EERI

Economics and Econometrics Research Institute

Strongly Consistent Determination of the Rank of Matrix.

Zaka Ratsimalahelo

EERI Research Paper Series No 4/2003

EERI
Economics and Econometrics Research Institute
Avenue de Beaulieu
1160 Brussels
Belgium

Tel: +322 299 3523
Fax: +322 299 3523
www.eeri.eu

Copyright © 2003 by Zaka Ratsimalahelo

Strongly Consistent Determination of the Rank of Matrix.

Zaka Ratsimalahelo
University of Franche-Comté, U.F.R. Science Economique, 45D, Av., de l'Observatoire - 25030 Besançon - France, E-mail: Zaka.Ratsimalahelo@univ-fcomte.fr.

First version: April 15, 2003.
This version: June 21, 2003.

Abstract

In this paper, we develop methods of the determination of the rank of random matrix. Using the matrix perturbation theory to construct or find a suitable bases of the kernel (null space) of the matrix and to determine the limiting distribution of the estimator of the smallest singular values. We propose a new rank test for an unobserved matrix for which a root-N-consistent estimator is available and construct a Wald-type test statistic (generalized Wald test). The test, based on matrix perturbation theory, enable to determine how many singular values of the estimated matrix are insignificantly different from zero and we fully characterise the asymptotic distribution of the generalized Wald statistic under the most general conditions. We show that it is chi-square distribution under the null. In particular case, when the asymptotic covariance matrix has a Kronecker product form, the test statistic is equivalent to likelihood ratio test statistic and to Multiplier Lagrange test statistic. Two approaches to be considered are sequential testing strategy and information theoretic criterion. We establish a strongly consistent of the determination of the rank of matrix using the two approaches.

Key words: Rank Testing; Matrix Perturbation Theory; Rank Estimation; Subspace Methods; Singular Value Decomposition; Weighting Matrices; Sequential Testing Strategy; Information Theoretic Criterion.

JEL classification: C12, C13, C30.

1 Introduction

The main purpose of this paper is to propose some new tests for determining the rank of random matrix. The rank of a matrix is of interest in a number of applications in econometrics; for instance, the classical identification problem in linear simultaneous-equations models involves the rank of a particular submatrix of the reduced-form parameters and in a likelihood setting, the rank of the information matrix relates to the identifiability of a vector of parameters [Hsiao (1986), Anderson and Kunitomo (1992)]. Lewbel (1991) and Lewbel and Perraudin (1995) have shown that several results in consumer theory can depend on the rank of certain estimable matrices. In principal-component and factor models, the number of factors or components in the model equals the rank of covariance matrix, [Lawley and Maxwell (1971)]. Also, in ARMA models, the maximum order of the AR and MA processes equals the rank of a Hankel matrix of autocovariances and, following the Kronecker theory, the rank of the Hankel matrix equals the number of unobserved state variables in the state-space representation of the time-series generating process, see Kailath (1980); Aoki and Havenner (1991), Ratsimalahelo and Lardies (1998).

Determining the rank of a matrix is a difficult task made more so if the matrix is contaminated with errors, which is always the case in econometrics and statistical applications based on estimated matrices. Gill and Lewbel (1992), Cragg and Donald (1996) used a rank test based on the Gaussian elimination Lower-Diagonal-Upper triangular (LDU) decomposition. Their test has the advantage of possessing a limiting chi-squared distribution. Unfortunately the Gaussian elimination test tends to be conservative with type 1 error close to 0 when the sample size is small (see Ratsimalahelo (2002)). Also, Cragg and Donald (1997) propose another test for the rank of matrix based on a minimum chi-squared criterion. The procedure need to minimize the objective function numerically which is often very difficult (see Ferguson (1996)).

However it is well known that the rank of the matrix is equal to its number of the non zero eigenvalues. Thus a formal test of rank can be expressed as a test of the number of zero eigenvalues of the matrix. For this we shall construct a test of rank based on the smallest eigenvalues of the estimated matrix. One important property of the eigenvalues is that they are sufficient statistics invariant with respect to the multiplication of the matrix from the left and right by any nonsingular matrices. But the test is more complicated when the smallest eigenvalues has multiplicity. The asymptotic null distribution of Bartlett's statistic is not chi-squared [see Schoot (1988)]. He used an approximation based on some of its moments to obtain an asymptotically chi-squared [see also Lawley (1956) and Anderson (1963)]. More recently Anderson and Kunitomo (1994), Robin and Smith (2000) used a criterion which is N times a smooth function of the smallest eigenvalues as the test statistics

$$CRT = N \sum_{j=k+1}^{n} f(\widehat{\lambda}^2{}_j)$$

namely characteristic roots test (CRT) statistic, where N is the sample size and the λ_j^2 are the eigenvalues. The univariate function $f(.)$ is require to be smooth but otherwise arbitrary. This class of statistics includes many test statistics as special cases including the likelihood ratio statistic, the Lagrange Multiplier statistic and the Wald statistic, [see Anderson (1984), Anderson and Kunitomo (1994)]. Follows, Robin and Smith (2000) the CRT statistic is distributed asymptotically as a linear combination of independent chi-squared random variables rather than a chi-square. The disadvantages of this statistic are that the weights are unknown and must be estimated from the sample, their estimation introduces variability and hence potentially less accuracy, to the testing procedure. The test requires the percentiles of a weighted chi-squared distribution for which computationally intensive algorithms need be used. There is substantial literature to assist in computing the tail probabilities of linear combination of chi-squared random variables, (see Field (1993) for an introduction).

In this paper we shall consider the singular value decomposition of a matrix which allows us to use the orthogonal reduction of the matrix. The smallest singular value of a matrix can be seen as its distance to singularity. We shall use the matrix perturbation theory to construct or find a suitable bases of the kernel (null space) of the matrix and to determine the limiting distribution of the estimator of the smallest singular values [see also Eaton and Tyler (1994)]. For an unobserved matrix for which a $root-N$ consistent estimator is available, the result of the matrix perturbation show us that the smallest singular values converge asymptotically to zero in the order $O_p(N^{-1})$

and the corresponding left and right singular vectors converge asymptotically in the order $O_p(N^{-1/2})$. So we shall give a rank test for an unobserved matrix for which a $root - N$ consistent estimator is available and construct a Wald-type tests statistic or in terms more appropriate generalized Wald tests namely $L(k)$. The test, based on matrix perturbation theory, allows us to determine how many singular values of the estimated matrix are insignificantly different from zero and we shall show also that the test statistics is asymptotically distributed as chi-square under the null.

Ratsimalahelo (2000) (2002) has been shown that the $L(k)$ test statistics is more appropriate as direct tests to determine the relevant instrumental variables using the canonical correlations. Moreover, the performance of the $L(k)$ test statistics is similar to the statistic based on the Gaussian elimination decomposition (LDU) by Gill and Lewbel (1992), Cragg and Donald (1996). This paper extends the results of Ratsimalahelo (2000) (2002) to the case of weighting matrices and analysis the effects of the weighting matrices on the left and right singular vectors. We will completely characterise the asymptotic distribution of the $L(k)$ test statistics under general conditions. By general conditions we mean that the estimator's asymptotic covariance matrix may be singular or more general setting unknown rank. In particular case, when the asymptotic covariance matrix has a Kronecker product structure, the test statistics is asymptotically equivalent to likelihood ratio test statistic and to Lagrange Multiplier test statistic.

It is well known that a sequential testing strategy does not lead to a consistent estimate of the true rank matrix unless some adjustment is made to the significance level, Robin and Smith (2000), Cragg and Donald (1997) used the results of Potscher (1983) to establish the weakly consistent of the sequential testing strategy. A more general result is presented in this paper, we propose an appropriate significance level to obtain a strongly consistent determination of the rank of matrix using the sequential testing procedure. We shall present also an alternative approach to the information theoretic criterion.

The remainder of the paper is structured as follows: In section 2, we present the basic framework, the relevant material from hypothesis testing and the matrix-perturbation results. In section 3, we derive the asymptotic distribution of the smallest singular value. Based on this result a new rank test is developed and we examine its properties. In section 4, we generalise the previous results to the case of weighting matrices. We shall show that the test statistics is asymptotically equivalent to Likelihood ratio (LR) test and to Lagrange Multiplier (LM) test when the covariance matrix has a Kronecker structure. In sections 5, and 6 we show successively the strongly consistent determination of the rank of matrix using the sequential testing procedure and the information theoretic criterion. In section 7, we compare the $L(k)$ statistic with the CRT statistic of Anderson and Kunitomo (1994), Robin and Smith (2000). Section 8 offers some concluding remarks. Proofs of the fundamental theorems, and propositions that provide the foundation of the technique are assembled in the Appendix.

The following terminology and notation is used throughout the paper: $vec(A)$ stands for the vectorization of the $m \times n$ matrix A. The trace and the rank of the matrix A are denoted by $\mathrm{tr}(A)$ and $\mathrm{r}(A)$. For a singular matrix C, C^+ denotes its Moore-Penrose generalized inverse. Convergence in probability is denoted "$\longrightarrow^p$" and convergence in distribution by "$\longrightarrow^d$". For any matrix A, the linear space spanned by the columns (range) of A is noted by $\mathcal{R}(A)$ and by $\mathcal{N}(A)$ the null space (kernel) of A; the linear space spanned by the rows of A is noted by $\mathcal{R}(A')$ and the kernel of A' by $\mathcal{N}(A')$.

2 Basic Framework

2.1 Hypothesis testing

Consider an unobserved matrix A $(m \times n)$ with unknown true rank $k > 0$, without loss of generality, we assume $m \geq n$.

Assumption 1: Let $\widehat{A}$ be a $root - N$ consistent estimator of A, such that

$$N^{1/2} vec(\widehat{A} - A) \longrightarrow^d \mathcal{N}(0, \Sigma). \tag{1}$$

where the $mn \times mn$ covariance matrix Σ is non zero but possibly singular.

We wish to construct a test for the rank k of A, $r(A)$. Thus, we wish to test the null hypothesis

$$H_0 : r(A) = k \tag{2}$$

against the alternative

$$H_1 : r(A) > k.$$

which are also the hypotheses considered by Gill and Lewbel (1992), Gragg and Donald (1996), Gragg and Donald (1997) and Robin and Smith (2000).

The true rank of A is unknown but, with probability one, the rank of $\widehat{A}$, a consistent estimator of A has full rank. Thus, the null hypothesis based on a test using $\widehat{A}$ is satisfied only asymptotically in N. Therefore, interest attaches to a statistical test enabling us to determine the rank of A, given the estimator $\widehat{A}$, from

$$H_0 : r[\widehat{A}] = k^*. \tag{3}$$

A well-conditioned means for evaluating the rank is to use the singular value decomposition (SVD) and count the number of nonzero singular values. Let the SVD of the $m \times n$ $(m \geq n)$ real matrix A and with a rank k be denoted [Golub and Van Loan (1996)].

$$A = UDV' = [U_1, U_2] \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \tag{4.1}$$

where $U = [U_1, U_2]$ of order $(m \times m)$ and $V = [V_1, V_2]$ of order $(n \times n)$ are orthogonal matrices and $D = diag(D_1, D_2)$ is an $m \times n$ rectangular diagonal matrix with decreasing non-negative diagonal elements λ_i called the singular values.

In fact $D_2 = O$ of order $(m-k) \times (n-k)$ zero matrix and $D_1 = diag(\lambda_1, \lambda_2, ..., \lambda_k)$ is order k with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k > 0$ the non-zero singular values of A. The number of positive singular values is the rank of A that is k. Thus the SVD of A can also be written as

$$A = U_1 D_1 V_1' \tag{4.2}$$

The k columns of U_1 are the left singular vectors corresponding to the non-zero singular values D_1, they span the range (column) space of $A : \mathcal{R}(A)$, the $m-k$ columns of U_2 are the left singular vectors corresponding to the zero singular values D_2 and span the null space (kernel) of $A' : \mathcal{N}(A')$. Similarly, the k columns of V_1 are the right singular vectors corresponding to the non-zero singular values and span the row space of A (or the column space of A') : $\mathcal{R}(A')$, the $n-k$ columns of V_2 are the right singular vectors corresponding to the zero singular values and span the null space (kernel) of $A : \mathcal{N}(A)$.

Because the orthogonality of U and V, the matrix D may be written as

$$D = \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} U_1'AV_1 & U_1'AV_2 \\ U_2'AV_1 & U_2'AV_2 \end{bmatrix} \tag{5}$$

Thus

$$D_1 = U_1'AV_1 = diag(\lambda_1, \lambda_2, ..., \lambda_k) \tag{6.1}$$

$$D_2 = U_2'AV_2 = O_{m-k,n-k} \tag{6.2}$$

and the off diagonal terms $U_1'AV_2 = O_{k,n-k}$ and $U_2'AV_1 = O_{m-k,k}$ are satisfied because $AV_2 = 0$ (the columns of V_2 span the null space of A) and $U_2'A = (A'U_2)' = 0$ (the columns of U_2 span the null space of A').

Thus under H_0,the $(m-k)(n-k)$ diagonal matrix D_2 is identical to zero. The hypothesis of rank condition H_0 is equivalent to the hypothesis on the smallest singular values

$$H_0^{ssv} : D_2 = 0. \tag{7}$$

According (6.2) the existence of a matrix V_2 (or equivalently U_2) such that $D_2 = 0$ is equivalent to the hypothesis of rank condition H_0. This testing problem is also mathematically equivalent to the hypothesis for the rank test. This testing problem may be considered as the test of the hypothesis: "Are the columns of V_2 span the null space of A (equivalently are the columns of U_2 span the orthogonal complement of the range of A)"?

A) Subspaces

We shall first introduces the notions of subspace, column space, and rank of matrix. As the $(m \times n)$ matrix A have a k-dimensional range or column space ($k = dim\mathcal{R}(A)$) which is a subspace of the $m-$dimensional Euclidean space R^m. The rank of the matrix is the dimension of this subspace that is $r(A) = k = dim\mathcal{R}(A)$. Following the properties of the SVD of matrices [Golub and Van Loan (1996)], for any vector $x \in R^m$, it can be decomposed into mutually orthogonal vectors x_1 and x_2 in the spaces spanned by the columns of U_1 and U_2, respectively. These two spaces are respectively $k-$dimensional and $(m-k)-$dimensional orthogonal subspaces in R^m, and their direct sum is equal to R^m. Therefore, the orthogonal complement in R^m of the column space of A is spanned by the columns of the $m \times (m-k)$ orthonormal matrix U_2. Thus the null space (kernel) of A' denoted by $\mathcal{N}(A')$ is the orthogonal complement of $\mathcal{R}(A)$ in R^m and vice versa, hence U_2 is the orthogonal complement of U_1. Moreover the rank of A is also equal to $k = m - \dim \mathcal{N}(A')$. We obtain the following important property

$$U_2'A = 0 \Longleftrightarrow \mathcal{R}(A) = \mathcal{R}(U_1) \tag{8.1}$$

The relation (8.1) means that the subspace spanned by the columns of A is equal to the subspace spanned by the columns of U_1 which implies that the subspace spanned by the columns of U_2 is equal to the orthogonal complement of the column space of $A : \mathcal{R}(U_2) = \mathcal{N}(A')$ and vice versa.

The same holds for Euclidean space R^n of which the row space of A (or the column space of A') is a $k-$dimensional subspace and $dim\mathcal{R}(A') = dim\mathcal{R}(A) = r(A)$. The two spaces spanned by the columns of V_1 and V_2 are respectively $k-$dimensional and $(n-k)-$dimensional orthogonal subspaces in R^n and their direct sum is equal to R^n. The orthogonal complement in R^n of the row space of A is spanned by the columns of the $n \times (n-k)$ orthonormal matrix V_2. The null space (kernel) of $A : \mathcal{N}(A)$ is the orthogonal complement of $\mathcal{R}(A')$ in R^n, hence V_2 is the orthogonal complement of V_1. Therefore $r(A) = k = n - \dim \mathcal{N}(A)$. We have also the following property

$$AV_2 = 0 \Longleftrightarrow \mathcal{R}(A') = \mathcal{R}(V_1) \tag{8.2}$$

The properties (8.1) and (8.2) of the SVD of A lead to test the rank of A. The hypothesis for the rank test is also equivalent to the hypothesis for the tests on the kernel, or on the range of a matrix [see also Gourieroux, Monfort and Renault (1993)].

B) Test on the kernel of A.

The subspace spanned by the rows of A denoted by $\mathcal{R}(A')$ is called the right principal subspace and its dimension is equal to the rank of A. According (8.2) A' and V_1 matrix of the principal right singular vectors of A span the same subspace. Hence inference on the dimension of $\mathcal{R}(A')$ is equivalent to inference on the dimension of $\mathcal{R}(V_1)$ (the right principal subspace of A), so an implicit form of the null hypothesis is given by:

$$H_0^{\ker} : AV_2 = 0. \tag{9.1}$$

where V_2 is a matrix of the right singular vectors corresponding to the zero singular values and span the null space (kernel) of A. The columns of V_2 are orthonormal so that $V_2'V_2 = I_{n-k}$. In other words V_2 is an orthonormal basis of $\mathcal{N}(A)$. This linear constraint (9.1) is often used in literature to determine the rank of a matrix without reference to the SVD [see e.g. Gourieroux, Monfort and Renault (1993), Gragg and Donald (1997)]. We see that the SVD gives an interpretation of the constraint in terms on the orthonormal basis for the kernel of A.

The null hypothesis may be put in a explicit form, since the $(n \times k)$ matrix V_1 has a full column rank and it is the orthogonal complement of V_2 then there exist a $(m \times k)$ matrix F has a rank k such that $A = FV_1'$. Hence A is decomposed from a product of rank-k matrices. Then the null hypothesis can be written in explicit form

$$H_0^{\text{ker}} : A = FV_1'. \tag{9.2}$$

This parameterization of A corresponds to (4) or (6.1) with $F = U_1 D_1$ where U_1 is a matrix of the principal left singular vectors.

C) Test on the range of A.

Similarly, the left principal subspace is given by $\mathcal{R}(A)$ the subspace spanned by the range of A. Its dimension is also equal to the rank of A, then inference on the dimension of $\mathcal{R}(A)$ or equivalently on the dimension of $\mathcal{R}(U_1)$ (the left principal subspace of A) is based by Eq. (8.1) which constitutes an implicit form of the null hypothesis:

$$H_0^{ran} : U_2'A = 0. \tag{10.1}$$

where U_2 is a matrix of the left singular vectors corresponding to the zero singular values and span the orthogonal complement of the range of A. The columns of U_2 are orthonormal that is $U_2'U_2 = I_{m-k}$. (U_2 is an orthonormal basis of $\mathcal{N}(A')$). Since U_1 is the orthogonal complement of U_2 then there exist a $(n \times k)$ matrix G has a rank k such that $A = U_1 G'$. So the null hypothesis can be written in explicit form

$$H_0^{ran} : A = U_1 G'. \tag{10.2}$$

It follows (4) or (6.1) that $G' = D_1 V_1'$.

Follows the both implicit forms of the null hypothesis of the kernel (9.1) and the range (10.1), the rank of A is equal to the dimension of $\mathcal{R}(V_1)$ and

$\mathcal{R}(U_1)$ respectively. Orthonormal right and left bases for the null-spaces of A are V_2 and U_2 respectively.

The both explicit forms of the null hypothesis of the kernel (9.2) and the range (10.2) of A are the SVD of the product of rank k matrices. One such parameterization of the reduced rank of the matrix is commonly used in the literature as $A = \alpha\beta'$ where α and β are full column rank of dimension $(m \times k)$ and $(n \times k)$, respectively. Thus, we may write the null hypothesis (reduced rank) as in explicit form

$$H_0^{rr} : A = \alpha\beta',$$

where α, β have respective a dimension $(m \times k)$, $(n \times k)$ and a rank equal to k.

Now we can of course identify α and β in terms of principal singular vectors and singular values. It follows (4) or (6.1) that in more general

$$\alpha\beta' = (U_1 D_1^{\theta})(D_1^{1-\theta} V_1')$$

for some specified $0 \leq \theta \leq 1$. We see that the SVD of matrix is a natural normalisation to determine the unknown matrices. For $\theta = 0$ then $\alpha = U_{1,}$ matrix of the principal left singular vectors and $\beta' = D_1 V_1'$. For $\theta = 1$ then $\beta' = V_1'$ matrix of the principal right singular vectors and $\alpha = U_1 D_1$. And for $\theta = 1/2$ the scales of the left and right principal singular vectors are identical that is $\alpha = U_1 D_1^{1/2}$ and $\beta' = D_1^{1/2} V_1'$.

The test of the rank of the matrix A can be testing of one these null hypotheses $(H_0, H_0^{ssv}, H_0^{\ker}, H_0^{ran})$. For the tests on the kernel and range of matrix based either on the implicit or explicit forms of the hypotheses, we must estimate the left and right singular vectors which are U_1, U_2, V_1, and V_2. Note that the SVD of a real matrix is not unique, hence the matrices U and V are not uniquely defined. However, the singular values are uniquely determined, thus the matrix D is uniquely defined. In this paper, we are interesting to test the null hypothesis on the smallest singular values H_0^{ssv}.

The test the null hypothesis H_0^{ssv} is based on estimating small singular values and the associated right and left null vectors respectively.

2.2 Matrix Perturbation Results

Matrix perturbation analysis is concerned with the sensitivity of the eigenelements of a matrix to perturbations in its components. In this section, we will introduce the matrix perturbation results used to define a detection scheme that allows estimation of the number of zero singular values of A.

There is a well-developed mathematical theory on such matrix perturbations as presented by Kato (1982), Golub and Van Loan (1996), Stewart and Sun (1990). But this theory cannot easily be rewritten in a form suitable for statistical inference. One is then interested in identifying the leading first-order terms determining asymptotic distributions, as well as in establishing easily interpretable bounds on second-order terms.

Following the singular value decomposition of a matrix A, this rank is equal the number of nonzero singular values. If only an estimator $\widehat{A}$ of A is available, and the estimation error is small, one would still hope to be able to recognize the 'zero' singular values. Indeed, the smaller the estimation error, the easier such a decision would be. Matrix perturbation analysis formalizes and confirms this intuition.

Let us write the perturbed matrix

$$\widehat{A} = A + \varepsilon B \tag{11}$$

where the matrix εB is the perturbation of the matrix A. While the $m \times n$ matrix A is assumed to have rank $k \leq n$, the perturbation matrix B is assumed to have small values but is of full rank. Then, using (1) and following the central limit theorem, the perturbation matrix can be seen to be Gaussian with elements having zero mean and variance of order N^{-1}. This result implies that the dominant term in the matrix εB is of order $N^{-1/2}$, denoted

$$\varepsilon B = O_p(N^{-1/2}). \tag{12}$$

where $O_p(N^{-a})$ $(for\ \alpha \geq 0)$ thus denotes a term whose elements have a standard deviation of the order of N^{-a}.

We will consider the SVD of $\widehat{A}$ by

$$\widehat{A} = \widehat{U}\widehat{D}\widehat{V'} = \begin{bmatrix} \widehat{U}_1 & \widehat{U}_2 \end{bmatrix} \begin{bmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_2 \end{bmatrix} \begin{bmatrix} \widehat{V'}_1 \\ \widehat{V'}_2 \end{bmatrix} \tag{13}$$

We partition the matrices conformably to the partitioning of U, V and D that is $\widehat{U}_1$ has k columns and $\widehat{U}_2$ has $m-k$ columns.

The estimated matrix $\widehat{A}$ almost certainly has full rank, although it will be " close " to a matrix of rank k. Moreover, the spaces spanned by $\widehat{U}_1$, $\widehat{U}_2$,

$\widehat{V}_1$, and $\widehat{V}_2$ are approximations to the subspaces corresponding U_1, U_2, V_1, and V_2. The matrix diagonal can be written as

$$\widehat{D} = \begin{bmatrix} \widehat{U}_1' \\ \widehat{U}_2' \end{bmatrix} \widehat{A} \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix} = \begin{bmatrix} \widehat{U'}_1 \widehat{A} \widehat{V}_1 & \widehat{U'}_1 \widehat{A} \widehat{V}_2 \\ \widehat{U'}_2 \widehat{A} \widehat{V}_1 & \widehat{U'}_2 \widehat{A} \widehat{V}_2 \end{bmatrix} \tag{14}$$

which satisfies the equations:
$\widehat{D}_1 = \widehat{U'}_1 \widehat{A} \widehat{V}_1 = diag(\hat{\lambda}_1, ..., \hat{\lambda}_k)$ such that $\hat{\lambda}_i \geq \hat{\lambda}_{i+1}$.
$\widehat{D}_2 = \widehat{U'}_2 \widehat{A} \widehat{V}_2 = diag(\hat{\lambda}_{k+1,}..., \hat{\lambda}_n)$ the smallest singular values estimated.
$\widehat{U'}_2 \widehat{A} \widehat{V}_1 = 0 = \widehat{U'}_1 \widehat{A} \widehat{V}_2$.

To study the stochastic behaviour of the smallest singular values estimated of the matrix $\widehat{A}$, we must define ε. The perturbation matrix εB has zero mean and converges to zero (with probability one) as N increases. Each of its components multiplied by $N^{1/2}$ follows a central limit theorem and is asymptotically Gaussian. Following (12), the difference between corresponding components of $\widehat{A}$ and A is order $N^{-1/2}$ and one must take ε as being of this order of magnitude. We can write

$$\widehat{A} = A + (\widehat{A} - A) = A + \varepsilon[(\widehat{A} - A)/\varepsilon] = A + \varepsilon B$$

where $\varepsilon = N^{-1/2}$. Then for large N, $0 < \varepsilon < 1$ and the elements of B's are bounded almost surely because of (1).

We use the matrix perturbation theory to construct or find a suitable bases of the kernel (null space) of the matrix and to determine the limiting distribution of the estimator of the smallest singular values.

This perturbation of A induces corresponding perturbations to the singular values $\{\lambda_i,\ i = 1, .., n\}$ and singular vectors $\{U_i$, and $V_i\ i = 1, .., n\}$ of A. We need to know more about the stochastic behaviour of the estimated singular vectors corresponding to the $n - k$ smallest singular values. To this end we would state the following proposition:

Proposition 1 . *Let $\widehat{A} = A + \varepsilon B$ a perturbed matrix with $\varepsilon B = O_p(N^{-1/2})$. Let $\widehat{U}_2$ and $\widehat{V}_2$ be the estimator of the left and right singular vectors associated with the $n - k$ smallest singular values estimated $\widehat{D}_2$ of $\widehat{A}$, where $\widehat{D}_2 = diag(\hat{\lambda}_{k+1,}..., \hat{\lambda}_n)$. Then for sufficiently large N, there are bases denoted $\widetilde{U}_2$ (respectively $\widetilde{V}_2$) of null-spaces of AA' (respectively $A'A$) for which the following relations hold:*

$$\widehat{U}_2 = \widetilde{U}_2 + O_p(N^{-1/2}) \tag{15.1}$$

$$\widehat{V}_2 = \widetilde{V}_2 + O_p(N^{-1/2}) \tag{15.2}$$

$$\widehat{D}_2 = \varepsilon \widetilde{U'}_2 B \widetilde{V}_2 + O_p(N^{-1}) \tag{16}$$

where $\widetilde{U}_2 = U_2 M$ and $\widetilde{V}_2 = V_2 P$, with $M : (m-k) \times (m-k)$ and $P : (n-k) \times (n-k)$ are orthogonal matrices respectively..■

Eq. (15) shows that given a perturbed matrix $\widehat{A}$ and thus $\widehat{U}_2$ and $\widehat{V}_2$, there exist perturbation-dependent exact bases $\widetilde{U}_2$ and $\widetilde{V}_2$ of the null-spaces (kernel) of A that are close to $\widehat{U}_2$ and $\widehat{V}_2$. Eq. (16) shows that the same bases realize the SVD of the matrix B up to first-order terms in ε. Then the smallest singular values $\widehat{D}_2$ of $\widehat{A}$ have the same asymptotic distribution as the right hand side in (16).

In other words, Eq. (15) shows that the estimated singular vectors $\widehat{U}_2$ (respectively $\widehat{V}_2$) corresponding to the smallest singular values are equal to the singular vectors U_2 (respectively V_2) premultiplied by the orthogonal matrices M (respectively P) plus other elements that disappear asymptotically.

To determine the distribution of the smallest singular values $\widehat{D}_2$, we express $\widehat{D}_2$ as function of the matrix $\widehat{A}$. To this end pre- and post-multiply both sides of (11) by $\widetilde{U_2'}$ and $\widetilde{V}_2$ respectively. Since the columns of V_2 span the null space of A, and the columns U_2 span the null space of A' implies $A\widetilde{V}_2 = AV_2P = 0$ and $\widetilde{U_2'}A = M'U_2'A = 0$, thus we have

$$\widetilde{U'}_2 \widehat{A} \widetilde{V}_2 = \varepsilon \widetilde{U'}_2 B \widetilde{V}_2. \tag{17}$$

Using (17) we get the following corollary

Corollary 1 . *The estimates of the perturbed smallest singular values can be written as*

$$\widehat{D}_2 = \widetilde{U'}_2 \widehat{A} \widetilde{V}_2 + O_p(N^{-1}) \tag{18}$$

where $\widetilde{U_2'}$ and $\widetilde{V}_2$ are perturbation-dependent exact bases that are close to U_2 and V_2.■

We will now derive the statistical properties of perturbed smallest singular values of $\widehat{A}$ which will enable us to develop a statistical test.

3 Inference statistical of the smallest singular values(SSV)

In this section, we will develop the statistical properties of perturbed smallest singular values. Since the $(m-k)\times(m-k)$ and $(n-k)\times(n-k)$ matrices M and P are orthogonal respectively, and the singular values are invariant under left and right multiplication by orthogonal matrices, then we can expressed the estimates of the smallest singular values as:

$$\widehat{D}_2 = U_2'\widehat{A}V_2 + O_p(N^{-1}) \tag{19}$$

where U_2 and V_2 are an arbitrary bases of the null-spaces of A.

The smallest singular values $\widehat{D}_2$ of $\widehat{A}$ have the same asymptotic distribution as the right hand side in (19). The perturbed zero singular values can be approximated (19), we propose to investigate the statistical properties of this matrix. We denote by L

$$L = U_2'\widehat{A}V_2 \tag{20}$$

Eq. (20) defines a random matrix L as a function of the random matrix $\widehat{A}$, the matrices U_2 and V_2 are non-random matrices since they contain singular vectors of A.

We can now state the main result: first, we shall consider the limiting distribution of $N^{1/2}L$ as $N \longrightarrow \infty$.

Theorem 1 . *Let $l = vecL$, then the vector $N^{1/2}l$ is asymptotically normally distributed with zero mean and $(m-k)(n-k)\times(m-k)(n-k)$ covariance matrix given by*

$$Q = (V_2' \otimes U_2')\Sigma(V_2 \otimes U_2)$$

$$N^{1/2}l \sim \mathcal{N}(0, Q)$$

where Q is finite and positive definite.■

Let $\widehat{Q}$ the estimate of Q which can be obtained by replacing the unknown values U_2, V_2 and Σ by the corresponding estimate $\widehat{U}_2$, $\widehat{V}_2$ and $\widehat{\Sigma}$ the asymptotic distribution derived above can be used to construct a Wald-type test of null hypothesis.

Theorem 2 . *1) Let $\widehat{\Sigma}$, $\widehat{U}_2$ and $\widehat{V}_2$ be consistent estimates of Σ, U_2 and V_2, then*

$$\widehat{Q} \longrightarrow^{P} Q$$

where $\widehat{Q} = \left(\widehat{V'}_2 \otimes \widehat{U}'_2\right) \widehat{\Sigma} \left(\widehat{V}_2 \otimes \widehat{U}_2\right)$.

2) Under the null hypothesis, and given the conditions of Theorem 1, then (i) the test statistic is given by

$$L_1(k) = N\hat{l}'\widehat{Q}^{-1}\widehat{l} \longrightarrow^{d} \chi^2_{(m-k)(n-k)}$$

(ii) $L_1(k)$ is a sufficient statistic invariant under the orthogonal transformations on the bases U_2 and V_2.■

The theorem shows that the test statistic $L_1(k)$ is independent of the orthogonal transformations on the bases U_2 and V_2. We can thus evaluate it using the perturbation-dependent bases $\widetilde{U}_2$ and $\widetilde{V}_2$.

Recall that the estimator of L is

$\widehat{L} = \widehat{U'}_2\widehat{A}\widehat{V}_2 = \widehat{D}_2$ the diagonal matrix of the $n-k$ smallest singular values of $\widehat{A}$ and $\widehat{L}$ is a consistent estimator of L.

3.1 The singular covariance matrix case

In subsection 3.2. we have assumed that the asymptotic covariance matrix Σ of the matrix $\widehat{A}$ is non singular and in consequence the asymptotic covariance matrix Q of the vector $N^{1/2}l$ is finite and positive definite. In many situations in econometrics and statistics this assumption may be seriously violated. In this section the singular covariance matrix case is addressed.

We assume that the rank of the asymptotic covariance matrix Σ of dimension $mn\times$ mn is smaller than the number column mn that is $r(\Sigma) = mf < mn$. In consequence, the asymptotic covariance matrix Σ is singular. Since the asymptotic covariance matrix Q is a function of a matrix Σ then in this case the matrix Q may be singular. So the test statistic may not have an asymptotic χ^2 distribution under the null hypothesis. In this, we can generalise the above result to take account the singularity of the matrix Q.

We shall now examine the condition in which the covariance matrix Q may be singular. The covariance matrix $Q = (V_2' \otimes U_2')\Sigma(V_2 \otimes U_2)$ is of dimension $(m-k)(n-k) \times (m-k)(n-k)$. The matrices U_2 and V_2 are full column rank $(m-k)$ and $(n-k)$ respectively, hence the matrix $(V_2 \otimes U_2)$ has full column rank $(m-k)(n-k)$. In addition the columns of $(V_2 \otimes U_2)$ are orthonormal so that it verifies $(V_2 \otimes U_2)'(V_2 \otimes U_2) = (I_{m-k} \otimes I_{n-k})$. The rank of Q is equal to the minimum between the number of columns in $(V_2 \otimes U_2)$ and the rank of Σ.

$$r(Q) = \min\{(m-k)(n-k); mf\} \tag{21}$$

Therefore, the covariance matrix Q will be nonsingular if the rank of Σ is greater than or equal to $(m-k)(n-k)$.

We need the following proposition

Proposition 2 . *Suppose that $r(\widehat{\Sigma})$ converge almost surely to $r(\Sigma)$.*

Then:
(i) $\widehat{\Sigma}^{+} \longrightarrow^{P} \Sigma^{+}$.
(ii) $\widehat{Q}^{+} \rightarrow^{p} Q^{+}$.
where $\widehat{Q}^{+} = \left(\widehat{V'}_2 \otimes \widehat{U}_2'\right) \widehat{\Sigma}^{+} \left(\widehat{V}_2 \otimes \widehat{U}_2\right)$.

The matrices $\widehat{\Sigma}^{+}$ and $\widehat{Q}^{+}$are the Moore-Penrose generalized inverses of the matrices $\widehat{\Sigma}$ and $\widehat{Q}$ respectively.

Indeed it is worth emphasizing the fact that the consistency of $\widehat{\Sigma}$ for Σ does not imply consistency of $\widehat{\Sigma}^{+}$ for Σ^{+} since the Moore-Penrose inverse is not a continuous function of its elements.

By using Proposition 2, we are able to construct a statistic which has a central chi-squared distribution if, and only if, $H_{0,}$ is true.

Theorem 3 . *Under the null hypothesis and given the conditions of Theorem* 1 *and Proposition* 2*, then, the test statistic is given by*

$$L_1(k) = N\hat{l}'\widehat{Q}^{+}\widehat{l} \longrightarrow^{d} \chi^2{}_v \tag{22}$$

where $\widehat{Q}^{+} = \left(\widehat{V'}_2 \otimes \widehat{U}_2'\right)\widehat{\Sigma}^{+}\left(\widehat{V}_2 \otimes \widehat{U}_2\right)$ and v are degrees of freedom with $v = \min\{(m-k)(n-k); mf\}$ the rank of Q.

Note that the covariance matrix Q will be nonsingular if the rank of Σ is greater than or equal to $(m-k)(n-k)$ and we will use the inverse of Q instead a Moore-Penrose generalized inverse.

3.2 More general settings

In this section, we will consider the most general case in which we do not explicit the assumptions about the rank or the structure of the asymptotic covariance matrix Σ. The rank condition of the matrix $\widehat{\Sigma}$ is a sufficient condition to obtain an estimator consistent of $\widehat{\Sigma}^{+}$ its Moore-Penrose inverse, see Andrews (1987), however there is a situation where this condition may not be satisfied and consequently the asymptotic distribution of the $L_1(k)$ statistic is not a chi-square. It may be possible to overcome the problem by constructing or finding a suitable reduced rank estimator $\widehat{Q}$. We will use the singular value decomposition and the perturbation matrix results to find a suitable reduced rank estimator $\widehat{Q}$.

Theorem 4 . *Let $\widehat{Q}$ be a consistent estimator of Q and let $\widehat{S}$ and $\widehat{T}$ be the $(m-k)(n-k) \times (m-k)(n-k)$ orthogonal matrices such that $\widehat{S}'\widehat{Q}\widehat{T} = \widehat{\Lambda} = diag(\widehat{\gamma}_1, ..., \widehat{\gamma}_{vc})$ where $\widehat{\gamma}_1 \geqslant \widehat{\gamma}_2. \geqslant ... \geqslant \widehat{\gamma}_{vc} \geqslant 0$ and $vc = (m-k)(n-k)$.*

Under the null hypothesis, then the statistic test is given by
$L_1(k) = N \sum_{j=1}^{v} (\widehat{t}_j\widehat{l'})(\widehat{s_j}'\widehat{l})/\widehat{\gamma}_j \longrightarrow^{d} \chi^2{}_v$
where v are degrees of freedom and being the rank of Q.

4 Extension. Weighting matrices

Let the matrix A^* be defined as the following scaled version of the matrix A

$$A^* = \Phi A \Gamma$$

where the$(m \times m)$ and $(n \times n)$ weighting matrices Φ and Γ are positive definite. Since the rank of a matrix is invariant by pre-and post-multiplication by any nonsingular matrices, the rank of A^* is the same as that the rank of A, namely, k. We note, also, that the singular values are invariant with respect to the multiplication of the matrix A from the left and right by any nonsingular matrices. The important property of the $L(k)$ test statistics is that it is invariant with respect to orthogonal transformations on the bases U_2 and V_2. A nonsingular transformations on the bases U_2 and V_2 would in general affect the outcome of the $L(k)$ test statistics. We therefore particularly interested in left and right kernels of A computed from singular value decomposition of matrix A^*

$$A^* = \Phi A \Gamma = (U_1, U_2) \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1' \\ V_2' \end{pmatrix}$$

the partition of matrix corresponding of the rank of A then the matrices are partitioned in the same way as that (6) that is the order of the diagonal matrix D_1 is $(k \times k)$ and the order of the null matrix in the diagonal matrix is $(m-k) \times (n-k)$.

From Eq.(1), the estimator $\widehat{A}$ is a root-N consistent estimator of A, thus the limiting behavior of $\widehat{A^*} = \Phi \widehat{A} \Gamma$ is characterised by

$$N^{1/2} vec(\widehat{A^*} - A^*) = N^{1/2} (\Gamma' \otimes \Phi) vec(\widehat{A} - A) \longrightarrow^d \mathcal{N}(0, \Sigma^*). \qquad (23)$$

where $\Sigma^* = (\Gamma' \otimes \Phi)\Sigma(\Gamma \otimes \Phi')$ is the $mn \times mn$ covariance matrix non zero but possibly singular.

We apply a similar singular value decomposition to the matrix estimated of $\widehat{A^*}$

$$\widehat{A^*} = (\widehat{U}_1, \widehat{U}_2) \begin{pmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_2 \end{pmatrix} \begin{pmatrix} \widehat{V}_1' \\ \widehat{V}_2' \end{pmatrix}$$

the matrices are partitioned in the same way as that (6) and hence the sizes of $\widehat{U}_1$, $\widehat{U}_2, \widehat{D}_1,$ $\widehat{D}_2,$ $\widehat{V}_1$ and $\widehat{V}_2$ are exactly the same as that of U_1, $U_2, D_1,$ D_2, V_1 and V_2.

The expression of the $(m-k)(n-k)$ matrix L intervening in the perturbation analysis is then

$$L = U_2' \Phi \widehat{A} \Gamma V_2$$

Following Theorem 1, the vector $N^{1/2}vecL$ is asymptotically normal with zero mean and $(m-k)(n-k) \times (m-k)(n-k)$ covariance matrix Q^* defined as

$$Q^* = (V_2' \otimes U_2')\Sigma^*(V_2 \otimes U_2)$$

$$N^{1/2} l \sim \mathcal{N}(0, Q^*)$$

This asymptotic covariance matrix Q^* is more general than defined in the Theorem 1.

We now examine the condition in which the covariance matrix Q^* may be singular. Since the matrix $(V_2 \otimes U_2)$ has full column rank $(m-k)(n-k)$ then in general case, the rank of the $(m-k)(n-k) \times (m-k)(n-k)$ asymptotic covariance matrix Q^* is equal to

$$r(Q^*) = \min\{(m-k)(n-k); r(\Sigma^*)\}.$$

and in particular case where the $m \times m$ and $n \times n$ weighting matrices Γ and Φ are nonsingular then the rank of Q^* is equal to

$$r(Q^*) = \min\{(m-k)(n-k); r(\Sigma)\}.$$

Also, in that case the covariance matrix Q^* will be nonsingular if the rank of Σ is greater than or equal to $(m-k)(n-k)$ this condition is the same rank condition as the covariance matrix Q.

Let $\widehat{\Phi}$ and $\widehat{\Gamma}$ be consistent estimator of Φ and Γ then following Proposition 2, the consistent estimator $\widehat{Q^*}$of Q^* can be obtained by replacing U_2, V_2 and Σ^* by the corresponding estimate $\widehat{U}_2$, $\widehat{V}_2$ and $\widehat{\Sigma^*}$ that is

$$\widehat{Q^*} = (\widehat{V}_2' \otimes \widehat{U}_2')\widehat{\Sigma^*}(\widehat{V}_2 \otimes \widehat{U}_2)$$

where $\widehat{\Sigma^*} = (\widehat{\Gamma}' \otimes \widehat{\Phi})\widehat{\Sigma}(\widehat{\Gamma} \otimes \widehat{\Phi}')$.

If we assume that the $m-k)(n-k) \times (m-k)(n-k)$ covariance matrix Q^* is nonsingular then the test statistic is defined by

$$L_2(k) = N\hat{l}'\widehat{Q^*}^{-1}\hat{l} \longrightarrow^d \chi^2{}_{(m-k)(n-k)}$$

if the null hypothesis is true.

In the more general settings, in which we do not explicit the assumptions about the rank or the structure of the asymptotic covariance matrix Σ^*. We use the singular value decomposition and the perturbation matrix result to find a suitable reduced rank estimator $\widehat{Q^*}$. We can establish the following theorem analogously to the Theorem 4.

Theorem 5 . *Let $\widehat{Q^*}$ be a consistent estimator of Q^* and let $\widehat{S^*}$ and $\widehat{T^*}$ be the $(m-k)(n-k) \times (m-k)(n-k)$ orthogonal matrices, then the singular value decomposition of $\widehat{Q^*}$ is*

$$\widehat{S^*}'\widehat{Q^*}\widehat{T^*} = \widehat{\Lambda^*} = diag(\widehat{\gamma}^*{}_1, ..., \widehat{\gamma}^*{}_{vc})$$

where $\widehat{\gamma}^*{}_1 \geqslant \widehat{\gamma}^*{}_2. \geqslant ... \geqslant \widehat{\gamma}^*{}_{vc} \geqslant 0$ and $vc = (m-k)(n-k)$.

Under the null hypothesis, then the statistic test is given by

$L_2(k) = N \sum\limits_{j=1}^{v} (\widehat{t^*}'_j \hat{l})(\widehat{s_{*j}}'\widehat{l})/\widehat{\gamma}^*{}_j \longrightarrow^d \chi^2{}_v$

where v are degrees of freedom and being the rank of Q^*.

The Proof is similar to the Proof of Theorem 4.

4.1 A particular case: Kronecker structure of the covariance matrix Σ

In this section, we will consider that the asymptotic covariance matrix Σ admits a Kronecker product form $\Sigma = \Psi \otimes \Omega$. In many applications in econometrics, one such Kronecker structure of the covariance matrix occurs, for example, in linear multivariate regression model, in seemingly unrelated regression (SURE) model and in linear simultaneous equations models (and equivalently instrumental variables models), among others see Greene (2000). The matrix A as be seen as a regression coefficient matrix and the matrix Σ its asymptotic covariance matrix. When the covariance matrix has a Kronecker structure, the $L_2(k)$ test statistics can be simplified because the matrix L has a multivariate normal standard distribution follows the appropriate choice of the two weigthing matrices. In the sequel we will consider a more general condition where the matrix Σ may be singular. We will, firstly, examine the conditions to normalise the asymptotic covariance matrix and show that the asymptotic distribution of the $L(k)$ statistic is χ^2 variable. In that case we will establishes a relation of the $L(k)$ statistic with other tests statistics such likelihood ratio and Lagrange multiplier statistics. Secondly, we will indicate how the weighting matrices Γ and Φ should be chosen to normalise the matrix Σ^*. The choice of the two weighting matrices depends on the properties of the matrices Ψ and Ω as well as the singularity or not of the matrix Σ. These choice leads to satisfy that the matrix Σ^* is equal to

$$\Sigma^* = (\Gamma' \otimes \Phi)\Sigma(\Gamma \otimes \Phi') = I$$

the identity matrix . As the asymptotic covariance matrix Σ may be decomposed into the Kronecker product form, we can be expressed the matrix Σ^* by

$$\Sigma^* = (\Gamma' \otimes \Phi)(\Psi \otimes \Omega)(\Gamma \otimes \Phi') = (\Gamma'\Psi\Gamma \otimes \Phi\Omega\Phi') = I$$

Thus the matrix $\Sigma^*$is equal to the identity matrix if and only if

$$(\Gamma'\Psi\Gamma) = I \text{ and } (\Phi\Omega\Phi') = I$$

These conditions provides a natural standardization for the elements of $vec(\widehat{A})$, hence

$$N^{1/2}(\Gamma' \otimes \Phi)vec(\widehat{A} - A) \longrightarrow^d \mathcal{N}(0, I)$$

Consequently the asymptotic covariance matrix Q^* of the vector $N^{1/2}l$ is also the identity matrix because the matrix $V_2 \otimes U_2$ is orthonormal, that is

$$Q^* = (V_2' \otimes U_2')I(V_2 \otimes U_2) = I \otimes I = I.$$

So that the asymptotic distribution of the random vector $N^{1/2}l$ is normal with mean vector zero and variance matrix unity.

$$N^{1/2}l \longrightarrow^d N(0, I).$$

Thus we have the following theorem.

Theorem 6 . *Let $\widehat{\Phi}$, $\widehat{\Gamma}$, $\widehat{U}_2$ and $\widehat{V}_2$, be the consistent estimator of the matrices Φ, Γ, U_2, and V_2 respectively and the vector $N^{1/2}vec\widehat{L}$ is asymptotically distributed in multivariate normal standard*

$$N^{1/2} vec\widehat{L} = N^{1/2} vec(\widehat{U}_2' \widehat{\Phi} \widehat{A} \widehat{\Gamma} \widehat{V}_2) \longrightarrow^d N(0, I).$$

Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq ... \geq \widehat{\lambda}_n$ be the ordered singular values of the matrix $\widehat{\Phi}\widehat{A}\widehat{\Gamma}$. Then, the test statistic is defined by

$$L_3(k) = N\hat{l}'\widehat{l} = N \sum_{j=k+1}^{n} \widehat{\lambda}_j^2 \longrightarrow^d \chi^2{}_v \tag{24}$$

where v are degrees of freedom with v being the rank of Q^*.

We note that the $\widehat{\lambda}_j^2$ are also the eigenvalues of the matrix $\widehat{A}'\widehat{A}$ or $\widehat{A}\widehat{A}'$. This theorem shows that when the asymptotic covariance matrix Σ has a Kronecker structure then, the $L_3(k)$ test statistics can be simplified by N times the sum of the $n-k$ smallest eigenvalues of $\widehat{A}'\widehat{A}$ or $\widehat{A}\widehat{A}'$.

This result is more important, it shows that the $L_3(k)$ statistic is invariant with respect the weigthing matrices $\widehat{\Phi}$, and $\widehat{\Gamma}$, but the both matrices influence the variance of the estimator of the matrix $\widehat{A}$. The both weigthing matrices normalise the covariance matrix Σ^*, and the rank of the identity matrix depends on the properties of Σ.

We will discuss below the number of degrees of freedom of χ^2 distribution of $L_3(k)$ test statistics. It depends the rank of identity matrix, the variance of the random vector $N^{1/2}l$ (multivariate normal standard). Before we will establish the relation of the $L_3(k)$ test statistics with other statistics.

Corollary 2 . *Under the null hypothesis, and given the conditions of Theorem 6, then (i) the $L_3(k)$ test statistic is equivalent to the likelihood ratio (LR) statistic $LR = N\sum_{j=k+1}^{n} \ln(1+\widehat{\lambda}_j^2)$,and to the Lagrange Multiplier (LM) statistic which is identical to the Rao score statistic in general $LM = N\sum_{j=k+1}^{n} \frac{\widehat{\lambda}_j^2}{1+\widehat{\lambda}_j^2}$*

(ii) the likelihood ratio (LR) and the Lagrange Multiplier (LM) statistics are asymptotically distributed as a $\chi^2{}_v$ random variable.

Comments: 1. The normality assumption is not necessary to derive the asymptotic distribution of these statistics.

2. The likelihood ratio statistic is derived by Anderson (1951) (1980) and (1984) for testing the rank of a regression coefficient matrix in multivariate normal linear model.

3. These statistics have been proposed in the literature on canonical correlation analysis, for testing the significance of the $(n-k)$ smallest canonical correlations where $\widehat{\lambda}_j = \widehat{\rho}_j/(1-\widehat{\rho}_j^2)^{1/2}$ and $\widehat{\rho}_j$ is the jth smallest sample canonical correlation [Anderson (1984, ch. 12) (1999), Robinson (1973)] among others.

4. These statistics have been proposed, also, in the multivariate regression model for testing linear restrictions [Anderson (1984, ch.8)] and in simultaneous equation models for testing overidentification and predeterminedness [Anderson and Kunitomo (1992), (1994)].

So far, we have made no other assumption about Σ except that it is decomposed into Kronecker product form. We next develop another properties of the $L_3(k)$ test statistic. We will indicate how the weighting matrices Γ and Φ should be chosen to normalise the asymptotic covariance matrix Σ^* and we will also determine the number of degrees of freedom of the chi-square distribution of test statistics. The development that follows depends on whether the covariance matrix Σ is nonsingular or singular. The two cases will be considered separately.

Since the covariance matrix Σ can be represented by the Kronecker product of two matrices Ψ and Ω, we have then the following possibilities:

1) both Ψ and Ω nonsingular,

2) (i) Ψ nonsingular and Ω singular,

(ii) Ψ singular and Ω nonsingular

(iii) both Ψ and Ω singular.

4.1.1 The covariance matrix Σ is nonsingular

First, we will consider that the two matrices Ψ and Ω are positive definite. Since the Kronecker product of two positive definite matrices is positive definite, so that Σ is positive definite, then in this case the choice of the two weighting matrices Γ and Φ is easy. If the $(m \times m)$ and $(n \times n)$ matrices Ψ and Ω are positive definite then there exist two nonsingular matrices Γ and Φ such that

$$\Gamma'\Psi\Gamma = I_m \text{ and } \Phi\Omega\Phi' = I_n$$

thus the matrix Σ^* is equal to $\Sigma^* = I_m \otimes I_n = I_{mn}$ the identity matrix. Often, in practice one use the square root of the matrices Ψ and Ω such that $\Psi = (\Psi^{1/2})(\Psi^{1/2})'$ and $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$ where the matrices $\Psi^{1/2}$

and $\Omega^{1/2}$ are the square-root matrices respectively. In this case one can choice $\Gamma = \Psi^{-1/2}$ and $\Phi = \Sigma^{-1/2}$ such that $(\Psi^{-1/2})\Psi(\Psi^{-1/2})' = I_m$ and $(\Omega^{-1/2})\Omega(\Omega^{-1/2})' = I_n$ then $\Sigma^* = I_m \otimes I_n = I_{mn}$.

In consequence the asymptotic covariance matrix Q^* of the vector $N^{1/2}l$ is also the identity matrix

$$Q^* = (V_2' \otimes U_2')I_{mn}(V_2 \otimes U_2) = I_{n-k} \otimes I_{m-k} = I_{(n-k)(m-k)}. \quad (25)$$

and the $L_3(k)$ test statistic has a limiting chi-square distribution with $v = (m-k)(n-k)$ as the degrees of freedom.

4.1.2 The covariance matrix Σ is singular

Next, we will consider the case where the asymptotic covariance matrix Σ is singular, in this case unless one the two matrices Ψ and Ω are positive semidefinite.

(i) Suppose that the $(m \times m)$ matrix Ψ is positive definite and the $(n \times n)$ matrix Ω positive semidefinite and $r(\Omega) = s < n$. Using the square root of the matrix Ω that is $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$ where the $(s \times n)$ matrix $\Omega^{1/2}$ is the square-root matrix. In this case, one can choice $\Phi = \Omega^{+1/2}$ the Moore-Penrose generalized inverse of $\Omega^{1/2}$ and $\Gamma = (\Psi^{1/2})^{-1}$. Thus, we have $(\Omega^{+1/2})\Omega(\Omega^{+1/2})' = I_s$. Then the matrix Σ^* is equal to $\Sigma^* = I_m \otimes I_s = I_{ms}$ and the asymptotic covariance matrix Q^* of the vector $N^{1/2}l$ is

$$Q^* = (V_2' \otimes U_2')I_{ms}(V_2 \otimes U_2) = I_v \quad (26)$$

The rank of the matrix Q^* is equal to $r(Q^*) = v = \min\{(m-k) \times (n-k); ms\}$. Thus the $L_3(k)$ test statistic has a limiting chi-square distribution with $v = \min\{(m-k) \times (n-k); ms\}$ degrees of freedom.

(ii) Alternatively, suppose that the $(m \times m)$ matrix Ψ is positive semidefinite and $r(\Psi) = r < m$ and the matrix Ω positive definite. One can choice $\Phi = \Omega^{-1/2}$ and $\Gamma = \Psi^{+1/2}$ the Moore-Penrose generalized inverse of the $(r \times m)$ matrix $\Psi^{1/2}$ such that $(\Psi^{+1/2})\Psi(\Psi^{+1/2})' = I_r$. The rank of the matrix $Q^*$is equal to $r(Q^*) = v = \min\{(m-k) \times (n-k); rn\}$ which is also the degrees of freedom of the test statistic $L_3(k)$.

(iii) More generally, the $(m \times m)$ and $(n \times n)$ matrices Ψ and Ω are positive semidefinite with $r(\Psi) = r < m$, and $r(\Omega) = s < n$. Also using the square root of the two matrices Ψ and Ω. One can choice $\Gamma = \Psi^{+1/2}$ and $\Phi = \Omega^{+1/2}$ the Moore-Penrose generalized inverse of the $(r \times m)$ and $(s \times n)$ matrices

$\Psi^{1/2}$ and $\Omega^{1/2}$ respectively. Then $r(Q^*) = v = \min\{(m-k)\times(n-k); rs\}$ which is also the degrees of freedom of the $L_3(k)$ test statistic.

Remark 1 *These results are particularly important, the $L_3(k)$ statistic generalise the canonical correlation analysis statistics and the statistics for testing linear restrictions in the multivariate regression model in which the covariance matrix Σ is singular.*

When the asymptotic covariance matrix has a Kronecker structure, the weighting matrices Γ and Φ can be chosen in different ways following the property of the matrix Σ. All these results are gathered in the following table

Table 1. Weighting matrices corresponding to Kronecker structure of the covariance matrix Σ.

Γ	Φ	Covariance matrix $\Sigma = \Psi \otimes \Omega$.
$\Psi^{-1/2}$	$\Omega^{-1/2}$	Ψ and Ω are non singular
$\Psi^{+1/2}$	$\Omega^{-1/2}$	Ψ singular and Ω non singular
$\Psi^{-1/2}$	$\Omega^{+1/2}$	Ψ nonsingular and Ω singular
$\Psi^{+1/2}$	$\Omega^{+1/2}$	Ψ and Ω are singular

5 Strong consistency for a sequential testing procedure

It is well known that a sequential testing procedure does not lead to a consistent estimate of the true rank matrix unless some adjustment is made to the significance level, [see Cragg and Donald (1997); Robin and Smith (2000)]. In this section we will formalise the testing procedures that empirical researchers often use in a less formal, and sometimes vague fashion. We will also establish sufficient conditions for strong consistency.

We consider tests based on the $L(k)$ statistics where $L(k)$ denoted all tests statistics $L_1(k)$, $L_2(k)$, and $L_3(k)$. Starting with $k = 1$, we carry out tests with progressively larger k until we find a test that does not reject the null hypothesis that the rank of the matrix A is k. Let $\widehat{k}$ be the value of k for the first test we find that does not rejected This is a sequential testing procedure of the rank of the matrix.

The $L(k)$ test statistic has an asymptotic chi-square distribution with v degrees of freedom. We take $\gamma_k C_N \succ 0$ the critical value employed with the test statistic $L(k)$ when $\widehat{k} = k$ and the sample size is N. The value γ_k represent the quantile of a chi-squared distribution with v degrees of freedom and C_N is a predetermined sequence of numbers whose choice is discussed below. So that $\gamma_k C_N$ be the $(1-\alpha)-$quantile of a chi-squared distribution.

Let $\widehat{k} \in \{1, 2, ...k-1\}$ be such that $L(k) \succ \gamma_k C_N$ $\forall k > \widehat{k}$ with $k \in \{1, 2, ...n\}$, $L(k) \leq \gamma_k C_N$ and $\widehat{k} = k$. In words, $\widehat{k}$ the estimator of the true rank of the matrix A is the smallest value of k for which some $L(k)$ tests do not reject the null hypothesis.

For a strongly consistent estimate of the rank of the matrix A, we assume the function C_N satisfy:

Assumption C (i) $C_N > 0$; (ii)$\frac{C_N}{N} \longrightarrow 0$; and (iii) $\frac{C_N}{\log\log N} \longrightarrow \infty$.

Remark 2 *:*

1) The critical value $\gamma_k C_N$ is similar to the significance level α_N satisfying $\alpha_N \rightarrow 0$ and $ln\alpha_N = o(N)$ of Theorem 5.8 of Potscher (1983) for weak consistency of Lagrange multiplier (LM) tests for lag selection in ARMA models.

2) The conditions (i)-(iii) admit a large range of possible choices of C_N. In the simulation results of Ratsimalahelo (2002), we took $C_N = \sqrt{\log N}$.

The strongly consistent estimate of the rank of matrix requires the law of iterated logarithm. As Gragg and Donald (1997) and Nishii (1988) and without loss of generality, we assume that the matrix estimated $\widehat{A}$ follows the law of iterated logarithm (LIL). Zhao, Krishnaiah and Bai (1986) have been shown that if the matrix estimated follows the law of iterated logarithm then the corresponding singular values follows also the law of iterated logarithm.

We now state the assumption under which the results below hold.

Assumption LIL. For the matrix estimated $\widehat{A}$ and its singular values $\widehat{\lambda}_i$, the following relations hold with probability one

$$\widehat{A} - A = O(\log\log N/N)^{1/2}$$

$$\widehat{\lambda}_i - \lambda_i = O(\log\log N/N)^{1/2} \qquad \text{for } i = 1, ..., n.$$

where $\lambda_i = 0$ for $i > k$ then we have $\widehat{\lambda}_{k+1} = O(\log\log N/N)^{1/2}$.

For any $\widehat{k} = 1, ..., k-1$, the null hypothesis $r(\widehat{A}) = k$ is rejected almost surely as $N \rightarrow \infty$ and let $\widehat{k} = k$, then the null hypothesis does not rejected almost surely as $N \rightarrow \infty$. Thus $\widehat{k}$ is the value for which $L(k)$ test does not reject. The strong consistency of $\widehat{k}$ is established in the following theorem.

Theorem 7 . *Let γ_k the value of the quantile of a chi-squared distribution with v degrees of freedom and k_0 denote the true rank of A. Suppose Assumptions C and LIL hold. Then under the null hypothesis, with probability one, $lim_{N\to\infty}L(k)/C_N \succ \gamma_k$ if $\widehat{k} \prec k_0$ and $lim_{N\to\infty}L(k)/C_N \leq \gamma_k$ if $\widehat{k} = k_{0.}$*

This theorem shows that the $L(k)$ statistics provides a strongly consistent estimate of k_0 the true rank of A.

6 The model selection approach

We now consider the approach of the information criteria to estimate the rank of matrix

Define the following criteria

$IC(k, C_N) = \widehat{L}(k) + \varphi(k)C_N$ for $k = 0, 1, ..., n-1$

where the function $\varphi(k)$ is strictly increasing and the constants C_N is deterministic penalty term.

Let $\widehat{k}$ the estimator of k that minimise the criteria function

$\widehat{k} = \arg\min IC(k, C_N)$ for $k = 0, 1, ..., n-1$

Next we will prove that $\widehat{k}$ is a consistent estimator of k. This depends on the functions $\varphi(k)$ and C_N. Thus the essential part of the criteria function is the penalty term $\varphi(k)C_N$. The higher we specify the penalty term, the lower the risk of overestimating the rank and the higher the risk of underestimating the rank. They are assumed to satisfy:

Assumption IC: (a) $\varphi(k)$ is strictly increasing (b) C_N satisfies the conditions (i) - (iii) of the *Assumption C.*

In application of the $IC(k, C_N)$ criteria, practitioners will have to specify the functions $\varphi(k)$ and C_N and the Theorem provides an asymptotic justification for many different values of the penalty terms. In statistical literature, some fixed choices of C_N have been suggested such as $C_N = 2$ by Akaike (1974) (AIC); $C_N = \log(N)$ by Schwarz (1978) (BIC) and $C_N = c\log\log(N)$ for some $c > 2$ by Hannan and Quinn (1979) ($HQIC$).

In the simulation results of Ratsimalahelo (2002), we took $c = 2.02$ in the specification of the $HQIC$criterion.

Some comments on the choice of C_N have been made by Bai, Krisnhaiah, and Zhao (1989). The AIC procedure does not satisfy *Assumption IC(b)* because $C_N \nrightarrow \infty$, in consequence AIC is not consistent. As a result there will generally be a positive probability of overestimating the rank of the matrix.

Now we will give the following theorem concerning strong consistency of the estimator k.

Theorem 8 . *Let k_0 denote the true rank of A and $\widehat{k}$ the estimated rank obtained from Eq.* (3)*. Suppose Assumptions LIL and IC hold, then $lim_{N\to\infty}$ $\widehat{k}=k_0$ a.s.*

Note that the strong consistency of $\widehat{k}$, follows $P(lim_{N\to\infty}\ \widehat{k}=k_0)=1$. This result shows that the underestimation and the overestimation of the true rank is not possible asymptotically.

7 Relation with the characteristic roots test (CRT)

The characteristic roots test statistic is based of the eigenvalues of the matrix quadratic form $\Phi A\Gamma A'$ where the $m\times m$ and $n\times n$ matrices Φ and Γ are positive definite. Let $\lambda_1^2\geq\lambda_2^2\geq...\geq\lambda_n^2\geq 0$ be the roots of the matrix quadratic form, the λ_j^2 for $j=1,...,n$ are also the solutions of the determinantal equation $\mid A\Gamma A'-\lambda^2\Phi^{-1}\mid=0$. Because the matrices Φ and Γ are non singular, the rank of the matrix quadratic form is equal to the rank of the matrix A which characterise the null hypothesis H_0 $r(A)=k$. Procedures used to estimate the rank from H_0 involve testing a sequence of *rank hypotheses* defined in terms of the nullity of some of the eigenvalues of the matrix quadratic form. Anderson and Kunitomo (1994) and Robin and Smith (2000) used a criterion which is N times a smooth function of the $n-k$ smallest roots $f(\lambda^2{}_{k+1},...,\lambda^2{}_n)$ as the test statistic

$CRT=N\sum_{j=k+1}^{n}f(\widehat{\lambda}^2{}_j)$

namely characteristic roots test statistic, where the function $f(.)$ satisfies (i) $f(0,...,0)=0$, (ii) $f(\lambda^2{}_{k+1},...,\lambda^2{}_n)$ is totally differentiable at $(\lambda^2{}_{k+1},...,\lambda^2{}_n)=(0,...,0)$ and (iii) $\frac{\partial f(\lambda^2{}_{k+1},...,\lambda^2{}_n)}{\partial\lambda^2{}_{j1}}\rfloor_{\lambda^2{}_{k+1}=...=\lambda^2{}_n=0}=1, \qquad j=k+1,...,n.$ [see Anderson and Kunitomo (1994)].

This class of statistics includes many test statistics as special cases including the likelihood ratio statistic, the Lagrange Multiplier statistic, and the Wald statistic [1].

We will shown in this section the relation between the CRT statistic and our $L_2(k)$ statistic (using the weighting matrices).

Consider the following decomposition of the matrix weigthed A^*

$$\Phi A\Gamma = W\Delta O' \qquad (27)$$

where the $m \times m$ and $n \times n$ matrices W and O are nonsingular and $\Delta = diag(\lambda_1, \lambda_2, ..., \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq, ..., \geq \lambda_n$.

Since the $m \times m$ and $n \times n$ matrices Φ and Γ are nonsingular, pre-and post-multiply (27) by Φ^{-1} and Γ^{-1} respectively, we obtain

$$A = \Phi^{-1}W\Delta O'\Gamma^{-1} \qquad (28)$$

If we impose the orthonormalization conditions

$$W'\Phi^{-1}W = I \text{ and } O'\Gamma^{-1}O = I \qquad (29)$$

then, pre-and post-multiply (28) by W' and O respectively, we gets

$$W'AO = \Delta \qquad (30)$$

Accordingly from equations (29) and (30), let o_i and w_i obey the orthonormality conditions $w_i'\Phi^{-1}w_j = o_i'\Gamma^{-1}o_j = \delta ij$ where δij is the Kronecker delta and $w_i'Ao_i = \lambda_i = o_i'A'w_i$. The resulting equations are

$$Ao_i - \lambda_i\Phi^{-1}w_i = 0 \qquad (31.1)$$

$$-\lambda_i\Gamma^{-1}o_i + A'w_i = 0 \qquad (31.2)$$

The following proposition show the relation between the eigenvalues of the matrix quadratic form $\Phi A\Gamma A'$ (or equivalently $\Gamma A'\Phi A$) and the singular values of the matrix $A^* = \Phi A\Gamma$.

Proposition 3 . *Let the $m \times m$ matrix W of eigenvectors of $A\Gamma A'$ in the metric of Φ^{-1} and the $n \times n$ matrix O of eigenvectors of $A'\Phi A$ in the metric of Γ^{-1} satisfies the orthonormalization conditions $W'\Phi^{-1}W = I$ and $O'\Gamma^{-1}O = I$. Then the singular values of the matrix $\Phi A\Gamma : \lambda_1, \lambda_2, ..., \lambda_n$ are the square*

[1] It is possible to consider more general class of statistics. (See ch. 8 of Anderson (1984)).

roots of $\lambda_1^2, \lambda_2^2, ..., \lambda_n^2$ *the latent roots of* $\Phi A \Gamma A'$ *or equivalently the latent roots of* $\Gamma A' \Phi A$.

The relation between the CRT statistic and our $L_2(k)$ statistic follows from the fact the singular values of the weighted matrix $\Phi A \Gamma$ that is $\Delta = diag(\lambda_1, \lambda_2, ..., \lambda_n) = D$ are the square roots of eigenvalues of the matrix quadratic form $\Phi A \Gamma A'$ (or equivalently $\Gamma A' \Phi A$). The smallest singular values are contained in D_2 It follows from (30) that the limiting distribution of the smallest singular values $\widehat{D}_2$ is the same as the limiting distribution of the corresponding singular values of the $(m-k) \times (n-k)$ matrix

$$\sqrt{N} W_2'(\widehat{A} - A) O_2 \tag{32}$$

where W_2 and O_2 are the matrices of eigenvectors corresponding to the $n-k$ smallest eigenvalues.

The asymptotic distribution of CRT is the same as that of the sum of the squares of the singular values of (32) which can be expressed as $N \times trace\{[W_2'(\widehat{A}-A)O_2][W_2'(\widehat{A}-A)O_2]'\}$. Therefore, the asymptotic distribution of CRT is the same as the asymptotic of $Nvec[W_2'(\widehat{A} - A)O_2]'vec[W_2'(\widehat{A} - A)O_2]$. Because $\sqrt{N}vec[W_2'(\widehat{A} - A)O_2]$ has a normal limiting distribution with zero means and covariance matrix $(O_2' \otimes W_2)\Sigma(O_2 \otimes W_2')$ then the CRT statistic is asymptotically distributed as linear combination of independent chi-squared random variables with one degree of freedom [see Johnson and Kotz (1970) p. 150. e.g. Lemma 3.2 of Vuong (1989)][2]. The weights are the nonzero eigenvalues of the matrix $(O_2' \otimes W_2)\Sigma(O_2 \otimes W_2')$.

This procedure is more complicated because the asymptotic critical values of the test statistic are not tabulated. The test requires the percentiles of a weighted chi-squared distribution for which computationally intensive algorithms need be used. Moreover the weight estimation introduces variability and hence potentially less accuracy, to the testing procedure.

To construct our $L_2(k)$ statistic we use the matrix perturbation theory to determine the limiting distribution of the smallest singular values $\widehat{D}_{2.}$ The result show that the limiting distribution of the smallest singular values of $\widehat{D}_2$ is the same as the limiting distribution the $(m-k)(n-k)$ matrix $L = U_2' \Phi \widehat{A} \Gamma V_2$. We use the properties statistical of the matrix L to construct

[2]see also theorem 3.2 of Robin and Smith (2000)

the $L_2(k)$ statistic. So the test statistic can be obtain straightforwardly without a particular smooth function and without the orthonormalization conditions, as is well known the orthonormalization condition have no economic interpretation. Moreover the limiting distribution of the $L_2(k)$ statistic is a standard chi square distribution. In addition, when the covariance matrix has a Kronecker structure, the $L_2(k)$ statistic can be simplified even if the covariance matrix is singular and in that case it is asymptotically equivalent to the LR and LM statistics.

8 Conclusion

In this paper, we have proposed new and general methods for determining the rank of an unobserved matrix for which a $root - N$ consistent estimator is available. We use the matrix perturbation theory to construct or find a suitable bases of the kernel of the matrix and to determine the limiting distribution of the estimator of the smallest singular values. The statistic, based on matrix perturbation results, is asymptotically distributed as chi-square under the null. We have fully characterized the asymptotic of the generalized Wald statistic under the most general conditions. The test statistic has desirables properties that it is (i) a statistic sufficient invariant under the orthogonal transformations, (ii) asymptotically equivalent to LR and LM statistics when the covariance matrix has a Kronecker structure even if the asymptotic covariance matrix is singular.

Two approaches have been considered to estimate the rank of random matrix, sequential testing strategy and information theoretic criterion procedures, we have been established conditions for strong consistency of both procedures.

This study provides a theoretical justification for a number of statistical tests for which one can determine how many singular values of an estimated matrix should be declared equal to zero.

We have assumed that the dimensions m and n of the matrix A are finite, however most of the results will still hold when the dimensions are not finite, provide A is a *compact operator.*

The test proposed in this paper may be particularly useful in determining the dimension of the cointegrating space which are very commonly used in modelling time series.

References

[ANDERSON] T. W. (1951): "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics,* 22, 327-351.

[ANDERSON] T. W. (1963): "Asymptotic Theory for Principal Component Analysis," *The Annals of Mathematical Statistics*, 34, 122-148.

[ANDERSON] T. W. (1980): Correction to "estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Statistics*, 8, 1400.

[ANDERSON] T. W. (1984): "*An Introduction to Multivariate Statistical Analysis"*, Second Edition, Wiley, New-York.

[ANDERSON] T. W., AND N. KUNITOMO (1992): "Tests of overidentification and predeterminess in simultaneous equation models," *Journal of Econometrics*, 54, 49-78.

[ANDERSON] T. W., AND N. KUNITOMO (1994): "Asymptotic robustness of tests of overidentification and predeterminess," *Journal of Econometrics*, 62, 383-414.

[ANDERSON] T. W. (1999): "Asymptotic Theory for canonical correlation analysis," *Journal of Multivariate Analysis*, 70, 1-29.

[ANDREWS] D.W.K. (1987): "Asymptotic results for generalized Wald tests," *Econometric theory*, 3, 348-358.

[ANDREWS] D.W.K. (1988): "Chi-square diagnostic tests for econometric models: Theory," *Econometrica*, 56, 1419-1453.

[AOKI] M.,AND A. HAVENNER (1991): "State Space Modelling of Multiple Time Series," *Econometric Reviews*, 10, 1-59.

[BAI] Z. D., P. R. KRISNHAIAH, AND L. C., Zhao (1989): "On rates of convergence of efficient detection criteria in signal processing with white noise," *IEEE Transactions On Information Theory*. 35, 380-388.

[CRAGG] J. G., AND S. G. DONALD (1996): "On the Asymptotic Properties of LDU Based Tests for the Rank of a Matrix," *Journal of the American Statistical Association*, 91, 1301-1309.

[CRAGG] J. G., AND S. G. DONALD (1997): "Inferring the Rank of Matrix," *Journal of Econometrics*, 76, 223-250.

[EATON] M. L. AND D. E. TYLER (1994): "The asymptotic distribution of singular values with applications to canonical correlations and correspondance analysis," *Journal of Multivariate Analysis*, 34, 439-446.

[FERGUSON] T. S. (1996): "*A Course in Large Sample Theory*" Chapman and Hall, London.

[FIELD] C. (1993): "Tail areas of linear combinations of chi-squares and non-central chi-squares," *Journal of Statistical Computation and Simulation*, 45, 243-248.

[FUCHS] J. J. (1990): "Structure and Order Estimation of Multivariable Stochastic Processes," *IEEE Transactions On Automatic Control*, 35, 1338-1341.

[GILL] L., AND A. LEWBEL (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., AND C. F. VAN LOAN (1996): *Matrix Computations*, 3rd ed. John Hopkins University Press. Baltimore.

[GOURIEROUX] C., A. MONFORT, AND E. RENAULT (1993): "Tests sur le noyau, l'image et le rang de la matrice des coefficients d'un modèle linéaire multivarié," *Annales d'Economie et de Statistique*, 32, 81-111.

[HALL] A. R., G. D. RUDEBUSCH AND D. W. WILCOX (1996): "Judging Instrument relevance in instrument variables estimation," International Economic Review, 37, 283-298.

[HANNAN] E. J. AND B. G. QUINN (1979): "The determination of the order of an autoregression," *Journal of Royal Statistical Society Series B,* 41, 190-195.

[HANNAN] E. J., AND M. DEISTLER (1988): *The Statistical Theory of Linear Systems.* Wiley New York.

[HSIAO] C. (1986): Identification, *Handbook of Econometrics*, Volume 1, eds Z. Griliches and M.D Intrilligator. Amsterdam: North-Holland.

[JOHNSON] , N. L., AND S. KOTZ (1970): *Continuous Univariate Distributions-2.* New-York: John Wiley and Sons.

[KAILATH] T. (1980): *Linear systems*, Prentice-Hall, Englewood Cliffs.

[KATO] T. (1982): *A Short Introduction to Perturbation Theory for Linear Operators.* Springer Verlag, New-York.

[LARDIES] J., AND Z. RATSIMALAHELO (2001): "Model Order Estimation of A Multivariable Stochastic Process", Submitted for publication.

[LAWLEY] D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method.* London.

[LEWBEL] A. (1991): "The rank of Demand Systems: Theory and nonparametric estimation," *Econometrica*, 59, 711-730.

[LEWBEL] A., AND PERRAUDIN (1995): "A Theorem on Portfolio Separation with General Preferences," *Journal of Economic Theory*, 65, 624-626.

[LUTKEPOHL] H., AND M. M. BURDA (1997): "Modified Wald tests under nonregular conditions," *Journal of Econometrics*, 78, 315-332.

[MOORE] D. S. (1977): "Generalized inverses, Wald's method, and the construction of chi-squared tests of fit," *Journal of American Statistical Association,* 72, 131-137.

[NISHII] R.(1988): "Maximum likelihood principle and model selection when the true model is unspecified," *Journal of Multivariate Analysis*, 27, 392-403.

[PHILLIPS] P.C.B. (1996): "Econometric model determination," *Econometrica*, 64, 763-812.

[RAO] C. R. (1973) Linear Stastical Inference and Its Applications, 2nd ed. Wiley, New-York.

[RAO] C. R. (1979). "Separation Theorems for Singular Values of Matrices and Their Applications in Multivariate Analysis," *Journal of Multivariate Analysis,* 9, 362-377.

[RATSIMALAHELO] Z., AND J. LARDIES (1998): "Comparison of Tests for Model Specification in Multivariate Time Series Analysis" in Holly and Greenblat (eds) *Computation in Economics, Finance and Engineering: Economic Systems*, Elsevier, North-Holland Publisher, pp 291-306.

[RATSIMALAHELO] Z. (2000): "Rank Test Based On Matrix Perturbation Theory", mimeo.

[RATSIMALAHELO] Z. (2001): "Specification of VARMA Echelon Form Models", in Neck R. (eds) *Proceedings IFAC Symposium on Modelling and Control of Economic Systems*, Elsevier, North-Holland Publisher, pp 129-137.

[RATSIMALAHELO] Z. (2002): "Rank Test Based On Matrix Perturbation Theory", Submitted for publication.

[RATSIMALAHELO] Z. (2003): "Strongly Consistent Determination of the Rank" of Matrix, mimeo.

[ROBIN] J-M., AND R. J. SMITH (2000): "Tests of Rank". *Econometric Theory,* 16. 151-175.

[ROBINSON] P. M (1973): "Generalized canonical analysis for time series," *Journal of Multivariate Analysis*, 3, 141-160.

[SCHOTT] J. R. (1984): "Optimal Bounds for the Distributions of Some Test Criteria for Tests of Dimensionality," *Biometrika*, 71, 561-567.

[SCHOTT] J. R. (1988): "Testing the equality of the smallest latent roots of a correlation matrix," *Biometrika*, 75, 794-796.

[SHIBATA] R. (1976): "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika*, 63, 117-126.

[SIN] C-H AND H. WHITE (1996): "Information criteria for selecting possibly misspecified parametric models", *Journal of Econometrics*, 71, 207-225.

[STEWART] G. W. (1984): Rank Degeneracy, *SIAM Journal on Scientific and Statistical Computing*, 5, 403-413.

[STEWART] G. W., AND J. G. SUN (1990): *Matrix Perturbation Theory*, Academic Press, Boston.

[VUONG] Q.H. (1987): "Generalized Inverses and Asymptotic Properties of Wald Tests," *Economics Letters*, 24, 343-347.

[VUONG] Q. H.(1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," Econometrica, 57, 307-333.

[WHITE] H. (1984): *Asymptotic Theory for Econometricians*, Academic Press, Boston.

[ZHAO] L. C.; P. R. KRISHNAIAH, AND Z., D. BAI (1986): "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, 20, 1-25.

Appendix of Proofs.

Proof of Theorem 4.

Suppose that the $(m-k)(n-k) \times (m-k)(n-k)$ covariance matrix Q have rank $v = \min\{(m-k)(n-k); mf\}$ and let $S = \begin{bmatrix} S_1, & S_2 \end{bmatrix}$ and $T = \begin{bmatrix} T_1, & T_2 \end{bmatrix}$ are the $(m-k)(n-k) \times (m-k)(n-k)$ orthogonal matrices. Then the singular value decomposition of Q is

$$Q = S_1 \Lambda_1 T_1'$$

where Λ_1 is vxv nonsingular diagonal matrix with diagonal elements $\gamma_1 \geqslant ... \geqslant \gamma_v > 0$. The Moore-Penrose generalized inverse of Q is given by

$$Q^+ = T_1 \; \Lambda_1^{-1} S_1'$$

and it is unique Golub and Van Loan (1996).

The singular value decomposition of the estimator of the covariance matrix $\widehat{Q}$ is

$$\widehat{Q} = \widehat{S}_1 \widehat{\Lambda}_1 \widehat{T}_1' + \widehat{S}_2 \widehat{\Lambda}_2 \widehat{T}_2'$$

Following Theorem 2, $\widehat{Q} \longrightarrow^p Q$ implies $\widehat{S} \longrightarrow^p S$, $\widehat{T} \longrightarrow^p T$ and $\widehat{\Lambda} \longrightarrow^p \Lambda$ and hence $\widehat{\Lambda^+} \longrightarrow^p \Lambda^+$. It follows that

$\widehat{Q^+} \longrightarrow^p Q^+$ where $\widehat{Q^+} = \widehat{T}_1 \widehat{\Lambda}_1^{-1} \widehat{S}_1'$.

Hence

$N\hat{l}'\widehat{Q}^+\widehat{l} = N\widehat{l}'\widehat{T}_1\widehat{\Lambda}_1^{-1}\widehat{S}_1'\widehat{l} \longrightarrow^d \chi^2{}_v$.

Proof of Theorem 6.

The Proof is immediate, if $(\Gamma'\Psi\Gamma) = I$ and $(\Phi\Omega\Phi') = I$ then $N^{1/2}\widehat{A^*} \longrightarrow^d N(0, I)$ and consequently $N^{1/2}l \longrightarrow^d N(0, I)$. The test statistic $L(k)$ is defined by $L_3(k) = N\hat{l}'\widehat{l} = N(vec\widehat{D}_2)'(vec\widehat{D}_2) = Ntr(\widehat{D}_2'\widehat{D}_2)$.

Proof of Corollary.

Part (i), under the null hypothesis, the $n-k$ smallest eigenvalues converge to zero, then the $L_3(k)$ is asymptotically equivalent to LR and to LM because $\ln(1 + \widehat{\lambda}_j^2) \approx \widehat{\lambda}_j^2$ and $\widehat{\lambda}_j^2 \approx \frac{\widehat{\lambda}_j^2}{1+\widehat{\lambda}_j^2}$.

Part (ii), the likelihood ratio (LR) and the Lagrange Multiplier (LM) have identical limiting distributions to that of the $L_3(k)$ statistic.

Proof of Proposition 3.

According (26.2) we have $\lambda_i o_i = \Gamma A' w_i$ premultiply by A, one obtain $\lambda_i A o_i = A\Gamma A' w_i$ then following (26.1) one gets $\lambda_i^2 \Phi^{-1} w_i = A\Gamma A' w_i$ thus we have the problem of the generalized eigenvalues $(A\Gamma A' - \lambda_i^2 \Phi^{-1}) w_i = 0$ where the $\lambda_i^2$are the root of the determinantal equation $\mid A\Gamma A' - \lambda_i^2 \Phi^{-1} \mid = 0$ and w_i are the corresponding eigenvectors of $A\Gamma A'$ in the metric of Φ^{-1}.

Similarly $A'\Phi A o_i = \lambda_i A' w_i = \lambda_i^2 \Gamma^{-1} o_i$ thus $(A'\Phi A - \lambda_i^2 \Gamma^{-1}) o_i = 0$ where the $\lambda_i^2$are the root of the determinantal equation $\mid A'\Phi A - \lambda_i^2 \Gamma^{-1} \mid = 0$ and o_i are the corresponding eigenvectors of $A'\Phi A$ in the metric of Γ^{-1}.