$, we conduct the same LDU-decomposition with Gaussian elimination (28)–(30) on $\hat{\Pi}$ as on Π . Cragg and Donald (1996) show that under $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$, Assumptions 1, 2, and an assumption on the Gaussian elimination that $\mathrm{vec}(\hat{D}_{22})$ has a normal limiting distribution with mean zero. The LDU-statistic is equal to a quadratic form of $\mathrm{vec}(\hat{D}_{22})$ with the inverse of its asymptotic covariance matrix and converges under H_0 and the before-mentioned assumptions to a $\chi^2((k-q)(m-q))$ distributed random variable.

The LDU-decomposition is not unique and different orderings of the data result in different LDU-decompositions. Before conducting the LDU-decomposition, Cragg

[3] In the LDU-decomposition that is used by Gill and Lewbel (1992), the matrices L_{22} and U_{22}, which are identity matrices in (30), are also lower triangular with ones on the diagonal and D_{22} is a diagonal matrix. Cragg and Donald (1996) show that the estimator of D_{22}, $\hat{D}_{22}$, in this case has not a normal limiting distribution under $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$. We therefore use the LDU-decomposition that is proposed by Cragg and Donald (1996).

and Donald (1996) therefore permute $\hat{\Pi}$ q times using Gaussian elimination. The Gaussian elimination re-orders the columns and rows of $\hat{\Pi}$ such that the first q rows and columns of $\hat{\Theta}$ contain the largest elements of $\hat{\Pi}$. The Gaussian elimination accomplishes that the elements of $\hat{D}_{11}$ are significantly different from zero under H_0. In that case, $\hat{D}_{22}$ can be used to test $\mathrm{H}_0 : \mathrm{rank}(\Theta) = q$. Because of the triangular structure of $\hat{L}$ and $\hat{U}$ with ones on the diagonal, both $\hat{D}_{11}$ and $\hat{D}_{22}$ reflect whether $\hat{\Theta}$ is close to a lower rank value. The Gaussian elimination is needed to ensure that only $\hat{D}_{22}$ reflects whether $\mathrm{rank}(\hat{\Theta})$ is close to q. This explains why Cragg and Donald (1996) need an additional assumption, alongside Assumptions 1 and 2, to ensure that the Gaussian elimination performs adequately. Since this assumption is not made for the rk-statistic, the LDU-statistic applies in a more restricted setting than the rk-statistic although their limiting distributions are identical.

Ratsimalahelo (2001,2002). When we express λ_q and Ω_q in terms of the elements of U, S and V we obtain: $\lambda_q = [(V_{22}V_{22}')^{-1/2}V_{22} \otimes (U_{22}U_{22}')^{-1/2}U_{22}]\mathrm{vec}(S_2)$ and

$$\Omega_q = ((V_{22}V_{22}')^{-1/2}V_{22} \otimes (U_{22}U_{22}')^{-1/2}U_{22})^{-1\prime}\left(\begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} \otimes \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}\right)' \mathscr{W} \times \left(\begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} \otimes \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}\right)((V_{22}V_{22}')^{-1/2}V_{22} \otimes (U_{22}U_{22}')^{-1/2}U_{22})^{-1}. \tag{31}$$

The expression for the rk(q) (22) then becomes

$$\mathrm{rk}(q) = T\mathrm{vec}(S_2)'\left[\left(\begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} \otimes \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}\right)' \mathscr{W} \left(\begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} \otimes \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}\right)\right]^{-1} \mathrm{vec}(S_2), \tag{32}$$

which equals the statistic that is proposed in Ratsimalahelo (2002). Theorem 1 states that rk(q) has a $\chi^2(k-q)(m-q)$ limiting distribution. In the construction of this limiting distribution, we cannot use asymptotic normality of the limiting distribution of S_2 because S_2 is a diagonal matrix with non-negative diagonal elements so it cannot have a zero-mean normal limiting distribution. Robin and Smith (2000) therefore opt for the limiting distribution of $\mathrm{vec}(S_2)'\mathrm{vec}(S_2) = \mathrm{tr}(S_2'S_2)$ directly without any reference to asymptotic normality of $\mathrm{vec}(S_2)$. Ratsimalahelo (2002) obtains the limiting distribution of rk(q) using the asymptotic normality of an orthogonal transformation of S_2. The orthogonal transformation used by Ratsimalahelo converges, however, to the identity matrix when the sample size gets large, see Proposition 1 in Ratsimalahelo (2002). His construction of the limiting distribution is therefore based on asymptotic normality of S_2. Surprisingly, the limiting distribution of rk(q) is not affected by the invalid intermediate result of asymptotic normality of S_2.

Our orthogonal transformation of S_2 given by Λ_q adds a sign to the non-negative singular values S_2. This can easily be demonstrated when $k = m = 2$ and $q = 1$. The two possible orthogonal specifications for either U and V in the singular value decomposition are given by

$$U^{(1)} = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \quad \text{or} \quad U^{(2)} = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ \sin(\varphi) & -\cos(\varphi) \end{pmatrix},$$
$$V^{(1)} = \begin{pmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix} \quad \text{or} \quad V^{(2)} = \begin{pmatrix} \cos(\psi) & \sin(\psi) \\ \sin(\psi) & -\cos(\psi) \end{pmatrix} \tag{33}$$

with $0 \leqslant \varphi < 2\pi$ and $0 \leqslant \psi < 2\pi$. There is only one of the four combinations of the U and V specifications that results in a non-negative value of S_1 and S_2 This specification is used by the singular value decomposition to determine S_1 and S_2 and is therefore not known beforehand.

When we use both specifications of U and V to construct $A_{q,\perp}$ and $B_{q,\perp}$, we obtain

$$1.\ A_{q,\perp}^{(1)} = \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix} \frac{\cos(\varphi)}{\sqrt{\cos(\varphi)^2}} = \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}$$

$$2.\ A_{q,\perp}^{(2)} = \begin{pmatrix} \sin(\varphi) \\ -\cos(\varphi) \end{pmatrix} \frac{-\cos(\varphi)}{\sqrt{\cos(\varphi)^2}} = \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix} \frac{\cos(\varphi)}{\sqrt{\cos(\varphi)^2}}$$
$$= \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix} \tag{34}$$

and

$$1.\ B_{q,\perp}^{(1)} = \frac{\cos(\psi)}{\sqrt{\cos(\psi)^2}} \begin{pmatrix} -\sin(\psi) \\ \cos(\psi) \end{pmatrix}' = \begin{pmatrix} -\sin(\psi) \\ \cos(\psi) \end{pmatrix}'$$

$$2.\ B_{q,\perp}^{(2)} = \frac{-\cos(\psi)}{\sqrt{\cos(\psi)^2}} \begin{pmatrix} \sin(\psi) \\ -\cos(\psi) \end{pmatrix}' = \frac{\cos(\psi)}{\sqrt{\cos(\psi)^2}} \begin{pmatrix} -\sin(\psi) \\ \cos(\psi) \end{pmatrix}'$$
$$= \begin{pmatrix} -\sin(\psi) \\ \cos(\psi) \end{pmatrix}'. \tag{35}$$

Hence, the specifications of $A_{q,\perp}$ and $B_{q,\perp}$ are invariant to the specification of U and V. It implies that for $k-q=m-q=1$, $\Lambda_q = (U_{22}U_{22}')^{-1/2}U_{22}S_2V_{22}'(V_{22}V_{22}')^{-1/2}$ simplifies to $\Lambda_q = \text{sign}(U_{22} \times V_{22})S_2$ since $(U_{22}U_{22}')^{-1/2}U_{22}$ and $V_{22}'(V_{22}V_{22}')^{-1/2}$ equal ± 1. Hence, the orthogonal transformation in Λ_q gives a sign to S_2. The limiting distribution of Λ_q can therefore be normal with a zero mean, while the limiting distribution of S_2 cannot. The above reasoning extends to higher dimensional settings as well.

4. Rank test: the non-stationary cointegration case

In vector autoregressive (VAR) models, cointegration implies a reduced rank value of the long-run multiplier, see, for example, Engle and Granger (1987) and Johansen (1991, 1995). A convenient specification of a VAR(1) model to reflect cointegration is the so-called error correction specification

$$\Delta y_i = \Pi y_{i-1} + \varepsilon_i \quad \text{for } i = 1, \ldots, T, \tag{36}$$

where y_i is a $k \times 1$ vector that contains time-series observations of the variable y at time i, $\Delta y_i = y_i - y_{i-1}$, ε_i is a $k \times 1$ vector of disturbances which we assume to be uncorrelated over time. Cointegration occurs when the $k \times k$ matrix Π has a reduced rank value, which we can denote as $\Pi = A_q B_q$ with A_q a $k \times q$ matrix and B_q a $q \times k$ matrix. The canonical correlation rank test for the rank of the long-run multiplier is in this case identical to the Johansen trace statistic, see Johansen (1991). These statistics apply when the covariance matrix of the disturbances ε_i is constant over time.

In Kleibergen and van Dijk (1994), the LDU decomposition is used to decompose the long-run multiplier $\hat{\Pi}$ to construct an alternative cointegration test. This cointegration test has however the same variable ordering problem as the LDU-based test in the stationary case. Therefore we construct in this paper a test for rank reduction in case of non-stationary variables, where we follow the same approach as in the previous section. Because the dependent and explanatory variables are different realizations of the same (economic) variables over time, we do not need to normalize the estimator $\hat{\Pi}$ to obtain a rank statistic. As mentioned before, an appropriately normalized rank statistic corresponds with the canonical correlation rank test that is identical to the Johansen trace statistic. We use the SVD to decompose the unrestricted least squares estimator $\hat{\Pi} = \sum_{i=1}^{T} \Delta y_i y_{i-1}' (\sum_{i=1}^{T} y_{i-1} y_{i-1}')^{-1}$ as

$$\hat{\Pi} = \hat{A}_q \hat{B}_q + \hat{A}_{q,\perp} \hat{\Lambda}_q \hat{B}_{q,\perp}. \tag{37}$$

The rank statistic corresponds to a quadratic form in $\lambda_q = \text{vec}(\Lambda_q)$. To derive the limiting distribution of the rank statistic, we use the Representation Theorem for cointegrated series, see, for example, Johansen (1991, 1995). Theorem 2 states the limiting behavior of the rank statistic under the assumptions that the disturbances ε_i are a white noise process with constant covariance matrix Σ.

Theorem 2. *When the disturbances ε_i are white noise with constant covariance matrix Σ, $\Pi = A_q B_q$ and $A_{q,\perp}' B_{q,\perp}'$ is non-singular, the limiting behavior of the rank statistic*

$$\text{rk}(q) = T \hat{\lambda}_q' \hat{\Omega}_q^{-1} \hat{\lambda}_q \tag{38}$$

with $\hat{\lambda}_q = \text{vec}(\hat{\Lambda}_q)$,

$$\hat{\Omega}_q = (\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}') \hat{\mathscr{W}} (\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}')' \tag{39}$$

and

$$\hat{\mathcal{W}} = \left(\left(\frac{1}{T} \sum_{i=1}^{T} y_{i-1} y_{i-1}' \right)^{-1} \otimes \frac{1}{T} \sum_{i=1}^{T} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \right), \tag{40}$$

where $\tilde{\varepsilon}_i = \Delta y_i - \hat{\Pi} y_{i-1}$, *is characterized by*

$$\begin{aligned} \mathrm{rk}(q) \underset{\mathrm{d}}{\rightarrow} \mathrm{tr} \Bigg[& \left(\int_0^1 B_{k-q}(t)\, \mathrm{d}B_{k-q}(t)' \right)' \left(\int_0^1 B_{k-q}(t) B_{k-q}(t)'\, \mathrm{d}t \right)^{-1} \\ & \times \left(\int_0^1 B_{k-q}(t)\, \mathrm{d}B_{k-q}(t)' \right) \Bigg], \end{aligned} \tag{41}$$

where $B_{k-q}(t)$ *is a* $(k-q)$*-dimensional Brownian motion defined on the unit interval with identity covariance matrix at time* t.

Proof. See the appendix. □

The limiting distribution in Theorem 2 is identical to the limiting distribution of the Johansen trace statistic, see Johansen (1991, 1995). The rank statistic (38) results directly from a SVD of the least squares estimator. It can therefore be considered as a multivariate generalization of the Dickey–Fuller t-statistic for testing unit roots in univariate autoregressive models. The Johansen trace statistic is a likelihood ratio statistic while the Dickey–Fuller statistic and the rank statistic (38) are Wald statistics. The rank statistic (38) is similar to the cointegration statistic proposed in Saikkonen (1999). Rank statistic (38) results from an unrestricted specification of the long run multiplier and is therefore a Wald statistic while Saikkonen's (1999) statistic results from an auxiliary regression and is therefore a Lagrange multiplier statistic. Hence, the statistics differ in the employed estimators of the covariance matrix.

The derivation of the limiting distribution of the rank test in the appendix shows that we can also obtain the rank test from

$$\begin{aligned} \mathrm{rk}(q) = \frac{1}{T} & \left[(\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}') \mathrm{vec} \left(\sum_{i=1}^{T} \hat{\varepsilon}_i y_{i-1}' \right) \right]' [(\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}') \hat{\mathcal{V}} (\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}')']^{-1} \\ & \times \left[(\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}') \mathrm{vec} \left(\sum_{i=1}^{T} \hat{\varepsilon}_i y_{i-1}' \right) \right], \end{aligned} \tag{42}$$

where $\hat{\varepsilon}_i = \Delta y_i - \hat{A}_q \hat{B}_q y_{i-1}$ and $\hat{\mathcal{V}} = (\frac{1}{T} \sum_{i=1}^{T} y_{i-1} y_{i-1}' \otimes \frac{1}{T} \sum_{i=1}^{T} \tilde{\varepsilon}_i \tilde{\varepsilon}_i')$. Based on this specification of the rank test, we can define a set of assumptions, which are more general than the white noise constant variance assumption, under which the test has the same limiting distribution as in (41), that is, the asymptotic distribution of the Johansen trace statistic.

Assumption 3. The limiting behavior of vec $(\sum_{i=1}^{T}\hat{\varepsilon}_i y_{i-1}')$ is given by

$$\frac{1}{T}\left[(\hat{B}_{q,\perp}\otimes\hat{A}_{q,\perp}')\text{vec}\left(\sum_{i=1}^{T}\hat{\varepsilon}_i y_{i-1}'\right)\right]\underset{d}{\rightarrow}\Psi_q^{1/2}\left(\int_0^1 B_{k-q}(t)\,\mathrm{d}B_{k-q}(t)'\right)'\times\Sigma_q^{1/2}(A_{q,\perp}'B_{q,\perp}')^{-1\prime}, \tag{43}$$

while the limiting behavior of $\hat{\Omega}_q=(\hat{B}_{q,\perp}\otimes\hat{A}_{q,\perp}')\hat{\mathscr{V}}(\hat{B}_{q,\perp}\otimes\hat{A}_{q,\perp}')'$ accords with

$$\frac{1}{T}\hat{\Omega}_q\underset{d}{\rightarrow}\left((A_{q,\perp}'B_{q,\perp}')^{-1}\Sigma_q^{1/2}\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)\Sigma_q^{1/2}(A_{q,\perp}'B_{q,\perp}')^{-1\prime}\otimes\Psi_q\right), \tag{44}$$

where Σ_q and Ψ_q are $(k-q)\times(k-q)$ dimensional non-singular matrices.

This assumption holds under the conditions given in Phillips and Solo (1992) and is valid for processes with time-varying conditional variances for which the unconditional variance is finite. Hence, Assumption 3 allows us to use the limiting distribution given in (41) for our rank statistic (42) for a more general set of processes than white noise with constant covariance matrix.

The decomposition in (37) can also be applied to other estimators of the long-run multiplier, for instance to those that result when we incorporate higher order lags or use different weights for the observations to account for time varying conditional variances or non-normal distributed errors.

5. Testing rank in stochastic discount factor models

Jagannathan and Wang (1996) base the stochastic discount factor (SDF) in a SDF model on a conditional capital asset pricing model (CAPM). A SDF model for asset returns is represented by, see, for example, Campbell et al. (1997),

$$\mathrm{E}[r_{ji+1}s_{i+1}|I_i]=1,\quad j=1,\ldots,k, \tag{45}$$

where r_{ji} is the return at time i on asset j, $j=1,\ldots,k$, I_i is the information set at time i and s_i is the SDF at time i. Based on the conditional CAPM, Jagannathan and Wang (1996) specify the SDF s_{i+1} as a linear function of a set of additional variables:

$$s_{i+1}=f_{i+1}'\gamma, \tag{46}$$

where f_{i+1} is an $m\times 1$ vector of additional variables and γ is an $m\times 1$ vector of risk premia. We substitute the specification of the SDF (46) into the SDF model (45) to obtain

$$\{\mathrm{E}[r_{i+1}f_{i+1}'|I_i]\}\gamma=\iota_k, \tag{47}$$

where $r_i=(r_{1i}\ldots r_{ki})'$ and ι_k is a $k\times 1$ vector of ones. The vector of risk premia γ can be estimated using Hansen's (1982) generalized method of moments (GMM), see, for example, Jagannathan et al. (2002). The GMM estimator of γ is consistent if the

$k \times m$ matrix

$$\mathrm{E}[r_{i+1}f'_{i+1}|I_i] \tag{48}$$

has full rank, see Hansen (1982), Newey and McFadden (1994) and Wright (2003). Tests of the rank of (48) therefore indicate whether all elements of γ are identified.

The asset returns that are used in SDF models are typically returns on portfolios. These portfolios are constructed using some kind of criterion, like, for example, the market capitalization or size of stocks. It is well known now that the returns on portfolios that consist of small firm stocks exhibit autocorrelation, see, for example, Campbell et al. (1997). This implies that the canonical correlation rank statistic to test for the rank of (48) is essentially inappropriate since the (asymptotic) covariance matrix of an estimator of (48) does not possess a Kronecker structure. In the next sub-sections, we therefore use the previously discussed rank tests, which can be applied when the covariance matrix does not have a Kronecker structure, to test for the rank of (48) in SDF models.

5.1. Application

Jagannathan and Wang (1996) construct return series on hundred size and beta sorted portfolios to estimate the risk premia in the SDF model (45). The return series begin in July 1963 and end in December 1990 so $T = 330$. The SDF (46) is a linear function of a constant, the return on a value-weighted portfolio, a corporate bond yield spread and per capita labor income growth. The size and beta sorted portfolios are obtained by ranking all non-financial stocks listed in the NYSE and AMEX and covered by CRSP with respect to size. For every decile of size sorted stocks, the stocks are once more sorted with respect to their beta, that is their correlation with the market return. All stocks in such a beta sorted decile constitute one portfolio. Since we have both ten size and beta sorted deciles, the total number of portfolios equals hundred. Inverting a 100×100 leads to a large numerical error. To decrease the number of portfolios, we use only the returns on the smallest beta portfolio within each size decile. This reduces the number of portfolios to ten which are sorted with respect to size.

When the expectation of $r_{i+1}f'_{i+1}$ does not depend on the information set I_i, the expectation in (48) is an unconditional expectation. Instead of testing the rank of (48), we can then also analyze the rank of the $k \times m$ matrix Π in a linear factor model (9) for the returns on the ten size sorted portfolios,

$$r_i = \Pi f_i + \varepsilon_i, \quad i = 1, \ldots, T, \tag{49}$$

where $k = 10, f_i = (f_{1i} \ldots f_{mi})'$ with $m = 4$, $T = 330$, and the $k \times 1$ vector ε_i contains the disturbances. The returns contained in r_i are such that r_{1i} is the return on the smallest size portfolio and r_{ki} is the return on the largest size portfolio at time i. The factors in f_i consist of a constant term (c), a return on a value-weighted portfolio (vw), a corporate bond yield spread (int) and a measure of per capita labor income growth ($labor$).

Returns on portfolios of small firm stocks are known to exhibit autocorrelation, see, for example, Campbell et al. (1997). We test for first order autocorrelation in the disturbances using a Lagrange multiplier (LM) statistic. The LM statistic results from regressing the residuals on the lagged residuals. The LM statistic equals 141.4 which is well above the 5% significance level of the $\chi^2(10^2)$ distribution that is equal to 124.2. The (asymptotic) significance of the LM statistic for first order autocorrelation indicates that we have to be careful with interpreting the canonical correlation rank statistic. Alongside the canonical correlation rank statistic, we therefore compute rank statistics that allow for non-Kronecker covariance matrices. Using a consistent covariance matrix estimator, the limit distributions of these rank statistics also apply when the disturbances exhibit autocorrelation. The rank statistics that we compute are the LDU rank statistic, the rk-statistic (22) and the CD-statistic (27). We do not consider the rank statistic of Robin and Smith (2000) because, as shown in Section 3, it consists of exactly the same elements as the rk-statistic and has a non-standard limiting distribution which makes a direct comparison more difficult.

To compute the rank statistics, we first transform or normalize the least-squares estimator of Π, $\hat{\Pi} = \sum_{i=1}^{T} r_i f_i' (\sum_{i=1}^{T} f_i f_i')^{-1}$, towards $\hat{\Theta} = G\hat{\Pi}F$ (16) by using a specification of F and G corresponding to

$$\left(\left(\sum_{i=1}^{T} f_i f_i'\right)^{-1} \otimes \sum_{i=1}^{T} r_i r_i'\right) = ((F'F)^{-1} \otimes (G'G)^{-1}) \tag{50}$$

such that the covariance matrix estimator for $\hat{\theta}$ becomes

$$\hat{\mathscr{W}} = (F \otimes G)\hat{\mathscr{V}}(F \otimes G)'$$

with $\hat{\mathscr{V}}$ the covariance matrix estimator for $\hat{\pi}$. When the covariance matrix estimator $\hat{\mathscr{V}}$ accords with (50), the rk-statistic is identical to the canonical correlation rank statistic, as in Proposition 1. Hence, the specification of F and G is such that the rk-statistic uses the same elements as the canonical correlation rank statistic to discriminate between different rank values of Θ but with a consistent covariance matrix estimator. We therefore consider the rk-statistic with this specification of F and G as a size-corrected canonical correlation rank statistic. We use HACC matrix estimators for the covariance matrix estimator $\hat{\mathscr{V}}$.[4]

Table 1 lists the values of statistics for testing the rank of Π using different HACC matrix estimators. The LDU rank statistic results from the same specification of $\hat{\Theta}$ as we used for the rk-statistic. The applied HACC matrix estimators are the Newey–West (1987) estimator with one lag, the White (1980) estimator and the West (1997) estimator with one lag. The latter estimator is intended for moving average processes.

The values of the rank statistics in Table 1 indicate that rank zero and rank one for Π are rejected with 5% asymptotic significance. The value of the LDU-statistic to

[4]The specification of F and G that transforms $\hat{\mathscr{V}}$ towards $\hat{\mathscr{W}}$ is also convenient to control the numerical error. The covariance matrix $\hat{\mathscr{V}}$ is a 40×40 matrix which can be difficult to invert when it is not scaled appropriately. This is achieved by using the specification of F and G from (50).

Table 1
Rank statistics with different covariance matrix estimators (can. cor. stands for canonical correlation, NW stands for Newey–West and CV stands for critical value)

Rank	Test statistic						5%CV
	rk-NW	rk-White	rk-West	LDU-West	CD-West	Can. cor.	
0	91.80	114.7	107.4	107.4	107.4	286.4	55.7
1	31.45	34.76	33.78	35.72	33.63	31.24	40.1
2	12.79	13.66	14.22	8.70	21.87	12.78	26.3
3	2.21	2.06	2.76	0.64	3.39	1.97	14.1

test for rank two is quite different from the rank value of the other statistics. This is due to the sensitivity of the LDU-statistic to the ordering of the variables. Depending on the ordering of the series in f_i, the values of the rank tests can differ substantially. This explains why Cragg and Donald (1996) permute $\hat{\Theta}$ using Gaussian elimination. All other statistics are insensitive to the ordering of the variables. The CD-statistic (27) is obtained using numerical optimization. The rank two case of the CD-statistic differs somewhat from the other statistics but does not lead to a different conclusion with respect to the hypothesis of interest. The CD statistic is computed using numerical optimization and the number of parameters for the rank 2 and 3 cases, 24 and 33, is so large that numerical optimization does not always perform satisfactorily. This shows that the need for numerical optimization is a drawback for practical purposes.

5.2. *Size comparison*

The canonical correlation rank statistic that tests for rank one is below its asymptotic 5% significance level. Because of the autocorrelated disturbances, the critical value is, however, inappropriate such that we cannot assess it. To assess the validity of the asymptotic critical values of the different rank statistics from Table 1, we bootstrap the distributions of these statistics. We estimate the parameters of the factor model (49) with disturbances that are first order moving averages. The estimated parameters are used in the data generating process to obtain the distributions of the rank statistics. The parameter values of the data generating process are given in the appendix. The true value of Π in the data generating process is the canonical correlation estimate of Π with rank one and we use the same values for f_i, $i = 1, \ldots, T$ as in the original model. The disturbances in the model are simulated from a normal distribution with mean zero, where we incorporate their moving average property in the data generating process.

Fig. 1 shows the empirical distribution functions of the rank statistics obtained from the outlined data generating process. Fig. 1 shows that the distribution of the canonical correlation rank statistic lies below its asymptotic distribution which is a $\chi^2(27)$ distribution. The empirical distribution functions of the rank statistics that employ the White and West covariance matrix estimators, that is, the rk-West,

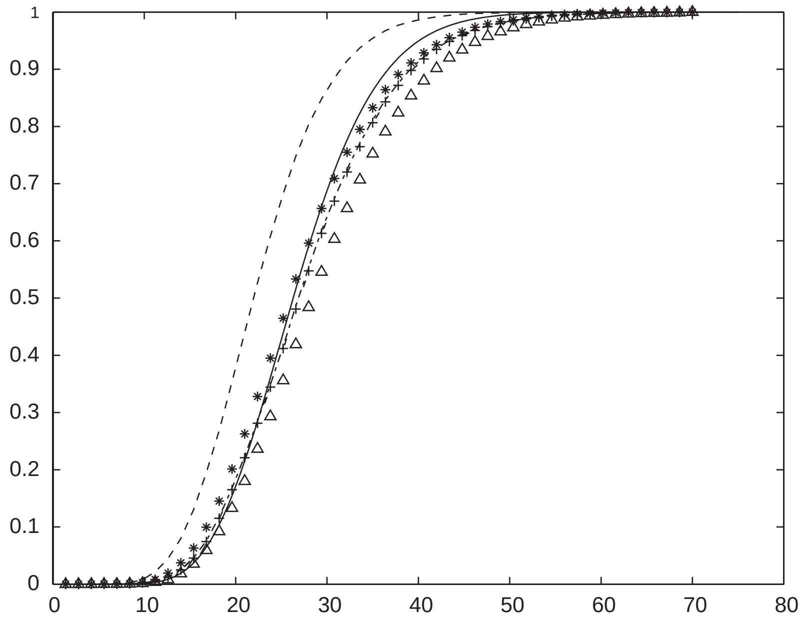

Fig. 1. Empirical distribution functions of rank statistics that test for rank one while true rank is equal to one: asymptotic distribution-$\chi^2(27)$ (solid line), can. cor. (dashed), CD-West (dotted), LDU-West (dashed-dotted), rk-West (plusses), rk-White (stars), rk-NW (triangles).

rk-White, CD-West and LDU-West, lie more on less on top of their asymptotic distribution function. This indicates that these HACC estimators perform adequately in restoring the distribution of the canonical correlation rank statistic. The empirical distribution function of the rk-statistic with the Newey–West covariance matrix estimator lies above its asymptotic distribution. For this data generating process, the Newey–West covariance matrix estimator therefore performs less satisfactory than the West and White covariance matrix estimators.

Fig. 1 supports the validity of the asymptotic critical values for the rank statistics with the HACC matrix estimators reported in Table 1. We conclude from Table 1 that we cannot reject rank values of Π equal to two and three. Hence, we cannot reject lower rank values of Π and also not the non-identifiedness of (elements of) the vector of risk premia γ. GMM estimates of γ can therefore be estimates of non-identified parameters which implies that their value can be more or less anything. This is reflected by the value and standard error (s.e.) of the GMM estimates of γ in Jagannathan and Wang (1996, Table IIC): $\gamma_c = 2.26$ (s.e. $= 0.35$), $\gamma_{vw} = 1.81$ (s.e. $= 1.44$), $\gamma_{int} = -65.72$ (s.e. $= 21.2$) and $\gamma_{labor} = -97.72$ (s.e. $= 33.2$). As the scale of the factors in f_i is about the same, very large values of parameters and standard errors may indicate that the parameters are badly identified. The value of the rank statistics further indicates the issue of identification of the parameters. This also holds for most of the other specifications of the SDF in Table II of Jagannathan and Wang (1996).

5.3. Power comparison

We compare the power of the different rank statistics using the same data generating process that we employed for the empirical distribution functions in Fig. 1. Hence, we generate data from the model

$$r_i = \Pi(\lambda^*) f_i + \varepsilon_i, \quad i = 1, \ldots, T, \tag{51}$$

where ε_i are moving average disturbances of order 1 with normal distributed innovations and

$$\Pi(\lambda^*) = \tilde{\Pi}_1 + \lambda^*(\hat{\Pi} - \tilde{\Pi}_1), \tag{52}$$

with $\hat{\Pi}$ the least squares estimate of Π for the Jagannathan and Wang (1996) data and $\tilde{\Pi}_1$ is the canonical correlation estimate of Π when it has rank one. Fig. 2 shows the power curves of the different statistics that test for a rank one value of Π with 5% asymptotic significance for various values of λ^*. Fig. 2 shows, just like Fig. 1, that the rank statistics which involve the West or White covariance matrix estimators are approximately size correct. The canonical correlation rank statistic and the rank statistic that uses the Newey–West covariance matrix estimator are size distorted.

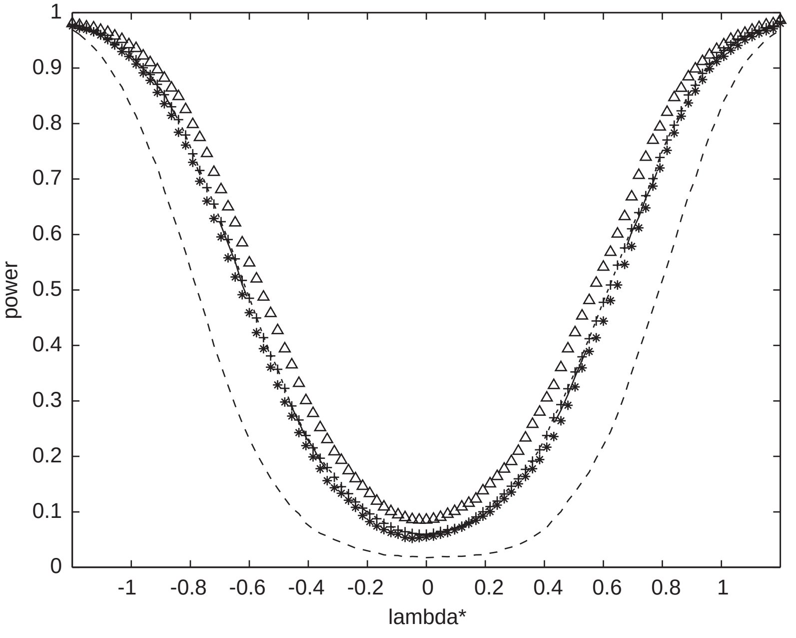

Fig. 2. Power curves of rank statistics that test for rank one with 5% asymptotic significance: can. cor. (dashed), CD-West (dotted), LDU-West (dashed-dotted), rk-West (plusses), rk-White (stars), rk-NW (triangles).

The power of the rank statistics at other values of λ^* behaves in a similar way, that is, the power curves of the rank statistics that use the West and White covariance matrix estimators are very similar and the power curves of the canonical correlation rank statistic and the rank statistic that involves the Newey–West covariance matrix estimator lie consistently below or above these.

In case of a non-Kronecker product form covariance matrix, the CD-statistic is the most powerful procedure because it uses the optimal estimate of the reduced rank matrix. The similarity of the power curves of the CD, rk and LDU statistics which use the West (1997) covariance matrix estimator shows that the covariance matrix almost has a Kronecker product form structure since the power curves of the rk and LDU statistics coincide with the power curve of the CD-statistic. The Kronecker product form structure of the covariance matrix results since the disturbances are simulated while the factors f_i remain fixed. Hence, the factors and simulated disturbances are independent so a Kronecker product form covariance matrix results. The canonical correlation rank statistic is size distorted since is does not account for the autocorrelated disturbances.

6. Concluding remarks

To overcome some of the deficiencies of statistics that test the rank of a matrix, we propose a novel one. The rank statistic requires a consistent estimator of the unrestricted matrix, which does not need to have a Kronecker covariance matrix. This allows one, for example, to use HACC estimators for the covariance matrix. The unrestricted matrix is decomposed using a singular value decomposition. The rank statistic is a quadratic form of an orthogonal transformation of the smallest singular values with the inverse of the respective covariance matrix. If the limiting distribution of the estimator of the unrestricted matrix is normal, the proposed rank statistic has a standard χ^2 limiting distribution. The rank statistic simplifies to the canonical correlation rank statistic if the covariance matrix of the unrestricted matrix has a Kronecker covariance matrix. In case of non-stationary cointegrated variables, the limiting distribution of the rank statistic equals the limiting distribution of the Johansen trace statistic.

In many econometric models, the rank of a matrix governs the identification of the parameters. The limiting distribution of estimators of these parameters are only valid if this matrix has full rank. For example, in order to obtain the limiting distributions of GMM estimators, it is assumed that a matrix of derivatives has full rank, see, for example, Hansen (1982) and Newey and McFadden (1994). In such cases, rank statistics can be used to test for the identification of the parameters. We use rank statistic to test for the identification of the parameters in the SDF model of Jagannathan and Wang (1996). We find that lower rank values of the identifying matrix cannot be rejected. The parameters in the SDF model of Jagannathan and Wang are therefore badly identified. This has consequences for their empirical results and the resulting conclusions.

Acknowledgements

We thank three anonymous referees and participants of ESEM 2003 in Stockholm and the EC[2] meeting in Bologna for helpful comments and suggestions. The first authors' research documented in this article has been partially funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models".

Appendix

Proof of Theorem 1. Under H_0, $\Theta = A_q B_q$, and Assumption 1,

$$\|\hat{\Theta} - \Theta\| \underset{p}{\to} 0,$$

where $\| \ \|$ is a distance function. We pre-multiply this expression by $(A_q \vdots A_{q,\perp})'$ and post-multiply it by $(B_q' \vdots B_{q,\perp}')$ to obtain

$$\left\| \begin{pmatrix} A_q'(\hat{\Theta} - A_q B_q)B_q' & A_q'\hat{\Theta}B_{q,\perp}' \\ A_{q,\perp}'\hat{\Theta}B_q' & A_{q,\perp}'\hat{\Theta}B_{q,\perp}' \end{pmatrix} \right\| \underset{p}{\to} 0.$$

It results from the SVD that $\hat{\Theta} = \hat{A}_q\hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}$ and so

$$\left\| \begin{pmatrix} A_q'(\hat{A}_q\hat{B}_q - A_q B_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp})B_q' & A_q'(\hat{A}_q\hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp})B_{q,\perp}' \\ A_{q,\perp}'(\hat{A}_q\hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp})B_q' & A_{q,\perp}'(\hat{A}_q\hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp})B_{q,\perp}' \end{pmatrix} \right\| \underset{p}{\to} 0.$$

Since the q largest singular values of $\hat{\Theta}$ are contained in $\hat{A}_q\hat{B}_q$, $\hat{\Lambda}_q$ is an orthogonal transformation of the smallest singular values and $\hat{A}_{q,\perp}'\hat{A}_{q,\perp} = I_{k-q}$, $\hat{B}_{q,\perp}\hat{B}_{q,\perp}' = I_{m-q}$, the above expression implies that $\hat{A}_q\hat{B}_q \underset{p}{\to} A_q B_q$, $\hat{B}_q B_{q,\perp}' \underset{p}{\to} 0$, $A_{q,\perp}'\hat{A}_q \underset{p}{\to} 0$ and $\hat{\Lambda}_q \underset{p}{\to} 0$.Further usage of these results and Assumption 1 gives

$$\begin{aligned} &\sqrt{T}\mathrm{vec}(\hat{\Theta} - A_q B_q) \underset{d}{\to} \mathrm{N}(0, \mathscr{W}) \Leftrightarrow \\ &\sqrt{T}\mathrm{vec}(\hat{A}_q\hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp} - A_q B_q) \underset{d}{\to} \mathrm{N}(0, \mathscr{W}) \Leftrightarrow \\ &\sqrt{T}\mathrm{vec}(\hat{A}_q\hat{B}_q - A_q B_q) + \sqrt{T}\mathrm{vec}(\hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}) \underset{d}{\to} \mathrm{N}(0, \mathscr{W}). \end{aligned}$$

This implies that $\hat{A}_q\hat{B}_q$ is a root-T consistent estimator of $A_q B_q$. Using the normalization imposed on either $\hat{A}_q$ or $\hat{B}_q$, we can show the root-T consistency of the estimators involved in $\hat{A}_q\hat{B}_q$. For example, in case that $\hat{B}$ is normalized as $\hat{B} = [I_q \vdots \hat{B}_{q,2}]$ with $\hat{B}_{q,2}$ a $q \times (m-q)$ matrix,

$$\begin{aligned} \hat{A}_q\hat{B}_q - A_q B_q &= [\hat{A}_q - A_q \vdots \hat{A}_q\hat{B}_{q,2} - A_q B_{q,2}] \\ &= [\hat{A}_q - A_q \vdots A_q(\hat{B}_{q,2} - B_{q,2}) + (\hat{A}_q - A_q)\hat{B}_{q,2}] \\ &= (\hat{A}_q - A_q)[I_q \vdots \hat{B}_{q,2}] + A_q(\hat{B}_{q,2} - B_{q,2})[0 \vdots I_q]. \end{aligned}$$

Since $\hat{A}_q\hat{B}_q$ is a root-T consistent estimator of A_qB_q, $\hat{A}_q$ and $\hat{B}_{q,2}$ are root-T consistent estimators of A_q and $B_{q,2}$, respectively. Furthermore, by pre-multiplying the above expression by either I_k or $(A_q'A_q)^{-1}A_q'$ and post-multiplying by either $B_q'(B_qB_q')^{-1}$ or $[0\vdots I_q]'$, we obtain that

$$\sqrt{T}\mathrm{vec}(\hat{A}_q - A_q) \underset{d}{\rightarrow} \mathrm{N}(0, \mathscr{W}_A),$$
$$\sqrt{T}\mathrm{vec}(\hat{B}_{q,2} - B_{q,2}) \underset{d}{\rightarrow} \mathrm{N}(0, \mathscr{W}_B)$$

with $\mathscr{W}_A = ((B_qB_q')^{-1}B_q \otimes I_k)\mathscr{W}((B_qB_q')^{-1}B_q \otimes I_k)'$, and $\mathscr{W}_B = ([0\vdots I_q] \otimes (A_q'A_q)^{-1}A_q')\mathscr{W}([0\vdots I_q]) \otimes (A_q'A_q)^{-1}A_q')'$.

If $\hat{A}_q$ and $\hat{B}_q$ are root-T consistent estimators of A_q and B_q, $\hat{A}_{q,\perp}$ and $\hat{B}_{q,\perp}$ are also root-T consistent estimators of $A_{q,\perp}$ and $B_{q,\perp}$, respectively. This follows directly from

$$\hat{A}_q'\hat{A}_{q,\perp} \equiv 0 \Leftrightarrow$$
$$A_q'(\hat{A}_{q,\perp} - A_{q,\perp}) + (\hat{A}_q - A_q)'\hat{A}_{q,\perp} \equiv 0,$$

where we have used that $A_q'A_{q,\perp} \equiv 0$. The fact that $\hat{A}_q$ is a root-T consistent estimator of A_q now implies that $\hat{A}_{q,\perp}$ is a root-T consistent estimator of $A_{q,\perp}$. The same result can be obtained for $\hat{B}_{q,\perp}$.

The limiting behavior of $\hat{\lambda} = \mathrm{vec}(\hat{\Lambda})$, with $\hat{\Lambda} = \hat{A}_{q,\perp}'\hat{\Theta}\hat{B}_{q,\perp}'$, then results from Assumption 1 and the root-T consistency of $\hat{A}_{q,\perp}$ and $\hat{B}_{q,\perp}$,

$$\sqrt{T}\hat{\lambda}_q \underset{d}{\rightarrow} \mathrm{N}(0, \Omega_q),$$

with $\Omega_q = (B_{q,\perp} \otimes A_{q,\perp}')\mathscr{W}(B_{q,\perp} \otimes A_{q,\perp}')$. □

Proof of Proposition 1. When $\mathscr{V} = ((F'F)^{-1} \otimes (G'G)^{-1})$,

$$\mathscr{W} = (F(F'F)^{-1}F' \otimes G(G'G)^{-1}G') = I_{km}.$$

As a consequence,

$$\begin{aligned}\Omega_q &= (B_{q,\perp} \otimes A_{q,\perp}')\mathscr{W}(B_{q,\perp}' \otimes A_{q,\perp})' \\ &= (B_{q,\perp} \otimes A_{q,\perp}')(B_{q,\perp}' \otimes A_{q,\perp}) \\ &= I_{(k-q)(m-q)},\end{aligned}$$

and

$$\begin{aligned}\hat{\lambda}'\Omega_q^{-1}\hat{\lambda} &= \mathrm{vec}(\hat{\Lambda})'\mathrm{vec}(\hat{\Lambda}) \\ &= \mathrm{vec}(\hat{A}_{q,\perp}'\hat{\Theta}\hat{B}_{q,\perp}')'\mathrm{vec}(\hat{A}_{q,\perp}'\hat{\Theta}\hat{B}_{q,\perp}') \\ &= \mathrm{tr}(\hat{B}_{q,\perp}\hat{\Theta}'\hat{A}_{q,\perp}\hat{A}_{q,\perp}'\hat{\Theta}\hat{B}_{q,\perp}').\end{aligned}$$

If we write for notational convenience $U = [U_1 \vdots U_2]$ with U_1 a $k \times q$ and U_2 a $k \times (k-q)$ matrix and $V = [V_1 \vdots V_2]$ with V_1 an $m \times q$ and V_2 an $m \times (m-q)$ matrix, the expressions in (12) simplify to $A_{q,\perp} = U_2U_{22}^{-1}(U_{22}U_{22}')^{1/2}$ and $B_{q,\perp} = (V_{22}V_{22}')^{1/2}V_{22}'^{-1}V_2'$. Substituting these values in the above expression and using the

SVD expression of $\hat{\Theta}$ $(= USV')$, we obtain

$$\begin{aligned}\hat{\lambda}'\Omega_q^{-1}\hat{\lambda} &= \mathrm{tr}[(V_{22}V_{22}')V_{22}'^{-1}V_2'\hat{\Theta}'U_2U_{22}^{-1}(U_{22}U_{22}')U_{22}^{-1\prime}U_2'\hat{\Theta}V_2V_{22}^{-1}] \\ &= \mathrm{tr}[V_2'VSU'U_2U_2'USV'V_2] \\ &= \mathrm{tr}[V_2'V_2S_2U_2'U_2S_2V_2'V_2] \\ &= \mathrm{tr}[S_2'S_2] = \mathrm{tr}[S_2S_2'],\end{aligned}$$

where we have used that $U_1'U_2 = 0$, $V_1'V_2 = 0$, $U_2'U_2 = I_{k-q}$, $V_2'V_2 = I_{m-q}$. S_2 is a $(k-q)\times(m-q)$ dimensional rectangular matrix. When $k>m$ it contains the square roots of the $m-q$ smallest eigenvalues of $\hat{\Theta}'\hat{\Theta}$. When $m>q$ it contains the square roots of the $k-q$ smallest eigenvalues of $\hat{\Theta}\hat{\Theta}'$ on its main diagonal. As

$$\begin{aligned}\hat{\Theta}'\hat{\Theta} &= VS'U'USV' \\ &= VS'SV' \\ &= [V_1 \vdots V_2]\begin{pmatrix} S_1'S_1 & 0 \\ 0 & S_2'S_2\end{pmatrix}[V_1 \vdots V_2]', \\ \hat{\Theta}\hat{\Theta}' &= USV'VS'U' \\ &= [U_1 \vdots U_2]\begin{pmatrix} S_1S_1' & 0 \\ 0 & S_2S_2'\end{pmatrix}[U_1 \vdots U_2]',\end{aligned}$$

it follows that $\mathrm{tr}(S_2'S_2)$ is equal to the sum of the smallest $m-q$ eigenvalues of $\hat{\Theta}'\hat{\Theta}$ and $\mathrm{tr}(S_2S_2')$ is equal to the sum of the smallest $k-q$ eigenvalues of $\hat{\Theta}\hat{\Theta}'$. The eigenvalues of $\hat{\Theta}'\hat{\Theta}$ and $\hat{\Theta}\hat{\Theta}'$ correspond with the roots of the polynomial equations $|\mu\hat{F}'\hat{F} - \hat{\Pi}\hat{G}'\hat{G}\hat{\Pi}'| = 0$ and $|\mu\hat{G}'\hat{G} - \hat{\Pi}\hat{F}\hat{F}'\hat{\Pi}'| = 0$, respectively. This shows the relationship with the canonical correlations when $\hat{\Pi}$ is a least squares estimator.

Proof of Theorem 2. When $\Pi = A_qB_q$ and $A_{q,\perp}'B_{q,\perp}'$ is non-singular, the Representation Theorem for the VAR(1) (36), implies that

$$y_i = B_{q,\perp}'(A_{q,\perp}'B_{q,\perp}')^{-1}A_{q,\perp}'\sum_{j=1}^{i}\varepsilon_j + \xi_i,$$

where ξ_i is a stationary error term and $\varepsilon_j = \Delta y_j - A_qB_qy_{j-1}$, see, for example, Johansen (1991). In case of white noise disturbances ε_i with constant covariance matrix Σ,

$$\frac{1}{\sqrt{T\tau}}y_{\tau T} \underset{\mathrm{d}}{\rightarrow} B_{q,\perp}'(A_{q,\perp}'B_{q,\perp}')^{-1}\Sigma_q^{1/2}B_{k-q}(\tau), \quad 0<\tau<1,$$

with $\Sigma_q = A_{q,\perp}'\Sigma A_{q,\perp}$ and $B_{k-q}(\tau)$ is a $(k-q)$-dimensional Brownian motion defined on the unit interval with identity covariance matrix at time τ.

From the first order condition for the least squares estimator, we obtain that

$$\sum_{i=1}^{T}(\Delta y_i - \hat{\Pi} y_{i-1})y'_{i-1} = 0 \Leftrightarrow$$
$$\sum_{i=1}^{T}(\Delta y_i - \hat{A}_q\hat{B}_q y_{i-1} - \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}y_{i-1})y'_{i-1} = 0 \Leftrightarrow$$
$$\sum_{i=1}^{T}(\hat{\varepsilon}_i - \hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}y_{i-1})y'_{i-1} = 0 \Leftrightarrow$$
$$\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1} = \sum_{i=1}^{T}\hat{A}_{q,\perp}\hat{\Lambda}_q\hat{B}_{q,\perp}y_{i-1}y'_{i-1},$$

where $\hat{\varepsilon}_i = \Delta y_i - \hat{A}_q\hat{B}_q y_{i-1}$. Since $I_k = P_{\hat{A}_q} + P_{\hat{A}_{q,\perp}}$, it follows that

$$\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1} = \hat{A}_{q,\perp}\hat{A}'_{q,\perp}\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1},$$

where we use that $\hat{A}'_{q,\perp}\hat{A}_{q,\perp} = I_{k-q}$. We can express $\hat{\Lambda}_q$ as

$$\hat{\Lambda}_q = \hat{A}'_{q,\perp}\left[\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1}\right]\left[\sum_{i=1}^{T}y_{i-1}y'_{i-1}\right]^{-1}\hat{B}'_{q,\perp}.$$

To derive the limiting behavior of the rank statistic, we first consider the limiting behavior of $\sum_{i=1}^{T}y_{i-1}\hat{\varepsilon}'_i$. We specify $\hat{\varepsilon}_i$ as $\Sigma^{1/2}u_i$ with u_i a $k\times 1$ random vector with an identity covariance matrix. When we use that $\hat{A}_{q,\perp}$ is a consistent estimator of $A_{q,\perp}$, we obtain that

$$\frac{1}{T}\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1} \underset{d}{\rightarrow} A_{q,\perp}\Sigma_q^{1/2}\left(\int_0^1 B_{k-q}(t)\,\mathrm{d}B_{k-q}(t)'\right)'\Sigma_q^{1/2}(A'_{q,\perp}B'_{q,\perp})^{-1\prime}B_{q,\perp},$$

where we used that $\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1} = \hat{A}_{q,\perp}\hat{A}'_{q,\perp}\sum_{i=1}^{T}\hat{\varepsilon}_i y'_{i-1}$ and $\Sigma_q = A'_{q,\perp}\Sigma A_{q,\perp}$. The limiting behavior of the second term in the expression of $\hat{\Lambda}_q$, $\frac{1}{T^2}\sum_{i=1}^{T}y_{i-1}y'_{i-1}$, is given by

$$\frac{1}{T^2}\sum_{i=1}^{T}y_{i-1}y'_{i-1} \underset{d}{\rightarrow} B'_{q,\perp}(A'_{q,\perp}B'_{q,\perp})^{-1}\Sigma_q^{1/2}\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)$$
$$\times\Sigma_q^{1/2\prime}(A'_{q,\perp}B'_{q,\perp})^{-1\prime}B_{q,\perp},$$

and

$$\left(\frac{1}{T^2}\sum_{i=1}^{T}y_{i-1}y'_{i-1}\right)^{-1} \underset{d}{\rightarrow} A_{q,\perp}\Sigma_q^{-1/2}\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)^{-1}\Sigma_q^{-1/2\prime}A'_{q,\perp}.$$

The limiting behavior of $\hat{\Lambda}_q$ is therefore

$$T\hat{\Lambda}_q \underset{d}{\rightarrow} \Sigma_q^{1/2}\left(\int_0^1 B_{k-q}(t)\,\mathrm{d}B_{k-q}(t)'\right)'\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)^{-1}\Sigma_q^{-1/2\prime}(B_{q,\perp}A_{q,\perp})'.$$

To estimate the covariance matrix of $\hat{\pi} = \mathrm{vec}(\hat{\Pi})$, we use

$$\hat{\mathscr{W}} = \left(\left(\frac{1}{T}\sum_{i=1}^{T} y_{i-1}y_{i-1}'\right)^{-1} \otimes \frac{1}{T}\sum_{i=1}^{T}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right),$$

where $\tilde{\varepsilon}_i = \Delta y_i - \hat{\Pi}y_{i-1}$ and hence the estimator of the covariance matrix of $\hat{\lambda}_q = \mathrm{vec}(\hat{\Lambda}_q)$ is

$$\hat{\Omega}_q = (\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}')\hat{\mathscr{W}}(\hat{B}_{q,\perp} \otimes \hat{A}_{q,\perp}')'.$$

The limiting behavior of $\hat{\Omega}_q$ corresponds with

$$\frac{1}{T}\hat{\Omega}_q \underset{d}{\rightarrow} \left(B_{q,\perp}A_{q,\perp}\Sigma_q^{-1/2}\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)^{-1}\Sigma_q^{-1/2\prime}A_{q,\perp}'B_{q,\perp}' \otimes \Sigma_q\right),$$

and of its inverse

$$\left(\frac{1}{T}\hat{\Omega}_q\right)^{-1} \underset{d}{\rightarrow} \left((A_{q,\perp}'B_{q,\perp}')^{-1}\Sigma_q^{1/2}\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)\Sigma_q^{1/2}(A_{q,\perp}'B_{q,\perp}')^{-1\prime} \otimes \Sigma_q^{-1}\right).$$

From this we obtain the limiting behavior of the rank statistic $\mathrm{rk}(q) = \frac{1}{T}\hat{\lambda}_q'\hat{\Omega}_q^{-1}\hat{\lambda}_q$,

$$\mathrm{rk}(q) \underset{d}{\rightarrow} \mathrm{tr}\left[\left(\int_0^1 B_{k-q}(t)\,\mathrm{d}B_{k-q}(t)'\right)'\left(\int_0^1 B_{k-q}(t)B_{k-q}(t)'\,\mathrm{d}t\right)^{-1} \times \left(\int_0^1 B_{k-q}(t)\,\mathrm{d}B_{k-q}(t)'\right)\right].$$

Data generating process. The data generating process in the simulation experiment is given by

$$r_t = \Pi f_t + \varepsilon_t, \text{ with } \varepsilon_t = v_t + \Gamma v_{t-1},$$

where $\Pi = \alpha\beta'$ with

$$\alpha = (0.0813 \;\; -0.0271 \;\; -0.6203 \;\; -0.0460)',$$

$$\beta' = (-0.3477 \;\; -0.1277 \;\; -0.3838 \;\; -0.5312 \;\; -0.2728 \;\; -0.3527 \;\; -0.2188 \;\; -0.2934 \;\; -0.2035 \;\; -0.3427),$$

$$\Gamma = \begin{pmatrix} 0.0312 & 0.0255 & -0.0185 & 0.0591 & 0.0389 \\ 0.0346 & -0.0166 & -0.0608 & 0.0743 & 0.0794 \\ -0.0304 & 0.0624 & -0.1347 & 0.1054 & -0.0369 \\ -0.0414 & 0.0951 & -0.0029 & -0.0497 & -0.0586 \\ -0.0570 & -0.0845 & 0.0606 & -0.0143 & -0.1971 \\ -0.0649 & -0.0738 & 0.0030 & 0.0335 & 0.0346 \\ -0.0334 & -0.1163 & -0.0139 & -0.0218 & -0.0390 \\ -0.1029 & 0.0368 & 0.0737 & -0.0005 & -0.1686 \\ -0.1153 & 0.0008 & 0.0373 & 0.0185 & -0.0927 \\ -0.0737 & -0.0669 & 0.0500 & 0.1466 & -0.1359 \end{pmatrix.$$

$$\begin{matrix} 0.0953 & -0.1515 & 0.2286 & -0.0806 & -0.1659 \\ -0.0043 & -0.2194 & 0.2959 & -0.0043 & 0.0016 \\ -0.0187 & -0.0989 & 0.3571 & 0.0133 & -0.1731 \\ 0.0910 & -0.0903 & 0.1850 & 0.0616 & -0.0865 \\ 0.0528 & 0.0403 & 0.1935 & -0.0114 & 0.1141 \\ -0.0432 & -0.0787 & 0.2199 & -0.0266 & -0.0013 \\ 0.0128 & -0.0645 & 0.1299 & 0.1105 & 0.0097 \\ 0.0254 & 0.0184 & 0.0966 & -0.0176 & 0.0596 \\ 0.1029 & 0.0546 & 0.0529 & -0.1792 & 0.0798 \\ 0.0617 & 0.1090 & 0.0402 & -0.0659 & -0.0440 \end{matrix} \right)$$

and the disturbances $v_t \sim \mathrm{N}(0, \Sigma)$ with

$$\Sigma = \frac{1}{100} \begin{pmatrix} 0.19 & 0.09 & 0.07 & 0.05 & 0.04 & 0.03 & 0.02 & -0.01 & -0.00 & -0.01 \\ 0.09 & 0.11 & 0.06 & 0.05 & 0.04 & 0.04 & 0.03 & 0.01 & 0.02 & 0.01 \\ 0.07 & 0.06 & 0.10 & 0.05 & 0.04 & 0.04 & 0.03 & 0.03 & 0.02 & 0.01 \\ 0.05 & 0.05 & 0.05 & 0.08 & 0.04 & 0.04 & 0.04 & 0.03 & 0.02 & 0.01 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.08 & 0.05 & 0.05 & 0.05 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.04 & 0.05 & 0.08 & 0.06 & 0.05 & 0.05 & 0.03 \\ 0.02 & 0.03 & 0.03 & 0.04 & 0.05 & 0.06 & 0.08 & 0.06 & 0.05 & 0.03 \\ -0.01 & 0.01 & 0.03 & 0.03 & 0.05 & 0.05 & 0.06 & 0.10 & 0.07 & 0.05 \\ -0.00 & 0.02 & 0.02 & 0.02 & 0.04 & 0.05 & 0.05 & 0.07 & 0.09 & 0.04 \\ -0.01 & 0.01 & 0.01 & 0.01 & 0.03 & 0.03 & 0.03 & 0.05 & 0.04 & 0.07 \end{pmatrix}.$$

References

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Andrews, D.W.K., 1987. Asymptotic results for generalized Wald tests. Econometric Theory 3, 345–358.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Campbell, J.Y., Lo, A.L., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Cragg, J.G., Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. Econometric Theory 9, 222–240.

Cragg, J.C., Donald, S.G., 1996. On the asymptotic properties of LDU-based tests of the rank of a matrix. Journal of the American Statistical Association 91, 1301–1309.

Cragg, J.C., Donald, S.G., 1997. Inferring the rank of a matrix. Journal of Econometrics 76, 223–250.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Gill, L., Lewbel, A., 1992. Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA Models. Journal of the American Statistical Association 87, 766–776.

Golub, G.H., van Loan, C.F., 1989. Matrix Computations. The Johns Hopkins University Press, Baltimore.

Hansen, L.P., 1982. Large sample properties of generalized method moments estimators. Econometrica 50, 1029–1054.

Horn, R.A., Johnson, C.A., 1991. Topics in Matrix Analysis. Cambridge University Press, New York.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Jagannathan, R., Skoulakis, G., Wang, Z., 2002. Generalized method of moments: applications in finance. Journal of Business & Economic Statistics 20, 470–481.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Johansen, S., 1995. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

Kleibergen, F., van Dijk, H.K., 1994. Direct cointegration testing in error correction models. Journal of Econometrics 63, 61–103.

Kleibergen, F., van Dijk, H.K., 1998. Bayesian simultaneous equation analysis using reduced rank structures. Econometric Theory 14, 701–743.

Kleibergen, F., Paap, R., 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. Journal of Econometrics 111, 223–249.

Lewbel, A., 1991. The rank of demand systems: theory and nonparametric estimation. Econometrica 59, 711–730.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. Elsevier Science, North-Holland, pp. 2113–2148.

Newey, W.K., West, K.D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Ratsimalahelo, Z., 2001. Specification of VARMA echelon form models. In: Neck, R. (Ed.), Proceedings IFAC Symposium on Modelling and Control of Economic Systems. Elsevier, North-Holland, pp. 186–191.

Ratsimalahelo, Z., 2002. Rank test based on matrix perturbation theory. Unpublished working paper, U.F.R. Science Economique, University de Franche-Comté.

Reinsel, G.C., Velu, R.P., 1998. Multivariate Reduced Rank Regression: Theory and Applications. Springer, New York.

Robin, J.-M., Smith, R.J., 2000. Tests of rank. Econometric Theory 16, 151–175.

Saikkonen, P., 1999. Testing normalization and overidentification of cointegrating vectors in vector autoregressive processes. Econometric Reviews 18, 235–257.

West, K.D., 1997. Another heteroscedasticity- and autocorrelation-consistent covariance matrix estimator. Journal of Econometrics 76, 171–191.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–838.

Wright, J.H., 2003. Detecting lack of identification in GMM. Econometric Theory 19, 322–330.