STATISTICAL TESTS AND ESTIMATORS OF THE RANK OF A MATRIX AND THEIR APPLICATIONS IN ECONOMETRIC MODELLING

by Gonzalo Camba-Méndez
and George Kapetanios

STATISTICAL TESTS AND ESTIMATORS OF THE RANK OF A MATRIX AND THEIR APPLICATIONS IN ECONOMETRIC MODELLING[1]

by Gonzalo Camba-Méndez[2]
and George Kapetanios[3]

CONTENTS

Abstract

Testing and estimating the rank of a matrix of estimated parameters is key in a large variety of econometric modelling scenarios. This paper describes general methods to test for and estimate the rank of a matrix, and provides details on a variety of modelling scenarios in the econometrics literature where such methods are required. Four different methods to test the true rank of a general matrix are described, as well as one method that can handle the case of a matrix subject to parameter constraints associated with definiteness structures. The technical requirements for the implementation of the tests of rank of a general matrix differ and hence there are merits to all of them that justify their use in applied work. Nonetheless, we review available evidence of their small sample properties in the context of different modelling scenarios where all, or some, are applicable.

Keywords: Multiple Time Series, Model Specification, Tests of Rank.
JEL classification: C12, C15 and C32.

Non-technical summary

Tests and estimators of the rank of a matrix are key in a large variety of statistical and econometric multivariate modelling scenarios. The purpose of this paper is to describe some general methods to test and estimate the rank of a matrix and review their use in econometric modelling.

In most cases tests of rank are carried out on matrices of parameter estimates rather that data matrices. Of course the particular context of such tests varies greatly but certain common threads are discernible. Most models that rely on rank deficient parameter matrices do so in order to reduce the channels of effects from one set of variables to another. In this sense many instances of rank reduction can be related to factor structures where a small number of observed or unobserved factors affect a larger set of variables.

There is a large variety of modelling scenarios where these tests of rank are useful for specification purposes. The modelling scenarios range from linear and stationary models such as standard VARs, factor analysis, dynamic factor models, instrumental variable estimation, and dynamic principal component models, to nonlinear frameworks such as nonparametric factor models and also to nonstationary frameworks such as cointegrated systems. We conclude that these methods are of increasing relevance given the focus of econometric and statistical work on multivariate systems.

Four different methods to test the true rank of a general matrix are described, as well as one method that can handle the case of a matrix subject to parameter constraints associated with definiteness structures. Alternative methods for the estimation of the rank of a matrix that do not use statistical tests but information criteria methods are also reviewed.

The technical requirements for the implementation of the tests of rank of a general matrix differ and hence there are merits to all of them that justify their use in applied work. Nonetheless, we review available evidence of their small sample properties in the context of different modelling scenarios where all, or some, are applicable. Monte Carlo evidence suggests that statistical tests of rank may have an advantage over standard information criteria methods for a number of modelling scenarios. Additionally, Monte Carlo evidence reviewed in this paper suggests that bootstrapped procedures of those tests of rank significantly improved upon the performance of the corresponding asymptotic tests.

1 Introduction

Tests and estimators of the rank of a matrix are key in a large variety of statistical and econometric multivariate modelling scenarios. In most cases tests of rank are carried out on matrices of parameter estimates rather that data matrices. Of course the particular context of such tests varies greatly but certain common threads are discernible. Most models that rely on rank deficient parameter matrices do so in order to reduce the channels of effects from one set of variables to another. For example, reduced rank VAR models restrict the coefficient matrices of a VAR model to have reduced rank so as to reduce the number of channels via which lags of variables can affect their present values. In this sense many instances of rank reduction can be related to factor structures where a small number of observed or unobserved factors affect a larger set of variables. The purpose of this paper is to describe some general methods to test and estimate the rank of a matrix and review their use in econometric modelling.

For a general $m \times n$ matrix $\boldsymbol{A}$, the problem is to identify its unknown true rank which will be denoted by $\rho[\boldsymbol{A}] = r^*$, where $0 \leq r^* \leq \min(m, n)$, and $\rho[.]$ denotes the rank of a matrix. For a sample of size T, we define an estimate of $\boldsymbol{A}$ by $\hat{\boldsymbol{A}}$. This paper reviews different methods to test the true rank of a general matrix not subject to parameter constraints associated with definiteness structures, and one method that can handle such parameter constraints. The technical requirements for the implementation of the tests of rank of a general matrix differ and hence there are merits to all of them that justify their use in applied work. Nonetheless, we review available evidence of their small sample properties in the context of different modelling scenarios where all, or some, are applicable. The tests can be used as building blocks for estimators of the rank of a matrix. We discuss estimators based on tests of rank as well as estimators based on information criteria.

The structure of the paper is as follows. Section 2 reviews various methods to test the null hypothesis $H_0 : \{\rho[\boldsymbol{A}] = r^*\}$ against the alternative hypothesis $H_1 : \{\rho[\boldsymbol{A}] > r^*\}$ in the case of a general matrix. Section 3 concentrates on testing this same hypothesis when $\boldsymbol{A}$ is a hermitian positive semidefinite matrix (with $m = n$). Section 4 reviews sequential testing procedures for the null hypothesis $H_0 : \{\rho[\boldsymbol{A}] = r\}$ against the alternative hypothesis $H_1 : \{\rho[\boldsymbol{A}] > r\}$ for $r = 0, 1, \ldots, \min(m, n)$ that estimate the true rank, r^* of $\boldsymbol{A}$. This section also reviews some plausible information criteria methods for this estimation problem. Section 5 presents a large variety of modelling scenarios where the methods we discuss are of immediate relevance. In this context, this section further reviews available evidence on

the small sample properties of the tests of rank discussed in section 2. Finally, Section 6 concludes.

2 Rank of a General Matrix

This section reviews four different methods to test the rank of general matrix $\boldsymbol{A}$. The first method is a minimum discrepancy test proposed by Cragg and Donald (1997) and will be denoted by MD. Implementation of this method relies on assuming that by an application of some suitable central limit theorem $\sqrt{T}vec(\hat{\boldsymbol{A}} - \boldsymbol{A}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$, where $\boldsymbol{V}$ is non-singular. It is further assumed that a consistent estimate of $\boldsymbol{V}$ is available. The second test has been proposed by Cragg and Donald (1996). It is based on the implementation of gaussian elimination on matrix $\boldsymbol{A}$ and will hence be denoted as GE. This method also requires the existence of an estimate of $\boldsymbol{V}$ but it is not necessary for it to be non-singular, however, knowledge of its true rank is needed. The third method, proposed by Robin and Smith (2000), is computed from the characteristic roots of a quadratic form built from $\boldsymbol{A}$, and will be denoted as CRT. Once more an estimate of $\boldsymbol{V}$ is needed but it can be rank deficient and its rank unknown. The fourth test, proposed by Bartlett (1947) and denoted by BA in the text, does not rely at all on the existence of an estimate of $\boldsymbol{V}$.

Before starting the discussion of the tests it is worth making a comment on computational aspects of calculating the rank of an observed matrix. In particular, we note that we abstract from issues concerned with rank calculation due to rounding errors that arise from the finite precision of computer based matrix computations. This is a large area that is both nontrivial and interesting especially for large matrices. Issues related to such matrices have recently come to the fore with the increased availability of large datasets in econometrics. For discussions on rounding errors see Golub and Van Loan (1996, sec. 2.4) and Highham (1996).

2.1 A Minimum Discrepancy Function Test

This section presents a minimum discrepancy function (MDF) method to test whether a $q \times 1$ parameter vector $\boldsymbol{\theta}$ can be represented as a function of a $p \times 1$ parameter vector $\boldsymbol{\mu}$ where $p < q$. That is, a test of the null hypothesis $H_0 : \{\boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\mu}_0)\}$, where $\boldsymbol{\mu}_0$ is used to denote the true value of $\boldsymbol{\mu}$. We further make the following assumption:

Assumption 1 *a. There exists a consistent estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, such that $\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega})$, and where $\boldsymbol{\Omega}$ is a non-singular matrix.*

b. $\boldsymbol{\mu}$ belongs to a specified compact parameter space $\boldsymbol{\Xi} \subset \Re^p$ that contains $\boldsymbol{\mu}_0$. The

parameter space $\boldsymbol{\Xi}$ contains a neighborhood $\boldsymbol{\Psi}$ of $\boldsymbol{\mu}_0$ in which $\boldsymbol{h}(\boldsymbol{\mu})$ has continuous second partial derivatives, and the $q \times p$ Jacobian matrix $\boldsymbol{\Delta} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\mu}'}$ at $\boldsymbol{\mu}_0$ is of full column rank p.

c. $\boldsymbol{h}(\boldsymbol{\mu}^*) = \boldsymbol{h}(\boldsymbol{\mu}_0)$ for $\boldsymbol{\mu}^* \in \boldsymbol{\Xi}$ implies $\boldsymbol{\mu}^* = \boldsymbol{\mu}_0$.

The assumptions above are the usual regularity conditions for a minimum discrepancy type test being chi-squared distributed. A minimum discrepancy function test statistic could thus be formulated as:

$$MD = T \min_{\mu} \left\{ \left(\hat{\boldsymbol{\theta}} - \boldsymbol{h}(\boldsymbol{\mu}) \right)' \boldsymbol{\Omega}^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{h}(\boldsymbol{\mu}) \right) \right\} \qquad (1)$$

and the following result follows:

Proposition 1 . *Under Assumption 1 above and under the null H_0, it holds that i) the minimizer $\hat{\boldsymbol{\mu}} \xrightarrow{a.s.} \boldsymbol{\mu}_0$, and ii) $MD \xrightarrow{p} \chi^2_{q-p}$, where χ^2_{q-p} denotes the χ^2 distribution with degrees of freedom $q - p$.*

Proof: See Chamberlain (1982, Propositions 6 and 8). $\square$

It remains to show that this testing strategy can be applied to the problem of testing the rank of a matrix. We define for this purpose $\boldsymbol{\theta} = vec(\boldsymbol{A})$, and note that assuming $m < n$, under H_0 it is possible, after a certain reordering of the columns, to write the last $n - r^*$ columns of $\boldsymbol{A}$ as a linear function of the first r^* columns.[1] This allows us to write $\boldsymbol{A} = [\boldsymbol{A}_1 \ \boldsymbol{A}_1\boldsymbol{S}]$, where $\boldsymbol{A}_1$ and $\boldsymbol{S}$ are matrices of dimension $m \times r^*$ and $r^* \times (n - r^*)$ respectively. A test of rank of a matrix is then a test of the null hypothesis $\boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\mu})$, where $\boldsymbol{\mu} = (\boldsymbol{a_1}', \boldsymbol{s}')'$ and:

$$\boldsymbol{h}(\boldsymbol{\mu}) = \begin{pmatrix} \boldsymbol{a}_1 \\ (\boldsymbol{S}' \otimes \boldsymbol{I}) \boldsymbol{a}_1 \end{pmatrix} \quad \boldsymbol{\Delta}_h = \begin{pmatrix} \boldsymbol{I}_{mr} & \boldsymbol{0}_{mr,r(n-r)} \\ (\boldsymbol{S}' \otimes \boldsymbol{I}_m) & (\boldsymbol{I}_{n-r} \otimes \boldsymbol{A}_1) \end{pmatrix} \qquad (2)$$

where $\boldsymbol{s} = vec(\boldsymbol{S})$ and $\boldsymbol{a}_1 = vec(\boldsymbol{A}_1)$. The parameter constraints imposed by $\boldsymbol{h}(\boldsymbol{\mu})$ as defined in (2), to test for the rank of a matrix are in line with the functional constraints stated in assumption 1 above. This issue has been addressed in Cragg and Donald (1997). The MD statistic would have in this case a limiting chi-square distribution with $(m - r^*)(n - r^*)$ degrees of freedom. The implementation of this method relies of course on further assuming that by an application of some suitable central limit theorem $\sqrt{T}vec(\hat{\boldsymbol{A}} - \boldsymbol{A}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$, and that a consistent estimate of $\boldsymbol{V}$ is available.

---

[1]The reordering can be accomplished by using the pivoting matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ obtained from the r steps of Gaussian elimination, see Golub and Van Loan (1996). To avoid excessive notation pivoting matrices will be ignored in this section. For details on Gaussian elimination with complete pivoting see Cragg and Donald (1996) or Golub and Van Loan (1996).

2.2 Cragg and Donald (1996)

The procedure proposed by Cragg and Donald (1996) is based on the transformation of the matrix $\boldsymbol{A}$ using Gaussian elimination with complete pivoting[2]. r^* steps of Gaussian elimination with full pivoting on matrix $\boldsymbol{A}$ amounts to the following operations:

$$\boldsymbol{Q}_{r^*}\boldsymbol{R}_{r^*}\boldsymbol{Q}_{r^*-1}\boldsymbol{R}_{r^*-1}\ldots\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\ldots\boldsymbol{C}_{r^*-1}\boldsymbol{C}_{r^*} = \left[\begin{array}{cc} \boldsymbol{A}_{11}(r^*) & \boldsymbol{A}_{12}(r^*) \\ \boldsymbol{0} & \boldsymbol{A}_{22}(r^*) \end{array}\right]$$

where $\boldsymbol{R}_i$ and $\boldsymbol{C}_i$ are pivoting matrices for step i and $\boldsymbol{Q}_i$ are Gauss transformation matrices. The pivoting matrices used to perform the first r^* steps of Gaussian elimination are applied to $\boldsymbol{A}$ to obtain the following relation

$$\boldsymbol{R}_{r^*}\boldsymbol{R}_{r^*-1}\ldots\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\ldots\boldsymbol{C}_{r^*-1}\boldsymbol{C}_{r^*} = \boldsymbol{R}\boldsymbol{A}\boldsymbol{C} = \boldsymbol{F} = \left[\begin{array}{cc} \boldsymbol{F}_{11}(r^*) & \boldsymbol{F}_{12}(r^*) \\ \boldsymbol{F}_{21}(r^*) & \boldsymbol{F}_{22}(r^*) \end{array}\right]$$

where $\boldsymbol{F}$ is partitioned accordingly, i.e. $\boldsymbol{F}_{11}(r^*)$ is of dimension $r^* \times r^*$. Note that in this case $\boldsymbol{F}_{11}(r^*)$ has full rank, under the null hypothesis that $\rho[\boldsymbol{A}] = r^*$. It then follows, (see Cragg and Donald (1996)), that $\boldsymbol{F}_{22}(r^*) - \boldsymbol{F}_{21}(r^*)\boldsymbol{F}_{11}^{-1}(r^*)\boldsymbol{F}_{12}(r^*) = 0$. The estimated counterpart of the above relation, i.e. $\hat{\boldsymbol{F}}_{22} - \hat{\boldsymbol{F}}_{21}\hat{\boldsymbol{F}}_{11}^{-1}\hat{\boldsymbol{F}}_{12} = \hat{\boldsymbol{\Lambda}}_{22}(r^*)$, may be used as a test statistic of the hypothesis that the rank of $\boldsymbol{A}$ is r^*. Under regularity conditions, including the requirement that $\sqrt{T}vec(\hat{\boldsymbol{A}} - \boldsymbol{A}) \overset{d}{\to} N(\boldsymbol{0}, \boldsymbol{V})$ where $\boldsymbol{V}$ has full rank, the following result can be shown, under H_0.

$$\sqrt{T}vec(\hat{\boldsymbol{\Lambda}}_{22}(r^*)) \overset{d}{\to} N(\boldsymbol{0}, \boldsymbol{\Gamma}\boldsymbol{V}\boldsymbol{\Gamma}')$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Phi}_2 \otimes \boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_1 = \left[-\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1} \quad \boldsymbol{I}_{m-r^*}\right]\boldsymbol{R}$, $\boldsymbol{\Phi}_2 = \left[-\boldsymbol{F}_{12}'\boldsymbol{F}_{11}^{-1'} \quad \boldsymbol{I}_{n-r^*}\right]\boldsymbol{C}'$ and $\overset{d}{\to}$ denotes convergence in distribution. Then,

$$GE = Tvec\,\hat{\boldsymbol{\Lambda}}_{22}(r^*)'(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Gamma}}')^{-1}vec\,\hat{\boldsymbol{\Lambda}}_{22}(r^*) \overset{d}{\to} \chi^2_{(m-r^*)(n-r^*)}$$

where $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{V}}$ are the sample estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{V}$ and χ^2_l denotes the χ^2 distribution with l degrees of freedom. This test computes the inverse of the covariance matrix $\boldsymbol{V}$. However, in many modelling scenarios this matrix is singular. The use of a generalized inverse, $\boldsymbol{V}^+$, may still be feasible in some instances.[3] Extension to such cases is stated in the following proposition.

---

[2]The foundations behind this strategy follow the work of Gill and Lewbel (1992). The asymptotic distribution of the test suggested by Gill and Lewbel (1992) was incorrect, nonetheless, it provided researchers with an ingenious strategy to test for the rank.

[3]In this context it is important to note that $Prob(\hat{\boldsymbol{V}} = \boldsymbol{V}) \to 1$ as $T \to \infty$ does not guarantee $Prob(\hat{\boldsymbol{V}}^+ = \boldsymbol{V}^+) \to 1$ as $T \to \infty$. This is due to the fact that generalized inverses are not continuous. Andrews (1987) has shown that the condition $Prob(\rho[\hat{\boldsymbol{V}}] = \rho[\boldsymbol{V}]) \to 1$ as $T \to \infty$ is a sufficient condition to avoid this issue. To enforce this condition sometimes it is possible to follow the solution suggested in Lütkepohl and Burda (1997), namely that if the rank of $\boldsymbol{V}$ is r^v (known), then use as an estimator $\hat{\boldsymbol{V}}_{r^v} = \hat{\boldsymbol{E}}\hat{\boldsymbol{\Lambda}}_{r^v}\hat{\boldsymbol{E}}'$, where $\hat{\boldsymbol{E}}$ is a matrix with the eigenvectors of $\hat{\boldsymbol{V}}$, and $\hat{\boldsymbol{\Lambda}}_{r^v} = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_{r^v}, 0, \ldots, 0)$, where $\hat{\lambda}_j$ for $j = 1, \ldots, r^v$ are the r^v largest eigenvalues of $\hat{\boldsymbol{V}}$.

Proposition 2 *Under the general conditions in Cragg and Donald (1996), if additionally the rank of $\boldsymbol{V}$ is known and $\rho\left[\hat{\boldsymbol{V}}\right] = \rho\left[\boldsymbol{V}\right]$, $\forall T$, then*

$$GE^g = T vec\ \hat{\boldsymbol{\Lambda}}_{22}(r^*)'(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Gamma}}')^+ vec\ \hat{\boldsymbol{\Lambda}}_{22}(r^*) \xrightarrow{d} \chi^2_\varpi$$

where $^+$ denotes the Moore-Penrose inverse of a matrix, and the number of degrees of freedom ϖ is given by the minimum between the number of rows in $\hat{\boldsymbol{\Gamma}}$ and the rank of $\hat{\boldsymbol{V}}$.

Proof: See Camba-Mendez and Kapetanios (2001) □

2.3   Robin and Smith (2000)

The testing procedure suggested by Robin and Smith (2000) focuses on the eigenvalues of quadratic forms of $\boldsymbol{A}$. The quadratic form $\boldsymbol{\Upsilon A \Pi A}'$ where $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ are positive definite matrices, is considered. It follows that $\rho\left[\boldsymbol{A}\right] = \rho\left[\boldsymbol{\Upsilon A \Pi A}'\right] = r^*$, and therefore this quadratic form has $\min(m,n) - r^*$ zero eigenvalues. Additionally, the eigenvalues of the estimator of the above quadratic form converge in probability to their population counterparts. Robin and Smith (2000) consider the statistic

$$CRT = T\sum_{i=r^*+1}^{\min(m,n)} \hat{\lambda}_i$$

where $\hat{\lambda}_i$ are the eigenvalues of $\hat{\boldsymbol{\Upsilon}}\hat{\boldsymbol{A}}\hat{\boldsymbol{\Pi}}\hat{\boldsymbol{A}}'$ in descending order, $\hat{\boldsymbol{\Upsilon}}$ and $\hat{\boldsymbol{\Pi}}$ are estimates of $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ respectively. Under the null hypothesis, the above statistic converges in distribution to a weighted sum of independent χ^2_1 random variables. The weights are given by the eigenvalues of $(\boldsymbol{D}'_{r^*} \otimes \boldsymbol{C}'_{r^*})\boldsymbol{V}(\boldsymbol{D}_{r^*} \otimes \boldsymbol{C}_{r^*})$, τ_i, $i = 1, \ldots, (m-r^*)(n-r^*)$. $\boldsymbol{D}_{r^*}$ and $\boldsymbol{C}_{r^*}$ are $n \times (n-r^*)$ and $m \times (m-r^*)$ matrices containing the eigenvectors corresponding to the $n-r^*$ and $m-r^*$ smallest eigenvalues of $\boldsymbol{\Pi A}'\boldsymbol{\Upsilon A}$ and $\boldsymbol{\Upsilon A \Pi A}'$ respectively. The sample counterparts of the above matrices may be obtained straightforwardly to estimate the asymptotic distribution of the test statistic. A few comments are in order for this test. Choices for $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ are not discussed in much detail by Robin and Smith (2000). This choice can depend crucially on the application considered. An obvious choice that can be made irrespective of application is to set both $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Pi}$ equal to the identity. Robin and Smith (2000) also consider another choice for their Monte Carlo but they do not elaborate on their motivation. Finally, it is worth noting that Robin and Smith (2000) claim that a big advantage of their test is that neither full nor known rank for $\boldsymbol{V}$ is needed or, therefore, assumed.

2.4   Bartlett (1947)

Applicability of this test to the problem of testing the rank of matrix $\hat{\boldsymbol{A}}$ relies on whether it is possible to define two random vectors $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$, such that $\boldsymbol{A} = E\{\boldsymbol{y}_t \boldsymbol{x}'_t\}$. That being

the case, it is possible to make use of a well known result in canonical correlation analysis; namely, that given two random stationary vector series $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ of dimensions m and n respectively, the rank of the covariance matrix between those two random vectors is equal to the number of nonzero canonical correlations, see Anderson (2003) for further details. Define the matrices $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)'$, compute the QR decomposition of the matrices $\boldsymbol{Y}$ and $\boldsymbol{X}$, i.e. $\boldsymbol{Y} = \boldsymbol{Q}_1 \boldsymbol{R}_1$ and $\boldsymbol{X} = \boldsymbol{Q}_2 \boldsymbol{R}_2$. The canonical correlations between the vectors $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$, are the singular values of $\boldsymbol{Q}_1' \boldsymbol{Q}_2$. We denote the canonical correlations as ρ_i, $i = 1, \ldots, \min(m,n)$. Bartlett (1947) provided a likelihood ratio criterion for testing the null hypothesis that the last $r_{\min(m,n)} - r^*$ canonical correlations are zero, i.e., $H_0 : \rho_{r^*+1} = \cdots = \rho_{\min(m,n)} = 0$. Under the null hypothesis and assuming stationarity

$$BA = \left[\frac{m+n+1}{2} - T\right] \ln \prod_{i=r^*+1}^{\min(m,n)} \left(1 - \hat{\rho}_i^2\right) \xrightarrow{d} \chi^2_{(m-r^*)(n-r^*)}$$

Fujikoshi (1974) proved that this test procedure is based on the likelihood ratio method. Bartlett's test was developed under independence and normality assumptions, but his result remains valid asymptotically following arguments by Kohn (1979) on the likelihood ratio tests for dependent observations.

Lawley (1959) provided a Bartlett (scale) correction to the LR statistic, the moments of which equal those of the nominal asymptotic chi-square distribution, apart from errors of order T^{-2}. We refer to this corrected test as the BC test. Under H_{0,r^*}, and assuming for simplicity that $m < n$,

$$BC = [(T - r^*) - \frac{1}{2}(m+n+1) + \sum_{i=1}^{r^*} \hat{\lambda}_i^{-2}] \sum_{i=r^*+1}^{m} \ln(1 + \hat{\lambda}_i^2)$$

has a limiting chi-square distribution with $(m - r^*)(n - r^*)$ degrees of freedom, and where $\hat{\lambda}_i = \hat{\rho}_i/(1 - \hat{\rho}_i^2)^{\frac{1}{2}}$; see Glynn and Muirhead (1978).

Before concluding this section, it is instructive to briefly investigate the theoretical power properties of the MD, GE and CRT procedures.[4] As all these tests are consistent, the use of local alternatives is of relevance. Since the procedures are based on different properties of rank deficient matrices, we need to provide a common framework. The null hypothesis in our framework is given by

$$H_0 : \boldsymbol{A} = \boldsymbol{\Psi}, \quad \rho\left[\boldsymbol{\Psi}\right] = r^*$$

and the alternative by

$$H_1 : \boldsymbol{A} = \boldsymbol{\Psi} + \boldsymbol{B}, \quad \rho\left[\boldsymbol{\Psi} + \boldsymbol{B}\right] > r^*$$

---

[4]We not do explore the BA test here as it is useful in more specific circumstances that the rest of the tests, since it is only applicable if it is possible to define two random vectors $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$, such that $\boldsymbol{A} = E\{\boldsymbol{y}_t \boldsymbol{x}_t'\}$.

Then, the local alternatives may be expressed as

$$H_{1T} : \boldsymbol{A} = \boldsymbol{\Psi} + \frac{\boldsymbol{B}}{\sqrt{T}}, \quad \rho\left[\boldsymbol{\Psi} + \frac{\boldsymbol{B}}{\sqrt{T}}\right] > r^* \; \forall \text{ finite } T$$

Proposition 4 in the appendix provides local power results for this local alternative hypothesis. The local power of the three tests depends on the true parameters of the model considered and therefore it is not possible to provide a general conclusion concerning their relative asymptotic performance. In general, the power of the test rises as the elements of $\boldsymbol{B}$ deviate further from zero. Also, again intuitively, the influence of each element of $\boldsymbol{B}$ depends inversely on the relative variance of each element of $\hat{\boldsymbol{A}}$.

3 Rank of a Hermitian Positive Semidefinite Matrix

In what follows we assume that in the following partition of $\boldsymbol{A}$ the $r^* \times r^*$ submatrix $\boldsymbol{A}_{11}$ is of full rank.

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$$

If $\boldsymbol{A}_{11}$ is not initially of full rank r^*, a valid reordering of the columns and rows of $\boldsymbol{A}$ would guarantee this without affecting the overall rank of the matrix. As stated above, Cragg and Donald (1996) proposed the application of r^* steps of Gaussian elimination with complete pivoting on $\boldsymbol{A}$ to achieve the required result. This manipulation guarantees that $\boldsymbol{A}_{11}$ in the finally reordered matrix is of full rank r^*. In the case of the hermitian positive semidefinite matrix we need to preserve the symmetry of $\boldsymbol{A}$ and hence symmetric pivoting should be implemented.[5] Without lack of generality we avoid the issue of pivoting in this section for ease of notation.

Given the linear dependance of the last $n - r^*$ columns on the first r^* columns it must hold that $\boldsymbol{\Lambda} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} = \boldsymbol{0}$. This implies that a test of rank $H_0 : \rho[\boldsymbol{A}] = r^*$ is equivalent to a test of the null hypothesis $H_0 : \boldsymbol{\Lambda} = \boldsymbol{0}$. Camba-Mendez and Kapetanios (2005a) show that $\boldsymbol{\Lambda} = \boldsymbol{0}$ if and only if $\Lambda_{i,i} = 0$, $i = 1, \ldots, n - r^*$ where $\Lambda_{i,i}$ denotes the i-th diagonal element of $\boldsymbol{\Lambda}$. This simplifies the test because it is thus only necessary to concentrate on testing the null hypothesis $H_0 : \boldsymbol{\theta} = 0$ where $\boldsymbol{\theta} = (\Lambda_{1,1}, \ldots, \Lambda_{n-r^*,n-r^*})'$. Under the null hypothesis we show in the appendix that $\sqrt{T}\, vec(\hat{\boldsymbol{\Lambda}}) \xrightarrow{d} N^C(\boldsymbol{0}, \boldsymbol{W})$ where $\boldsymbol{W}$ is defined in the appendix. Hence

$$\sqrt{T}\, \hat{\boldsymbol{\theta}} = \sqrt{T}\, \boldsymbol{L} vec(\hat{\boldsymbol{\Lambda}}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{L}\boldsymbol{W}\boldsymbol{L}')$$

---

[5]An algorithm to compute the factorization $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}' = \boldsymbol{G}\bar{\boldsymbol{G}}'$, where $\boldsymbol{P}$ is an $n \times n$ pivoting matrix and $\boldsymbol{G}$ is an $n \times r^*$ lower triangular matrix is available in the LINPACK, see Dongarra, Bunch, Moler, and Stewart (1979), and subroutine CCHDC for details.

where $\boldsymbol{L}$ is a $n - r \times (n - r^*)^2$ selector matrix that picks the diagonal elements of $\hat{\boldsymbol{\Lambda}}$. Then, using the results of Kudo (1963) we can construct the test statistic for the null hypothesis $H_0 : \boldsymbol{\theta} = 0$ against the alternative $H_1 : \theta_i \geq 0$, $i = 1, \ldots n - r^*$ where at least one inequality is strict. This is stated as follows:

Proposition 3 *Under the null hypothesis, $H_0 : \rho[\boldsymbol{A}] = r^*$ the test statistic, $CK = T \, \hat{\boldsymbol{\theta}}' \boldsymbol{\Psi}^{-1} \hat{\boldsymbol{\theta}}$, where $\boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{W}\boldsymbol{L}'$, is distributed as a weighted mixture of χ_q^2, $q = 1, \ldots, n - r^*$, i.e.*

$$Pr\left(\bar{\chi}^2 \geq \bar{\chi}_0^2\right) = \sum_{q=0}^{n-r^*} w_q Pr\left(\chi_q^2 \geq \bar{\chi}_0^2\right)$$

where $\chi_0^2 = 0$, and w_q are nonnegative weights.

Proof: See Camba-Mendez and Kapetanios (2005a) □

Following results in Kudo (1963) these weights are given by:

$$w_i = \sum_{Q_i} P\{(\boldsymbol{\Omega}_{Q_i'})^{-1}\} P\{\boldsymbol{\Omega}_{Q_i:Q_i'}\} \tag{3}$$

where the summation runs over all subsets Q_i of $K = \{1, \ldots, q\}$ of size i, and Q_i' is the complement of Q_i where $\boldsymbol{\Omega}_{Q_i}$ is the variance matrix of θ_j, $j \in Q_i$, and $\boldsymbol{\Omega}_{Q_i:Q_i'}$ is the same under the condition $\theta_j = 0$, $j \notin Q_i$, and $P\{\boldsymbol{\Omega}\}$ is the probability that the variables distributed in a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}$ are all positive; finally, $P\{\boldsymbol{\Omega}_{\emptyset:K}\} = 1$ and $P\{(\boldsymbol{\Omega}_{K'})^{-1}\} = P\{(\boldsymbol{\Omega}_\emptyset)^{-1}\} = 1$. The probabilities in (3) can be easily computed by means of the algorithm proposed in Sun (1988). Note that a simple expression for $\boldsymbol{\Omega}_{Q_i:Q_i'}$ is given by $\boldsymbol{\Omega}_{Q_i} - \boldsymbol{\Omega}_{Q_i,Q_i'} \boldsymbol{\Omega}_{Q_i'}^{-1} \boldsymbol{\Omega}'_{Q_i,Q_i'}$ where $\boldsymbol{\Omega}_{Q_i,Q_i'}$ is the covariance matrix of θ_j, $j \in Q_i$ and θ_k, $k \in Q_i'$ (see e.g. Anderson (2003, pp. 33-35)). It is worth noting that the multivariate one sided test has been generalized by Kudo and Choi (1975) to cases where $\boldsymbol{\Psi}$ is singular. A generalization of the test of rank presented here hence also follows.

4 Methods to identify the true rank

In the previous sections we have discussed tests for the null hypothesis that the rank of a matrix is equal to a particular value. This section discusses the related problem of estimating consistently the rank of a matrix. We consider two classes of estimators. The first class considers estimators that are based on a sequence of tests of rank. The second class is based on information criteria.

4.1 Sequential Testing Methods

Starting with the null hypothesis of $r = 1$, a sequence of tests is performed. If the null hypothesis is rejected, r is augmented by one and the test is repeated. When the null cannot

be rejected, r is adopted as the estimate of the rank of $\boldsymbol{A}$. However, the rank estimate provided by this approach will not converge in probability to the true value of the rank of the matrix r^*. The reason is that even if the null hypothesis tested is true, the testing procedure will reject it with probability α, where α is the chosen significance level. The rank estimate will converge to its true value, r^*, as T goes to infinity, if α is made to depend on T and goes to zero as T goes to infinity but not faster than a given rate. We denote this α by α_T, where the subscript T now denotes dependence of the significance level on the sample size. Hosoya (1989) shows that if α_T goes to zero as the sample size T goes to infinity and also $\lim_{T \to \infty} \ln \alpha_T / T = 0$, then the rank estimate provided by the sequential testing procedure will converge in probability to r^*, see also Cragg and Donald (1997) and Potcher (1983).

Although we have couched the problem in the form of a test, we also review methods that rely on information criteria to determine the rank of a matrix.

4.2 Information Criteria Methods

Information Criteria methods to test for the rank of a matrix can be defined. These method suggest to choose the rank r that minimizes a criterion function that takes the form:

$$IC(r) = TL + f(T)F(r)$$

where L denotes the log of the pseudo maximum likelihood estimator of $\boldsymbol{A}$ subject to its rank being restricted to r, $F(r)$ denotes the number of freely estimated parameters. Alternative specifications have been proposed for $f(T)$. Akaike (1976) adopted the formulation $f(T) = 2$, and their criteria is usually denoted as AIC. Schwarz (1978) proposed $f(T) = ln(T)$ and the standard notation for this criterion is BIC. Hannan and Quinn (1979) used $f(T) = 2 * ln(ln(T))$, and the notation used is HQ. Note that these criteria penalizes models with large number of parameters, and by extension large rank, and favor parsimonious representations.

Akaike (1974) and Akaike (1976) showed that the number of linearly independent components of the projections of the previously defined $\boldsymbol{y}_t$ onto the linear space spanned by the components of $\boldsymbol{x}_t$ is identical to the number of nonzero canonical correlations between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$. When both $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ are Gaussian, canonical correlation analysis between $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ is equivalent to maximum likelihood estimation of the linear model: $\boldsymbol{y}_t = \boldsymbol{\Psi} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t$, see Anderson (2003). The number of free parameters for this model is: $F(r) = \{[m(m+1)]/2\} + \{[n(n+1)]/2\} + r(m+n-r)$ where m denotes the dimension of the vector $\boldsymbol{y}_t$ and n de-

notes the dimension of $\boldsymbol{x}_t$. The first two terms are the number of free parameters of the covariance matrices of $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ respectively, and the last term gives the number of free parameters in matrix $\boldsymbol{\Psi}$. The value of pseudo likelihood is defined as $L = \ln \prod_{i=1}^{r}(1 - \hat{\rho}_i^2)$. where $\hat{\rho}_i$ are the estimated canonical correlation coefficients previously defined. Note that, as discussed in Anderson (2003, pp. 505), when $\rho_i = 0$ then $\hat{\rho}_i^2 = O_p(T^{-1})$, implying that $\ln(1 - \hat{\rho}_i^2) = O_p(T^{-1})$ where $O_p(.)$ denotes order in probability. This suggests that there is a positive probability that AIC will be minimised for some $r > r^*$ since the probability that $T \sum_{i=r^*+1}^{r} \ln(1 - \hat{\rho}_i^2) < 2(F(r^*) - F(r))$ is greater than zero. Therefore, the estimated rank will not converge in probability to r^* when AIC is used. The penalty used by BIC is much more severe than that used by AIC. In fact, it is easy to see that the rank estimate obtained by BIC will converge in probability to r^*. Nevertheless, BIC is likely to underestimate the rank in small samples.

Information criteria rank selection methods can also be formulated with the elements of the MDF test of rank. Cragg and Donald (1997) showed that information criterion methods defined with $TL = MD$ and $F(r) = r(m + n - r)$ provided also a consistent method to search for the rank of a matrix.

5 Applications of tests of rank

5.1 Identification and Specification of IV Models

5.1.1 Theoretical Considerations

Cragg and Donald (1993) studied the problem of identifiability and specification in instrumental variable models. For Ordinary Least Square Estimators to yield consistent estimates, the error terms must be orthogonal to the regressors. This condition is violated in the context of simultaneous equation models. These models can be written in their *structural form* as,

$$\boldsymbol{B}\boldsymbol{y}_t = \boldsymbol{\Gamma}\boldsymbol{x}_{1t} + \boldsymbol{\varepsilon}_t \tag{4}$$

where $\boldsymbol{y}_t$ is a m-vector of endogenous variables, $\boldsymbol{x}_{1t}$ is a k_1-vector of predetermined variables, $\boldsymbol{\varepsilon}_t$ is a m-vector random process of zero mean and covariance matrix $\boldsymbol{\Omega}$ and $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are matrices of parameters of dimension $m \times m$ and $m \times k_1$ respectively. Saying that $\boldsymbol{x}_{1t}$ is predetermined means that $E(\boldsymbol{x}_{1t}\boldsymbol{\varepsilon}_t') = \boldsymbol{0}$. In the context of this section we further assume that the matrix of second moments of predetermined variables, e.g. $E(\boldsymbol{x}_{1t}\boldsymbol{x}_{1t}')$, is nonsingular. Alternatively, the model could be written in *reduced form* as:

$$\boldsymbol{y}_t = \boldsymbol{\Pi}\boldsymbol{x}_{1t} + \boldsymbol{u}_t$$

where $\boldsymbol{\Pi} = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}$ and $\boldsymbol{u}_t$ is a zero mean m-vector random process of zero mean and, assuming that $\boldsymbol{B}$ is invertible, covariance matrix $\boldsymbol{B}^{-1}\boldsymbol{\Omega}\boldsymbol{B}^{-1\prime}$. Estimation of the m equations in (4) by means of Least Square is not feasible due to the non orthogonality of some of the regressors. There is no orthogonality problem though in estimating the system in its reduced form. The only problem with this strategy is that it may not always be possible to recover the structural parameter matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ from the relationship $\boldsymbol{B}\boldsymbol{\Pi} = \boldsymbol{\Gamma}$. This is referred to as the problem of identification and is well documented in the literature. Conditions for identification usually translate into zero restrictions for some of the elements of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$. Write the first equation in (4) as:

$$y_{1t} = -\boldsymbol{b}'_{12}\boldsymbol{y}_{2t} + \boldsymbol{\gamma}'_1\boldsymbol{x}_{1t} + \varepsilon_{1t} \tag{5}$$

where we have partitioned $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ in line with y_{1t} and $\boldsymbol{y}_{2t}$ as follows,

$$\boldsymbol{B} = \left[\begin{array}{cc} b_{11} & \boldsymbol{b}'_{12} \\ \boldsymbol{b}_{21} & \boldsymbol{B}_{22} \end{array}\right] \quad \boldsymbol{\Gamma} = \left[\begin{array}{c} \boldsymbol{\gamma}'_1 \\ \boldsymbol{\Gamma}_2 \end{array}\right]$$

where it is further assumed that $b_{11} = 1$. If we assume that there are no zero restrictions on the $\boldsymbol{\gamma}_1$ there is an identification problem. In this setting, it is necessary to find a vector of instrumental variables, $\boldsymbol{x}_{2t}$, uncorrelated with y_{1t} but correlated with $\boldsymbol{y}_{2t}$. The dimension of $\boldsymbol{x}_{2t}$ should be at least $\bar{m} \geq m - 1$. This condition on the number of instruments is usually referred to as the *order condition*. We could then write equation (5) as:

$$y_{1t} = -\boldsymbol{b}'_{12}\boldsymbol{y}_{2t} + \boldsymbol{\gamma}'_1\boldsymbol{x}_{1t} + \boldsymbol{\delta}'_1\boldsymbol{x}_{2t} + \varepsilon_{1t} \tag{6}$$

where the central specification hypothesis is that the $\bar{m} \times 1$ parameter vector $\boldsymbol{\delta}_1$ is equal to a vector of zeroes. We could define the vector $\boldsymbol{x}_t = (\boldsymbol{x}'_{1t}, \boldsymbol{x}'_{2t})'$, and estimate the reduced form:

$$\boldsymbol{y}_t = \boldsymbol{K}\boldsymbol{x}_t + \boldsymbol{u}_t$$

If the *rank condition* stated below is satisfied, then the structural parameters can be recovered from $\boldsymbol{B}\boldsymbol{K} = \boldsymbol{\Gamma}^*$, where

$$\boldsymbol{\Gamma}^* = \left[\begin{array}{cc} \boldsymbol{\gamma}'_1 & \boldsymbol{\delta}'_1 \\ \boldsymbol{\Gamma}_2 & \boldsymbol{\Delta}_2 \end{array}\right]$$

In particular, for equation (6) these are given by,

$$\begin{aligned} \boldsymbol{k}'_{11} &= \boldsymbol{\gamma}'_1 - \boldsymbol{b}'_{12}\boldsymbol{K}_{21} \\ \boldsymbol{k}'_{12} &= \boldsymbol{\delta}'_1 - \boldsymbol{b}'_{12}\boldsymbol{K}_{22} \end{aligned}$$

where, as before, $\boldsymbol{K} = \left[\begin{array}{cc} \boldsymbol{k}'_{11} & \boldsymbol{k}'_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{array}\right]$ has been partitioned in four blocks, comformable with y_{1t} and $\boldsymbol{y}_{2t}$ for the rows and conformable with $\boldsymbol{x}_{1t}$ and $\boldsymbol{x}_{2t}$ for the columns. It follows that

the condition $\boldsymbol{\delta}_1 = \boldsymbol{0}$ implies that $\boldsymbol{k}'_{12} = -\boldsymbol{b}'_{12}\boldsymbol{K}_{22}$ and hence the rank of $[\boldsymbol{k}_{12}\ \boldsymbol{K}'_{22}]'$ must be strictly less than the number of endogenous variables, m. Further, identification of the parameters, or in other words, recovery of $\boldsymbol{b}_{12}$ from $\boldsymbol{k}'_{12} = -\boldsymbol{b}'_{12}\boldsymbol{K}_{22}$, requires the rows of $\boldsymbol{K}_{22}$ to be linearly independent. These two conditions together imply that testing for the identifiability and specification of the instrumental variable model is equivalent to testing for the rank of $[\boldsymbol{k}_{12}\ \boldsymbol{K}'_{22}]'$ being equal to $m - 1$. This is the *rank condition*. If $\bar{m} > m - 1$ the instrumental variable model would of course be over-identified.[6]

5.1.2 Small Sample Properties

There is only limited published work with information on the small sample properties of statistical tests of rank in the context of the identification and specification of IV models. The only known exception to the authors is Cragg and Donald (1993), who provided an analysis for the MD method. In particular they explored the small sample properties of two alternative MD tests: a) a test of the null $H_0 : \rho[\boldsymbol{k}_{12}\ \boldsymbol{K}'_{22}] = m - 1$ and b) a test of the null $H_0 : \rho[\boldsymbol{K}_{22}] = m - 1$. For very small samples the sizes of the tests were too large, especially for the first alternative studied. This study further provided asymptotic expansions for the MD test in the context of a model like that in equation (6). This improved the size properties considerably particularly so for the second alternative. The second alternative was also the more powerful when using experimentally determined critical values.

5.2 Demand Systems

5.2.1 Theoretical Considerations

Tests of rank have been used in the context of the estimation of the Engel curve relationship, i.e. the relationship between budget shares and total expenditure (income). Engel curves are relevant to model the impact of policy measures on consumer responses, and in addition the welfare impact of such measures. Also the Engel curve serves as a tool to study the impact of fiscal policy measures on the relative demand of goods. The Engel curve is as follows:

$$\boldsymbol{w}_i = \boldsymbol{B}\boldsymbol{G}(x_i) + \boldsymbol{\varepsilon}_i \quad for\ \ i = 1, \ldots, N \tag{7}$$

where $\boldsymbol{w}_i$ is a $k \times 1$ vector of the budget shares of individual i, $\boldsymbol{B}$ is an $k \times m$ matrix of parameters, where $\boldsymbol{G}(x_i)$ is a $m \times 1$ vector where the functional form of $\boldsymbol{G}(.)$ may be unknown, and x_i is total expenditure of individual i, and $\boldsymbol{\varepsilon}_i$ is a $k \times 1$ zero mean random vector

---

[6]Hamilton (1994, Prop. 9.1) shows that this formulation of the rank condition is equivalent to its more commonly formulated form, namely that the rows of $\boldsymbol{M}$ are linearly independent, where:

$$\boldsymbol{M} = \left[\begin{array}{cc} E\left(\boldsymbol{x}_{1t}\boldsymbol{x}'_{1t}\right) & E\left(\boldsymbol{x}_{1t}\boldsymbol{x}'_{2t}\right) \\ E\left(y_{1t}\boldsymbol{x}'_{1t}\right) & E\left(y_{1t}\boldsymbol{x}'_{2t}\right) \end{array} \right]$$

independent of x_i. Note that the sum of the elements of the vector of budget shares sums to 1, i.e. $\boldsymbol{\varepsilon}_i'\iota = 0$ where ι is a $k \times 1$ vector of ones; this obviously implies certain restrictions on $E\{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\}$.[7] Tests of rank in this setting are relevant to find m, the number of unknown factors.

The rank of the demand system has important implications for demand theory, see Lewbel (1991) for a detailed review. Under the setting in (7) a rank of 1 implies that the demands are homothetic, i.e. budget shares are independent of the level of income. If the rank is two the demands are generalized linear. The PIGLOG specification, see Muellbauer (1975), is an example of rank two demand system in which budget shares are linear in the log of total expenditure. The clear advantage of the PIGLOG demand system is that they can be aggregated across individuals of different income. It is clear that the rank or structure of demand system has direct implications for the structure of aggregate demand equations. The PIGLOG would imply that the resulting aggregate demand equation is equivalent to the representative agent model.

Lewbel (1991) suggested the following strategy to estimate m nonparametrically. Let $\boldsymbol{Q}(x_i)$ be a $k \times 1$ (or larger than k) vector of functions having finite mean, and denote $\boldsymbol{A} = E\{\boldsymbol{w}_i\boldsymbol{Q}(x_i)'\}$. Given that x_i is independent of $\boldsymbol{\varepsilon}_i$, it holds that $\boldsymbol{A} = E\{\boldsymbol{B}\boldsymbol{G}(x_i)\boldsymbol{Q}(x_i)'\}$, and so it follows that rank of $\boldsymbol{A}$ is equal to m, unless some component of $\boldsymbol{G}$ is orthogonal to all the elements of $\boldsymbol{Q}$, which should be a very remote coincidence.[8]

Empirical Studies on the estimation of Engel curves on household data have been conducted among others by Atkinson, Gomulka, and Stern (1990), Banks, Blundell, and Lewbel (1997), Blundell, Duncan, and Pendakur (1998), Blundell and Duncan (1998), Hausman, Newey, Ichimura, and Powell (1991) and Hausman, Newey, and Powell (1995). Blundell, Duncan, and Pendakur (1998) estimated a semiparametric Engel curve in which household composition is modelled using an extended partially linear framework. Previous work, relied on trimming the sample of households to have an homogenous group. Banks, Blundell, and Lewbel (1997) provided a demand system model which was able to provide a detailed welfare

---

[7]Gorman (1981) suggested the following alternative specification for demand systems

$$\boldsymbol{w}_i = \boldsymbol{B}(P)\boldsymbol{G}(x_i) + \boldsymbol{\varepsilon}_i \quad for \quad i = 1, \dots, N$$

where additionally P is a vector of prices. Under this specification, the rank must be smaller than three for demands to be aggregable. See Lewbel (1991) and references therein for further studies of exactly aggregable demands.

[8]Note that the Barlett test could be implemented as $\boldsymbol{A}$ is nothing but the covariance matrix between $\boldsymbol{w}_i$ and $\boldsymbol{Q}(x_i)$. A consistent estimator of $\boldsymbol{A}$ is given by $\hat{\boldsymbol{A}} = T^{-1}\sum_{i=1}^{N}\boldsymbol{w}_i\boldsymbol{Q}(x_i)'$, so that $\sqrt{N}(\hat{\boldsymbol{A}}-\boldsymbol{A}) \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{V})$, and where a consistent estimator for $\boldsymbol{V}$ can be easily obtained, and hence the other tests of rank presented in section 2 can also be applied.

analysis of shifts in relative prices.

5.2.2 Small Sample Properties

The small sample properties of some sequential testing procedures based on the MD, GE and CRT methods have been explored in Cragg and Donald (1996) and Robin and Smith (2000). Cragg and Donald (1996) compared the small sample properties of the MD and GE tests in the context of controlled simulations of an estimated Engel curve like that in (7) where $\boldsymbol{G}(x_i) = \{1, x_i, x_i^2, x_i^3, x_i^4\}$.[9] Their results showed that the GE test tended to have a larger size particularly in exercises where the non-zero eigenvalues of $\boldsymbol{B}$ were all large. When some of the non-zero eigenvalues were small the size of the GE test was closer to the nominal size. The power properties were equally related to the size of the smallest non-zero eigenvalue. In a similar exercise, Robin and Smith (2000) showed that the statistical properties of the CRT method were dependent on the weighting matrices used. They further showed for the simulation exercise presented that the size properties of the CRT appear superior to those of the GE test, and importantly, the size properties of the CRT test displayed significant improvements when increasing the sample size from 250 to 2000, while the GE test showed relatively little improvement.

5.3 Reduced Rank VAR Models

5.3.1 Theoretical Considerations

Consider a conventional VAR of the form:

$$\boldsymbol{y}_t = \sum_{k=1}^{p} \boldsymbol{C}_k \boldsymbol{y}_{t-k} + \boldsymbol{\varepsilon}_t \tag{8}$$

each of the $\boldsymbol{C}_k$ is an $m \times m$ matrix, and $\boldsymbol{\epsilon}_t$ is an *iid* process. It is often the case that such VARs include a large number of insignificant coefficients; one can impose zero restrictions in a relatively *ad hoc* way so as to make the model more parsimonious. Velu, Reinsel, and Wichern (1986) proposed a reduced rank VAR model which provides a parsimonious method to model multivariate time series. This model has the following structure:

$$\boldsymbol{y}_t = \boldsymbol{F} \left[\sum_{k=1}^{p} \boldsymbol{G}_k \boldsymbol{y}_{t-k} \right] + \boldsymbol{\varepsilon}_t \tag{9}$$

---

[9]To control for the rank of $\boldsymbol{B}$ in the demand system, when they imposed a rank r in the simulation exercise, they performed r steps in the Gaussian elimination of the estimated matrix $\boldsymbol{B}$, and imposed the restriction that the resulting right-hand side and lower part submatrix of dimension $(5-r) \times (5-r)$ was a matrix of zeros. Undoing the Gaussian elimination of this resulting matrix provided an estimate of the rank r that was then used in simulating model (7).

Here each of the $\boldsymbol{G}_k$ is an $r \times m$ matrix ($r < m$) and $\boldsymbol{F}$ is an $m \times r$ matrix, where r is the rank of the system.[10] Velu, Reinsel, and Wichern (1986) suggested a method for estimating the parameters $\boldsymbol{F}$ and $\begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{G}_2 & ... & \boldsymbol{G}_k \end{bmatrix}$ in (9) conditional on a given r. Denote $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \boldsymbol{y}'_{t-2}, \boldsymbol{y}'_{t-k})'$ and $\boldsymbol{\Omega_\varepsilon} = \boldsymbol{\Omega_{yy}} - \boldsymbol{\Omega_{yx}}\boldsymbol{\Omega_{xx}^{-1}}\boldsymbol{\Omega_{xy}}$ where $\boldsymbol{\Omega_\varepsilon}$ is the covariance of the residuals of the OLS unrestricted regression of (8) and $\boldsymbol{\Omega_{xy}}$ is the covariance matrix between $\boldsymbol{x}$ and $\boldsymbol{y}$. Additionally denote $\boldsymbol{\Pi} = \boldsymbol{\Omega_\varepsilon^{-1}}$ and set $\boldsymbol{v}_j$ to be the eigenvector corresponding to the j-th largest eigenvalue of $\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega_{yx}}\boldsymbol{\Omega_{xx}^{-1}}\boldsymbol{\Omega_{xy}}\boldsymbol{\Pi}^{\frac{1}{2}}$, λ_j^2. If $\boldsymbol{\Upsilon}_r = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & ... & \boldsymbol{v}_r \end{bmatrix}$ then

$$\boldsymbol{F} = \boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Upsilon}_r, \quad [\boldsymbol{G}_1 \ \boldsymbol{G}_2 \ \ldots \ \boldsymbol{G}_k] = \boldsymbol{\Upsilon}'_r\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega_{yx}}\boldsymbol{\Omega_{xx}^{-1}}$$

are the solutions which minimize $tr\left\{\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\boldsymbol{\Pi}^{\frac{1}{2}}\right\}$. To determine r is equivalent to determine the rank of any of the $\boldsymbol{C}_k$'s which are assumed to have common rank. Consider the RRVAR model (9) re-expressed as

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \tag{10}$$

$t = 1, ..., T$, where the (m, mp) matrix $\boldsymbol{A} \equiv \boldsymbol{\alpha}\boldsymbol{\beta}'$.[11]

Reduced rank regression models like that in (10) have been used by Bekker, Dobbelstein, and Wansbeek (1996) to estimate Arbitrage Pricing models. Camba-Mendez, Kapetanios, Smith, and Weale (2003) presented a Monte Carlo exercise comparing the forecasting performance of reduced rank and unrestricted VAR models in which the former appear superior. They further estimated reduced rank VAR models for leading indicators of UK economic activity. Their results show that these more parsimonious multivariate representations display an improvement in forecasting performance over that of unrestricted VAR models.

5.3.2 Small Sample Properties

Camba-Mendez, Kapetanios, Smith, and Weale (2003) conducted a Monte Carlo exercise with a VAR model like that in (8) with $p = 1$ and $m = 5$, and where the rank of $\boldsymbol{C}$ is

---

[10] An alternative, and more general, representation for a reduced rank VAR model is the following:

$$\boldsymbol{y}_t = \sum_{k=1}^{p} \boldsymbol{F}_k\boldsymbol{G}_k\boldsymbol{y}_{t-k} + \boldsymbol{\varepsilon}_t$$

where each of the $\boldsymbol{G}_k$ is an $r_j \times m$ matrix ($r_j \leq m$) for $j = 1, \ldots, p$ and each $\boldsymbol{F}_k$ is an $m \times r_j$ matrix. It is further assumed that the r_j's are non-increasing. The $\boldsymbol{y}_t$ are simply output variables. This model was suggested by Ahn and Reinsel (1988) and was named *nested reduced rank autoregressive model*. This model has been extended by Ahn and Reinsel (1990) to incorporate error correction forms. Reinsel and Ahn (1992) provided the asymptotic distribution for testing for the number of unit roots in a vector autoregressive model with unit roots and the additional reduced rank structure of the nested reduced rank model.

[11] Note that Bartlett's (1947) test can then be easily computed from the ordered squared sample canonical correlations between $\{\boldsymbol{y}_t\}$ and $\{\boldsymbol{x}_t\}$. Note that under suitable regularity conditions, $T^{1/2}vec(\hat{\boldsymbol{A}} - \boldsymbol{A}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma_{XX}^{-1}} \otimes \boldsymbol{\Sigma_{\epsilon\epsilon}})$, where $\boldsymbol{\Sigma_{XX}} \equiv E\{\boldsymbol{x}_t\boldsymbol{x}'_t\}$ is assumed positive definite. Given this distribution, computation of the GE and CRT tests follows.

controlled by three sets of eigenvalues. In that exercise they compared the performance of the GE versus the performance of the BA test. Their results showed that the GE test significantly overrejects for many sample sizes, particularly if the relative magnitude of the eigenvalues is large. Furthermore, the size properties of the GE test did not improve very rapidly as the sample size increased. This finding was similar to those reported in Cragg and Donald (1996) and Robin and Smith (2000) in the context of a different modelling scenario, namely a demand system. The size properties of the BA test, on the other hand, appeared satisfactory at moderate to large sample sizes and were substantially better than those of the GE.

Further to evaluating the performance of these asymptotic tests of rank, Camba-Mendez, Kapetanios, Smith, and Weale (2003) also assessed the performance of their corresponding bootstrapped versions. They found that the bootstrap version of the GE test had clearly superior size properties that were not sensitive to the magnitude of the eigenvalues. The bootstrapped version of the BA test also offered an improvement and displayed good size properties similar to those of the bootstrapped version of the GE test. The power of the bootstrapped version of the BA test appeared to be slightly better than the bootstrapped version of the GE test.

5.4 State Space models

5.4.1 Theoretical Considerations

We focus on the state space representation in the *innovation form*, i.e.:

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{C}\boldsymbol{s}_t + \boldsymbol{e}_t \\
\boldsymbol{s}_{t+1} &= \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{e}_t
\end{aligned}
\tag{11}
$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are $r \times r$, $r \times m$ and $m \times r$ parameter matrices respectively, $\boldsymbol{s}_t$ is a r-vector of unobservable state variables, and $\boldsymbol{e}_t$ is an m-vector of random variables with mean zero and positive definite covariance matrix $\boldsymbol{\Omega}$.[12] This system can be characterized by a system transfer function $\boldsymbol{G}(z) = \sum_{i=1}^{\infty} \boldsymbol{G}_i z^{-1}$, where $\boldsymbol{G}_i$ are the impulse response matrices. The order of the system, is defined as the order of the minimal state-space realization, i.e. the minimal dimension of the state vector that replicates the transfer function. Corresponding to the transfer function $\boldsymbol{G}(z)$ above, the infinite dimensional Hankel matrix is defined as:

$$
\boldsymbol{H} = \mathcal{O}\mathcal{C} =
\begin{bmatrix}
\boldsymbol{G}_1 & \boldsymbol{G}_2 & \boldsymbol{G}_3 & \cdots \\
\boldsymbol{G}_2 & \boldsymbol{G}_3 & \cdots & \cdots \\
\boldsymbol{G}_3 & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{CB} & \boldsymbol{CAB} & \boldsymbol{CA}^2\boldsymbol{B} & \cdots \\
\boldsymbol{CAB} & \boldsymbol{CA}^2\boldsymbol{B} & \cdots & \cdots \\
\boldsymbol{CA}^2\boldsymbol{B} & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\tag{12}
$$

---

[12]For further details on the innovation form representation see Brockwell and Davis (1991, Sec. 12.4)

where the so called observability matrix is defined as $\mathcal{O} = \left[\boldsymbol{C}', \boldsymbol{A}'\boldsymbol{C}', (\boldsymbol{A}^2)'\boldsymbol{C}', \cdots \right]'$ and the so called controllability matrix as $\mathcal{C} = \left[\boldsymbol{B}, \boldsymbol{AB}, (\boldsymbol{A}^2)\boldsymbol{B}, \cdots \right]$. Kronecker's theorem can be used to show that the order of the system is equal to the rank of the Hankel matrix (see Kailath (1980)).

Searching for the rank of the Hankel matrix, however, is not conducted directly on an estimate of (12) but rather on some *pseudo*-Hankel matrices. For example, an alternative characterization of this system is in terms of a Hankel matrix of the covariances of the output vector, $\boldsymbol{y}_t$.

$$\boldsymbol{H}^a = \mathcal{O}\overline{\mathcal{C}} = \begin{bmatrix} \boldsymbol{\Delta}_1 & \boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_3 & \cdots \\ \boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_3 & \cdots & \cdots \\ \boldsymbol{\Delta}_3 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where $\boldsymbol{\Delta}_i$ is the autocorrelation matrix of $\boldsymbol{y}_t$ for lag i. Where $\mathcal{O}$ is the observability matrix defined above, and $\overline{\mathcal{C}} = \left[\overline{\boldsymbol{C}}, \boldsymbol{A}\overline{\boldsymbol{C}}, (\boldsymbol{A}^2)\overline{\boldsymbol{C}}, \cdots \right]$, and $\overline{\boldsymbol{C}} = \boldsymbol{B} + \boldsymbol{APC}'$ where $\boldsymbol{P}$ is the covariance matrix of the state vector defined as $E\{\boldsymbol{s}_t\boldsymbol{s}_t'\}$. It follows that the rank of $\boldsymbol{H}^a$ is equivalent to the rank of $\boldsymbol{H}$, see Faurre (1976). Obviously one cannot use the infinite dimensional matrix above, and when working with finite data will have to resort to a finite truncation of the Hankel matrix. Note that this Hankel Covariance matrix can be defined as the covariance matrix between the vectors $\boldsymbol{y}_+^t$ and $\boldsymbol{y}_-^t$, and defined as follows:

$$\boldsymbol{H}^a = E\left(\boldsymbol{y}_+^t \boldsymbol{y}_-^{t\,'} \right) = E\left(\begin{array}{c} \boldsymbol{y}_{t+1} \\ \dots \\ \boldsymbol{y}_{t+k} \end{array} \right) \left(\begin{array}{ccc} \boldsymbol{y}_t' & \dots & \boldsymbol{y}_{t-p+1}' \end{array} \right) = \begin{bmatrix} \boldsymbol{\Delta}_1 & \boldsymbol{\Delta}_2 & \dots & \boldsymbol{\Delta}_p \\ \boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_3 & \dots & \boldsymbol{\Delta}_{p+1} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{\Delta}_k & \boldsymbol{\Delta}_{k+1} & \dots & \boldsymbol{\Delta}_{k+p-1} \end{bmatrix} \quad (13)$$

The truncation parameters k and p must be fixed, and setting them implies a trade off between generality in model specification and modeling $\boldsymbol{\Delta}_i$ at very distant lags; see Aoki and Havenner (1991) for further details.[13]

This type of state space model has been used to model exchange rates, Dorfman (1997), economic interdependence between countries, Aoki (1987), build a small macroeconometric model for the Dutch Economy, Otter and Dal (1987) and forecasting commodity prices, Foster, Havenner, and Walburger (1995). Dorfman and Havenner (1992) developed a Bayesian

---

[13]The representation of the Hankel matrix stated in equation (13) suggests that the Bartlett test could be used to test for the rank of this matrix, and by extension also the information criteria procedures and the Bias Correction Bartlett test are valid. The Cragg and Donald (1996) procedure is also feasible. Under the assumption of stationarity of $\boldsymbol{y}_t$ it can be shown, Brockwell and Davis (1991, Ch. 7), that $\sqrt{T}vec(\hat{\boldsymbol{H}}^a - \boldsymbol{H})$ is asymptotically distributed as $N(\boldsymbol{0}, \boldsymbol{V}^H)$. While the matrix $\boldsymbol{V}^H$ is of reduced rank, the rank of a consistent estimate may only be of reduced rank asymptotically. As stated in section 2 this is problematic for the Cragg and Donald (1996) procedure. An estimator of $\boldsymbol{V}^H$ with equal rank to $\boldsymbol{V}^H$ can be constructed as in Camba-Mendez and Kapetanios (2001).

approach to state space multivariate modelling. More recently, Kapetanios (2004) and Camba-Mendez and Kapetanios (2005b) have used this type of model to compute measures of underlying inflation extracted from a vector series that contained all available subcomponents of consumer price indices. These core inflation measures were proved better than other traditional measures to track inflation developments over the medium to long term.

5.4.2 Small Sample Properties

Camba-Mendez and Kapetanios (2004) conducted a Monte Carlo exercise with a state space model like that in (11) to assess the performance of the GE and BA statistical tests of rank. The Monte Carlo exercise dealt with several values for the truncation parameters k and p, and different degrees of persistence of the shocks in the system.[14]

Their results showed that the performance of the tests worsens with respect to the size of the Hankel matrix, for experiments with smaller moduli of the eigenvalues of matrix $\boldsymbol{A}$, and when the sample size is small. The size of the GE test were not good when the dimension of the Hankel matrix was large. The BA method was less sensitive to all dimensions in the study, and more robust than the other methods when the dimension of the Hankel matrix was large.

In line with the findings of Camba-Mendez, Kapetanios, Smith, and Weale (2003), results in Camba-Mendez and Kapetanios (2004) also showed that bootstrapped procedures of those tests of rank significantly improved upon the performance of the corresponding asymptotic tests. Furthermore, these procedures were also shown to have in general a better performance than standard information criteria methods. The performance of information criteria methods did not deteriorate much when increasing the dimension of the Hankel matrix, however, they appeared more sensitive than statistical tests to sample size. This was particularly the case for the Schwarz (1978) criteria which underestimated the rank for samples of size $T = 200$.

---

[14]Matrix $\boldsymbol{A}$ in (11) is the key matrix to explain the dynamics of $\boldsymbol{y}_t$; the degree of persistence of shocks will depend on the eigenvalues of $\boldsymbol{A}$. To control for this in the experiment Camba-Mendez and Kapetanios (2004) used $\boldsymbol{A} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}'$, where $\boldsymbol{\Lambda}$ is a 3×3 quasi upper triangular matrix; the last element of the diagonal corresponds to the modulus assigned to that experiment, and the 2×2 block matrix in the left upper corner was computed so that the modulus of the complex pair of eigenvalues of this 2×2 block was also equal to the modulus assigned to the eigenvalues of that experiment; with the remaining values fixed to a value of one. $\boldsymbol{E}$ was an orthonormal matrix generated from a standard normal matrix using Gram-Schmidt orthogonalization. 3 different moduli making three alternative experiments were used: i) with moduli given by (0.8,0.8,0.8); ii) (0.4,0.4,0.4) and iii) (0.8,0.8,0.2), which allowed to check for the robustness of the procedures when one of the eigenvalues was small.

5.5 Cointegration

5.5.1 Theoretical Considerations

Phillips (1986) showed that a necessary condition for cointegration is that the spectral density matrix of the innovation sequence of an I(1) multivariate process has deficient rank at frequency zero. The equivalence of time-domain and frequency-domain analysis of time series is well documented in the statistical and econometric literature. Nevertheless, the use of spectral densities is by far less widespread than the use of covariances in the econometric analysis of time series.[15] Phillips and Ouliaris (1988) suggested two procedures for detecting the presence of cointegration. The drawback of their method was that they were tests of the null of 'no cointegration', namely a test of the hypothesis that the r smallest eigenvalues are greater than zero. Tests of the rank of that matrix at frequency zero are tests of the null of 'cointegration', i.e. tests of the null that the r smallest eigenvalues are equal to zero. Brillinger (1981, pp. 262) or Brockwell and Davis (1991, pp. 447) show how to construct an estimate of the spectral density matrix at any frequency together with its distribution. This allows the implementation of the CK test of rank described in section 3.

5.5.2 Small Sample properties

The class of finite order VECM models is not the most appropriate class to assess nonparametric procedures. Therefore, linear and nonlinear cointegrating systems will be considered. The data generation process for the vector simulated series $\boldsymbol{y}_t$ is defined as follows:

$$\Delta\boldsymbol{y}_t = F(\Delta\boldsymbol{y}_{t-1})\boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \tag{14}$$

where we allow for three alternative specifications for $F(.)$:

$$F(\Delta\boldsymbol{y}_{t-1}) = \boldsymbol{I} \tag{15}$$

$$F(\Delta\boldsymbol{y}_{t-1}) = 1 - e^{-(\sum_{i=1}^{m} \Delta y_{i,t-1})^2} \tag{16}$$

$$F(\Delta\boldsymbol{y}_{t-1}) = 1\{|\sum_{i=1}^{m} \Delta y_{i,t-1}| > 2\} \tag{17}$$

These specifications lead to a linear model if (15), a STAR-type model if (16), and a SETAR-type model if (17). The last two models lead to nonlinear VECM models where the speed of convergence to equilibrium depends on $\Delta\boldsymbol{y}_{t-1}$. As their name indicate the STAR-type model is inspired by univariate smooth transition autoregressive (STAR) models, while the SETAR-type by self-exciting threshold autoregressive (SETAR) models. Note that these nonlinear

---

[15]The methods to test for cointegration most usually encountered in applied economic work are those of Johansen (1988), Stock and Watson (1988), Gregoir and Laroque (1994) and Snell (1999). Their tests are reviewed in many econometric textbooks. In this section we will focus instead on the strategy proposed by Camba-Mendez and Kapetanios (2005a), reviewed above.

models still imply the existence of a Wold decomposition for the differenced data, since they are covariance stationary (see, e.g., Granger and Teräsvirta (1993) and Tong (1990)), and therefore our suggested procedure is appropriate.

We concentrate on a multivariate model with 3 variables. We control the rank of the coefficient matrix, $\boldsymbol{\Pi}$ in the error correction representation by specifying the vector of its eigenvalues. Two different vectors are considered: $(-0.6, 0, 0)$, i.e. one cointegrating vector, and $(-0.6, -0.6, 0)$, i.e. two cointegrating vectors. Note that all the eigenvalues are negative given the requirement that the eigenvalues of $\boldsymbol{I} + \boldsymbol{\Pi}$ are less than or equal to one. We then construct a standard normal random matrix of eigenvectors, $\boldsymbol{E}$ which are almost surely linearly independent. These are transformed into an orthonormal basis, $\tilde{\boldsymbol{E}}$, using the Gram-Schmidt process. The coefficient matrix is then given by $\tilde{\boldsymbol{E}} \boldsymbol{\Lambda} \tilde{\boldsymbol{E}}'$ where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of the required coefficient matrix. Two alternative types of random disturbances are used for simulating $\boldsymbol{\epsilon}_t$. First, random normal disturbances with identity covariance matrix. Second, iid MA(1) processes with correlation coefficient 0.9. Using these random numbers a sample from a process following the error correction representation in (14) is obtained.

The sample sizes considered are 200 and 600. For each simulated sample, 200 initial observations have been discarded to minimise the effect of starting values. For each Monte Carlo experiment 10000 replications have been carried out. Bias and Mean Square Error (MSE) statistics for these simulation exercises are shown in table 1. For illustration purposes, this table also reports simulation results for Johansen (1988) maximum eigenvalue test (JM) and also his trace test (JT). Generally speaking the performance of the CK method described in section 2 is satisfactory for most cases under study. The only exceptions are exercises run with samples of size 200, rank 2 and a SETAR-type model. The test appears always best in terms of Bias and MSE for exercises of rank equal to 1, sample size equal to 600 and MA(1) errors. But for minor exceptions, the Johansen's procedures are always best for exercises conducted with normally distributed shocks.

5.6 Other potential applications

5.6.1 Dynamic Factor Models

Denote a zero mean, wide sense stationary m-vector process by $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$, and assume that there exists a representation such as:

$$\boldsymbol{y}_t = \boldsymbol{P} \boldsymbol{z}_t + \boldsymbol{\varepsilon}_t \tag{18}$$

where $\boldsymbol{P}$ is a $m \times r$ matrix of parameters, $\boldsymbol{\varepsilon}_t$ is an m-vector of *iid* zero mean processes with covariance matrix $\boldsymbol{\Sigma}_\varepsilon$, and $\boldsymbol{z}_t$ is a r-vector stationary process, with $r < m$, i.e. there is a

reduction in dimensionality, which follows an ARMA(p,q) process

$$\boldsymbol{\Phi}(L)\boldsymbol{z}_t = \boldsymbol{\Theta}(L)\boldsymbol{u}_t$$

where $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ are matrix polynomials in the lag operator L with all the roots of the determinant polynomials $|\boldsymbol{\Phi}(L)|$ and $|\boldsymbol{\Theta}(L)|$ outside the unit circle, and $\boldsymbol{u}_t$ is an iid random process with zero mean and positive definite covariance matrix $\boldsymbol{\Sigma}_u$. A further identification restriction imposed in this model is that the r factors are independent, and that all $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Theta}_i$ matrices are diagonal.[16] Matrix $\boldsymbol{P}$ is usually refer to as the factor loadings. For identification purposes it is assumed that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$. Denote $\boldsymbol{\Gamma}_y(k) = E\{\boldsymbol{y}_t\boldsymbol{y}'_{t-k}\}$, and $\boldsymbol{\Gamma}_z(k) = E\{\boldsymbol{z}_t\boldsymbol{z}'_{t-k}\}$. Under the representation in equation (18), it follows $\boldsymbol{\Gamma}_y(k) = \boldsymbol{P}\boldsymbol{\Gamma}_z(k)\boldsymbol{P}'$ for $k \geq 1$. The rank of $\boldsymbol{\Gamma}_y(k)$ for $k \geq 1$ is equal to r, the number of the common driving forces.[17]

Early applications of dynamic factor models to macroeconomic research include Sargent and Sims (1977) and Geweke (1977). Sargent and Sims (1977) proposed a dynamic factor model that was consistent with the idea of co-movement in macroeconomic series. They assumed that there was an underlying force behind the fluctuations of macroeconomic series. Rather than working under the assumption of a unique underlying force, Geweke and Singleton (1981) used a dynamic factor model with two latent variables (factors) to explain the business cycle. They identified those two factors with unanticipated aggregate demand shocks and innovations to anticipated aggregate demand shocks. In line with Sargent and Sim's work, Stock and Watson (1989) used a dynamic factor model to extract a latent variable that could be identified as the state of the economy. Their assumption was that the fluctuations of certain macroeconomic variables have an underlying common factor, and this common factor could be identified as the 'state of the economy'. The use of dynamic factor models in forecasting macroeconomic series is not new. Engle and Watson (1981) used a traditional dynamic factor model to forecast sectorial wage rates in Los Angeles. They compared the forecasting performance of that dynamic factor model with a regression model without latent variables, and found that the dynamic factor model was better. Recent work by Camba-Mendez, Kapetanios, Smith, and Weale (2001) and Stock and Watson (2000) address the problem of forecasting a single time series with many possible predictors. They showed that the predictors could be summarized by a small number of dynamic factors and

---

[16]An alternative equivalent representation with solid $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Theta}_i$ matrices is also explained in Pena and Box (1987).

[17]Having established the number of common driving forces, it is still necessary to identify the type of VARMA process followed by the vector of driving forces. To do so, it is possible to use a transformation of the vector series $\boldsymbol{y}_t$. Note that the columns of $\boldsymbol{P}$ are the eigenvectors $\boldsymbol{\Gamma}_y(k)$ associated with the nonzero eigenvalues. If we denote by $\boldsymbol{P}^+$ the Moore-Penroe generalized inverse of $\boldsymbol{P}$, then it follows that $\boldsymbol{P}^+\boldsymbol{y}_t = \boldsymbol{z}_t - \boldsymbol{P}^+\boldsymbol{\varepsilon}_t$, i.e. equal to the vector of common driving forces plus an added noise. This transformation can be used to identify the VARMA structure underlying the common driving forces.

that forecasts based on these factors outperformed various benchmark models.[18]

5.6.2 Dynamic Principal Components

The problem behind Dynamic Principal Components is that of approximating an m-vector stationary process $\boldsymbol{y}_t$, that without loss of generality it is assumed to have zero mean, by a filter series of itself, but having a filter which has reduced rank. A dynamic principal component model takes the form:

$$\boldsymbol{y}_t = \boldsymbol{C}(L)\boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_t \tag{19}$$

where $\boldsymbol{C}(L)$ is a polynomial lag and forward operator, i.e. a double sided filter, with $\boldsymbol{C}_i$ matrices of parameters of order $m \times k$; $\boldsymbol{\zeta}_t$ is a $k \times 1$ vector of principal components, and where $\boldsymbol{\varepsilon}_t$ is a $m \times 1$ error process. The dynamic principal components are a filter version of $\boldsymbol{y}_t$ given by $\boldsymbol{\zeta}_t = \boldsymbol{B}(L)\boldsymbol{y}_t$ where $\boldsymbol{B}(L)$ is a polynomial lag and forward operator, i.e. a double sided filter, with $\boldsymbol{B}_i$ matrices of parameters of order $k \times m$. The polynomial operators $\boldsymbol{B}_i$ and $\boldsymbol{C}_i$ which minimize:

$$E\{(\boldsymbol{y}_t - \boldsymbol{C}(L)\boldsymbol{\zeta}_t)^\tau \, (\boldsymbol{y}_t - \boldsymbol{C}(L)\boldsymbol{\zeta}_t)\}$$

and where $^\tau$ serves to denote transpose conjugate, are given by:

$$\boldsymbol{B}_u = (2\pi)^{-1} \int_0^{2\pi} \boldsymbol{\Upsilon}_k(\alpha)^\tau e^{iu\alpha} d\alpha$$

and

$$\boldsymbol{C}_u = (2\pi)^{-1} \int_0^{2\pi} \boldsymbol{\Upsilon}_k(\alpha) e^{iu\alpha} d\alpha$$

where $\boldsymbol{\Upsilon}_k(\alpha)$ are the k eigenvectors of the spectral density matrix of $\boldsymbol{y}_t$ at frequency α associated with the k largest eigenvalues, see Brillinger (1981) for further details. Tests of the rank of the spectral density matrix at frequency α could then be used to help in identifying k.

In a recent paper Forni and Reichlin (1998) suggested the use of a generalized dynamic factor model to describe the dynamics of sectoral industrial output and productivity for the US economy from 1958 to 1986. Their model was similar to that in (19), but without the

---

[18]Standard factor models, i.e. a model like that in (18), but where $\boldsymbol{z}_t$ is a non serially correlated random vector with mean 0, covariance matrix $\boldsymbol{\Sigma}_z$ and independent from $\boldsymbol{\varepsilon}_t$, have also been used in the Econometrics literature. This standard factor model has been used among other things for testing the Arbitrage Pricing Theory. In testing for the number of factors, and in the context of testing the Arbitrage Pricing Theory, Cragg and Donald (1997) suggested to use a k-vector of macro variables $\boldsymbol{x}_t$, where $k \geq r$. One could then estimate the equation,

$$\boldsymbol{y}_t = \boldsymbol{\gamma} + \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{y}_t$ is an m-vector of asset returns. The rank of $\boldsymbol{A}$ gives the number of factors.

idiosyncratic error component. By aggregating across a large number of sectors the idiosyncratic component vanishes. Under this setting the number of common shocks driving those series is equal to the rank of their spectral density matrix. The foundations for this result are to be found in the literature on dynamic principal components, see Brillinger (1981). This issue is further explored in Forni, Hallin, Lippi, and Reichlin (1999).

6   Conclusion

This review has concentrated on statistical methods that relate to the determination of the rank of a matrix. We describe several general tests of rank of a matrix. Further, we examine how these can be used to estimate the rank of a matrix. Alternative methods for this estimation that do not use tests but information criteria have also been reviewed. In addition, a large variety of modelling scenarios where these tests of rank are useful for specification purposes have been presented. The modelling scenarios range from linear and stationary models such as standard VARs, factor analysis, dynamic factor models, instrumental variable estimation, and dynamic principal component models, to nonlinear frameworks such as nonparametric factor models and also to nonstationary frameworks such as cointegrated systems. We conclude that these methods are of increasing relevance given the focus of econometric and statistical work on multivariate systems. We expect that this focus will only increase in the near future in line with the size of datasets under investigation.

References

AHN, S. K., AND G. C. REINSEL (1988): "Nested Reduced-Rank Autoregressive Models for Time Series," *Journal of the American Statistical Association*, 83, 849–856.

——— (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

AKAIKE, H. (1974): "Stochastic Theory of Minimal Realisations," *IEEE Transactions on Automatic Control*, AC-19(6), 667–674.

——— (1976): "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in *System Identification*, ed. by R. Mehra, and D. Lainiotis. Academic Press.

ANDERSON, T. W. (2003): *Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 3rd edn.

ANDREWS, D. W. K. (1987): "Asymptotic Results for Generalised Wald Tests," *Econometric Theory*, 3, 348–358.

AOKI, M. (1987): *State Space Modelling of Time Series*. Springer-Verlag.

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

ATKINSON, A. B., J. GOMULKA, AND N. H. STERN (1990): "Spending on Alcohol: Evidence from the Family Expenditure Survey 1970-1983," *Economic Journal*, 100, 808–827.

BANKS, J., R. BLUNDELL, AND A. LEWBEL (1997): "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79(4), 527–539.

BARTLETT, M. (1947): "Multivariate Analysis," *Journal of the Royal Statistical Society, Series B*, 9, 176–197.

BEKKER, P., P. DOBBELSTEIN, AND T. WANSBEEK (1996): "The APT Model as Reduced-Rank Regression," *Journal of Business and Economic Statistics*, 14, 199–202.

BERKOWITZ, J., AND F. X. DIEBOLD (1998): "Bootstrapping multivariate spectra," *Review of Economics and Statistics*, 80(4), 664–666.

BLUNDELL, R., AND A. DUNCAN (1998): "Kernel Regression in Empirical Microeconomics," *Journal of Human Resources*, 33(1), 62–87.

BLUNDELL, R., A. DUNCAN, AND K. PENDAKUR (1998): "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13(5), 435–461.

BRILLINGER, D. (1981): *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer Series in Statistics, Springer, New York, 2nd edition.

CAMBA-MENDEZ, G., AND G. KAPETANIOS (2001): "Testing the Rank of the Hankel Covariance Matrix: A Statistical Approach," *IEEE Transactions on Automatic Control*, 46, 331–336.

———— (2004): "Bootstrap Statistical Tests of Rank Determination for System Identification," *IEEE Transactions on Automatic Control*, 49.

———— (2005a): "Estimating the rank of the Spectral Density Matrix," *Journal of Time Series Analysis*, 26(1), 37–48.

———— (2005b): "Forecasting euro area Inflation using Dynamic Factor Measures of Underlying inflation," *Journal of Forecasting*, 24, 491–503.

CAMBA-MENDEZ, G., G. KAPETANIOS, R. J. SMITH, AND M. R. WEALE (2001): "An Automatic Leading Indicator of Economic Activity: Forecasting GDP growth for European Countries," *Econometrics Journal*, 4, 56–90.

———— (2003): "Tests of rank in Reduced rank regression models," *Journal of Business and Economic Statistics*, 21(1), 145–155.

CHAMBERLAIN, G. (1982): "Multivariate Regression models for panel data," *Journal of Econometrics*, 18, 5–46.

CRAGG, J. G., AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variables models," *Econometric Theory*, 9, 222–240.

———— (1996): "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association*, 91, 1301–1309.

———— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76(1–2), 223–250.

DONGARRA, J. J., J. R. BUNCH, C. B. MOLER, AND G. W. STEWART (1979): *LINPACK Users' Guide*. SIAM.

DORFMAN, J. H. (1997): "Competing exchange rate models: a state space model vs structural time series alternatives," in *Applications of computer aided time series modeling. Lecture Notes in Statistics, Vol. 119*, ed. by M. Aoki, and A. M. Havenner. Springer, New York.

DORFMAN, J. H., AND A. M. HAVENNER (1992): "A Bayesian approach to state space multivariate time series modeling," *Journal of Econometrics*, 52(3), 315–346.

ENGLE, R. F., AND M. WATSON (1981): "A One Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the Americal Statistical Association*, 76, 774–781.

FAURRE, P. L. (1976): "Stochastic Realization Algorithms," in *System Identification: Advances and Case Studies*, ed. by R. K. Mehra, and D. G. Lainiotis. New York, Academic Press.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (1999): "The Generalised Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82(4), 540–554.

FORNI, M., AND L. REICHLIN (1998): "Let's Get Real: A factor Analytical Approach to Disaggregated Business Cycle Dynamics," *Review of Economic Studies*, 65, 453–473.

FOSTER, K. A., A. M. HAVENNER, AND A. M. WALBURGER (1995): "System Theoretic Time Series Forecasts of Weekly Live Cattle Prices," *American Journal of Agricultural Economics*, 77(4), 1012–1023.

FUJIKOSHI, Y. (1974): "The Likelihood Ratio Tests for the Dimensionality of Regression Coefficients," *Journal of Multivariate Analysis*, 4, 327–340.

GEWEKE, J. F. (1977): "The Dynamic Factor Analysis of Economic Time Series Models," in *Latent Variables in Socioeconomic Models*, ed. by D. J. Aigner, and A. S. Goldberger. North Holland.

GEWEKE, J. F., AND K. J. SINGLETON (1981): "Maximum Likelihood Confirmatory Factor Analysis of Economic Time Series," *International Economic Review*, 22, 37–54.

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(419), 766–776.

GLYNN, W. J., AND R. J. MUIRHEAD (1978): "Inference in Canonical Correlation Analysis," *Journal of the Multivariate Analysis*, 8, 468–478.

GOLUB, G. H., AND C. F. VAN LOAN (1996): *Matrix Computations*. The John Hopkins University Press.

GORMAN, W. M. (1981): "Some Engel curves," in *Essays in the theory and measurement of consumer behaviour in honor of sir Richard Stone*, ed. by A. Deaton. Cambridge University Press, Cambridge.

GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press.

GREGOIR, S., AND G. LAROQUE (1994): "Polynomial Cointegration: Estimation and Test," *Journal of Econometrics*, 63, 183–214.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.

HANNAN, E. J., AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society. Series B*, 41, 190–195.

HAUSMAN, J., W. NEWEY, H. ICHIMURA, AND J. POWELL (1991): "Identification and estimation of polynomial errors in variables models," *Journal of Econometrics*, 50, 273–296.

HAUSMAN, J., W. NEWEY, AND J. POWELL (1995): "Nonlinear errors in variables: estimation of some Engel Curve models," *Journal of Econometrics*, 65, 205–234.

HIGHHAM, N. J. (1996): *Accuracy and Stability of Numerical Algorithms*. SIAM Publications.

HOSOYA, Y. (1989): "Hierarchical Statistical Models and a Generalised Likelihood Ratio Test," *Journal of the Royal Statistical Society B*, 51(3), 435–448.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

KAILATH, T. (1980): *Linear Systems*. Englewood Cliffs, N. J.: Prentice-Hall.

KAPETANIOS, G. (2004): "A Note on Modelling Core Inflation for the UK Using a New Dynamic Factor Estimation Method and a Large Disaggregated Price Index Dataset," *Economics Letters*, 85, 63–69.

KOHN, R. (1979): "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models," *Econometrica*, 47(4), 1005–1030.

KUDO, A. (1963): "A Multivariate Analogue of the one-sided test," *Biometrika*, 50(3), 403–418.

KUDO, A., AND J. CHOI (1975): "A Generalized Multivariate Analogue of the one-sided test," *Memoirs of the Faculty of Science, Kyushu University, Ser. A*, 29(2), 303–328.

LAWLEY, D. N. (1959): "Tests of Significance in Canonical Analysis," *Biometrika*, 46, 59–66.

LEWBEL, A. (1991): "The rank of demand systems: theory and nonparametric estimation," *Econometrica*, 59(3), 711–730.

LÜTKEPOHL, H. (1996): *Handbook of Matrices.* John Wiley and sons.

LÜTKEPOHL, H., AND M. M. BURDA (1997): "Modified Wald Tests Under Nonregular Conditions," *Journal of Econometrics*, 78, 315–332.

MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.

MUELLBAUER, J. (1975): "Aggregation, Income Distribution, and Consumer Demand," *Review of Economic Studies*, 62, 269–283.

OTTER, P. W., AND R. V. DAL (1987): "State Space and Distributed Lag Modelling of Dynamic Economic Processes Based on Singular Value Decompositions: with Applications to the Dutch Economy," *Annales D'Economie et de Statistique*, 6/7, 253–277.

PENA, D., AND G. E. P. BOX (1987): "Identifying a Simplifying Structure in Time Series," *Journal of the American Statistical Association*, 82(399), 836–843.

PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions," *Journal of Econometrics*, 33, 311–340.

PHILLIPS, P. C. B., AND S. OULIARIS (1988): "Testing for cointegration using principal components methods," *Journal of Economic Dynamics and Control*, 12, 205–230.

POTCHER, B. M. (1983): "Order Estimation in ARMA-Models by Langrangian Multiplier Tests," *Annals of Statistics*, 11, 872–885.

REINSEL, G. C., AND S. K. AHN (1992): "Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting," *Journal of Time Series Analysis*, 13, 353–375.

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

SARGENT, T. J., AND C. SIMS (1977): "Business Cycle Modelling Without Pretending to Have Too Much a Priori Theory," in *New Methods of Business Cycle Research*, ed. by C. Sims. Federal Reserve Bank of Minneapolis.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

SNELL, A. (1999): "Testing for r versus $r-1$ Cointegrating Vectors," *Journal of Econometrics*, 88, 151–191.

STOCK, J. H., AND M. WATSON (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

——— (1989): "New Indexes (Sic) of Coincident and Leading Indicators," *NBER Macroeconomics annual*, 4, 351–394.

——— (2000): "Macroeconomic Forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20(2), 147–162.

SUN, H. J. (1988): "A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine," *Communications in statistics - Simulation and computation*, 17(3), 1097–1111.

TONG, H. (1990): *Nonlinear time series: A dynamical system approach.* Oxford University Press.

VELU, R. P., G. C. REINSEL, AND D. W. WICHERN (1986): "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105–118.

A Appendix

A.1 Local Power Results

We have the following proposition concerning the local power of the MD, GE and CRT procedures under the hypothesis H_{1T}

Proposition 4 *Under H_{1T},*

$$
MD \xrightarrow{d} \chi^2_{(m-r^*)(n-r^*)} \left(vec(\boldsymbol{B})' \left(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1} \boldsymbol{\Delta}_h \left(\boldsymbol{\Delta}_h' \boldsymbol{V}^{-1} \boldsymbol{\Delta}_h \right)^{-1} \boldsymbol{\Delta}_h' \boldsymbol{V}^{-1} \right) vec(\boldsymbol{B}) \right)
$$

$$
GE \xrightarrow{d} \chi^{2\prime}_{(m-r^*)(n-r^*)} \left(vec(\boldsymbol{B})' \boldsymbol{\Gamma}' (\boldsymbol{\Gamma} \boldsymbol{V} \boldsymbol{\Gamma}')^{-1} \boldsymbol{\Gamma} vec(\boldsymbol{B}) \right)
$$

$$
CRT \xrightarrow{d} \sum_{i=1}^{(m-r^*)(n-r^*)} \tau_i \chi^{2\prime}_1 \left(\left[\boldsymbol{\Theta}((\boldsymbol{D}_{r^*}' \otimes \boldsymbol{C}_{r^*}') \boldsymbol{V} (\boldsymbol{D}_{r^*} \otimes \boldsymbol{C}_{r^*}))^{-1/2} \ (\boldsymbol{D}_{r^*}' \otimes \boldsymbol{C}_{r^*}') vec(\boldsymbol{B}) \right]_i^2 \right)
$$

where $\chi^2_l(\delta)$ denotes a non-central χ^2 variate with l degrees of freedom and non-centrality parameter δ, and where for a vector $\boldsymbol{\alpha}$ the notation $[\boldsymbol{\alpha}]_i^2$ denotes the square of the i-th element of $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$ is an orthogonal matrix containing the eigenvectors of $(\boldsymbol{D}_{r^*}' \otimes \boldsymbol{C}_{r^*}') \boldsymbol{V} (\boldsymbol{D}_{r^*} \otimes \boldsymbol{C}_{r^*})$.

Proof:

Proof for the MD test is given in Cragg and Donald (1997, Th. 2). Proof for GE follows easily from the fact that under H_{1T}, $\sqrt{T} vec(\hat{\boldsymbol{A}} - \boldsymbol{\Psi}) \xrightarrow{d} N(vec(\boldsymbol{B}), \boldsymbol{V})$. Proof for the CRT test requires the following Lemma

Lemma 1 *For a $m \times 1$ vector random variable $\boldsymbol{y} \sim N(\boldsymbol{\alpha}, \boldsymbol{\Omega})$, and a symmetric matrix $\boldsymbol{Q}$ the quadratic form $\boldsymbol{y}' \boldsymbol{Q} \boldsymbol{y}$ is distributed as a weighted sum of $\chi^{2\prime}_1$ random variables where the weights are the eigenvalues of $\boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2}$ and the noncentrality parameters are given by the squares of the elements of the mean vector in (20) below.*

Proof of Lemma 1. $\boldsymbol{y}' \boldsymbol{Q} \boldsymbol{y} = \tilde{\boldsymbol{y}}' \boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2} \tilde{\boldsymbol{y}}$ where $\tilde{\boldsymbol{y}} \sim N(\boldsymbol{\Omega}^{-1/2} \boldsymbol{\alpha}, \boldsymbol{I})$. By symmetry, we may write $\boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2} = \boldsymbol{D}' \boldsymbol{\Lambda} \boldsymbol{D}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2}$ and $\boldsymbol{D}$ is an orthogonal matrix containing the eigenvectors of $\boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2}$. Then $\tilde{\boldsymbol{y}}' \boldsymbol{\Omega}^{1/2} \boldsymbol{Q} \boldsymbol{\Omega}^{1/2} \tilde{\boldsymbol{y}} = \tilde{\tilde{\boldsymbol{y}}}' \boldsymbol{\Lambda} \tilde{\tilde{\boldsymbol{y}}}$ where

$$
\tilde{\tilde{\boldsymbol{y}}} \sim N(\boldsymbol{D} \boldsymbol{\Omega}^{-1/2} \boldsymbol{\alpha}, \boldsymbol{I}) \tag{20}
$$

and the conclusion of the Lemma easily follows.

Given Lemma 1 the conclusion of Proposition 4 for the CRT test easily follows from Theorem 3.2 in Robin and Smith (2000).

A.2 Distribution of Λ for a hermitian positive semidefinite matrix.

As $vec(\mathbf{\Lambda})$ is not analytic, it cannot be expanded as a Taylor series. We define instead for a hermitian complex matrix $\mathbf{A}$, a $2n \times 2n$ real symmetric matrix $\mathbf{A}^R$ which is an arrangement of the real and imaginary parts of the elements of $\mathbf{A}$. Details on $\mathbf{A}^R$ are given in Brillinger (1981, pp. 71). By Brillinger (1981, Lemma 3.7.1(i),(ii),(iv)), if $\mathbf{\Lambda} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ then $\mathbf{\Lambda}^R = \mathbf{\Sigma}_{22}^R - \mathbf{\Sigma}_{21}^R\mathbf{\Sigma}_{11}^{R\,-1}\mathbf{\Sigma}_{12}^R$. Note that $(Re\ vec(\mathbf{\Sigma})', Im\ vec(\mathbf{\Sigma})')' \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^r)$. Let $\mathbf{d}_{ij}$ be the vector of distinct elements of $\mathbf{\Sigma}_{ij}^R$. Define $\mathbf{J}_1$, $\mathbf{J}_2$, $\mathbf{J}_j^h$, $\mathbf{J}_{ij}^h$ and $\mathbf{D}_i$, $i,j = 1,2$, as $\mathbf{s} \equiv \left(vec(\mathbf{\Sigma}_{11}^R)', vec(\mathbf{\Sigma}_{21}^R)', vec(\mathbf{\Sigma}_{12}^R)', vec(\mathbf{\Sigma}_{22}^R)' \right)' = \mathbf{J}_1 (Re\ vec(\mathbf{\Sigma})', Im\ vec(\mathbf{\Sigma})')'$, $\mathbf{J}_2 vec(\mathbf{\Lambda}^R) = (Re\ vec(\mathbf{\Lambda})', Im\ vec(\mathbf{\Lambda})')'$, $\mathbf{J}_j^h \mathbf{d}_{jj} = vech(\mathbf{\Sigma}_{jj}^R)$, $\mathbf{J}_{ij}^h \mathbf{d}_{ij} = vec(\mathbf{\Sigma}_{ij}^R)$ and $vec(\mathbf{\Sigma}_{ii}^R) = \mathbf{D}_i vech(\mathbf{\Sigma}_{ii}^R)$. Then

$$\mathbf{R} \equiv \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial \mathbf{s}} = \left[\frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{11}^R)'}, \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{21}^R)'}, \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{12}^R)'}, \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{22}^R)'} \right]$$

Since $vec(\mathbf{\Sigma}_{21}^R\mathbf{\Sigma}_{11}^{R\,-1}\mathbf{\Sigma}_{12}^R) = \left(\mathbf{\Sigma}_{12}^{R\,'} \otimes \mathbf{\Sigma}_{21}^R \right) vec\left(\mathbf{\Sigma}_{11}^{R\,-1} \right)$, $\mathbf{\Sigma}_{11}^R$ and $\mathbf{\Sigma}_{22}^R$ are symmetric and $\mathbf{\Sigma}_{21}^R = \mathbf{\Sigma}_{12}^{R\,'}$, from Brillinger (1981, Lemma 3.7.1(v)), we have

$$\frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{11}^R)'} = \left(\mathbf{\Sigma}_{12}^{R\,'} \otimes \mathbf{\Sigma}_{21}^R \right) \mathbf{D}_1 \mathbf{D}_1^+ \left(\mathbf{\Sigma}_{11}^{R\,-1} \otimes \mathbf{\Sigma}_{11}^{R\,-1} \right) \mathbf{D}_1 \mathbf{J}_1^h \mathbf{J}_1^{h+} \mathbf{D}_1^+ \tag{21}$$

$$\frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{21}^R)'} = -\left(\mathbf{I}_{4(n-r)^2} + \mathbf{K}_{2(n-r),2(n-r)} \right) \left(\mathbf{\Sigma}_{21}^R\mathbf{\Sigma}_{11}^{R\,-1} \otimes \mathbf{I}_{2(n-r)} \right) \mathbf{J}_{21}^h \mathbf{J}_{21}^{h\,+} \tag{22}$$

$$\frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{12}^R)'} = \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{21}^R)'} \mathbf{K}_{2r,2(n-r)}, \qquad \frac{\partial vec(\mathbf{\Lambda}^R)}{\partial vec(\mathbf{\Sigma}_{22}^R)'} = \mathbf{D}_2 \mathbf{J}_2^h \mathbf{J}_2^{h+} \mathbf{D}_2^+ \tag{23}$$

where for a matrix $\mathbf{A}$, $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, $\mathbf{K}_{m,n}$ is a commutation matrix (see Lütkepohl (1996, Sec. 9.2)). (21), (22) and (23) follow from Lütkepohl (1996, 10.6(2) and 9.5.3(1)(ii)), Lütkepohl (1996, 10.5.1(7)) and Lütkepohl (1996, 10.4.1(1)(iii) and 9.5.3(1)(ii)) respectively[19]. Then, $\sqrt{2M+1} \left(Re\ vec(\hat{\mathbf{\Lambda}})', Im\ vec(\hat{\mathbf{\Lambda}})' \right)' \xrightarrow{d} N(\mathbf{0}, \mathbf{W}^r)$ where $\mathbf{W}^r = \mathbf{J}\mathbf{V}^r\mathbf{J}'$ and $\mathbf{J} = \mathbf{J}_2\mathbf{R}\mathbf{J}_1$. Finally, $\sqrt{2M+1} vec(\hat{\mathbf{\Lambda}}) \xrightarrow{d} N^C(\mathbf{0}, \mathbf{W})$. An alternative to the above is the use of numerical derivatives, or the use of the bootstrapped methods for the multivariate spectra described in Berkowitz and Diebold (1998).

---

[19]Results on the commutation matrix and more details on the facts used to derive (21), (22) and (23) may be found in Magnus and Neudecker (1988) which is the original source of the results quoted from Lütkepohl (1996).

Table 1: Bias and MSE of Estimated rank. Linear Model.[a]

Model	Noise	Test	rank	Bias 200	Bias 600	MSE 200	MSE 600
Linear	Normal	CK	1	0.191	0.134	0.206	0.135
			2	-0.366	-0.218	0.418	0.230
		JM	1	0.058	0.055	0.061	0.059
			2	0.059	0.060	0.059	0.060
		JT	1	0.060	0.056	0.073	0.070
			2	0.059	0.060	0.059	0.060
	MA(1)	CK	1	0.196	0.135	0.206	0.137
			2	-0.369	-0.206	0.425	0.218
		JM	1	0.158	0.162	0.183	0.187
			2	0.095	0.078	0.095	0.078
		JT	1	0.169	0.172	0.218	0.217
			2	0.095	0.078	0.095	0.078
STAR	Normal	CK	1	0.158	0.130	0.209	0.136
			2	-0.543	-0.307	0.670	0.341
		JM	1	0.055	0.056	0.059	0.059
			2	0.063	0.060	0.063	0.060
		JT	1	0.057	0.056	0.068	0.066
			2	0.063	0.060	0.063	0.060
	MA(1)	CK	1	0.173	0.125	0.206	0.129
			2	-0.492	-0.275	0.596	0.295
		JM	1	0.156	0.150	0.188	0.170
			2	0.093	0.081	0.093	0.081
		JT	1	0.145	0.160	0.229	0.196
			2	0.093	0.081	0.093	0.081
SETAR	Normal	CK	1	-0.125	0.081	0.342	0.156
			2	-1.019	-0.647	1.452	0.825
		JM	1	-0.115	0.055	0.208	0.058
			2	-0.041	0.060	0.177	0.060
		JT	1	-0.161	0.057	0.273	0.069
			2	-0.028	0.060	0.156	0.060
	MA(1)	CK	1	0.036	0.123	0.257	0.143
			2	-0.824	-0.466	1.110	0.558
		JM	1	-0.144	0.152	0.384	0.175
			2	-0.151	0.081	0.366	0.081
		JT	1	-0.166	0.159	0.489	0.202
			2	-0.055	0.081	0.236	0.081

[a]Sample sizes for Monte Carlo experiments are 200 and 600. CK denotes the test described in section 3 in this paper. JM refers to Johansen's maximum eigenvalue test and JT to Johansen's trace test. rk denotes the cointegrating rank which is 1 or 2 for the different exercises conducted as described in the text.

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.europa.eu).

817 "Convergence and anchoring of yield curves in the euro area" by M. Ehrmann, M. Fratzscher, R. S. Gürkaynak and E. T. Swanson, October 2007.

818 "Is time ripe for price level path stability?" by V. Gaspar, F. Smets and D. Vestin, October 2007.

819 "Proximity and linkages among coalition participants: a new voting power measure applied to the International Monetary Fund" by J. Reynaud, C. Thimann and L. Gatarek, October 2007.

820 "What do we really know about fiscal sustainability in the EU? A panel data diagnostic" by A. Afonso and C. Rault, October 2007.

821 "Social value of public information: testing the limits to transparency" by M. Ehrmann and M. Fratzscher, October 2007.

822 "Exchange rate pass-through to trade prices: the role of non-linearities and asymmetries" by M. Bussière, October 2007.

823 "Modelling Ireland's exchange rates: from EMS to EMU" by D. Bond and M. J. Harrison and E. J. O'Brien, October 2007.

824 "Evolving U.S. monetary policy and the decline of inflation predictability" by L. Benati and P. Surico, October 2007.

825 "What can probability forecasts tell us about inflation risks?" by J. A. García and A. Manzanares, October 2007.

826 "Risk sharing, finance and institutions in international portfolios" by M. Fratzscher and J. Imbs, October 2007.

827 "How is real convergence driving nominal convergence in the new EU Member States?" by S. M. Lein-Rupprecht, M. A. León-Ledesma and C. Nerlich, November 2007.

828 "Potential output growth in several industrialised countries: a comparison" by C. Cahn and A. Saint-Guilhem, November 2007.

829 "Modelling inflation in China: a regional perspective" by A Mehrotra, T. Peltonen and A. Santos Rivera, November 2007.

830 "The term structure of euro area break-even inflation rates: the impact of seasonality" by J. Ejsing, J. A. García and T. Werner, November 2007.

831 "Hierarchical Markov normal mixture models with applications to financial asset returns" by J. Geweke and G. Amisano, November 2007.

832 "The yield curve and macroeconomic dynamics" by P. Hördahl, O. Tristani and D. Vestin, November 2007.

833 "Explaining and forecasting euro area exports: which competitiveness indicator performs best?" by M. Ca' Zorzi and B. Schnatz, November 2007.

834 "International frictions and optimal monetary policy cooperation: analytical solutions" by M. Darracq Pariès, November 2007.

835 "US shocks and global exchange rate configurations" by M. Fratzscher, November 2007.

836 "Reporting biases and survey results: evidence from European professional forecasters" by J. A. García and A. Manzanares, December 2007.

837 "Monetary policy and core inflation" by M. Lenza, December 2007.

838 "Securitisation and the bank lending channel" by Y. Altunbas, L. Gambacorta and D. Marqués, December 2007.

839 "Are there oil currencies? The real exchange rate of oil exporting countries" by M. M. Habib and M. Manolova Kalamova, December 2007.

840 "Downward wage rigidity for different workers and firms: an evaluation for Belgium using the IWFP procedure" by P. Du Caju, C. Fuss and L. Wintr, December 2007.

841 "Should we take inside money seriously?" by L. Stracca, December 2007.

842 "Saving behaviour and global imbalances: the role of emerging market economies" by G. Ferrucci and C. Miralles, December 2007.

843 "Fiscal forecasting: lessons from the literature and challenges" by T. Leal, J. J. Pérez, M. Tujula and J.-P. Vidal, December 2007.

844 "Business cycle synchronization and insurance mechanisms in the EU" by A. Afonso and D. Furceri, December 2007.

845 "Run-prone banking and asset markets" by M. Hoerova, December 2007.

846 "Information combination and forecast (st)ability. Evidence from vintages of time-series data" by C. Altavilla and M. Ciccarelli, December 2007.

847 "Deeper, wider and more competitive? Monetary integration, Eastern enlargement and competitiveness in the European Union" by G. Ottaviano, D. Taglioni and F. di Mauro, December 2007.

848 "Economic growth and budgetary components: a panel assessment for the EU" by A. Afonso and J. González Alegre, January 2008.

849 "Government size, composition, volatility and economic growth" by A. Afonso and D. Furceri, January 2008.

850 "Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling" by G. Camba-Méndez and G. Kapetanios, January 2008.