Econometric Theory, **16**, 2000, 151–175. Printed in the United States of America.

TESTS OF RANK

Jean-Marc Robin
INRA and CREST-INSEE

Richard J. Smith
University of Bristol

This paper considers tests for the rank of a matrix for which a root-T consistent estimator is available. However, in contrast to tests associated with the minimum chi-square and asymptotic least squares principles, the estimator's asymptotic variance matrix is not required to be either full or of known rank. Test statistics based on certain estimated characteristic roots are proposed whose limiting distributions are a weighted sum of independent chi-squared variables. These weights may be simply estimated, yielding convenient estimators for the limiting distributions of the proposed statistics. A sequential testing procedure is presented that yields a consistent estimator for the rank of a matrix. A simulation experiment is conducted comparing the characteristic root statistics advocated in this paper with statistics based on the Wald and asymptotic least squares principles.

1. INTRODUCTION

Establishing the rank of a matrix is an important problem in a wide variety of econometric and statistical contexts. For example, the classical identification problem in linear simultaneous equation models involves the rank of particular submatrices of the reduced form parameters (see, among others, Rothenberg, 1973). Moreover, in a more general likelihood setting, there is an intimate relationship between the rank of the information matrix and the identifiability of a vector of parameters (see Hsiao, 1983). Such notions may be suitably adapted for the identifiability of parameters estimated by other methods. The need for knowledge of the rank of particular matrices also arises in many other situations in econometrics; for example, the rank of the substitution matrix in systems of demand equations (Lewbel, 1991; for other examples, see, among others, Cragg and Donald, 1997; Gill and Lewbel, 1992). Moreover, in misspecified models, asymptotic covariance matrices required for inference purposes may be singular. Hence, for statistics to possess the usual limiting chi-squared behavior, a consistent estima-

We are grateful to the Editor and three anonymous referees for their helpful and constructive suggestions on and criticisms of an earlier version of this paper, which have considerably assisted us in improving the exposition of the results herein. We also thank L. Gill, A. Lewbel, M.H. Pesaran, and K.J. van Garderen, participants at the 1994 Econometric Study Group Conference, University of Bristol, and seminar participants at the Universities of Cambridge and Exeter, where earlier versions of this paper were presented, for their comments. The second author is grateful to the Departement de la Recherche, CREST-INSEE, Paris, for its hospitality and to ESRC projects R 00023 4072 and R 00023 7386 for partial financial support. Address correspondence to: Richard J. Smith, Department of Economics, University of Bristol, Alfred Marshall Building, 8 Woodland Road, Bristol BS8 1TN, United Kingdom.

 0266-4666/00 $9.50

tor for a generalized inverse is needed that may be problematic if the asymptotic covariance matrix is of unknown rank. (On conditions for the consistent estimation of generalized inverses of matrices, see Andrews, 1987.)

This paper is concerned with tests for the rank of a matrix that is unobserved but for which a root-T consistent (RTC) estimator is available. Gill and Lewbel (1992) were the first authors to consider this problem. However, their solution based on estimators for the zero pivots obtained from an LDU decomposition of the RTC matrix estimator is in error as was shown by Cragg and Donald (1996). Cragg and Donald (1997) provided tests for the rank of a matrix in this framework based on a minimum chi-squared (MC) criterion (Ferguson, 1958; Rothenberg 1973). However, as they recognized, an essential feature for the application of the MC principle to testing for rank is knowledge of the rank of the asymptotic variance matrix of the limiting normal distribution of the RTC matrix estimator.

Therefore, an important departure for the approach taken in this paper is that no explicit assumptions are made concerning the rank or structure of the asymptotic covariance matrix of the limiting normal distribution of the RTC matrix estimator. In particular, we allow the asymptotic variance matrix to be less than full rank, and, moreover, this rank may be unknown. For example, the matrix of interest may be subject to (possibly unknown) a priori nonlinear restrictions and may have been estimated as such which will yield an asymptotic covariance matrix for the RTC estimator that is of less than full rank. Unlike our method, the MC and other test procedures for rank such as asymptotic least squares (ALS) (Gouriéroux, Monfort, and Trognon, 1985) require that the form of these a priori restrictions and, hence, the rank of the metric employed in these procedures are *known.* (For further discussion of tests of rank based on MC and ALS criteria in such circumstances, see Robin and Smith, 1995.) However, we do require an assumption that the rank of a matrix involving the asymptotic covariance matrix for the RTC estimator and matrices of certain characteristic vectors is nonzero; this assumption is empirically nonverifiable without further information on the constituent matrices.

The main focus of the paper is variants of a statistic originally proposed by Anderson (1951), which is a functional of certain estimated characteristic roots. Anderson's statistic is a likelihood ratio test for the rank of a regression coefficient matrix in a multivariate normal linear (MNL) model. In the MNL model, Anderson's statistic has a limiting chi-squared distribution under the null hypothesis that the matrix possesses a given rank. This result arises because the asymptotic variance matrix of the limit normal distribution of the maximum likelihood (least squares) estimator for the regression coefficient matrix has a special Kronecker product structure. We relax this assumption to allow for a general asymptotic variance matrix that may not necessarily be of full rank and the rank of which may not be known. However, the limit distributions of the proposed test statistics under the null hypothesis of given rank depend on certain nuisance parameters that may be consistently estimated, and, thus, an estimator for these limit distributions may be obtained. A sequential testing procedure that is specific to general is advocated that, in an asymptotic sense, should never accept a lower rank for the matrix of interest than the true rank of the matrix and that, given a particular dependence of the asymptotic sizes at each stage of the sequence on the sample size, results in a consistent estimator for the rank of the matrix.[1]

Section 2 outlines the problem under consideration and defines notation together with certain assumptions. The basis for the testing procedures of this paper is also introduced in Section 2. Test statistics for the rank of a matrix based on certain characteristic roots are described in Section 3 and their limiting distributions derived. An estimation procedure for these limiting distributions is provided in Section 4, and critical regions for tests of the rank hypothesis are also provided. Section 5 details the consistency properties of the critical regions of Section 4 and provides a sequential procedure for the consistent estimation of the rank of a matrix. To evaluate the size and power properties of tests based on characteristic roots and other methods, Section 6 presents a simulation experiment. Section 7 concludes the paper. All proofs are relegated to the Appendix.

2. SOME PRELIMINARIES

2.1. The Problem

The unobserved (p,q) matrix $\mathbf{B}$ has unknown true rank r^* where $0 \leq r^* \leq \min(p,q)$, which we state formally as the following assumption.

Assumption 2.1. (Rank of $\mathbf{B}$) The (p,q) matrix $\mathbf{B}$, $p \geq q$, is finite and has rank r^* where $0 \leq r^* \leq q$.

The problem of interest concerns constructing tests for the rank of the unobserved matrix $\mathbf{B}$. We denote the hypothesis that the rank of $\mathbf{B}$ is equal to r by

$$H_r : rk(\mathbf{B}) = r, \tag{2.1}$$

where $0 \leq r \leq q$. Initially, Section 3 discusses tests for the null hypothesis $H_{r^*} : rk(\mathbf{B}) = r^*$ against the alternative hypothesis $H'_{r^*} : rk(\mathbf{B}) > r^*$, which are also the hypotheses considered by Cragg and Donald (1993, 1996, 1997). Furthermore, in Section 5, we consider the use of a sequential testing procedure for the null hypothesis $H_r : rk(\mathbf{B}) = r$ against the alternative hypothesis $H'_r : rk(\mathbf{B}) > r$, $r = 0,1,\ldots,q-1$, to reveal (at least asymptotically) the true rank r^* of the matrix $\mathbf{B}$.

2.2. Notation and Further Assumptions

Our second assumption concerns the information on the matrix $\mathbf{B}$ available to the researcher.

Assumption 2.2. (Root-T consistent estimator for $\mathbf{B}$) The estimator $\hat{\mathbf{B}}$ is root-T consistent for the (p,q) matrix $\mathbf{B}$, where $p \geq q$; that is,

$$T^{1/2}\,\mathrm{vec}(\hat{\mathbf{B}} - \mathbf{B}) \to^L N_{pq}(\mathbf{0},\boldsymbol{\Omega}), \tag{2.2}$$

where $\boldsymbol{\Omega}$ is finite and $rk(\boldsymbol{\Omega}) = s$, $0 < s \leq pq$.

The notation $\to^L$ denotes convergence in distribution, whereas convergence in probability is indicated by $\to^P$. The theoretical development of this paper makes no explicit a priori assumption regarding the rank s of the (pq, pq) asymptotic covariance matrix $\boldsymbol{\Omega}$. In particular, $\boldsymbol{\Omega}$ may be less than full rank. More importantly, our procedure does not require explicit knowledge of the rank or structure of $\boldsymbol{\Omega}$, unlike other procedures based on the MC (Cragg and Donald, 1997) and ALS (Gouriéroux et al., 1985) principles. Hence, we allow for the possibility that the rank and structure of $\boldsymbol{\Omega}$ are unknown.

The statistical basis of the procedure adopted in this paper involves a matrix quadratic form in the RTC estimator $\hat{\mathbf{B}}$ that is similar in structure to the statistic considered by Anderson (1951). Anderson's statistic is a likelihood ratio (LR) statistic for the rank of a matrix and is a functional of certain characteristic roots (CR's) of a matrix quadratic form. Anderson was concerned with a particular multivariate normal problem in which the matrix $\boldsymbol{\Omega}$ in (2.2) has a special Kronecker product structure; this special case of our results is discussed subsequently. To derive the large sample properties of the requisite functionals, in particular the CR's, of the matrix quadratic form employed in our procedure, it is necessary first to describe the properties of its population analogue.

Let $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ denote (p,p) and (q,q) positive definite matrices, respectively. Consider the (p,p) matrix quadratic form in $\mathbf{B}$ given by $\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$, which, under Assumption 2.2, is the population analogue of the statistic considered in (3.1) subsequently. Now, $rk(\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}') = rk(\mathbf{B})$. Therefore, testing for the rank of $\mathbf{B}$ is equivalent to testing for the rank of $\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$. Moreover, under Assumption 2.1, the matrix quadratic form $\boldsymbol{\Sigma}\,\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$ has r^* nonzero and $(p - r^*)$ zero CR's. We denote the ordered CR's of $\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$ by $\tau_1^2 \geq \cdots \geq \tau_{r^*}^2 > 0$ and $\tau_{r^*+1}^2 = \cdots = \tau_p^2 = 0$, which are the solutions to the determinantal equation

$$|\mathbf{B}\boldsymbol{\Psi}\mathbf{B}' - \tau^2\,\boldsymbol{\Sigma}^{-1}| = 0. \tag{2.3}$$

The characteristic vector (CV) associated with τ_i^2 from (2.3) is denoted by $\mathbf{c}_i$, such that $\mathbf{c}_i'\boldsymbol{\Sigma}^{-1}\mathbf{c}_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta, $i, j = 1,\ldots,p$. The (p,p) matrix $\mathbf{C} \equiv (\mathbf{c}_1,\ldots,\mathbf{c}_p)$ collects as columns the CV's $\mathbf{c}_i$, $i = 1,\ldots,p$, and the columns of $\mathbf{C}$ are partitioned as $\mathbf{C} = (\mathbf{C}_{r^*}, \mathbf{C}_{p-r^*})$ conformably with respect to the r^* nonzero and $(p - r^*)$ zero CR's $\{\tau_i^2\}$ of $\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$. Hence, $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}' = \mathbf{C}_{r^*}\mathbf{C}_{r^*}' + \mathbf{C}_{p-r^*}\mathbf{C}_{p-r^*}'$. For a unique CR τ_i^2, the corresponding CV $\mathbf{c}_i$ is identified up to a normalization on its length, for example, $\|\mathbf{c}_i\| = 1$, whereas additionally for multiple roots, including $\tau_i^2 = 0$, $i = r^* + 1,\ldots,p$, the corresponding CV's are identified up to an orthonormal matrix of dimension equal to the multiplicity of the roots.

An alternative representation for the CR's $\tau_1^2 \geq \cdots \geq \tau_{r^*}^2 > 0$ and $\tau_{r^*+1}^2 = \cdots = \tau_q^2 = 0$ corresponding to the (q,q) matrix quadratic form $\mathbf{\Psi B' \Sigma B}$ is obtained from the determinantal equation

$$|\mathbf{B}'\mathbf{\Sigma B} - \tau^2 \mathbf{\Psi}^{-1}| = 0. \tag{2.4}$$

The CV associated with τ_i^2 from (2.4) is denoted by $\mathbf{d}_i$, such that $\mathbf{d}_i'\mathbf{\Psi}^{-1}\mathbf{d}_j = \delta_{ij}$, $i,j = 1,\ldots,q$, and we define the (q,q) matrix of CV's $\mathbf{D} = (\mathbf{d}_1,\ldots,\mathbf{d}_q)$. Similarly to $\mathbf{C}$, we partition the columns of $\mathbf{D}$ conformably with respect to the r^* nonzero and $(q - r^*)$ zero CR's $\{\tau_i^2\}$ of $\mathbf{\Psi B' \Sigma B}$ as $\mathbf{D} = (\mathbf{D}_{r^*}, \mathbf{D}_{q-r^*})$. Thus, $\mathbf{\Psi} = \mathbf{DD}' = \mathbf{D}_{r^*}\mathbf{D}_{r^*}' + \mathbf{D}_{q-r^*}\mathbf{D}_{q-r^*}'$. Similar comments to those earlier for the CV's $\{\mathbf{c}_i\}$ are also applicable for the identification of the CV's $\{\mathbf{d}_i\}$.

The next assumption allows for the possibility that for particular unknown positive definite matrices $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ consistent estimators $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Psi}}$ are available as arises, for example, in the statistics proposed by Anderson (1951).

Assumption 2.3. (Consistent estimators for $\mathbf{\Sigma}$ and $\mathbf{\Psi}$) The (p,p) and (q,q) estimators $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Psi}}$ are positive semidefinite and weakly consistent for the finite positive definite matrices $\mathbf{\Sigma}$ and $\mathbf{\Psi}$, respectively; that is, $\hat{\mathbf{\Sigma}} - \mathbf{\Sigma} = o_P(1)$ and $\hat{\mathbf{\Psi}} - \mathbf{\Psi} = o_P(1)$.

As will become apparent in the next section, particular importance is attached to an interaction between the asymptotic variance matrix $\mathbf{\Omega}$ of (2.2) and the matrices of CV's $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$ corresponding to the zero CR's $\{\tau_i^2\}_{i=r^*+1}^p$ of (2.3) and $\{\tau_i^2\}_{i=r^*+1}^q$ of (2.4), respectively. The associated condition is only of relevance when the true rank of $\mathbf{B}$ is such that $r^* < q$.

Assumption 2.4. (Rank condition) If $r^* < q \leq p$, the $[(p - r^*)(q - r^*), (p - r^*)(q - r^*)]$ matrix $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})'\mathbf{\Omega}(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})$ is nonzero; that is,

$$rk[(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})'\mathbf{\Omega}(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})] > 0, \tag{2.5}$$

where the $(p, p - r^*)$ and $(q, q - r^*)$ matrices $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$ are defined following (2.3) and (2.4), respectively.

Assumption 2.4 equivalently states that $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*}) \notin \mathcal{N}(\mathbf{\Omega})$, where $\mathcal{N}(\cdot)$ denotes the null space (or kernel) of the matrix $(\cdot)$.[2] In other words, the columns of the matrix $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})$ do not all lie in the space spanned by the characteristic vectors associated with the zero characteristic roots of $\mathbf{\Omega}$. Under Assumption 2.2, the nullity of $\mathbf{\Omega}$ or the dimension of $\mathcal{N}(\mathbf{\Omega})$ is $pq - s$. As $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$ are full column rank $p - r^*$ and $q - r^*$, respectively, the matrix $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})$ is full column rank $(p - r^*)(q - r^*)$. Therefore, Assumption 2.4 is automatically satisfied if $r^{*2} - (p + q)r^* + s > 0$, which is true if and only if r^* lies inside the interval $[0,\{(p+q)-((p+q)^2-4s)^{1/2}\}/2)$. In general, however, as both r^* and s are unknown, or, more precisely, without more explicit knowledge of the characteristic vector structure of $\boldsymbol{\Omega}$, it is impossible to guarantee that either this condition or (2.5) of Assumption 2.4 will not be violated.[3] Of course, if $\boldsymbol{\Omega}$ is positive definite, that is, $s = pq$, Assumption 2.4 is automatically satisfied.

3. TEST STATISTICS FOR THE RANK OF A MATRIX USING CHARACTERISTIC ROOTS

The tests for the rank of $\mathbf{B}$ considered in this paper are based on functionals of the matrix quadratic form

$$\hat{\boldsymbol{\Sigma}}\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}'. \tag{3.1}$$

The ordered estimators of the CR's derived from $\hat{\boldsymbol{\Sigma}}\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}'$ of (3.1) are denoted as $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$, which solve the determinantal equation corresponding to (2.3),

$$|\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}' - \lambda\hat{\boldsymbol{\Sigma}}^{-1}| = 0. \tag{3.2}$$

LEMMA 3.1. (Consistency of the CR estimators) *If Assumptions* 2.2 *and* 2.3 *hold, then the ordered* CR *estimators* $\{\hat{\lambda}_i\}_{i=1}^p$ *that solve the determinantal equation* (3.2) *are consistent estimators for the corresponding ordered* CR*'s* $\{\tau_i^2\}_{i=1}^p$ *that solve the determinantal equation* (2.3).

Therefore, from Lemma 3.1, under Assumption 2.1, $\hat{\lambda}_i \to^P 0$, $i = r^*+1,\ldots,p$. As $p \geq q$, $\hat{\lambda}_i = 0$, $i = q+1,\ldots,p$.

The consistency of the CR estimators $\{\hat{\lambda}_i\}$ for their population counterparts $\{\tau_i^2\}$ obtained from (2.3) and (2.4) suggests basing a test for the null hypothesis that the rank of $\mathbf{B}$ is $r^*, H_{r^*}: rk(\mathbf{B}) = r^*$, against the alternative that the rank of $\mathbf{B}$ exceeds $r^*, H'_{r^*}: rk(\mathbf{B}) > r^*$, on a suitable functional of the CR estimators $\{\hat{\lambda}_i\}_{i=r^*+1}^q$ defined via (3.2).

First, however, it is necessary to derive a representation for the limiting distribution of the CR's $\{\hat{\lambda}_i\}_{i=r^*+1}^q$.

THEOREM 3.1. (Limiting distribution of the CR estimators for the zero CR's) *If Assumptions* 2.1–2.4 *hold, then the* CR *estimators* $T\hat{\lambda}_i$, $i = r^*+1,\ldots,q$, *from* (3.2), *have the same limiting distribution as the first* $(q-r^*)$ *ordered* CR*'s of the determinantal equation*

$$|\mathbf{C}'_{p-r^*}T^{1/2}(\hat{\mathbf{B}}-\mathbf{B})\mathbf{D}_{q-r^*}\mathbf{D}'_{q-r^*}T^{1/2}(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{C}_{p-r^*} - T\lambda\mathbf{I}_{p-r^*}| = 0, \tag{3.3}$$

where $\mathbf{C}_{p-r^*}$ *and* $\mathbf{D}_{q-r^*}$ *are defined following* (2.3) *and* (2.4), *respectively*.

Note that $(p-q)$ CR's of the determinantal expression (3.3) are also automatically zero as $p \geq q$.

As it stands, Theorem 3.1 is not particularly useful for formulating a test of $H_{r^*} : rk(\mathbf{B}) = r^*$ against $H'_{r^*} : rk(\mathbf{B}) > r^*$. However, a number of test statistics may be formulated in terms of particular functions of the CR estimators $\{\hat{\lambda}_i\}_{i=r^*+1}^{q}$.

Assumption 3.1. (Functions) The function $h(\cdot)$ is nonnegative, $h(z) \geq 0$, $0 \leq z < \infty$, and possesses continuous derivatives at least up to the first order such that $h(0) = 0$ and $h'(0) = 1$.

Particular examples of functions $h(\cdot)$ satisfying the conditions of Assumption 3.1 are $h(z) = [\exp(\mu z) - 1]/\mu$, $\mu \geq 0$, and the Box–Cox transformation $h(z) = [(1+z)^{\mu} - 1]/\mu$, $\mu \geq 0$. Other familiar examples are the logarithmic function $h(z) = \ln(1+z)$ and the identity function $h(z) = z$, which are special cases of the Box–Cox transformation; these functional forms are considered further later.

We are concerned with tests for the null hypothesis $H_{r^*} : rk(\mathbf{B}) = r^*$ against the alternative hypothesis $H'_{r^*} : rk(\mathbf{B}) > r^*$. Consider a CR statistic based on the functionals $h(\hat{\lambda}_i)$, $i = r^* + 1, \ldots, q$, defined by

$$\mathcal{CRT}_{r^*} \equiv T \sum_{i=r^*+1}^{q} h(\hat{\lambda}_i). \tag{3.4}$$

Note that from Theorem 3.1 and Assumption 3.1 $\mathcal{CRT}_{r^*} = T \sum_{i=r^*+1}^{q} \hat{\lambda}_i + o_P(1)$. Hence, this form of statistic (3.4) is analogous to the trace form of Anderson's (1951) LR statistic for testing the null hypothesis $H_{r^*} : rk(\mathbf{B}) = r^*$ against the alternative hypothesis $H'_{r^*} : rk(\mathbf{B}) > r^*$.

THEOREM 3.2. (Limiting distribution of the CR statistic) *If* $r^* < q$ *and Assumptions* 2.1–2.4 *and* 3.1 *hold*, *then the* CR *statistic*, $\mathcal{CRT}_{r^*} = T \sum_{i=r^*+1}^{q} h(\hat{\lambda}_i)$ *of* (3.4), *has a limiting distribution described by*

$$\sum_{i=1}^{t^*} \lambda_i^{r^*} Z_i^2,$$

where $t^* \leq \min\{s, (p - r^*)(q - r^*)\}$, $\lambda_1^{r^*} \geq \cdots \geq \lambda_{t^*}^{r^*}$ *are the nonzero ordered* CR'*s of the matrix*

$$(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})' \boldsymbol{\Omega} (\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*}) \tag{3.5}$$

and $\{Z_i\}_{i=1}^{t^*}$ *are independent standard normal variates.*

As $\{Z_i^2\} \sim \chi^2(1)$ are mutually independent, the limiting distribution given in Theorem 3.2 for the CRT statistic (3.4) is that of a weighted sum of t^* independent chi-squared variables, each with one degree of freedom, where the weights are given by the t^* nonzero CR's $\{\lambda_i^{r^*}\}$ of the matrix $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})' \boldsymbol{\Omega} (\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})$ in (3.5), which are identical to those of $(\mathbf{D}_{q-r^*}\mathbf{D}'_{q-r^*} \otimes \mathbf{C}_{p-r^*}\mathbf{C}'_{p-r^*})\boldsymbol{\Omega}$. Hence, as the matrices $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$ are identified up to postmultiplication by $(p - r^*, p - r^*)$ and $(q - r^*, q - r^*)$

orthonormal matrices, respectively, the result given in Theorem 3.2 is unaffected by the choice of identifying constraints.[4]

The limiting distribution for the CR test statistic (3.4) may be derived straightforwardly from Theorem 3.2 in the special case when $\mathbf{\Omega}$ takes the special Kronecker product structure $(\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1})$ as in Anderson (1951).

COROLLARY 3.1. (Limiting distribution of the CR statistic when $\mathbf{\Omega} = (\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1})$). *If $r^* < q$ and Assumptions 2.1–2.4 and 3.1 hold and $\mathbf{\Omega} = (\mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1})$, then the CR statistic, $\mathcal{CRT}_{r^*} = T\sum_{i=r^*+1}^{q} h(\hat{\lambda}_i)$ of (3.4), has a limiting $\chi^2[(p-r^*)(q-r^*)]$ distribution.*

Analogously to Anderson's (1951) LR test, a statistic may be based on the function $h(z) = \ln(1+z)$, which yields the LR form for the CR test statistic

$$\mathcal{CRT}_{r^*}^{LR} \equiv T \sum_{i=r^*+1}^{q} \ln(1+\hat{\lambda}_i). \tag{3.6}$$

Moreover, second, also in an analogous fashion, a Wald form for the CRT may also be defined using $h(z) = z$, namely,

$$\mathcal{CRT}_{r^*}^{W} \equiv T \sum_{i=r^*+1}^{q} \hat{\lambda}_i. \tag{3.7}$$

The next result immediately follows from the conditions of Theorem 3.2.

COROLLARY 3.2. (Limiting distribution of the LR and Wald forms of the CR statistic) *If the conditions of Theorem 3.2 hold, then the LR and Wald forms of the CR statistic, $\mathcal{CRT}_{r^*}^{LR} = T\sum_{i=r^*+1}^{q}\ln(1+\hat{\lambda}_i)$ of (3.6) and $\mathcal{CRT}_{r^*}^{W} = T\sum_{i=r^*+1}^{q}\hat{\lambda}_i$ of (3.7), have identical limiting distributions to that of the CR statistic given in Theorem 3.2.*

4. ESTIMATION OF THE LIMITING DISTRIBUTION OF THE CR TEST STATISTICS

To apply the results of Theorem 3.2 and Corollary 3.2, an estimator for the limiting distribution of the CR statistics, (3.4), (3.6), and (3.7), is required. We adopt the following notation for the cumulative distribution function (c.d.f.) of the random variable $\sum_{i=1}^{t^*}\lambda_i^{r^*}Z_i^2$, where $\{Z_i\}_{i=1}^{t^*}$ are independent standard normal variates, which characterizes the limiting distribution of these CR statistics:

$$F_{r^*}^{CRT}(c) \equiv \mathcal{P}\left\{\sum_{i=1}^{t^*}\lambda_i^{r^*}Z_i^2 \le c\right\}, \qquad c \ge 0. \tag{4.1}$$

Initially, suppose that $\mathbf{B}$ and the asymptotic variance matrix $\mathbf{\Omega}$ of (2.2) are known. Knowledge of the positive definite matrices $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ would imply that, under Assumption 2.1, the CV matrices $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$, given following (2.3) and (2.4), respectively, are available and, thus, the t^* nonzero CR's $\{\lambda_i^{r^*}\}_{i=1}^{t^*}$ of $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})'\boldsymbol{\Omega}(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*})$ in (3.5) are also known. Therefore, the c.d.f. $F_{r^*}^{CRT}(\cdot)$ of (4.1) may be obtained using the methods described in Davies (1980) and Farebrother (1980, 1984). Alternatively, this c.d.f. may be straightforwardly simulated given knowledge of the characteristic roots $\{\lambda_i^{r^*}\}_{i=1}^{t^*}$.

However, in general, the asymptotic variance matrix $\boldsymbol{\Omega}$ of (2.2) is not known. Hence, we assume that a positive semidefinite and weakly consistent estimator for $\boldsymbol{\Omega}$ is available to the researcher.

Assumption 4.1. (Consistent estimator for the asymptotic variance matrix $\boldsymbol{\Omega}$) The estimator $\hat{\boldsymbol{\Omega}}$ is positive semidefinite and weakly consistent for the asymptotic variance matrix $\boldsymbol{\Omega}$ of (2.2); that is, $\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} = o_P(1)$.

Consider the following estimator for the c.d.f. $F_{r^*}^{CRT}(\cdot)$ of (4.1):

$$\hat{F}_{r^*}^{CRT}(c) \equiv \mathcal{P}\left\{\sum_{i=1}^{(p-r^*)(q-r^*)} \hat{\lambda}_i^{r^*} Z_i^2 \leq c\right\}, \qquad c \geq 0, \tag{4.2}$$

where $\{\hat{\lambda}_i^{r^*}\}_{i=1}^{(p-r^*)(q-r^*)}$ are the ordered CR's of $(\hat{\mathbf{D}}_{q-r^*} \otimes \hat{\mathbf{C}}_{p-r^*})'\hat{\boldsymbol{\Omega}}(\hat{\mathbf{D}}_{q-r^*} \otimes \hat{\mathbf{C}}_{p-r^*})$ and $\{Z_i\}_{i=1}^{(p-r^*)(q-r^*)}$ are independent standard normal variates. The $(p, p - r^*)$ and $(q, q - r^*)$ matrices $\hat{\mathbf{C}}_{p-r^*} = (\hat{\mathbf{c}}_{r^*+1}, \ldots, \hat{\mathbf{c}}_p)$ and $\hat{\mathbf{D}}_{q-r^*} = (\hat{\mathbf{d}}_{r^*+1}, \ldots, \hat{\mathbf{d}}_q)$ are the estimated counterparts of the matrices of CV's $\mathbf{C}_{p-r^*}$ and $\mathbf{D}_{q-r^*}$ obtained from the characteristic equations $(\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}' - \hat{\lambda}_t \hat{\boldsymbol{\Sigma}}^{-1})\hat{\mathbf{c}}_i = \mathbf{0}$, $i = r^* + 1, \ldots, p$ and $(\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}\hat{\mathbf{B}} - \hat{\lambda}_i \hat{\boldsymbol{\Psi}}^{-1})\hat{\mathbf{d}}_i = \mathbf{0}$, $i = r^* + 1, \ldots, q$.

THEOREM 4.1. (Estimation of the limiting distribution of the CR statistics) *If $r^* < q$ and Assumptions* 2.1–2.4, 3.1, *and* 4.1 *hold, then the c.d.f. $\hat{F}_{r^*}^{CRT}(\cdot)$ of* (4.2) *converges to the c.d.f. $F_{r^*}^{CRT}(\cdot)$ of* (4.1); *that is*, $\hat{F}_{r^*}^{CRT}(c) - F_{r^*}^{CRT}(c) = o(1)$, $c \geq 0$.

Let $\hat{c}_{1-\alpha}^{r^*}$ denote the $100(1-\alpha)$ percentile of the c.d.f. $\hat{F}_{r^*}^{CRT}(\cdot)$ of (4.2); that is,

$$\mathcal{P}\left\{\sum_{i=1}^{(p-r^*)(q-r^*)} \hat{\lambda}_i^{r^*} Z_i^2 \geq \hat{c}_{1-\alpha}^{r^*}\right\} = \alpha.$$

Using Theorem 4.1, we may define critical regions based on the CR statistics (3.4), (3.6), and (3.7) of Section 3.

THEOREM 4.2. (Critical region of tests based on the CR statistics) *If $r^* < q$ and Assumptions* 2.1–2.4, 3.1, *and* 4.1 *hold, a test for the null hypothesis $H_{r^*} : rk(\mathbf{B}) = r^*$ against the alternative hypothesis $H'_{r^*} : rk(\mathbf{B}) > r^*$ with asymptotic size α, $0 < \alpha < 1$, is given by the critical region $\{\mathcal{CRT}_{r^*} \geq \hat{c}_{1-\alpha}^{r^*}\}$, where $\hat{c}_{1-\alpha}^{r^*}$ is the* $100(1-\alpha)$ *percentile of the c.d.f. $\hat{F}_{r^*}^{CRT}(\cdot)$ of* (4.2).

5. CONSISTENT ESTIMATION OF THE RANK OF A MATRIX

Of course, the true rank r^* of $\mathbf{B}$ is unknown. The next result is informative for constructing a sequential procedure to ascertain the true rank of the matrix $\mathbf{B}$.

THEOREM 5.1. (Consistency of the critical region of the CR statistic) *If* $r < r^* \leq q$ *and Assumptions* 2.1–2.4, 3.1, *and* 4.1 *hold*, *a test for the null hypothesis* $H_r: rk(\mathbf{B}) = r$ *against the alternative hypothesis* $H_r': rk(\mathbf{B}) > r$ *with critical region* $\{\mathcal{CRT}_r \geq \hat{c}^r_{1-\alpha_r}\}$ *is consistent*, *where* $\hat{c}^r_{1-\alpha_r}$ *is the* $100(1-\alpha_r)$ *percentile of the c.d.f.* $\hat{F}_r^{CRT}(\cdot)$ *defined in* (4.2).

Theorem 5.1 emphasizes that a sequential procedure testing $H_r: rk(\mathbf{B}) = r$ against $H_r': rk(\mathbf{B}) > r$ based on the CR statistics (3.4), (3.6), and (3.7), for $r = 0,1,\dots,q-1$, and halting at the first value for r for which the CR statistic indicates nonrejection of $H_r: rk(\mathbf{B}) = r$ will never asymptotically choose a value of r less than r^*. However, at the stage $r = r^*$, if $r^* < q$, there is a positive asymptotic probability α_{r^*} that the true hypothesis $H_{r^*}: rk(\mathbf{B}) = r^*$ will be rejected. Therefore, such a sequential procedure will not deliver a weakly consistent estimator for the true rank r^*, if $r^* < q$, without further elaboration. However, as the critical regions $\{\mathcal{CRT}_r \geq \hat{c}^r_{1-\alpha_r}\}$ are only defined for values of $r = 0,\dots,q-1$, if $r^* = q$, this sequential procedure does provide a weakly consistent estimator for r^*.

A weakly consistent estimator for the true rank r^* of the matrix $\mathbf{B}$ may be obtained with an appropriate adjustment dependent on T to the asymptotic size α_r of the CR test at each stage r of the sequential procedure, $r = 0,\dots,q-1$, based on the results of Pötscher (1983) and Bauer, Pötscher, and Hackl (1988). Cragg and Donald (1997, Sect. 3.2) used a similar approach to estimate r^* using statistics based on minimum chi-squared that unlike our method, however, required that the rank of the asymptotic variance matrix $\mathbf{\Omega}$ be known (for further discussion of this point, see Robin and Smith, 1995).

The revised critical region at stage r is given by $\{\mathcal{CRT}_r \geq \hat{c}^r_{1-\alpha_{rT}}\}$ with asymptotic size α_{rT} under $H_r: rk(\mathbf{B}) = r$, $r = 0,\dots,q-1$, and we define the estimator for $rk(\mathbf{B})$ as

$$\hat{r} \equiv \min_{r\in\{0,\dots,q-1\}} \{r: \mathcal{CRT}_i \geq \hat{c}^i_{1-\alpha_{iT}}, i = 0,\dots,r-1, \mathcal{CRT}_r < \hat{c}^r_{1-\alpha_{rT}}\}. \tag{5.1}$$

THEOREM 5.2. (A consistent estimator for $rk(\mathbf{B}) = r^*$.) *If* $r^* < q$ *and Assumptions* 2.1–2.4, 3.1, *and* 4.1 *hold and if* (*a*) $\alpha_{rT} = o(1)$ *and* (*b*) $-T^{-1}\ln\alpha_{rT} = o(1)$, *then the estimator* $\hat{r}$ *defined in* (5.1) *is weakly consistent for* $rk(\mathbf{B}) = r^*$; *that is*, $\hat{r} - r^* = o_P(1)$.

Similar estimators for the true rank r^* of the matrix $\mathbf{B}$ can be defined in a likewise fashion to $\hat{r}$ of (5.1) using the CR statistics $\mathcal{CRT}_r^{LR}$ of (3.6) and $\mathcal{CRT}_r^{W}$ of (3.7), $r = 0,\dots,q-1$.[5]

6. A SIMULATION EXPERIMENT

6.1. Experimental Design

The Monte-Carlo experiments reported subsequently use data drawn from the UK Family Expenditure Survey (FES) for the period 1974–1992, which are the data used in the applications reported in Blundell and Robin (1997a, 1997b). We have chosen these data as a basis for our simulations to provide a degree of realism for our study. Thus, the results reported subsequently should bear some relation to the type of situation that applied workers are likely to encounter in practice. The sample selected is reasonably homogeneous, with 4,981 households each of which consists of a married couple with two children. To avoid the potential problem of zero expenditures in the tobacco and gasoline categories the sample includes only car owning households in which at least one adult smokes. The data comprise purchases of 14 nondurable and service goods: alcohol, food consumed at home, food consumed outside the home, energy, clothing, household services, personal goods and services, leisure goods, entertainment, other leisure services, fares, tobacco, motoring, and gasoline.

Our study examines a linearized version of Deaton and Muellbauer's (1980) almost ideal demand system, where each budget share is regressed on a constant, a set of three seasonal indicators, the logarithm of relative prices, and the logarithm of real total expenditure; the deflator used is the Stone price index. Because all shares sum to one, one equation is redundant, and consequently the final equation is eliminated. Thus, relative prices are computed as the ratio of the price index for the commodity group and that of the excluded commodity, which in this study is gasoline.

The system of demand equations may be written as $\mathbf{w}_t = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{p}_t + \varepsilon_t$, where $\mathbf{w}_t$ is the vector consisting of the 13 linearly independent budget shares of household t, $t = 1, \ldots, 4{,}981$; $\mathbf{z}_t$ comprises the constant term, the three seasonal indicators, and logged real total expenditure; $\mathbf{p}_t$ is the vector of relative prices; $\{\varepsilon_t\}$ are uncorrelated error terms with zero mean and constant positive definite variance matrix and $\mathbf{A}$ and $\mathbf{B}$ conformable matrices of unknown parameters. Our study concerns the rank of the matrix of relative price effects $\mathbf{B}$; hence, $p = q = 13$.

Economic theory indicates that the matrix of relative price effects $\mathbf{B}$ is symmetric. Therefore, the symmetry restriction on $\mathbf{B}$ is imposed in the second stage of estimation of the parameter matrices $\mathbf{A}$ and $\mathbf{B}$ by minimum chi-squared. The rank of $\mathbf{B}$ is assessed using this estimate. Table 1 summarizes the results of four rank tests.

The first column $\mathcal{CRT}$ corresponds to the Wald form of the CR statistic, $\mathcal{CRT}_r{}^W$ of (3.7), with the weighting matrices $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Sigma}}$ both set equal to (p, p) identity matrices. The second column $\mathcal{WCRT}$ is a weighted version of the Wald form of the CR statistic when the $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Sigma}}$ matrices are chosen by analogy with the within and between variance estimators in linear panel data models (Hsiao, 1986), namely, $\hat{\boldsymbol{\Psi}}^{-1} = (\mathbf{I}_p \otimes \boldsymbol{\iota}_p')\hat{\boldsymbol{\Omega}}(\mathbf{I}_p \otimes \boldsymbol{\iota}_p)/(\boldsymbol{\iota}_p'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\iota}_p)$ and

Table 1. *P*-values

	Statistics			
r	$\mathcal{CRT}$	$\mathcal{WCRT}$	$\mathcal{ALS}$	$\mathcal{W}$
0	0.046	0.000	0.000	0.000
1	0.300	0.002	0.000	0.000
2	0.240	0.008	0.000	0.000
3	0.298	0.074	0.000	0.000
4	0.410	0.142	0.000	0.000
5	0.638	0.130	0.052	0.000
6	0.788	0.174	0.499	0.000
7	0.946	0.432	0.618	0.000
8	0.946	0.772	0.780	0.000
9	0.896	0.866	0.687	0.001
10	0.798	0.950	0.971	0.228
11	0.470	0.904	0.920	0.139
12	0.162	0.650	0.598	0.229

$\hat{\mathbf{\Sigma}}^{-1} = (\boldsymbol{\iota}_p' \otimes \mathbf{I}_p)\hat{\mathbf{\Omega}}(\boldsymbol{\iota}_p \otimes \mathbf{I}_p)/[(\boldsymbol{\iota}_p' \otimes \boldsymbol{\iota}_p')\hat{\mathbf{\Omega}}(\boldsymbol{\iota}_p \otimes \boldsymbol{\iota}_p)/(\boldsymbol{\iota}_p'\hat{\mathbf{\Sigma}}^{-1}\boldsymbol{\iota}_p)]$, where $\boldsymbol{\iota}_p$ is a p-vector of units.[6] If $\mathbf{\Omega} = \mathbf{\Psi}^{-1} \otimes \mathbf{\Sigma}^{-1}$, which in our set of simulation experiments is the structure of the asymptotic variance of the estimator for $\mathbf{B}$ prior to imposing symmetry, this choice of $\mathbf{\Psi}$ and $\mathbf{\Sigma}$ results in a weighted CR statistic identical to that of Anderson (1951) (cf. Corollary 3.1).

The last two columns of Table 1 relate to tests proposed by Cragg and Donald (1996, 1997) and Gouriéroux et al. (1985). Under $H_r: rk(\mathbf{B}) = r$, to apply tests based on the MC and ALS principles, it is necessary first to isolate asymptotically an (r, r) submatrix of $\mathbf{B}$ that is nonsingular. Under certain conditions, Cragg and Donald (1996, Secs. 3 and 5; see also Gill and Lewbel, 1992) demonstrated that a suitable estimator for such a submatrix may be achieved by performing r row and column permutations via an LU decomposition with complete pivoting (CP) on the RTC estimator $\hat{\mathbf{B}}$ without affecting the limiting distribution of the MC statistic for testing the hypothesis $H_r: rk(\mathbf{B}) = r$ against $H_r': rk(\mathbf{B}) > r$, even in the presence of ties (in particular, see Cragg and Donald, 1996, Corollary 1, p. 1305). Hence, we partition the (p, p) symmetric matrix $\mathbf{B} = [(\mathbf{B}_{.j}^r)] = [(\mathbf{B}_{ij}^r)]$ into its first r and last $p - r$ rows and columns with the (p, r) and (r, r) submatrices $\mathbf{B}_{.1}^r$ and $\mathbf{B}_{11}^r$ assumed full column rank and nonsingular, respectively, under $H_r: rk(\mathbf{B}) = r$. Consequently, the $(p - r)(p - r)$ dimensional rank hypothesis $H_r: rk(\mathbf{B}) = r$ may be equivalently stated in freedom equation form as $\mathbf{B}_{.2}^r = \mathbf{B}_{.1}^r\mathbf{B}_r$, where $\mathbf{B}_r = (\mathbf{B}_{11}^r)^{-1}\mathbf{B}_{12}^r$, and in constraint equation form as $\mathbf{B}_{22}^r = \mathbf{B}_{21}^r\mathbf{B}_{11}^{r-1}\mathbf{B}_{12}^r$. The third column $\mathcal{ALS}$ is the ALS (Gouriéroux et al., 1985) statistic based on the freedom equation form $\mathbf{B}_{.2}^r = \mathbf{B}_{.1}^r\mathbf{B}_r$.[7] The final column $\mathcal{W}$ corresponds to the Wald test (Cragg and Donald, 1996) of the constraint equation form $\mathbf{B}_{22}^r = \mathbf{B}_{21}^r \mathbf{B}_{11}^{r-1} \mathbf{B}_{12}^r$.[8] Both statistics are obtained subject to the a priori symmetry constraint on $\mathbf{B}$. Under $H_r : rk(\mathbf{B}) = r$, both statistics have a limiting $\chi^2[(p-r)(p-r+1)/2]$ distribution, $r = 0,1,\dots,q-1$, although the distributions of these statistics may differ in finite samples.

Examining the p-values reported in Table 1, testing sequentially the hypotheses $H_r : rk(\mathbf{B}) = r$, $r = 0,1,\dots,q-1$, it is immediately clear that the tests select very different values for the rank of $\mathbf{B}$; at the 0.05 nominal level, 1, 3, 5, and 10, respectively. To appreciate this disparity in outcome, although the sample size seems large, households surveyed during the same month appear to face the same prices because the FES does not record individual prices. Consequently, we have imputed to each household the National Account Retail Price Indices of the month in which it is surveyed. Hence, there are effectively only 228 ($=19 \times 12$) different prices for each commodity. In the following simulations, we have chosen $\mathbf{B}$ to be the estimated matrix of relative price effects corresponding to a rank equal to 6, that is, the true rank $r^* = 6$. To provide differing relative price variables at each data point in the simulations, the original prices are perturbed by the addition of a sample drawn from a normal distribution with zero mean and variance given by that of the corresponding price in the original sample that yields an observed sample of exogenous variables $\mathbf{z}_t, \mathbf{p}_t$, $t = 1,\dots,4{,}981$. The variance matrix $\mathbf{V}$ for the simulated error terms is calculated as $1/4{,}981 \sum_{t=1}^{4{,}981} (\mathbf{w}_t - \mathbf{A}\mathbf{z}_t - \mathbf{B}\mathbf{p}_t)(\mathbf{w}_t - \mathbf{A}\mathbf{z}_t - \mathbf{B}\mathbf{p}_t)'$ divided by 9 where $\mathbf{A}$ is the corresponding sample estimate. This choice for the variance matrix of the simulated errors reduces the signal-noise ratio in the full sample to such a level that each of the preceding rank tests gives identical inferences.[9]

6.2. Monte-Carlo Results

We consider sample sizes $T = 250$, 900, and 2,000. Each experiment comprises 2,500 replications. Within each replication $s = 1,\dots,2{,}500$, a random sample $\{\varepsilon_t^s\}_{t=1}^T$ is drawn from a $N_{13}(\mathbf{0}, \mathbf{V})$ population and the simulated budget share vectors computed from $\mathbf{w}_t^s = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{p}_t + \varepsilon_t^s$. The parameter matrices $\mathbf{A}$ and $\mathbf{B}$ are initially estimated unrestrictedly equation by equation by ordinary least squares, and then symmetry is imposed on the estimator for $\mathbf{B}$. Finally, the rank of $\mathbf{B}$ is evaluated using each of the preceding four procedures.

Figures 1–6 present PP plots for sample sizes $T = 250$, 900, and 2,000. Given a probability p and its associated nominal critical value $c_r(p)$ obtained from the limiting distribution of the statistic under $H_r : rk(\mathbf{B}) = r$, the simulations estimate the exact probability $\pi_r(p)$ that the statistic exceeds $c_r(p)$. Each panel in the figures provided the PP plot of p on the vertical axis against $\pi_r(p)$ on the horizontal axis, $r = 0,\dots,12$.

First, consider the results presented in Figures 1 and 2 for $T = 250$. The panel corresponding to the null hypothesis $H_6 : rk(\mathbf{B}) = 6$, the true rank of $\mathbf{B}$, in Figure 2 indicates that, for the upper tail of its distribution, the unweighted CR statistic ($\mathcal{CRT}$) has empirical size substantially below the nominal size predicted by the

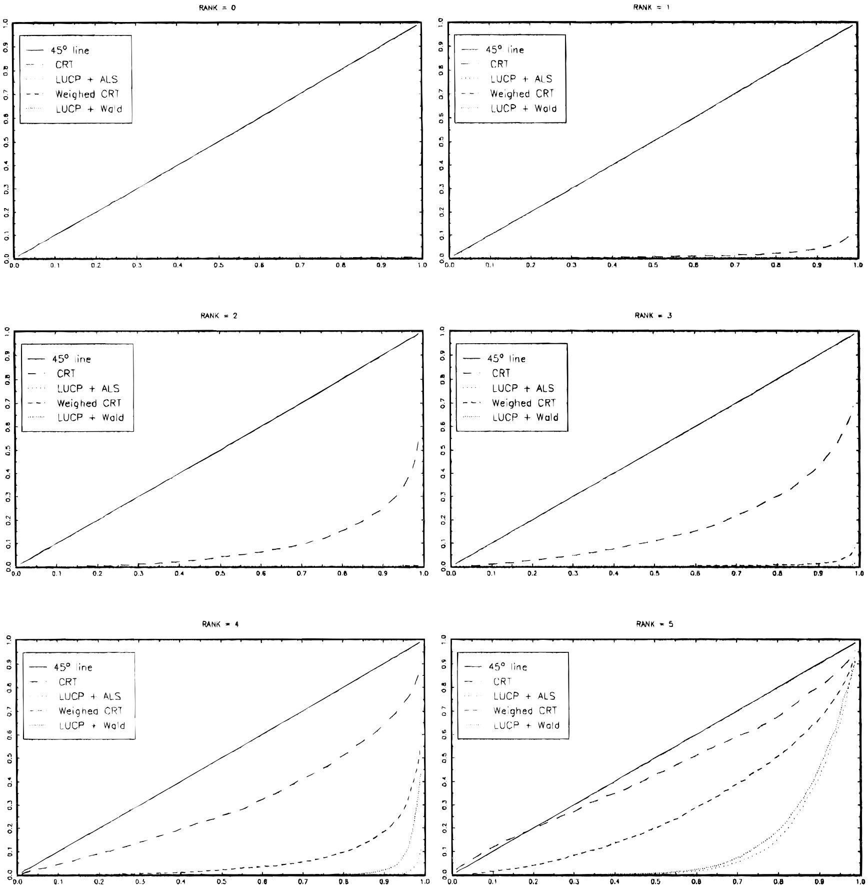

FIGURE 1. PP plots: $T = 250$.

asymptotic theory of Section 3. However, the deviation between empirical and nominal sizes for the weighted CR statistic $(\mathcal{WCRT})$ is less pronounced. Consequently, both CR statistics are likely to accept too low a rank for **B**. For the ALS, $\mathcal{ALS}$ and the Wald, $\mathcal{W}$ statistics, the reverse situation occurs, with empirical size much in excess of nominal size obtained by use of the limiting $\chi^2(28)$ distribution. Hence, the ALS and Wald procedures with LUCP are likely to induce acceptance of too high a rank for **B**. Examination of the power properties of these tests displayed in the panels of Figure 1 for the null hypotheses $H_r : rk(\mathbf{B}) = r$, $r = 0, \ldots, 5$ reveals that both the CR statistics have poor power characteristics, with the unweighted CR statistics $\mathcal{CRT}$ particularly bad in this regard, whereas both $\mathcal{ALS}$ and $\mathcal{W}$ appear to perform well, confirming the preceding observation. The relatively poor performance of $\mathcal{CRT}$ vis-à-vis $\mathcal{WCRT}$ may reflect the population values of the nonzero CRs reported in note 9. However, given the divergence between empirical and nominal sizes for both statistics discussed earlier,

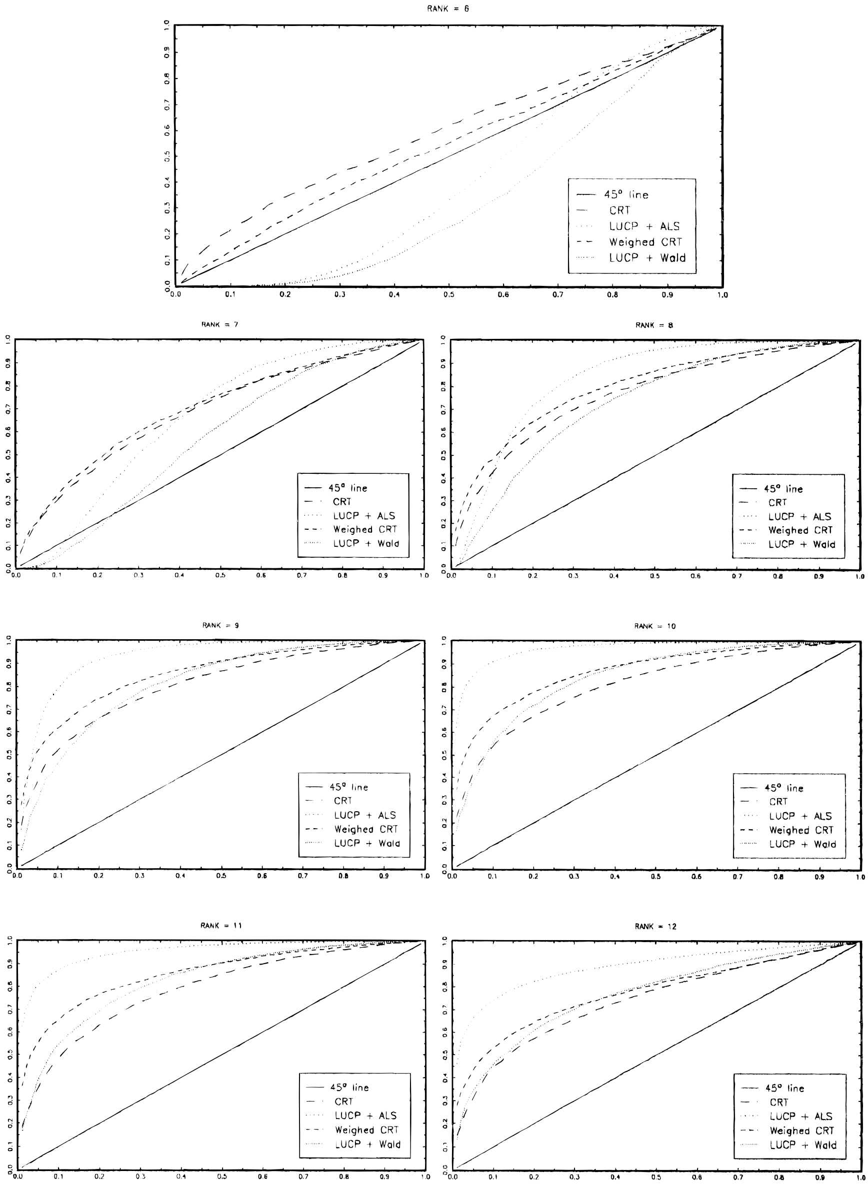

FIGURE 2. PP plots: $T = 250$ (plot continued from Figure 1).

one would expect that if the results in these panels had been based on size-corrected critical values the CR statistics would perform somewhat better, whereas the ALS and Wald statistics would perform somewhat worse than indicated in Figure 1. Given the relatively small price variation present in the actual sample

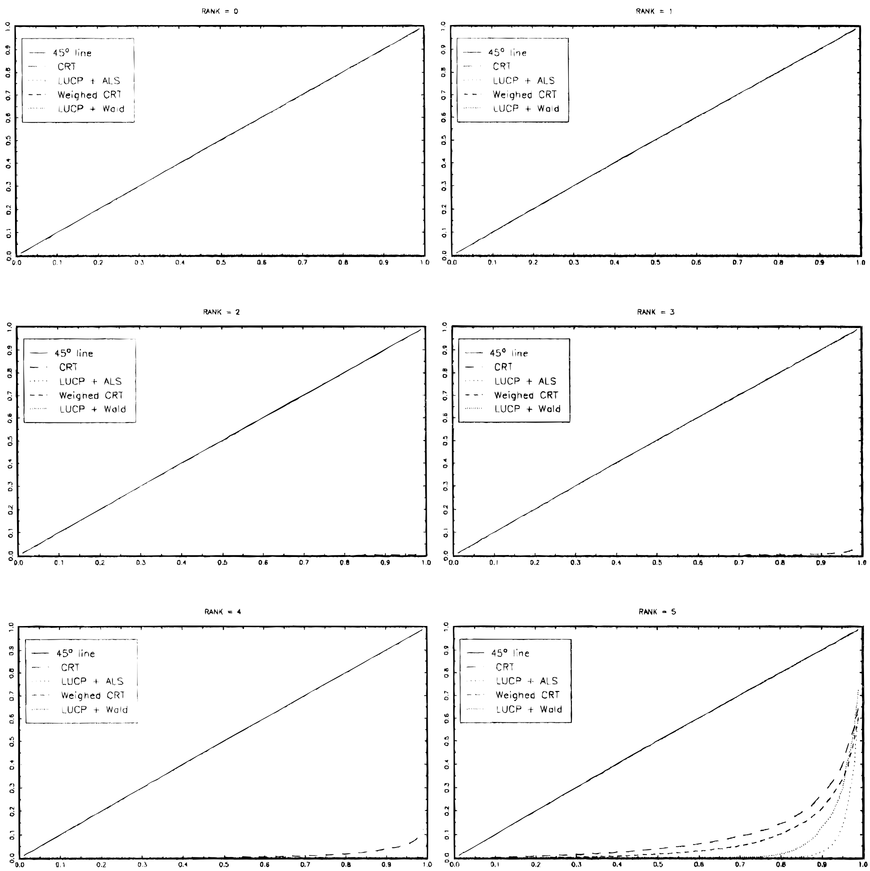

FIGURE 3. PP plots: $T = 900$.

discussed in Section 6.1, the disparity between the statistics exhibited in Figures 1 and 2 underlines the differences in inferences from the tests displayed in Table 1. The remaining panels in Figure 2 for the null hypotheses $H_r : rk(\mathbf{B}) = r$, $r = 7,\dots,12$ reflect the nonstandard distributions of the various rank statistics when $r > r^*$ (see also Cragg and Donald, 1996, Sect. 3, pp. 1303–1304).

Second, Figures 3 and 4 present results for $T = 900$. The situation for both of the CR statistics is now much improved, as is revealed by the $r = 6$ panel of Figure 4, in which the empirical and nominal sizes differ by very little. Moreover, the power characteristics for both statistics given in the $r = 0,\dots,5$ panels of Figure 3 are correspondingly much improved, and, consequently, nonrejection of too low a rank for $\mathbf{B}$ appears to be much less of a problem. Examination of the $r = 6$ panel for the ALS and Wald statistics shows some but, surprisingly, relatively little improvement. The power of these statistics based on the nominal

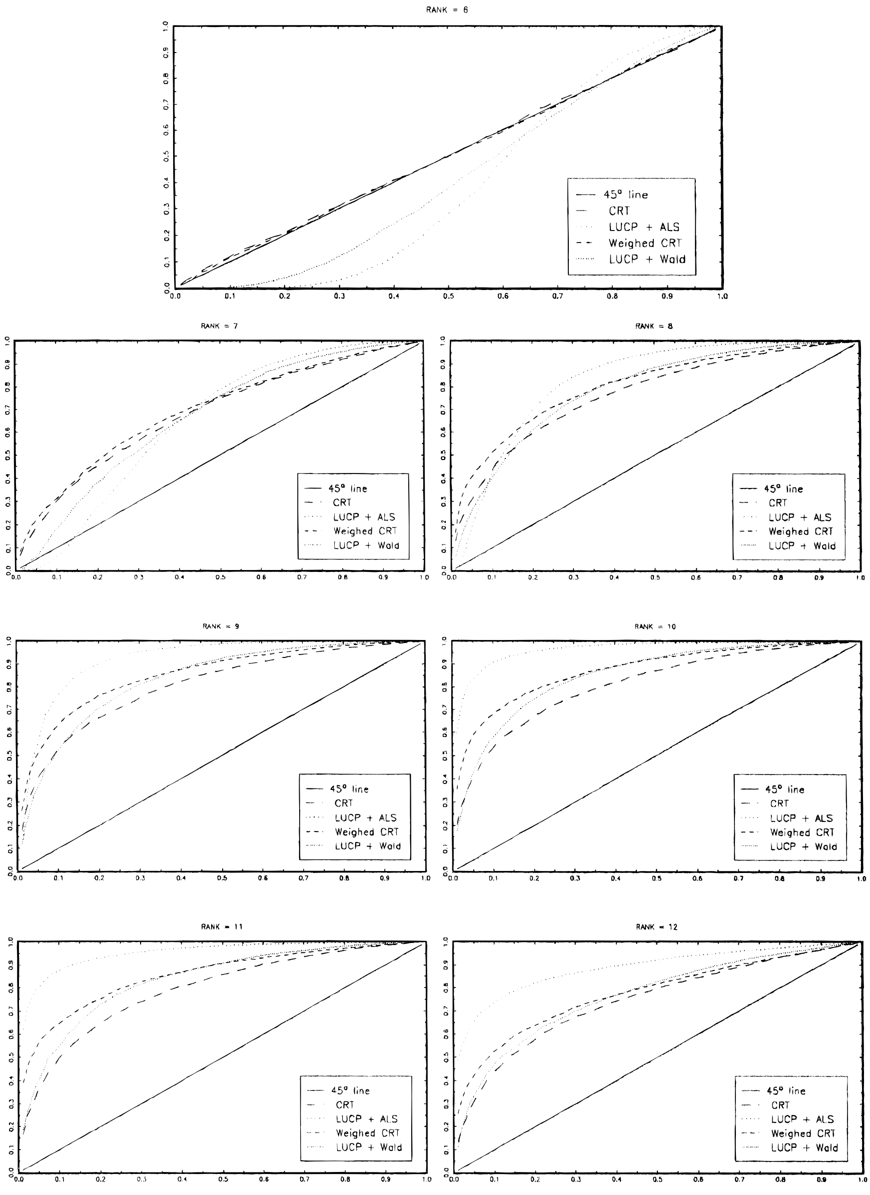

FIGURE 4. PP plots: $T = 900$ (plot continued from Figure 3).

critical values given in the $r = 0,\ldots,5$ tableaus has also improved vis-à-vis Figure 1 but not nearly as dramatically as compared to the CR statistics.

Third, the results reported in Figures 5 and 6 for the $T = 2{,}000$ case indicate that the empirical size properties given in the $r = 6$ panel of Figure 6 for both CR

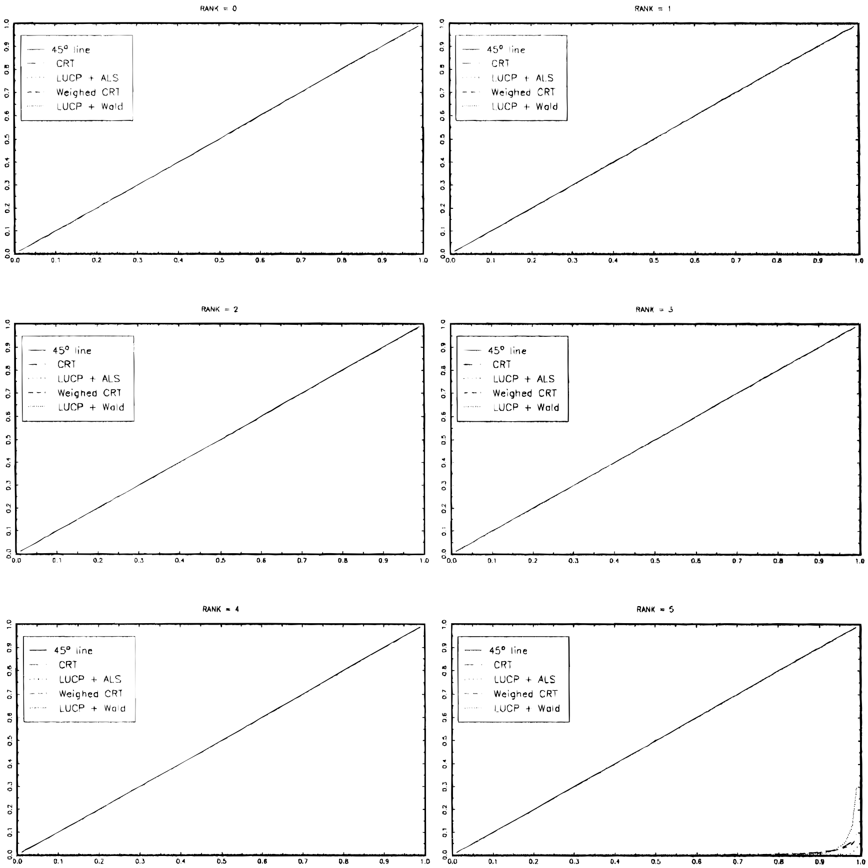

FIGURE 5. PP plots: $T = 2{,}000$.

statistics are approximated well by those indicated by the asymptotic theory of Section 3. More surprisingly, those for the ALS and Wald statistics again show relatively little improvement. The panels of Figure 5 corresponding to $r = 0, \ldots, 5$ demonstrate that all four statistics have good power characteristics.

7. SUMMARY AND CONCLUSIONS

This paper considers tests for the rank of a matrix for which a RTC estimator is available but where the rank of the estimator's asymptotic variance matrix may be neither full nor known. Test statistics based on certain estimated characteristic roots are proposed. Under the null hypothesis of a given rank, their limiting distribution is shown to be a weighted sum of independent chi-squared variables, each of which has one degree of freedom. The limiting null distribution of the

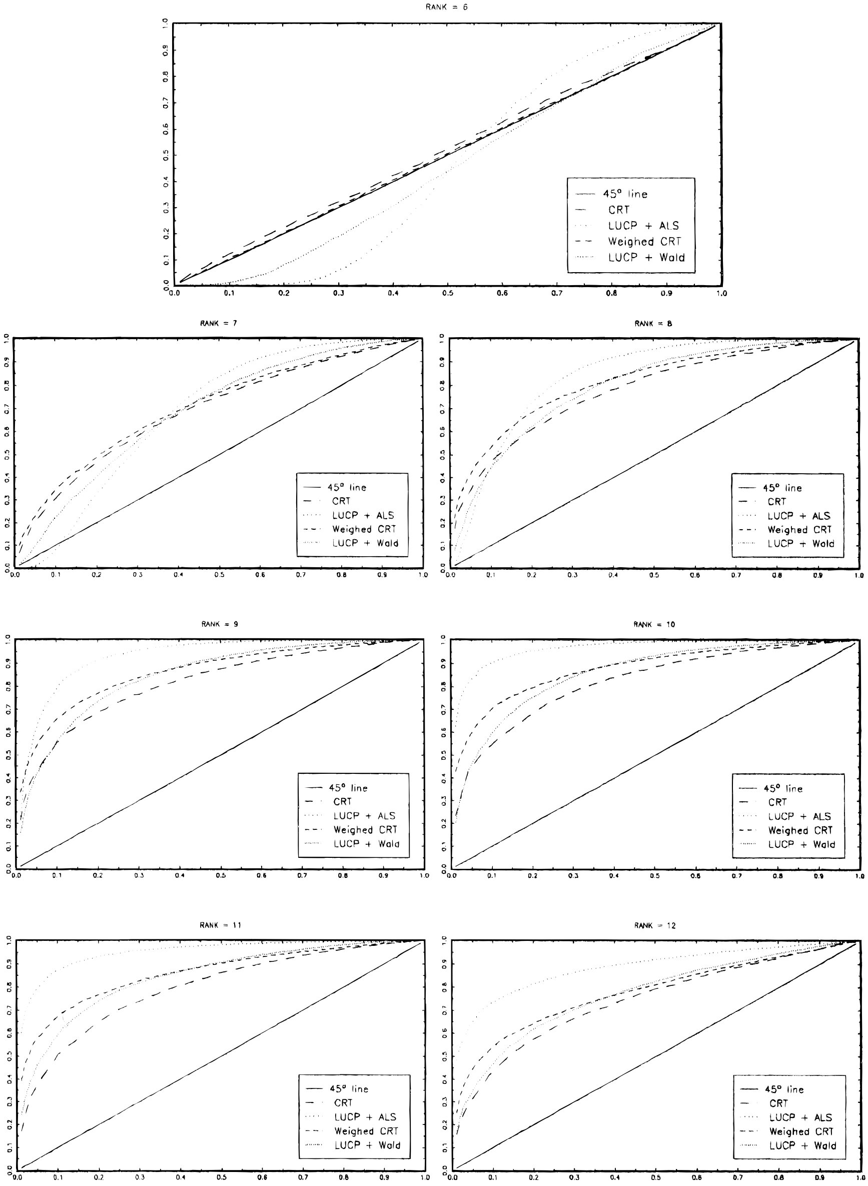

FIGURE 6. PP plots: $T = 2{,}000$ (plot continued from Figure 5).

characteristic root tests may be estimated either by simulation or by use of widely available algorithms. In an asymptotic sense, these test procedures will never accept a value for the rank of the matrix less than the true rank. Consistent estimation procedures for the rank of the matrix are proposed.

A simulation experiment conducted on a system of budget share equations indicates that for moderate sample sizes the empirical size of characteristic root tests is less than that indicated by asymptotic theory, whereas statistics based on asymptotic least squares and the Wald principle appear to be oversized. Hence, for such sample sizes, the former statistics are likely to accept a value of the rank of the matrix lower than the true rank, whereas the latter statistics are likely to accept too high a rank. It is only when the sample size is large that empirical and nominal sizes are similar for characteristic root statistics. Both forms of statistic have good power properties for such sample sizes. Consequently, in such circumstances, characteristic root statistics may possibly offer useful tests for the rank of a matrix. However, it would appear that for smaller sample sizes it may prove efficacious to examine more than one form of test for the rank of a matrix.

NOTES

1. Model selection methods offer an alternative approach to the consistent estimation of the rank of a matrix (see, e.g., Phillips, 1996). Chao and Phillips (1999) applied such methods to the joint determination of cointegration rank and lag length within a vector autoregressive system.

2. For example, let $\hat{\mathbf{B}} = \mathbf{B}(\hat{\phi})$, where $\mathbf{B}$ obeys the a priori constraints $\mathbf{B} = \mathbf{B}(\phi)$ continuously differentiable to the first order, ϕ is an s-vector of parameters, and $T^{1/2}(\hat{\phi} - \phi) \to^{L} N(0, \mathbf{V})$, $\mathbf{V}$ finite and positive definite. Hence, $\mathbf{\Omega} = \mathbf{\Phi}\mathbf{V}\mathbf{\Phi}'$, where $\mathbf{\Phi} = \nabla'_{\phi}\,\mathrm{vec}(\mathbf{B}(\phi))$ is assumed full column rank (cf. Assumption 2.2). Therefore, Assumption 2.4 is equivalent to $(\mathbf{D}_{q-r^*} \otimes \mathbf{C}_{p-r^*}) \notin \mathcal{N}(\mathbf{\Phi}')$.

3. In the simulations of Section 6, $p = q = 13$ and $s = 91$. Assumption 2.4 is therefore automatically satisfied if $r^* \in \{0,1,2,3,4\}$.

4. The matrix quadratic form $\hat{\mathbf{\Sigma}}\hat{\mathbf{B}}\hat{\mathbf{\Psi}}\hat{\mathbf{B}}'$ has the same characteristic roots as

$$\mathbf{F}^{-1}\hat{\mathbf{\Sigma}}\hat{\mathbf{B}}\hat{\mathbf{\Psi}}\hat{\mathbf{B}}'\mathbf{F} = (\mathbf{F}^{-1}\hat{\mathbf{\Sigma}}\mathbf{F}'^{-1})(\mathbf{F}'\hat{\mathbf{B}}\mathbf{G})(\mathbf{G}^{-1}\hat{\mathbf{\Psi}}\mathbf{G}'^{-1})(\mathbf{F}'\hat{\mathbf{B}}\mathbf{G})',$$

where $\mathbf{F}$ and $\mathbf{G}$ are arbitrary conformable nonsingular matrices. Therefore, the CR statistic (3.4) is invariant under the transformation $\hat{\mathbf{B}} \to \tilde{\mathbf{B}} = \mathbf{F}'\hat{\mathbf{B}}\mathbf{G}$ for given $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Psi}}$ if and only if $\mathbf{F} = \hat{\mathbf{C}}\mathbf{H}\hat{\mathbf{C}}^{-1}$ and $\mathbf{G} = \hat{\mathbf{D}}\mathbf{K}\hat{\mathbf{D}}^{-1}$, where $\hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$ are, respectively, the (p,p) and (q,q) matrices of estimated CV's (cf. following (2.3) and (2.4)) and $\mathbf{H}$ and $\mathbf{K}$ are arbitrary conformable orthonormal matrices, particular examples of which are permutation matrices. In the case when $\hat{\mathbf{\Sigma}} = \mathbf{I}_p$ and $\hat{\mathbf{\Psi}} = \mathbf{I}_q$, the CR statistic (3.4) is invariant to such transformations if and only if $\mathbf{F}$ and $\mathbf{G}$ themselves are constrained to be arbitrary conformable orthonormal matrices.

5. Alternative conservative procedures for testing $H_r : rk\{\mathbf{B}\} = r$ against $H'_r : rk\{\mathbf{B}\} > r$ that avoid computing the CR estimators $\{\hat{\lambda}_i^r\}_{i=1}^{(p-r)(q-r)}$ described in Theorem 4.1 for the CR's $\{\lambda_i^r\}_{i=1}^{(p-r)(q-r)}$ of the matrix $(\mathbf{D}_{q-r} \otimes \mathbf{C}_{p-r})'\mathbf{\Omega}(\mathbf{D}_{q-r} \otimes \mathbf{C}_{p-r})$ may be constructed by noting $\mathbf{\Omega}^{1/2\prime}(\mathbf{\Psi} \otimes \mathbf{\Sigma})\mathbf{\Omega}^{1/2} - \mathbf{\Omega}^{1/2\prime}(\mathbf{D}_{q-r} \otimes \mathbf{C}_{p-r})(\mathbf{D}_{q-r} \otimes \mathbf{C}_{p-r})'\mathbf{\Omega}^{1/2} \geq \mathbf{0}$, where $\mathbf{\Omega} = \mathbf{\Omega}^{1/2}\mathbf{\Omega}^{1/2\prime}$. Hence, $\lambda^*_{\max} \geq \lambda^*_i \geq \lambda^r_i$, $i = 1,\ldots,(p-r)(q-r)$, where $\lambda^*_{\max}$ and $\{\lambda^*_i\}_{i=1}^{s}$ are the maximum CR and nonzero ordered CR's of $(\mathbf{\Psi} \otimes \mathbf{\Sigma})\mathbf{\Omega}$, respectively. Therefore, $\lambda^*_{\max}\sum_{i=1}^{(p-r)(q-r)} Z_i^2 \geq \sum_{i=1}^{(p-r)(q-r)} \lambda^*_i Z_i^2 \geq \sum_{i=1}^{t} \lambda^r_i Z_i^2$, and, hence, $\mathcal{P}\{\lambda^*_{\max}\sum_{i=1}^{(p-r)(q-r)} Z_i^2 \leq c\} \leq \mathcal{P}\{\sum_{i=1}^{(p-r)(q-r)} \lambda^*_i Z_i^2 \leq c\} \leq F_r^{CRT}(c)$, where $\{Z_i\}_{i=1}^{(p-r)(q-r)}$ are independent standard normal variates. Consistent estimators for $\lambda^*_{\max}$ and λ^*_i, $\hat{\lambda}^*_{max}$ and $\hat{\lambda}^*_i$, $i = 1,\ldots,(p-r)(q-r)$ are obtained from the ordered CR's of $(\hat{\mathbf{\Psi}} \otimes \hat{\mathbf{\Sigma}})\hat{\mathbf{\Omega}}$ (cf. Lemma 3.1). Hence, conservative critical regions for testing the hypothesis $H_r : rk(\mathbf{B}) = r$ against $H'_r : rk(\mathbf{B}) > r$ together with associated sequential procedures for the consistent estimation of r^* may be constructed along similar lines to those of Theorems 4.2 and 5.2, respectively. In particular, the critical region $\{\hat{\lambda}^{*-1}_{\max}\mathcal{CRT}_r < \chi^2_{1-\alpha}[(p-r)(q-r)]\}$ has asymptotic size no greater than α, $0 < \alpha < 1$, under $H_r : rk(\mathbf{B}) = r$, where $\chi^2_{1-\alpha}(\cdot)$ denotes the $100(1-\alpha)$ percentile of the $\chi^2(\cdot)$ distribution. However, it is possible that in practice such critical regions will be too narrow, leading to rejection of $H_r : rk(\mathbf{B}) = r$, $r < r^*$, too infrequently (see Section 6).

6. The scalar divisors in $\hat{\mathbf{\Psi}}^{-1}$ and $\hat{\mathbf{\Sigma}}^{-1}$ may be ignored, as the critical regions based on the CR statistics are invariant to scale transformations.

7. That is, $T\hat{\boldsymbol{\xi}}'(\hat{\mathbf{M}}\hat{\mathbf{\Omega}}\hat{\mathbf{M}}')^{-1}\hat{\boldsymbol{\xi}}$, the sum of squared weighted generalized least squares residuals $\hat{\boldsymbol{\xi}}$ computed via the auxiliary regression $\mathbf{P}_r\,\mathrm{vec}(\hat{\mathbf{B}}^{r\prime}_{.2}) = \mathbf{P}_r(\hat{\mathbf{B}}^{r}_{.1} \otimes \mathbf{I}_{p-r})\mathrm{vec}(\mathbf{B}_r') + \boldsymbol{\xi}$, where $\mathbf{P}_r = \mathrm{diag}(\mathbf{I}_{r(p-r)}, \mathbf{L}_r), \mathbf{L}_r$ is the $((p-r)(p-r+1)/2, (p-r)^2)$ elimination matrix (Magnus and Neudecker, 1980), $\boldsymbol{\xi}$ is treated as having mean zero and variance matrix $\mathbf{M}\hat{\mathbf{\Omega}}\mathbf{M}'$ with $\hat{\mathbf{\Omega}}$ the estimated variance matrix of $\mathbf{B}$ and $\mathbf{M} = \mathbf{P}_r(\mathbf{I}_p \otimes (-\mathbf{B}_r', \mathbf{I}_{p-r}))\mathbf{K}_{p^2}$ with $\mathbf{K}_{p^2}$ the (p^2, p^2) commutation matrix (Magnus and Neudecker, 1988, Sect. 3.7, pp. 46–48). In our experiments, an initial estimator for $\mathbf{B}_r$ is obtained using ordinary least squares and the preceding equation then iterated three times.

8. Writing the rank restrictions as $\mathbf{h}_r[\mathrm{vec}(\mathbf{B})] = \mathrm{vec}(\mathbf{B}^r_{22} - \mathbf{B}^r_{21}\mathbf{B}^{r-1}_{11}\mathbf{B}^r_{12}) = \mathbf{0}$, the derivative matrix of $\mathbf{h}_r(\cdot)$ is given by $\mathbf{H}_r'(\cdot) \equiv \nabla'_{\mathrm{vec}(\mathbf{B})}\mathbf{h}_r(\cdot) = (-\mathbf{B}^{r\prime}_{12}(\mathbf{B}^r_{11})'^{-1}, \mathbf{I}_{p-r}) \otimes (-\mathbf{B}^r_{21}\mathbf{B}^{r-1}_{11}, \mathbf{I}_{p-r})$, which is full row rank $(p-r)^2$. The Wald statistic for $H_r: rk(\mathbf{B}) = r$ is given by $T\mathbf{h}_r[\mathrm{vec}(\hat{\mathbf{B}})]'\mathbf{L}_r'(\mathbf{L}_r\hat{\mathbf{H}}_r'\hat{\mathbf{\Omega}}\hat{\mathbf{H}}_r\mathbf{L}_r')^{-1}\mathbf{L}_r\mathbf{h}_r[\mathrm{vec}(\hat{\mathbf{B}})]$, where $\mathbf{L}_r$ is the $((p-r)(p-r+1)/2, (p-r)^2)$ elimination matrix and $\hat{\mathbf{H}}_r = \mathbf{H}_r'(\mathrm{vec}(\hat{\mathbf{B}}))$.

9. Assumption 2.4 was satisfied for both forms of CR statistics. The ordered nonzero CR's (normalizing the maximum CR to be unity) of $\mathbf{\Sigma B \Psi B}'$ for the $\mathcal{CRT}$ and $\mathcal{WCRT}$ statistics were 1.000, 0.5139, 0.1928, 0.1376, 0.0759, 0.0644 and 1.000, 0.5267, 0.4013, 0.3174, 0.2307, 0.0787, respectively.

REFERENCES

Anderson, T.W. (1951) The asymptotic distribution of certain characteristic roots and vectors. In J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 103–130. Berkeley: University of California Press.

Andrews, D.W.K. (1987) Asymptotic results for generalized Wald tests. *Econometric Theory* 3, 348–358.

Bauer, P., B.M. Pötscher, & P. Hackl (1988) Model selection by multiple tests. *Statistics* 19, 39–44.

Blundell, R.W. & J.-M. Robin (1997a) Estimation in Large and Disaggregated Systems: An Estimator for Conditionally Linear Systems. Discussion paper 97-2, University College, London.

Blundell, R.W. & J.-M. Robin (1997b) Latent Separability: Grouping Goods without Weak Separability. Mimeo, University College, London.

Chao, J.C. & P.C.B. Phillips (1999) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91, 227–271.

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

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

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

Davies, R.B. (1980) The distribution of a linear combination of χ^2 random variables. *Applied Statistics*, *Series C* 29, 323–333.

Deaton, A.S. & J. Muellbauer (1980) An almost ideal demand system. *American Economic Review* 70, 312–336.

Farebrother, R.W.F. (1980) Pan's procedure for the tail probabilities of the Durbin–Watson statistic. *Applied Statistics*, *Series C* 29, 224–227; 30, 189.

Farebrother, R.W.F. (1984) Remark on Pan's procedure for the tail probabilities of the Durbin–Watson statistic. *Applied Statistics*, *Series C* 33, 363–366.

Ferguson, T. (1958) A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics* 29, 1046–1061.

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

Gouriéroux, C., A. Monfort, & A. Trognon (1985) Moindres carres asymptotiques, *Annales de l'INSEE* 58, 91–122.

Hsiao, C. (1983) Identification. In Z. Griliches and M.D. Intrilligator (eds.), *Handbook of Econometrics*, vol. 1, pp. 223–283. Amsterdam: North-Holland.

Hsiao, C. (1986) *Analysis of Panel Data.* Econometric Society Monograph 11. New York: Cambridge University Press.

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

Lewbel, A. (1991) The rank of demand systems: Theory and nonparametric estimation. *Econometrica* 59, 711–730.

Magnus, J.R. & H. Neudecker (1980) The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic and Discrete Methods* 1, 422–449.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* New York: Wiley.

Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 763–812.

Pötscher, B.M. (1983) Order estimation in ARMA models by Lagrange multiplier tests. *Annals of Statistics* 11, 872–885.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley.

Robin, J.-M. & R.J. Smith (1995) Tests of Rank. Document de travail 9505, CREST-INSEE, Paris, and D.A.E. working paper 9521, University of Cambridge.

Rothenberg, T.J. (1973) *Efficient Estimation with a Priori Information.* New Haven: Yale University Press.

Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.

APPENDIX

Proof of Lemma 3.1. Under Assumptions 2.2 and 2.3, from (2.2), we have that $\hat{\mathbf{B}} \to^P \mathbf{B}$, $\hat{\mathbf{\Sigma}} \to^P \mathbf{\Sigma}$, and$\hat{\mathbf{\Psi}} \to^P \mathbf{\Psi}$. Therefore, from (3.1), $\hat{\mathbf{\Sigma}}\hat{\mathbf{B}}\hat{\mathbf{\Psi}}\hat{\mathbf{B}}' \to^P \mathbf{\Sigma B \Psi B}'$. As the CR's $\{\tau_i^2\}_{i=1}^p$ defined in (2.3) and (2.4) are continuous functions of the elements of $\mathbf{\Sigma B \Psi B}'$, we therefore have that $\hat{\lambda}_i \to^P \tau_i^2$, $i = 1, \ldots, p$. ■

Define the (r^*, r^*) diagonal matrix $\mathbf{\Upsilon}_{r^*} = \text{diag}(\tau_1, \ldots, \tau_{r^*})$, $\mathbf{C}'^{-1} = (\mathbf{C}^{r^*}, \mathbf{C}^{p-r^*})$ and $\mathbf{D}^{-1} = (\mathbf{D}^{r^*}, \mathbf{D}^{q-r^*})'$, where $\mathbf{C}$ and $\mathbf{D}$ are defined via (2.3) and (2.4), respectively, and $\mathbf{C}'^{-1}$ and $\mathbf{D}^{-1}$ are partitioned conformably with respect to the r^* nonzero and $(p - r^*)$ and $(q - r^*)$ zero characteristic roots of (2.3) and (2.4), respectively.

LEMMA A.1. *A (p, q) matrix $\mathbf{B}$ of rank r^* may be expressed as*

$$\mathbf{B} = \mathbf{C}^{r^*}\mathbf{\Upsilon}_{r^*}\mathbf{D}^{r^*\prime}. \tag{A.1}$$

Proof. The proof is similar to that of Rao (1973, 1c.3(v), pp. 42–43). We may write $\mathbf{\Sigma} = \sum_{i=1}^p \mathbf{c}_i \mathbf{c}_i'$. Hence

$$\mathbf{B} = \mathbf{\Sigma}^{-1}\left(\sum_{i=1}^{p} \mathbf{c}_i \mathbf{c}_i'\right)\mathbf{B} = \mathbf{\Sigma}^{-1}\left(\sum_{i=1}^{r^*} \mathbf{c}_i \mathbf{c}_i'\right)\mathbf{B}$$

$$= \sum_{i=1}^{r^*} \tau_i \mathbf{\Sigma}^{-1}\mathbf{c}_i \mathbf{d}_i' \mathbf{\Psi}^{-1} = \mathbf{\Sigma}^{-1}\mathbf{C}_{r^*}\mathbf{\Upsilon}_{r^*}\mathbf{D}_{r^*}'\mathbf{\Psi}^{-1}$$

$$= \mathbf{C}^{r^*}\mathbf{\Upsilon}_{r^*}\mathbf{D}^{r^*\prime}.$$

The second equality follows from $\mathbf{c}_i'\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'\mathbf{c}_i = 0$, $i = r^* + 1, \dots, p$, which implies $\mathbf{B}'\mathbf{c}_i = 0$, $i = r^* + 1, \dots, p$, as $\boldsymbol{\Psi}$ is positive definite; cf. (2.3). The third equality follows as $\mathbf{d}_i = \tau_i^{-1}\boldsymbol{\Psi}\mathbf{B}'\mathbf{c}_i$, $i = 1, \dots, r^*$ are CV's of $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}$ in the metric $\boldsymbol{\Psi}^{-1}$ (cf. (2.4)). The final equality follows as $\boldsymbol{\Sigma}^{-1} = \mathbf{C}'^{-1}\mathbf{C}^{-1}$ and $\boldsymbol{\Psi}^{-1} = \mathbf{D}'^{-1}\mathbf{D}^{-1}$. ∎

Remark. Note that $\mathbf{C}_{r^*}'\mathbf{C}^{r^*} = \mathbf{I}_{r^*}$ and $\mathbf{C}_{p-r^*}'\mathbf{C}^{r^*} = \mathbf{0}$.

Proof of Theorem 3.1. The proof follows along similar lines to that of Johansen (1991, proof of Theorem 2.1, pp. 1569–1571). We are concerned with the solutions to the determinantal equation $|\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}' - \lambda\hat{\boldsymbol{\Sigma}}^{-1}| = 0$ corresponding to the zero CR's $\{\tau_i^2\}_{i=r^*+1}^q$ of $\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Psi}\mathbf{B}'$. First, define

$$\hat{\mathbf{S}}(\lambda) \equiv \hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}' - \lambda\hat{\boldsymbol{\Sigma}}^{-1}. \tag{A.2}$$

Second, writing $\hat{\mathbf{B}} = \mathbf{B} + (\hat{\mathbf{B}} - \mathbf{B})$, we have from Lemma A.1 (cf. (A.1) and the preceding remark) that, under Assumptions 2.1 and 2.2,

$$\mathbf{C}_{r^*}'\hat{\mathbf{B}} = \boldsymbol{\Upsilon}_{r^*}\mathbf{D}^{r^*\prime} + O_P(T^{-1/2}), T^{1/2}\mathbf{C}_{p-r^*}'\hat{\mathbf{B}} = \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B}). \tag{A.3}$$

Next, from Lemma 3.1, $\hat{\lambda}_i \to^P 0$, $i = r^* + 1, \dots, q$. Hence, from (A.2) and (A.3), under Assumption 2.3, consider

$$\begin{aligned}
&(\mathbf{C}_{r^*}, T^{1/2}\mathbf{C}_{p-r^*})'\hat{\mathbf{S}}(\hat{\lambda}_i)(\mathbf{C}_{r^*}, T^{1/2}\mathbf{C}_{p-r^*})\\
&\quad= \begin{bmatrix} \boldsymbol{\Upsilon}_{r^*}^2 + o_P(1) & \boldsymbol{\Upsilon}_{r^*}\mathbf{D}_{r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{C}_{p-r^*} + o_P(1) \\ \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{D}_{r^*}\boldsymbol{\Upsilon}_{r^*} + o_P(1) & \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\boldsymbol{\Psi}T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{C}_{p-r^*} + o_P(1) \end{bmatrix}\\
&\qquad -T\hat{\lambda}_i\begin{bmatrix} O_P(T^{-1}) & o_P(T^{-1/2}) \\ o_P(T^{-1/2}) & \mathbf{I}_{p-r^*} + o_P(1) \end{bmatrix},
\end{aligned} \tag{A.4}$$

$i = r^* + 1, \dots, q$. Therefore,

$$\begin{aligned}
0 &= |\hat{\mathbf{S}}(\hat{\lambda}_i)| = |(\mathbf{C}_{r^*}, T^{1/2}\mathbf{C}_{p-r^*})'\hat{\mathbf{S}}(\hat{\lambda}_i)(\mathbf{C}_{r^*}, T^{1/2}\mathbf{C}_{p-r^*})|\\
&= \left|\begin{bmatrix} \boldsymbol{\Upsilon}_{r^*}^2 & \boldsymbol{\Upsilon}_{r^*}\mathbf{D}_{r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{C}_{p-r^*} \\ \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{D}_{r^*}\boldsymbol{\Upsilon}_{r^*} & \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\boldsymbol{\Psi}T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{C}_{p-r^*} \end{bmatrix}\right.\\
&\qquad \left. - T\hat{\lambda}_i\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-r^*} \end{bmatrix}\right| + o_P(1)\\
&= |\boldsymbol{\Upsilon}_{r^*}^2||\mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{D}_{q-r^*}\mathbf{D}_{q-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{C}_{p-r^*} - T\hat{\lambda}_i\mathbf{I}_{p-r^*}| + o_P(1),
\end{aligned} \tag{A.5}$$

$i = r^* + 1, \dots, q$, where the third equality follows from (A.4) and the final equality from Rao (1973, complements and problems 2.4, p. 32), also noting that $\boldsymbol{\Psi} - \mathbf{D}_{r^*}\mathbf{D}_{r^*}' = \mathbf{D}_{q-r^*}\mathbf{D}_{q-r^*}'$. Hence, from Assumption 2.4 and (A.5), the limiting distribution of $T\hat{\lambda}_i$, $i = r^* + 1, \dots, q$ is the same as that of the $(q - r^*)$ nonzero CR's of $\mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B}) \times \mathbf{D}_{q-r^*}\mathbf{D}_{q-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{C}_{p-r^*}$. Note that $(p - q)$ CR's in the second determinant of (A.5) are automatically zero as $p \geq q$. ∎

Proof of Theorem 3.2. Define

$$\mathbf{X}_{r^*} \equiv \mathbf{C}_{p-r^*}'T^{1/2}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{D}_{q-r^*}. \tag{A.6}$$

Hence, from Assumptions 2.2 and 2.4 and (A.6),

$$\mathrm{vec}(\mathbf{X}_{r^*}) \to^{L} N_{(p-r^*)(q-r^*)}(\mathbf{0},(\mathbf{D}'_{q-r^*}\otimes\mathbf{C}'_{p-r^*})\boldsymbol{\Omega}(\mathbf{D}_{q-r^*}\otimes\mathbf{C}_{p-r^*})). \tag{A.7}$$

Now the CR statistic of (3.4) may be written from Theorem 3.1 and (A.6) as

$$\begin{aligned}\mathcal{CRT}_{r^*} &= \mathrm{tr}\{\mathbf{X}_{r^*}\mathbf{X}'_{r^*}\} + o_P(1)\\ &= \mathrm{vec}(\mathbf{X}_{r^*})'\mathrm{vec}(\mathbf{X}_{r^*}) + o_P(1);\end{aligned} \tag{A.8}$$

that is, $\mathcal{CRT}_{r^*}$ is a quadratic form in the asymptotically normally distributed random vector $\mathrm{vec}(\mathbf{X}_{r^*})$. Therefore, from (A.7) and (A.8), it immediately follows from Vuong (1989, Lemma 3.2, p. 312) that $\mathcal{CRT}_{r^*}$ has the limiting distribution stated. ■

Remark. Note that

$$\begin{aligned} &rk((\mathbf{D}'_{q-r^*}\otimes\mathbf{C}'_{p-r^*})\boldsymbol{\Omega}(\mathbf{D}_{q-r^*}\otimes\mathbf{C}_{p-r^*}))\\ &\quad\le \min\{rk(\mathbf{D}_{q-r^*}\mathbf{D}'_{q-r^*}\otimes\mathbf{C}_{p-r^*}\mathbf{C}'_{p-r^*}), rk(\boldsymbol{\Omega})\}\\ &\quad= \min\{s,(p-r^*)(q-r^*)\}.\end{aligned}$$

Proof of Corollary 3.1. The result immediately follows as $\mathbf{C}'_{p-r^*}\boldsymbol{\Sigma}^{-1}\mathbf{C}_{p-r^*} = \mathbf{I}_{p-r^*}$ and $\mathbf{D}'_{q-r^*}\boldsymbol{\Psi}^{-1}\mathbf{D}_{q-r^*} = \mathbf{I}_{q-r^*}$. Hence, $(\mathbf{D}'_{q-r^*}\otimes\mathbf{C}'_{p-r^*})\boldsymbol{\Omega}(\mathbf{D}_{q-r^*}\otimes\mathbf{C}_{p-r^*}) = (\mathbf{I}_{q-r^*}\otimes\mathbf{I}_{p-r^*})$. ■

Proof of Theorem 4.1. Without loss of generality, we impose the normalizations $\|\hat{\mathbf{c}}_i\| = 1$, $i = r^*+1,\dots,p$ and $\|\hat{\mathbf{d}}_i\| = 1$, $i = r^*+1,\dots,q$. Hence, $\hat{\mathbf{C}}_{p-r^*} = O_P(1)$ and $\hat{\mathbf{D}}_{q-r^*} = O_P(1)$. Now, $\hat{\mathbf{C}}'_{p-r^*}\hat{\mathbf{B}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{B}}'\hat{\mathbf{C}}_{p-r^*} = \hat{\boldsymbol{\Lambda}}_{p-r^*}$, where $\hat{\boldsymbol{\Lambda}}_{p-r^*} = \mathrm{diag}\ (\hat{\lambda}_{r^*+1},\dots,\hat{\lambda}_p)$. As $\hat{\boldsymbol{\Lambda}}_{p-r^*} = O_P(T^{-1})$ from Lemma 3.1 and Theorem 3.2, using Assumption 2.3, $\hat{\mathbf{B}}'\hat{\mathbf{C}}_{p-r^*} = o_P(1)$ and, from Assumption 2.2, $\mathbf{B}'\hat{\mathbf{C}}_{p-r^*} = o_P(1)$. That is, the columns of $\hat{\mathbf{C}}_{p-r^*}$ are a weakly consistent estimator for a basis of the null space $\mathcal{N}(\mathbf{B}')$ of $\mathbf{B}'$. Hence, subject to normalization and identifying constraints, $\hat{\mathbf{C}}_{p-r^*} - \mathbf{C}_{p-r^*} = o_P(1)$, and, by a similar argument, $\hat{\mathbf{D}}_{q-r^*} - \mathbf{D}_{q-r^*} = o_P(1)$. Therefore, from Assumption 4.1,

$$(\hat{\mathbf{D}}'_{q-r^*}\otimes\hat{\mathbf{C}}'_{p-r^*})\hat{\boldsymbol{\Omega}}(\hat{\mathbf{D}}_{q-r^*}\otimes\hat{\mathbf{C}}_{p-r^*}) - (\mathbf{D}'_{q-r^*}\otimes\mathbf{C}'_{p-r^*})\boldsymbol{\Omega}(\mathbf{D}_{q-r^*}\otimes\mathbf{C}_{p-r^*}) = o_P(1).$$

Consequently, by a similar argument to that used in the proof of Lemma 3.1, the ordered CR estimators $\{\hat{\lambda}_i^{r^*}\}_{i=1}^{(p-r^*)(q-r^*)}$ are consistent for their ordered counterparts $\{\lambda_i^{r^*}\}_{i=1}^{(p-r^*)(q-r^*)}$. Therefore, noting the continuity of the c.d.f. $F_{r^*}^{CRT}(\cdot)$ of (4.1), the result follows as $\sum_{i=1}^{(p-r^*)(q-r^*)}(\hat{\lambda}_i^{r^*} - \lambda_i^{r^*})Z_i^2 = o_P(1)$. ■

Proof of Theorem 4.2. Let $c_{1-\alpha}^{r^*}$ denote the $100(1-\alpha)$ percentile of the c.d.f. $F_{r^*}^{CRT}(\cdot)$ of (4.1); that is, $\mathcal{P}\{\sum_{i=1}^{t^*}\lambda_i^{r^*}Z_i^2 \ge c_{1-\alpha}^{r^*}\} = \alpha$. Consider the identical events $\{\mathcal{CRT}_{r^*} \ge \hat{c}_{1-\alpha}^{r^*}\}$ and $\{\mathcal{CRT}_{r^*} - (\hat{c}_{1-\alpha}^{r^*} - c_{1-\alpha}^{r^*}) \ge c_{1-\alpha}^{r^*}\}$. From Theorem 4.1, as the c.d.f. $F_{r^*}^{CRT}(\cdot)$ is continuous, $\hat{c}_{1-\alpha}^{r^*} - c_{1-\alpha}^{r^*} = o(1)$. Hence, the statistics $\mathcal{CRT}_{r^*}$ and $\mathcal{CRT}_{r^*} - (\hat{c}_{1-\alpha}^{r^*} - c_{1-\alpha}^{r^*})$ have identical limiting distributions. Therefore, $\lim_{T\to\infty}\mathcal{P}\{\mathcal{CRT}_{r^*} \ge \hat{c}_{1-\alpha}^{r^*}\} = \lim_{T\to\infty}\mathcal{P}\{\mathcal{CRT}_{r^*} \ge c_{1-\alpha}^{r^*}\}$. ■

Proof of Theorem 5.1. From Lemma 3.1, $\hat{\lambda}_i \to^{p} \tau_i^2 > 0$, $i = 1,\dots,r^*$. Therefore, $\mathcal{CRT}_r \to^{P} \infty$, $r = 0,\dots,r^*-1$. ■

Proof of Theorem 5.2. Define the event $\mathcal{A}_{rT} \equiv \{\mathcal{CRT}_r \ge \hat{c}_{1-\alpha_{rT}}^{r}\}$, $r = 0,\dots,q-1$. Hence, $\mathcal{P}\{\hat{r} = r\} = \mathcal{P}\{\cap_{i=0}^{r-1}\mathcal{A}_{iT}\cap\mathcal{A}_{rT}^{c}\}$, where $\mathcal{A}_{rT}^{c}$ denotes the complement of $\mathcal{A}_{rT}$. From

Pötscher (1983) and Theorem 4.1, if (a) $\alpha_{rT} = o(1)$, then $\hat{c}^r_{1-\alpha_{rT}} \to \infty$, and, if (b) $-T^{-1} \ln \alpha_{rT} = o(1)$, then $T^{-1}\hat{c}^r_{1-\alpha_{rT}} = o(1)$.

First, consider the case $r < r^*$. Hence, by Lemma 3.1 and (b),

$$\begin{aligned}\mathcal{P}\{\hat{r} = r\} &\leq \mathcal{P}\{\mathcal{A}^c_{rT}\} = 1 - \mathcal{P}\{\mathcal{A}_{rT}\}\\ &= 1 - \mathcal{P}\{T^{-1}\mathcal{CRT}_r \geq T^{-1}\hat{c}^r_{1-\alpha_{rT}}\} \to 0\end{aligned}$$

as $T \to \infty$. Second, consider the case $r > r^*$. Thus, by Theorem 3.2 and (a),

$$\mathcal{P}\{\hat{r} = r\} \leq \mathcal{P}\{\mathcal{A}_{r^*T}\} = \mathcal{P}\{\mathcal{CRT}_{r^*} \geq \hat{c}^{r^*}_{1-\alpha_{r^*T}}\} \to 0$$

as $T \to \infty$. Therefore, $\lim_{T\to\infty} \mathcal{P}\{\hat{r} = r^*\} = 1$. ■