Exact Properties of the Conditional Likelihood Ratio Test in an IV Regression Model

Grant Hillier

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP23/06

Exact properties of the conditional likelihood ratio test in an IV regression model*

By

Grant Hillier

CeMMAP and Economics Division, University of Southampton.

May 4, 2007

Abstract

For a simplified structural equation/IV regression model with one right-side endogenous variable, we derive the exact conditional distribution function of Moreira's (2003) conditional likelihood ratio (CLR) test statistic. This is used to obtain the critical value function needed to implement the CLR test, and reasonably comprehensive graphical versions of this function are provided for practical use. The analogous functions are also obtained for the case of testing more than one right-side endogenous coefficient, but in this case for a similar test motivated by, but not generally the same as, the likelihood ratio test. Next, the exact power functions of the CLR test, the Anderson-Rubin test, and the Lagrange multiplier test suggested by Kleibergen (2002) are derived and studied. The CLR test is shown to clearly conditionally dominate the other two tests for virtually all parameter configurations, but no test considered is either inadmissable or uniformly superior to the other two. The unconditional distribution function of the likelihood ratio test statistic is also derived using the same argument. This shows that both exactly, and under Staiger/Stock weak-instrument asymptotics, the test based on the usual asymptotic critical value is always oversized, and can be very seriously so when the number of instruments is large.

*This is a revised version of an earlier paper that bore the title "Exact critical value and power functions for the conditional likelihood ratio and related tests in the IV regression model with known covariance". I thank the editor, Peter Phillips, and two anonymous referees for useful comments that helped improve on that version.

1 Introduction

Interest in problems of inference in the IV regression/structural equation model has seen a huge revival in the past few years, motivated by the realization that, in weakly identified models of this type, standard first-order asymptotic theory can be an extremely misleading guide to successful inference. See Stock, Wright, and Yogo (2002), and Andrews and Stock (2006), for surveys of much of this work, and Phillips (1989) for an early pointer to the problems under discussion.

One strand of the recent literature that is particularly welcome is the revival of interest in the problem of hypothesis testing in these models. In particular, because the properties of the standard tests depend on nuisance parameters which, in a weak-instrument context, may induce poor quality inference procedures, there has been great interest in eliminating these effects by invoking similarity and/or invariance arguments. Moreira (2003), Andrews, Moreira, and Stock (2006), Kleibergen (2002), (2005), and Chamberlain (2005), are some of the main contributors. These authors have shown that, at least when the error covariance matrix is known and the null hypothesis involves all coefficients of the endogenous variables, similar tests exist, and can be characterized. Moreover, the likelihood ratio (LR) test, which has emerged as the test most likely to be near-optimal, can be rendered similar by conditioning on a suitable statistic, and choosing the critical value so that the conditional size of the test does not depend on the conditioning variate. This is referred to as the conditional likelihood ratio (CLR) test.

One purpose of this paper is to provide the critical value functions that are needed to implement this testing procedure. An explicit expression for the conditional distribution function of the LR statistic - needed to define the critical value function - has hitherto been unavailable. Thus, we first give an exact expression for the conditional distribution of the LR statistic in the case of a single right-hand-side endogenous variable, given the statistic upon which the test must be conditioned in order to ensure similarity. The critical value function is then defined implicitly by the requirement that the conditional size of the test is constant, say α. This function cannot be written down explicitly, but is easily rendered graphically by the implicit-plot facility of a symbolic computer-algebra package. Detailed graphics are included to enable the practitioner to read off the critical value needed to produce an exact test.

We then extend these results, and again provide graphical critical value functions based on the relevant distribution function, for the case where the test involves the coefficients of several endogenous variables. In this case results for the true LR test, which involves the smallest root of a (random) polynomial of degree at least three, are much more difficult. Instead, we propose a test - a new similar test - that is motivated by, but different from, the true LR test, but for which generalized versions of the results obtained earlier can be used.

Our other main purpose is to provide analytical results on the power properties of the CLR and two closely related tests. No such results have hitherto been available, and the literature mentioned above has studied power properties entirely by simulation methods. The conditional power functions for the CLR test, the Anderson-Rubin (AR) test, and the Lagrange multiplier (LM) test are derived and studied in Section 5.

Even in the very simplified model discussed here, there is no uniformly best test, nor a uniformly best similar or invariant test. Thus, optimality results can only be obtained by placing further restrictions on the class of tests considered, or modifying the objective function to, say, average power, rather than actual. Andrews, Moreira, and Stock (2006) present detailed results of this type for the same problem as is discussed here. Some of the analytical results reported below are closely related to the results in that paper. Here we concentrate on the conditional power of the tests discussed, given a statistic that measures the sample Fisher information on the interest-parameter. However, the argument for conditioning is now different: we argue that the conditional power function is a more useful measure of test performance than the unconditional power function, which averages over values of the information measure that are not relevant. A related argument focusing on estimation in this model is given in detail in Forchini and Hillier (2003).

We begin by describing the model and notation to be used. After making the usual standardizing transformations, Section 2 gives the results dicussed above for the case of testing a single endogenous coefficient. In Section 3 we briefly discuss the properties of the unconditional LR test. We show that, when there are many instruments and the Fisher information is small, the unconditional size of the test (based on the usual asymptotically-justified critical value) can be very much larger than its nominal size. This result is also shown to hold under the so-called "weak instrument asymptotics" introduced by Staiger and Stock (1997), and is, of course, the motivation for seeking similar tests. Section 4 gives results for the case of testing several coefficients, and gives a new similar test for this case, a close relative of the one-parameter test in Section 3. Section 5 contains the power function results, and Section 6 provides some concluding remarks, including some comments on the case where the covariance matrix is unknown.

Naturally, although these tests and their properties are motivated by, and analysed under, Gaussian assumptions, we expect that their properties will be reasonably robust to departures from those assumptions. This, of course, demands investigation; there is Monte Carlo evidence in some of the literature cited above that it is so, but it is not an issue to which this paper contributes directly.

Throughout the paper we make extensive use of the generalized hypergeometric functions defined, for non-negative integers p, q $(p \leq q+1)$, by:

$$
{}_pF_q\left(a_1, ..., a_p; c_1, ..., c_q; z\right) = \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{\Pi_{i=1}^{p}(a_i)_j}{\Pi_{i=1}^{q}(c_i)_j},
$$

where $(a)_j = a(a+1)....(a+j-1)$ is the forward factorial or Pocchammer symbol. The series converges uniformly for $p \leq q$, and on the interval $|z| < 1$ for $p = q+1$, diverging in all other cases unless one of the a_i is a negative integer, in which case the series terminates.

1.1 Model and Assumptions

We consider the simple Gaussian structural-equation/IV-regression model:

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

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

where Z is $n \times k$, Π is $k \times m$, and we assume always that $m \leq k$. The corresponding reduced form model is of the form

$$Y = (y_1, Y_2) \sim N(Z\Pi(\beta, I_m), I_n \otimes \Omega), \tag{3}$$

where $Y = (y_1, Y_2)$ $(n \times (m+1))$ contains the observations on all the endogenous variables. The case in which equation (1) also contains exogenous variables on the right (that also appear in the reduced form (2)) is easily transformed into this simpler case. In common with most of the current literature on the problem, we assume that the reduced form covariance matrix Ω is known. Some consequences of relaxing this assumption are discussed briefly in the concluding comments at the end of the paper.

The log-likelihood is, apart from constants,

$$l(\Pi, \beta; \Omega) = -\frac{1}{2} trace \left\{ \Omega^{-1} Y'Y + \Sigma_{\beta\beta} \Pi' Z' Z \Pi - 2(\beta, I_m)\Omega^{-1} Y' Z \Pi \right\}, \tag{4}$$

where $\Sigma_{\beta\beta} = (\beta, I_m)\Omega^{-1}(\beta, I_m)'$. The null hypothesis of interest is

$$H_0 : \beta = \beta_0, \tag{5}$$

and thus specifies the entire vector β. There is no loss of generality in taking $\beta_0 = 0$, and we do so from now on.

The MLE for Π for fixed β and Ω is

$$\hat{\Pi}_\beta = (Z'Z)^{-1} Z' Y \Omega^{-1} \begin{pmatrix} \beta' \\ I \end{pmatrix} \Sigma_{\beta\beta}^{-1}, \tag{6}$$

and the concentrated log-likelihood is, apart from constants,

$$l_c(\beta; \Omega) = -\frac{1}{2} \frac{\binom{1}{-\beta}' Y' P_Z Y \binom{1}{-\beta}}{\binom{1}{-\beta}' \Omega \binom{1}{-\beta}} = -\frac{1}{2} r(\beta),$$

say, where $P_Z = Z(Z'Z)^{-1}Z'$. If H_0 is true this is $-\frac{1}{2}r(0)$, and for variable β it has a maximum equal to $-\frac{1}{2}f_1$, where f_1 is the smallest root of $|Y'P_ZY - f\Omega| = 0$. The likelihood ratio test therefore rejects H_0 for large values of

$$T = r(0) - f_1. \tag{7}$$

1.2 Standardization

It is clear from (4) that $Z'Y$ is a minimal sufficient statistic for the model, so all inferential procedures may be based on this matrix without essential loss of information. However, it is convenient to transform the sufficient statistic so that both its rows and columns are independent. Thus, let

$$U_\Omega = \begin{bmatrix} 1/\sqrt{\omega_{11}}, & -\omega_{21}'\Omega_{22.1}^{-\frac{1}{2}}/\omega_{11} \\ 0, & \Omega_{22.1}^{-\frac{1}{2}} \end{bmatrix}, \tag{8}$$

where $\Omega_{22.1} = \Omega_{22} - \frac{\omega_{21}\omega_{21}'}{\omega_{11}}$, so that $U_\Omega'\Omega U_\Omega = I_{m+1}$. Postmultiplication of $Z'Y$ by U_Ω therefore produces an identity covariance matrix for the rows of the transformed matrix (see Phillips (1983) for further discussion of this standardizing transform).

Multiplying on the left by $(Z'Z)^{-\frac{1}{2}}$ to produce an identity covariance matrix for the columns of P, we now define,

$$(p_1, P_2) = (Z'Z)^{-\frac{1}{2}}Z'YU_\Omega. \tag{9}$$

It then follows that p_1 and P_2 are independent, and, under H_0, $p_1 \sim N(0, I_k)$, and $P_2 \sim N(M_0, I_k)$, where $M_0 = (Z'Z)^{\frac{1}{2}}\Pi\Omega_{22.1}^{-\frac{1}{2}}$. Clearly, P_2 is sufficient for M_0 under H_0, so we have the following characterization of the class of similar tests for H_0 (cf. Hillier (1987b) for background, and Moreira (2003)):

Proposition 1 *P_2 is a complete sufficient statistic for Π (or M_0) when H_0 is true. Hence, every similar test must have fixed size in the distribution of p_1, i.e., fixed size conditional on P_2.*

2 The CLR Test, $m = 1$

Assume now that $m = 1$, and replace the matrix P_2 by the vector p_2 $(k \times 1)$ (and Π by π). If we put $W = (p_1, p_2)'(p_1, p_2)$, the components of the LR statistic T above can be written explicitly in terms of the elements of W (or in terms of (p_1, p_2)) as follows:

$$r(\beta_0) = p_1'p_1,$$

and

$$f_1 = \frac{1}{2}\left(t - \sqrt{t^2 - 4d}\right), \tag{10}$$

where $t = trace[W] = p_1'p_1 + p_2'p_2$ and $d = \det[W] = (p_1'p_1)(p_2'p_2) - (p_1'p_2)^2$.

Now let

$$\left.\begin{array}{c} q_1 = p_1'M_{p_2}p_1, \\ q_2 = p_1'P_{p_2}p_1, \\ q = q_1 + q_2 = p_1'p_1, \\ b = q_1/q, \ 0 < b < 1, \end{array}\right\} \tag{11}$$

where $P_{p_2} = p_2(p_2'p_2)^{-1}p_2'$ and $M_{p_2} = I_k - P_{p_2}$. Also set $w = w_{22} = p_2'p_2$, so that $t = q + w$ and $d = q_1 w = bqw$. Note that T depends on p_2 only through $w = p_2'p_2$.

Remark 1 *The statistic $b = q_1/q$ can be interpreted as a specification diagnostic for the maintained hypothesis that $E(p_1)$ is a multiple of $E(p_2)$: $1 - b$ is the value of R^2 in the regression of p_1 on p_2, so we expect b to be small if that maintained hypothesis is valid in the data. Kleibergen (2002) has proposed the Lagrange Multiplier statistic $LM = q_2 = (1 - b)q$ as an alternative test statistic for H_0. This statistic is discussed further in Section 5 below.*

The following distribution properties of (q, b, w) are easily deduced:

Proposition 2 *(Null distributions) Under H_0, and conditional on p_2, q_1 and q_2 are independent, $q_1|p_2 \sim \chi^2(k-1)$, $q_2|p_2 \sim \chi^2(1)$. Hence, these properties also hold unconditionally, so that q and b are unconditionally independent, $q \sim \chi^2(k)$, and $b \sim Beta(\frac{k-1}{2}, \frac{1}{2})$, (the Beta distribution with parameters $\frac{k-1}{2}$ and $\frac{1}{2}$). Both q and b are independent of w.*

Remark 2 *For the case where the covariance matrix is known the analogue of the Anderson-Rubin (1949) (AR) test statistic is q, which is $\chi^2(k)$ under H_0, but noncentral $\chi^2(k)$ under H_1 with noncentrality parameter $c_\beta^2 \pi' Z' Z \pi / \omega_{22}$, where $c_\beta = \sqrt{\frac{\omega_{22}}{\omega_{11}}}\beta$. Clearly, this can vanish when H_0 is false if π can be zero. That is, evidence that q is central $\chi^2(k)$ does not, by itself, suggest that H_0 is true: evidence that $\pi \neq 0$ is also needed. This observation is discussed more fully in Breusch (1985) and Hillier (1987a).*

From the above results it is clear that any test based on (q, b) alone will be similar. The AR test based on q is one such test, and the LM test based on $q_2 = (1 - b)q$ another. The properties of these tests will be examined further in Section 5 below. However, as Moreira (2003) has essentially pointed out, Proposition 1 implies that any test statistic that is a function of (q, b) *and* p_2 - like the likelihood ratio test statistic T - will be similar if and only if the critical value for the test is chosen in such a way that the conditional size - $\Pr\{T > z|p_2; H_0\} = \alpha$ - is not a function of p_2. In the case of T, which depends on p_2 only through w, this requires that the conditional size given w must not depend on w.

Remark 3 *The (expected) partial Fisher information for β is:*

$$i_n(\beta) = \pi' Z' Z \pi / \omega^2, \tag{12}$$

where $\omega^2 = \omega_{11}(1 - \rho^2)$, with $\rho = \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$ the correlation between the endogenous variables. Substituting the MLE for π when H_0 is true into this expression shows that the 'empirical information' at β_0 is proportional to w. Thus, to condition on w is to condition on the 'empirical information' under the null hypothesis.

Now, in terms of the mutually independent variates (q, b, w),

$$T = \frac{1}{2}\left(q - w + \sqrt{(q+w)^2 - 4wqb}\right). \tag{13}$$

This is obviously quite a complicated function of (q, b, w), so a direct attempt to deduce its density is difficult. However, it is easy to see that T is monotonic increasing in q, so the inequality $T < z$ is easily seen to be equivalent to the inequality

$$q < z(1 - a(z, w)b)^{-1}, \tag{14}$$

where $a = a(z, w) = w/(w + z)$ $(0 < a < 1)$. Thus, the conditional *cdf* of T, given b and w, is trivially obtained as

$$\Pr\{T < z|b, w\} = G_k(z(1 - ab)^{-1}), \tag{15}$$

where $G_k(\cdot)$ denotes the *cdf* of the central $\chi^2(k)$ distribution. The conditional *cdf*, given either b or w, can obviously be obtained from this by averaging with respect to the other variable.[1] Our interest next will be in the average with respect to b, and in the next section we also consider the average with respect to both b and w, *i.e.*, the unconditional density.

2.1 Conditional Distribution Function given w

In this Section we derive the conditional distribution function of T given w under H_0,

$$P_k(z; w) = \Pr\{T < z|w, H_0\}, \tag{16}$$

Having done that, we would like to use the *cdf* to obtain the critical value function, $z_\alpha(k; w)$ say, defined (implicitly) by the equation $P_k(z; w) = 1 - \alpha$. That is, $z_\alpha(k; w)$ is the critical value needed to obtain a conditional size for the test, given w, that is free of w. Before stating the main result we record a simple consequence of the fact that the inequality $T < z$ is equivalent to $q < z(1 - ab)^{-1}$.

For fixed w the critical regions in $(q, b)-$space corresponding to the AR, LM, and CLR tests are as follows:

$$\begin{aligned} C_{AR} &: \{q, b : q > c_\alpha(k)\} \\ C_{LM} &: \{q, b : q(1 - b) > c_\alpha(1)\}, \\ C_{CLR} &: \{q, b : q(1 - a_\alpha b) > z_\alpha(k; w)\}, \end{aligned} \tag{17}$$

where $a_\alpha = w/(w + z_\alpha(k; w))$, and each region has size α. Since $1 - b < 1 - a_\alpha b < 1$, it is easy to see that these definitions imply that, for fixed α, and all k and $w > 0$,

$$c_\alpha(1) \le z_\alpha(k; w) \le c_\alpha(k). \tag{18}$$

We shall see later that, for w small, $z_\alpha(k;w)$ is close to $c_\alpha(k)$, while for w large it is close to $c_\alpha(1)$.

We now seek to evaluate $P_k(z;w)$. Conditioning first on both b and w, the required *cdf* is, as we have seen, $G_k(z(1-ab)^{-1})$, and the unconditional *cdf* is the expectation of this with respect to b. That is,

$$P_k(z;w) = E_b\left[G_k(z(1-ab)^{-1})\right], \tag{19}$$

with $b \sim Beta(\frac{k-1}{2}, \frac{1}{2})$. Evaluating the expectation yields the main result of this section:

Theorem 1 *Let $a = w/(w+z)$ and $P_k(z;w) = \Pr\{T < z|w, H_0\}$. Then*

$$P_k(z;w) = (1-a)^{\frac{1}{2}} \sum_{l=0}^{\infty} \frac{a^l (\frac{1}{2})_l}{l!} G_{k+2l}(w+z). \tag{20}$$

All proofs are in the Appendix.

Remark 4 *Although the result here expresses the cdf as an infinite series of chi-square cdf's, because $a < 1$ the series converges very rapidly; the first ten or twenty terms are usually sufficient to achieve sufficient accuracy to use the result to, for instance, compute a conditional $p-$value. It is worth noting too that equation (19), expressing $P_k(z;w)$ as the mean of a function of a Beta variate, provides a very efficient procedure for simulating conditional $p-$values. Given z and w, one can simply average the values of $G_k(z(1-ab)^{-1})$ over repeated i.i.d. draws for b from the $Beta(\frac{k-1}{2}, \frac{1}{2})$ distribution. A relatively small number of repetitions (far less than the 10,000 needed by Moreira's (2003) simulation method) is sufficient to achieve acceptable accuracy. This simulation approach can be extended to the case of testing m parameters, as discussed in Hillier (2006b).*

2.2 Properties of the Distribution Function

The conditional density function of T, given w, can also be obtained from the relation $P_k(z;w) = E_b\left[G_k(z(1-ab)^{-1})\right]$ by first differentiating $G_k(z(1-ab)^{-1})$ with respect to z, then evaluating the expectation with respect to b. Since the density is probably of little interest, we omit this result and concentrate on the conditional *cdf*.

First note that, since it easy to see that the argument of $G_k(\cdot)$ in (15) is monotonic increasing in b, an immediate implication of equation (19) is:

$$G_k(z) \leq P_k(z;w) \leq G_k(z+w). \tag{21}$$

Thus, we have the intuitively sensible result that $P_k(z;w)$ will be close to $G_k(z)$ when w is small.

Next, it is clear that when $w = 0$, $P_k(z; 0) = G_k(z)$. The following result - which can be seen clearly in Figure 2 below - shows that the null distribution of T approaches that of a $\chi^2(1)$ random variable as $w \to \infty$ (see Moreira (2003) for a different proof of this result):

Proposition 3 *For fixed z and k, $P_k(z; w) < G_1(z)$ for all $w > 0$, and*

$$P_k(z; w) \to G_1(z) \text{ as } w \to \infty. \tag{22}$$

The function $G_1(z)$ is, of course, the conventional (unconditional) reference distribution based on standard asymptotics for the LR test. The standard result arises from the assumption that the information parameter $i_n(\beta)$, upon which the (null) distribution of w depends, goes to infinity as the sample size increases. Proposition 3 shows that this is also the correct *conditional* reference distribution for large *empirical information* w, but also shows that this choice will be incorrect for smaller values of w : the actual conditional size will exceed the nominal size of the test for small values of w (see also Section 3 below).

Finally, the following result, which gives a recursion for computing the functions $P_k(z; w)$, may be useful for computation. It is well-known that the *cdf* of the chi-square distribution satisfies a recursive relation (Abramowitz and Stegun (1972), Section 26.4.8). In view of Theorem 1, it is not surprising that $P_k(z; w)$ satisfies a similar recursion. This is given in:

Proposition 4 *The conditional distribution function $P_k(z; w)$ satisfies the recursion:*

$$P_{k+2}(z; w) = P_k(z; w) - \frac{z^{\frac{1}{2}}(w+z)^{\frac{k-1}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2}+1)} \exp\{-\frac{w+z}{2}\}{}_1F_1\left(\frac{1}{2}, \frac{k}{2}+1; \frac{w}{2}\right). \tag{23}$$

It follows that these functions can be generated recursively for all k from the cases $k = 1, 2$. When $k = 1$ it is easy to see that $T = q$, so that $P_1(z; w) = G_1(z)$, the $\chi^2(1)$ *cdf,* which is free of w. When $k = 2$, $P_2(z; w)$ is easily computed from the expression given in Theorem 1 above, and the recursion can then be used for even values of k.

Examples of the conditional distribution function are shown in Figure 1 for several values of w, and for various (odd) values of k. It is evident from the figures that when w is small the distribution is extremely sensitive to the value of k, but it is less sensitive to k when w is large. The observations made in Proposition 3 above are clearly visible in Figure 1: the actual conditional *cdf* can be very different from the reference $\chi^2(1)$ *cdf* when w is small. Clearly, these properties also imply that it is impossible to obtain an approximation to the conditional distribution function of T that will be accurate for all w and k.

2.3 Critical Value Function

To obtain a similar test based on T we clearly need to choose the critical value z, $z_\alpha(k;w)$ say, so that $P_k(z;w) = 1-\alpha$, and the result in Theorem 1 enables us to do so exactly, thus producing a version of the likelihood ratio test that is similar, the CLR test.

For fixed α the equation $P_k(z;w) = 1-\alpha$ implicitly defines z as a function of w, but that function cannot, it seems, be written down explicitly. We know from the results above that $z_\alpha(k;0) = c_\alpha(k)$, and, from Proposition 4, that, for all k,

$$z_\alpha(k;w) \to c_\alpha(1) \; as \; w \to \infty. \tag{24}$$

For intermediate values of the conditioning variate w the critical value function is unknown. Nevertheless, it is relatively straightforward to produce implicit plots of the function involved using a symbolic computer algebra package. We can do so for each combination of the two parameters that the distribution depends on - k, and the observed value of w - and the appropriate critical value can then be read off from the graph. In fact, since w is continuous, graphical presentation of the critical value function is more useful than conventional tabulations would be.

Figure 2 provides detailed plots of the critical value function for tests of size $\alpha = .05$, for odd values of k from 3 through 21. Figure 3 gives the corresponding results for even values of k from 2 through to 20. It is evident from the figure that for fixed k, $z_{.05}(k;w)$ decreases in w, but quite slowly, while for fixed w it increases quite rapidly in k.

Figures 2 and 3 provide reasonable coverage of (k,w)-space, but of course the above formulae can also be used (with the aid of a symbolic computer algebra package) to compute the required critical value for configurations of (k,w) not covered by the figures.

3 The LR Test and Weak Instruments

Although somewhat removed from our main purpose, it is of interest at this point to briefly discuss the unconditional properties of the LR statistic. In particular, it is relatively easy to adapt the methods used above to obtain an expression for the unconditional *cdf* of T, and use this to examine the size behaviour of the test under various assumptions about the behaviour of k and/or the information quantity $i_n(\beta)$. Here we provide the analogue for the LR test of the "weak instruments" results established for estimators in Staiger and Stock (1997), who assume that $i_n(\beta)$ fails to increase indefinitely as the sample size n increases. We shall show that, under this assumption, the LR test is oversized, and can be seriously so if k is large.

In principle the unconditional *cdf* of T can be obtained from the conditional *cdf* $P_k(z;w)$ by averaging with respect to the density of w (the non-central χ^2 density with k degrees of freedom and noncentrality parameter Δ_n, say). However, in the expression for the unconditional *cdf* (implied by equation (14)),

$$\Pr\{T < z|H_0\} = E_b E_w \left(G_k(U)\right), \tag{25}$$

where

$$U = U(w,b) = \frac{z(z+w)}{z+w(1-b)}, \tag{26}$$

it turns out to be more convenient to evaluate the expectation with respect to w, rather than b, first. This produces the following result:

Theorem 2 *Under the null hypothesis,*

$$\Pr\{T < z|H_0\} = G_1(z) - E_b\left[\int_z^{\frac{z}{1-b}} g_k(u) G_k^{\Delta_n}\left(\frac{z(u-z)}{(z-u(1-b))}\right) du\right], \tag{27}$$

where $G_k^{\Delta_n}(\cdot)$ denotes the cdf of a noncentral χ^2 variate with k degrees of freedom and noncentrality parameter Δ_n, and Δ_n is proportional to $i_n(\beta)$ in (12).

Several properties of the unconditional size of the test,

$$\begin{aligned} \alpha_{LR}(z;k,\Delta_n) &= \Pr\{T > z|H_0\} \\ &= 1 - G_1(z) + E_b\left[\int_z^{\frac{z}{1-b}} g_k(u) G_k^{\Delta_n}\left(\frac{z(u-z)}{(z-u(1-b))}\right) du\right], \end{aligned} \tag{28}$$

say, follow easily from this expression. Beginning with the standard asymptotic result for the LR test, these are gathered in:

Corollary 1 *(1) For fixed k,*

$$\alpha_{LR}(z;k,\Delta_n) \to \Pr\{\chi^2(1) > z\} \text{ as } \Delta_n \to \infty. \tag{29}$$

(2) Uniformly in k and Δ_n,

$$\alpha_{LR}(z;k,\Delta_n) > \Pr\{\chi^2(1) > z\}, \tag{30}$$

and

$$\alpha_{LR}(z;k,\Delta_n) < \Pr\{\chi^2(1) > z\} + E_b\left[\int_z^{\frac{z}{1-b}} g_k(u) G_k\left(\frac{z(u-z)}{(z-u(1-b))}\right) du\right]. \tag{31}$$

i.e., the size of the LR test is uniformly above its asymptotic nominal size, but is bounded above by the expression on the right in equation (31).
(3) For all k and Δ_n,

$$\Pr\{\chi^2(1) > z\} < \alpha_{LR}(z;k,\Delta_n) < \Pr\{\chi^2(k) > z\}. \tag{32}$$

That is, the size of the LR test using a fixed critical value is bounded below by that of a $\chi^2(1)$ r.v., and above by that of a $\chi^2(k)$ r.v..

In the asymptotic sequence adopted by Staiger and Stock (1997) to model "weak instruments" (or poor information on β), k is fixed and $\Delta_n \to \Delta$ ($0 < \Delta < \infty$) as $n \to \infty$. Since $G_k^{\Delta_n}(\cdot)$ is continuous in Δ_n, we easily obtain the following:

Proposition 5 *Under Staiger/Stock weak-instrument asymptotics, i.e, k fixed and $\Delta_n \to \Delta$ as $n \to \infty$,*

$$\alpha_{LR}(z; k, \Delta_n) \to \alpha_{LR}(z; k, \Delta) > \Pr\{\chi^2(1) > z\}, \tag{33}$$

i.e., $\alpha_{LR}(z; k, \Delta)$ is strictly greater than the asymptotic nominal size.

Thus, not surprisingly, under Staiger/Stock "weak-instruments asymptotics" the usual asymptotic size result for the LR test fails, and the size of the test is analytically of the same form as it is in finite samples. These are analogous to the results given for estimators by Staiger and Stock (1997).

In Figure 4 we plot the upper bound on the size of the LR test given in equation (31) as a function of k. The points were computed by simulating the expectation $E_b E_w(G_k\left(\frac{z(z+w)}{z+w(1-b)}\right)$, with $w \sim \chi^2(k)$ (when the bound is attained), and $b \sim Beta(\frac{k-1}{2}, \frac{1}{2})$, for values of k from $k = 2$ to $k = 82$ in steps of 2, each point being generated using a sample size of 1000.[2] The actual size of the test will be close to this upper bound for small values of Δ_n (or Δ in the Staiger/Stock story). It is clear from Figure 4 that the upper bound rises quite sharply as k increases, so the true size of the LR test can be considerably above its nominal level when k is large and Δ_n small. This, of course, is the motivation for the conditional approach.

It would be interesting to extend this analysis to study the behaviour of the LR test under the "many weak instruments" asymptotic sequences like those assumed by Chao and Swanson (2005), Han and Phillips (2006), and Hansen, Hausman, and Newey (2006), for instance. In these papers both k and Δ_n increase with n, but not necessarily at the same rate, and this can considerably modify the asymptotic properties of the commonly used estimators for β. Further study of these effects for the LR test is, however, a subject for another paper.

4 Several RHS Endogenous variables

Although, in principle, the argument used for the case $m = 1$ can be applied when $m > 1$, because T involves the characteristic root f_1 - which is now the smallest root of a polynomial of degree $m + 1$ - this is much more difficult. In this section we discuss a conditional exact test that is motivated by the LR test, but which, except in special cases, is not the true LR test. This test is designed specifically so that its properties - both size and power - can be analysed by essentially the same methods as above.

4.1 An Approximate LR Test

In view of Proposition 1, to obtain a similar test the critical value for the test needs to be chosen so that the conditional size of the test, given P_2, is constant. As noted above, this is difficult for the true LR test, but we now suggest an alternative to the true LR test that is closely related to it, but for which the similarity condition can be implemented more easily.

The true LR test rejects for large values of $T = q - f_1$. We seek a statistic of the form $T^* = q - f_1^*$, say, such that (i) f_1^* is close to f_1, and, to preserve the analogy with the case $m = 1$, (ii) f_1^* is the smaller of the two roots of a quadratic of the type:

$$p_\lambda(f) = f^2 - f(\lambda + q) + \lambda q_1, \tag{34}$$

where λ is a function only of P_2, and the statistics

$$\left.\begin{array}{l} q_1 = p_1' M_{P_2} p_1, \\ q_2 = p_1' P_{P_2} p_1, \end{array}\right\} \tag{35}$$

and $q = q_1 + q_2$, are analogues of the statistics defined above for the case $m = 1$. In terms of the matrix

$$W = (p_1, P_2)'(p_1, P_2) = \begin{bmatrix} w_{11} & w_{21}' \\ w_{21} & W_{22} \end{bmatrix}, \tag{36}$$

$q = w_{11}$ and $q_1 = w_{11.2} = w_{11} - w_{21}' W_{22}^{-1} w_{21}$. The smaller of the two roots of $p_\lambda(f)$ is

$$f_1^*(\lambda) = \frac{1}{2}\left\{(\lambda + q) - \sqrt{(\lambda + q)^2 - 4\lambda q_1}\right\}. \tag{37}$$

In the case $m = 1$, the choice $\lambda = w$ produces $f_1^*(w) = f_1$, and so yields the true likelihood ratio test when $m = 1$. Let $f_1 \leq f_2 \leq ... \leq f_{m+1}$ denote the ordered characteristic roots of W, and $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_m$ those of W_{22}. If $\lambda_i = \lambda$ for all i, the characteristic polynomial of W is (see the proof of Lemma 1 in the Appendix, particularly equation (80))

$$\bar{p}(f) = (\lambda - f)^{m-1} p_\lambda(f),$$

and it easy to see that $f_1 = f_1^*(\lambda)$, $f_{m+1} = f_2^*(\lambda)$, and $f_i = \lambda$ for $i \neq 1, m+1$. Thus, in this special case the test based on $T^* = q - f_1^*(\lambda)$ is the true LR test.

In the general case we obtain the following result, proved in the Appendix:

Lemma 1 *Let f_1 be the smallest characteristic root of W, λ_1 be the smallest characteristic root of W_{22}, and $f_1^*(\lambda_1)$ the smaller of the two roots of $p_{\lambda_1}(f) = 0$. Then:*

$$0 < f_1^*(\lambda_1) \leq f_1 \leq \lambda_1. \tag{38}$$

Thus, the suggested test is based on the statistic

$$T^* = q - f_1^*(\lambda_1), \tag{39}$$

and the critical value z must be chosen so that $\Pr\{T^* \leq z|\lambda_1\} = 1 - \alpha$ to render the test similar. Intuitively, Lemma 1 suggests that, at least in large samples, the tests based on T and T^* should agree, since f_1 will be close to zero (because W converges to a rank-m matrix) as the sample size increases. Nevertheless, the test based on T^* - which we shall show can also be rendered similar by conditioning - should be thought of as a new similar test, motivated by, but usually not the same as, the true LR test.[3]

Let, for the moment, λ be any positive scalar that is a function of P_2 alone. The following Theorem generalizes Theorem 1:

Theorem 3 *Under H_0 q_1 and q_2 are conditionally independent, $q_1 \sim \chi^2(k-m)$ and $q_2 \sim \chi^2(m)$. These are therefore also the unconditional distributions. Define $T^* = q - f_1^*(\lambda)$, with λ a function only of P_2, and define*

$$P_k^m(z;\lambda) = \Pr\{T^* < z|\lambda; H_0\} = E_b\left[G_k(z(1-ab)^{-1})\right], \tag{40}$$

with $a = \lambda/(z+\lambda)$ and $b \sim Beta(\frac{k-m}{2}, \frac{m}{2})$. Then:

$$P_k^m(z;\lambda) = (1-a)^{\frac{m}{2}} \sum_{l=0}^{\infty} \frac{a^l (\frac{m}{2})_l}{l!} G_{k+2l}(\lambda + z). \tag{41}$$

The functions $P_k^m(z;\lambda)$ occur not only in the conditional distribution function of T^*, but also in the power functions for these tests, and in the distribution theory for the analogous tests when Ω is unknown (see Section 5 below, and Hillier (2005)). The generalized version of Proposition 3 is easily seen to hold:

Proposition 6 *For fixed $k \geq m$ and z, $P_k^m(z;\lambda) < G_m(z)$ for all $\lambda > 0$, and*

$$P_k^m(z;\lambda) \to G_m(z) \text{ as } \lambda \to \infty. \tag{42}$$

That is, the null cdf of T^* approaches that of a $\chi^2(m)$ random variable as the conditioning variate $\lambda \to \infty$.

And, the functions $P_k^m(z;\lambda)$ also satisfy a recursive relation analogous to that given in Proposition 4 for the case $m = 1$:

Proposition 7 *$P_m^m(z,\lambda) \equiv G_m(z)$, and, for $k > m$, the functions $P_k^m(z;\lambda)$ defined in equation (41) satisfy the recursion:*

$$P_{k+2}^m(z;\lambda) = P_k^m(z;\lambda) - \frac{z^{\frac{m}{2}}(\lambda+z)^{\frac{k-m}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2}+1)} \exp\{-\frac{\lambda+z}{2}\} {}_1F_1\left(\frac{m}{2}, \frac{k}{2}+1; \frac{\lambda}{2}\right). \tag{43}$$

Remark 5 *Since the null hypothesis imposes what appear to be m constraints, the asymptotic reference distribution for the LR test statistic is that of a $\chi^2(m)$ random variable. Proposition 6 shows that the conditional distribution given λ does indeed converge to $\chi^2(m)$ as $\lambda \to \infty$, but it can be much closer to $\chi^2(k)$ for small values of λ.*

Some examples of the functions $P_k^m(z; \lambda_1)$ are shown in Figures 5 and 6. In Figure 5 $\lambda_1 = 1$, and in Figure 6 $\lambda_1 = 8$. Again we see that the conditional *cdf* is sensitive to k when λ_1 is small, but is less so when λ_1 is large.

Since this paper was written I have succeeded in showing that the conditioning argument used in this paper can be generalized to the case $m > 1$ - see (Hillier (2006b)). Thus, it is now possible to implement the CLR test when $m > 1$, although the test suggested in this section is still much simpler to implement. As a referee has suggested, it would be interesting to compare these two tests, but this is beyond the scope of the present paper.

4.2 Critical Value Function for the Approximate LR Test

We have already established that the critical value function $z_\alpha^m(k; \lambda)$, say, satisfies $z_\alpha^m(k; 0) = c_\alpha(k)$ and, in Proposition 6, that

$$z_\alpha^m(k; \lambda) \to c_\alpha(m) \text{ as } \lambda \to \infty. \tag{44}$$

However, as before, the critical value function $z_\alpha^m(k; \lambda)$, implicitly defined by the equation

$$P_k^m(z; \lambda) = 1 - \alpha,$$

cannot be obtained explicitly, but is readily graphed using a symbolic computer algebra package. Figure 7 provides some examples of these functions for the case $\alpha = .05$, values $m = 2$, 3, and 4, respectively, and, in each case, for $k = m+2, m+4, ..., m+10$, thus providing a fairly extensive coverage of the parameter set (k, m, λ).

5 Exact Power Functions: $m = 1$

Because we now have critical value functions giving, for each pair (k, w) and test size α, the critical value $z_\alpha(k; w)$, it is possible to calculate the power function of the CLR test, and compare it to that of other tests. Thus, in this section we analyse the power properties of the CLR and AR tests discussed in Section 2, as well as Kleibergen's (2002) suggested test statistic $LM = (1 - b)q$, dealing throughout only with the case $m = 1$. For $m > 1$ the results for tests based on (b, q) alone are straightforward extensions of those for the case $m = 1$, and conditional results for the test based on the statistic T^* are exactly analogous to those for $m = 1$.

We focus here on the *conditional* power properties of these tests for w fixed at its observed sample value. That is, we condition on the observed partial information on β *in the sample actually available.* As we shall see, the power properties of these (and other) tests depend on w, so, extending the argument made in Forchini and Hillier (2003) in the context of estimation in this model, it can be argued that the conditional power functions are the relevant basis for comparisons between the tests. But, even if one dismisses this argument, unconditional properties of the tests can also be deduced from their conditional counterparts, so the conditional results are useful from both points of view.

5.1 Conditional and Unconditional Power Functions

When H_0 is false p_1 and p_2 remain independent, and of course normally distributed with identity covariance matrices, but now $E(p_1) = c_\beta\mu$ and $E(p_2) = d_\beta\mu$, where (replacing Π by π when $m = 1$) $\mu = (Z'Z)^{\frac{1}{2}}\pi/\sqrt{\omega_{22}}$, $c_\beta = \sqrt{\frac{\omega_{22}}{\omega_{11}}}\beta$, and

$$d_\beta = \frac{1-\rho c_\beta}{\sqrt{1-\rho^2}}, \tag{45}$$

where $\rho = \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$ is the correlation between the two endogenous variables. Note that $d_\beta = 0$ at $c_\beta = \rho^{-1}$; we shall see below that it is this that induces the "quirky" behaviour of the power functions of the CLR and LM tests near $c_\beta = \rho^{-1}$ noted by Andrews, Moreira, and Stock (2006).

Remark 6 *Confining attention to the distribution of the matrix* $P = (p_1, p_2) \sim N(\mu(c_\beta, d_\beta), I_k \otimes I_2)$, *the problem can be thought of as testing* $H_0 : c_\beta = 0$ *in this distribution. This problem is invariant under the group of transformations* $P \to HP$, $H \in O(k)$, *and a maximal invariant under this group is the matrix* $W = P'P$. *The LR test is, of course, an invariant test, but there is no best invariant test (because the best invariant test for fixed* μ *depends on* μ*). Thus, to obtain an optimal test one must modify the definition of optimality (to, say, average power, not actual), or further constrain the class of tests considered (by, say, adding similarity to the restriction to invariant tests). Andrews, Moreira, and Stock (2006) present an extensive discussion of the problem discussed here using such arguments, and obtain some optimality results for weighted average and point optimal power criteria. In Hillier (1987a) and Hillier (2005) I explore the invariance properties of quite general hypothesis testing problems in much more general structural models with unknown covariance matrix.*

From the distribution properties of (p_1, p_2) we easily deduce the following conditional results for the distributions of (q_1, q_2), and hence (q, b), given p_2 :

Proposition 8 *(Non-Null Distributions) Under H_1, q_1 and q_2 are conditionally independent, given p_2,*

$$\begin{aligned} q_1|p_2 &\sim \chi'^2(k-1,\delta_1), \text{ with } \delta_1 = c_\beta^2\mu' M_{p_2}\mu, \\ q_2|p_2 &\sim \chi'^2(1,\delta_2), \text{ with } \delta_2 = c_\beta^2\mu' P_{p_2}\mu. \end{aligned}$$

Thus, under H_1,

$$\begin{aligned} pdf(q,b|p_2) &= \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{k-1}{2})} exp\{-\frac{1}{2}c_\beta^2\Delta_n\} exp\{-\frac{1}{2}q\} q^{\frac{k}{2}-1} b^{\frac{k-1}{2}-1}(1-b)^{-\frac{1}{2}} \\ &\times {}_0F_1\left(\frac{k-1}{2};\frac{1}{4}qb\delta_1\right) {}_0F_1\left(\frac{1}{2};\frac{1}{4}q(1-b)\delta_2\right), \end{aligned} \tag{46}$$

where

$$\Delta_n = \mu'\mu = \pi' Z' Z\pi/\omega_{22}. \tag{47}$$

Thus, in the non-null case q and b are no longer conditionally independent.

Here, $\chi'^2(v,\delta)$ denotes the non-central chi-square distribution with degrees of freedom v and noncentrality parameter δ. To implement the same strategy as was used above for the null *cdf* we now need, first, the conditional density of q given both b and p_2, then the conditional density of b given p_2. From the results in Proposition 8, these are (in reverse order) as follows:

Theorem 4 *The conditional density of b given p_2 is:*

$$\begin{aligned} pdf(b|p_2) &= [B(\frac{k-1}{2},\frac{1}{2})]^{-1} exp\{-\frac{1}{2}c_\beta^2\Delta_n\} b^{\frac{k-1}{2}-1}(1-b)^{-\frac{1}{2}} \\ & \sum_{j,l=0}^{\infty} \frac{(\delta_1 b)^j(\delta_2(1-b))^l(\frac{k}{2})_{j+l}}{j!l!2^{j+l}(\frac{k-1}{2})_j(\frac{1}{2})_l}, \end{aligned} \tag{48}$$

and, since p_1 and p_2 are independent, q is independent of p_2 and $q \sim \chi'^2(k, c_\beta^2\Delta_n)$, so that

$$pdf(q) = \frac{exp\{-\frac{1}{2}c_\beta^2\Delta_n\}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} exp\{-\frac{1}{2}q\} q^{\frac{k}{2}-1} {}_0F_1\left(\frac{k}{2};\frac{1}{4}qc_\beta^2\Delta_n\right). \tag{49}$$

Hence,

$$\begin{aligned} pdf(q|b,p_2) &= pdf(q,b|p_2)/pdf(b|p_2) \\ &= \frac{exp\{-\frac{1}{2}q\}q^{\frac{k}{2}-1}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} \left\{ \frac{{}_0F_1(\frac{k-1}{2};\frac{1}{4}qb\delta_1){}_0F_1(\frac{1}{2};\frac{1}{4}q(1-b)\delta_2)}{\sum_{j,l=0}^{\infty} \frac{(\delta_1 b)^j(\delta_2(1-b))^l(\frac{k}{2})_{j+l}}{j!l!2^{j+l}(\frac{k-1}{2})_j(\frac{1}{2})_l}} \right\}. \end{aligned} \tag{50}$$

Remark 7 *It is clear from these results that when $\mu = 0$ the joint distribution of (b, q, w) is free of all parameters, and no test based on them can have power. Thus, we henceforth assume that $\mu \neq 0$.*

Before examining the power properties of the CLR test, note that the power function for the Anderson-Rubin test is, from (49):

$$\mathcal{P}_k^{AR}(c_\beta^2 \Delta_n) = \exp\{-\frac{1}{2}c_\beta^2\Delta_n\} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2\Delta_n)^j}{j!} \{1 - G_{k+2j}(c_\alpha(k))\} . \tag{51}$$

This can be regarded as the power function of the LR test conditioned on both b and w and, for fixed α and k is easily seen to be monotonically increasing in $c_\beta^2\Delta_n$. Note that c_β^2 and Δ_n appear only as the product $c_\beta^2\Delta_n$.

Theorem 4 enables us to use exactly the same methods as were used above for the null case to compute the conditional power function of the CLR test, given p_2. The result is given in:

Theorem 5 *The conditional power function of the CLR test, given p_2, is given by:*

$$\Pr\{T > z_\alpha(k; w)|p_2\} = \exp\{-\frac{1}{2}c_\beta^2\Delta_n\} \sum_{j,l=0}^{\infty} \frac{\delta_1^j \delta_2^l}{j!l!2^{j+l}} \left\{1 - P_{k+2(j+l)}^{1+2l}(z_\alpha(k; w); w)\right\} . \tag{52}$$

From this result it is very easy to obtain bounds on the conditional power function of the CLR test. For, it is clear from equation (40) that, for all j, l, k, $z > 0$ and $w > 0$,

$$G_{k+2(j+l)}(z) < P_{k+2(j+l)}^{1+2l}(z; w) < G_{k+2(j+l)}(z + w).$$

This implies:

Proposition 9 *The conditional power function of the CLR test, given p_2, is bounded above by the function*

$$\begin{aligned}\mathcal{P}_k^U(c_\beta^2\Delta_n; w) &= \exp\{-\frac{1}{2}c_\beta^2\Delta_n\} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2\Delta_n)^j}{j!} \{1 - G_{k+2j}(z_\alpha(k; w))\} \\ &= \Pr\{\chi'^2(k, c_\beta^2\Delta_n) > z_\alpha(k; w)\},\end{aligned} \tag{53}$$

and below by the function

$$\begin{aligned}\mathcal{P}_k^L(c_\beta^2\Delta_n; w) &= \exp\{-\frac{1}{2}c_\beta^2\Delta_n\} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2\Delta_n)^j}{j!} \{1 - G_{k+2j}(w + z_\alpha(k; w))\} \\ &= \Pr\{\chi'^2(k, c_\beta^2\Delta_n) > w + z_\alpha(k; w)\}.\end{aligned} \tag{54}$$

Since these functions depend on p_2 only through w, they also bound the conditional power function given w.

It is also easy to see from the results in Section 2 that $z_\alpha(k;w) < c_\alpha(k) < w + z_\alpha(k;w)$, which implies:

$$\mathcal{P}_k^L(c_\beta^2\Delta_n;w) < \mathcal{P}_k^{AR}(c_\beta^2\Delta_n) < \mathcal{P}_k^U(c_\beta^2\Delta_n;w).$$

That is, the power function for the AR test also lies between these bounds. However, the AR and CLR tests are both admissable: we shall see below that their power functions cross under certain parameter configurations (see Chernozhukov, Hansen, and Jansson (2006) for theoretical results on this point). Some examples of these three functions are shown in Figure 8 for the cases $k = 3, 11$, $w = 1, 10$, and $\Delta_n = 1$. For larger values of Δ_n the pictures are similar but rise more sharply to one as $c_\beta^2\Delta_n$ increases. It is clear from these simple results that the CLR test cannot conditionally improve on the AR test when w - the sample information on β - is small, but that there is scope for considerable improvement when w is large.

To address the question more precisely we need to obtain the exact conditional power function, given w, rather than p_2. To do so, define (recall that we are assuming that $\mu \neq 0$) :

$$\left.\begin{array}{c} w_1 = p_2' M_\mu p_2 \sim \chi^2(k-1), \\ w_2 = p_2' P_\mu p_2 \sim \chi'^2(1, d_\beta^2\Delta_n), \end{array}\right\} \tag{55}$$

these variates being independent. Then $\delta_1 = c_\beta^2\Delta_n w_1/(w_1+w_2)$ and $\delta_2 = c_\beta^2\Delta_n w_2/(w_1+w_2)$. Setting $x = w_1/(w_1+w_2)$, $0 \le x \le 1$, and $w = w_1+w_2$, we have $\delta_1 = c_\beta^2\Delta_n x$ and $\delta_2 = c_\beta^2\Delta_n(1-x)$. Thus, the conditional power function given p_2 depends on p_2 only through (w, x). Note that w and x are analogues of q and b from Section 2. Hence:

Proposition 10 *When $\mu \neq 0$, the conditional power function given p_2 depends on p_2 only through w and x. The joint density of (x, w) is*:

$$\begin{aligned} pdf(x,w) &= \frac{\exp\{-\frac{1}{2}w\}w^{\frac{k}{2}-1}}{[B(\frac{k-1}{2},\frac{1}{2})]2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k-1}{2}-1}(1-x)^{\frac{1}{2}-1} \\ &\quad \times \exp\{-\frac{1}{2}d_\beta^2\Delta_n\}{}_0F_1(\frac{1}{2};\frac{1}{4}d_\beta^2\Delta_n w(1-x)), \end{aligned} \tag{56}$$

marginally $w \sim \chi'^2(k, d_\beta^2\Delta_n)$, and the conditional density of x given w is:

$$pdf(x|w) = \frac{x^{\frac{k-1}{2}-1}(1-x)^{\frac{1}{2}-1}}{[B(\frac{k-1}{2},\frac{1}{2})]}\left\{\frac{{}_0F_1(\frac{1}{2};\frac{1}{4}d_\beta^2\Delta_n w(1-x))}{{}_0F_1(\frac{k}{2};\frac{1}{4}d_\beta^2\Delta_n w)}\right\} \tag{57}$$

By expressing the conditional power function $\Pr\{T > z_\alpha(k;w)|p_2\}$ in terms of (x, w), multiplying by $pdf(x|w)$, and integrating out x, we obtain the conditional power function given w alone:

Theorem 6 *The conditional power function of the CLR test, given w, is:*

$$\begin{aligned}\mathcal{P}_k^{CLR}(\beta, \Delta_n; w) &= \exp\{-c_\beta^2 \Delta_n/2\} \sum_{j,l=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2 \Delta_n)^{j+l}}{j!l!} \\ &\times \left\{1 - P_{k+2(j+l)}^{1+2l}(z_\alpha(k;w);w)\right\} \mu_{j,l}(d_\beta^2 \Delta_n w), \end{aligned} \tag{58}$$

with

$$\begin{aligned}\mu_{j,l}(d_\beta^2 \Delta_n w) &= E_{x|w}[x^j(1-x)^l] \\ &= \frac{(\frac{k-1}{2})_j(\frac{1}{2})_l}{(\frac{k}{2})_{j+l}} \times \frac{{}_1F_2(l+\frac{1}{2}; j+l+\frac{k}{2}, \frac{1}{2}; \frac{1}{4}d_\beta^2 \Delta_n w)}{{}_0F_1(\frac{k}{2}; \frac{1}{4}d_\beta^2 \Delta_n w)}. \end{aligned} \tag{59}$$

It seems unlikely that the unconditional power function can be computed from this conditional result, because the critical values $z_\alpha(k;w)$ themselves depend on w.

The conditional power function for the statistic suggested by Kleibergen (2002), $LM = q_2 = (1-b)q$, and indeed other simple functions of (q_1, q_2), or, equivalently, (q, b), such as $f = q_2/q_1 = (1-b)/b$, can be computed by methods similar to those used for the CLR test. Conditional on p_2, $LM \sim \chi'^2(1, \delta_2)$, so that

$$\Pr(LM > c_\alpha(1)|p_2\} = \exp\{-\frac{1}{2}\delta_2\} \sum_{l=0}^{\infty} \frac{(\delta_2/2)^l}{l!} \{1 - G_{1+2l}(c_\alpha(1))\}.$$

Thus, setting $\delta_2 = c_\beta^2 \Delta_n(1-x)$, expanding the exponential term $\exp\{\frac{1}{2}c_\beta^2 \Delta_n x\}$, and averaging with respect to $pdf(x|w)$, as above, we obtain:

Theorem 7 *The conditional power function for the LM statistic $LM = q_2 = (1-b)q$, given w, is, for the case $m = 1$:*

$$\begin{aligned}\mathcal{P}_k^{LM}(\beta, \Delta_n; w) &= \exp\{-\frac{1}{2}c_\beta^2 \Delta_n\} \sum_{j,l=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2 \Delta_n)^{j+l}}{l!} \{1 - G_{1+2l}(c_\alpha(1))\} E_{x|w}[x^j(1-x)^l] \\ &= \exp\{-\frac{1}{2}c_\beta^2 \Delta_n\} \sum_{j,l=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2 \Delta_n)^{j+l}}{l!} \{1 - G_{1+2l}(c_\alpha(1))\} \mu_{j,l}(d_\beta^2 \Delta_n w). \end{aligned} \tag{60}$$

Here, it is straightforward to obtain the unconditional power function from this conditional result:

Theorem 8 *The unconditional power function for the LM test is given by:*

$$\mathcal{P}_k^{LM}(\beta, \Delta_n) = \exp\{-\frac{1}{2}c_\beta^2 \Delta_n\} \sum_{j,l=0}^{\infty} \frac{(\frac{1}{2}c_\beta^2 \Delta_n)^{j+l}}{l!} \{1 - G_{1+2l}(c_\alpha(1))\} \bar{\mu}_{j,l}(d_\beta^2 \Delta_n), \tag{61}$$

with

$$\begin{aligned}\bar{\mu}_{j,l}(d_\beta^2\Delta_n) &= E_w[\mu_{j,l}(d_\beta^2\Delta_n w)] \\ &= \frac{(\frac{k-1}{2})_j(\frac{1}{2})_l}{(\frac{k}{2})_{j+l}}\exp\{-\frac{1}{2}d_\beta^2\Delta_n\}{}_2F_2\left(l+\frac{1}{2},\frac{k}{2};j+l+\frac{k}{2},\frac{1}{2};\frac{1}{2}d_\beta^2\Delta_n\right).\end{aligned} \tag{62}$$

Power functions for various other statistics that are functions only of (b,q) are readily obtained by similar methods.

Remark 8 *Both $\mathcal{P}_k^{LM}(\beta,\Delta_n;w)$ and $\mathcal{P}_k^{LM}(\beta,\Delta_n)$ depend on c_β through both the term $c_\beta^2\Delta_n$, and the term $d_\beta^2\Delta_n = \Delta_n(1-\rho c_\beta)^2/(1-\rho^2)$, as does the function $\mathcal{P}_k^{CLR}(\beta,\Delta_n;w)$. As a consequence, in contrast to the power function of the AR test, these functions are not symmetric about zero, and they exhibit noticeable non-monotinicity in c_β near the point $c_\beta=\rho^{-1}$ (at which $d_\beta^2=0$) for some parameter configurations. This was noted by Andrews, Moreira, and Stock (2006), and is a precise analogue of the bimodality of the density of the two stage least squares estimator for β that has been remarked on in the literature (see, for instance, Phillips (2006), Stock, Wright, and Yogo (2002), and Hillier (1990) and (2006a), and the references therein).*

5.2 Power Comparisons

The conditional power functions for the CLR test, the AR test, and the LM test evidently depend on the known number of instruments, k, the assumed - known degree of endogeneity, ρ, the unknown concentration parameter Δ_n, and the observed sample information on β, w, as well as the interest parameter c_β. Although there is no uniformly best similar test, by studying the conditional power functions derived here, a clear preference ordering can be established.

Some examples of the power functions, computed from the formulae above, are depicted in Figures 9 - 12. In Figures 9 and 10 $k=3$ and $\Delta_n=1$ and 5, respectively. The values of w chosen are $w=1$, indicating a low value of the sample information, and $w=E(w)=k+\Delta_n/(1-\rho^2)$, the mean of w when $c_\beta=0$. When $E(w)>25$ w is set at 25. These values therefore represent an "average" sample information figure, given the other relevant parameters. In Figures 11 and 12 $k=12$ and again $\Delta_n=1$ and $\Delta_n=5$, respectively, and the same choices are made for w. From these, and more extensive study of the (conditional) power functions not reproduced, the following conclusions emerge:

1. None of these tests uniformly dominates the other two, nor is any of the tests inadmissable relative to the others.

2. When the observed information on β, w, is low, the power properties of the AR and CLR tests are almost indistinguishable, whatever k, Δ_n, and ρ, and both (almost) uniformly dominate the LM test. The LM test can be conditionally better than both the CLR and AR tests when $\rho c_\beta < 0$ and ρ is large, but when it is the margin is slight. When ρ is only moderate, the power *disadvantage* of the LM test can be substantial.

3. When w is larger the relationships between the power functions are more complex. For $\rho c_\beta < 0$, the CLR and LM tests are almost indistinguishable, and both dominate the AR test, whatever k, Δ_n, and ρ. For $\rho c_\beta > 0$ the power curves are distorted by the kink at $c_\beta = \rho^{-1}$, but whatever the configuration of the other parameters, the CLR test is either the most powerful test, or close to being so.

In view of these results it is safe to conclude that, of these three tests, the CLR test is the preferred test. This conclusion is, of course, based on comparisons of *conditional* power properties of the CLR and LM tests, given w, and we would argue that, since there is clear evidence here that the observed information *is* pertinent to the properties of the tests, it makes no sense to average over outcomes for w that have not occurred. Nevertheless, if one insists on comparisons based on unconditional power, these results still support the conclusion that the CLR test is to be preferred: the averaging required to produce the unconditional power functions from the reported conditional functions must preserve this ordering.

As a referee has pointed out, it would be interesting to extend the power function calculations to the case $m > 1$, and in particular to compare the power of the approximate CLR test introduced in Section 4 to that of the true CLR test. As noted earlier, the results in Hillier (2006b) mean that this exercise is now feasible. It is, however, beyond the scope of the present paper.

6 Concluding Comments

A satisfactory theory of hypothesis testing for the structural equation/IV regression model has been sorely lacking in econometrics. The recent upsurge of interest in this model has sparked renewed interest in the testing problem, and papers by Moreira (2003), Kleibergen (2002), (2005), Andrews, Moreira and Stock (2006), and Chamberlain (2005), have made some progress on the problem for the case where the null hypothesis specifies all coefficients of right-hand-side endogenous variables, and where the reduced form covariance matrix is known (or also specified by the null). In particular, a class of similar tests for this context can be characterized, and a conditional version of the likelihood ratio test can be shown to be a member of that class.

This paper provides some key results needed to implement these results, namely, the exact conditional distribution function of the LR statistic, and, more importantly, the implicit critical value function needed to render the LR test exact. In the case of tests for more than one coefficient the results apply to an approximation to the LR test, rather than the true LR test. The tabulations provided are certainly not complete, but should cover many situations met by applied workers.

Although such results represent a significant step forward, many important and difficult problems remain to be solved before we can claim to have a complete theory of testing for this model. Specifically, we need to develop methods of comparable effectiveness for the more general cases where the hypothesis specifies only a subvector of β, and/or the case where the covariance matrix is unknown.

In the first case - testing a subvector of β, but with known covariance matrix, the null hypothesis restricts the density of y_1 in (1) by asserting that $E(y_1)$ depends on only a submatrix of Π, rather than the full matrix. This restriction is not strong enough to ensure that the analogue of p_1 has mean zero under the null, which suggests that (exact) similarity may be unattainable in this more realistic problem. Likewise, when Ω is unknown, the transformation matrix U_Ω in (8) is unknown, and one cannot transform to achieve independence between the columns of what has here been called P, the standardised sufficient statistic, although of course this can be achieved asymptotically. Thus, the conditional mean $E(p_1|P_2)$ is no longer free of P_2 (zero) under the null, and again exact similarity seems likely to be unattainable (though asymptotic similarity will be). This, incidentally, explains why power results under the known-covariance assumption have much in common with results under the null when the covariance matrix is unknown.

In Hillier (1987a) and Hillier (2005) I analyse the most general model and testing problems - testing subvectors of parameters, including the coefficients of both endogenous and/or exogenous variables, with an unknown reduced form covariance matrix - from an invariance point of view. Likelihood ratio tests are members of the class of invariant tests, but the challenge remains to provide accurate critical values for the LR test (conditional or otherwise) that can be used in practice, and that are not rendered inaccurate in situations where the instruments are weak.

Notes

[1]One way to achieve similarity for the LR test would be to condition on both b and w, which would imply choosing the critical value z so that $z(1-ab)^{-1} = c_\alpha(k)$, where $c_\alpha(k)$ is such that $\Pr\{\chi^2(k) > c_\alpha(k)\} = \alpha$. But this is equivalent to a test based on q - the Anderson-Rubin statistic - alone. That is, when conditioned on both b and w, the similar version of the LR test is equivalent to the Anderson-Rubin test. This approach would therefore yield nothing new.

[2]The expectation on the right in equation (31) can be evaluated analytically, but

is rather messy. The simulation approach used here is a considerably simpler way to compute the bounding size function of interest.

[3]The inequality $f_1^*(\lambda_1) \leq f_1$ (in quite different notation) was also asserted by Kleibergen (2005), who also sought an approximate version of the LR test, for similar reasons, but his original proof of the result contained an error.

7 Appendix: Proofs of Main Results

Proof of Theorem 1:

We make use of the following well-known result for the *cdf* of the Chi-square distribution:

Lemma 2 *(Abramowitz and Stegun (1972), Section 6.5) The cdf of the $\chi^2(k)$ distribution, $G_k(c) = \Pr\{\chi^2(k) < c\}$, is given by:*

$$G_k(c) = \frac{(c/2)^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)} {}_1F_1(\frac{k}{2}, \frac{k}{2}+1; -\frac{c}{2}). \tag{63}$$

For fixed (b, w) and k it follows from the Lemma that:

$$\begin{aligned}
\Pr\{T &< z|(b,w)\} = \Pr\{q < z(1-ab)^{-1}|(b,w)\} \\
&= \frac{(z/2)^{\frac{k}{2}}(1-ab)^{-\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)} {}_1F_1(\frac{k}{2}, \frac{k}{2}+1; -\frac{z}{2}(1-ab)^{-1}).
\end{aligned} \tag{64}$$

On multiplying by *pdf*(b) and integrating with respect to b we find:

$$\begin{aligned}
P_k(z; w) &= \frac{(z/2)^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)} \sum_{j=0}^{\infty} \frac{(\frac{k}{2})_j(-z/2)^j}{j!(\frac{k}{2}+1)_j} {}_2F_1\left(j+\frac{k}{2}, \frac{k-1}{2}; \frac{k}{2}; a\right) \\
&= \frac{(z/2)^{\frac{k}{2}}(1-a)^{-\frac{k-1}{2}}}{\Gamma(\frac{k}{2}+1)} \sum_{j=0}^{\infty} \frac{(\frac{k}{2})_j\left(-\frac{z}{2(1-a)}\right)^j}{j!(\frac{k}{2}+1)_j} {}_2F_1(-j, \frac{1}{2}; \frac{k}{2}; a) \\
&= \frac{(z/2)^{\frac{k}{2}}(1-a)^{-\frac{k-1}{2}}}{\Gamma(\frac{k}{2}+1)} \sum_{j=0}^{\infty} \frac{(\frac{k}{2})_j\left(-\frac{w+z}{2}\right)^j}{j!(\frac{k}{2}+1)_j} \sum_{l=0}^{j} \binom{j}{l} \frac{(\frac{1}{2})_l}{(\frac{k}{2})_l}(-a)^l \\
&= \frac{(z/2)^{\frac{k}{2}}(1-a)^{-\frac{k-1}{2}}}{\Gamma(\frac{k}{2}+1)} \sum_{j,l=0}^{\infty} \frac{(\frac{k}{2})_{j+l}(\frac{1}{2})_l}{j!l!(\frac{k}{2}+1)_{j+l}(\frac{k}{2})_l} \left(-\frac{w+z}{2}\right)^j (\frac{w}{2})^l \\
&= \frac{(z/2)^{\frac{k}{2}}(1-a)^{-\frac{k-1}{2}}}{\Gamma(\frac{k}{2}+1)} \sum_{l=0}^{\infty} \frac{(w/2)^l(\frac{1}{2})_l}{l!(\frac{k}{2}+1)_l} {}_1F_1\left(\frac{k}{2}+l, \frac{k}{2}+l+1; -\frac{(w+z)}{2}\right) \\
&= (1-a)^{\frac{1}{2}} \sum_{l=0}^{\infty} \frac{a^l(\frac{1}{2})_l}{l!} G_{k+2l}(w+z).
\end{aligned} \tag{65}$$

These steps depend on the following standard results/techniques: (1) Gauss's identity for the hypergeometric function:

$$ {}_2F_1(a;b,c;z) = (1-z)^{c-a-b}{}_2F_1(c-a;c-b,c;z), \tag{66}$$

(line 1 $\rightarrow$ line 2), (2) the observation that the series ${}_2F_1(-j,b,c;z)$ terminates at the $j-th$ term,

$$ {}_2F_1(-j,b,c;z) = \sum_{l=0}^{j} \binom{j}{l} \frac{(b)_l}{(c)_l}(-z)^l, \tag{67}$$

(line 2 $\rightarrow$ line 3), and (3) the device of summing double series 'by diagonals':

$$\sum_{j,l=0}^{\infty} c(j,l) = \sum_{j=0}^{\infty}\sum_{l=0}^{j} c(j-l,l), \tag{68}$$

(line 3 $\rightarrow$ line 4). The final step uses Lemma 1 in reverse.

Proof of Proposition 3:

The first statement follows from equation (19) and the fact that $z(1-ab)^{-1} \leq z(1-b)^{-1}$ for all $w > 0$ and all b, together with Lemma 3 below, which asserts that $E_b[G_k(z(1-b)^{-1})] = G_1(z)$. The second result can be obtained directly by using standard results on the asymptotic behaviour of the confluent hypergeometric function (Abramowitz and Stegun (1972), Chapter 13, Slater (1960)). But, since $a \rightarrow 1$ as $w \rightarrow \infty$ for all finite $z > 0$, this also follows from equation (19) and Lemma 3 below.

Proof of Proposition 4:

The function $G_k(c)$ satisfies the recursion:

$$G_{k+2}(c) = G_k(c) - \frac{(\frac{c}{2})^{\frac{k}{2}} \exp\{-\frac{c}{2}\}}{\Gamma(\frac{k}{2}+1)} \tag{69}$$

(Abramowitz and Stegun (1972), Section 26.4.8). Using this in the expression for $P_{k+2}(z;w)$ we have:

$$\begin{aligned} P_{k+2}(z;w) &= (1-a)^{\frac{1}{2}} \sum_{l=0}^{\infty} \frac{a^l(\frac{1}{2})_l}{l!} G_{k+2+2l}(w+z) \\ &= (1-a)^{\frac{1}{2}} \sum_{l=0}^{\infty} \frac{a^l(\frac{1}{2})_l}{l!} \left\{ G_{k+2l}(w+z) - \frac{(\frac{w+z}{2})^{\frac{k}{2}+l} \exp\{-\frac{w+z}{2}\}}{\Gamma(\frac{k}{2}+l+1)} \right\} \\ &= P_k(z;w) - \frac{z^{\frac{1}{2}}(w+z)^{\frac{k-1}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2}+1)} \exp\{-\frac{w+z}{2}\}{}_1F_1(\frac{1}{2}, \frac{k}{2}+1; \frac{w}{2}). \end{aligned}$$

Proof of Theorem 2:

We first obtain a simple result that is of interest in its own right. We prove the result directly, but it also follows from the fact that, if $b \sim Beta(\frac{k-1}{2}, \frac{1}{2})$ is independent of $q \sim \chi^2(k)$, then $(1-b)q \sim \chi^2(1)$ (see Section 2).

Lemma 3 *If $b \sim Beta(\frac{k-1}{2}, \frac{1}{2})$,*

$$E_b\left[G_k\left(\frac{z}{1-b}\right)\right] = G_1(z). \tag{70}$$

Proof. We have

$$\begin{aligned} E_b\left[G_k\left(\frac{z}{1-b}\right)\right] &= [2^{\frac{k}{2}}\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{1}{2}\right)]^{-1} \\ &\times \int_0^1 \int_0^{\frac{z}{1-b}} \exp\{-\frac{1}{2}y\}y^{\frac{k}{2}-1}b^{\frac{k-1}{2}-1}(1-b)^{\frac{1}{2}-1}dydb. \end{aligned} \tag{71}$$

Replace y by $\tilde{y} = y(1-b)$ (Jacobian $(1-b)^{-1}$). The integral here becomes

$$\int_0^1 \int_0^z \exp\{-\frac{1}{2}\frac{y}{1-b}\}y^{\frac{k}{2}-1}b^{\frac{k-1}{2}-1}(1-b)^{-\frac{k+1}{2}}dydb.$$

Now put $f = b/(1-b)$ ($0 < f < \infty$, Jacobian $(1+f)^{-2}$), giving

$$\int_0^\infty \int_0^z \exp\{-\frac{1}{2}y(1+f)\}y^{\frac{k}{2}-1}f^{\frac{k-1}{2}-1}dydf.$$

Interchanging the order of integration, and evaluating the integral over $0 < f < \infty$ gives

$$[2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)]^{-1}\int_0^z \exp\{-\frac{1}{2}y\}y^{\frac{1}{2}-1}dy = G_1(z), \tag{72}$$

establishing the result. ■

Now, the expectation with respect to w in equation (25) can be regarded as the expectation with respect to the conditional density of U given b. And if, for fixed b, the random variable U in (26) satisfies $\underline{u} < U < \bar{u}$ as w varies over $w \geq 0$, we would have

$$\Pr\{T < z|b\} = E_w\left[G_k\left(U(w,b)\right)\right] = \int_{\underline{u}}^{\bar{u}} pdf_U(u|b)G_k(u)du, \tag{73}$$

where $pdf_U(u|b)$ is the conditional density of U for fixed b. Integrating by parts,

$$\Pr\{T < z|b\} = F_U(u|b)G_k(u)|_{\underline{u}}^{\bar{u}} - \int_{\underline{u}}^{\bar{u}} F_U(u|b)g_k(u)du, \tag{74}$$

where $F_U(u|b)$ is the conditional *cdf* of U for fixed b. But, for fixed $b \in (0,1)$, the function $U(w,b)$ is monotonically increasing in w, and satisfies

$$z \leq U(w,b) < \frac{z}{1-b}, \tag{75}$$

so that, for u in this interval, the inequality $U(w,b) < u$ corresponds to the inequality

$$w < \frac{z(u-z)}{(z-u(1-b))}. \tag{76}$$

Thus $\underline{u} = z$ and $\bar{u} = z/(1-b)$, and, since under the null hypothesis $w \sim \chi'^2(k, \Delta_n)$, with noncentrality parameter Δ_n proportional to $i_n(\beta)$ in (12),

$$\begin{aligned} F_U(u|b) &= \Pr\left\{\chi'^2(k;\Delta_n) < \frac{z(u-z)}{(z-u(1-b))}\right\} \\ &= G_k^{\Delta_n}\left(\frac{z(u-z)}{(z-u(1-b))}\right) \quad (z \leq u < \frac{z}{1-b}). \end{aligned} \tag{77}$$

Hence,

$$\Pr\{T < z|H_0\} = E_b\left[G_k\left(\frac{z}{1-b}\right) - \int_z^{\frac{z}{1-b}} G_k^{\Delta_n}\left(\frac{z(u-z)}{(z-u(1-b))}\right) g_k(u)du\right]. \tag{78}$$

Applying the result in the Lemma above completes the proof. ■

Proof of Corollary 1 (informal):

(1) This follows simply from equation (27) and the well-known fact that for each k and c, $G_k^{\Delta}(c) \to 0$ as $\Delta \to \infty$ (see Ghosh (1970), for instance). The second term in (27) therefore vanishes as $\Delta_n \to \infty$.
(2) The first inequality follows directly from equation (27), the second from (27) and the fact that $G_k^{\Delta_n}(\cdot)$ is decreasing in Δ_n.
(3) Since $G_k^{\Delta_n}(\cdot) \leq 1$, in the conditional formula, given b, the second term

$$\int_z^{\frac{z}{1-b}} G_k^{\Delta_n}\left(\frac{z(u-z)}{(z-u(1-b))}\right) g_k(u)du \leq \int_z^{\frac{z}{1-b}} g_k(u)du = G_k\left(\frac{z}{1-b}\right) - G_k(z).$$

This immediately gives the stated result.

Proof of Lemma 1:

There is an orthogonal $m \times m$ matrix H such that $W_{22} = HD^2H'$, where $D = diag\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_m}\}$. Define $r = D^{-1}H'w_{21}$, and note that $q_1 = q - r'r$, and that the characteristic roots of W are those of

$$\tilde{W} = \begin{bmatrix} q, & r'D \\ Dr, & D^2 \end{bmatrix}. \tag{79}$$

The characteristic polynomial of $\tilde{W}$ is

$$\begin{aligned}\tilde{p}(f) &= \det[\tilde{W} - fI_{m+1}] \\ &= (q-f)\Pi_{i=1}^{m}(\lambda_i - f) - \sum_{i=1}^{m}\lambda_i r_i^2\left[\prod_{j\neq i}(\lambda_j - f)\right]. \end{aligned} \tag{80}$$

Now, assume that all of the roots f_j differ from the λ_i. Then, for a particular value of j,

$$\tilde{p}(f_j) = [\Pi_{i=1}^{m}(\lambda_i - f_j)](q - f_j - g(f_j)), \tag{81}$$

where

$$g(f) = \sum_{i=1}^{m}\frac{r_i^2\lambda_i}{\lambda_i - f}. \tag{82}$$

Hence, in particular, since $\tilde{p}(f_1) = 0$, $q = f_1 + g(f_1)$.

To show that $f_1^*(\lambda_1) \leq f_1$ it is enough to show (because of the shape of $p_\lambda(f)$) that $p_{\lambda_1}(f_1) \leq 0$. But, using the above identity,

$$\begin{aligned} p_{\lambda_1}(f_1) &= f_1^2 - f_1(\lambda_1 + q) + \lambda_1 q_1 \\ &= f_1^2 - f_1(\lambda_1 + f_1 + g(f_1)) + \lambda_1(f_1 + g(f_1) - r'r) \\ &= (\lambda_1 - f_1)g(f_1) - \lambda_1 r'r \\ &= f_1\sum_{i=2}^{m}\frac{r_i^2(\lambda_1 - \lambda_i)}{(\lambda_i - f_1)}. \end{aligned} \tag{83}$$

Now $\lambda_1 \leq \lambda_i$ for all $i > 1$, and it is well known that $f_1 \leq \lambda_1 \leq \lambda_i$, so that the coefficients in the sum here are non-positive for all i. If any of the f_j is equal to a λ_i the fact that $p_{\lambda_1}(f_1) \leq 0$ is obvious. The result follows.

Proof of Theorem 3:

The proof is virtually identical to that of Theorem 1, beginning from equation (19), but with b now distributed as $Beta(\frac{k-m}{2}, \frac{m}{2})$. We omit the details.

Proofs of Propositions 6 and 7:

The Proofs of Propositions 6 and 7 are exactly analogous to those of Propositions 3 and 4 and are omitted.

Proof of Proposition 8 and Theorem 4:

These follow from standard properties of quadratic forms in noncentral normal variables, followed by elementary manipulations. We omit the details.

Proof of Theorem 5:

From the results in Theorem 4 for the conditional density $pdf(q|b,p_2)$, we compute:

$$\Pr\{q < z(1-ab)^{-1}|b,p_2\} = \left\{ \frac{\sum_{j,l=0}^{\infty} \frac{(\delta_1 b)^j(\delta_2(1-b))^l(\frac{k}{2})_{j+l}}{j!l!2^{j+l}(\frac{k-1}{2})_j(\frac{1}{2})_l} G_{k+2(j+l)}(z(1-ab)^{-1})}{\sum_{j,l=0}^{\infty} \frac{(\delta_1 b)^j(\delta_2(1-b))^l(\frac{k}{2})_{j+l}}{j!l!2^{j+l}(\frac{k-1}{2})_j(\frac{1}{2})_l}} \right\}. \tag{84}$$

Multiplying this by

$$pdf(b|p_2) = \frac{exp\{-\frac{1}{2}c_\beta^2\Delta_n\}}{B(\frac{k-1}{2},\frac{1}{2})} b^{\frac{k-1}{2}-1}(1-b)^{-\frac{1}{2}} \sum_{j,l=0}^{\infty} \frac{(\delta_1 b)^j(\delta_2(1-b))^l(\frac{k}{2})_{j+l}}{j!l!2^{j+l}(\frac{k-1}{2})_j(\frac{1}{2})_l},$$

and integrating out b, we have

$$\Pr\{T < z|p_2\} = exp\{-\frac{1}{2}c_\beta^2\Delta_n\} \sum_{j,l=0}^{\infty} \frac{(\delta_1/2)^j(\delta_2/2)^l}{j!l!} E_{b(j,l)}\left[G_{k+2(j+l)}(z(1-ab)^{-1})\right], \tag{85}$$

where $E_{b(j,l)}$ denotes the expectation with respect to a $Beta(j+\frac{k-1}{2}, l+\frac{1}{2})$ variate. But, by definition,

$$E_{b(j,l)}\left[G_{k+2(j+l)}(z(1-ab)^{-1})\right] = P_{k+2(j+l)}^{1+2l}(z;w), \tag{86}$$

which gives the result.

Proof of Proposition 9:

The upper bound follows from the fact that, for all $w > 0$, and all $z > 0$,

$$P_{k+2(j+l)}^{1+2l}(z;w) > G_{k+2(j+l)}(z).$$

Using this in equation (52) with $z = z_\alpha(k;w)$ gives the upper bound, since, summing by diagonals,

$$\begin{aligned} \sum_{j,l=0}^{\infty} \frac{\delta_1^j\delta_2^l}{j!l!2^{j+l}}\left\{1-G_{k+2(j+l)}(z)\right\} &= \sum_{j=0}^{\infty} \frac{\{1-G_{k+2j}(z)\}}{j!2^j} \sum_{l=0}^{j} \binom{j}{l} \delta_1^{j-l}\delta_2^l \\ &= \sum_{j=0}^{\infty} \frac{(c_\beta^2\Delta_n/2)^j}{j!}\left\{1-G_{k+2j}(z)\right\}, \end{aligned} \tag{87}$$

because $\delta_1 + \delta_2 = c_\beta^2\Delta_n$. The lower bound is obtained in a similar way.

Proofs of Proposition 10 and Theorem 6:

These are again straightforward applications of standard results, followed by elementary manipulations, and are omitted.

Proofs of Theorems 7 and 8:

These are straightforward, and omitted.

References

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

Anderson, T.W., and H. Rubin (1949), "Estimation.of the Parameters.of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, **20**, 46 - 63.

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., and J. Stock, (2005), "Inference with Weak Instruments", Invited paper presented at the Econometric Society World Congress, London, 2005, and mimeo, Harvard University.

Breusch, T. (1986), "Hypothesis Testing in Unidentified Models", *Review of Economic Studies*, **53**, 4 (Special issue on Econometrics).

Chao, J.C. and N.R. Swanson (2005), "Consistent estimation with a large number of weak instruments", *Econometrica*, **73**, 1, 1673-1692.

Chamberlain, G. (2005), "Decision theory applied to an instrumental variables model", mimeo, Harvard University.

Chernozhukov, V., C. Hansen, and M. Jansson, (2006), "Admissible Invariant Similar Tests for Instrumental Variables Regression", Manuscript, M.I.T.,

Forchini, G, and G.H. Hillier, (2003), "Conditional Inference for Possibly Unidentified Structural Equations", *Econometric Theory*, **19**, 707-743.

Ghosh, B.K. (1970), *Sequential Tests of Statistical Hypotheses,* Addison-Wesley, Reading, Mass.

Han, C., and Phillips, P.C.B., (2006), "GMM with many moment conditions", *Econometrica*, **74**, 1, 147-192.

Hansen, C., Hausman, J., and Newey, W., (2006), "Estimation with many instrumental variables", CeMMAP Working Paper CWP19/06, September 2006.

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. (2006a), "Yet More on the Exact Properties of IV Estimators", *Econometric Theory,* **22**, 913-931.

Hillier, G.H. (2006b), "On the conditional likelihood ratio test for several parameters in IV Regression", CeMMAP working Paper CW/2607, October 2006.

Kleibergen, F. (2002), "Pivotal statistics for testing Structural Parameters in Instrumental Variables Regression", *Econometrica*, **70,** 5, 1781-1803.

Kleibergen, F. (2005), "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics", *Journal of Econometrics*, forthcoming.

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

Phillips, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equation Model", in M.D. Intriligator and Z. Griliches, eds., *Handbook of Econometrics*, Amsterdam: North Holland.

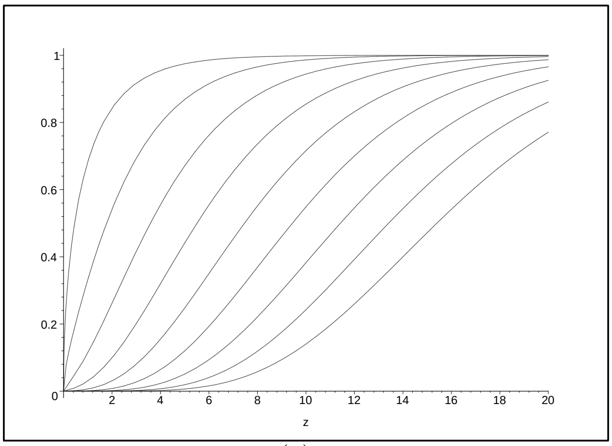
Phillips, P.C.B., (2006), "A Remark on Bimodality and Weak Instrumentation in Structural Equation Estimation", *Econometric Theory*, **22**, 947-960.

Phillips, P.C.B. (1989), "Partially Identified Econometric Models", *Econometric Theory*, 5, 181-240.

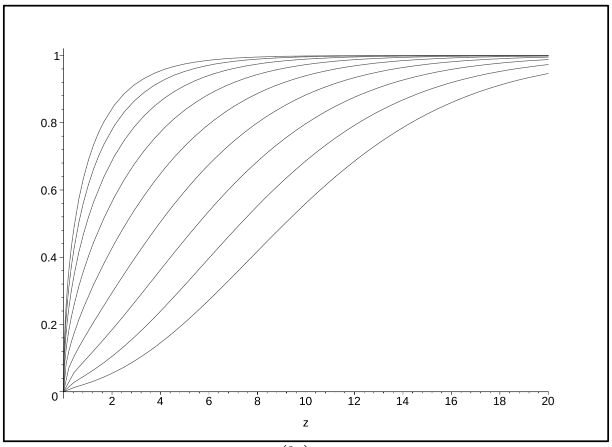
Slater, L.J. (1960), *Confluent Hypergeometric Functions,* Cambridge University Press, Cambridge, England.

Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, **65**, 557-586.

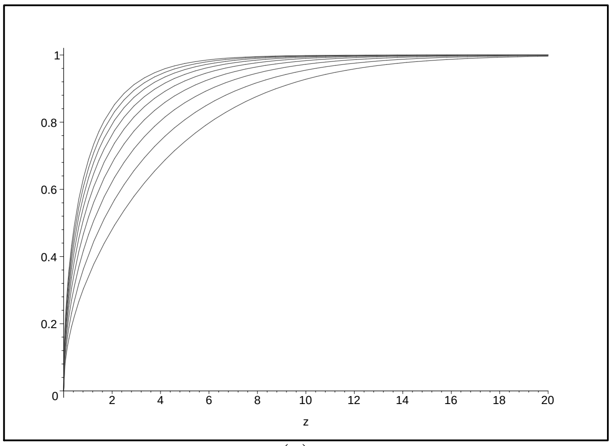
Stock, J.H., J.H. Wright, and M. Yogo, (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business and Economic Statistics,* ***20****, 518-529.*



(a)



(b)



(c)

Figure 1: Conditional Distribution Function of the LR statistic given w; k odd, $k = 1, 3, .., 17$. (a) $w = 1$, (b) $w = 8$, (c) $w = 20$. The case $k = 1$ is on the left, the *cdf* moving to the right as k increases.

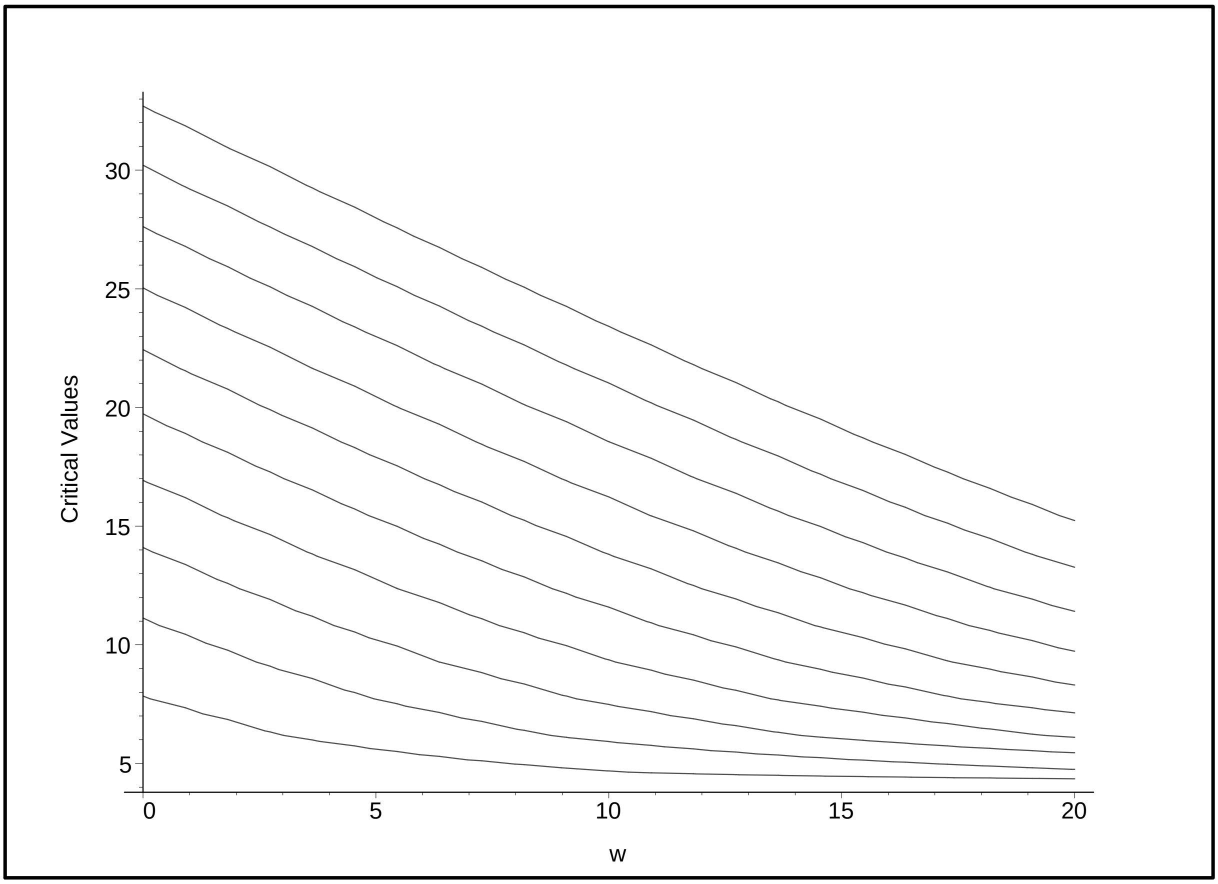


Figure 2: Critical Value Function for the CLR test given w; k odd, $k = 3, .., 21$. The lowest line is the case $k = 3$, the topmost $k = 21$.

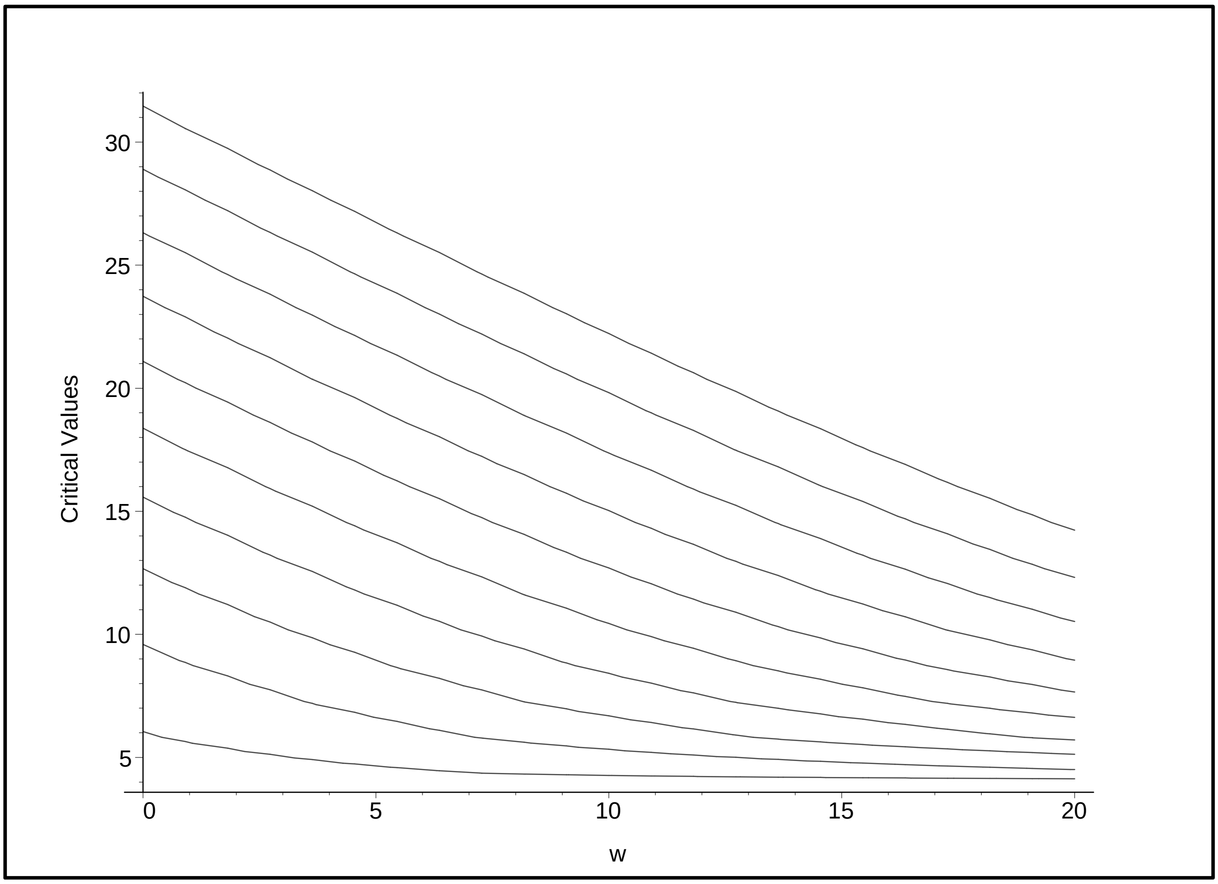


Figure 3: Critical Value Function for the CLR test given w; k even, $k = 2, 4, .., 20$. The lowest line is the case $k = 2$, the topmost $k = 20$.

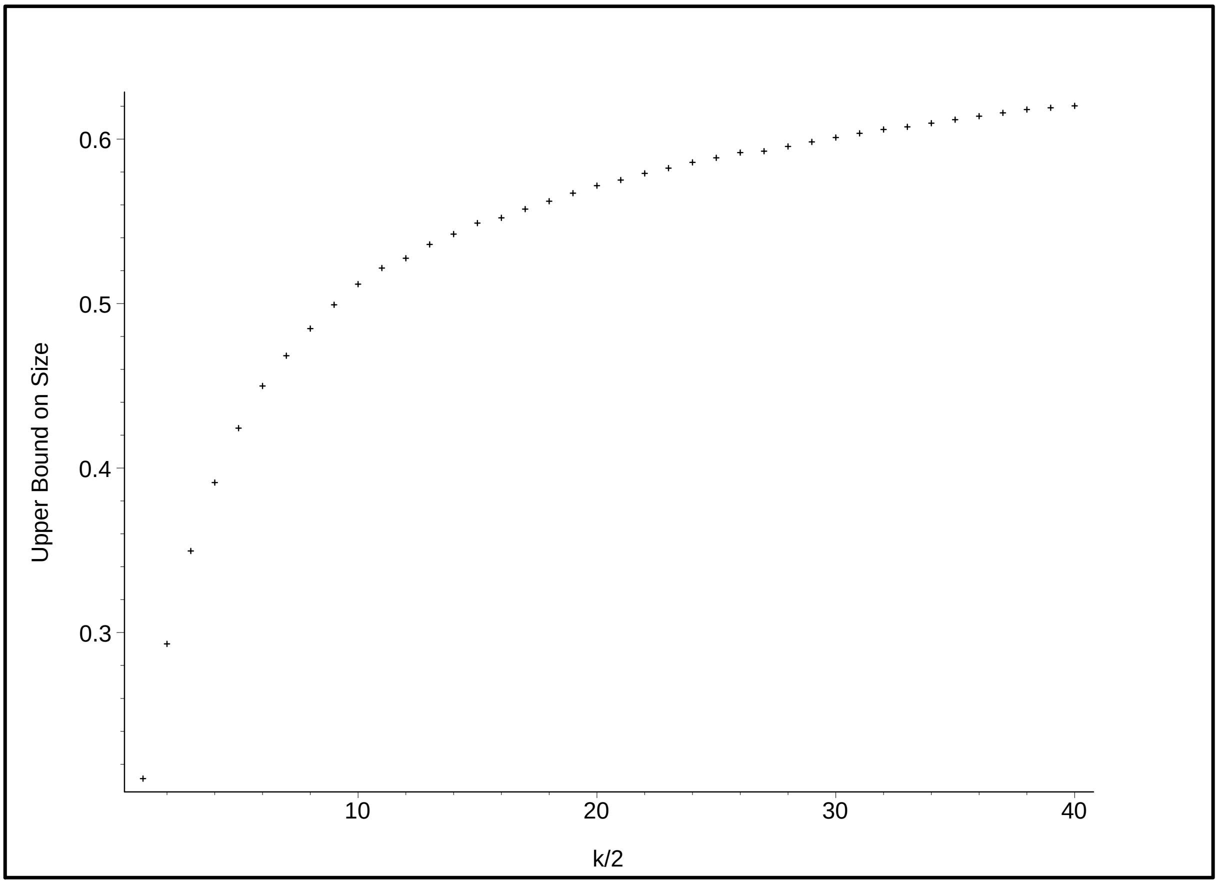
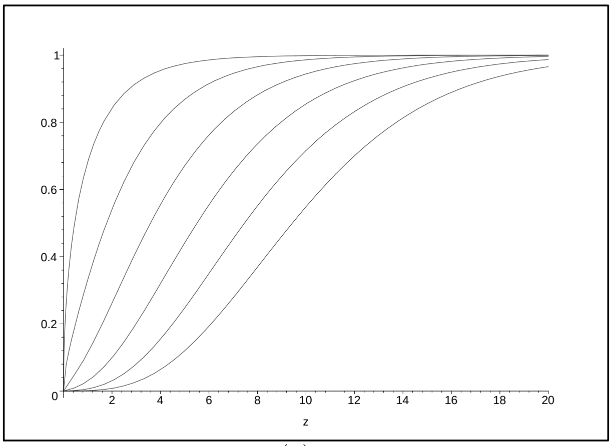
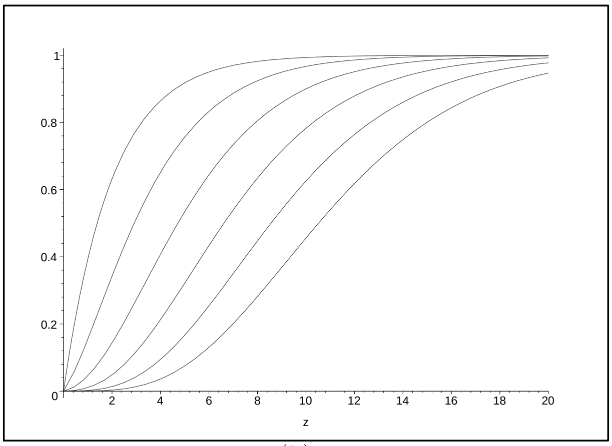


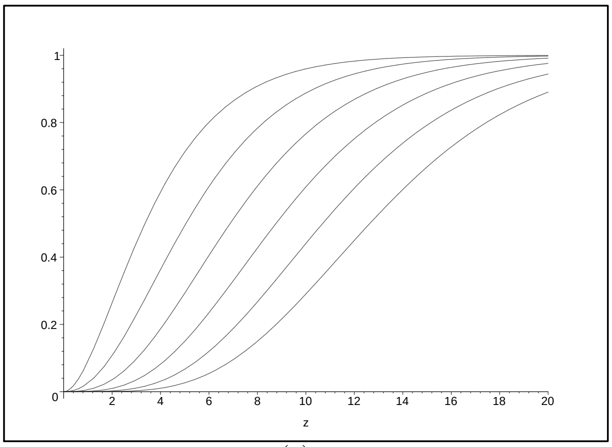
Figure 4: Upper Bound on the Size of the LR test; $k = 2, .., 82$.



(a)

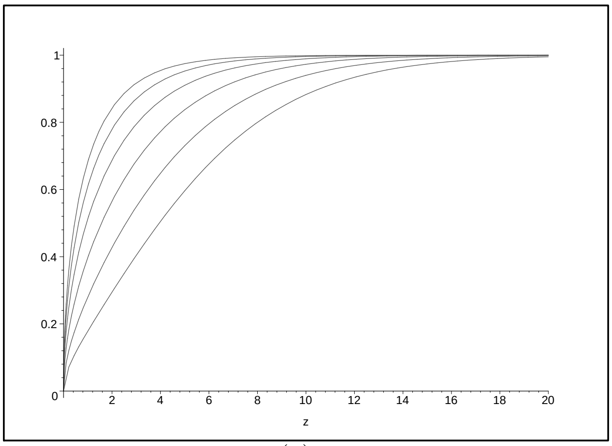


(b)

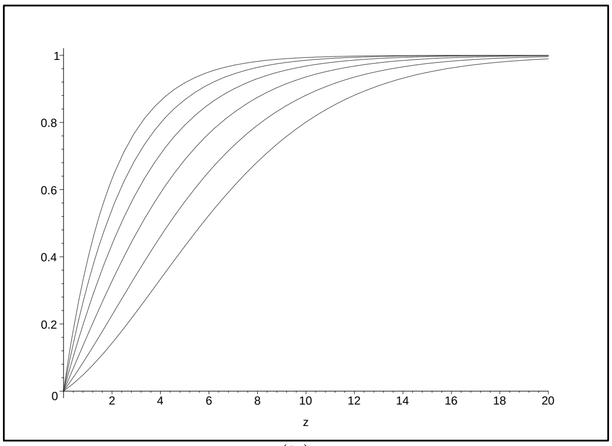


(c)

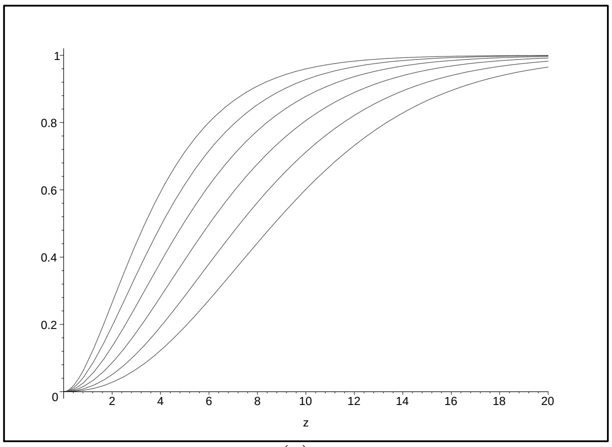
Figure 5: Generalised Conditional CDF given λ_1: $\lambda_1 = 1$; $k = m, m+2, .., m+10$. (a) $m = 1$, (b) $m = 2$, (c) $m = 4$. The *cdf* moves to the right as k increases.



(a)

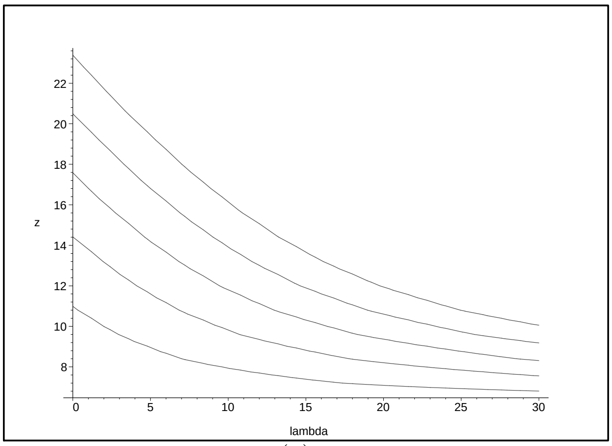


(b)

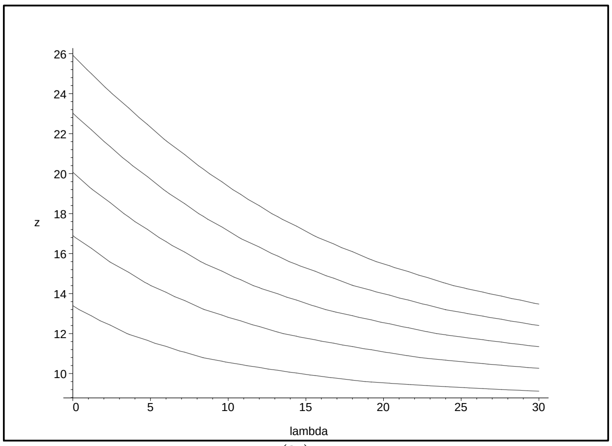


(c)

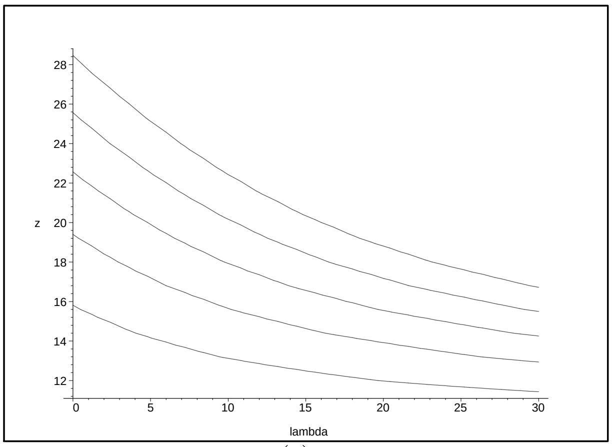
Figure 6: Generalised Conditional CDF given λ_1: $\lambda_1 = 8$; $k = m, m+2, .., m+10$. (a) $m = 1$, (b) $m = 2$, (c) $m = 4$. The *cdf* moves to the right as k increases.



(a)



(b)



(c)

Figure 7: Critical Value Function for the Generalized CLR test; $k = m+2, .., m+10$. (a) $m = 2$, (b) $m = 3$, (c) $m = 4$. The higher lines correspond to larger values of k.

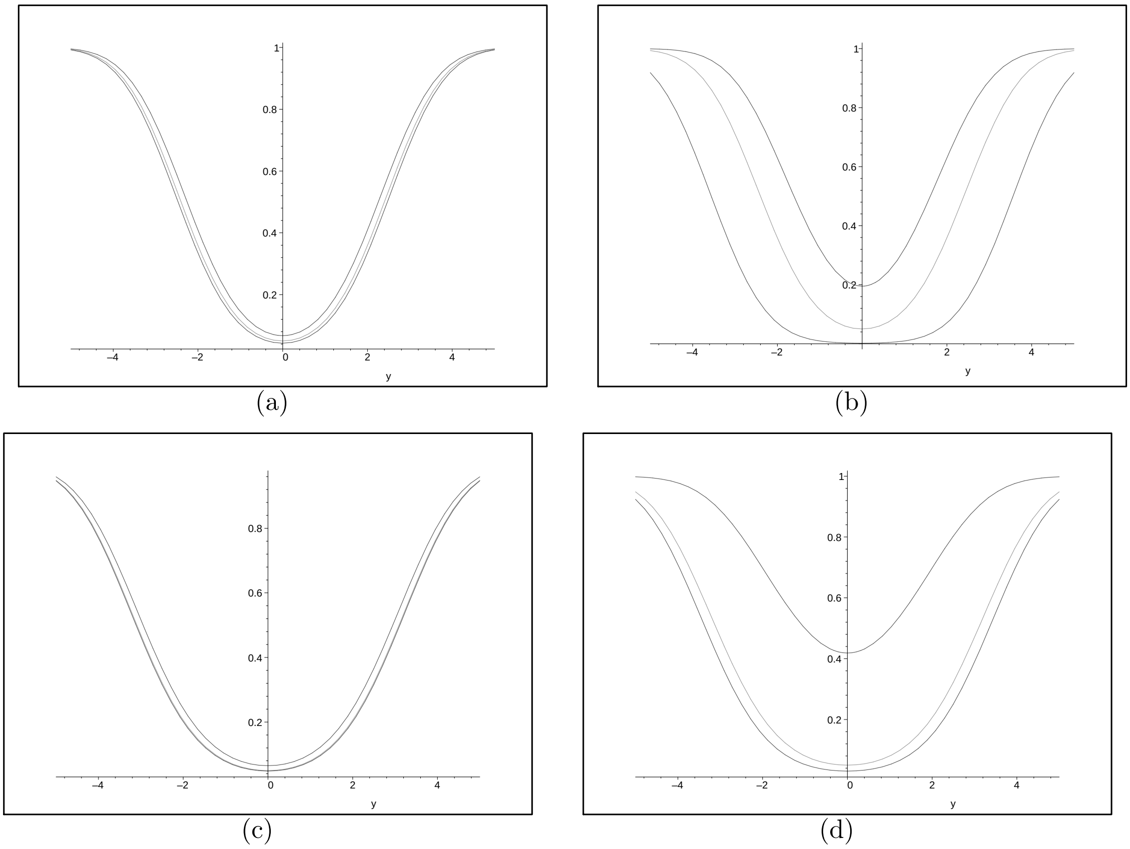
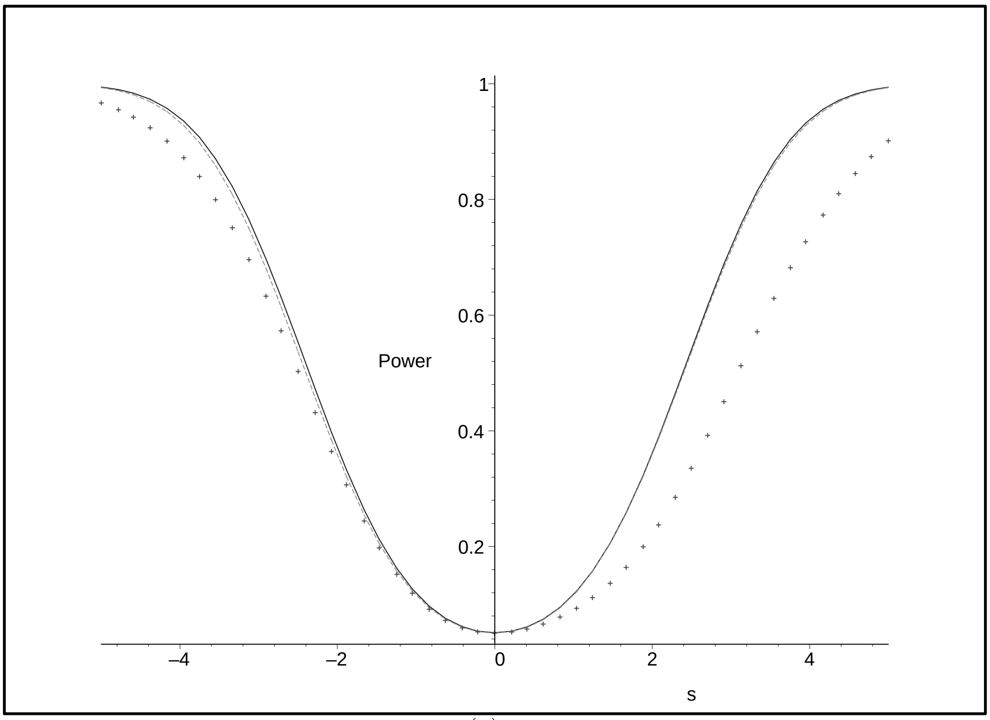
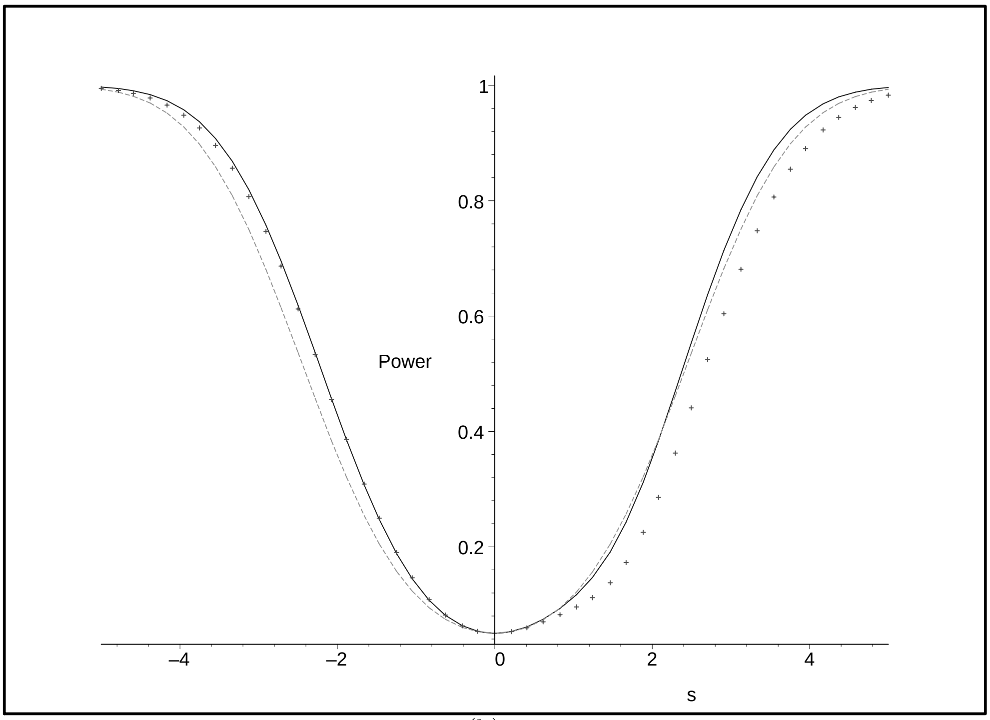


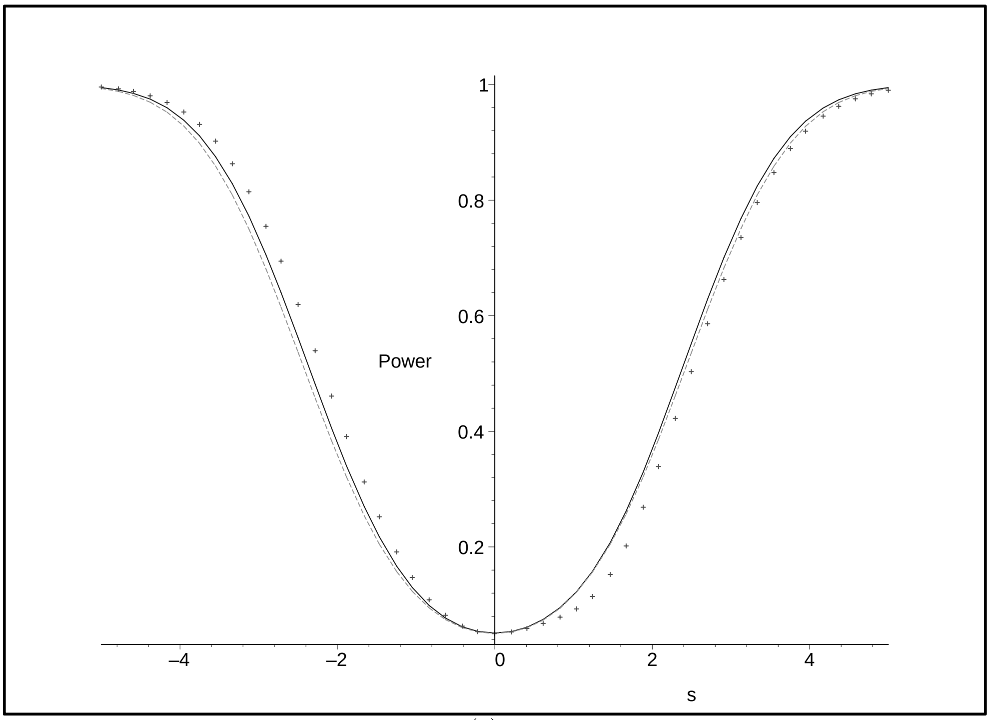
Figure 8: Bounding Power Functions for the CLR test and AR Power Function: (a) $k = 3, w = 1$, (b) $k = 3, w = 10$, (c) $k = 11, w = 1$, (d) $k = 11, w = 10$.



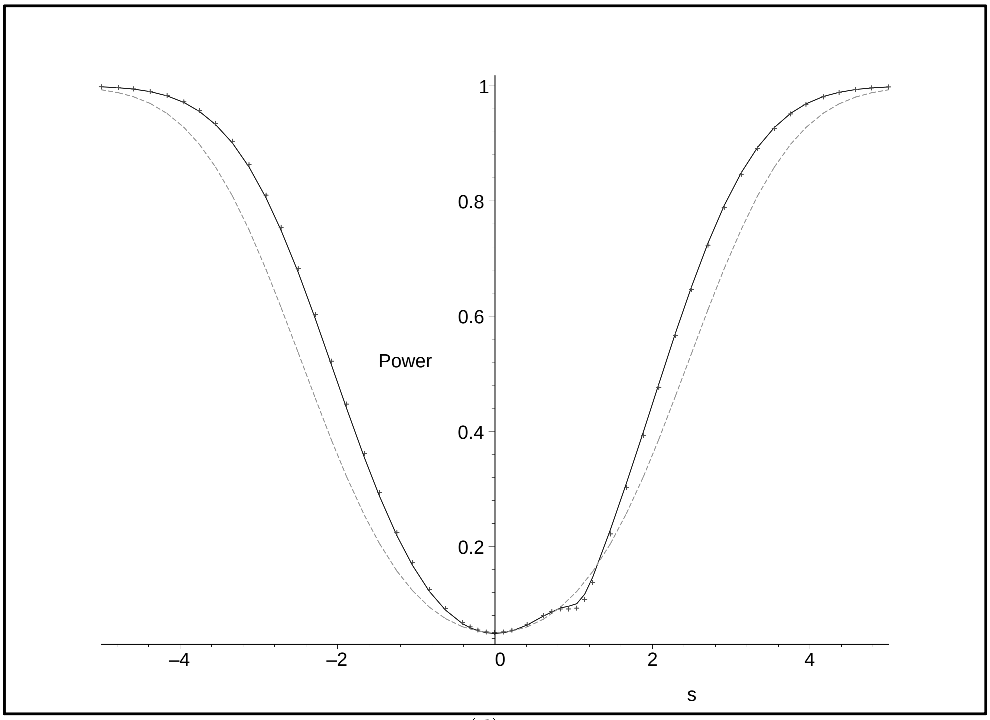
(a)



(b)

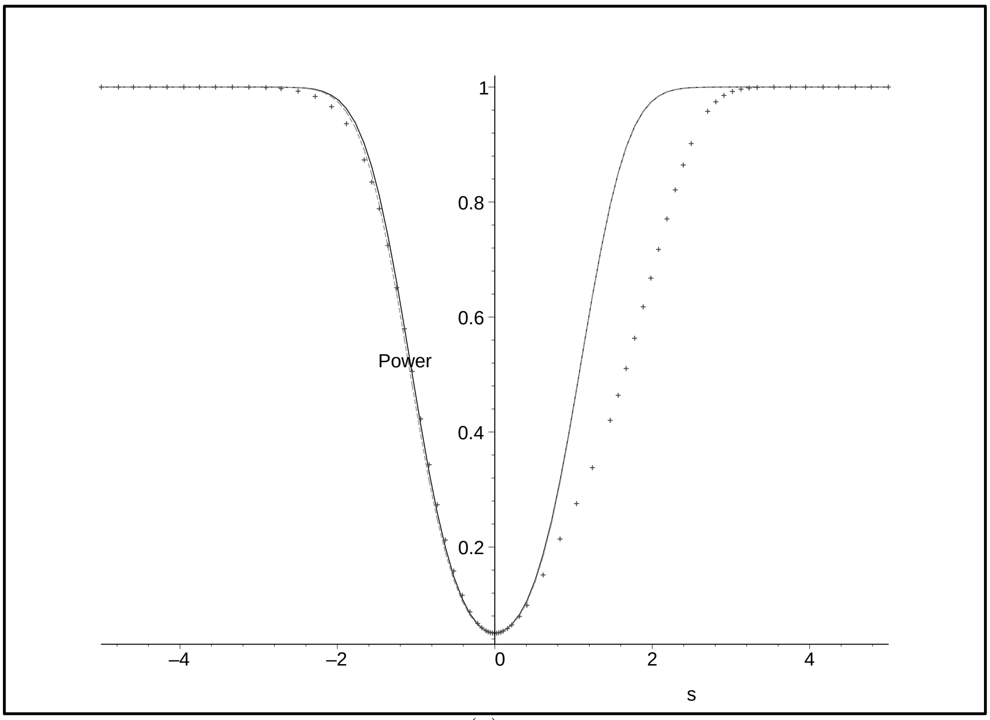


(c)

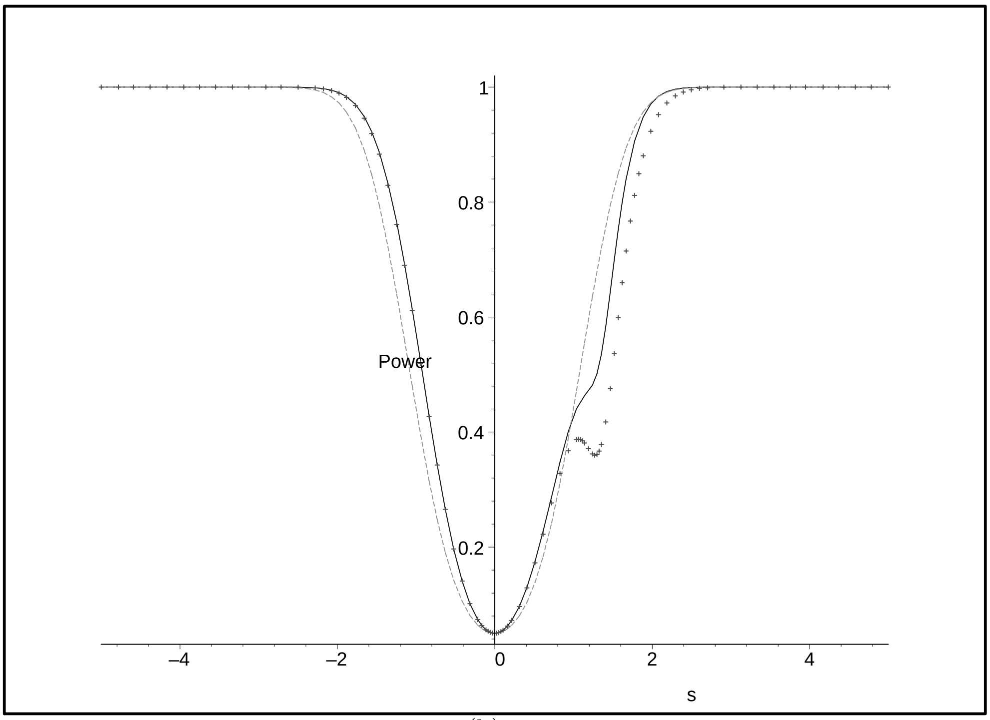


(d)

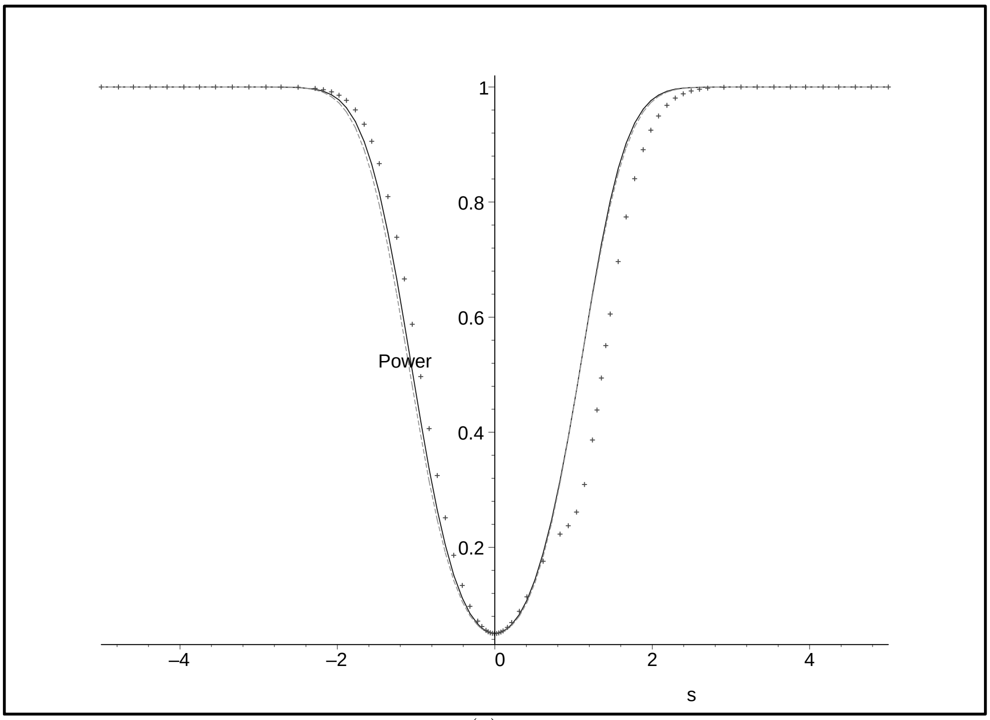
Figure 9: Power Functions for CLR test (solid line), AR test (dash), and LM test (crosses): $k = 3, \Delta_n = 1$; (a) $\rho = .75, w = 1$, (b) $\rho = .75, w = 5.3$, (c) $\rho = .95, w = 1$, (d) $\rho = .95, w = 13.25$.



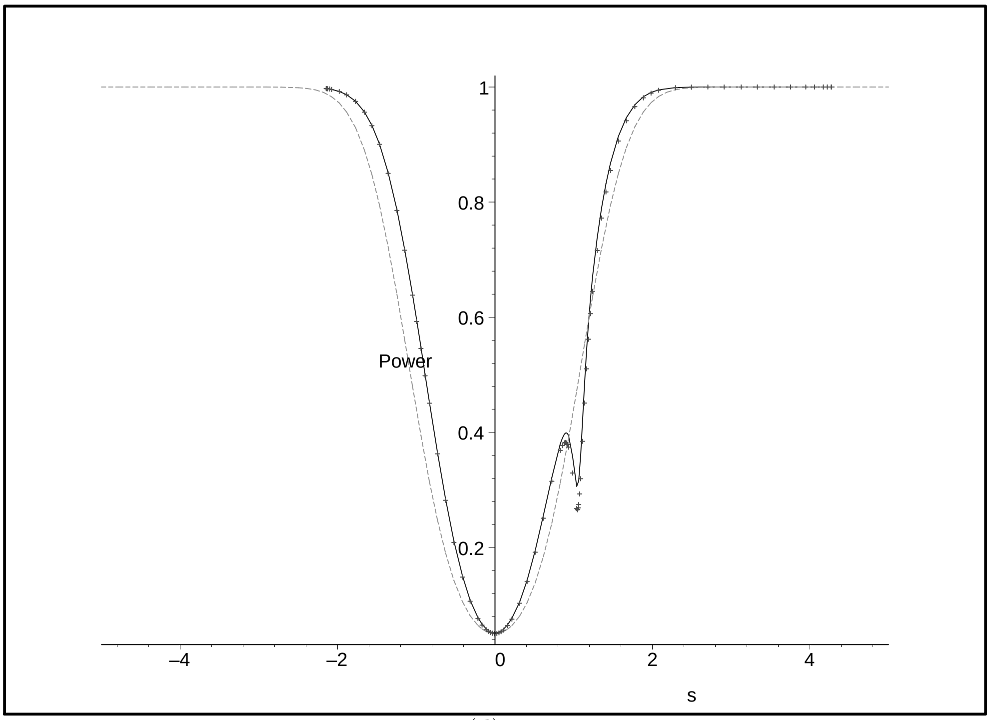
(a)



(b)

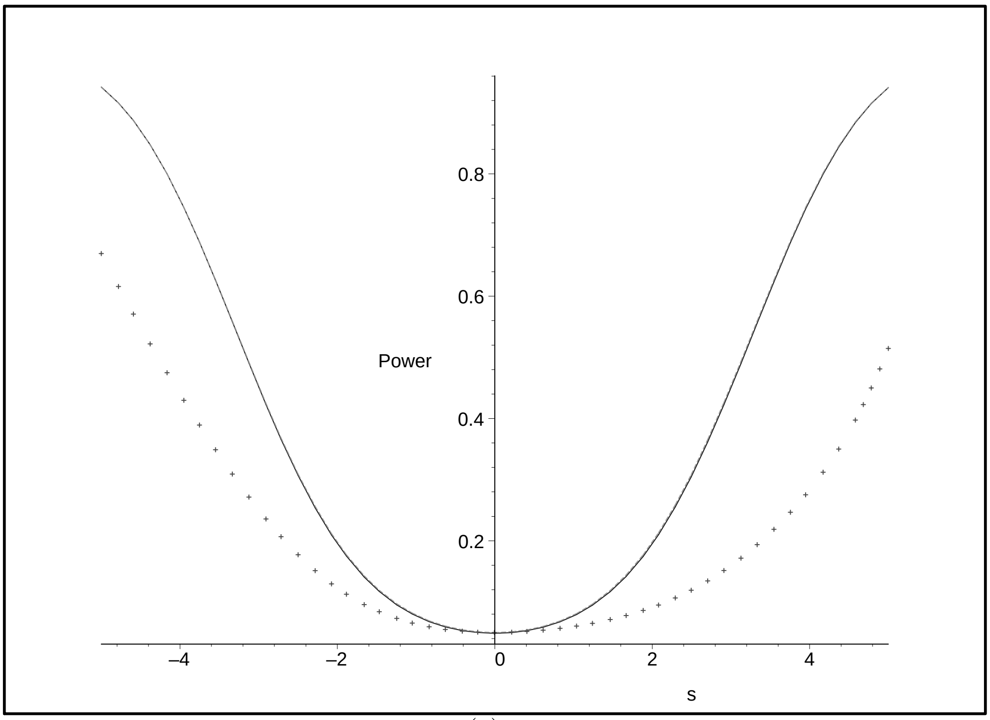


(c)

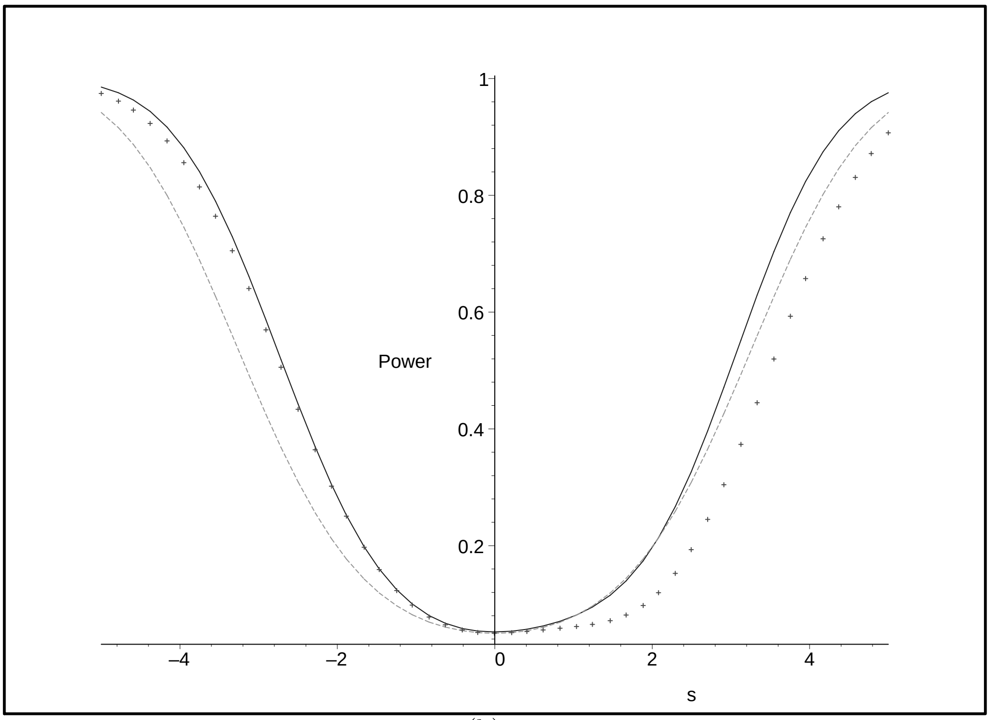


(d)

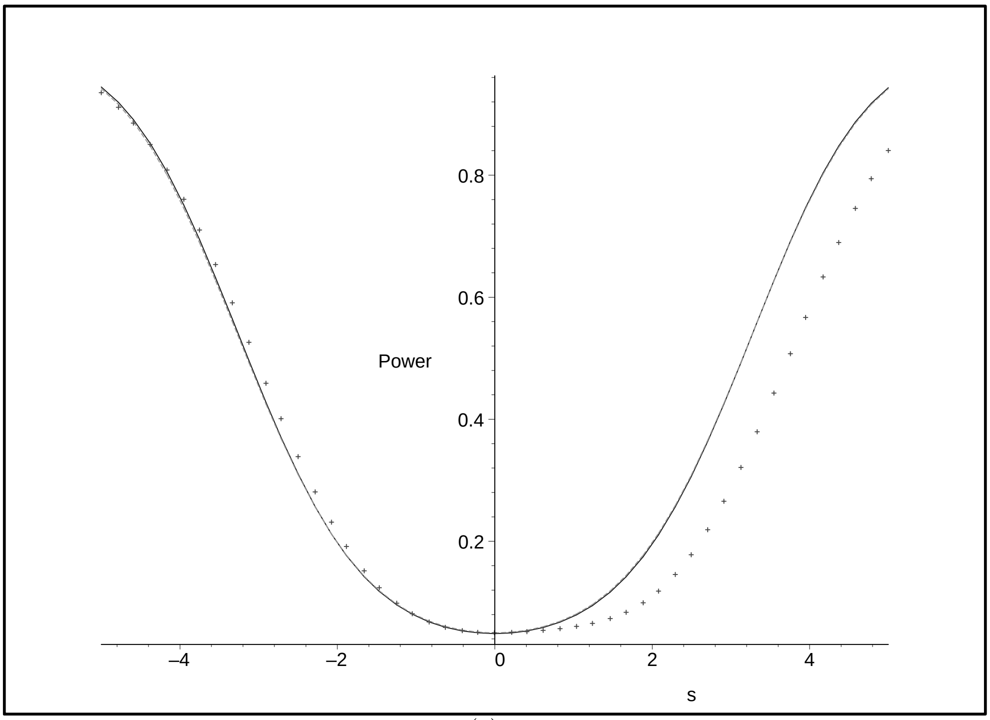
Figure 10: Power Functions for CLR test (solid line), AR test (dashes), and LM test (crosses): $k = 3, \Delta_n = 5$ (a) $\rho = .75, w = 1$, (b) $\rho = .75, w = 14.4$, (c) $\rho = .95, w = 1$, (d) $\rho = .95, w = 25$.



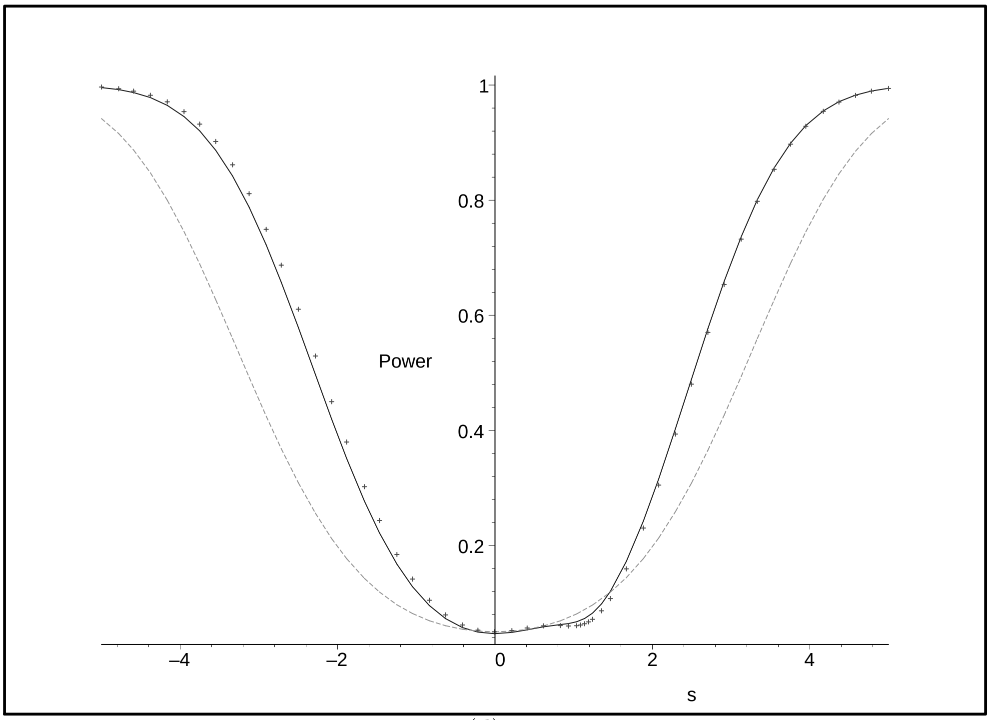
(a)



(b)

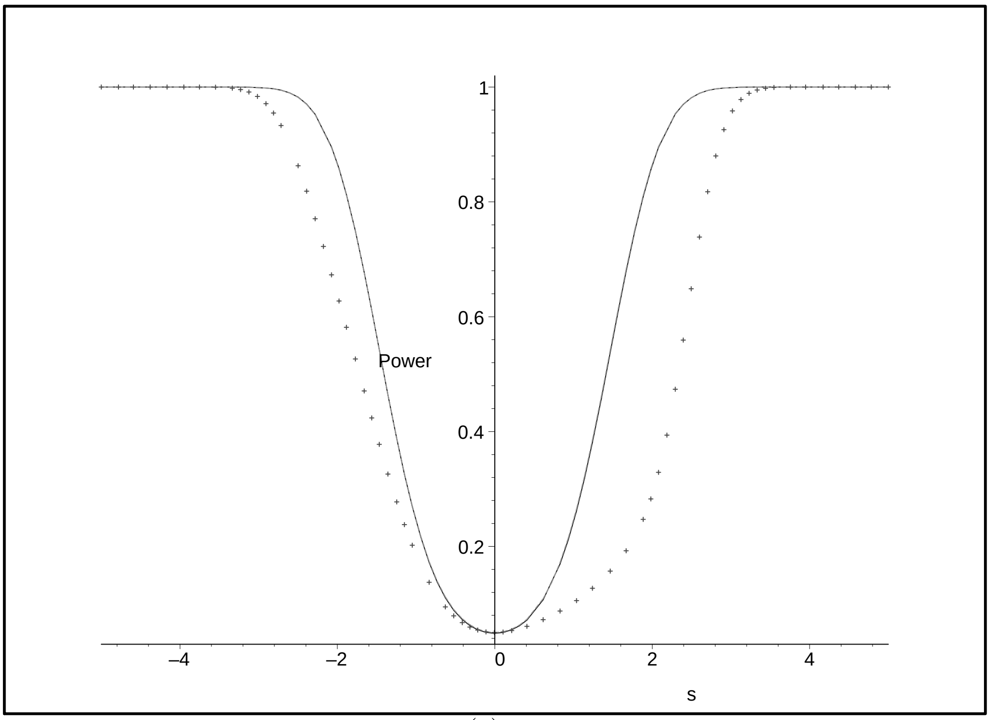


(c)

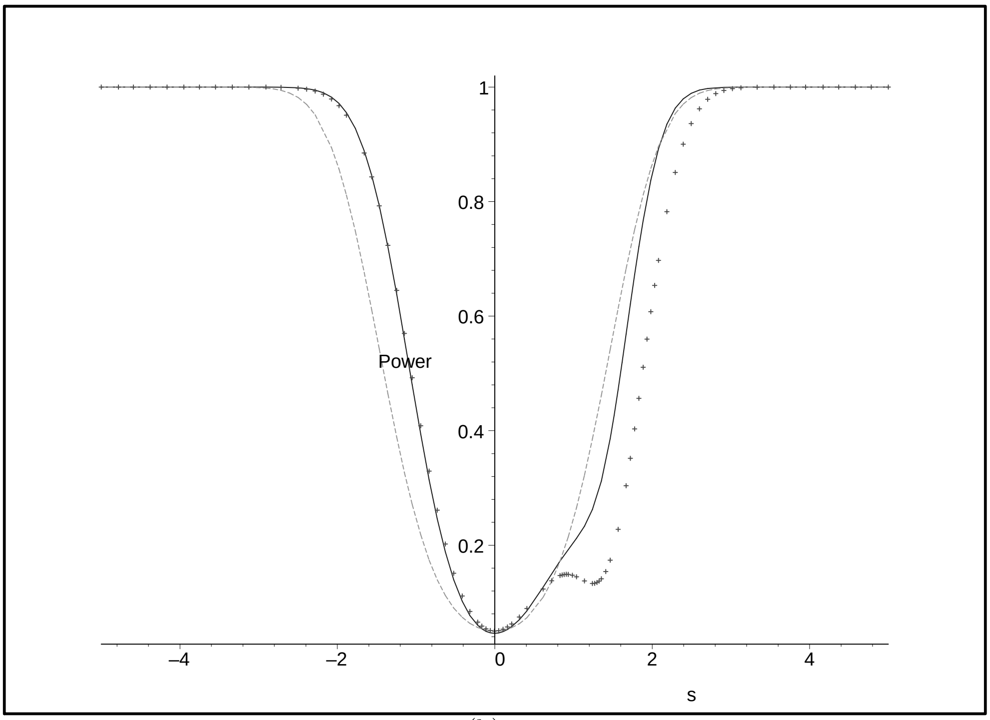


(d)

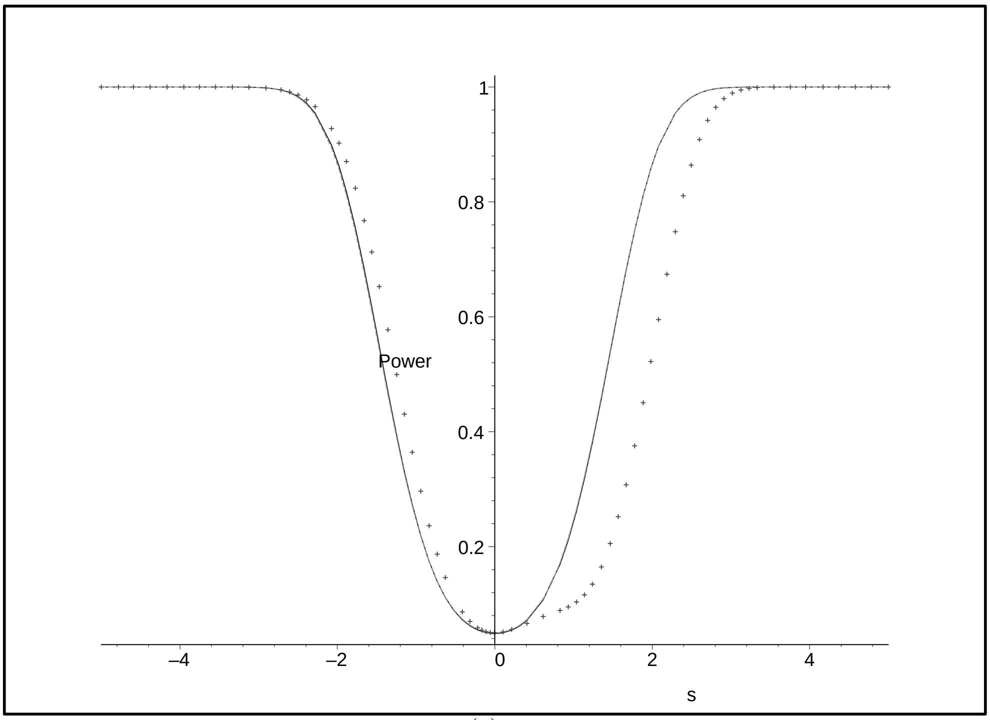
Figure 11: Power Functions for CLR test (solid line), AR test (dashes), and LM test (crosses): $k = 12, \Delta_n = 1$ (a) $\rho = .75, w = 1$, (b) $\rho = .75, w = 14.3$, (c) $\rho = .95, w = 1$, (d) $\rho = .95, w = 22.25$.



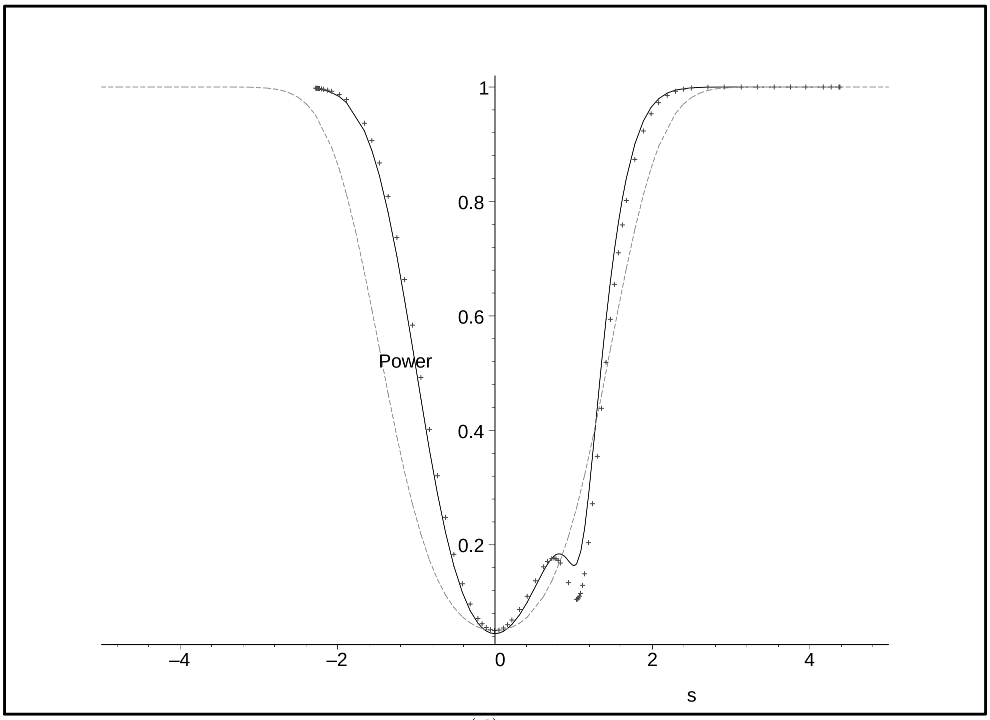
(a)



(b)



(c)



(d)

Figure 12: Power Functions for CLR test (solid line), AR test (dashes), and LM test (crosses): $k = 12, \Delta_n = 1$ (a) $\rho = .75, w = 1$, (b) $\rho = .75, w = 23.4$, (c) $\rho = .95, w = 1$, (d) $\rho = .95, w = 25$.