cemmap
centre for microdata methods and practice

On the Conditional Likelihood Ratio Test for Several Parameters in IV Regression

Grant Hillier

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP26/06

On the Conditional Likelihood Ratio Test for Several Parameters in IV Regression*

By

Grant Hillier

CeMMAP and Economics Division, University of Southampton.

Department of Economics
University of Southampton
Highfield, Southampton SO17 1BJ
United Kingdom
email: ghh@soton.ac.uk
tel: (44) 02380-592659

*I am grateful to Peter Phillips and three anonymous referees for very useful comments on earlier versions of this paper, and to Ray O'Brien for helpful conversations on the computational issues discussed in Section 5

Address for Proofs: Professor Grant Hillier
Department of Economics
University of Southampton
Highfield, Southampton SO17 1BJ
United Kingdom

Abstract

For the problem of testing the hypothesis that all m coefficients of the right-hand-side endogenous variables in an IV regression are zero, the likelihood ratio (LR) test can, if the reduced form covariance matrix is known, be rendered similar by a conditioning argument. To exploit this fact requires knowledge of the relevant conditional *cdf* of the LR statistic, but the statistic is a function of the smallest characteristic root of an $(m+1)-$square matrix, and is therefore analytically difficult to deal with when $m > 1$. We show in this paper that an iterative conditioning argument used by Hillier (2006) and Andrews, Moreira, and Stock (2007) to evaluate the *cdf* in the case $m = 1$ can be generalized to the case of arbitrary m. This means that we can completely bypass the difficulty of dealing with the smallest characteristic root. Analytic results are obtained for the case $m = 2$, and a simple and efficient simulation approach to evaluating the *cdf* is suggested for larger values of m.

1 Introduction

Interest in the problem of hypothesis testing in the classical IV-regression/structural-equation model has seen a resurgence in the last few years, largely because recent work on weak instruments has shown that inference procedures founded on purely asymptotic arguments can be extremely unreliable. When errors are Gaussian with known covariance matrix, and we are testing the coefficients of all m (say) right-hand-side endogenous variables jointly, we know from the work of Moreira (2003), and Andrews, Moreira, and Stock (2006, 2007), and others, that the likelihood ratio (LR) test can be rendered similar by a conditioning argument. And, it seems from simulation results that replacing the assumed-known covariance matrix by an estimator may not be too damaging to the results. Thus, in principle at least, the size of the LR test need not be sensitive to weak instruments, or inadequate information in any sense, though of course its power will be. The problem is to provide the critical values needed to implement this so-called conditional LR (CLR) test, and it is to this issue that this paper is addressed.

The main part of the paper is a purely theoretical exercise. The reader interested only in using the test in practice should finish reading this Introduction, and then skip directly to Section 5, which explains in detail how the results in the paper can be implemented.

In the case $m = 1$, Hillier (2006) has recently derived the exact conditional *cdf* of the LR statistic, and used this to provide plots of the critical value function needed to implement the CLR test. He also uses the same analytical approach to derive the exact conditional power functions for the CLR test, and some other tests that have been proposed for the problem. Andrews, Moreira, and Stock (2007) had earlier used an equivalent argument, together with numerical integration, to compute the conditional null *cdf* of the LR statistic for the case $m = 1$, but power calculations have hitherto relied entirely on rather crude simulation methods.

When $m > 1$ the LR statistic involves the smallest characteristic root of an $(m + 1)-$square symmetric matrix, i.e., the smallest root of a polynomial of degree $m+1$. It would therefore seem on the face of it that little analytical progress could be made for the case $m > 1$. Even in the case $m = 2$ we have to deal with the smallest root of a cubic - a formidable formula in itself. And, for $m \geq 4$, of course, no explicit formula for the roots is available. Both Hillier (2006, Section 3) and Kleibergen (2007) discuss - as a way round this problem - a similar test for several parameters that is based upon, but not in general identical to, the CLR test. This approach is easy to implement, but since most of the power calculations so far available suggest that the true CLR test is generally near-optimal, it seems desirable to provide methods that facilitate its use in practice, and that is our purpose here.

The main accomplishment of this paper is to show that a generalized version of the approach used in Hillier (2006) and Andrews, Moreira, and Stock (2007) for the case $m = 1$ can also be used in the case $m > 1$. That is, we shall show that, *notwithstanding the fact that the LR statistic cannot be written down directly as a function of the underlying data, its conditional cdf can be computed by a sequential conditioning argument.*

The argument is iterative. The first step is to show that, after a suitable transformation of variables, the statistic of interest - a complicated (in general, implicit) function of the sufficient statistics for the problem - is monotonically related to a random variable that is independent of all other variables and has very simple distribution properties.[1] Conditioning on those remaining variables, the null conditional *cdf* of the LR statistic is therefore easily written down. The conditional *cdf* we seek is, after this first step, the conditional *expectation* of a well-behaved function of the remaining variables. In the case $m = 1$ the first step leaves only the evaluation of a one dimensional integral to yield the required conditional *cdf*. In Hillier (2006) this integral is evaluated analytically, and Andrews, Moreira, and Stock (2007) evaluate it by numerical integration.

In the case $m > 1$ the first-step result provides simple bounds on the conditional *cdf* that can be useful without further calculation. But, continuing with the discussion of the exact *cdf*, the first step in the process leaves an $m-$dimensional integration that in turn involves the smallest root of a polynomial, this time of degree m. Using an argument which exactly parallels that used for the first step, we show that this root is again a monotonic function of a random variable with simple distribution properties, and which is again independent of the remaining variables. Thus, the integral required after the first step can be evaluated reasonably easily, given the remaining $m - 1$ variables. In the case $m = 2$ this second step leaves only the evaluation of a one-dimensional integral - again, the expectation of a fairly simple function of that variable - and we give the analytic details for this case in Section 4.

Remarkably, when $m > 2$ the function whose expectation remains to be evaluated after this second step, is an *explicit, relatively simple, function of the remaining random variables.* Thus, even when $m > 2$, the second step in this iterative conditioning process is sufficient to completely bypass the complexity arising from the fact that we are dealing with a characteristic root. Evidently, this result has general applicability well beyond the confines of the problem to which it is applied here.

Of course, when $m > 2$ (and even when $m = 2$ - see Section 4.2 below) the integrations required to produce an analytical expression for the *cdf* are formidable. A second purpose of the paper is to suggest a simple simulation procedure for evaluating the integrals involved, based on the fact that they are essentially the *expectations* of relatively simple random functions, with nice properties. Section 5 discusses the implementation of the tests in practice, and the associated computational issues. Throughout the paper we denote the *cdf* of the $\chi^2(k)$ distribution by $G_k(c) = \Pr\{\chi^2(k) < c\}$.

2 Model, Assumptions, LR Test

We use, as is common in this literature, the stripped-down IV model consisting of a single linear structural equation (or IV regression equation), together with a reduced form equation for the right-hand-side endogenous variables that appear in it:

$$y_1 = Y_2\beta + u, \tag{1}$$

$$Y_2 = Z\Pi + V, \tag{2}$$

with y_1 $n \times 1, Y_2$ $n \times m$, Z $n \times k$, and β $m \times 1$. In general we might wish to test the hypothesis $H_0 : \beta = \beta_0$, but there is no loss of generality in taking $\beta_0 = 0$, which we do from now on (simply replace y_1 by $y_1 - Y_2\beta_0$ if $\beta_0 \neq 0$). The changes needed to accomodate the more realistic case where both equations contain additional exogenous variables are described in Section 5. Combining (1) and (2) produces, under Gaussian assumptions, a joint distribution for $Y = (y_1, Y_2)$ of the form:

$$Y \sim N(Z\Pi(\beta, I_m), I_n \otimes \Omega), \tag{3}$$

and we assume that the covariance matrix for the rows of Y, Ω, is known. In this model the matrix $Z'Y$ is minimal sufficient for the unknown parameters (Π, β). Under these assumptions there is no loss of generality in assuming that the variables have been scaled so that Ω has the form of a correlation matrix, $\Omega = \{\rho_{ij}; i, j = 1, ..., m{+}1\}$, with $\rho_{ii} = 1$. We assume that this has been done, and merely note that, although this rescales the parameters too, it does not affect the truth or falsity of the hypothesis of interest. Finally, to avoid degeneracies, we assume throughout that $k > m$, so that β is overidentified.

Remark 1 *The assumptions of joint normality and known covariance matrix are, of course, unrealistic. To motivate these assumptions we may (as usual in this literature) appeal to results in Staiger and Stock (1997), where it is shown that, under so-called weak-instrument asymptotics, the asymptotic properties of inferential procedures based on the sufficient statistics for the model (3) correspond to those that apply under the assumptions made above. Alternatively, we may note the simulation evidence that, in the Gaussian model with unknown covariance matrix, using an estimator for Ω seems to leave the main results that follow approximately intact. Note, though, that the problem of exact inference in the Gaussian model with unknown covariance matrix is not the subject of this paper, at least not directly.*

There is a block-triangular matrix U_Ω of the form

$$U_\Omega = \begin{pmatrix} 1 & a' \\ 0 & A \end{pmatrix}$$

with the property that $U_\Omega' \Omega U_\Omega = I_{m+1}$. We define the standardised sufficient statistic

$$(p_1, P_2) = (Z'Z)^{-\frac{1}{2}} Z'YU_\Omega. \tag{4}$$

Under the null hypothesis we then have that p_1 and P_2 are independent, $p_1 \sim N(0, I_k)$, and $P_2 \sim N(M, I_m \otimes I_k)$, with M (a transformation of Π) a nuisance parameter. Since the the null distribution of P_2 is complete, similarity for tests of H_0 can only be achieved by tests that have Neyman structure with respect to P_2, i.e., whose size *conditional on* P_2 does not depend on P_2 (see Hillier (1987b) for background, and Moreira (2003)). In what follows, therefore, we shall always be concerned with the *conditional* properties of the LR test given P_2.

The LR test rejects the null hypothesis when the statistic

$$T = q - f_1 \tag{5}$$

is large, where f_1 is the smallest characteristic root of

$$W = (p_1, P_2)'(p_1, P_2), \tag{6}$$

and $q = w_{11} = p_1'p_1$. Note that, if $k = m$, $f_1 = 0$ and the LR test is identical to a test based on the Anderson-Rubin statistic q, which is $\chi^2(k)$ under H_0. It is for this reason that we assume that k is strictly larger than m. Under this condition, because T is a function of P_2 the usual LR test is non-similar: its size depends on M. However, exactly as in the case $m = 1$, the test can be rendered similar by choosing the critical value in such a way that, conditional on P_2, the size of the test is free of P_2. In fact, we shall see below that the distribution of T depends on P_2 only through the characteristic roots of $W_{22} = P_2'P_2$, so we need only consider the conditional properties of T for fixed values of those roots.

In general we cannot write f_1 explicitly as a function of W. Nevertheless, as indicated in the Introduction, the conditional distributional properties of T, given P_2, can be deduced by an indirect argument.[2] First, let

$$W_{22} = P_2'P_2 = HD^2H'$$

be the spectral decomposition of W_{22}, where $D = diag\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_m}\}$, the λ_i being the characteristic roots of W_{22}, and H is an $m \times m$ orthogonal matrix containing the eigenvectors of W_{22}. We assume throughout that the λ_i are labelled in *increasing* order, and, to avoid repeated diversions, we assume that the λ_i are distinct (except where otherwise indicated). All of the results to follow can be obtained without this assumption, but it does simplify the proofs. Next define

$$r = D^{-1}H'P_2'p_1, \tag{7}$$

so that

$$W = \begin{pmatrix} q & r'DH' \\ HDr & HD^2H' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix} \begin{pmatrix} q, & r'D \\ Dr, & D^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix}',$$

and f_1 is also the smallest characteristic root of

$$\tilde{W} = \begin{pmatrix} q, & r'D \\ Dr, & D^2 \end{pmatrix}. \tag{8}$$

It follows from these observations that:

Proposition 1 *The properties of the LR test statistic depend on the properties of* W *only through the properties of* q, r, *and the characteristic roots* $\lambda_1, ..., \lambda_m$ *of* W_{22}.

We can therefore focus on the variates q, r, and $\lambda_1, ..., \lambda_m$, and conditioning on P_2 is equivalent to conditioning on the characteristic roots $\lambda_1, ..., \lambda_m$. Now, the characteristic polynomial of $\tilde{W}$ is, up to sign,

$$\psi(f) = \det[fI_{m+1} - \tilde{W}] = \det \begin{pmatrix} f-q, & -r'D \\ -Dr, & fI_m - D^2 \end{pmatrix}.$$

Direct expansion of the determinant yields:

$$\psi(f) = (f-q)\Pi_{i=1}^m (f-\lambda_i) - \sum_{i=1}^m \lambda_i r_i^2 \left(\prod_{j \neq i=1}^m (f-\lambda_j) \right), \tag{9}$$

and f_1 is the smallest root of the equation $\psi(f) = 0$. We now show that, after a suitable transformation to new variables, this characteristic polynomial can be written as a simple function of q - in fact, it is linear in q. This will enable us to show (in the next Section) that T is strictly monotonically increasing in q when all of the remaining variables - which are constructed so as to be independent of q - are held fixed.

Define

$$q_1 = p_1' M_{P_2} p_1, \tag{10}$$

where $M_{P_2} = I_k - P_2(P_2'P_2)^{-1}P_2'$. Under the null hypothesis, q_1 and r are conditionally independent, given P_2,

$$\begin{aligned} q_1|P_2 &\sim \chi^2(k-m), \text{ and} \\ r|P_2 &\sim N(0, I_m). \end{aligned}$$

Since these distributional properties are free of the conditioning variables P_2 they also hold unconditionally. Thus, unconditionally, q_1 and r are independent and are independent of P_2, and hence also of the characteristic roots of W_{22}, the λ_i.

Next, let

$$q_{2i} = r_i^2, \ i = 1, ..., m.$$

Under the null hypothesis the q_{2i} are independent of q_1, of each other, and of the λ_i, and each is $\chi^2(1)$. Now define the $m+1$ new variates

$$\left.\begin{aligned} &q = q_1 + \Sigma_{i=1}^m q_{2i}, \\ &b = q_1/q, \\ &c_i = q_{2i}/(\Sigma_{i=1}^m q_{2i}), \ i = 1, ..., m-1. \end{aligned}\right\} \tag{11}$$

The following null distribution properties for $(q, b, c_1, ...c_{m-1})$ follow easily:

Proposition 2 *Under the null hypothesis, q, b, and $c_1, ..., c_{m-1}$ are independent, $q \sim \chi^2(k)$, $b \sim Beta(\frac{k-m}{2}, \frac{m}{2})$, and $(c_1, ..., c_{m-1})$ have the Dirichlet distribution with parameters $(\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2}; \frac{1}{2})$. These $m+1$ variables are independent of $\lambda_1, ..., \lambda_m$, and we have*

$$\begin{aligned} q_1 &= bq, \\ q_{2i} &= c_i(\Sigma_{i=1}^m q_{2i}) = c_i q(1-b),\ i = 1, ..., m-1. \end{aligned}$$

We also put $c_m = 1 - \Sigma_{i=1}^{m-1} c_i$, so that $q_{2m} = q(1-b)c_m$.

We may now rewrite the characteristic polynomial (9) in terms of these variables as:

$$\psi(f) = f\psi_1(f) - q\psi_2(f), \tag{12}$$

where

$$\psi_1(f) = \Pi_{i=1}^m (f - \lambda_i) \tag{13}$$

is essentially the characteristic polynomial of W_{22} (of degree m in f), and

$$\psi_2(f) = \psi_1(f) + (1-b)\sum_{i=1}^m c_i\lambda_i \left(\prod_{j\neq i=1}^m (f - \lambda_j) \right), \tag{14}$$

again a polynomial of degree m in f.

Notice that $\psi_1(f)$ is determined entirely by $(\lambda_1, ..., \lambda_m)$, which we will condition on (i.e., hold fixed), and that $\psi_2(f)$ is determined by $(b, c_1, ..., c_{m-1})$ and $(\lambda_1, ..., \lambda_m)$, all of which are independent of q under H_0. If we initially hold all of these variates fixed the roots of $\psi(f) = 0$ are the points where the *fixed* polynomial $f\psi_1(f)$ intersects one member of the *family* of polynomials indexed by q :

$$\mathcal{F}_{\psi_2} = \{q\psi_2(f) : q > 0\}.$$

Note that all members of the family $\mathcal{F}_{\psi_2}$ vanish at the same points, the m roots of the equation $\psi_2(f) = 0$; this is the key to the monotonicity property that we seek to establish.

3 First Monotonicity Property

The argument in Hillier (2006), which deals with the case $m = 1$, is based on showing that the LR statistic T is (for fixed values of the other variables involved) monotonic-increasing in q. This is easily established in the case $m = 1$ by using the explicit formula for the statistic directly. Only the formula for the smallest root of a quadratic is needed in the case $m = 1$. In the general case a direct approach is unavailable, because the smallest root f_1 cannot usually be written explicitly as a function of the other variables. In this section we shall show that, nevertheless, an exactly analogous monotonicity property holds when $m > 1$. Some care is needed to make the argument precise.

For fixed $(b, c_1, ..., c_{m-1})$ and $(\lambda_1, ..., \lambda_m)$, the equation $\psi(f) = f\psi_1(f) - q\psi_2(f) = 0$ may be interpreted as defining a relation between the scalar variables q and f. It is this relation that we first need to study. In the direction $f \to q$ the relation is defined by the equation $q(f) = f\psi_1(f)/\psi_2(f)$. Here, since we require that $q > 0$, f must be restricted to values for which ψ_1 and ψ_2 have the same sign. Subject to this restriction, and the requirement that ψ_2 must not vanish, this is a well-defined function for all $f > 0$.

In the direction $q \to f$, however, there is a difficulty: for each $q > 0$ we know that the equation $\psi(f) = 0$ has $m+1$ real roots, so in this direction the relation (which is only implicitly defined, in general) cannot be represented by a well-defined function. The key to resolving this difficulty is to recall the Cauchy interlacing inequalities (Courant and Hilbert (1953), p.454), which, together with the fact that W is positive-definite symmetric, assert that

$$0 \le f_1 \le \lambda_1 \le .... \le f_m \le \lambda_m \le f_{m+1} < \infty. \tag{15}$$

Since, for our purposes, the λ_i will be treated as fixed, these inequalities may be used to partition the domain of f into non-overlapping intervals, and we may represent the relation $q \to f$ by a function with $m + 1$ *branches*, one in each interval.

Thus, let $L_i = [\lambda_{i-1}, \lambda_i)$, $i = 1, ..., m+1$ (with $\lambda_0 \equiv 0$ and $\lambda_{m+1} \equiv \infty$) denote $m+1$ disjoint intervals whose union is R^+. In each such interval we may define a branch of the function $q \to f$ (see Figure 1 below for the case $m = 2$). Correspondingly, we may restrict the domain of $q(f)$ to L_i, so that we have a matching set of functions $q_i(f) : L_i \mapsto R$ defined by $q_i(f_i) = f_i\psi_1(f_i)/\psi_2(f_i)$, with $f_i \in L_i$. Note that, with a slight abuse of notation, f_i here denotes a variable with domain L_i (not just a root of $\psi(f) = 0$). It is now at least plausible that each branch of the function, relating q and an f_i, is monotonic (one-to-one), and we shall in fact show that this is the case.

Before doing so we need to establish some properties of the polynomial $\psi_2(f)$ for fixed b, $c_1, ..., c_{m-1}$, and $\lambda_1, ..., \lambda_m$. The results needed are collected in the following Lemma:

Lemma 1 *(i) The polynomial $\psi_2(f)$ has exactly m real roots f_{2i}, say, with precisely one root in each of the intervals L_i, $i = 1, ..., m$. (ii) If the f_{2i} are labelled in increasing order, for each $i = 1, ..., m$ we have $\lambda_{i-1} \le f_i \le f_{2i} \le \lambda_i$. (iii) For $i = 1, ..., m$, on each of the sub-intevals $\lambda_{i-1} < f < f_{2i}$ containing the roots f_i of $\psi(f) = 0$, $\psi_2(f) \neq 0$ and $q(f) > 0$. For $f > \lambda_m$ (i.e., $f \in L_{m+1}$), $\psi_2(f) \neq 0$ and $q(f) > 0$.*

Proof. All proofs are in the Appendix. ∎

Let $R_i = [\lambda_{i-1}, f_{2i}) \subset L_i$, $i = 1, ..., m + 1$, with $\lambda_0 = 0$ and $f_{2,m+1} = +\infty$. According to the Lemma, $\psi_2(f_i) \neq 0$ on R_i, and $q(f_i) > 0$. From now on we denote by f_i only values of f in R_i. We are now in a position to show that each of the relations between q and $f_i \in R_i$ defined by $\psi(f) = 0$ is strictly monotonic-increasing, and hence one-to-one. Further, we will show that this is also true for the function $T(q) = q - f_1(q)$, the function corresponding to the LR statistic.

Consider first the function $q(f)$ defined by

$$q(f) = \frac{f\psi_1(f)}{\psi_2(f)}, \tag{16}$$

and the function

$$T(f) = q(f) - f, \tag{17}$$

which corresponds to the LR statistic when f is restricted to R_1. The following result shows that both $q(f)$ and $T(f)$ are strictly monotonic-increasing.

Lemma 2 *Let $q(f) = f\psi_1(f)/\psi_2(f)$ and $T(f) = q(f) - f$. Then, $q'(f) > 0$ and $T'(f) > 0$. i.e., both $q(f)$ and $T(f)$ are strictly monotonic-increasing in f.*

An immediate consequence of this simple Lemma is the following:

Theorem 1 *For fixed values of $(b, c_1, ..., c_{m-1})$ and $(\lambda_1, ..., \lambda_m)$, and for $i = 1, ..., m+1$, each of the functions $f_i(q) : R^+ \to R_i$ is strictly monotonic-increasing in q. The relationship between q and any root f_i of $\psi(f) = 0$ - i.e., any branch of the function $f(q)$ - is therefore one-to-one.*

In Figure 1 we illustrate these results for the case $m = 2$. The two solid horizontal lines are the roots f_{21} and f_{22} of $\psi_2(f) = 0$, the two dashed lines λ_1 and λ_2. The functions were plotted using the ImplicitPlot facility in Maple, and each branch is clearly monotonic-increasing in q. Figure 2 shows why this monotonicity arises: we show the *fixed* cubic $f\psi_1(f)$, and several members of the family of quadratics $\digamma_{\psi_2} = \{q\psi_2(f) : q > 0\}$ generated by varying q. The solid quadratic corresponds to a small value of q, the dotted to a larger value, and the dashed to a larger value still. The monotonicity arises because each of the functions $q\psi_2(f)$ must cross the axis at the same points f_{21} and f_{22}.

We now focus on the branch where $f \in R_1$. Theorem 1 means that, for fixed values of all other variables, we may regard $T = q - f_1$ as a function of f_1 alone: $T(f_1) = q(f_1) - f_1$. This function $T(f_1)$ is monotonic increasing, and it is easy to see that $T(f_1) \to +\infty$ as $f_1 \to f_{21}$. Thus, for any $z > 0$, the inequality $T(f_1) < z$ corresponds to an inequality $f_1 < T^{-1}(z)$, with $f_1 \in R_1$. But this in turn corresponds to an inequality for q, because $f_1(q)$ is monotonic-increasing. We thus have:

Proposition 3 *For any $z > 0$ the inequality $T < z$ corresponds to the inequality*

$$q < z + e_1, \tag{18}$$

where e_1 is the smallest root of the equation

$$\tau(e) = z\psi_1(e) + (z + e)(1 - b) \sum_{i=1}^{m} c_i \lambda_i \left[\prod_{j \neq i} (e - \lambda_j) \right] = 0. \tag{19}$$

Note that part of the proof of Proposition 3 entails showing that e_1 is the *smallest* root of the equation $\tau(e) = 0$. Note also that $\tau(e)$ is a polynomial of degree m in e, not $m+1$.

3.1 The conditional distribution of T

Now, let

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) = \Pr\{T < z | \lambda_1, ..., \lambda_m\}, \tag{20}$$

denote the conditional *cdf* of T, given $(\lambda_1, ..., \lambda_m)$. Also let $E_x[h(x)]$ denote the expectation of any function of x (which may be a vector) with respect to the distribution of x. Since, under H_0, q is independent of the remaining variables, and $q \sim \chi^2(k)$, the following Theorem follows at once from Proposition 3 and Theorem 1:

Theorem 2 *For all $m < k$,*

$$\Pr\{T < z | b, c_1, ..., c_{m-1}; \lambda_1, ..., \lambda_m\} = G_k(z + e_1). \tag{21}$$

Hence,

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) = E_{b,c_1,...,c_{m-1}}[G_k(z + e_1)]. \tag{22}$$

Equation (22) generalizes equation (19) in Hillier (2006). Note that it expresses the conditional *cdf* $P_{k,m}(z; \lambda_1, ..., \lambda_m)$ as the *expectation* of a relatively simple function of the m variates $(b, c_1, ..., c_{m-1})$.

Example 1 *The case $m = 1$.*

When $m = 1$ the $c's$ are missing and $\psi(f)$ is the quadratic

$$\psi(f) = f^2 - f(q + w) + qwb = f(f - w) - q(f - wb), \tag{23}$$

where $w = W_{22}$ replaces the roots λ_i of W_{22}. So, $\psi_1(f) = f - w$ and $\psi_2(f) = f - wb$. Hence e_1 is the solution to the linear equation

$$T(e) = \frac{e\psi_1(e)}{\psi_2(e)} - e = \frac{ew(b-1)}{e - wb} = z,$$

so $e_1 = zwb/(z + w - wb)$ and

$$z + e_1 = \frac{z(z + w)}{(z + w - wb)} = z(1 - ab)^{-1},$$

with $a = w/(z + w)$. This is the result used in Hillier (2006), and is equivalent to the result used in Andrews, Moreira, and Stock (2007). In this case the expectation $E_b[G_k(z(1 - ab)^{-1})]$ can be evaluated explicitly.

Example 2 *The case $m = 2$.*

In this case $\psi(f)$ is the cubic

$$\psi(f) = f^3 - f^2 s_2 + f s_1 - s_0 = 0, \tag{24}$$

where

$$\left.\begin{aligned} s_2 &= trace[\tilde{W}], \\ s_1 &= trace[adj(\tilde{W})], \\ s_0 &= \det[\tilde{W}]. \end{aligned}\right\} \tag{25}$$

This becomes, in terms of the new variables (q, b, c),

$$\psi(f) = (f-q)(f-\lambda_1)(f-\lambda_2) - q(1-b)[c\lambda_1(f-\lambda_2) + (1-c)\lambda_2(f-\lambda_1)], \tag{26}$$

and in this case $c \sim Beta(\frac{1}{2}, \frac{1}{2})$. Thus,

$$\psi_1(f) = (f-\lambda_1)(f-\lambda_2),$$

$$\psi_2(f) = \psi_1(f) + (1-b)[c\lambda_1(f-\lambda_2) + (1-c)\lambda_2(f-\lambda_1)],$$

and e_1 is the smaller root of the quadratic equation

$$\tau(e) = z(e-\lambda_1)(e-\lambda_2) + (1-b)(e+z)[c\lambda_1(e-\lambda_2) + (1-c)\lambda_2(e-\lambda_1)] = 0.$$

Remaining details for this case are given in Section 4.2 below.

Example 3 *The case of equal λ_i.*

If $\lambda_i = \lambda$ for $i = 1, ..., m$ then

$$\tau(e) = (e-\lambda)^{m-1}\{z(e-\lambda) + \lambda(z+e)(1-b)\},$$

which does not depend on the c_i. The smallest root of $\tau(e) = 0$ is thus seen to be $e_1 = z\lambda b/(z + \lambda(1-b))$, and so

$$z + e_1 = \frac{z(z+\lambda)}{(z+\lambda-\lambda b)} = z(1-ab)^{-1},$$

with $a = \lambda/(z+\lambda)$, and $b \sim Beta(\frac{k-m}{2}, \frac{m}{2})$. The unconditional cdf is, in this case,

$$Q_{k,m}(z;\lambda) = \Pr\{T < z|\lambda\} = E_b\left[G_k(z(1-ab)^{-1})\right], \tag{27}$$

with $b \sim Beta(\frac{k-m}{2}, \frac{m}{2})$. The expectation here is given (in different notation) in Hillier (2006), Section 4, in the form:

$$Q_{k,m}(z;\lambda) = (1-a)^{\frac{m}{2}} \sum_{j=0}^{\infty} \frac{a^j (\frac{m}{2})_j}{j!} G_{k+2j}(z+\lambda). \tag{28}$$

See Kleibergen (2007) for a related discussion.

4 Evaluating the Expectation

In order to complete the calculation of the conditional *cdf* $P_{k,m}(z;\lambda_1,...,\lambda_m)$ when the λ_i are not all equal we need to evaluate the expectation in equation (22) with respect to the joint distribution of the variates $(b, c_1, ..., c_{m-1})$. In principle this can be done analytically, or it can be done by numerical integration, generalizing the approach taken for the case $m = 1$ by Andrews, Moreira, and Stock (2007). However, the variables $(b, c_1, ..., c_{m-1})$ enter the problem only through e_1, which is itself the smallest root of a polynomial of degree m, so the problem still seems analytically complex. In fact, as noted in the Introduction, we can eliminate this difficulty (whichever approach is adopted) by iterating the argument used so far, as we explain in this Section.

In the equation defining e_1 we may set $u = z + e$, and rewrite the equation as an equation defining $u_1 = z + e_1$, which of course is a random variable - a function of $(b, c_1, ..., c_{m-1})$. Equation (22) then has the form of an expectation with respect to the distribution of u_1 :

$$P_{k,m}(z;\lambda_1,...,\lambda_m) = E_{u_1}[G_k(u_1)]. \tag{29}$$

As before, we may first evaluate this with respect to the *conditional distribution* of u_1 given $(c_1, ..., c_{m-1})$ (and $(\lambda_1, ..., \lambda_m)$), but it will be convenient to write the expectation in terms of the conditional *cdf* of u_1, rather than its density. The means of doing so is provided by the following Lemma:

Lemma 3 *Suppose that u_1 is supported on the interval $[\underline{u}_1,\bar{u}_1]$ with conditional density $p_{u_1}(v|c,\lambda)$, given $c = (c_1, ..., c_{m-1})$ and $\lambda = (\lambda_1, ..., \lambda_m)$, and conditional cdf*

$$F_{u_1}(v|c,\lambda) = \Pr\{u_1 < v|c,\lambda) \tag{30}$$

for $\underline{u}_1 \leq v \leq \bar{u}_1$, with $F_{u_1}(\underline{u}_1|c,\lambda) = 0$ and $F_{u_1}(\bar{u}_1|c,\lambda) = 1$. Then

$$\begin{aligned} P_{k,m}(z;\lambda) &= E_c E_{u_1|c}[G_k(u_1)] \\ &= E_c\left[G_k(\bar{u}_1) - \int_{\underline{u}_1}^{\bar{u}_1} g_k(v)F_{u_1}(v|c,\lambda)dv\right], \end{aligned} \tag{31}$$

where $g_k(v) = \exp\{-\frac{1}{2}v\}v^{\frac{k}{2}-1}/[2^{\frac{k}{2}}\Gamma(\frac{k}{2})]$ is the density function of the $\chi^2(k)$ distribution, $E_c[\cdot]$ denotes expectation with respect to the joint distribution of $(c_1, ..., c_{m-1})$, and $E_{u_1|c}[\cdot]$ conditional expectation with respect to u_1 given $(c_1, ..., c_{m-1})$.

We now use methods exactly analogous to those used above to evaluate the (conditional) *cdf* $F_{u_1}(v|c,\lambda)$, in this case by showing that, for fixed c and λ, the equation defining the relation between b and e is, when restricted to suitable intervals for e, monotonic. To do so we shall establish a second monotonicity property - analogous to Theorem 1 above - but in this case dealing with the roots of $\tau(e) = 0$ as functions of b.

4.1 A Second Monotonicity Property

Since the argument imitates that given above for $\psi(f)$, we shall be somewhat briefer. The argument rests on the fact that we can write

$$\tau(e) = \tau_1(e) - b\tau_2(e), \tag{32}$$

with

$$\tau_1(e) = z\Pi_{i=1}^m(e-\lambda_i) + (z+e)\sum_{i=1}^m c_i\lambda_i[\Pi_{j\neq i}(e-\lambda_j)], \tag{33}$$

and

$$\tau_2(e) = (z+e)\sum_{i=1}^m c_i\lambda_i[\Pi_{j\neq i}(e-\lambda_j)]. \tag{34}$$

Note that both τ_1 and τ_2 are of degree m in e, and that $\tau_1(0) = 0$ (i.e., the term of degree zero in $\tau_1(e)$ vanishes).

In this case we have the following analogue of the Cauchy interlacing inequalities (15):

Lemma 4 *The polynomial $\tau(e)$ has exactly m real, non-negative, roots $e_1, ..., e_m$ (labelled in increasing order), and these satisfy the interlacing inequalities*

$$\lambda_{i-1} \leq e_i \leq \lambda_i, \;\; i = 1, ..., m, \tag{35}$$

where $\lambda_0 = 0$. The e_i are not otherwise restricted.

As before, this result enables us to partition the domain of e in such a way that the relationship between b and e has a single branch in each interval $[\lambda_{i-1}, \lambda_i)$, $i = 1, ..., m$. Next, using the above representation for $\tau(e)$, we may express b as a function of e :

$$b(e) = \frac{\tau_1(e)}{\tau_2(e)} = 1 - zh(e), \tag{36}$$

where

$$h(e) = \left[\frac{\Pi_{i=1}^m(\lambda_i - e)}{(z+e)\sum_{i=1}^m c_i\lambda_i[\Pi_{j\neq i}(\lambda_j - e)]}\right]. \tag{37}$$

However, as before there is a problem in the direction $b \to e$, because the equation $\tau(e) = 0$ has m real roots. But, using Lemma 4, we partition the domain of e, the interval $[0, \lambda_m)$, into a union of intervals $L_i = [\lambda_{i-1}, \lambda_i)$, $i = 1, ..., m$, $(\lambda_0 \equiv 0)$, and consider the m distinct branches of the relationship seperately. In this case it is straightforward to check that $b'(e) > 0$, so, writing e_i for the restriction of e to L_i, and $b_i(e_i)$ for the branch with $e_i \in L_i$, we have the following analogue of Theorem 1:

Theorem 3 *On each interval L_i the function $b_i(e_i)$ is well-defined and monotonic-increasing in e_i, and is therefore invertible on L_i.*

It follows from Lemma 4 that, in Lemma 3, $\underline{u}_1 = z$ and $\bar{u}_1 = z + \lambda_1$. And, for v in the interval $(z, z + \lambda_1)$, the inequality $u_1 < v$ corresponds to the inequality $b < 1 - zh(v - z)$. Thus, we can now state:

Theorem 4 *Let*

$$F_{u_1}(v|c, \lambda) = \Pr\{u_1 < v|c_1, ..., c_{m-1}; \lambda_1, ..., \lambda_m) \tag{38}$$

denote the conditional cdf of u_1*, given* $c = (c_1, ..., c_{m-1})$ *and* $\lambda = (\lambda_1, ..., \lambda_m)$*, where* $u_1 = z+e_1$*, with* e_1 *the smallest root of* $\tau(e) = 0$*. Then, for* v *in the interval* $(z, z+\lambda_1)$*,*

$$F_{u_1}(v|c, \lambda) = \Pr\{b < 1 - zh(v - z)|c, \lambda\}, \tag{39}$$

and $F_{u_1}(v|c, \lambda) = 0$ *otherwise. Here,* $b \sim Beta\left(\frac{k-m}{2}, \frac{m}{2}\right)$.

An explicit form for the incomplete Beta function integral is well-known, and is given in the following Proposition (see Abramowitz and Stegun (1972), Equation 6.6.8, together with Gauss's transformation formula, 15.3.3):

Proposition 4 *For* $b \sim Beta(s, t)$*, and* $0 < \varepsilon < 1$*,*

$$\Pr\{b < \varepsilon\} = \frac{\Gamma(s+t)\varepsilon^s(1-\varepsilon)^t}{\Gamma(t)\Gamma(s+1)} {}_2F_1(s+t, 1, s+1; \varepsilon). \tag{40}$$

It can easily be checked that $1 - zh(0) = 0$ and $1 - zh(\lambda_1) = 1$, which imply that $F_{u_1}(z|c, \lambda) = 0$ and $F_{u_1}(z + \lambda_1|c, \lambda) = 1$. Thus, combining Lemma 3, Theorem 4, and Proposition 4, we may state:

Theorem 5 *The conditional cdf of the likelihood ratio statistic* T*, given* $(c_1, ..., c_{m-1})$ *and* $(\lambda_1, ..., \lambda_m)$*, is given by:*

$$\Pr\{T < z|c_1, ..., c_{m-1}; \lambda_1, ..., \lambda_m) = G_k(z + \lambda_1) - \int_z^{z+\lambda_1} g_k(v)F_{u_1}(v|c, \lambda)dv, \tag{41}$$

and so

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) = G_k(z + \lambda_1) - E_c\left[\int_z^{z+\lambda_1} g_k(v)F_{u_1}(v|c, \lambda)dv\right], \tag{42}$$

with

$$\begin{aligned} F_{u_1}(v|c, \lambda) \quad &= \quad \frac{\Gamma(\frac{k}{2})(1 - zh(v-z))^{\frac{k-m}{2}}(zh(v-z))^{\frac{m}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{k-m+2}{2})} \\ &\quad \times {}_2F_1\left(\frac{k}{2}, 1, \frac{k-m+2}{2}; 1 - zh(v-z)\right). \end{aligned} \tag{43}$$

Remark 2 *The function $F_{u_1}(v|c, \lambda)$ that appears in this expression is an explicit function of the random variables involved, $c_1, ..., c_{m-1}$. This contrasts with the earlier cases where both f_1 and e_1 were, in general, only implicitly defined. Thus, after this second step in the iterative process, no further complications arise from the fact that we were initially dealing with a characteristic root.*

Remark 3 *When $\lambda_i = \lambda$ for all i then*

$$F_{u_1}(v|c, \lambda) = \Pr\left\{b < \frac{(v-z)(z+\lambda)}{\lambda v}\right\}$$

does not depend on the c_i. Thus,

$$P_{k,m}(z; \lambda, ..., \lambda) = G_k(z+\lambda) - \int_z^{z+\lambda} g_k(v) B_{\frac{1}{2}(k-m), \frac{1}{2}m}\left(\frac{(v-z)(z+\lambda)}{\lambda v}\right) dv,$$

where $B_{s,t}(\cdot)$ denotes the cdf of the Beta(s,t) distribution. The result is the function $Q_{k,m}(z; \lambda)$ given in equation (28) above.

In principle the integral remaining in equation (42) can be evaluated analytically, but that is likely to be a formidable task for general m. Alternatively, it can be evaluated by numerical integration, or by a simulation approach which we discuss in the next Section. In the case $m = 2$, however, it turns out that the *cdf* $F_{u_1}(v|c, \lambda)$ is particularly simple, and only one variable $c = c_1$ is involved. This permits a fairly simple analytical evaluation of the result, which we turn to next. Before doing so we note the following bounds on $P_{k,m}(z; \lambda_1, ..., \lambda_m)$ that follow from the results above:

Proposition 5 *For all $z > 0$, the cdf $P_{k,m}(z; \lambda_1, ..., \lambda_m)$ is bounded below by*

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) \geq Q_{k,m}(z; \lambda_1). \tag{44}$$

where $Q_{k,m}(z; \lambda_1$ is as defined in Equation (28) with $\lambda = \lambda_1$ and $a = \lambda_1/(z+\lambda_1)$, and $b \sim$ Beta $\left(\frac{k-m}{2}, \frac{m}{2}\right)$, and, since $G_k(\cdot)$ is non-decreasing and $0 < e_1 < \lambda_1$, is bounded above by $G_k(z+\lambda_1)$.

Remark 4 *The bounds on the cdf,*

$$Q_{k,m}(z; \lambda_1) \leq P_{k,m}(z; \lambda_1, ..., \lambda_m) \leq G_k(z+\lambda_1) \tag{45}$$

may be enough to provide the inference required : for an observed value, t, of T, the conditional $p-$value that can be assigned to t under the null hypothesis is between $1 - G_k(t+\lambda_1)$ and $1 - Q_{k,m}(t; \lambda_1)$. Thus, if the nominal (conditional) size of the test sought for H_0 is α, H_0 can be rejected without further calculation if $Q_{k,m}(t; \lambda_1) > 1-\alpha$, and can be accepted if $G_k(t+\lambda_1) < 1-\alpha$. These bounding $p-$values are quite easy to compute (see Section 5 below for discussion of computational issues). More precise calculations are really needed only for intermediate cases.

In Figure 3 we give some examples of these bounds for the case $m = 3$, various values of k, and, in each case, a small and large (relative to k) value of λ_1. It can be seen that, in each case, the bounds are tighter (for given k and m) when λ_1 is small, and the $p-$values implied by the bounds will be considerably smaller for larger values of λ_1. Thus, the indecisive region (i.e., the region between the curves) based on these bounds will, for a given size of test, be quite small when λ_1 is small, but can be substantial when λ_1 is larger.

4.2 Analytic evaluation for the case $m = 2$

When $m = 2$ there is only one variable $c_1 = c$, and $c \sim Beta(\frac{1}{2}, \frac{1}{2})$. And, using the above result for the incomplete Beta function with $m = 2$, we obtain the remarkably simple result that, when $m = 2$,

$$F_{u_1}(v|c,\lambda) = (1 - zh(v - z))^{\frac{k-2}{2}}, \;\; z < v < z + \lambda_1. \tag{46}$$

Thus

$$\begin{aligned} \Pr\{T &< z|c, \lambda_1, \lambda_2\} = G_k(z + \lambda_1) - \int_z^{z+\lambda_1} g_k(v)(1 - zh(v - z))^{\frac{k-2}{2}} dv \\ &= G_k(z + \lambda_1) - \int_z^{z+\lambda_1} \frac{\exp\{-\frac{1}{2}v\}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} [v(1 - zh(v - z))]^{\frac{k-2}{2}} dv. \end{aligned} \tag{47}$$

Before considering the general case, the following result for the special case in which $\lambda_1 = \lambda_2 = \lambda$ is easily obtained:

Proposition 6 *For $m = 2$, and in the special case where $\lambda_1 = \lambda_2 = \lambda$,*

$$P_{k,2}(z; \lambda, \lambda) = G_k(z + \lambda) - a^{-\frac{k-2}{2}} \exp\{-\frac{1}{2}z\} G_k(\lambda), \tag{48}$$

where $a = \lambda/(z + \lambda)$.

This result for the case $m = 2$ is slightly simpler than the expression given in equation (28) for general m, but the two versions can easily be shown to agree. In the general case with $m = 2$ and $\lambda_1 < \lambda_2$, the lower bound given in Proposition 5 has the form:

$$P_{k,2}(z; \lambda_1, \lambda_2) \geq G_k(z + \lambda_1) - a_1^{-\frac{k-2}{2}} \exp\{-\frac{1}{2}z\} G_k(\lambda_1) \tag{49}$$

where $a_1 = \lambda_1/(z + \lambda_1)$. It is easy to check that the lower bound on $P_{k,2}(z; \lambda_1, \lambda_2)$ approaches $G_2(z)$ for all k as $\lambda_1 \to \infty$ (cf. Hillier (2006), Section 4).

In the general case the following explicit form for the conditional *cdf*

$$P_{k,2}(z; \lambda_1, \lambda_2) = E_c \left[\Pr\{T < z|c; \lambda_1, \lambda_2\}\right] \tag{50}$$

with $c \sim Beta(\frac{1}{2}, \frac{1}{2})$, is derived in the Appendix:

Theorem 6 *In the case $m = 2$ the conditional cdf of T given (λ_1, λ_2) (or, equivalently, given W_{22}) is given by:*

$$\Pr\{T < z|\lambda_1, \lambda_2\} = G_k(z + \lambda_1) - (a_1 a_2)^{-\frac{k-2}{2}} \exp\{-\frac{1}{2}z\} \times \sum_{j,l=0}^{\infty} \frac{(-\frac{k-2}{2})_j(\frac{k-2}{2})_l}{j!l!} H_{j,l}(z; \lambda_1) W_{j,l}(z; \lambda_1, \lambda_2), \quad (51)$$

where $a_i = \lambda_i/(z + \lambda_i)$, $i = 1, 2$,

$$H_{j,l}(z; \lambda_1) = \left(\frac{z}{z + \lambda_1}\right)^j \left(\frac{2}{\lambda_1}\right)^l \sum_{s=0}^{j} \binom{j}{s} \left(\frac{2}{z}\right)^s \left(\frac{k}{2}\right)_{s+l} G_{k+2(s+l)}(\lambda_1), \quad (52)$$

and

$$W_{j,l}(z; \lambda_1, \lambda_2) = \sum_{s=0}^{j} \sum_{t=0}^{l} \binom{j}{s} \binom{l}{t} \frac{(-1)^{s+t}(\frac{1}{2})_{s+t}}{(1)_{s+t}} \left(\frac{\Delta}{z + \lambda_2}\right)^s \left(\frac{\Delta}{\lambda_2}\right)^t, \quad (53)$$

where $\Delta = \lambda_2 - \lambda_1$.

In principle this result can be used to produce, for each combination of values of k, λ_1, and λ_2, a set of tables, or graphs, of the critical values, $z_\alpha(k; \lambda_1, \lambda_2)$ say, needed to give a test of size exactly α. However, such tables would obviously be extremely cumbersome, so the result is probably more useful as a means of computing the $p-$value associated with the given values $(k, \lambda_1, \lambda_2)$, and the observed value, t, of the statistic T.

In Figure 4 we show (for the upper tail of the distribution) the bounds in equation (45) (shown as the solid and dashed lines), together with the exact *cdf* computed using the formula in Theorem 6 (crosses), for the case $k = 18$, $\lambda_1 = 18$, $\lambda_2 = 20$. It turns out that the exact *cdf* is sensitive to the value of λ_1, but fairly insensitive to the value of λ_2. Even the detailed calculations needed to plot the functions shown here take only a few seconds, so the exact formula is extremely efficient as a means of computing the single $p-$value needed to implement the test when $m = 2$.

5 Implementation and Computation of $p-$values

5.1 Implementation - Preliminaries

From the point of view of a practitioner two issues need to be addressed before the results presented above can be implemented. These are: the fact that the model in equations (1) - (2) is overly simplistic, and the fact that in practice the covariance matrix Ω will be unknown.

On the first point, the structural/IV equation (1) would typically involve an additional matrix of exogenous variables, X, of dimension $n \times k_1$, say, which would also appear in equation (2). Thus, a more realistic model would be:

$$\begin{aligned} y_1 &= Y_2\beta + X\gamma + u, & \text{(1a)} \\ Y_2 &= Z\Pi + X\Psi + V. & \text{(2a)} \end{aligned}$$

The only effects of this extension of the model on the formulae for the various statistics that have been introduced above are these: the vectors and matrices y_1, Y_2, and Z should be interpreted as the residuals after the original variables (with these names) are regressed on X. That is, if we denote, for any matrix A of full column rank, the idempotent matrix $I - A(A'A)^{-1}A'$ by M_A, (y_1, Y_2, Z) are actually $(M_X y_1, M_X Y_2, M_X Z)$.

Next, the assumption that Ω is known is, of course, unrealistic. In practice the matrix Ω needs to be replaced by an estimator, and there are various possible estimators that could be used. Probably the simplest choice is to use the unconstrained estimator in the reduced form, which, when we use the extended model (1a) - (2a), is a multiple of the matrix

$$S = (y_1, Y_2)' M_{Z,X} (y_1, Y_2). \tag{54}$$

The estimator $\hat{\Omega} = (n - k - k_1)^{-1} S$ is easily seen to be unbiased.

The matrix U_{Ω} may be defined in terms of $\hat{\Omega}$ exactly as U_{Ω} was in terms of Ω. However, if (in contrast to the text above) we do not assume that variables have previously been scaled by their standard deviations, then U_{Ω} has the form

$$U_{\Omega} = \begin{pmatrix} \frac{1}{\sqrt{\omega_{11}}} & -\frac{\omega_{21}' \Omega_{22.1}^{-\frac{1}{2}}}{\sqrt{\omega_{11}}} \\ 0 & \Omega_{22.1}^{-\frac{1}{2}} \end{pmatrix}. \tag{55}$$

Here, $\Omega_{22.1} = \Omega_{22} - \omega_{21}\omega_{21}'/\omega_{11}$, and $\Omega_{22.1}^{-\frac{1}{2}}$ is any matrix satisfying $\Omega_{22.1}^{-\frac{1}{2}} \Omega_{22.1} \Omega_{22.1}^{-\frac{1}{2}} = I_m$.

After both of these modifications, the matrix W in equation (6) will in practice be:

$$W = U_{\hat{\Omega}}' (y_1, Y_2)' M_X Z (Z' M_X Z)^{-1} Z' M_X (y_1, Y_2) U_{\hat{\Omega}}. \tag{56}$$

The key variables q and f_1 that define the LR statistic, as well as the other variables introduced above, are defined in terms of this matrix[3]. Here, if the original sample size were, say, N, then n in the earlier part of the paper corresponds to $N - k_1$, the rank of M_X.

Of course, transformation by $U_{\hat{\Omega}}$ does not produce the independence properties that are used to obtain the main results in the paper, nor do the distributional results employed in the argument hold exactly when Ω is replaced by an estimator. Nevertheless, the simulation results in Moreira (2003) suggest that, at least in moderate sized samples, the above procedure should produce a test which is very close to being exact. In a separate, as yet unfinished, paper I am studying this aspect of the testing problem analytically, rather than by simulations, and will report those results elsewhere when they are available.

5.2 Implementation - Computation

It remains to consider the computational aspects of implementing the results. Several methods are suggested by the results in the body of the paper, and we discuss these in turn.

First, according to Remark 4, the conditional $p-$value associated with an observed value t of T, given the roots $(\lambda_1, ..., \lambda_m)$, $1 - P_{k,m}(t; \lambda_1, ..., \lambda_m)$, satisfies

$$1 - G_k(t + \lambda_1) \leq 1 - P_{k,m}(t; \lambda_1, ..., \lambda_m) \leq 1 - Q_{k,m}(t; \lambda_1), \tag{57}$$

and it is natural to begin by checking whether this, by itself, implies that H_0 can be either accepted or rejected. This would be so if either $G_k(t + \lambda_1) < 1 - \alpha$ (which would imply acceptance of H_0, because the conditional $p-$value must then be greater than the nominal size, α), or if $Q_{k,m}(t; \lambda_1) > 1 - \alpha$ (which would imply rejection of H_0, because the conditional $p-$value must then be less than α). We need to discuss the evaluation of $Q_{k,m}(t; \lambda_1)$ for this part of the procedure.

Three methods for computing $Q_{k,m}(t; \lambda_1)$ are available. First, we have the exact, infinite series, representation of the function in equation (28). Second, we have a representation of the function as the expectation of $G_k(z(1 - a_1 b)^{-1})$ with respect to $b \sim Beta(\frac{k-m}{2}, \frac{m}{2})$ given in equation (27). And, third, we may observe that ($cf.$ Hillier (2006), Kleibergen (2007)) $Q_{k,m}(z; \lambda_1) = \Pr\{T^* < z|\lambda_1\}$, where T^* is the statistic

$$T^* = \frac{1}{2}\left\{q - \lambda_1 + \sqrt{(q + \lambda_1)^2 - 4qb\lambda_1}\right\}, \tag{58}$$

q and b are independent, $q \sim \chi^2(k)$, and $b \sim Beta\left(\frac{k-m}{2}, \frac{m}{2}\right)$, and λ_1 is fixed. These yield the following possible computational procedures, respectively:

1. Truncate the infinite series at a point designed to produce the required accuracy,

2. Estimate the mean $E_b[G_k(z(1 - a_1 b)^{-1})]$ by simulation, using the mean of a sequence of values $G_k(z(1-a_1 b_i)^{-1})$, with the b_i $i.i.d$ draws from the distribution of b, and

3. Estimate $\Pr\{T^* < z|\lambda_1\}$ directly by simulation, using $i.i.d.$ draws from the joint distribution of (q, b) to compute values of T^* (this is essentially the method suggested by Kleibergen (2007) in discussing the statistic T^*).

We show in the Appendix that the third of these methods is unequivocally inferior to the second. Of the other two, the first method is more efficient than the second, in the sense that far fewer terms in the infinite series are needed, relative to the size of the simulation sample, to achieve a given level of accuracy. For example, one can show (see Appendix) that the error incurred by truncating the series in equation (28) at the $r-th$ term is no greater than $a_1^{r+1}(\frac{m}{2})_{r+1}/(r+1)!$. So, for $m=3$ and $a_1 = \lambda_1/(t+\lambda_1) = .5$, say, less than 15 terms are needed to give third-figure accuracy - much lower than the sample size required to simulate the expected value to this accuracy (in a confidence interval sense). Thus, of these three methods for computing $Q_{k,m}(t;\lambda_1)$, direct evaluation is preferred, and is extremely quick and easy.

Now if, after evaluating both $G_k(t+\lambda_1)$ and $Q_{k,m}(t;\lambda_1)$, it turns out that no decision can be made, a more precise calculation of the $p-$value $1-P_{k,m}(t;\lambda_1,...,\lambda_m)$ is needed. For $m=2$ this can be done using the exact formula for $P_{k,m}(t;\lambda_1,...,\lambda_m)$ given in Theorem 6, which, like $Q_{k,m}(t;\lambda_1)$, is extremely efficient. For $m>2$, however, we do not have an equivalent analytical expression, so in this case the $p-$value must be evaluated by simulation.[4] We now describe this procedure in more detail.

The probability $\Pr\{T<t|\lambda_1,...,\lambda_m\}$, where t is the observed value of T, is given by the expectation

$$\Pr\{T<t|\lambda_1,...,\lambda_m\} = E_{u_1}[G_k(u_1)], \tag{59}$$

where $u_1 = t+e_1$, and e_1 is the smallest root of the $m-$th degree polynomial:

$$\tau(e) = t\prod_{i=1}^{m}(e-\lambda_i) + (z+e)(1-b)\sum_{i=1}^{m} c_i\lambda_i\left[\prod_{j\neq i}(e-\lambda_j)\right]. \tag{60}$$

In this formula for $\tau(e)$ the notation is as follows: (i) The λ_i are the characteristic roots of W_{22}. These remain fixed at their observed values. (ii) b is a random variable distributed as $Beta(\frac{k-m}{2},\frac{m}{2})$, and (iii) $c_1,...,c_m$ are random variables independent of b (and the λ_i); $c_m = 1-\Sigma_{i=1}^{m-1}c_i$, and the remaining c_i are defined as follows. Let g_i, $i=1,...,m$, be m independent $\chi^2(1)$ random variables. For $i=1,...,m-1$, define c_i by

$$c_i = \frac{g_i}{\Sigma_{i=1}^{m} g_i}, \tag{61}$$

and for $i=m$ set $c_m = 1-\Sigma_{i=1}^{m-1}c_i$. The c_i constructed by this procedure have a Dirichlet distribution with parameters $(\frac{1}{2},...,\frac{1}{2};\frac{1}{2})$.

The value of e_1 is a random variable because it is a function of b and $c_1,...,c_{m-1}$, and hence so also is the value of $G_k(t+e_1)$. But, what we require is simply the mean of $G_k(t+e_1)$, and this is easy to simulate with relatively few repetitions. Thus, we take, say, M independent draws from the distributions of b and $c_1,...,c_{m-1}$. Each of these yields a polynomial $\tau_i(e)$, say, and from each we compute the smallest root, e_{1i}, of the equation $\tau_i(e)=0$. From these we obtain a sequence of values $G_k(t+e_{1i})$, and the estimate of the required expectation is simply

$$p(t;\lambda_1,...,\lambda_m) = \frac{1}{M}\sum_{i=1}^{M} G_k(t+e_{1i}). \tag{62}$$

The hypothesis is rejected at size α if $p(t; \lambda_1, ..., \lambda_m) > 1-\alpha$, and accepted otherwise. The chief difficulty is the need to compute, at each repetition, the smallest root of the degree-m polynomial equation $\tau(e) = 0$, but as long as m is not large - and it is unlikely to be in practice - this is not a major problem.

6 Concluding Comments

The ability to compute conditional $p-$values efficiently evidently provides a complete solution to the problem of obtaining an exact test for the problem we have considered. However, for completeness it is worth mentioning that the critical value function required to implement the CLR test with exact size α is the solution, $z_\alpha^m(k; \lambda_1, ..., \lambda_m)$ say, to the equation:

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) = 1 - \alpha. \tag{63}$$

In general this equation cannot be solved explicitly, and in any case its tabulation would be an extremely cumbersome affair, because one needs a critical value for each possible value of the conditioning vector $(\lambda_1, ..., \lambda_m)$. Hillier (2006) gives some graphs of the required function for the case $m = 1$, where it is a function of a single characteristic root, but this is the only case where graphical tabulation is reasonably easy. Nevertheless, figures such as those presented here can be used to obtain the value needed - for the observed values of $(\lambda_1, .., \lambda_m)$ - reasonably easily, and do provide an alternative to using $p-$values.

The results in this paper relate entirely to the conditional *size* of the LR test. However, the decomposition of the characteristic polynomial given in equation (12) can also be used to generalize these results in two directions. First, one can use similar arguments to those used here to obtain expressions for the power function of the CLR test. The main difficulty is that the variates involved are no longer independent under the alternative hypothesis, nor are their distribution properties so simple. For instance, at the first step in the process, the variate q is no longer independent of the remaining variates, and distributed as $\chi^2(k)$. Its conditional distribution given the remaining variates is non-central $\chi^2(k)$, with noncentrality parameter that is a function of the other variates. Nevertheless, the conditional distribution of T can be obtained, and subsequent steps carried out in a manner similar to those above, but with much more complex results. Calculations of this type are carried out for the power function in the case $m = 1$ in Hillier (2006). Here, analytic results can be obtained, but for $m > 1$ a simulation approach after the first step is almost certainly essential.

A second direction of generalization based on (12) is to the more realistic case in which the error covariance matrix is unknown. Here, too, the transformations we have used to produce the independence properties that the argument depends on are unavailable, nor are the distributions involved so simple. Again, though, initial conditioning can be productive, and the argument used here can be generalized.

These matters are the subject of ongoing research.

Notes

[1]Throughout the paper, whenever we use the term monotonic we mean strictly monotonic.

[2]There is a large literature on the *unconditional* distribution theory for the characteristic roots of a Wishart matrix (the case we are dealing with). See, for instance, Muirhead (1982) for some of this theory, and many references. However, the methods used in that theory depend upon the spectral decomposition for positive definite symmetric matrices, and hence cannot be used to develop the *conditional* properties that we are concerned with here.

[3]The statistic T produced by this procedure is not, in fact, the likelihood ratio statistic in a model with an unknown covariance matrix. However, the two statistics will coincide in large samples.

[4]An alternative would be to use numerical integration in equation (42). I have not compared the efficiency of this approach with that of a direct simulation approach.

7 Appendix: Proofs.

Proof of Lemma 1.

First observe that $\psi_2(0) = b[\Pi_{i=1}^m \lambda_i](-1)^m$ and $\psi_1(0) = (-1)^m[\Pi_{i=1}^m \lambda_i]$, so that $sign[\psi_2(0)] = sign[\psi_1(0)] = (-1)^m$. Next, consider

$$\psi_2(\lambda_r) = (1-b)c_r\lambda_r \left(\prod_{i \neq r} (\lambda_r - \lambda_i) \right). \tag{64}$$

Clearly, $sign[\psi_2(\lambda_r)]$ is determined by the sign of the last term. It is $(-1)^{m-1}$ for $r = 1$, and alternates with r because the products $\prod_{i \neq r}(\lambda_r - \lambda_i)$ contain odd and even numbers of negative terms. Since $\psi_2(f)$ is continuous in f, these properties imply that there is precisely one real root of $\psi_2(f) = 0$ in each interval L_i, establishing part (i) of the Lemma.

Now, $\psi_2(f)$ changes sign at the f_{2i}, and $\psi_1(f)$ at the λ_i, which we know from part (i) satisfy $\lambda_i > f_{2i}$ (the inequality is strict with probability 1). On the other hand, the roots f_i of $\psi(f) = 0$ occur where the functions $f\psi_1(f)$ and $q\psi_2(f)$ cross. Consider the interval L_1. Since $sign[\psi_2(\lambda_1)] = (-1)^{m-1}$, but sign $\psi_1(f) = (-1)^m$ for $f \leq \lambda_1$, ψ_1 and ψ_2 have opposite signs on the interval $f_{21} < f < \lambda_1$, so $f\psi_1$ and $q\psi_2$ cannot cross there. Hence, if $f\psi_1(f)$ and $q\psi_2(f)$ cross at all in the interval L_1, and we know that they do, they must do so for $f < f_{21}$. This argument can be repeated for each interval L_i, proving part (ii) of the Lemma. Part (iii) follows from the fact that the signs of ψ_1 and ψ_2 are the same on the sub-intervals $\lambda_{i-1} < f_i < f_{2i}$, $i = 1, ..., m$, and ψ_2 does not vanish there.

Proof of Lemma 2.

Let

$$g(f) = \sum_{i=1}^m \frac{c_i\lambda_i}{\lambda_i - f},$$

so that $\psi_2(f) = \psi_1(f)\{1 - (1-b)g(f)\}$. Consider first

$$T(f) = \frac{f\psi_1(f)}{\psi_2(f)} - f = \frac{(1-b)fg(f)}{1-(1-b)g(f)}.$$

Then,

$$T'(f) = \frac{(1-b)[g(f) + fg'(f) - (1-b)g(f)^2]}{(1-(1-b)g(f))^2}.$$

But,

$$\begin{aligned} g(f) + fg'(f) - (1-b)g(f)^2 &= \sum_{i=1}^{m} \frac{c_i\lambda_i^2}{(\lambda_i - f)^2} - (1-b)\left[\sum_{i=1}^{m} \frac{c_i\lambda_i}{(\lambda_i - f)}\right]^2 \\ &> \sum_{i=1}^{m} \frac{c_i\lambda_i^2}{(\lambda_i - f)^2} - \left[\sum_{i=1}^{m} \frac{c_i\lambda_i}{(\lambda_i - f)}\right]^2 \\ &= \sum_{i=1}^{m} c_i \left[\frac{\lambda_i}{\lambda_i - f} - \sum_{i=1}^{m} \frac{c_i\lambda_i}{(\lambda_i - f)}\right]^2 \geq 0, \end{aligned}$$

the third line following since $\Sigma_{i=1}^{m} c_i = 1$. This establishes that $T'(f) > 0$. Since $q(f) = T(f) + f$, this in turn implies $q'(f) > 0$. Note that, if the λ_i are all equal to λ, the first line here reduces to $b\lambda^2/(\lambda - f)^2 > 0$, so the property also holds in that case. This can also be shown more directly using the results in Hillier (2006), Section 4.1.

Proof of Proposition 3.

From Theorem 1, for any $z > 0$ the inequality $T(f_1) < z$ corresponds to the inequality $f_1 < T^{-1}(z) = e_1$, with $e_1 \in R_1$, and this also corresponds to the inequality $q(f_1) < q(e_1)$. But $q(e_1) = T(e_1) + e_1 = z + e_1$, which establishes the result. The equation $T(e) = z$ is easily seen to correspond to the equation $\tau(e) = 0$. To see that e_1 is the smallest root of this equation, first observe that $\tau(0) = zb[\Pi_{i=1}^{m}\lambda_1](-1)^m$. But also, $\tau(f_{21}) = -f_{21}\psi_1(f_{21})$, and we know from Lemma 1 that $sign[\psi_1(f_{21})] = (-1)^m$, so that $sign[\tau(f_{21})]$ is opposite to that of $\tau(0)$. Hence, $\tau(\cdot)$ changes sign between the origin and f_{21}, which implies that $e_1 \in R_1$.

Proof of Lemma 3.

The expectation required is:

$$\Pr\{T < z|c, \lambda) = \int_{\underline{u}_1}^{\bar{u}_1} p_{u_1}(v|c, \lambda)G_k(v)dv. \tag{65}$$

Integrating by parts we obtain

$$\Pr\{T < z|c, \lambda) = G_k(v)F_{u_1}(v|c, \lambda)|_{\underline{u}_1}^{\bar{u}_1} - \int_{\underline{u}_1}^{\bar{u}_1} F_{u_1}(v|c, \lambda)g_k(v)dv. \tag{66}$$

This yields the formula given if $F_{u_1}(\underline{u}_1|c, \lambda) = 0$ and $F_{u_1}(\bar{u}_1|c, \lambda) = 1$.

Proof of Lemma 4.

The details are almost identical to those for the proof of Lemma 1. It is first established that $sign[\tau(0)] = (-1)^m$, and that $sign[\tau(\lambda_r)] = (-1)^{m-r}$, $r = 1, ..., m$.

Continuity then implies that there is a real root in each interval $[\lambda_{i-1}, \lambda_i]$, $i = 1, ..., m$. To see that a given root can be arbitrarily close to the upper limit of its respective interval, simply observe that as $b \to 1$, $\tau(e) \to z\Pi_{i=1}^m (e - \lambda_i)$, the roots of which are the λ_i.

Proof of Proposition 5.

The term $\sum_{i=1}^m \frac{c_i\lambda_i}{\lambda_i - e}$ that arises in the expression for $h(e)$ is a convex combination of the $\lambda_i/(\lambda_i - e)$, and therefore satisfies, for $0 < e < \lambda_1$,

$$\lambda_m/(\lambda_m - e) \leq \sum_{i=1}^m \frac{c_i\lambda_i}{\lambda_i - e} \leq \lambda_1/(\lambda_1 - e)$$

for all c. Therefore, for all c, and all $0 < e < \lambda_1$,

$$1 - zh(e) \leq 1 - \frac{z(\lambda_1 - e)}{\lambda_1(z + e)} = \frac{e(z + \lambda_1)}{\lambda_1(z + e)}.$$

But, from this it also follows that, for all c, and v satisfying $z < v < z + \lambda_1$,

$$F_{u_1}(v|c, \lambda) = \Pr\{b < 1 - zh(v - z)|c, \lambda\} \leq \Pr\{b < \frac{e(z + \lambda_1)}{\lambda_1(z + e)}|\lambda\}, \tag{67}$$

and so

$$P_{k,m}(z; \lambda_1, ..., \lambda_m) \geq G_k(z+\lambda_1) - \int_z^{z+\lambda_1} g_k(v) B_{\frac{1}{2}(k-m), \frac{1}{2}m} \left(\frac{(v - z)(z + \lambda_1)}{\lambda_1 v} \right) dv. \tag{68}$$

But the right-hand-side is $Q_{k,m}(z; \lambda_1)$ in equation (28) with $\lambda = \lambda_1$.

Proof of Proposition 6.

When $\lambda_1 = \lambda_2 = \lambda$ we have

$$v(1 - zh(v - z)) = x/a, \tag{69}$$

where $x = v - z$ and $a = \lambda/(z + \lambda)$. Since this does not depend on c, we have at once:

$$P_{k,2}(z; \lambda_1, \lambda_2) = G_k(z + \lambda) - a^{-\frac{k-2}{2}} \exp\{-\frac{1}{2}z\} \int_0^\lambda g_k(x)dx, \tag{70}$$

which is the stated result.

Proof of Theorem 6.

Setting $x = v - z$ $(0 < x < \lambda_1)$ we have, when $m = 2$,

$$h(v - z) = \frac{(\lambda_1 - x)(\lambda_2 - x)}{(x + z)[c\lambda_1(\lambda_2 - x) + (1 - c)\lambda_2(\lambda_1 - x)]}. \tag{71}$$

Thus, in terms of x,

$$\begin{aligned} v(1-zh(v-z)) &= \left[x+z-\frac{z(\lambda_1-x)(\lambda_2-x)}{[\lambda_1\lambda_2-c\lambda_1 x-(1-c)\lambda_2 x]}\right] \\ &= x\left[1+\frac{z(\lambda_1-x+c\Delta)}{\lambda_2(\lambda_1-x)+c\Delta x)}\right] \\ &= x\left[\frac{(z+\lambda_2)(\lambda_1-x)+c\Delta(z+x)}{\lambda_2(\lambda_1-x)+c\Delta x}\right] \\ &= \frac{x}{a_1a_2}\left[\frac{1-\frac{z+x}{z+\lambda_1}\left(1-\frac{c\Delta}{z+\lambda_2}\right)}{1-\frac{x}{\lambda_1}(1-\frac{c\Delta}{\lambda_2})}\right], \end{aligned} \tag{72}$$

where $\Delta=\lambda_2-\lambda_1$ and $a_i=\lambda_i/(z+\lambda_i)$, $i=1,2$. This obviously reduces to x/a, with $a=\lambda/(z+\lambda)$, when $\lambda_1=\lambda_2=\lambda$ (so $\Delta=0$), which then produces Proposition 6.

We thus have, for the integral

$$\int_z^{z+\lambda_1} g_k(v)(1-zh(v-z))^{\frac{k-2}{2}}dv,$$

the expression

$$(a_1a_2)^{-\frac{k-2}{2}}\exp\{-\frac{1}{2}z\}\int_0^{\lambda_1} g_k(x)\left[\frac{1-\frac{z+x}{z+\lambda_1}\left(1-\frac{c\Delta}{z+\lambda_2}\right)}{1-\frac{x}{\lambda_1}(1-\frac{c\Delta}{\lambda_2})}\right]^{\frac{k-2}{2}} dx. \tag{73}$$

The term $\left[1+\frac{z(\lambda_1-x+c\Delta)}{\lambda_2(\lambda_1-x)+c\Delta x)}\right]^{\frac{k-2}{2}}$ that arises in the second line of the development above is, for all $c\in(0,1)$, bounded above by $a_1^{-\frac{k-2}{2}}$ on the interval $0<x<\lambda_1$. This implies the lower bound for $P_{k,2}(z;\lambda_1,\lambda_2)$ given in Equation (49):

$$P_{k,2}(z;\lambda_1,\lambda_2)\geq G_k(z+\lambda_1)-a_1^{-\frac{k-2}{2}}\exp\{-\frac{1}{2}z\}G_k(\lambda_1). \tag{74}$$

It follows that the conditional $p-$value $\Pr\{T>t|\lambda_1,\lambda_2\}$ is bounded above by:

$$1-G_k(t+\lambda_1)+a_1^{-\frac{k-2}{2}}\exp\{-\frac{1}{2}t\}G_k(\lambda_1), \tag{75}$$

where $a_1=\lambda_1/(t+\lambda_1)$. Rejecting H_0 when this is less than α clearly corresponds to using a test that is conservative at level α.

In the general case, note that, on the region of integration $0<x<\lambda_1$, $0<c<1$, both

$$0<\frac{x}{\lambda_1}\left(1-\frac{c\Delta}{\lambda_2}\right)<1,$$

and

$$0 < \frac{z+x}{z+\lambda_1}\left(1-\frac{c\Delta}{z+\lambda_2}\right) < 1,$$

so we may expand both numerator and denominator terms in the expression

$$\left[\frac{1-\frac{z+x}{z+\lambda_1}\left(1-\frac{c\Delta}{z+\lambda_2}\right)}{1-\frac{x}{\lambda_1}(1-\frac{c\Delta}{\lambda_2})}\right]^{\frac{k-2}{2}}$$

in power series, both converging uniformly on the region of integration. This gives

$$\sum_{j,l=0}^{\infty}\frac{(-\frac{k-2}{2})_j(\frac{k-2}{2})_l}{j!l!}\left(\frac{z+x}{z+\lambda_1}\right)^j\left(\frac{x}{\lambda_1}\right)^l\left(1-\frac{c\Delta}{z+\lambda_2}\right)^j\left(1-\frac{c\Delta}{\lambda_2}\right)^l \tag{76}$$

Note that the series indexed by j terminates if $k-2$ is an even integer.

Assuming, for the moment, that term-by-term integration with respect to c is legitimate we thereby obtain:

$$\sum_{j,l=0}^{\infty}\frac{(-\frac{k-2}{2})_j(\frac{k-2}{2})_l}{j!l!}\left(\frac{z+x}{z+\lambda_1}\right)^j\left(\frac{x}{\lambda_1}\right)^l W_{j,l}(\lambda_1,\lambda_2), \tag{77}$$

with

$$W_{j,l}(z;\lambda_1,\lambda_2) = E_c\left[\left(1-\frac{c\Delta}{z+\lambda_2}\right)^j\left(1-\frac{c\Delta}{\lambda_2}\right)^l\right],\ c\sim Beta(\frac{1}{2},\frac{1}{2}). \tag{78}$$

The expectation required in (78) is easily evaluated in the form:

$$W_{j,l}(z;\lambda_1,\lambda_2) = \sum_{s=0}^{j}\sum_{t=0}^{l}\binom{j}{s}\binom{l}{t}\frac{(-1)^{s+t}(\frac{1}{2})_{s+t}}{(1)_{s+t}}\left(\frac{\Delta}{z+\lambda_2}\right)^s\left(\frac{\Delta}{\lambda_2}\right)^t. \tag{79}$$

Finally, we need to evaluate also the integrals

$$H_{j,l}(z;\lambda_1) = \int_0^{\lambda_1} g_k(x)\left(\frac{z+x}{z+\lambda_1}\right)^j\left(\frac{x}{\lambda_1}\right)^l dx. \tag{80}$$

This is straightforward, and gives

$$H_{j,l}(z;\lambda_1) = \left(\frac{z}{z+\lambda_1}\right)^j\left(\frac{2}{\lambda_1}\right)^l\sum_{s=0}^{j}\binom{j}{s}\left(\frac{2}{z}\right)^s\left(\frac{k}{2}\right)_{s+l}G_{k+2(s+l)}(\lambda_1). \tag{81}$$

To confirm that these two term-by-term integrations of the multiple series are legitimate, first observe that, for all (j,l),

$$W_{j,l}(z;\lambda_1,\lambda_2) < \left(\frac{z+\lambda_1}{z+\lambda_2}\right)^j\left(\frac{\lambda_1}{\lambda_2}\right)^l. \tag{82}$$

Also, it is clear that $H_{j,l}(z;\lambda_1) < G_k(\lambda_1)$ for all (j,l) and (z,λ_1), so the resulting series is dominated termwise by

$$G_k(\lambda_1)\sum_{j,l=0}^{\infty}\frac{(-\frac{k-2}{2})_j(\frac{k-2}{2})_l}{j!l!}\left(\frac{z+\lambda_1}{z+\lambda_2}\right)^j\left(\frac{\lambda_1}{\lambda_2}\right)^l, \tag{83}$$

which certainly converges (to $G_k(\lambda_1)\left(\frac{\Delta}{z+\lambda_2}\right)^{\frac{k-2}{2}}\left(\frac{\Delta}{\lambda_2}\right)^{-\frac{k-2}{2}}$), justifying the term-by-term integrations.

We therefore have, finally,

$$\Pr\{T \; < \; z|\lambda_1,\lambda_2\} = G_k(z+\lambda_1) - (a_1a_2)^{-\frac{k-2}{2}}\exp\{-\frac{1}{2}z\} \tag{84}$$

$$\times\sum_{j,l=0}^{\infty}\frac{(-\frac{k-2}{2})_j(\frac{k-2}{2})_l}{j!l!}H_{j,l}(z;\lambda_1)W_{j,l}(\lambda_1,\lambda_2), \tag{85}$$

as stated.

Proofs of the results in Section 5

The second of the methods listed is a problem of estimating a mean, the third that of estimating a Binomial probability, and associated with each is a confidence interval whose width is determined by the relevant variance. But, it is easy to see that the variance involved for the second method, namely

$$Var\left[G_k(z(1-a_1b)^{-1})\right] = E_b\left[[G_k(z(1-a_1b)^{-1})]^2\right] - \left[E_b[G_k(z(1-a_1b)^{-1})]\right]^2, \tag{86}$$

is (very considerably) less than that involved for the third,

$$E_b[G_k(z(1-a_1b)^{-1})] - \left[E_b[G_k(z(1-a_1b)^{-1})]\right]^2, \tag{87}$$

simply because $E_b\left[[G_k(z(1-a_1b)^{-1})]^2\right]$ is very much less than $E_b[G_k(z(1-a_1b)^{-1})]$. Thus, the second of these methods is (very much) more efficient than the third, as claimed.

Finally, we shall show that, when using equation (28) to compute a $p-$value, trucation of the series component in equation (28) leads to an error in the $p-$value that is smaller than the coefficient of the first term in the series that is ignored. This effectively bounds the truncation error involved.

The remainder after truncating the series in equation (28) at the $r-$th term is

$$\begin{aligned} R_r &= \sum_{j=r+1}^{\infty} \frac{a^j (\frac{m}{2})_j}{j!} G_{k+2j}(z+\lambda) \\ &< \sum_{j=r+1}^{\infty} \frac{a^j (\frac{m}{2})_j}{j!} \\ &= \frac{a^{r+1}(\frac{m}{2})_{r+1}}{(r+1)!} {}_2F_1(1, r+1+\frac{m}{2}; r+2; a) \\ &= (1-a)^{-\frac{m}{2}} \frac{a^{r+1}(\frac{m}{2})_{r+1}}{(r+1)!} {}_2F_1(r+1, -(\frac{m-2}{2}); r+2; a). \end{aligned} \tag{88}$$

The term ${}_2F_1(r+1, -(\frac{m-2}{2}); r+2; a)$ is less than 1 on the interval $0 < a < 1$, so that the error in the value of the *cdf* due to series truncation is bounded above by $\frac{a^{n+1}(\frac{m}{2})_{n+1}}{(n+1)!}$, the coefficient in the first ignored term in the series.

References

Abramowitz, M., and I. Stegun, (1972), *Handbook of Mathematical Functions,* Dover, New York.

Andrews, D.W.K., M. Moreira, and J. Stock, (2006), "Optimal Two-Sided Invariant Similar tests for Instrumental Variables Regression", *Econometrica*, **74**, No. 3, 715-752.

Andrews, D.W.K., M. Moreira, and J. Stock, (2007), "Performance of conditional Wald tests in IV regression with weak instruments", *Journal of Econometrics*, **139**, 1, 116 - 132.

Courant, R. and D. Hilbert, (1953), *Methods of Mathematical Physics, Vol. 1.* Interscience, New York.

Hillier, G.H. (1987a), "Hypothesis Testing in a Structural Equation; Part I: Reduced Form Equivalences and Invariant Test Procedures", unpublished Manuscript, Monash University.

Hillier, G.H. (1987b), "Classes of Similar Regions and their Power Properties for some Econometric Testing Problems", *Econometric Theory*, **3**, 1-44.

Hillier, G.H. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica*, **58**, No. 5, 1181-1194.

Hillier, G.H. (2005), "Hypothesis Testing in a Structural Equation: Similar and Invariant Tests", mimeo, University of Southampton.

Hillier, G.H. (2006), "Exact properties of the conditional likelihood ratio test in an IV regression model", CeMMAP working paper #23/06, *Econometric Theory*, forthcoming.

Kleibergen, F. (2007), "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics", *Journal of Econometrics*, **139**, 1, 181 - 216.

Moreira, M., (2003), "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica*, **71,** 4, 1027-1048.

Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York.

Rainville, E.D. (1960), *Special Functions, New York:* Macmillan.

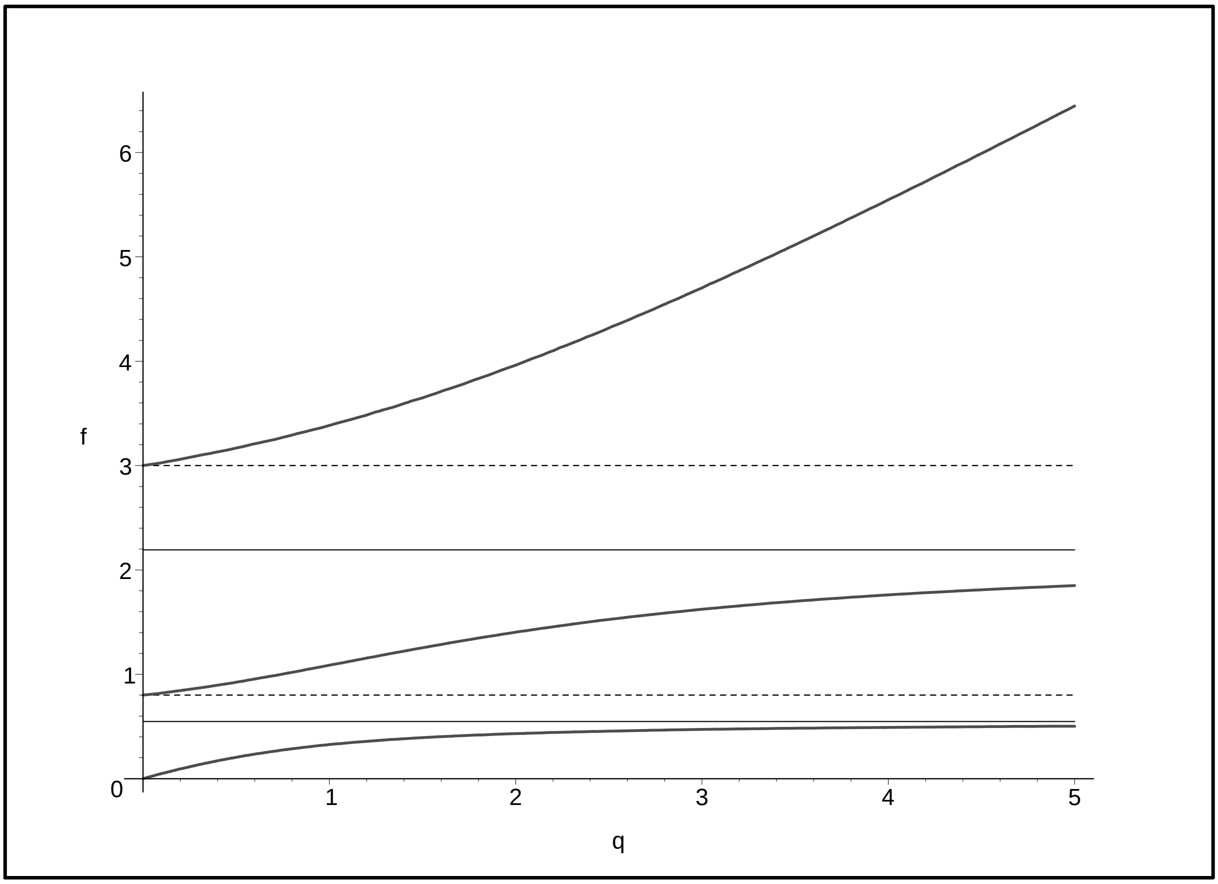


Figure 1: The three branches of the relation between f and q when $m = 2$. The dashed horizontal lines are λ_1 and λ_2, the solid horizontal lines are f_{21} and f_{22}.

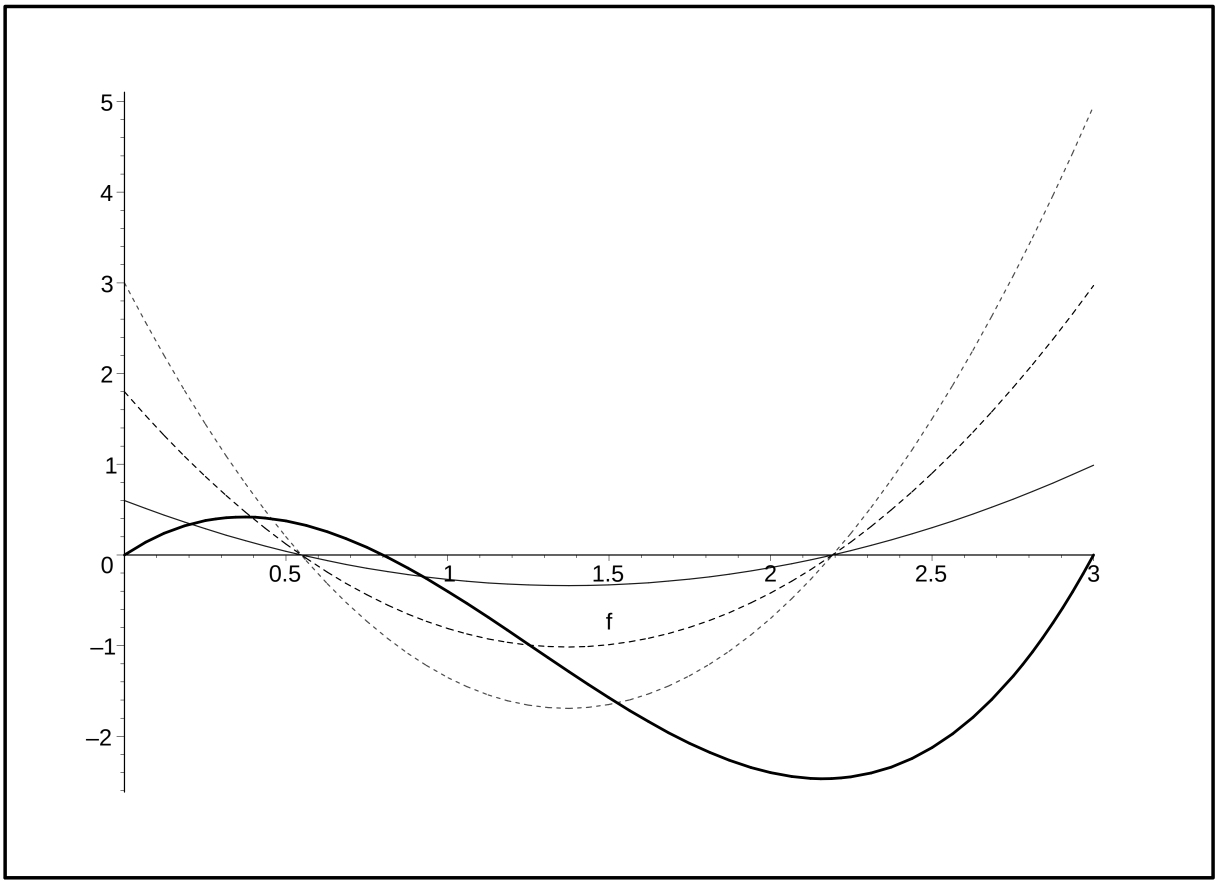
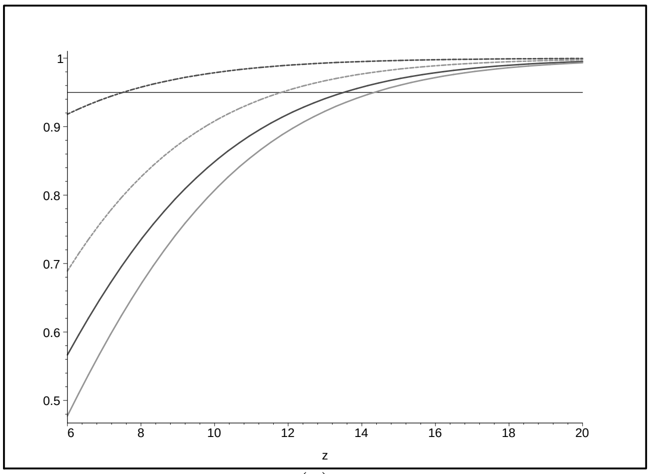
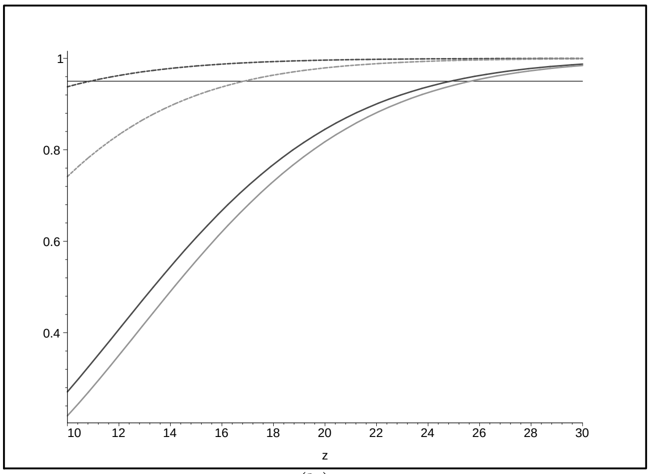


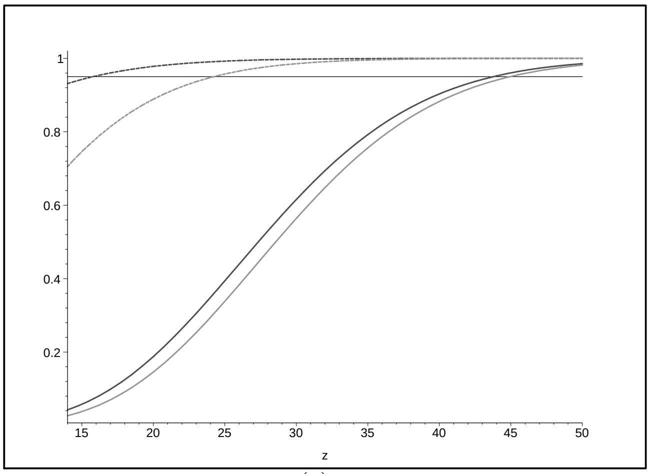
Figure 2: The Two Components of the Characteristic Polynomial, *m=2*: The cubic (dark line) is $f\psi_1(f)$, intersecting the axis at 0, λ_1, and λ_2.; the quadratics shown are members of the family $q\psi_2(f)$ for three values of q, all intersecting the axis at f_{21} and f_{22}.



(a)



(b)



(c)

Figure 3: Bounds on the cdf of the LR statistic, $m = 3$: (a) $k = 8$, $\lambda_1 = 2$ and $\lambda_1 = 8$; (b) $k = 18$, $\lambda_1 = 4$ and $\lambda_1 = 18$; (c) $k = 40$, $\lambda_1 = 12$ and $\lambda_1 = 40$. The lower pair of bounds corresponds, in each case, to the smaller value of λ_1. The horizontal line is at .95.

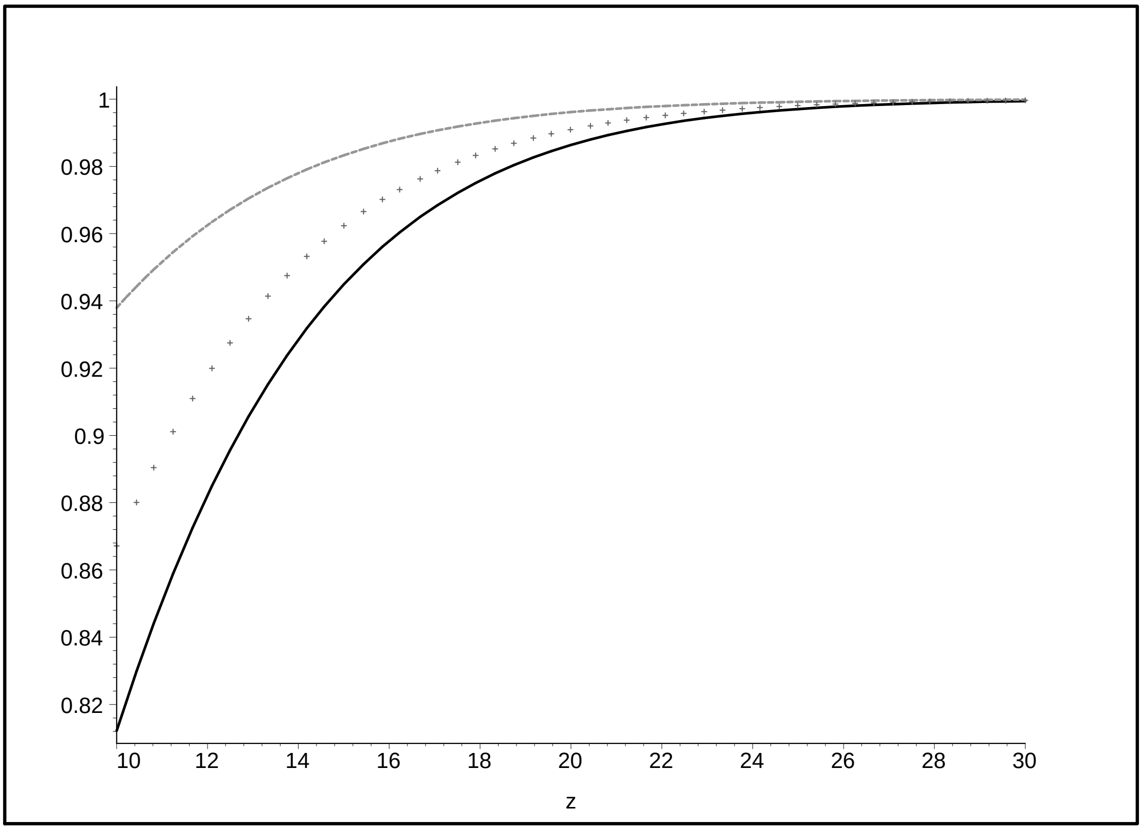


Figure 4: Bounds (solid and dashed lines) and exact (crosses) cdf of the LR statistic: $m = 2$, $k = 18$, $\lambda_1 = 18$, $\lambda_2 = 20$, computed using the exact formula in Theorem 6.