Department of Economics

Statistical Tests of the Rank of a Matrix and their Applications in Econometric Modelling

Gonzalo Camba-Mendez and George Kapetanios

Working Paper No. 541 May 2005 ISSN 1473-0278

Statistical tests of the rank of a matrix and their applications in econometric modelling

Gonzalo Camba-Mendez*and George Kapetanios†

March 29, 2005

Abstract

Testing 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 the rank of a matrix, and provides details on a variety of modelling scenarios in the econometrics literature where these tests are required.

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

*European Central Bank, Kaiserstrasse 29, D-60311, Frankfurt am Main, Germany, email: gonzalo.camba-mendez@ecb.int .

†Queen Mary, University of London, Mile End Road, London, E1 4NS, United Kingdom, email: G.Kapetanios@qmul.ac.uk

1 Introduction

Tests 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. So, 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 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 test $H_0 : \{\rho[\boldsymbol{A}] = r\}$ against $H_A : \{\rho[\boldsymbol{A}] > r\}$, where $\rho[.]$ denotes the rank of a matrix. For a sample of size T, we define an estimate of $\boldsymbol{A}$ by $\hat{\boldsymbol{A}}$. By an application of some suitable central limit theorem we assume that $\sqrt{T}vec(\hat{\boldsymbol{A}} - \boldsymbol{A}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$. 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, denoted by r^0. 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^0, 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^0, see also Cragg and Donald (1997). 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.

This provides a succinct definition of the general problem we address. The applications of these general rank estimation procedures ranges from identification of IV models to factor analysis and VAR modelling. Section 2 reviews the various procedures for determining the

rank of an estimated matrix in the case of a general matrix. Section 3 concentrates on determining the rank of a hermitian positive semidefinite matrix. Section 4 presents a large variety of modelling scenarios where the tests of rank we discuss are of immediate relevance. Finally, Section 5 concludes.

2 Rank of a General Matrix

2.1 A Generalized Minimum Discrepancy Function Test

This section presents a minimum discrepancy function (MDF) method[1] to test whether a $q \times 1$ parameter vector $\boldsymbol{\theta}$ can be represented as a function of $p \times 1$ parameter vector $\boldsymbol{\mu} \in \Re^p$, and where $p < q$. It is then of interest to formulate a general test of the hypothesis $H_0 : \{\boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\mu}_0)\}$, where $\boldsymbol{\mu}_0$ is used to denote the true value of $\boldsymbol{\mu}$, and $\boldsymbol{h}(\boldsymbol{\mu}) = \{h_1(\boldsymbol{\mu}), h_2(\boldsymbol{\mu}), \ldots, h_q(\boldsymbol{\mu})\}'$, where the functions $h_i(\boldsymbol{\mu})$ for $i = 1, \ldots, q$ are continuously differentiable. A minimum discrepancy function test statistic could thus be formulated as:

$$\hat{\psi}^g = \min_{\mu} F\left(\hat{\boldsymbol{\theta}}, \boldsymbol{h}(\boldsymbol{\mu})\right) = T \min_{\mu} \left\{ \left(\hat{\boldsymbol{\theta}} - \boldsymbol{h}(\boldsymbol{\mu})\right)' \boldsymbol{V}^+ \left(\hat{\boldsymbol{\theta}} - \boldsymbol{h}(\boldsymbol{\mu})\right) \right\} \tag{1}$$

Where the following assumptions are made:

Assumption 1 .

a. $\sqrt{T} vec(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$.

b. *(uniqueness). The parameter vector* $\boldsymbol{\mu}$ *is identified at* $\boldsymbol{\mu}_0$*, i.e.* $\boldsymbol{h}(\boldsymbol{\mu}^*) = \boldsymbol{h}(\boldsymbol{\mu}_0)$ *for* $\boldsymbol{\mu}^* \in \Re^p$ *implies* $\boldsymbol{\mu}^* = \boldsymbol{\mu}_0$*; and* $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ *is the unique minimizer of* $F\left(\hat{\boldsymbol{\theta}}, \boldsymbol{h}(\boldsymbol{\mu})\right)$.

c. *The mapping* $\boldsymbol{h}(\boldsymbol{\mu})$ *is defined in a neighbourhood of* $\boldsymbol{\mu}_0$*, 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.

d. *(inf-boundedness). There exists a number* $\alpha > F\left(\hat{\boldsymbol{\theta}}, \boldsymbol{h}(\boldsymbol{\mu}_0)\right)$ *and a compact subset* $\boldsymbol{\Xi}^* \subset \Re^p$ *such that* $\{\boldsymbol{\mu} \in \Re^p : F\left(\hat{\boldsymbol{\theta}}, \boldsymbol{h}(\boldsymbol{\mu})\right) < \alpha\} \subseteq \boldsymbol{\Xi}^*$ *whenever* $\hat{\boldsymbol{\theta}}$ *is in a neighbourhood of* $\boldsymbol{\theta}_0$.

e. *The* $p \times p$ *matrix* $\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\Delta}_0$ *is nonsingular, and where* $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}$.

Assumption 1.a can be usually justified by an application of the Central Limit Theorem. The assumptions above are the usual regularity conditions for a minimum discrepancy type test being chi-squared distributed. This is stated in the following proposition:

[1]Note that some notation in this paper is recycled in different subsections. The context should make the meaning of recycled symbols clear.

Proposition 1 *Under Assumptions 1.a-1.d above and under the null H_0, it holds that i) the minimizer $\hat{\boldsymbol{\mu}} \xrightarrow{a.s.} \boldsymbol{\mu}_0$, and ii) $\hat{\psi}^g \xrightarrow{p} \chi^2_{q-p} + \bar{\chi}^2$, where χ^2_β denotes the χ^2 distribution with degrees of freedom $\beta = r^v - p$.*

Proof: See appendix. □

$Prob\left(\hat{\boldsymbol{V}} = \boldsymbol{V}\right) \to 1$ as $T \to \infty$ does not guarantee $Prob\left(\hat{\boldsymbol{V}}^+ = \boldsymbol{V}^+\right) \to 1$ as $T \to \infty$. This is due to the fact that generalized inverses are not continuous. Andrews (1987) has shown that the condition $Prob\left(\rho\left[\hat{\boldsymbol{V}}\right] = \rho\left[\boldsymbol{V}\right]\right) \to 1$ as $T \to \infty$ is a sufficient condition to avoid this issue. To enforce this condition, we follow the solution suggested in Lutkepohl and Burda (1997), namely that if the rank of $\boldsymbol{V}$ is r^v, 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 $\hat{\boldsymbol{V}}$.

An MDF test of rank. 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[2]. 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:

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

where $\boldsymbol{s} = vec(\boldsymbol{S})$ and $\boldsymbol{a_1} = vec(\boldsymbol{A}_1)$. It only remains to show that $\boldsymbol{h}(\boldsymbol{\mu})$ is in line with the assumptions made above. This issue has been addressed by Cragg and Donald (1997) and we state it here in the following proposition.

Proposition 2 *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 assumptions 2-4 above.*

Proof: See appendix. □

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[3]. r^* steps of Gaussian

[2]The reordering can be accomplished by using the pivoting matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ obtained from the r steps of Gaussian elimination as explained above. To avoid excessive notation pivoting matrices will be ignored in this section.

[3]For details on Gaussian elimination with complete pivoting see Cragg and Donald (1996) or Golub and Loan (1983). The foundations behind this strategy follow the work of Gill and Lewbel (1992). The

elimination with full pivoting on matrix $\boldsymbol{A}$ amounts to the following operations:

$$\boldsymbol{Q}_{r^*}\boldsymbol{R}_{r^*}\boldsymbol{Q}_{r^*-1}\boldsymbol{R}_{r^*-1}\ldots\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\ldots\boldsymbol{C}_{r^*-1}\boldsymbol{C}_{r^*} = \begin{bmatrix} \boldsymbol{A}_{11}(r^*) & \boldsymbol{A}_{12}(r^*) \\ \boldsymbol{0} & \boldsymbol{A}_{22}(r^*) \end{bmatrix}$$

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

$$\boldsymbol{R}_{r^*}\boldsymbol{R}_{r^*-1}\ldots\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\ldots\boldsymbol{C}_{r^*-1}\boldsymbol{C}_{r^*} = \boldsymbol{R}\boldsymbol{A}\boldsymbol{C} = \boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_{11}(r^*) & \boldsymbol{F}_{12}(r^*) \\ \boldsymbol{F}_{21}(r^*) & \boldsymbol{F}_{22}(r^*) \end{bmatrix}$$

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 $rk(\boldsymbol{A}) = r^*$. 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 the covariance matrix of the asymptotically normally distributed matrix $\sqrt{T}vec(\hat{\boldsymbol{A}} - \boldsymbol{A})$ has full rank, the following result can be shown, under H_0.

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

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

$$\hat{\xi} = T vec\ \hat{\boldsymbol{\Lambda}}_{22}(r^*)'(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{V}}\hat{\boldsymbol{\Gamma}}')^{-1} vec\ \hat{\boldsymbol{\Lambda}}_{22}(r^*) \xrightarrow{d} \chi^2_{(m-r^*)(n-r^*)}$$

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 tests computes the inverse of the covariance matrix $\boldsymbol{V}$. However, in many modelling scenarios this matrix is singular. Extension to such cases is stated in the following proposition.

Proposition 3 *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*

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

where $^+$ *denotes the Moore-Penrose inverse of a matrix, and the number of degrees of freedom* β *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) □

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.

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. Under the null hypothesis, $\boldsymbol{A}$ has $\min(m,n) - r^*$ zero eigenvalues. It follows that the above quadratic form has $\min(m,n) - r^*$ zero eigenvalues as well. 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

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

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.

2.4 Other Methods

2.4.1 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 $r_{\min(\mathrm{m,n})} - r^*$ canonical correlations are zero, i.e., $H_{r^*} : \rho_{r^*+1} = \cdots = \rho_{\min(\mathrm{m,n})} = 0$. Under the null hypothesis and assuming stationarity

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

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).

2.4.2 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 denotes 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 $r^* > r^0$ since the probability that $T\sum_{i=r^0+1}^{r^*} \ln(1-\hat{\rho}_i^2) < 2(F(r^0) - F(r^*))$ is greater than zero. Therefore, the estimated rank will not converge in probability to r^0 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^0. 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 $L = \hat{\psi}$ and $F(r) = p$ provided also a consistent method to search for the rank of a matrix.

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.[4] 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 $H_0 : rk(\boldsymbol{A}) = r$ 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

[4] An algorithm to compute the factorization $\boldsymbol{PAP'} = \boldsymbol{G}\bar{\boldsymbol{G}}'$, 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.

show in the appendix that $\sqrt{T}\ vec(\hat{\mathbf{\Lambda}}) \xrightarrow{d} N^C(\mathbf{0}, \boldsymbol{W})$ where $\boldsymbol{W}$ is defined in the appendix. Hence

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

where $\boldsymbol{L}$ is a $n-r \times (n-r)^2$ selector matrix that picks the diagonal elements of $\hat{\mathbf{\Lambda}}$. 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 $H_0 : \theta_i \geq 0$, $i = 1, \ldots n-r$ where at least one inequality is strict. This is stated as follows:

Proposition 4 *Under the null hypothesis, $H_0 : r = r^*$ the test statistic, $\bar{\chi}^2 = T\ \hat{\boldsymbol{\theta}}' \boldsymbol{\Psi}^{-1} \hat{\boldsymbol{\theta}}$, where $\boldsymbol{\Psi} = \boldsymbol{LWL}'$, is distributed as a weighted mixture of χ_i^2, $i = 1, \ldots, n-r^*$, i.e.*

$$Pr\left(\bar{\chi}^2 \geq \bar{\chi}_0^2\right) = \sum_{i=0}^{q} w_i Pr\left(\chi_i^2 \geq \bar{\chi}_0^2\right)$$

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

Proof: See Camba-Mendez and Kapetanios (2005a) $\square$

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

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

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 $\mathbf{\Omega}_{Q_i}$ is the variance matrix of θ_j, $j \in Q_i$, and $\mathbf{\Omega}_{Q_i:Q_i'}$ is the same under the condition $\theta_j = 0$, $j \notin Q_i$, and $P\{\mathbf{\Omega}\}$ is the probability that the variables distributed in a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Omega}$ are all positive; finally, $P\{\mathbf{\Omega}_{\phi:K}\} = 1$ and $P\{(\mathbf{\Omega}_{K'})^{-1}\} = P\{(\mathbf{\Omega}_{\phi})^{-1}\} = 1$. The probabilities in (3) can be easily computed by means of the algorithm proposed in Sun (1988). Note that a simple expression for $\mathbf{\Omega}_{Q_i:Q_i'}$ is given by $\mathbf{\Omega}_{Q_i} - \mathbf{\Omega}_{Q_i,Q_i'} \mathbf{\Omega}_{Q_i'}^{-1} \mathbf{\Omega}_{Q_i,Q_i'}'$ where $\mathbf{\Omega}_{Q_i,Q_i'}$ is the covariance matrix of θ_j, $j \in Q_i$ and θ_j, $j \in Q_i'$ (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 $\mathbf{\Psi}$ is singular. A generalization of the test of rank presented here hence also follows.

4 Applications of tests of rank

4.1 Identification and Specification of IV Models

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 which are one of the most important models in econometrics. A simultaneous equation model can be written in its *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 exogenous variables, $\boldsymbol{\varepsilon}_t$ is a m-vector random process of zero mean and covariance matrix $\boldsymbol{\Omega}$. 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 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 in zero restrictions in 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 with y_{1t} and $\boldsymbol{y}_{2t}$ as follows,

$$\boldsymbol{B} = \begin{bmatrix} B_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Gamma}_2 \end{bmatrix}$$

and where it is further assumed that $B_{11} = 1$. If we assume that there are no zero restrictions on $\boldsymbol{\Gamma}_1$ there is an identification problem. In this setting, it is necessary to find certain instrumental variables, say $\boldsymbol{x}_{2t}$, uncorrelated with $\boldsymbol{y}_{1t}$ but correlated with $\boldsymbol{y}_{2t}$; we could then write equation (5) as:

$$y_{1t} = -\boldsymbol{B}_{12}\boldsymbol{y}_{2t} + \boldsymbol{\Gamma}_1\boldsymbol{x}_{1t} + \boldsymbol{\delta}_1\boldsymbol{x}_{2t} + \varepsilon_{1t} \tag{6}$$

and where the central specification hypothesis is $\boldsymbol{\delta}_1 = 0$. We could define the vector $\boldsymbol{x}_t = (\boldsymbol{x}'_{1t}, \boldsymbol{x}'_{2t})'$, and estimate the reduced form:

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

The structural parameters can be recover from $\boldsymbol{B}\boldsymbol{K} = \boldsymbol{\Gamma}^*$, where $\boldsymbol{\Gamma}^* = [\boldsymbol{\Gamma} \;\; \boldsymbol{\delta}]$ where $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2)'$. In particular for equation (6) these are given by,

$$\begin{aligned} \boldsymbol{K}_{11} &= \boldsymbol{\Gamma}_1 + \boldsymbol{K}_{12}\boldsymbol{B}_{12} \\ \boldsymbol{K}_{21} &= \boldsymbol{\delta}_1 + \boldsymbol{K}_{22}\boldsymbol{B}_{12} \end{aligned}$$

Where, as before, $\boldsymbol{K}$ has been partitioned in four blocks, according with y_{1t} and $\boldsymbol{y}_{2t}$ on the rows and according to $\boldsymbol{x}_{1t}$ and $\boldsymbol{y}_{2t}$ on the columns. It follows that the specifying condition $\boldsymbol{\delta}_1 = 0$ implies that $\boldsymbol{K}_{21} = \boldsymbol{K}_{22}\boldsymbol{B}_{12}$ and this implies that the rank of $[\boldsymbol{K}_{21} \ \boldsymbol{K}_{22}]$ must be strictly less than m the number of endogenous variables. Further, identification of the parameters implies that $\boldsymbol{K}_{22}$ must be full rank if one is to recover $\boldsymbol{B}_{12}$ from $\boldsymbol{K}_{21} = \boldsymbol{K}_{22}\boldsymbol{B}_{12}$. These two conditions together imply that testing for the identifiability and the specification of the instrumental variable model is testing for the rank of $[\boldsymbol{K}_{21} \ \boldsymbol{K}_{22}]$ being equal to $m-1$. Cragg and Donald (1993) developed alternative tests of rank for the identifiability of parameters apparently estimable by instrumental variables. Their method is less general than those presented above.

4.2 Factor Analysis

A factor analysis model describes a m-vector $\boldsymbol{y}_t$ of observable variables as:

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{K}\boldsymbol{f}_t + \boldsymbol{u}_t \tag{7}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{u}_t$ are vectors of dimension m, $\boldsymbol{K}$ is a matrix of parameters of order $m \times r$, and $\boldsymbol{f}_t$ are the common factors. $\boldsymbol{u}_t$ is a random vector independent of $\boldsymbol{f}_t$, with mean zero and covariance matrix $\boldsymbol{\Sigma_u}$. Equivalently, the factors $\boldsymbol{f}_t$ are random variables with mean zero and covariance matrix $\boldsymbol{\Sigma_f}$. Then, the covariance matrix of the observed vector $\boldsymbol{y}_t$ is:

$$\boldsymbol{\Sigma_y} = E\left(\boldsymbol{y}_t - \boldsymbol{\mu}\right)\left(\boldsymbol{y}_t - \boldsymbol{\mu}\right)' = \boldsymbol{K}\boldsymbol{\Sigma_f}\boldsymbol{K}' + \boldsymbol{\Sigma_u} \tag{8}$$

Identification of the factor model in (7) requires that a triplet $\boldsymbol{\Sigma_u}, \boldsymbol{\Sigma_f}$ and $\boldsymbol{K}$ that solves equation (8) exists and is unique. Existence refers to whether there is a nonnegative diagonal matrix $\boldsymbol{\Sigma_u}$ such that $\boldsymbol{\Sigma_y} - \boldsymbol{\Sigma_u}$ is positive definite and of rank equal to r. Identification requires to impose further restrictions on those matrices. A common restriction is to fix the covariance matrix $\boldsymbol{\Sigma_f}$ to be equal to an identity matrix, and for $\boldsymbol{\Gamma} = \boldsymbol{K}'\boldsymbol{\Sigma_u^{-1}}\boldsymbol{K}$ to be diagonal. An alternative identifying restriction would be for $\boldsymbol{\Sigma_f}$ to be diagonal and for $\boldsymbol{K} = (\boldsymbol{I}_r, \boldsymbol{K}_2')'$. Maximum Likelihood estimation of $\boldsymbol{\mu}$, $\boldsymbol{K}$ and $\boldsymbol{\Sigma_u}$ are obtained by maximizing the likelihood function:

$$L = (2\pi)^{-\frac{1}{2}mT}|\boldsymbol{\Sigma_y}|^{-\frac{1}{2}T} exp\{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\mu})'\boldsymbol{\Sigma_y^{-1}}(\boldsymbol{y}_t - \boldsymbol{\mu})\}$$

Models like (7) have been used in testing the Arbitrage Pricing Theory. There have been two major approaches to test the Arbitrage Pricing Theory. Firstly, the Theoretical Approach specifies certain macroeconomic and financial variables which are the factors $\boldsymbol{f}_t$, and are thought to capture the systematic risk of the economy. Secondly, the Statistical Approach, which is based on Factor Analysis or alternatively on Principal Component Analysis. Empirical Studies under this second approach include Lehmann and Modest (1988) and Connor

and Korajczyk (1988).

A likelihood ratio test can be constructed to test the hypothesis on adequacy of r factors. Under the null that r factors are adequate, the statistic

$$-\left(T-1-\frac{1}{6}(2T+5)-\frac{2}{3}r\right)\left[ln|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}|-ln|\hat{\boldsymbol{K}}\hat{\boldsymbol{K}}'+\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}|\right]$$

is distributed as a χ^2_l where $l=\frac{1}{2}\left[(T-r)^2-T-r\right]$ is the degrees of freedom. This particular structure for the APT model was not without criticism, Roll and Ross (1980), Roll (1984) and Dhrymes, Friend, and Gultekin (1984) showed that the number of factors selected by this test increased when the number of asset returns which were part of $\boldsymbol{y}_t$ was increased. The problem was associated with the requirement of matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ being diagonal. This translates in imposing that the diversifiable components of returns being uncorrelated across assets. However this restriction is too restrictive, and the APT model of Ross (1976) doesn't rely on this but on a weaker restriction, namely that the nonfactor risk $\boldsymbol{u}_t$, can be diversified away in asset portfolios. Chamberlain and Rothschild (1983) and Ingersoll (1984) generalized the factor model above to account for 'weak' correlation among assets in $\boldsymbol{u}_t$, they refer to this as the *approximate factor structure*. It is obvious that in the presence of an *approximate factor structure*, the Likelihood ratio test would overestimate the number of factors. Connor and Korajczyk (1993) developed a test for the number of factors in an approximate factor model. Engle, Ng, and Rothschild (1990) extended a traditional factor model to account for ARCH effects in the residuals, the factor-ARCH model. This model was used to study the relationship between asset risk premia and volatilities in a multivariate system.

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,

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

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

4.3 Demand Systems

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:

$$\boldsymbol{w}_i = \boldsymbol{A}\boldsymbol{G}(x_i) + \boldsymbol{\varepsilon}_i \quad for \;\; i = 1, \ldots, N \tag{9}$$

where $\boldsymbol{w}_i$ is a $k \times 1$ vector of the budget shares of individual i, $\boldsymbol{A}$ 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 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 $E\{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\}$. Tests of rank in this setting are relevant to find m, the number of unknown factors. Note that the structure of (9) resembles closely that of Factor Analysis. The factors in this setting are not linear on the variables, but rather should be referred to as Nonparametric factors.

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 $\boldsymbol{B} = E\{\boldsymbol{w}_i\boldsymbol{Q}(x_i)'\}$. Given that x_i is independent of $\boldsymbol{\varepsilon}_i$, it holds that $\boldsymbol{B} = E\{\boldsymbol{A}\boldsymbol{G}(x_i)\boldsymbol{Q}(x_i)'\}$, and so it follows that rank of $\boldsymbol{B}$ 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. Lewbel (1991) applied this nonparametric rank estimation method to individual household expenditures data, in particular he used the UK Department of Employment Family expenditure Survey, and the US Labor Statistics Consumer Expenditure Survey. For the US dataset total consumption expenditures were divided in seven categories: food, clothing, recreation, furnishing, health, care, transportation, and other. The list of instruments used as $\boldsymbol{Q}(x_i)$ were $1, x_i, ln(x_i), x_i^2, ln(x_i)^2, 1/x_i, 1/ln(x_i), 1/(x_i^2), x_i ln(x_i)$. Note that the Barlett test could be implemented as $\boldsymbol{B}$ is nothing but the covariance matrix between $\boldsymbol{w}_i$ and $\boldsymbol{Q}(x_i)$. A consistent estimator of $\boldsymbol{B}$ is given by $\hat{\boldsymbol{B}} = T^{-1}\sum_{i=1}^{N}\boldsymbol{w}_i\boldsymbol{Q}(x_i)'$, so that $\sqrt{N}(\hat{\boldsymbol{B}} - \boldsymbol{B}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$, 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.

A Related strategy which doesn't make use of the list of instruments $\boldsymbol{Q}(x_i)'$ is the following. Denote $\boldsymbol{F}(x_i) = \boldsymbol{A}\boldsymbol{G}(x_i)$. Estimation of $\boldsymbol{F}(x_i)$ can be accomplished by means of kernel estimation. Under the null that there is a representation $\boldsymbol{A}\boldsymbol{G}(x_i)$ for $\boldsymbol{F}(x_i)$ which is of reduced rank it must be the case that $\boldsymbol{B} = E\{\boldsymbol{F}(x_i)\boldsymbol{w}_i\}$ is of reduced rank. Testing for the number of positive canonical correlations of the vector series $\boldsymbol{F}(x_i)$ and $\boldsymbol{w}_i$ is equivalent to testing for the rank of $\boldsymbol{B}$, i.e. Barlett's test is applicable. Also, noting that an estimate for $\boldsymbol{B}$ is computed as $\hat{\boldsymbol{B}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_i\boldsymbol{F}(x_i)'$, and once more a consistent estimator for $\boldsymbol{V}$ can be constructed.

This second strategy is related to the approach described in Donald (1997). Notwithstanding, Donald (1997) rather than applying any of the tests reviewed above, suggested an

alternative test. If the number of nonparametric factors is smaller than m, this implies that there exists a matrix $\boldsymbol{c}$ of order $k \times (k-m)$ such that:

$$\boldsymbol{w}_i'\boldsymbol{c} = \boldsymbol{F}(x_i)'\boldsymbol{c} + \boldsymbol{\varepsilon}_i'\boldsymbol{c} = \mathbf{0} + \boldsymbol{e}_i$$

where $\boldsymbol{e}_i = \boldsymbol{\varepsilon}_i'\boldsymbol{c}$. Donald (1997) suggested to test for the number of nonparametric factors m by means of testing for the moment restriction $E\{\boldsymbol{c}'\boldsymbol{F}(x_i)\boldsymbol{w}_i'\boldsymbol{c}\} = \mathbf{0}$. Their proposed test was constructed from the eigenvalues of a sample analogue of this expression with an appropriate normalization, see Donald (1997) for further details.

The rank of the demand system has important implications for demand theory, see Lewbel (1991) for a detailed review. Under the setting in (9) 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. Gorman (1981) suggested the following alternative specification for demand systems

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

where additionally P is a vector of prices. Under this specification, the rank must be smaller than three for demands to be aggregable[5]. 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 analysis of shifts in relative prices.

[5]See Lewbel (1991) and references therein for further studies of exactly aggregable demands.

4.4 Reduced Rank VAR Model

Consider a conventional VAR of the form:

$$\boldsymbol{y}_t = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{\varepsilon}_t \tag{10}$$

each of the $\boldsymbol{A}_k$ is an $m \times m$ matrix. $\boldsymbol{\epsilon}_t$ 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:

$$\boldsymbol{y}_t = \boldsymbol{F}\left[\sum_{k=1}^{p} \boldsymbol{G}_k \boldsymbol{y}_{t-k}\right] + \boldsymbol{\varepsilon}_t \tag{11}$$

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. Velu, Reinsel, and Wichern (1986) suggested a method for estimating the parameters $\boldsymbol{F}$ and $\left[\begin{array}{cccc} \boldsymbol{G}_1 & \boldsymbol{G}_2 & \ldots & \boldsymbol{G}_k \end{array}\right]$ in (11) conditional on a given r. Denote $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \boldsymbol{y}'_{t-2}, \boldsymbol{y}'_{t-k})'$ and $\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}} = \boldsymbol{\Omega}_{\boldsymbol{yy}} - \boldsymbol{\Omega}_{\boldsymbol{yx}}\boldsymbol{\Omega}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Omega}_{\boldsymbol{xy}}$ where $\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}$ is the covariance of the residuals of the OLS unrestricted regression of (10) and $\boldsymbol{\Omega}_{\boldsymbol{xy}}$ is the covariance matrix between $\boldsymbol{x}$ and $\boldsymbol{y}$. Additionally denote $\boldsymbol{\Pi} = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}^{-1}$ and set $\boldsymbol{v}_j$ to be the eigenvector corresponding to the jth largest eigenvalue of $\boldsymbol{\Pi}^{\frac{1}{2}}\boldsymbol{\Omega}_{\boldsymbol{yx}}\boldsymbol{\Omega}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Omega}_{\boldsymbol{xy}}\boldsymbol{\Pi}^{\frac{1}{2}}$, λ_j^2. If $\boldsymbol{V}_r = \left[\begin{array}{cccc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_r \end{array}\right]$ then

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

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 $\boldsymbol{A}_k$'s which are assumed to have common rank. Consider the RRVAR model (11) re-expressed as

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \tag{12}$$

$t = 1, ..., T$, where the (m, mp) matrix $\boldsymbol{B} \equiv \boldsymbol{\alpha}\boldsymbol{\beta}'$. 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, $T^{1/2}vec(\hat{\boldsymbol{B}} - \boldsymbol{B}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX}^{-1} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon\epsilon}})$, where $\boldsymbol{\Sigma}_{XX} \equiv E\{\boldsymbol{x}_t\boldsymbol{x}'_t\}$ is assumed positive definite which holds if $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ is a non-defective process. Given this distribution, computation of the tests of Cragg and Donald or Robin and Smith follows.

Reduced rank regression models like that in (12) 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.

4.5 Nested reduced rank VAR models

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*.

To identify the ranks of the $\boldsymbol{A}_k$ matrices in (10) use is made of the canonical correlation approach described by Tsay and Tiao (1985) and Tiao and Tsay (1985). Define $\boldsymbol{Y}_{s,t} = \left(\boldsymbol{y}_t', \ldots, \boldsymbol{y}_{t-s}'\right)'$. The rank of $\boldsymbol{A}_s$ is equivalent to the number of non-zero canonical correlations between $\boldsymbol{Y}_{s-1,t}$ and $\boldsymbol{Y}_{s-1,t-1}$. This, as stated in section 2, is equivalent to the rank of the covariance matrix between $\boldsymbol{Y}_{s-1,t}$ and $\boldsymbol{Y}_{s-1,t-1}$. The Bartlett test described above is therefore of use in this context. This strategy allows also to identify the order of the VAR, since for $s > p$ there will be a minimum of m zero canonical correlations between $\boldsymbol{Y}_{s-1,t}$ and $\boldsymbol{Y}_{s-1,t-1}$. 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.

4.6 Dynamic Factor Models

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

$$\boldsymbol{y}_t = \boldsymbol{P}\boldsymbol{z}_t + \boldsymbol{\varepsilon}_t \tag{13}$$

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.[6] 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 (13), 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.

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 $\boldsymbol{z}_t = \boldsymbol{P}^+\boldsymbol{y}_t + \boldsymbol{P}^+\boldsymbol{\varepsilon}_t$, 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. Maximum likelihood estimation of this system is easily implemented once the model is formulated in state space form.

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 analysis 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

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

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 that forecasts based on these factors outperformed various benchmark models.

4.7 State Space models

We focus on the State Space representation in the *innovation form*, i.e.:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{C}\boldsymbol{s}_t + \boldsymbol{e}_t \\ \boldsymbol{s}_{t+1} &= \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{e}_t \end{aligned} \tag{14}$$

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 $\boldsymbol{\Omega}$. This system can be characterized by a system transfer function matrix $\boldsymbol{G}(z) = \sum_{i=1}^{\infty} \boldsymbol{G}_i z^{-1}$, 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. 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 approach to state space multivariate modelling. Corresponding to the transfer function matrix $\boldsymbol{G}(z)$ above, the infinite dimensional Hankel matrix is defined as:

$$\boldsymbol{H} = \mathcal{O}\mathcal{C} = \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{G}_2 & \boldsymbol{G}_3 & \cdots \\ \boldsymbol{G}_2 & \boldsymbol{G}_3 & \cdots & \cdots \\ \boldsymbol{G}_3 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^2\boldsymbol{B} & \cdots \\ \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^2\boldsymbol{B} & \cdots & \cdots \\ \boldsymbol{C}\boldsymbol{A}^2\boldsymbol{B} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{15}$$

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 $\mathcal{C} = \left[\boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, (\boldsymbol{A}^2)\boldsymbol{B}, \cdots\right]$. 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)). The computation of the rank of the Hankel matrix is not an easy task, as it is unlikely that the impulse response matrices are given exactly, and in a majority of cases they are estimated. Furthermore, searching for the rank of the Hankel matrix is not conducted directly on an estimate of (15) 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, $\boldsymbol{y}_t$.

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

where $\boldsymbol{\Delta}_i$ is the autocorrelation matrix of $\boldsymbol{y}_t$ for lag i. Where $\boldsymbol{\mathcal{O}}$ is the observability matrix defined above, and $\overline{\boldsymbol{\mathcal{C}}} = \left[\overline{\boldsymbol{C}}, \boldsymbol{A}\overline{\boldsymbol{C}}, (\boldsymbol{A}^2)\overline{\boldsymbol{C}}, \cdots\right]$, 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:

$$\boldsymbol{H}^a = E\left(\boldsymbol{y}_+^t \boldsymbol{y}_-^{t\,\prime}\right) = E\begin{pmatrix} \boldsymbol{y}_{t+1} \\ \cdots \\ \boldsymbol{y}_{t+k} \end{pmatrix}\begin{pmatrix} \boldsymbol{y}_t' & \cdots & \boldsymbol{y}_{t-p+1}' \end{pmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_1 & \boldsymbol{\Delta}_2 & \cdots & \boldsymbol{\Delta}_p \\ \boldsymbol{\Delta}_2 & \boldsymbol{\Delta}_3 & \cdots & \boldsymbol{\Delta}_{p+1} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{\Delta}_k & \boldsymbol{\Delta}_{k+1} & \cdots & \boldsymbol{\Delta}_{k+p-1} \end{bmatrix} \quad (16)$$

The truncation parameters k and p must be fixed, and setting them implies a trade off between generality in model specification and modelling $\boldsymbol{\Delta}_i$ at very distant lags; see Aoki and Havenner (1991) for further details. The representation of the Hankel matrix stated in equation (16) shows that the Barlett test could be used to test for the rank of this matrix, and by extension also the information criteria procedures and the Bias Correction Barlett test are valid. While the matrix $\boldsymbol{V}^a$ is of reduced rank, the rank of $\hat{\boldsymbol{V}}^a$ is only of reduced rank asymptotically. This is problematic for the Cragg and Donald (1996) procedure. An estimator of $\boldsymbol{V}^a$ with equal rank to $\boldsymbol{V}^a$ can be constructed as in Camba-Mendez and Kapetanios (2001). Results in Camba-Mendez and Kapetanios (2004) further showed that bootstrapped procedures of those tests or rank presented above significantly improve upon the performance of the corresponding asymptotic tests, and that statistical tests have in general a better performance than standard information criteria methods in the identification of the dimensionality of these systems.

Kapetanios (2004) and Camba-Mendez and Kapetanios (2005b) used a state space model like that in (14) to compute measures of underlying inflation extracted from a vector series that contained all available sub-components of consumer price indices. Measures of underlying inflation are commonly used to formulate monetary policy and assist in forecasting observed inflation. Camba-Mendez and Kapetanios (2005b) explored empirically the forecasting ability of core inflation measures built using state space models against those built using more

traditional techniques, and found them to perform better than traditional measures over medium to long forecasting horizons.

4.8 Cointegration

The methods to test for cointegration most usually encounter in applied economics work are those of Johansen (1988), Stock and Watson (1988), Gregoir and Laroque (1994) and Snell (1999). Their tests are review in many Econometrics textbook. In this section we will focus instead in the strategy proposed by Camba-Mendez and Kapetanios (2005a), reviewed above. 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. 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. Test 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. In what follows we present an estimate of the spectral density matrix at any frequency together with its distribution.

Denote a zero mean, wide sense stationary m-vector process by $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$. The spectral density matrix of $\boldsymbol{x}_t$ is defined as

$$\boldsymbol{\Sigma}(\omega) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_k e^{-ik\omega}$$

for $\theta \in [-\pi, \pi]$ where $\boldsymbol{\Gamma}_k = E\{\boldsymbol{x}_t \boldsymbol{x}'_{t-k}\}$. Given a sample of T observations an estimate of the spectral density matrix is given by:

$$\overline{\boldsymbol{\Sigma}}(\omega) = (2\pi)^{-1} \sum_{k=-(T-1)}^{T-1} \hat{\boldsymbol{\Gamma}}_k e^{-ik\omega}$$

where $\hat{\boldsymbol{\Gamma}}_k = \frac{1}{T} \sum_{t=1}^{T-|k|} \boldsymbol{x}_t \boldsymbol{x}'_{t-k}$. $2\pi\overline{\boldsymbol{\Sigma}}(\omega)$ is the periodogram. The periodogram provides an inconsistent but asymptotically unbiased estimate of the spectral density matrix, and is asymptotically distributed as a complex Wishart variable with 1 degree of freedom. A standard approach for consistent estimation of the spectral density matrix[7] relies on 'smoothing' the periodogram itself over the frequencies, i.e. averaging adjacent frequency ordinates.

[7] As we are mainly interested in the rank of the spectral density matrix, in the rest of the discussion we drop the normalizing constant 2π.

These estimates take the form,

$$\hat{\boldsymbol{\Sigma}}(\omega) = \frac{1}{2M+1} \sum_{k=-M}^{M} \overline{\boldsymbol{\Sigma}}(\omega + k/T)$$

For M fixed as $T \to \infty$ this estimate is still inconsistent, asymptotically unbiased for the spectral density matrix and asymptotically distributed as $(2M+1)^{-1}W^C(2M+1, \boldsymbol{\Sigma}(\omega_j))$, (see Brillinger (1981, pp. 245)). This is the simplest form of a smoothed periodogram estimate for the spectral density matrix. Different weights can be assigned to the periodogram coordinates $\bar{\boldsymbol{\Sigma}}(\omega + k/T)$, see Brillinger (1981, Chapter 7). If we allow $M \to \infty$ as $T \to \infty$ but impose $M^4/T \to 0$ we get a consistent and asymptotically normal estimate (see e.g. Newey and West (1987)). In particular we get that $\sqrt{2M+1}(vec(\hat{\boldsymbol{\Sigma}}(\omega)) - vec(\boldsymbol{\Sigma}(\omega)))$ is asymptotically complex normal[8] with a covariance matrix whose element giving the asymptotic covariance between $\hat{\boldsymbol{\Sigma}}_{i,j}(\omega)$ and $\hat{\boldsymbol{\Sigma}}_{u,v}(\omega)$, is given by:

$$\begin{array}{ll} \boldsymbol{\Sigma}_{i,u}(\omega)\boldsymbol{\Sigma}_{j,v}(\omega) + \boldsymbol{\Sigma}_{i,v}(\omega)\boldsymbol{\Sigma}_{j,u}(\omega) & \text{if } \omega = 0, \pm\pi \\ \boldsymbol{\Sigma}_{i,u}(\omega)\boldsymbol{\Sigma}_{j,v}(\omega) & \text{if } \omega \neq 0, \pm\pi \end{array}$$

where $\boldsymbol{\Sigma}_{i,j}(\omega)$ is the (i,j)-th element of $\boldsymbol{\Sigma}(\omega)$. We will denote this covariance matrix by $\boldsymbol{V}$ and its estimate, obtained by using the estimated spectral density matrix, by $\hat{\boldsymbol{V}}$. More details may be found in e.g. Brillinger (1981, pp. 262) or Brockwell and Davis (1991, pp. 447).

4.9 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:

$$\boldsymbol{y}_t = \boldsymbol{C}(L)\boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_t \tag{17}$$

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:

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

[8]For more details on the choice of M and its effect on the asymptotic bias and variance of the estimator see also Brillinger (1981, Chapter 2).

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

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

and

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

where $\boldsymbol{V}_k(\alpha)$ are the k eigenvectors of the spectral density matrix $\boldsymbol{\Sigma}(\alpha)$ associated with the largest eigenvalues, see Brillinger (1981) for further details.

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 (17), but without the 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) and Forni, Hallin, Lippi, and Reichlin (2000) where a 'generalized dynamic factor' model, novel to the literature, is proposed.

5 Conclusion

This paper has described several general tests of rank of a matrix. Furthermore, 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 Variables models, and dynamic principal component models, to nonlinear frameworks such as nonparametric factor models and also to nonstationary frameworks such as cointegrated systems.

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.

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

——— (2005b): "Forecasting Inflation using Dynamic Factor Measures of Underlying inflation," *Journal of Forecasting*, forthcoming.

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., and M. Rothschild (1983): "Arbitrage and mean variance analysis on large asset markets," *Econometrica*, 51, 1281–1304.

Connor, G., and R. Korajczyk (1988): "Performance measurement with the Arbitrage Pricing Theory: A new framework for analysis," *Journal of Financial Economics*, 21, 373–394.

Connor, G., and R. A. Korajczyk (1993): "A test for the number of factors in an approximate factor model," *Journal of Finance*, 48, 1263–1291.

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.

Dhrymes, P. J., I. Friend, and N. B. Gultekin (1984): "A critical reexamination of the empirical evidence on the arbitrage pricing theory," *Journal of Finance*, 39, 323–346.

Donald, S. G. (1997): "Inference concerning the number of factors in a multivariate nonparametric relationship," *Econometrica*, 65(1), 103–131.

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., V. K. Ng, and M. Rothschild (1990): "Asset Pricing with a FACTOR-ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213–237.

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.

——— (2000): "Reference Cycles: the NBER methodology revisited," CEPR Discussion Paper No 2400.

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. V. Loan (1983): *Matrix Computations*. North Oxford Academic.

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.

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

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.

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

Ingersoll, J. E. (1984): "Some results in the theory of arbitrage pricing," *Journal of Finance*, 39, 1021–1039.

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.

Lehmann, B., and D. Modest (1988): "The Empirical Foundations of the Arbitrage Pricing Theory," *Journal of Financial Economics*, 21, 213–254.

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.

Lutkepohl, H., and M. M. Burda (1997): "Modified Wald Tests Under Nonregular Conditions," *Journal of Econometrics*, 78, 315–332.

Moore, D. S. (1977): "Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests of Fit," *Journal of the American Statistical Association*, 72(357), 131–137.

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

Newey, W., and K. West (1987): "A Simple Positive Semi Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

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.

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.

Roll, R. (1984): "A critical reexamination of the empirical evidence on the arbitrage pricing theory," *Journal of Finance*, 39, 347–350.

Roll, R., and S. A. Ross (1980): "An empirical investigation of the arbitrage pricing theory," *Journal of Finance*, 35, 1073–1103.

Ross, S. A. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

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.

Shapiro, A. (1984): "A note on the consistency of estimators in the analysis of moment structures," *British Journal of Mathematical and Statistical Psychology*, 37, 84–88.

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.

Tiao, G. C., and R. S. Tsay (1985): "A Canonical Correlation Approach to Modeling Multivariate Time Series," In Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 112-120.

Tsay, R. S., and G. C. Tiao (1985): "Use of Canonical Analysis in Time Series Model Identification," *Biometrika*, 72, 299–315.

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 Proof of proposition 1

Statement i holds from the continuity of $F\left(\hat{\boldsymbol{\theta}}, \boldsymbol{h}(\boldsymbol{\mu})\right)$, and assumptions 1.a and 1.b as shown in Shapiro (1984). Proof of ii) goes as follows. By noting that $\hat{\boldsymbol{\mu}} \xrightarrow{a.s.} \boldsymbol{\mu}_0$, and by taking Taylor expansions of $\boldsymbol{h}(\boldsymbol{\mu})$ around $\boldsymbol{\mu}_0$ it follows that:

$$\hat{\psi} \xrightarrow{a.s.} \min_{\boldsymbol{\mu} \in \Re^p} F\left(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}_0(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right) = \min_{\boldsymbol{\eta} \in \Re^p} F\left(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}_0 \boldsymbol{\eta}\right)$$

where $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\theta}} - \boldsymbol{h}(\boldsymbol{\mu}_0)$. Define by $\tilde{\boldsymbol{z}}$ that $\boldsymbol{\eta}$ that solves that minimization problem, i.e. $\tilde{\boldsymbol{z}} = \left(\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\Delta}_0\right)^{-1} \boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\varepsilon}$, then it follows that:

$$\min_{\boldsymbol{\eta} \in \Re^p} F\left(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}_0 \boldsymbol{\eta}\right) = F\left(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}_0 \tilde{\boldsymbol{z}}\right) = T \boldsymbol{\varepsilon}' \left\{\boldsymbol{I} - \boldsymbol{M}\right\}' \boldsymbol{V}^+ \left\{\boldsymbol{I} - \boldsymbol{M}\right\} \boldsymbol{\varepsilon}$$

where now we define $\boldsymbol{M} = \boldsymbol{\Delta}_0 \left(\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\Delta}_0\right)^{-1} \boldsymbol{\Delta}_0' \boldsymbol{V}^+$. It is then easy to show that

$$\left\{\boldsymbol{I} - \boldsymbol{M}\right\}' \boldsymbol{V}^+ \left\{\boldsymbol{I} - \boldsymbol{M}\right\} = \boldsymbol{V}^+ - \boldsymbol{M}' \boldsymbol{V}^+$$

It is then possible to write:

$$\hat{\psi}^g \xrightarrow{a.s.} T \boldsymbol{\varepsilon}' \boldsymbol{V}^+ \boldsymbol{\varepsilon} - T \boldsymbol{\varepsilon}' \boldsymbol{V}^+ \boldsymbol{\Delta}_0 \left(\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\Delta}_0\right)^{-1} \boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\varepsilon}$$

Following arguments in Moore (1977), and the provisions made at the beginning of this section, it follows that the first summand in the equation above is distributed as a chi-squared with degrees of freedom r^v. It is easy to show that $\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\varepsilon}$ is a $p \times 1$ normally distributed vector process with mean zero and covariance matrix $\left(\boldsymbol{\Delta}_0' \boldsymbol{V}^+ \boldsymbol{\Delta}_0\right)$, from which it follows that the second summand is distributed as chi-squared with degrees of freedom p, and this completes the proof. □

It is also easy to show that if $\boldsymbol{V}$ is nonsingular $\hat{\psi}^g$ would converge to $T \boldsymbol{\varepsilon}' \boldsymbol{V}^{-\frac{1}{2}} \left(\boldsymbol{I} - \tilde{\boldsymbol{M}}\right) \boldsymbol{V}^{-\frac{1}{2}} \boldsymbol{\varepsilon}$, where $\tilde{\boldsymbol{M}} = \left(\boldsymbol{I} - \boldsymbol{V}^{-\frac{1}{2}} \boldsymbol{\Delta}_0 \left(\boldsymbol{\Delta}_0' \boldsymbol{V}^{-1} \boldsymbol{\Delta}_0\right)^{-1} \boldsymbol{\Delta}_0' \boldsymbol{V}^{-\frac{1}{2}}\right)$ is an idempotent matrix of rank equal to $(q - p)$. Further noting that $\boldsymbol{V} - \frac{1}{2} \boldsymbol{\varepsilon}$ is a $q \times 1$ normally distributed vector process with mean zero and covariance matrix $\boldsymbol{I}$, it thus follows that $\hat{\psi}^g$ would be distributed as a chi-squared with degrees of freedom $q - p$.

A.2 Proof of proposition 2

Proof of boundedness and uniqueness are provided in lemmas 1 and 2 in Cragg and Donald (1997). Full column rank of $\boldsymbol{\Delta}_h$ follows directly from (2). □

A.3 Distribution of Λ for a hermitian positive semidefinite matrix.

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

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

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

$$\frac{\partial vec(\boldsymbol{\Lambda}^R)}{\partial vec(\boldsymbol{\Sigma}_{11}^R)'} = \left(\boldsymbol{\Sigma}_{12}^{R\,\prime} \otimes \boldsymbol{\Sigma}_{21}^R\right) \boldsymbol{D}_1 \boldsymbol{D}_1^+ \left({\boldsymbol{\Sigma}_{11}^R}^{-1} \otimes {\boldsymbol{\Sigma}_{11}^R}^{-1}\right) \boldsymbol{D}_1 \boldsymbol{J}_1^h {\boldsymbol{J}_1^h}^+ \boldsymbol{D}_1^+ \tag{18}$$

$$\frac{\partial vec(\boldsymbol{\Lambda}^R)}{\partial vec(\boldsymbol{\Sigma}_{21}^R)'} = -\left(\boldsymbol{I}_{4(n-r)^2} + \boldsymbol{K}_{2(n-r),2(n-r)}\right) \left(\boldsymbol{\Sigma}_{21}^R{\boldsymbol{\Sigma}_{11}^R}^{-1} \otimes \boldsymbol{I}_{2(n-r)}\right) \boldsymbol{J}_{21}^h {\boldsymbol{J}_{21}^h}^+ \tag{19}$$

$$\frac{\partial vec(\boldsymbol{\Lambda}^R)}{\partial vec(\boldsymbol{\Sigma}_{12}^R)'} = \frac{\partial vec(\boldsymbol{\Lambda}^R)}{\partial vec(\boldsymbol{\Sigma}_{21}^R)'} \boldsymbol{K}_{2r,2(n-r)}, \qquad \frac{\partial vec(\boldsymbol{\Lambda}^R)}{\partial vec(\boldsymbol{\Sigma}_{22}^R)'} = \boldsymbol{D}_2 \boldsymbol{J}_2^h {\boldsymbol{J}_2^h}^+ \boldsymbol{D}_2^+ \tag{20}$$

where for a matrix $\boldsymbol{A}$, $\boldsymbol{A}^+ = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, $\boldsymbol{K}_{m,n}$ is a commutation matrix (see Lütkepohl (1996, Sec. 9.2)). (18), (19) and (20) 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. Then, $\sqrt{2M+1}\left(Re\ vec(\hat{\boldsymbol{\Lambda}})', Im\ vec(\hat{\boldsymbol{\Lambda}})'\right)' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{W}^r)$ where $\boldsymbol{W}^r = \boldsymbol{J}\boldsymbol{V}^r\boldsymbol{J}'$ and $\boldsymbol{J} = \boldsymbol{J}_2\boldsymbol{R}\boldsymbol{J}_1$. Finally, $\sqrt{2M+1}vec(\hat{\boldsymbol{\Lambda}}) \xrightarrow{d} N^C(\mathbf{0}, \boldsymbol{W})$. 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).

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm