Rank Test Based On Matrix Pertubation Theory.

Zaka Ratsimalahelo*

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

First version July 2000. This version, November 2002

Abstract

In this paper, we propose methods of the determination of the rank of matrix. We consider a rank test for an unobserved matrix for which an estimate exists having normal asymptotic distribution of order $N^{1/2}$where N is the sample size. The test statistic is based on the smallest estimated singular values. Using Matrix Perturbation Theory, the smallest singular values of random matrix converge asymptotically to zero in the order $O(N^{-1})$ and the corresponding left and right singular vectors converge asymptotically in the order $O(N^{-1/2})$. Moreover, the asymptotic distribution of the test statistic is seen to be chi-squared. The test has advantages over standard tests in being easier to compute. 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 both approaches. Some economic applications are discussed and simulation evidence is given

*We are grateful for the comments of D. Belsley, M-C. Pichery. The first draft of this paper was presented at the sixth International Conference of the Society on Computational Economics, Barcelona, Spain, July, 2000.

for this test. Its performance is compared to that of the LDU rank tests of Gill and Lewbel (1992) and Cragg and Donald (1996).

Key words: Rank Testing; Matrix Perturbation Theory; Rank Estimation; Estimated Matrices; Singular Value Decomposition.

JEL classification: C12, C13, C30.

1 Introduction

Many econometric models are specified through the rank of a certain matrix for which a consistent estimate exists. 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)]. 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); 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. The most commonly used methods for evaluating the rank of a matrix rely upon the eigenvalue decomposition on the QR factorization [Stewart G.W. (1984)]. These tools, however, only help in making decisions and do not constitute a test. The important paper by Gill and Lewbel (1992) introduces a rank test based on the Gaussian elimination Lower-Diagonal-Upper triangular (LDU) decomposition. Unfortunately, the asymptotic distribution given for this test statistic is incorrect except for a limited special case. Cragg and Donald (1996) provides an appropriate modification. Their test has the advantage of possessing a limiting chi-squared distribution. Also, Cragg and Donald (1997) propose another test for the rank of matrix based on a minimum

chi-squared criterion.

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 will construct a test of rank based on the smallest eigenvalue of the estimated matrix. Moreover the test is more complicated when the smallest eigenvalue has multiplicity. The asymptotic null distribution of Bartlett's statistic is not chi-squared, see Schott (1988). He used an approximation based on some of its moments to obtain an asymptotically chi-squared [see also Lawley (1956)]. More recently Robin and Smith (2000) proposed a test statistics which is distributed asymptotically as a linear combination of independent chi-squared random variables rather than a chi-square. Unfortunately, the test requires the percentiles of a weighted chi-squared distribution for which computationally intensive algorithms need be used. Moreover the weights are unknown and must be estimated from the sample. Thus weight estimation introduces variability and hence potentially less accuracy, to the testing procedure. In addition, 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. We shall give also a rank test for an unobserved matrix for which a root-N-consistent estimator is available and construct a test statistics based on estimate of the smallest singular value. 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 that the test statistic is asymptotically distributed as chi-square under the null.

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, 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 shal present also an alternative approach to the information theoretic criterion.

The remainder of the paper is structured as follows: Section 2 presents the relevant material from hypothesis testing. When several singular values are zeros it is necessary to construct a basis of the null space. Section 3 presents matrix-perturbation results needed for statistical inference of the smallest singular values. Based on these analysis results, a new rank test is developed and its strong consistency is proved in section 4. Section 5 presents an alternative method to determine the rank of matrix using the information theoretic criterion. Section 6 discusses some potential economic applications. To demonstrate the usefulness of the proposed test, in section 7, we present a Monte Carlo study comparing the finite sample behaviour of the proposed test with the LDU procedure and comment on the results. Section 8 offers concluding remarks. Proofs of the fundamental lemmas, theorems, and propositions that provide the foundation of the technique are assembled in the Appendix.

The following notation is used throughout the paper: vec(A) stands for the vectorization of the $m \times n$ matrix A by stacking the columns of A. For a singular matrix C, C^+ denotes its Moore-Penrose genralized inverse. Convergence in probability is denoted "$\longrightarrow^p$" and convergence in distribution by "$\longrightarrow^d$"

2 Hypothesis testing

Consider an unobserved matrix A $(m \times n)$ with unknown true rank $k > 0$. Without loss of generality, we assume $m \geq n$. Let $\widehat{A}$ be a root-N consistent estimator of A, then

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

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

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

Let us write the perturbed matrix

$$\widehat{A} = A + \varepsilon B \tag{4}$$

where the matrix εB is the perturbation of the matrix A. 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

$$\varepsilon B = O_p(N^{-1/2}). \tag{5}$$

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}.

3 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.

Thre 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.

3.1 Introduction and general results

Let A be an $m \times n$ real matrix with $m \geq n$ and rank k, where $0 < k \leq n$. One well-conditioned means for evaluating the rank is to use the singular value decomposition (SVD) and count 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. Before, giving the results which are use below, we will specify the notation further.

Let the SVD of the $m \times n$ real matrix A be denoted.

$$A = UDV' = U_1 D_1 V_1' \tag{6}$$

where $U = [u_1 u_2 ... u_m] = [U_1, U_2]$ of order $(m \times m)$and $V = [v_1 v_2 ... v_n] = [V_1, V_2]$ 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. The columns of U and V are the left and right singular vectors of the matrix A respectively.

The SVD gives orthonormal bases for the null space and range of A. Let us define $\mathcal{N}(A) = \text{span}\ \{v_{k+1}, ..., v_n\}$ be the null space (or Kernel) of $A : V_2 \in \mathcal{N}(A)$ and $\mathcal{R}(A) = \text{span}\{u_1, ..., u_k\}$ is the range (or column) space of $A : U_1 \in \mathcal{R}(A)$. The orthogonal complement of $\mathcal{R}(A)$ is $\mathcal{N}(A') = \text{span}\{u_{k+1}, ..., u_m\} : U_2 \in \mathcal{N}(A')$ hence U_1 is the orthogonal complement of U_2. Likewise the orthogonal complement of $\mathcal{N}(A)$ is $\mathcal{R}(A') = \text{span}\{v_1, ..., v_k\} : V_1 \in \mathcal{R}(A')$ hence V_1 is the orthogonal complement of V_2.

Thus U_1 order $(m \times k)$ (respectively, V_1 order $(n \times k)$) an orthonormal basis of $\mathcal{R}(A)$ (respectively $\mathcal{R}(A')$) and U_2 order $(m \times m-k)$ (respectively,

V_2 order $(n \times n-k)$) an arbitrary orthonormal basis of $\mathcal{N}(A')$ (respectively $\mathcal{N}(A)$).

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

$$D = \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} U_1'AV_1 & U_1'AV_2 \\ U_2'AV_1 & U_2'AV_2 \end{bmatrix} \tag{7}$$

Thus

$$D_1 = U_1'AV_1 = diag(\lambda_1, \lambda_2, ..., \lambda_k)$$

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

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 $AV_2 = 0$ (V_2 span the null space of A: $V_2 \in \mathcal{N}(A)$) and $U_2'A = (A'U_2)' = 0$ (U_2 span the null space of A': $U_2 \in \mathcal{N}(A')$).

Let $\widehat{A}$ be an estimator of the matrix A, then the SVD of $\widehat{A}$ is

$$\widehat{A} = \widehat{U}\widehat{D}\widehat{V'} \tag{8}$$

and

$$\widehat{D} = \widehat{U'}\widehat{A}\widehat{V}. \tag{9}$$

We now partition the matrices conformably to the partitioning of U, V and D.

$\widehat{U} = [\ \widehat{U}_1 \quad \widehat{U}_2\ ]$; $\widehat{V} = \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix}$; and $\widehat{D} = diag(\widehat{D}_1, \widehat{D}_2)$ where $\widehat{U}_1$ has k columns and $\widehat{U}_2$ has $m-k$ columns. Then the SVD of the matrix $\widehat{A}$ can be written as

$$\widehat{A} = [\ \widehat{U}_1 \quad \widehat{U}_2\ ] \begin{bmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_2 \end{bmatrix} \begin{bmatrix} \widehat{V'}_1 \\ \widehat{V'}_2 \end{bmatrix} \tag{10}$$

The estimated matrix $\widehat{A}$ almost certainly has full rank, although it will be " close " to a matrix of rank k. The matrix diagonal can be written as

$$\widehat{D} = \begin{bmatrix} \widehat{U}_1' \\ \widehat{U}_2' \end{bmatrix} \widehat{A} \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix} = \begin{bmatrix} \widehat{U'}_1\widehat{A}\widehat{V}_1 & \widehat{U'}_1\widehat{A}\widehat{V}_2 \\ \widehat{U'}_2\widehat{A}\widehat{V}_1 & \widehat{U'}_2\widehat{A}\widehat{V}_2 \end{bmatrix} \tag{11}$$

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 Equation (5), 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).

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 $\varepsilon B = O(N^{-1/2})$. 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:*

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

$$\widehat{V}_2 = \widetilde{V}_2 + O_p(N^{-1/2}) \tag{12}$$

$$\widehat{D}_2 = \varepsilon \widetilde{U'}_2 B \widetilde{V}_2 + O_p(N^{-1}) \tag{13}$$

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. (12) shows that, given $\widehat{A}$ 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 space of A which are close to $\widehat{U}_2$ and $\widehat{V}_2$. Eq. (13) 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 (13).

In other words, Eq. (12) 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 matrix 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 (4) by $\widetilde{U_2'}$ and $\widetilde{V}_2$, respectively. Since the columns of V_2 span the null space of A, and U_2 span the null space of A' that is $V_2 \in \mathcal{N}(A)$ and $U_2 \in \mathcal{N}(A')$ implies $A\widetilde{V}_2 = AV_2P = 0$ and $\widetilde{U_2'}A = M'U_2'A = 0$, we have

$$\widetilde{U'}_2 \widehat{A} \widetilde{V}_2 = \varepsilon \widetilde{U'}_2 B \widetilde{V}_2. \tag{14}$$

Using (14) we get the following corollary

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

$$\widehat{D}_2 = \widetilde{U'}_2 \widehat{A} \widetilde{V}_2 + O_p(N^{-1}), \tag{15}$$

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

From (1), we have $\widehat{A} \longrightarrow^p A$. Since the singular values λ_i for $i = 1, ..., n$ are continuous functions of the elements of A, we have

$$\widehat{\lambda}_i \rightarrow^p \lambda_i \text{ for } all \text{ i} \tag{16}$$

where $\lambda_i = 0 \; for \; i > k$.

Therefore, from (16) $\widehat{D}_1 \longrightarrow^p D_1$ and $\widehat{D}_2 \longrightarrow^p 0$.

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

3.2 Inference statistical of the smallest singular values.

In this section, we 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:

$$\widehat{D}_2 = U_2'\widehat{A}V_2 + O_p(N^{-1}) \tag{17}$$

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

The perturbed zero singular values can be approximated by the right hand side in(17), we would propose to investigate the statistical properties of this matrix. We denote by L

$$L = U_2'\widehat{A}V_2 \tag{18}$$

[1]

Eq. (18) 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. Then the smallest singular values $\widehat{D}_2$ of $\widehat{A}$ has the same asymptotic distribution as the right hand side of Eq. (18).

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 covariance matrix given by*

[1] If we use the perturbation-dependent exact bases $\widetilde{U}_2 = U_2M$ and $\widetilde{V}_2 = V_2P$ where the matrices M and P are orthogonal respectively, then one can, also, expressed the matrix L as $L = M\widehat{D}_2P'$ a rotation of the smallest singular values around the origin. We will show in the theorem 2 that test statistic is independent with respect to the orthogonal transformation of the bases U_2 and V_2,see also Ratsimalahelo (2000).

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

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

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

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

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

The Theorem shows that the test statistic $L(k)$ 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

$\widehat{L} = \widehat{U'}_2 \widehat{A} \widehat{V}_2 = \widehat{D}_2 = diag\left(\widehat{\lambda}_{k+1}, ..., \widehat{\lambda}_n\right)$ the smallest singular values estimated.

Hence a test for $H_0 : r(A) = k$ of (2) at the α level may be based on $L(k)$ statistic rejecting the null hypothesis if $L(k)$ exceeds the a critical value of chi-square.

3.3 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 columns 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 statistics 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 will 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 columns 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 orthogonal that is $(V_2 \otimes U_2)'(V_2 \otimes U_2) = (I_{m-k} \otimes I_{n-k})$. 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.

$$r(Q) = \min\{(m-k)\times(n-k); mf\} \tag{19}$$

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

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

$\widehat{\Sigma}^+$ and $\widehat{Q}^+$are the Moore-Penrose generalized inverses of the matrices $\widehat{\Sigma}$ and $\widehat{Q} = \left(\widehat{V'}_2 \otimes \widehat{U}_2'\right)\widehat{\Sigma}\left(\widehat{V}_2 \otimes \widehat{U}_2\right)$ 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 the proposition* 2*, then the statistic test is given by*

$$L(k) = N\hat{l}'\widehat{Q}^{+}\widehat{l} \longrightarrow^{d} \chi^{2}{}_{v} \tag{20}$$

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 being the rank of Q.

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

4 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 strongly consistency.

We consider tests based on the statistic $L(k)$. 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 reject This is a sequential testing procedure of the rank of the matrix.

The test statistic $L(k)$ 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 1 *:*

1) The critical value $\gamma_k C_N$ is similar to the significance level α_N satisfying $\alpha_N \to 0$ 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 Monte-Carlo experiment, 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 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 $N \to \infty$ and let $\widehat{k} = k$ then the null hypothesis does not reject almost surely as $N \to \infty$. 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 4 *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 statistics $L(k)$ provides a strongly consistent estimate of k_0 the true rank of A.

Test Algorithm

The strategy test is performed by the following steps:

Step 1: Estimate the matrix A.

Step 2 Perform their SVD and order its singular elements by increasing values. Using the smallest singular values to obtain $\widehat{D}_2$, $\widehat{U}_2$ and $\widehat{V}_2$.

Step 3 Construct $\widehat{L}$ from $\widehat{D}_2$ and evaluate $\widehat{Q}$ by using $\widehat{\Sigma}$, $\widehat{U}_2$, and $\widehat{V}_2$.

Step 4 Construct the $p-$ value for the statistic $L(k)$

$$p - \text{value } = P(L(k) > \gamma_k C_N)$$

5 The model selection approach.

An alternative approach to estimate the rank of matrix is the information criteria , the approach is analogous to the well-known model selection criteria (SC) used for choosing between competing models.

Define the following criteria

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

where function $\varphi(k)$ and the constants C_N for $N \geq 1$ are specified by the practitioners.

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

$$\widehat{k} = \arg\min IC(k, C_N) \qquad for\ k = 0, 1, ..., n-1 \tag{21}$$

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 *Assumption C.*

In application of the $IC(k, C_N)$ criteria, the 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).

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 would give the following theorem concerning strong consistency of the estimator k.

Theorem 5 *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 $lim_{N\to\infty} \widehat{k} = k_0$ almost surely.*

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.

6 Some applications

In this section we will discuss how the methods described in this paper can be used in two applications: determining the rank of demand systems and constructing direct tests for instrumental variables.

6.1 The rank of Demand Systems

The rank of demand systems is defined as the maximum dimension of the function space spanned by Engel curves, and its knowledge can provide information about the properties of consumer preferences such as functional form, separability, and aggregability.

Lewbel (1991) has shown that a number of propositions in demand theory relate to the rank of demand systems.

Roughly, any demand system has rank R if there exist R goods such that the Engel curve of any good equals a weighted average of the Engel curves

of those R goods. For example, a demand system having all linear Engel curves is rank two, unless the Engel curves are all rays from the origin, in which case it is homothetic and hence rank one. Quadratic Engel curves can be either rank two or rank three.

Formally, the rank of any given demand system $g(p, x, m)$ is the smallest value of R such that each g_i can be written as:

$$g_i(p, x, m) = \sum_{j=l}^{R} \phi_{ij}(p, m) f_j(p, x, m)$$

for some functions ϕ_{ij} and f_j and where p is the vector of prices of goods, x the total expenditure level and m is a vector of observable demographic or other non-income characteristics that affect the preferences of the household. For a fixed value of p, the rank of a demand system is defined to equal the dimension of the space spanned by the functions g_1 , $g_2, ...g_R$. All demand systems have rank $R \leq N$, the number of goods.

Demand system rank has many theoretical implications for separability, aggregation, welfare analysis, demand system specification, and even portfolio separation in asset demands [Lewbel and Perraudin (1995)].

6.2 Direct test for Instrumental Variables

It is well known that instrumental variables (IV) estimator provides consistent, asymptotically normal estimates even in the presence of endogenous regressors. To satisfy these properties an instrument should posses two key properties: (1) relevance, that is a high correlation with that portion of the endogenous regressors that cannot be explained by other instruments, and (2) exogeneity, that is no correlation with the errors in the dependent variable.

Recent research, Shea (1997), Staiger and Stock (1997) and Hall, Rudebusch and Wilcox (1996) has emphasized that an IV estimator will have poor finite-sample performance if the instruments have low relevance for the regressors. Thus it is necessary to clarify the link between instrument relevance and model identification and to establish a test of instrument relevance.

Consider the linear regression model

$$y = X\beta + \varepsilon$$

where y is a $(T \times 1)$ vector of observations on the dependent variable, X is a $(T \times k)$ matrix of regressors with rank $(X) = k$, ε is a $(T \times 1)$ vector of observations on the error process with $E(\varepsilon) = 0$ and var $(\varepsilon) = \sigma^2 I_T$ (homoskedastic) and β is an $(k \times 1)$ vector of unknown parameters.

Suppose that the right-hand side variables are correlated with the errors, $\text{plim}(\frac{1}{T}X'\varepsilon) \neq 0$, but that there exists a set of q instruments Z, $q \geq k$ with $\text{plim}(\frac{1}{T}Z'\varepsilon) = 0$.

Then, following White (1984), we have the standard regularity conditions:
(a): $\text{plim}(\frac{1}{T}Z'X) = Q_{zx}$ a matrix of finite constants with rank k.
(b): $\text{plim}(\frac{1}{T}Z'Z) = Q_{zz}$ a matrix of finite constants with rank q.
(c): $T^{-1/2}Z'\varepsilon \longrightarrow^d N(O, \sigma^2 Q_{zz})$.

The vector of unknown parameters β is commonly estimated by instrumental variables (equivalently two stage least squares TSLS or generalized method of moments GMM). The instrumental variable estimator is

$$\beta_{IV} = (X'P_zX)^{-1}X'P_zY$$

where $P_z = Z(Z'Z)^{-1}Z'$, the projection matrix for Z.
The asymptotic distribution of β_{IV} is given by

$$T^{1/2}(\beta_{IV} - \beta) \longrightarrow^d N(O, \sigma^2 V).$$

where $V = (Q_{zx}Q_{zz}^{-1}Q'_{zx})^{-1}$.

This asymptotic distribution depends in some way on the relationship between the instruments and regressors. If the rank condition in (a) is close to being violated in finite samples then the true parameter is "close" to being unidentified and this would have an adverse effect on the distribution of the IV estimator.

Hall, Rudebusch, and Wilcox (1996) propose a measure of the relevance by expressing the IV estimator and its associated statistics in terms of canonical correlations between the regressors and the instruments. The instrument relevance depends on the moment matrix $Z'X$. Then they suggest direct tests of the rank condition based on the size of the smallest canonical correlation of the matrix $Z'X$. They also point out that, for usual asymptotic Wald statistic screening based on such a pre-test, of which the first stage R^2 and

F statistics are special cases, an intended selection bias may be introduced that worsens rather than mitigates finite sample bias.

Their tests of the rank condition use the determinantal equation and their statistics is the likelihood ratio (LR). From a matrix analysis point view, the determinant is certainly not a good tool for assessing the conditioning to singularity of a matrix. In our context it easy to use the singular value perturbation results to determine the rank of the matrix $Z'X$.

7 Some Monte Carlo evidence

In this section we will present some simulation results to demonstrate both the effectiveness of our methods and the usefulness of the procedure. We will then report the results of a Monte Carlo study of the performance of the proposed test in determining the rank of the Hankel matrix of covariance and in testing whether the order of autoregression moving average has been correctly specified. Such empirical applications to ARMA model identification have been used by Gill and Lewbel (1992) to compare their test with that of Box and Jenkins, They have shown that the result of the LDU test performed better than that of the Box and Jenkins test.

Moreover Ratsimalahelo (2001) used the L test statistic to determine the Kronecker indices in the specification of the VARMA echelon form models. Lardies and Ratsimalahelo (2001) generalised the L test statistic in determining the rank of the product of three matrices (or the matrix weighted) and we have used this generalised L test statistic to determine the order of multivariate time series models.

We will also compare the behavior of our tests with the behavior of the test based on Gaussian elimination.

Before, we will summarise the procedure of a rank test of the matrix based on Gaussian Elimination with complete pivoting of the matrix $\hat{A}$. A development can be found in Cragg and Donald (1996) see also Gill and Lewbel (1992).

A Test based on Gaussian Elimination.

Cragg and Donald (1996) propose a Wald-type procedure for testing the null hypothesis $H_{0k} : r(A) = k$ against $H_{1k} : r(A) > k$.

Using the results described in Cragg and Donald (1996, Sections 3-5), we proceed as follows.

Perform Gaussian elimination k steps with rows and columns on $\widehat{A}$ the estimator of the matrix A so that we have

$$\widehat{P}(k)\widehat{A}\widehat{Q}(k) = \begin{bmatrix} \widehat{\Omega}_{11} & \widehat{\Omega}_{12} \\ 0 & \widehat{\Omega}_{22} \end{bmatrix} \tag{22}$$

for some matrices $\widehat{P}(k)$ and $\widehat{Q}(k)$ where $\widehat{\Omega}_{22}$ is a $(m-k)\times(n-k)$ matrix. Each step of Gaussian elimination involves a possible exchange of rows and columns (in search for the largest absolute value among the diagonal elements whose row numbers are no smaller than the current one). Under the null hypothesis the matrices $\widehat{P}(k)$ and $\widehat{Q}(k)$ converge to $P(k)$ and $Q(k)$ respectively as $N \to \infty$, so Eq.(22) holds in its limit as the k-step Gaussian elimination on A.

If no rows or columns need to be exchanged in the process of Gaussian elimination, we have a simple expression for $\widehat{\Omega}_{22} = \widehat{A}_{22} - \widehat{A}_{21}\widehat{A_{11}}^{-1}\widehat{A}_{12}$ when $\widehat{A}$ is naturally partitioned. So $\widehat{\Omega}_{22}$ may be used as a test statistics of the hypothesis that the rank of A is k. As the estimated matrix $\widehat{A}$ is consistent and asymptotically normally distributed, and according to Cragg and Donald (1996), under the null hypothesis

$$\sqrt{N}vec(\widehat{\Omega}_{22}) \longrightarrow^{d} \mathcal{N}(0, V).$$

where $V = \Gamma\Sigma\Gamma'$ and Γ is a function of the submatrices obtained by partitioning A.

Then,

$$\widehat{\zeta} = Nvec(\widehat{\Omega}_{22})'\widehat{V}^{+}vec(\widehat{\Omega}_{22})$$

where $\widehat{V}^{+}$ is the Moore-Penrose generalized inverse of $\widehat{V}$.

The test statistic $\widehat{\zeta}$ is refered to Gaussian elimination test (GE) and it converges to the chi-square distribution with v degrees of freedom with v being the rank of V. This test statistic is more general than Cragg and Donald's test statistic, since they consider only the nonsingularity of the asymptotic covariance matrix V i.e. the nonsingularity of the covariance matrix Σ of the matrix $\widehat{A}$ (because the matrix Γ is full row rank). In our procedure, we consider a situation more general, the covariance matrix Σ can be of reduced rank.

There are other methods for testing the null hypothesis that the rank of the matrix A is k. First more recently Robin and Smith (2000) proposed a test statistics which is distributed asymptotically as a linear combination of independent chi-squared random variables. 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. Next Cragg and Donald (1997) proposed a test statistic based on a minimum chi-square criterion. The procedure need to minimize the objective function numerically which is often very difficult. We will not include these two last tests in our comparisons.

We now turn to the Monte-Carlo study. Data are generated according to a univariate ARMA process of the form

$$a(L)y_t = b(L)u_t \qquad t = 1, ..., T$$

The design differs with respect to the sample size T and the value of parameters $a(L)$ and $b(L)$. We would specify two different ARMA models:

-(M1)- the first model: ARMA (2,1)
$a(L) = 1 + 0.64L + 0.7L^2$
$b(L) = 1 + 0.8L.$

-(M2)- the second model: ARMA (4,2)
$a(L) = 1 - 2.76L + 3.809L^2 - 2.654L^3 + 0.924L^4$
$b(L) = 1 - 0.2L + 0.04L^2$

In the two models, the errors are considered Gaussian white noise with zero mean and unit variance. These two models are considered as a starting-point for investigation of the performance of the different methods. In model M1, the roots of the polynomials $a(z)$ and $b(z)$ are inside the unit circle where z is a complexe variable. Thus the process y_t is stationary. For model M2, the situation is more delicate, the roots of $a(z)$ are situated much closer to the unit circle, the process y_t is nonstationary.

An estimate the rank of sample Hankel matrices is as follows : For each sample size, a sample Hankel matrix $\hat{A} = \hat{H}_p$ where the $(2p-1, 2p-1)$ matrix is composed of the estimated covariance matrices of the process. We consider $p = 4$ for first model (M1) and $p = 5$ for the second model (M2).

The matrix Σ is defined as the covariance matrix of $vec(\hat{H})$. Then a consistent estimate of Σ is obtained by replacing all quantities in the summation by their sample counterparts. However, for a small sample size, this will turn out to be quite a poor estimate of Σ. Accordingly in the LDU method we use a parameter to truncate the summation.

We used three different sample sizes T: 100, 300 and 1000. For each simulated sample, 200 initial observations have been discarded, to minimize the effect of initial condition. The first observation is set equal to the mean of the process, zero. The results are based on 1000 independent replications. In using the L test statistic, C_N was chosen to be $\sqrt{\log N}$. Clearly, $C_N = \sqrt{\log N}$ grows very slowly as N increases. In the specification of the HQIC criterion, we take $c = 2.02$.

We report the results that measure the performance of the model selection criteria SC (AIC, BIC, HQIC), L and GE procedures. For each procedure, we calculated the probabilities that the procedure

1. estimated correctly the true order r^0
2. overestimated the true order
3. underestimated the true order.

In the first case the probability refers to the percentage of times that the correct order is estimated for a certain number of runs, each on independent realisation of the data. In the other cases, (over-and underestimate the true order), the probabilities are defined in a similar manner. A procedure with a high probability to estimate correctly r^0 leads to a performance of the rank test method.

For each of these tests we computed the probability of rejecting the null at 5% nominal significance levels. Each probability is calculated as percentage of times in 1000 replications that the value of the test statistic exceeds the 5% critical value of the appropriate asymptotic distribution. All computations were carried out using the MATLAB programming language.

7.1 Monte Carlo results

Now we shall present the Monte Carlo results for the probabilities of estimate the order. Table 1 reports the probabilities for SC-AIC, SC-BIC, SC-HQIC, L and GE for the first model (M1) and for three different sample sizes N equal to (100, 300, 1000). For SC-BIC, the probability of selecting the true order r^0 increases from 0.732 to 0.853 to 0.990; while the probability of underestimate declines from 0.156 to 0.118 to 0.000. For SC-HQIC, the

corresponding changes are from 0.740 to 0.843 to 0.982 and from 0.136 to 0.098 to 0.000 respectively. For L, the corresponding changes are from 0.950 to 0.989 to 0.992 and from 0.015 to 0.001 to 0.000 respectively.. For GE, the corresponding changes are from 0.795 to 0.976 to 0.985 and from 0.130 to 0.004 to 0.000 respectively..

The probabilities of SC-AIC are much less sensitive to the sample size N than are those of the other four procedures. As the sample size N increases from 100 to 300 to 1000, the probability of selecting the true order by SC-AIC changes from 0.747 to 0.754 to 0.756 and the probability of the underestimate decreases from 0.048 to 0.037 to 0.000. The fact that the probability does not increase toward one as N increases, reflects the inconsistence of the SC-AIC procedure. For the smallest sample size SC-AIC performed better than the other two SC procedures. But for larger samples sizes, it does not perform as well as the other two SC.

Our results suggest that SC-AIC work best for the smallest sample size whereas SC-BIC, SC-HQIC work best for all other sample sizes. GE and L perform very well for all sample sizes. The GE procedure showed a behaviour very similar to that of the L procedure and its performance very closely. It is only for very smallest samples sizes, that a small difference in performance is visible.

Next, we shall consider the second model (M2). The motivation here is to evaluate the performance of the various procedures when the process is nonstationary. Table 2 presents the probabilities for SC-AIC, SC-BIC, SC-HQIC, L and GE for three different sample sizes N equal to (100, 300, 1000). The true model is in fact nonstationary ARMA(4, 2) with the order $r = 4$. The consistent procedures (SC-BIC, SC-HQIC, L and GE) show a clear and rapid tendency to converge towards a correct rank even if the process is nonstationary. But the performance of SC-AIC changes relatively little.

Overall, there are three main points that can be drawn from the probabilities of selecting the true order. The SC-AIC based correct decision display practically no variability across all samples sizes, pointing to the correct order approximately 55-75% of the times across all sample sizes. It has a strong tendency to overrank and no tendency to improve as N increases. The two other SC procedures (SC-BIC and SC-HQIC) show a relatively good performance. Finally The L and GE procedures appear to perform best in all-around sense, they showed a very similar behaviour.

Table 1

Model M1: ARMA(2,1)

True rank k = 2, the true parameter values: $a_1 = 0.64; a_2 = 0.7; b_1 = 0.8$					
	SC-AIC	SC-BIC	SC-HQIC	L	GE
Sample size: N = 100					
correct	0.747	0.732	0.740	0.950	0.795
overestimate	0.205	0.112	0.124	0.035	0.075
underestimate	0.048	0.156	0.136	0.015	0.130
Sample size: N = 300					
correct	0.754	0.853	0.843	0.989	0.976
overestimate	0.209	0.029	0.059	0.009	0.020
underestimate	0.037	0.118	0.098	0.001	0.004
Sample size: N = 1000					
correct	0.756	0.990	0.982	0.992	0.985
overestimate	0.244	0.010	0.018	0.008	0.015
underestimate	0.000	0.000	0.000	0.000	0.000

Table 2.

Model M2: ARMA(4,2)

True rank k = 4,
the true parameter values: $a_1 = -2.76; a_2 = 3.809; a_3 = -2.654; a_4 = 0.924;$
$b_1 = -0.2; b_2 = 0.04$

	SC-AIC	SC-BIC	SC-HQIC	L	GE
Sample size: N = 100					
Correct	0.452	0.405	0.456	0.412	0.388
Overestimate	0.303	0.095	0.149	0.000	0.000
Underestimate	0.245	0.500	0.395	0.588	0.612
Sample size: N = 300					
Correct	0.570	0.694	0.714	0.796	0.805
Overestimate	0.333	0.086	0.111	0.000	0.000
Underestimate	0.097	0.220	0.175	0.204	0.195
Sample size: N = 1000					
Correct	0.575	0.950	0.916	0.955	0.945
Overestimate	0.330	0.005	0.030	0.000	0.000
Underestimate	0.095	0.045	0.054	0.045	0.055

8 Conclusions

In this paper we proposed new methods for determining the rank of an unobserved matrix for which a root-N consistent estimator is available. It is based on matrix perturbation results applied to estimated matrix. It allows to handle more general situations where the covariance matrix is definite positive or semidefinite positive. Conditions for strongly consistency of sequential testing strategy and information theoretic criterion procedures are obtained. The performance of the methods is assessed by means of Monte-Carlo experiments to determine the order of scalar ARMA processes. Its performance appears to be similar to that of LDU decomposition.

The two methods gave similar results. Accordingly the proposed procedure provides an alternative to the LDU decomposition, which from a practical point of view involve a much higher computational burden. The test procedure which should be preferred in the present ARMA order determination context is the simpler one: test the smallest singular values of the Hankel matrices of increasing order. It can then be implemented without performing a full singular values decomposition of the different Hankel matrices since only the smallest singular values are needed. This drastically reduces the complexity of the test.

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.

The methods proposed in this paper may be particularly useful in selecting the correct moment condition for a generalized method of moments (GMM). The procedures also can consistently determine whether there is a sufficient number of correct moment conditions to identify the unknown parameters of interest.

References

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

[Anderson] T., W. (1984). An introduction of multivariate statistical analysis, 2eds, Wiley, New-York.

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

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

[Camba-Mendez] G. and Kapetanios G. (2001) Testing the rank of the Hankel covariance matrix. A statistical approach. *IEEE Transactions On Automatic Control,* Vol 46, No 2, 331-336.

[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*, Vol. 91, No 435, 1301-1309.

[Cragg] J., G. and S., G. Donald (1997). Inferring the Rank of Matrix, *Journal of Econometrics*, Vol 76, No 2, 223-250.

[Field] C; (1993). Tail areas of linear combinations of chi-squares and non-central chi-squares, *Journal of Statistical Computation and Simulation*, Vol 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,* Vol. 87, No , 766-776.

[Golub] G., H. and Van Loan, C. F. (1996): *Matrix Computations*, 3eds John Hopkins University Press. Baltimore.

[Hall] A., R., Rudebusch G., D. and Wilcox D., W. (1996). Judging Instrument Relevance in Instrumental Variables Estimation, *International Economic Review*, Vol 37, No 2, 283-298.

[Hannan] E., J. and Quinn B., G. (1979). The determination of the order of an autoregression, *Journal of Royal Statistical Society Series B* 41, 190-195.

[Hannan] E., J. and Deistler M. (1988). *The statistical Theory of Linear Systems*. New York, Wiley.

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

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

[Lardies J.] and Ratsimalahelo, Z. (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*, Vol 59, 711-730.

[Lewbel] A. and Perraudin (1995). A Theorem on Portfolio Separation with General Preferences, *Journal of Economic Theory*, 65, 624-626.

[Potscher] B., M. (1983). Order Estimation in ARMA-models by Lagrangian Multiplier Tests, *Annals of Statistics*, 11, 872-885.

[Potscher] B., M. (1989). Model selection under nonstationarity, autoregressive models and stochastic linear regression models. *Annals of Statistics*, 17, 1257-1274.

[Potscher] B., M. (1991). Effects of Model Selection on Inference, *Econometric Theory* 7, 163-185.

[Ratsimalahelo] Z. and Lardies J. (1998). Comparison of Tests for Model Specification in Multivariate Time Series Models in *Symposium in Computation in Economics, Finance and Engineering: Economic Systems*, Greenblatt and Holly, Eds. Elsevier, North-Holland, pp 291-306.

[Ratsimalahelo] (2000) Rank Test Based On Matrix Perturbation Theory, mimeo

[Ratsimalahelo] Z. (2001). Specification of VARMA Echelon Form Models, in Reinhard Neck ed. "*Modelling and Control of Economic Systems*" Proceedings IFAC Symposium (SME, 2001), Elsevier, North-Holland, pp 186-191.

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

[Schwarz] G. (1978) Estimating the dimension of a Model, *Annals of Statistics*, 6, 461-464.

[Shea] J. (1997). Instrument relevance in multivariate linear models: a simple measure, *Review of Economics and Statistics*, Vol. LXXIX, No. 2, 348-352.

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

[Sin] C-H and White H. (1996). Information criteria for selecting possibly misspecified parametric models, *Journal of Econometrics*, 71, 207-225.

[Staiger] D. and Stock J., H. (1997). Instrument variables regressions with weak instruments, *Econometrica*, 65, 557-586.

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

[Stewart] G., W. and Sun Ji Guang (1990). *Matrix Perturbation Theory*. Academic Press, Boston.

[White] H. (1984). *Asymptotic Theory for Econometricians*, Academic Press.

[Zhao] L., C.; Krishnaiah P. R. and Bai Z. D. (1986). On detection of the number of signals in presence of white noise, *Journal of Multivariate Analysis*, Vol 20, no 1, 1-25.

Appendix of Proofs.

Proof of Proposition 1.
Following the SVD of the matrix A, we have

$$U'AV = D = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{A1}$$

Let us consider the perturbed matrix

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

where $\varepsilon = N^{-1/2}$ and for large $N, 0 < \varepsilon < 1$.

The SVD of the matrix $\widehat{A}$ corresponding $\widehat{U}, \widehat{V}, \widehat{D}$ such that

$$\widehat{U}'\widehat{A}\widehat{V} = \widehat{D} = \begin{bmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_2 \end{bmatrix} \tag{A2}$$

with $\widehat{U} = \begin{bmatrix} \widehat{U}_1 & \widehat{U}_2 \end{bmatrix}$ and $\widehat{V} = \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix}$.

We will indicate mainly how U_2, V_2 and D_2, the zero singular elements are perturbed under the assumption that the zero singular value D_2 of multiplicity $n-k$ is well separated from the others, *i.e.*

$$min\lambda_i >> \varepsilon \ for \quad i = 1, ..., k \tag{A3}$$

This means that the gap between the zero and non-zero singular values must be order of magnitude greater than ε.

Applying the orthogonal transformations U and V to $\widehat{A}$, one obtains

$$U'\widehat{A}V = D + \varepsilon U'BV = D + \varepsilon G$$

where

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} U_1'BV_1 & U_1'BV_2 \\ U_2'BV_1 & U_2'BV_2 \end{bmatrix}$$

we can be written

$$U'\widehat{A}V = \begin{bmatrix} D_1 + \varepsilon G_{11} & \varepsilon G_{12} \\ \varepsilon G_{21} & \varepsilon G_{22} \end{bmatrix} \tag{A4}$$

a matrix that no longer possesses a block diagonal structure.

Before we try to recover this structure, let us introduce the SVD of the order $(n-k)$ matrix G_{22}:

$$G_{22} = U_2' B V_2 = M \Lambda_{22} P'$$

where M and P are orthogonal matrices and Λ_{22} is a diagonal matrix such that

$$\Lambda_{22} = M' U_2' B V_2 P \tag{A5}$$

The two modified orthogonal transformations are thus defined:

$$\widetilde{U} = [\ U_1 \quad \widetilde{U}_2\ ]; with\ \widetilde{U}_2 = U_2 M$$

$$\widetilde{V} = \begin{bmatrix} V_1 & \widetilde{V}_2 \end{bmatrix}; with\, \widetilde{V}_2 = V_2 P \tag{A6}$$

Since we only changed the bases of the null-space, we still have $\widetilde{U'} A \widetilde{V} = D$ as in (A1), but in addition, these new bases (A4) can be rewritten as

$$\widetilde{U'} \hat{A} \widetilde{V} = D + \varepsilon \widetilde{G} = \begin{bmatrix} D_1 + \varepsilon G_{11} & \varepsilon \widetilde{G}_{12} \\ \varepsilon \widetilde{G}_{21} & \varepsilon \Lambda_{22} \end{bmatrix} \tag{A7}$$

What we did is specify perturbation dependent bases $\widetilde{U}_2$ and $\widetilde{V}_2$ for the null-spaces of AA' and $A'A$ in place of generic bases U_2, and V_2. To this purpose we block diagonalized the second member of (A7). We would propose to use two further orthogonal transformations, R and Q which we would define in their partitioned form:

$$R = I + \varepsilon F + O_p(\varepsilon^2) = I + \varepsilon \begin{bmatrix} 0 & F_{12} \\ -F_{21} & 0 \end{bmatrix} + O_p(\varepsilon^2)$$

where $F_{12} = -D_1^{-1} \widetilde{G'}_{21}$

The matrix R is a perturbed identity matrix which allows us to diagonalize it up to first order terms in ε, and similarly

$$Q = I + \varepsilon H + O_p(\varepsilon^2) = I + \varepsilon \begin{bmatrix} 0 & H_{12} \\ -H_{21} & 0 \end{bmatrix} + O_p(\varepsilon^2)$$

where $H_{12} = -D_1^{-1}\widetilde{G}_{12}$.

One verifies that these matrices are orthogonal up to second order in ε. So that pre-and post-multiply equation (A7) by R' and Q, respectively, yields

$$R'\widetilde{U}'\widehat{A}\widetilde{V}Q = \begin{bmatrix} D_1 + \varepsilon\Lambda_{11} & 0 \\ 0 & \varepsilon\Lambda_{22} \end{bmatrix} + O_p(\varepsilon^2) \tag{A8}$$

Therefore

$$\widetilde{U}R = [\ U_1 - \varepsilon F_{12}' \quad \widetilde{U}_2 + \varepsilon U_1 F_{12}\ ] + O_p(\varepsilon^2) \tag{A9}$$

$$\tilde{V}Q = [\ V_1 - \varepsilon\widetilde{V}_2 H_{12}' \quad \widetilde{V}_2 + \varepsilon V_1 H_{12}\ ] + O_p(\varepsilon^2)$$

Comparing the relations (A8) and (A9) with (A2), we have

$$\widehat{U}_2 = \widetilde{U}_2 + \varepsilon U_1 F_{12} + O_p(\varepsilon^2)$$

$$\widehat{V}_2 = \widetilde{V}_2 + \varepsilon V_1 H_{12} + O_p(\varepsilon^2).$$

$$\widehat{D}_2 = \varepsilon\Lambda_{22} + O_p(\varepsilon^2)$$

Since the matrices U_1 and V_1 are orthonormal respectively, then U_1 and V_1 are order $O_p(1)$. Following the assumption (A3) ${D_1}^{-1}$ is bounded, $O_p(1)$ which implies that F_{12} and H_{12} are order $O_p(1)$. Consequently, the second terms in these first two equations, are order $O_p(\varepsilon)$. We then have

$$\widehat{U}_2 = \widetilde{U}_2 + O_p(\varepsilon)$$

$$\widehat{V}_2 = \widetilde{V}_2 + O_p(\varepsilon)$$

$$\widehat{D}_2 = \varepsilon \widetilde{U'}_2 B \widetilde{V}_2 + O_p(\varepsilon^2).$$

Proof of Theorem 1.

Let $l = vecL$, the vectorization of the matrix L

$$N^{1/2} l = ({V_2}' \otimes U_2') N^{1/2} vec\widehat{A}$$

The mean is null

$$E(N^{1/2} l) = ({V_2}' \otimes U_2') E(N^{1/2} vec(\widehat{A})) = N^{1/2}({U_2}' A V_2) = 0$$

because $AV_2 = 0$, the column of V_2 span the null space of A
And the covariance is

$$\begin{aligned} V(N^{1/2} l) &= NE(ll') = E\left[(V_2' \otimes U_2')(N^{1/2} vec\widehat{A})(N^{1/2} vec\widehat{A})'(V_2 \otimes U_2)\right] \\ &= (V_2' \otimes U_2') \left\{ E\left[\left(N^{1/2} vec\widehat{A} \right) (N^{1/2} vec\widehat{A})' \right] \right\} (V_2 \otimes U_2) \end{aligned}$$

Following the Eq.(1), one obtain

$$E[((N^{1/2} vec\widehat{A})(N^{1/2} vec\widehat{A})'] = \Sigma$$

thus the covariance matrix of the vector $N^{1/2} l$ is

$$V(N^{1/2} l) = Q = ({V_2}' \otimes U_2')\Sigma(V_2 \otimes U_2\,).$$

Proof of Theorem 2.

First, we will show that $L(k)$ does not depend on the orthogonal transformations on the bases U_2 and V_2. Consider the perturbation-dependent exact

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

Let $\widetilde{L} = \widetilde{U'}_2 \widehat{A} \widetilde{V}_2 = (U_2 M)' \widehat{A} (V_2 P)$ by the vectorization of the matrix $\widetilde{L}$, we get

$$vec\widetilde{L} = (P'V_2' \otimes M'U_2')vec\widehat{A} = (P' \otimes M')(V_2' \otimes U_2')vec\widehat{A}$$

This can be written by

$$vec\widetilde{L} = (P' \otimes M')vecL$$

The asymptotic covariance matrix $\widetilde{Q}$ of $N^{1/2} vec\widetilde{L}$ is equal to $\widetilde{Q} = (P' \otimes M')Q(P \otimes M)$ and the quadratic form is

$$\begin{aligned} \widetilde{L}(k) &= N(vec\widetilde{L})'[\widetilde{Q}]^{-1}(vec\widetilde{L}) \\ &= N(vecL)'(P \otimes M)[(P' \otimes M')Q(P \otimes M)]^{-1}(P' \otimes M')(vecL) \end{aligned}$$

Since $(P \otimes M)$ is an orthogonal matrix, thus we have

$$\begin{aligned} \widetilde{L}(k) &= N(vecL)'(P \otimes M)[(P' \otimes M')Q^{-1}(P \otimes M)](P' \otimes M')(vecL) \\ &= N(vecL)'Q^{-1}(vecL) \\ &= L(k). \end{aligned}$$

We have thus established that $L(k)$ is invariant with respect to the orthogonal transformations on the bases U_2 and V_2.

Next, without loss of generality, we will take in the sequel that the orthogonal matrices M and P to be the identity matrix.

Let $\widehat{U}_2$ and $\widehat{V}_2$ be the estimate of U_2 and V_2, then following proposition 1, we have :

$\widehat{U}_2 \rightarrow^p U_2$ and $\widehat{V}_2 \rightarrow^p V_2$ so that $\widehat{U}_2$ and $\widehat{V}_2$ are consistent estimator of U_2 and V_2 respectively. We get also $\left(\widehat{V}_2' \otimes \widehat{U}_2'\right) \longrightarrow^P (V_2' \otimes U_2')$.

Let $\widehat{\Sigma}$ be the estimate of Σ, since $\widehat{A} \longrightarrow^P A$, clearly we have $\widehat{\Sigma} \longrightarrow^P \Sigma$.

As the asymptotic variance Q is full rank, then

$$\widehat{Q} = \left(\widehat{V'}_2 \otimes \widehat{U}_2'\right) \widehat{\Sigma} \left(\widehat{V}_2 \otimes \widehat{U}_2\right) \longrightarrow^P Q = (V_2' \otimes U_2')\Sigma(V_2 \otimes U_2)$$

So $\widehat{Q}$ is a consistent estimator of Q.

Finally, since the rank of the asymptotic covariance matrix Q is equal to $r(Q) = (m-k) \times (n-k)$, and $\widehat{Q}$ is a consistent estimator of Q, then the quadratic form $Nvec(\widehat{L})'\widehat{Q}^{-1}vec(\widehat{L})$ converges in distribution to a chi-square random variable.

We note that $\widehat{L} = \widehat{U}_2'\widehat{A}\widehat{V}_2 = \widehat{D}_2$ is the diagonal matrix of the $n-k$ smallest singular values of $\widehat{A}$. As $\widehat{A}$ is consistent estimator of A then $\widehat{L}$ is consistent estimator of L (also by Slutsky theorem).

Proof of Proposition 2.

First, we establish Proposition 2(i). Note that 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. In fact if $r(\widehat{\Sigma})$ converges almost surely to $r(\Sigma)$ that is $P[r(\widehat{\Sigma}) = r(\Sigma)] \to 1$ as $N \to \infty$ and since $\widehat{\Sigma} \longrightarrow^P \Sigma$ then the Moore-Penrose inverse satisfies $\widehat{\Sigma}^+ \longrightarrow^P \Sigma^+$, (see Andrew (1987, Theorem 2)).

Now Proposition 2(ii), the $(mn) \times (m-k)(n-k)$ matrix $(V_2 \otimes U_2)$ has full column rank $(m-k)(n-k)$ since U_2 and V_2 are full column rank $(m-k)$ and $(n-k)$ respectively. Moreover, the columns of $(V_2 \otimes U_2)$ are orthonormal, so that $(V_2' \otimes U_2')(V_2 \otimes U_2) = (I_{m-k} \otimes I_{n-k}) = I_{(m-k)(n-k)}$ hence $(V_2 \otimes U_2)^+ = (V_2' \otimes U_2')$. Therefore the Moore-Penrose generalized inverse of Q is

$$Q^+ = [(V_2' \otimes U_2')\,\Sigma\,(V_2 \otimes U_2)]^+ = (V_2' \otimes U_2')\,\Sigma^+\,(V_2 \otimes U_2)\,.$$

Following proposition 1, $\left(\widehat{V}_2' \otimes \widehat{U}_2'\right) \longrightarrow^P (V_2' \otimes U_2')$ and since $r\left(\widehat{V}_2' \otimes \widehat{U}_2'\right) \longrightarrow^P (m-k)(n-k)$, then the convergences almost surely of the rank of the matrix $\widehat{\Sigma}$ to the rank of the matrix Σ imply that $r(\widehat{Q})$ converges almost surely to $r(Q)$. Thus, since $\widehat{Q} \longrightarrow^P Q$, we have $\widehat{Q}^+ \longrightarrow^P Q^+$ where $\widehat{Q}^+ = [\left(\widehat{V}_2' \otimes \widehat{U}_2'\right)\widehat{\Sigma}\left(\widehat{V}_2 \otimes \widehat{U}_2\right)]^+ = \left(\widehat{V}_2' \otimes \widehat{U}_2'\right)\widehat{\Sigma^+}\left(\widehat{V}_2 \otimes \widehat{U}_2\right)$.

Proof of Theorem 3.

Following the result of Moore (1977) see e.g. Andrews (1987) for any consistent estimator $\widehat{Q}^+$of Q^+, the quadratic form $N\ l'\widehat{Q}^+l$ will have an asymptotic chi-squared distribution with v degrees of freedom .

Proof of Theorem 4.

To prove the strong consistency, we need to show the following

1) The null hypothesis $H_0 : \widehat{k} = 1, 2, ..., k-1$ are rejected almost surely as $N \to \infty$ this is equivalent to $lim_{N\to\infty} L(k)/C_N \succ \gamma_k$.

2) However, the null hypothesis $H_0 : \widehat{k} = k$ is accepted almost surely as $N \to \infty$ or equivalent to $lim_{N\to\infty} L(k)/C_N \leq \gamma_k$.

Following the A*ssumption LIL,* if $k \prec k_0$ then there exists a constant $c \succ 0$ such that

$$\widehat{l}'\widehat{Q}^{+}\widehat{l} \geq c$$

and on the other hand for $k \succeq k_0$ we have

$$\widehat{l}'\widehat{Q}^{+}\widehat{l} = O(\log\log N/N)$$

Firstly, assuming that $\widehat{k} \prec k_0$, and C_N satisfies the *Assumption C*, then we have

$$\lim_{N\to\infty} \frac{L(k)}{C_N} = \lim_{N\to\infty} \frac{N}{C_N}\widehat{l}'\widehat{Q}^{+}\widehat{l} = \infty \quad a.s.$$

where the convergence holds by *Assumption C(ii)* and *a.s.* denote the abbreviate "almost surely" (i.e., with probability one). This result means that, for any positive number γ_k, $L(k)/C_N \succ \gamma_k$ almost surely, as $N \to \infty$, thus the null hypothesis is rejected with probability one.

Next let us assume that $\widehat{k} = k_0$. This implies

$$\lim_{N\to\infty} \frac{L(k)}{C_N} = \lim_{N\to\infty} \frac{O(\log\log N)}{C_N} = 0 \quad a.s.$$

by *Assumption C(iii).* This result means that, for any positive number γ_k, $L(k)/C_N \prec \gamma_k$ almost surely, as $N \to \infty$.

Proof of Theorem 5.

To prove the strong consistency of $\widehat{k}$, i.e. $P(lim_{N\to\infty}\ \widehat{k} = k_0) = 1$ or $lim_{N\to\infty}\ \widehat{k} = k_0$ *a.s.* It is enough to prove that the underestimation and the overestimation of the true rank is not possible asymptotically.

We can assert with probability one that for large N, for $k \leq k_0$

$$L(k) - L(k_0) \geqslant cN$$

where $c > 0$ is a constant.

On the other hand, under the assumption for $k > k_0$, we have

$$\begin{aligned} L(k) &= O(\log\log N) \\ L(k_0) &= O(\log\log N) \end{aligned}$$

Case 1. We will prove that if $k < k_0$ then the underestimation of the true rank is not possible asymptotically.

$$\begin{aligned} IC(k, C_N) - IC(k_0, C_N) &= L(k) - L(k_0) + [\varphi(k) - \varphi(k_0)]C_N \\ &\geqslant c + [\varphi(k) - \varphi(k_0)]\frac{C_N}{N} \quad \forall c > 0 \end{aligned}$$

Since the second term on the right-hand side converge to zero as N tends to infinity from *Assumption C(ii)* and since $c > 0$ thus asymptotically we have $IC(k, C_N)) - IC(k_0, C_N) > 0$ *a.s.* This implies $liminf\ \widehat{k} \geq k_0$ *a.s.*, for $N \to \infty$.

This result means that the minimum cannot be attained at $k < k_0$ and the underestimation of the true rank is not possible asymptotically.

Next we will show that the true rank k_0 will be preferred to $k \geq k_0$, asymptotically, this is the case of overestimation of the true rank .

Case 2. We will prove that if $k \geqslant k_0$ then the overestimation of the true rank is not possible asymptotically.

$$\begin{aligned} IC(k, C_N) - IC(k_0, C_N) &= O\left(\log\log N\right) + [\varphi(k) - \varphi(k_0)]C_N \\ \frac{IC(k, C_N) - IC(k_0, C_N)}{C_N} &= O\left(\frac{\log\log N}{C_N}\right) + [\varphi(k) - \varphi(k_0)] \end{aligned}$$

The term $\frac{\log\log N}{C_N}$tends to zero from *Assumption C(iii)* and $C_N > 0$ and $\varphi(k_0) < \varphi(k)$ then asymptotically, we have $IC(k, C_N) - IC(k_0, C_N) > 0$ *a.s.* This implies $limsup\ \widehat{k} \leq k_0$ *a.s* for $N \to \infty$. Consequently, the overestimation of the true rank does not occur asymptotically.

Finally, taking the above two cases, it imply that $lim_{N\to\infty}\ \widehat{k} = k_0$ *a.s.* The minimum of the criteria function is attained at $\widehat{k} = k_0$.