

A Conditional Likelihood Ratio Test for Structural Models

Marcelo J. Moreira*

January 9, 2002

Abstract

This paper develops a general method for constructing similar tests based on the conditional distribution of nonpivotal statistics in a simultaneous equations model with normal errors and known reduced-form covariance matrix. The test based on the likelihood ratio statistic is particularly simple and has good power properties. When identification is strong, the power curve of this conditional likelihood ratio test is essentially equal to the power envelope for similar tests. Monte Carlo simulations also suggest that this test dominates the Anderson-Rubin test and the score test. Dropping the restrictive assumption of disturbances normally distributed with known covariance matrix, approximate conditional tests are found that behave well in small samples even when identification is weak.

JEL Classification: C12, C31.

Keywords: instrumental variables, similar tests, Wald test, likelihood ratio test, power envelope, confidence regions, 2SLS estimator, LIML estimator.

*© 2001 Copyright. Lengthy discussions with Thomas Rothenberg were extremely important for this work and I am deeply indebted for all his help and support. This paper follows a suggestion by Peter Bickel. For comments and advice, I also would like to thank Kenneth Chay, Michael Jansson, James Powell and Paul Ruud.

1 Introduction

When making inferences about coefficients of endogenous variables in a structural equation, applied researchers often rely on asymptotic approximations. However, as emphasized in recent work by Bound, Jaeger and Baker (1995) and Staiger and Stock (1997), these approximations are not satisfactory when instruments are weakly correlated with the regressors. In particular, if identification can be arbitrarily weak, Dufour (1997) showed that Wald-type confidence intervals cannot have correct coverage probability while Wang and Zivot (1998) showed that the standard likelihood ratio test employing chi-square critical values do not have correct size. The problem arises because inference is based on nonpivotal statistics whose exact distributions depart substantially from their asymptotic approximations when identification is weak.

This paper develops a general procedure for constructing valid tests of structural coefficients based on the conditional distribution of nonpivotal statistics. When the reduced-form errors are normally distributed with a known covariance matrix, this procedure yields tests that are exactly similar; that is, their null rejection probabilities do not depend on the values of the unknown nuisance parameters. Simple modifications of these tests are shown to be approximately similar even when the errors are nonnormal and the reduced-form covariance matrix is unknown.

The conditional approach is employed to find critical value functions for Wald and likelihood ratio tests yielding correct rejection probabilities no matter how weak the instruments. Although the conditional Wald test has relatively poor power in some regions of the parameter space, the conditional likelihood ratio test has overall good power properties. Monte Carlo simulations suggest that this conditional likelihood ratio test not only has power close to the power envelope of similar tests when identification is good but it also dominates the test proposed by Anderson and Rubin (1949) and the score tests proposed by Kleibergen (2000) and Moreira (2001) when identification is weak. The conditional Wald and likelihood ratio tests can also be used to construct confidence regions centered around the 2SLS and LIML

estimators, respectively, that have correct coverage probability even when instruments are weak and that are informative when instruments are good.

This paper is organized as follows. In Section 2, exact results are developed for the special case of a two-equation model under the assumption that the reduced-form disturbances are normally distributed with known covariance matrix. Sections 3 and 4 extend the results to more realistic cases, although at the cost of introducing some asymptotic approximations. Monte Carlo simulations suggest that these approximations are quite accurate. Section 5 compares the confidence region based on the conditional likelihood ratio test with the confidence region based on a score test that is also approximately similar. Section 6 contains concluding remarks. All proofs are given in the appendix.

2 Known Covariance Matrix

2.1 The Model

To simplify exposition, consider a simple model in which the structural equation of interest is

$$(1) \quad y_1 = y_2\beta + u$$

where y_1 and y_2 are $n \times 1$ vectors of observations on two endogenous variables, u is an $n \times 1$ unobserved disturbance vector, and β is an unknown scalar parameter. This equation is assumed to be part of a larger linear simultaneous equations model which implies that y_2 is correlated with u . The complete system contains exogenous variables which can be used as instruments for conducting inference on β . Specifically, it is assumed that the reduced form for $Y = [y_1, y_2]$ can be written as

$$(2) \quad \begin{aligned} y_1 &= Z\pi\beta + v_1 \\ y_2 &= Z\pi + v_2 \end{aligned}$$

where Z is an $n \times k$ matrix of exogenous variables having full column rank k and π is a $k \times 1$ vector; the n rows of the $n \times 2$ matrix of reduced form

errors $V = [v_1, v_2]$ are i.i.d. normally distributed with mean zero and 2×2 nonsingular covariance matrix $\Omega = [\omega_{i,j}]$. It is assumed that $k > 1$ so the structural equation is “overidentified.” The goal here is to test the null hypothesis $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$.

Commonly used tests reject the null hypothesis when a test statistic \mathcal{T} takes on a value greater than a specified critical value c . The test is said to have size α if, when the null hypothesis is true,

$$\text{Prob}(\mathcal{T} > c) \leq \alpha$$

for all admissible values of the nuisance parameters π and Ω . Since π and Ω are unknown, finding a test with correct size is nontrivial. Of course, if the null distribution of \mathcal{T} does not depend on the nuisance parameters, the $1 - \alpha$ quantile of \mathcal{T} can be used for c and the null rejection probability will be identically equal to α . In that case, \mathcal{T} is said to be *pivotal* and the test is said to be *similar*.

Although structural coefficient tests based on pivotal statistics have been proposed in the literature (see, for example, Anderson and Rubin (1949), Kleibergen (2000), and Moreira (2001)), they sometimes have poor power properties. On the other hand, the Wald and likelihood ratio statistics most commonly employed in practice are not pivotal. However, under regularity conditions, these test statistics are asymptotically chi-square with one degree of freedom and tests using the $1 - \alpha$ chi-square quantile for c are asymptotically similar with size α . Unfortunately, if π cannot be bounded away from zero, their actual size can differ substantially from α since the asymptotic approximation will be very poor when the instruments are weakly correlated with y_2 .

One possible solution to the size problems that result when using non-pivotal statistics is to replace the chi-square critical value with some larger, conservative value guaranteeing that the null rejection probability is no larger than α . This is the approach taken by Wang and Zivot (1998) for the likelihood ratio test and the Hessian-based score test. Unfortunately, when π is not near the origin, these tests have null rejection probabilities much lower than α and waste power unnecessarily. Moreover, this approach is useless for

statistics that are not boundedly pivotal. Here we will develop an alternative method to construct similar tests based on nonpivotal statistics under the assumption Ω is known.

2.2 Similar Tests Based on Nonpivotal Statistics

When Ω is known and the errors are normal, the probability model is a member of the curved exponential family and the $k \times 2$ matrix $Z'Y$ is a sufficient statistic for the unknown parameters (β, π) . Hence, any test depends on the data only through $Z'Y$. However, for any known nonsingular, nonrandom 2×2 matrix D , the $k \times 2$ matrix $Z'YD$ is also sufficient. A convenient choice is the matrix $D = [b, \Omega^{-1}a]$, where $b = (1, -\beta_0)'$ and $a = (\beta_0, 1)'$. Then the sufficient statistic can be represented by the pair of $k \times 1$ vectors

$$S = Z'Yb = Z'(y_1 - y_2\beta_0) \quad \text{and} \quad T = Z'Y\Omega^{-1}a.$$

The vector S is normally distributed with mean $Z'Z\pi(\beta - \beta_0)$ and covariance matrix $(b'\Omega b)Z'Z$; T is independent of S and normally distributed with mean $(\bar{a}'\Omega^{-1}a)Z'Z\pi$ and covariance matrix $(a'\Omega^{-1}a)Z'Z$, where $\bar{a} = (\beta, 1)$. Thus we have partitioned the sufficient statistic into two independent, normally distributed vectors, T having a null distribution depending on π and S having a null distribution not depending on π .

Any test statistic that depends only on S will be pivotal and can be used to form a similar test. Likewise, as discussed in Moreira (2001), similar tests can be constructed from pivotal statistics of the form $g(T)'S/\sqrt{g(T)'Z'Zg(T)}$ where g is any (measurable) k -dimensional vector depending on T . The goal here instead is to find a similar test at level α based on a nonpivotal test statistic $\psi(S, T, \beta_0)$. The following approach is suggested by the analysis in Lehmann (1986, Chapter 4). Although the marginal distribution of ψ depends on π , the conditional null distribution of ψ *given that T takes on the value t* does not depend on π . As long as this distribution is continuous, its quantiles can be computed and used to construct a similar test. The following proposition follows immediately:

PROPOSITION 1: *Suppose that $\psi(S, t, \beta_0)$ is a continuous random variable under H_0 for every t . Define $c(t, \beta_0, \alpha)$ to be the $1 - \alpha$ quantile of the null distribution of $\psi(S, t, \beta_0)$. Then the test that rejects H_0 if $\psi(S, T, \beta_0) > c(T, \beta_0, \alpha)$ is similar at level $\alpha \in (0, 1)$.*

Moreira (2001) shows that $T = a'\Omega^{-1}a \cdot Z'Z\hat{\pi}$, where $\hat{\pi}$ is the maximum likelihood estimator of π when β is constrained to take the null value β_0 and Ω is known. Therefore, this method of finding similar tests can be interpreted as adjusting the critical value based on a preliminary estimate of π . Alternatively, this approach may be thought of as replacing the nonpivotal statistic $\psi(S, T, \beta_0)$ by the statistic $\psi(S, T, \beta_0) - c(T, \beta_0, \alpha)$ which yields a test of correct size no matter what the value of π .

Although Proposition 1 can be applied to any continuously distributed test statistic depending on S and T , the details will be worked out here only for two special cases: the Wald test statistic based on the two-stage least squares estimator and the likelihood ratio test statistic.

2.2.1 A Conditional Wald Test

Let $N_Z = Z(Z'Z)^{-1}Z'$. Consider the Wald statistic centered around the 2SLS estimator:

$$(3) \quad W_0 = \frac{(b_{2SLS} - \beta_0)^2 y_2' N_Z y_2}{\hat{\sigma}^2}$$

where $b_{2SLS} = (y_2' N_Z y_2)^{-1} y_2' N_Z y_1$ and $\hat{\sigma}^2 = [1 - b_{2SLS}] \Omega [1 - b_{2SLS}]'$. Here, the nonstandard structural error variance estimate exploits the fact that Ω is known. Since W_0 depends on the data only through $Z'Y$, it can be written as a function of S, T and β_0 . Although in principle the critical value function $c_W(t, \beta_0, \alpha)$ can be derived from the known distribution of S , a simple analytical expression seems out of reach. A Monte Carlo simulation from the null distribution of S is much simpler. Indeed, the applied researcher need only do a simulation for the actual value of T observed in the sample and for the particular β_0 being tested; there is no need to derive the whole function $c_W(t, \beta_0, \alpha)$.

Although a simple algebraic expression for the critical value function $c_W(t, \beta_0, \alpha)$ is not available, some of its properties are known. For any positive integer r , let $q_\alpha(r)$ denote the $1 - \alpha$ quantile of a chi-square distribution with r degrees of freedom. Let ρ_0 be the correlation coefficient between an element of $y_1 - y_2\beta_0$ and the corresponding element of y_2 . Then,

PROPOSITION 2: *The critical value function for the conditional W_0 test satisfies*

$$\begin{aligned} c_W(t, \beta_0, \alpha) &\rightarrow q_\alpha(1) \text{ as } t't \rightarrow \infty \\ c_W(t, \beta_0, \alpha) &\rightarrow q_\alpha(k) \frac{\rho_0^2}{1 - \rho_0^2} \text{ as } t't \rightarrow 0 \end{aligned}$$

Note that T is normal with a mean that is proportional to π . When π is far from the origin, $T'T$ will take on large values with very high probability; the relevant critical value for W_0 is then the standard chi-square-one quantile. When π is near the origin, $T'T$ is likely to be small and the relevant critical value for W_0 may be very large if ρ_0^2 is near one.

2.2.2 The Conditional Likelihood Ratio Test

When Ω is known, the likelihood ratio statistic, defined to be two times the log of the likelihood ratio, is given by:

$$(4) \quad LR_0 = \bar{\lambda}^{\max} - \frac{a'\Omega^{-1}Y'NY\Omega^{-1}a}{a'\Omega^{-1}a}$$

where $\bar{\lambda}^{\max}$ is the largest eigenvalue of $\Omega^{-1/2}Y'NY\Omega^{-1/2}$. This statistic can be written as a function of the sufficient statistics S and T . However, the expression is somewhat simpler when written in terms of the standardized statistics

$$\bar{S} = \frac{(Z'Z)^{-1/2}S}{\sqrt{b'\Omega b}} \quad \text{and} \quad \bar{T} = \frac{(Z'Z)^{-1/2}T}{\sqrt{a'\Omega^{-1}a}}$$

having covariance matrices equal to the identity matrix. Under the null hypothesis, \bar{S} has mean zero so $\bar{S}'\bar{S}$ is distributed as chi square with k degrees of freedom. The statistic $\bar{T}'\bar{T}$ is distributed as noncentral chi square with

noncentrality proportional to $\pi'Z'Z\pi$; it can be viewed as a natural statistic for testing the hypothesis that $\pi = 0$ under the assumption that $\beta = \beta_0$.

In Appendix B, the following expression for the likelihood ratio statistic is derived:

$$(5) \quad LR_0 = \frac{1}{2} \left[\bar{S}'\bar{S} - \bar{T}'\bar{T} + \sqrt{[\bar{S}\bar{S} + \bar{T}'\bar{T}]^2 - 4[\bar{S}'\bar{S} \cdot \bar{T}'\bar{T} - (\bar{S}'\bar{T})^2]} \right]$$

When $k = 1$, \bar{S} and \bar{T} are scalars and the LR_0 statistic collapses to the pivotal statistic $\bar{S}'\bar{S}$. In the overidentified case the LR_0 statistic depends also on $\bar{T}'\bar{T}$ and is no longer pivotal. Nevertheless, a similar test can be found by applying Proposition 1. Again, an analytic expression for the critical value function for the conditional LR_0 statistic is not available but the needed values can be computed by simulation. Some general properties of the function are known.

PROPOSITION 3: *The critical value function for the conditional LR_0 test depends only on α , k , and $\bar{t}'\bar{t}$. It satisfies*

$$\begin{aligned} c_{LR}(\bar{t}'\bar{t}, k, \alpha) &\rightarrow q_\alpha(1) \text{ as } \bar{t}'\bar{t} \rightarrow \infty \\ c_{LR}(\bar{t}'\bar{t}, k, \alpha) &\rightarrow q_\alpha(k) \text{ as } \bar{t}'\bar{t} \rightarrow 0 \end{aligned}$$

Table 1 presents the critical value function calculated from 10,000 Monte Carlo replications for the significance level of 5%. When $k = 1$, the true critical value function is a constant equal to 3.84 at level 5%. The slight variation in the first column of Table 1 is due to simulation error. For each k , the critical value function has approximately an exponential shape, decreasing from $q_\alpha(k)$ at $\bar{t}'\bar{t} = 0$ to $q_\alpha(1)$ as $\bar{t}'\bar{t}$ tends to infinity. For example, when $k = 4$, the approximation $c(\bar{t}'\bar{t}, 4, 0.05) = 3.84 + 5.65 \exp(-\bar{t}'\bar{t}/7)$ seems to fit reasonably well.

The shape of the critical value function explains why the method proposed by Wang and Zivot (1998) leads to a test with low power. Their proposed critical value of a chi-square- k quantile is the upper bound for the true critical value function. The method proposed here can also be seen as a refinement of the method proposed by Zivot, Startz and Nelson (1998) that selects for the critical value either $q_\alpha(k)$ or $q_\alpha(1)$ depending on a preliminary test of the

hypothesis $\pi = 0$. The conditional approach has the advantage that it is not *ad hoc* and the final test has correct size without unnecessarily wasting power. Figure 1 illustrates each method, sketching its respective critical values¹ for different values of $\bar{t}'\bar{t}$ when the number of instruments equals four.

2.3 Monte Carlo Simulations

To evaluate the power of the conditional W_0 and LR_0 tests, a 1,000 replication experiment was performed based on design I of Staiger and Stock (1997). The hypothesized value β_0 is zero. The elements of the 100×4 matrix Z are drawn as independent standard normal and then held fixed over the replications. Two different values of the π vector are used so that the “population” first-stage F-statistic (in the notation of Staiger and Stock) $\lambda'\lambda/k = \pi'Z'Z\pi/(\omega_{22}k)$ takes the values 1 (weak instruments) and 10 (good instruments). The rows of (u, v_2) are i.i.d. normal random vectors with unit variances and correlation ρ . Results are reported for ρ taking the values 0.00, 0.50 and 0.99. The critical values for the conditional likelihood ratio and Wald tests were based on 1,000 replications.

In addition to the two conditional tests, denoted as LR_0^* and W_0^* , two other similar tests were evaluated: the Anderson-Rubin test based on the statistic $AR_0 = \bar{S}'\bar{S}$ (modified to take into account that Ω is known) and the score test based on the statistic $LM_0 = (\bar{T}'\bar{S})^2/\bar{T}'\bar{T}$. The latter test is described in Kleibergen (2000) and Moreira (2001). Figures 2-3 graph, for a fixed value of π and ρ , the rejection probabilities of the AR_0 , LM_0 , conditional LR_0 and conditional W_0 tests as functions of the true value β .² The power envelope for similar tests is also included. In each figure, all four power curves are at approximately the 5% level when β equals β_0 . This reflects the fact that each test is similar. As expected, the power curves become steeper as the quality of instruments improve.

¹The pre-testing procedure proposed by Zivot, Nelson and Startz (1998) is based on the *OLS* estimator for π . Instead, figure 1 sketches the critical value function based on a pre-testing on the constrained maximum likelihood estimator for π .

²As β varies, ω_{11} and ω_{12} change to keep the structural error variance and the correlation between u and v_2 constant.

As expected, the AR_0 test has power considerably lower than the envelope power. The LM_0 test has relative low power either for the weak-instrument case or for some alternatives β for the good-instrument case. Figures 2-3 also suggest that the conditional W_0 test has poor power for some parts of the parameter space. These poor power properties are not shared by the conditional LR_0 test. The conditional likelihood ratio test not only seems to dominate the Anderson-Rubin test and the score test, but it also has power essentially equal to the power envelope for similar tests³ when identification is strong.

The good performance of the conditional LR_0 test is not surprising. Various authors have noticed that, in curved exponential models, the likelihood ratio test performs well for a wide range of alternatives. Van Garderen (2000) addresses this issue in the case the nuisance parameter is present only under the alternative.

3 Unknown Covariance Matrix

In practice, of course, the reduced-form covariance matrix Ω is not known and must be treated as an additional nuisance parameter. Under normality, the probability model remains curved exponential, but the sufficient statistic expands to $(Z'Y, Y'Y)$. However, the conditional approach developed in Section 2 cannot be easily applied in this context. Nevertheless, it seems plausible that the conditional tests can still be used after replacing the unknown Ω by some estimate since Ω can be well estimated regardless of the quality of the instruments. Furthermore, although motivated by the normal model, under relatively weak regularity conditions, the test statistics have limiting distributions that do not depend on the error distribution. Thus, modified versions of the tests developed under the restrictive assumptions of normal errors with known reduced form covariance matrix can be expected to behave

³Other tests that have been proposed in the literature such as the Wald test based on the LIML estimator and the GMM_0 test proposed by Wang and Zivot (1998) were also considered. However, Monte Carlo simulations suggest that their conditional counterparts have power no larger than the conditional likelihood ratio test.

well under much weaker assumptions.

3.1 A Conditional Wald Test

The OLS estimator of Ω is given by $\hat{\Omega} = Y'M_Z Y/(n - k)$ where $M_Z = I - Z(Z'Z)^{-1}Z'$. It is natural then to consider the test statistic

$$(6) \quad W = \frac{(b_{2SLS} - \beta_0)' y_2' N_Z y_2}{\tilde{\sigma}^2}$$

where $\tilde{\sigma}^2 = [1 - b_{2SLS}] \hat{\Omega} [1 - b_{2SLS}]'$. The critical value function derived for W_0 can then be applied here, but with $\tilde{T} = Z'Y\hat{\Omega}^{-1}a$ instead of T . That is, when Ω is unknown, one would reject the null hypothesis that $\beta = \beta_0$ if

$$W - c_W(\tilde{T}, \beta_0, \alpha) > 0.$$

3.2 The Likelihood Ratio Test

Likewise, the LR_0 test statistic also depends on Ω and, therefore, cannot be used when the covariance matrix of the reduced form disturbances is unknown. Again, a natural modification is to replace Ω by its OLS estimator:

$$LR_1 = \lambda^{\max} - \frac{a' \hat{\Omega}^{-1} Y' N_Z Y \hat{\Omega}^{-1} a}{a' \hat{\Omega}^{-1} a}$$

where λ^{\max} is the largest eigenvalue of $\hat{\Omega}^{-1/2} Y' N_Z Y \hat{\Omega}^{-1/2}$. Alternatively, one could use the likelihood ratio statistic for the case Ω is unknown:

$$LR = -\frac{n}{2} \ln \left(1 + \frac{\lambda^{\min}}{n - k} \right) + \frac{n}{2} \ln \left(1 + \frac{b' Y' N_Z Y b}{b' Y' M_Z Y b} \right)$$

where λ^{\min} is the smallest eigenvalue of $\hat{\Omega}^{-1/2} Y' N_Z Y \hat{\Omega}^{-1/2}$.

Even for relatively small samples, the LR_1 and LR statistics are close to the LR_0 statistic. Therefore, the critical values in Table 1 can be used for the conditional LR_1 and LR tests as well, after replacing Ω by $\hat{\Omega}$ in the expression for \bar{T} .

3.3 Monte Carlo Simulations

Although the modified conditional tests are not exactly similar, they appear to have good size properties even when the instruments may be weak. To evaluate the rejection probability under H_0 , the design used in Section 2.3 is once more replicated. Results are reported for the same parameter values except for sample size.

Tables II and III present rejection probabilities for the following tests: Anderson-Rubin⁴ (AR), the Hessian-based score test (LM) described by Zivot, Startz and Nelson (1998), the information-based score test (LM_2) described in Moreira (2001), the likelihood ratio test (LR), the conditional likelihood ratio test (LR^*), the Wald test centered around the 2SLS estimator (W), and the conditional Wald test (W^*). The AR , LM_2 , LR^* , and W^* tests are approximately similar, whereas the LM , LR , and W test are not.

Although the LM test does not present good size properties, the LM_2 test does. Likewise, the LR and W tests present worse size properties than the conditional LR^* and W^* tests. The null rejection probabilities of the LR test range from 0.048-0.220 and those of the W test range from 0.002-0.992 when the number of observations is 80. The null rejection probabilities of their conditional counterparts range from 0.046-0.075 and 0.030-0.072, respectively.

Results for non-normal disturbances are analogous⁵. Tables IV and V show the rejection probabilities of some 5% tests when Staiger and Stock's design II is used. The structural disturbances, u and v_2 , are serially uncorrelated with $u_t = (\xi_{1t}^2 - 1)/\sqrt{2}$ and $v_{2t} = (\xi_{2t}^2 - 1)/\sqrt{2}$ where ξ_{1t} and ξ_{2t} are normal with unit variance and correlation $\sqrt{\rho}$. The k instruments are indicator variables with equal number of observations in each cell. When the number of observations is 80, the rejection probabilities under H_0 of the LR^* and W^* tests are still close to 5% for all values of $\lambda'\lambda/k$ and ρ .

Finally, Tables VI and VII compare the power of the conditional LR_0^* test

⁴For the AR test, a $\chi^2(k)$ critical value was used.

⁵Once more, the critical value function is based on 1,000 Monte Carlo simulations as if the disturbances were normally distributed with known variance Ω .

(Ω known) with that of the conditional LR^* test (Ω unknown) when Staiger and Stock's design I with 100 observations is used. The difference between the two power curves is small, which suggests that the power comparison in Section 2.3 for the LR_0^* test is also valid for the LR^* test. Tables VIII and IX show that the same conclusion holds for the conditional W_0^* and W^* tests.

4 Extensions

The previous theory can easily be extended to a structural equation with more than two endogenous variables and with additional exogenous variables as long as inference is to be conducted on *all* the endogenous coefficients. Consider the structural equation

$$y_1 = Y_2\beta + X\gamma + u$$

where Y_2 is the $n \times l$ matrix of observations on the l explanatory endogenous variable and X is the $n \times r$ matrix of observations on r exogenous variables. This equation is part of a larger linear system containing the additional exogenous variables Z . The reduced form for $Y = [y_1, Y_2]$ is

$$\begin{aligned} y_1 &= Z\Pi\beta + X\delta + v_1 \\ Y_2 &= Z\Pi + X\Gamma + V_2 \end{aligned}$$

where $\delta = \Gamma\beta + \gamma$. The rows of $V = [v_1, V_2]$ are i.i.d. normal random vectors with mean zero and covariance matrix Ω . It is assumed that X and Z have full column rank. The problem is to test the vector hypothesis $H_0 : \beta = \beta_0$ treating Π, Γ, δ as nuisance parameters.

The unknown parameters associated with X can be eliminated by taking orthogonal projections. Define the $l + 1$ component column vector $b = (1, -\beta_0')'$. Let A be any $(l + 1) \times l$ matrix whose columns are orthogonal to b . Then, if $M_X = I - X(X'X)^{-1}X'$, the statistics

$$S = Z'M_X Y b \quad \text{and} \quad T = Z'M_X Y \Omega^{-1} A$$

are independent and normally distributed. For a nonpivotal statistic $\psi(S, T, \beta_0)$, the critical value can be found computing the $1 - \alpha$ quantile of the distribution of ψ conditioned on $T = t$. Again, Ω can be replaced with a consistent estimate and the normality assumption dropped without affecting the results significantly.

5 Confidence Regions

Confidence regions for β with approximately correct coverage probability can be constructed by inverting approximately similar tests. Although Dufour (1997) showed that Wald-type confidence intervals are not valid, the confidence regions based on the conditional Wald test has correct coverage probability in large samples no matter how weak the instruments. Likewise, if the LM_2 test or the conditional LR test is used, the resulting confidence regions have approximately correct level. Moreover, the regions based on the conditional Wald test necessarily contain the 2SLS estimator of β while the ones based on the conditional likelihood ratio test or on the LM_2 test are centered around the limited-information maximum likelihood estimator of β . Therefore, confidence regions based on these tests can be used as evidence of the accuracy of their respective estimators.

To illustrate how informative the confidence regions based on the conditional LR test are compared with the ones based on the LM_2 test, design I of Staiger and Stock (1997) is once more used. One sample was drawn where the true value of β is zero. Figures 4-6 plots the LR and LM_2 statistics and their respective critical value functions at significance level of 5% against β_0 ⁶. The region in which each statistic is below its critical value curve is the corresponding confidence set.

When the instruments are invalid (Figure 4), the confidence regions cover the real line. This is expected to happen about 95% of time since the confidence regions have correct coverage probability and β is unidentified. As the quality of the instruments increases, the confidence regions become narrower.

⁶The approximate critical value function $c(t) = 3.84 + 5.65 \exp(-\bar{t}t/7)$ was used.

Moreover, LR confidence regions are significantly smaller than the LM_2 ones, as a result of better power properties of the conditional likelihood ratio test. For example, when $\lambda' \lambda / k = 10$ and $\rho = 0$, the LR confidence region is the set $[0.0, 0.6]$ while the LM confidence region is the set $[-4.1, -2.3] \cup [0.0, 0.6]$ (Figure 6). This illustrates that the score test fails to reject values very far from the true value, whereas the conditional likelihood ratio test does not.

6 Conclusions

Monte Carlo simulations suggest that the conditional likelihood ratio test has good size and power properties. If identification is good, this test has power curve essentially equal to the upper bound power curve for similar tests. Monte Carlo simulations also suggest that this test dominates the test proposed by Anderson and Rubin (1949) and the score tests studied by Kleibergen (2000) and Moreira (2001). This test can also be used to construct informative confidence regions with correct coverage probability centered around the limited-information maximum likelihood estimator.

The conditional approach used in this paper for finding similar tests based on nonpivotal statistics can be applied to other statistical problems involving nuisance parameters. Improved inference should be possible whenever a subset of the statistics employed to form a test statistic has a nuisance-parameter free distribution and is independent of the remaining statistics under the null hypothesis.

Department of Economics, University of California at Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720-3880 USA; jovita@econ.berkeley.edu; <http://socrates.berkeley.edu/~jovita>

7 References

ANDERSON, T., 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.

BOUND, J., D. JAEGER AND R. BAKER (1995): “Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak,” *Journal of American Statistical Association*, 90, 443-50.

DUFOUR, J-M. (1997): “Some impossibility theorems in econometrics with applications to structural and dynamic models,” *Econometrica*, 65, 1365-88.

GARDEREN, K. J. VAN (2000): “An alternative comparison of classical tests: assessing the effects of curvature” in *Applications of Differential Geometry to Econometrics*, ed. by M. Salmon and P. Marriot. Cambridge: Cambridge University Press.

KLEIBERGEN, F.(2000): “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression,” Tinbergen Institute Discussion Paper TI2000-055/4.

LEHMANN, E. (1986): *Testing Statistical Hypothesis*. 2nd edition, Wiley Series in Probability and Mathematical Statistics.

MOREIRA, M. (2001): *Tests with Correct Size when Instruments Can Be Arbitrarily Weak*. Manuscript.

STAIGER, D., AND J. STOCK (1997): “Instrumental variables regression with weak instruments,” *Econometrica*, 65, 557-86.

WANG, J., AND E. ZIVOT (1998): “Inference on a structural parameter in instrumental variables regression with weak instruments,” *Econometrica*, 66, 1389-404.

ZIVOT, E., R. STARTZ, AND C. NELSON (1998): “Valid confidence intervals and inference in the presence of weak instruments,” *International Economic Review*, 39, 1119-44.

TABLE I
CRITICAL VALUE FUNCTION OF THE LIKELIHOOD RATIO TEST

$\bar{t} \bar{t}$	k							
	1	2	3	4	5	6	7	8
0	3.96	5.88	7.79	9.45	11.19	12.44	14.20	15.74
1	3.80	5.64	7.13	8.67	10.25	11.73	13.09	14.73
2	3.79	5.12	6.62	8.04	9.40	10.74	12.36	13.72
3	3.73	4.94	6.46	7.60	8.67	10.12	11.69	13.24
4	3.85	4.81	5.72	6.93	8.31	9.71	10.61	12.34
5	3.85	4.48	5.45	6.49	7.54	8.77	10.17	11.37
6	3.76	4.23	5.12	6.16	7.25	8.26	9.42	10.66
7	3.78	4.43	5.00	5.84	6.65	7.88	8.96	9.89
8	3.89	4.19	4.96	5.63	6.40	7.32	8.54	9.46
9	3.91	4.19	5.00	5.39	6.08	7.01	7.78	8.96
10	3.83	4.27	4.64	5.09	5.79	6.72	7.45	8.18
15	3.78	4.06	4.36	4.67	5.01	5.45	6.06	6.67
20	3.88	3.97	4.21	4.38	4.74	4.93	5.25	5.61
25	3.79	4.12	4.22	4.29	4.47	4.79	4.96	5.17
30	3.91	3.96	4.00	4.25	4.28	4.69	4.83	4.99
40	3.93	3.93	3.98	4.20	4.34	4.26	4.55	4.72
50	3.85	3.91	4.01	4.20	4.10	4.35	4.44	4.48
75	3.73	3.85	4.02	3.88	4.04	4.14	4.13	4.05
100	3.86	3.74	3.79	3.93	4.06	4.07	4.12	4.04
150	3.74	3.92	3.96	3.88	4.06	3.92	3.99	4.03
200	3.72	3.91	3.93	3.95	3.90	4.03	4.00	4.04
250	3.74	3.78	3.87	4.02	3.86	3.89	4.02	3.96
500	3.72	3.92	3.79	3.94	3.93	3.99	3.89	4.01
750	3.86	3.74	3.64	3.89	3.86	3.82	3.77	3.86
1000	3.87	3.90	3.82	3.94	3.84	3.94	3.68	3.87
5000	3.86	3.90	3.94	3.81	3.84	3.83	3.82	3.76
10000	3.76	3.79	3.78	3.86	3.82	3.88	3.88	3.78
50000	3.88	3.90	3.92	3.71	3.88	3.82	3.94	3.72

TABLE I (CONTINUED)
 CRITICAL VALUE FUNCTION OF THE LIKELIHOOD RATIO TEST

$\bar{t}\bar{t}$	k							
	9	10	15	20	25	50	100	200
0	16.85	18.26	24.95	31.23	37.92	67.42	123.85	234.11
1	15.96	17.39	23.98	30.72	36.69	66.18	123.45	232.64
2	15.18	16.57	23.15	29.68	35.83	65.89	122.33	232.26
3	14.40	16.04	22.35	28.40	34.84	64.04	121.60	230.63
4	13.59	14.77	21.22	27.27	33.81	63.73	120.42	229.70
5	12.92	13.90	20.57	26.74	33.01	62.35	119.61	228.79
6	11.97	13.34	19.66	26.04	31.88	61.49	118.52	228.29
7	11.48	12.46	18.90	24.66	31.15	60.20	117.61	227.04
8	10.74	11.86	17.58	24.06	30.05	59.49	116.30	226.37
9	9.84	10.99	16.99	22.99	29.18	58.62	115.21	225.48
10	9.30	10.37	15.89	21.87	28.23	58.35	114.48	224.54
15	7.07	7.91	12.52	18.10	23.66	53.46	109.76	218.54
20	5.82	6.41	9.58	14.07	19.38	47.98	104.72	214.16
25	5.53	5.88	7.88	11.35	15.77	43.23	99.43	209.40
30	5.05	5.51	7.09	9.11	12.67	38.70	94.78	204.80
40	4.65	4.86	5.71	7.05	9.00	29.47	84.80	194.81
50	4.60	4.73	5.19	6.15	7.28	21.44	75.25	184.99
75	4.42	4.36	4.56	5.11	5.67	10.13	51.29	159.19
100	4.10	4.22	4.45	4.76	4.90	7.36	30.38	135.05
150	4.10	4.13	4.24	4.47	4.62	5.72	10.68	86.07
200	3.98	3.98	3.97	4.34	4.28	5.03	7.65	42.16
250	3.89	3.92	3.99	4.21	4.29	4.54	6.40	17.79
500	3.92	3.99	3.93	4.10	4.04	4.26	4.78	6.40
750	3.92	3.95	4.02	3.97	3.97	4.02	4.49	5.27
1000	3.84	3.91	3.88	3.91	3.85	4.05	4.35	4.72
5000	3.76	3.73	3.93	3.88	3.94	3.81	4.03	4.10
10000	3.75	3.97	3.89	3.94	3.82	3.86	3.87	3.95
50000	3.85	3.74	3.84	3.84	3.84	3.88	3.80	3.77

TABLE II
 PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5% (20 OBS.)

ρ	$\lambda'\lambda/k$	AR	LM	LM_2	LR	LR^*	W	W^*
0.00	0.00	9.50	4.20	6.70	26.40	9.70	0.80	4.40
0.00	1.00	9.40	3.80	5.00	16.90	8.90	2.10	4.10
0.00	10.00	9.60	5.70	5.60	8.50	7.30	6.50	7.80
0.50	0.00	9.10	14.40	6.70	26.10	10.80	16.20	8.20
0.50	1.00	8.30	8.20	6.20	17.70	8.30	13.40	6.20
0.50	10.00	9.10	4.00	4.90	7.10	6.20	7.50	4.70
0.99	0.00	9.80	46.60	6.60	25.60	11.10	98.40	9.10
0.99	1.00	9.70	22.40	4.50	6.30	6.30	66.50	10.30
0.99	10.00	9.10	5.90	4.20	5.10	5.50	13.10	6.00

TABLE III
 PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5% (80 OBS.)

ρ	$\lambda'\lambda/k$	AR	LM	LM_2	LR	LR^*	W	W^*
0.00	0.00	6.20	4.00	5.30	20.70	5.00	0.20	3.00
0.00	1.00	5.30	4.90	5.50	16.20	6.30	1.00	5.20
0.00	10.00	6.10	4.60	4.60	5.80	4.60	3.30	4.00
0.50	0.00	6.70	13.00	5.30	22.00	5.60	13.00	5.10
0.50	1.00	6.10	9.00	5.00	13.80	5.60	12.30	6.10
0.50	10.00	6.10	4.20	4.40	4.80	4.60	5.10	4.00
0.99	0.00	7.30	41.60	5.80	21.80	7.50	99.20	7.20
0.99	1.00	6.50	22.00	4.80	5.00	4.80	60.50	7.00
0.99	10.00	6.40	6.60	5.50	6.00	6.10	13.40	5.80

TABLE IV
 PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5% (20 OBS.)

- NON-NORMAL DISTURBANCES AND BINARY INSTRUMENTS -

ρ	$\lambda\lambda/k$	AR	LM	LM_2	LR	LR^*	W	W^*
0.00	0.00	12.60	6.40	8.40	27.60	13.60	2.40	5.40
0.00	1.00	10.70	5.40	6.80	25.70	10.90	2.10	5.00
0.00	10.00	11.30	8.70	8.50	13.60	11.30	6.90	8.30
0.50	0.00	10.00	8.70	8.30	25.30	12.40	5.60	5.50
0.50	1.00	9.10	6.30	6.80	22.70	10.30	5.50	3.90
0.50	10.00	9.10	7.90	7.60	12.70	9.30	7.40	7.70
0.99	0.00	14.30	48.60	10.70	30.70	15.50	97.50	12.80
0.99	1.00	10.10	27.90	10.20	12.30	11.90	81.00	7.50
0.99	10.00	12.30	12.50	9.50	11.00	10.90	27.60	4.70

TABLE V
 PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5% (80 OBS.)

- NON-NORMAL DISTURBANCES AND BINARY INSTRUMENTS -

ρ	$\lambda\lambda/k$	AR	LM	LM_2	LR	LR^*	W	W^*
0.00	0.00	6.20	4.40	5.30	23.80	5.90	0.30	3.80
0.00	1.00	6.40	4.00	5.50	22.50	6.50	0.20	3.80
0.00	10.00	5.90	7.30	7.90	12.10	8.10	2.90	7.10
0.50	0.00	7.20	8.60	6.60	23.40	7.90	4.40	5.60
0.50	1.00	6.50	6.70	6.30	21.80	7.50	3.10	5.40
0.50	10.00	6.70	6.70	6.70	10.80	7.40	4.20	5.40
0.99	0.00	7.60	41.30	7.10	24.30	7.90	96.90	7.00
0.99	1.00	6.60	29.20	6.70	8.40	7.00	81.20	5.60
0.99	10.00	5.70	11.10	6.40	6.70	7.20	27.40	3.10

TABLE VI
 PERCENT REJECTED AT NOMINAL LEVEL OF 5%
CONDITIONAL LIKELIHOOD RATIO TEST - WEAK INSTRUMENTS

β	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	LR_0^*	LR^*	LR_0^*	LR^*	LR_0^*	LR^*
-10.00	30.10	31.70	34.20	35.10	56.90	57.70
-8.00	28.20	29.80	34.00	34.10	60.30	60.40
-6.00	28.90	29.60	34.70	34.90	62.30	61.80
-4.00	28.00	29.30	35.90	36.90	71.50	71.70
-2.00	24.00	24.80	37.80	38.00	96.70	96.10
0.00	5.40	5.70	5.90	6.40	6.30	7.00
2.00	26.60	26.00	21.70	23.30	24.90	26.10
4.00	27.10	28.00	25.60	26.00	33.70	33.80
6.00	29.00	31.40	27.60	28.90	37.00	37.20
8.00	30.50	30.90	30.20	29.70	41.00	41.60
10.00	28.80	30.50	28.20	30.10	43.20	45.00

TABLE VII
 PERCENT REJECTED AT NOMINAL LEVEL OF 5%
CONDITIONAL LIKELIHOOD RATIO TEST - GOOD INSTRUMENTS

β	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	LR_0^*	LR^*	LR_0^*	LR^*	LR_0^*	LR^*
-10.00	99.80	99.80	100.00	100.00	100.00	100.00
-8.00	99.70	99.80	100.00	100.00	100.00	100.00
-6.00	98.90	98.70	100.00	100.00	100.00	100.00
-4.00	95.10	95.10	100.00	99.90	100.00	100.00
-2.00	58.50	59.00	78.90	78.90	98.40	98.60
0.00	5.40	5.90	5.30	5.50	6.40	7.00
2.00	59.70	60.00	48.90	48.60	41.70	42.90
4.00	94.30	93.70	85.10	84.40	78.60	78.70
6.00	98.80	98.70	96.00	95.60	90.80	90.60
8.00	99.40	99.10	97.50	97.00	95.70	94.70
10.00	99.80	99.80	99.50	99.30	95.50	95.50

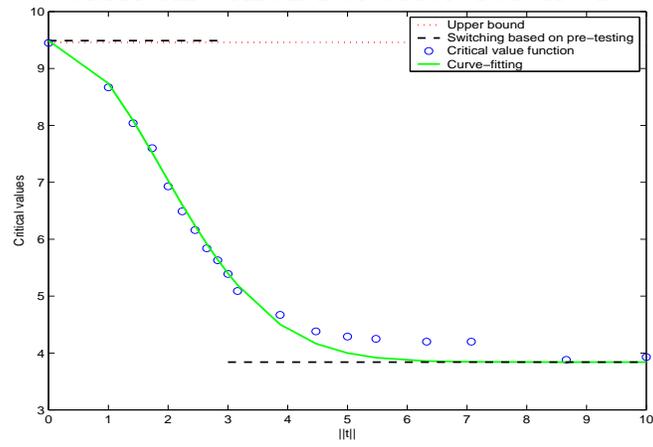
TABLE VIII
 PERCENT REJECTED AT NOMINAL LEVEL OF 5%
 CONDITIONAL WALD TEST - WEAK INSTRUMENTS

β	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	W_0^*	W^*	W_0^*	W^*	W_0^*	W^*
-10.00	29.70	30.70	30.40	30.80	2.00	2.80
-8.00	30.70	31.70	31.30	32.80	2.00	2.60
-6.00	29.80	31.40	32.30	32.70	2.50	3.30
-4.00	29.20	29.70	32.30	33.00	2.40	2.40
-2.00	25.20	26.50	33.60	35.60	25.40	27.70
0.00	4.90	4.70	3.60	4.50	3.70	5.20
2.00	27.50	28.00	22.60	22.50	0.90	1.10
4.00	28.00	30.60	23.70	24.60	1.30	1.90
6.00	29.30	29.10	25.80	25.80	1.90	1.90
8.00	31.80	32.20	29.10	29.10	1.80	2.10
10.00	29.10	30.20	26.30	27.50	1.70	1.80

TABLE IX
 PERCENT REJECTED AT NOMINAL LEVEL OF 5%
 CONDITIONAL WALD TEST - GOOD INSTRUMENTS

β	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	W_0^*	W^*	W_0^*	W^*	W_0^*	W^*
-10.00	99.80	99.90	100.00	100.00	100.00	100.00
-8.00	100.00	100.00	100.00	100.00	100.00	100.00
-6.00	99.70	99.40	100.00	100.00	100.00	100.00
-4.00	97.30	97.40	100.00	100.00	100.00	100.00
-2.00	70.20	69.90	73.60	74.10	43.70	43.80
0.00	5.50	5.60	4.30	4.50	5.00	5.40
2.00	70.10	70.20	58.80	59.40	48.00	47.80
4.00	96.20	95.90	87.30	86.80	75.10	73.70
6.00	99.20	99.00	94.80	93.70	83.90	82.20
8.00	99.70	99.40	96.40	95.30	85.60	83.30
10.00	99.80	99.70	97.20	96.60	80.00	78.90

FIGURE 1
CRITICAL VALUE FUNCTION FOR $K=4$



Student Version of MATLAB

FIGURE 2
 EMPIRICAL POWER OF TESTS: WEAK INSTRUMENTS

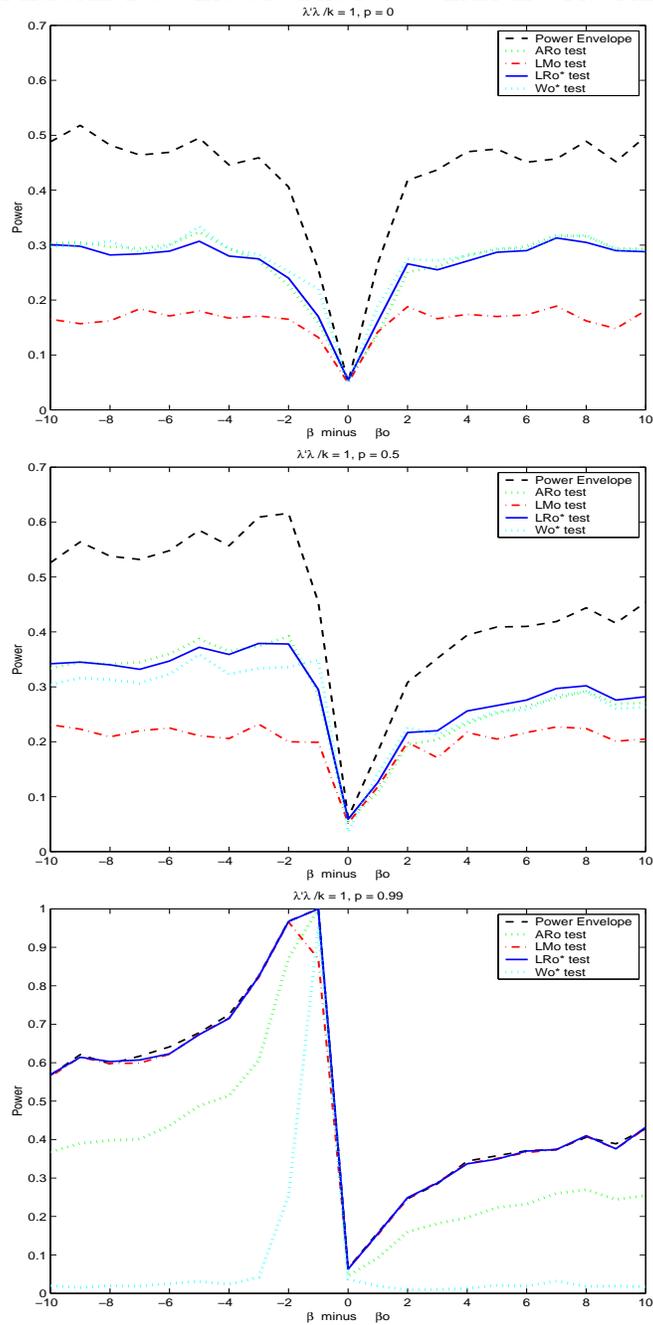


FIGURE 3
 EMPIRICAL POWER OF TESTS: GOOD INSTRUMENTS

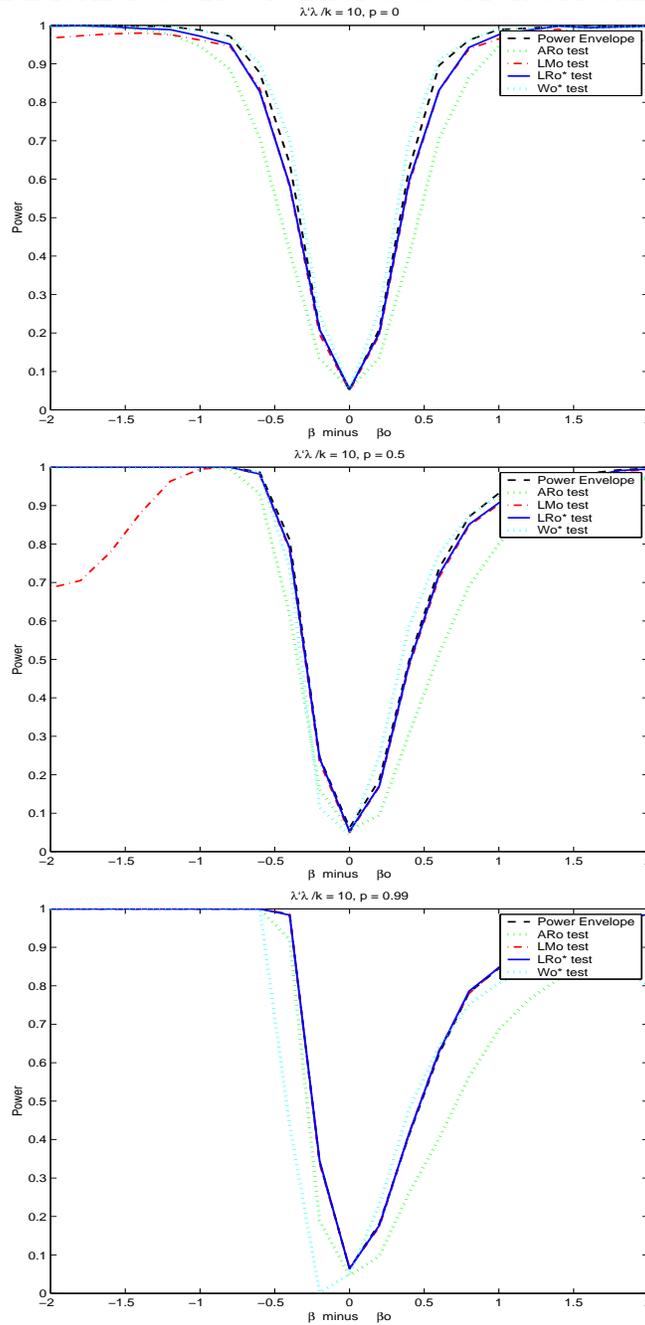


FIGURE 4
 CONFIDENCE REGIONS: INVALID INSTRUMENTS

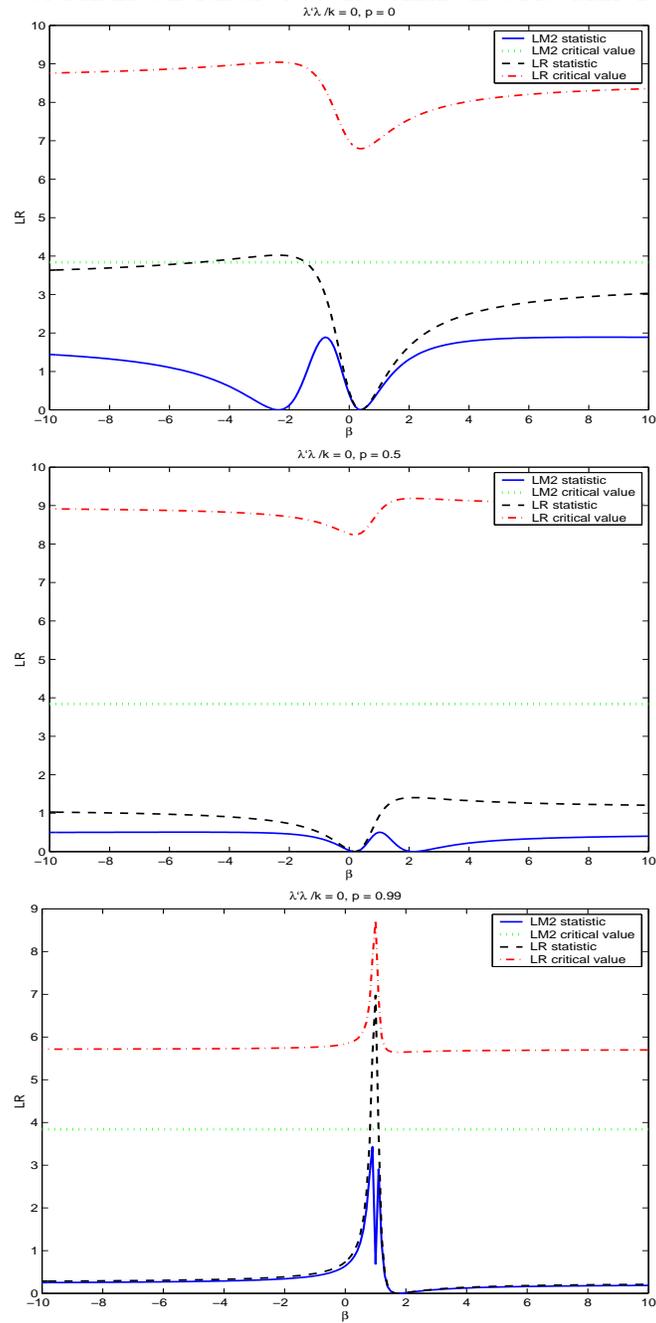


FIGURE 5

CONFIDENCE REGIONS: WEAK INSTRUMENTS

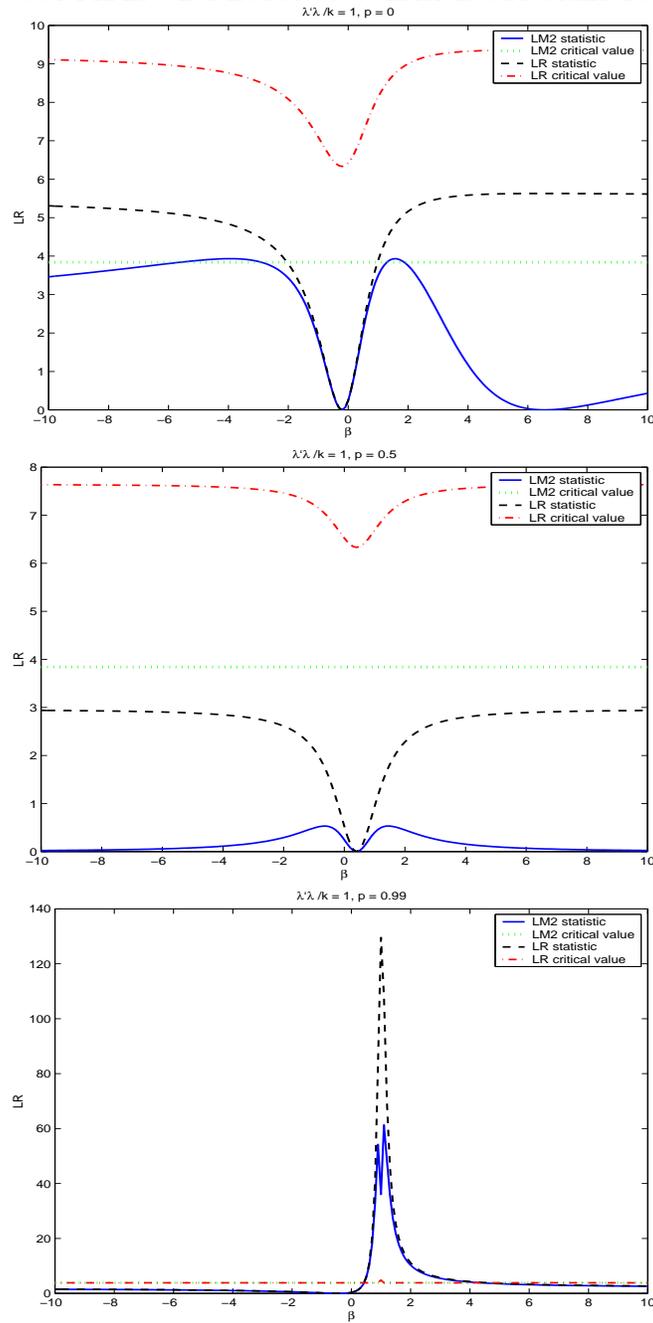
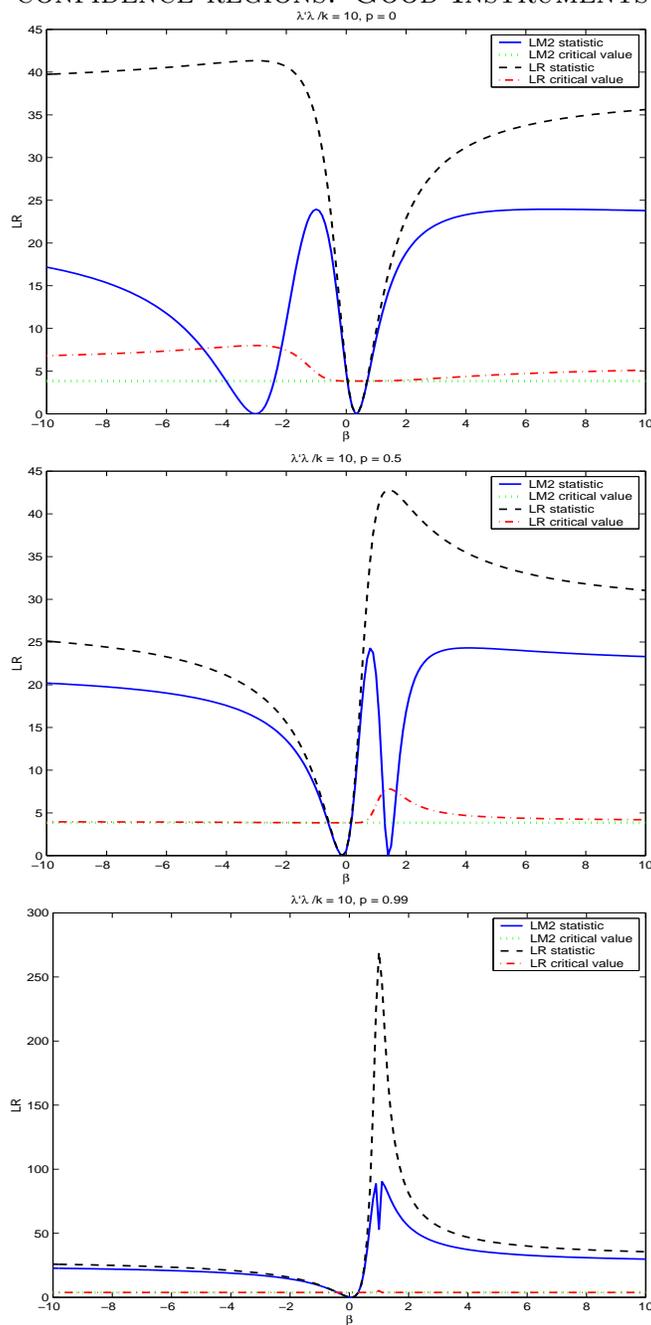


FIGURE 6
 CONFIDENCE REGIONS: GOOD INSTRUMENTS



8 Appendix A: Likelihood Ratio Derivation

Ignoring an additive constant, the log-likelihood function can be written as:

$$(A.1) \quad l(Y; \beta, \pi, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} \text{tr} (\Omega^{-1} V' V)$$

Maximizing with respect to π , one finds that $\pi(\beta, \Omega) = (Z'Z)^{-1} Z'Y\Omega^{-1}c/c'\Omega^{-1}c$ where $c \equiv (\beta, 1)'$. The concentrated log-likelihood function, $l_c(Y; \beta, \Omega)$, defined as $l(Y; \beta, \pi(\beta, \Omega), \Omega)$, is given by:

$$l_c(Y; \beta, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} \left[\text{tr} \Omega^{-1} Y' Y - \frac{c' \Omega^{-1} Y' N Y \Omega^{-1} c}{c' \Omega^{-1} c} \right]$$

When evaluated at $\hat{\beta}$, the maximum likelihood estimator of β when Ω is known, this becomes

$$l_c(Y; \hat{\beta}, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} [\text{tr} \Omega^{-1} Y' Y - \bar{\lambda}^{\max}]$$

where $\bar{\lambda}^{\max}$ is the largest eigenvalue of $(Z'Z)^{-1/2} Z'Y\Omega^{-1}Y'Z (Z'Z)^{-1/2}$ or, equivalently, the largest eigenvalue of $\Omega^{-1/2} Y' N_Z Y \Omega^{-1/2}$. Since the likelihood ratio statistic when Ω is known, LR_0 , is defined as $2 [l_c(Y; \hat{\beta}, \Omega) - l_c(Y; \beta_0, \Omega)]$, it follows that:

$$LR_0 = \bar{\lambda}^{\max} - \frac{a' \Omega^{-1} Y' N_Z Y \Omega^{-1} a}{a' \Omega^{-1} a}$$

To find the likelihood ratio when Ω is unknown, we maximize (A.1) with respect to Ω , obtaining $\Omega(\beta, \pi) = V'V/n$. Inserting this into (A.1). we obtain

$$l(Y; \beta, \pi, \Omega(\beta, \pi)) = -\frac{n}{2} \ln |V'V| + n \ln(n) - n$$

After considerable algebra we find that the maximum value of the log-likelihood function for a fixed β is given by:

$$-\frac{n}{2} \ln \left(\frac{u'u}{u' M_Z u} \right) - \frac{n}{2} \ln |Y' M_Z Y| + n \ln(n) - n$$

The concentrated log-likelihood function, $l_c(Y; \beta)$, defined as $l(Y; \beta, \pi(\beta), \Omega(\beta))$, is then given by:

$$l_c(Y; \beta) = -\frac{n}{2} \ln \left(1 + \frac{d' Y' N_Z Y d}{d' Y' M_Z Y d} \right) - \frac{n}{2} \ln |Y' M_Z Y| + n \ln(n) - n$$

where $d = (1, -\beta)'$. Moreover, the concentrated log-likelihood function evaluated at the maximum likelihood estimator β_{LIML} is then given by:

$$l_c(Y; \beta_{LIML}) = -\frac{n}{2} \ln \left(1 + \frac{\lambda^{\min}}{n-k} \right) - \frac{n}{2} \ln |Y' M_Z Y| + n \ln(n) - n$$

Since the LR when Ω is unknown is defined as $2[l_c(Y; \beta_{LIML}) - l_c(Y; \beta_0)]$, it follows that:

$$LR = -n \ln \left(1 + \frac{\lambda^{\min}}{n-k} \right) + n \ln \left(1 + \frac{b' Y' N_Z Y b}{b' Y' M_Z Y b} \right)$$

9 Appendix B: Proof of Propositions

PROOF OF PROPOSITION 1: In fact we need only assume that $\psi(S, t, \beta_0)$ is a continuous random variable for all t except for a set having T -probability zero. For any t where $\psi(S, t, \beta_0)$ is not a continuous random variable, define $c_\alpha(t)$ to be zero. Otherwise, let $c_\alpha(t)$ be the $1 - \alpha$ quantile of ψ . Then, by definition, $\Pr[\psi(S, T, \beta_0) > c_\alpha(T) | T = t] = \alpha$. Since this holds for all t , it follows that $\Pr[\psi(S, T, \beta_0) > c_\alpha(T)] = \alpha$ unconditionally.

Q.E.D.

PROOF OF PROPOSITION 2: By a linear change of variables, the system (2) can be transformed into one where Ω and $Z'Z$ are identity matrices. The new parameter $\bar{\beta}$ is related to the old parameter by $\bar{\beta} = (\beta\omega_{22} - \omega_{12})/|\Omega|^{1/2}$. In the new coordinates, we have

$$\begin{aligned} Z' y_1 &= (S + \bar{\beta}_0 T) / (1 + \bar{\beta}_0^2) \\ Z' y_2 &= (T - \bar{\beta}_0 S) / (1 + \bar{\beta}_0^2) \end{aligned}$$

Then

$$(\bar{b}_{2SLS} - \bar{\beta}_0)^2 y_2' N_Z y_2 = \frac{(\bar{\beta}_0 S' S - T' S)^2}{T' T + \bar{\beta}_0^2 S' S - 2\bar{\beta}_0 S' T}$$

and

$$\bar{b}_{2SLS} = \frac{\bar{\beta}_0 (T' T - S' S) + (1 - \bar{\beta}_0^2) T' S}{T' T + \bar{\beta}_0^2 S' S - 2\bar{\beta}_0 S' T}$$

Substituting into (3), we find $W_0 \rightarrow S'T(T'T)^{-1}T'S/(1+\bar{\beta}_0^2)$ as $T'T \rightarrow \infty$, But $S/(1+\bar{\beta}_0^2)^{1/2}$ is distributed as $N(0, I)$ under the null hypothesis. Hence, conditional on $T = t$, W_0 tends to a chi-square-one random variable. When $T = 0$, $W_0 = S'S\bar{\beta}_0^2/(1+\bar{\beta}_0^2)$. Transforming back to the original coordinates, we find that this is $\rho_0^2/(1-\rho_0^2)$ times a chi-square- k random variable.

Q.E.D.

PROOF OF PROPOSITION 3: Note that $(\bar{S}, \bar{T}) = (Z'Z)^{-1/2}Z'Y\Omega^{-1/2}W$ where

$$W = \begin{bmatrix} \Omega^{1/2}b & \Omega^{-1/2}a \\ \sqrt{b'\Omega b} & \sqrt{a'\Omega^{-1}a} \end{bmatrix}$$

is an orthogonal matrix. Thus the eigenvalues of $\Omega^{-1/2}Y'N_ZY\Omega^{-1/2}$ are the same as the eigenvalues of $(\bar{S}, \bar{T})'(\bar{S}, \bar{T})$. The largest eigenvalue then is given by

$$\bar{\lambda}_{\max} = \frac{1}{2} \left[\bar{T}'\bar{T} + \bar{S}'\bar{S} + \sqrt{(\bar{T}'\bar{T} + \bar{S}'\bar{S})^2 - 4[\bar{S}'\bar{S} \cdot \bar{T}'\bar{T} - (\bar{S}'\bar{T})^2]} \right]$$

Moreover, $a'\Omega^{-1}Y'NY\Omega^{-1}a/a'\Omega^{-1}a = \bar{T}'\bar{T}$. Therefore, the LR_0 test statistic is given by expression (5). For $\bar{T}'\bar{T} \neq 0$, LR_0 can be rewritten as

$$LR_0 = \frac{1}{2} \left[Q_1 + Q_{k-1} - \bar{T}'\bar{T} + \sqrt{(Q_1 + Q_{k-1} + \bar{T}'\bar{T})^2 - 4Q_{k-1} \cdot \bar{T}'\bar{T}} \right]$$

where $Q_1 = \bar{S}'\bar{T}(\bar{T}'\bar{T})^{-1}\bar{T}'\bar{S}$ and $Q_{k-1} = \bar{S}'[I - \bar{T}(\bar{T}'\bar{T})^{-1}\bar{T}']\bar{S}$. Conditioned on $\bar{T} = \bar{t}$, Q_1 and Q_{k-1} are independent and under H_0 have chi-square distributions with one and $k-1$ degrees of freedom, respectively. Therefore the critical value function just depends on k, α , and $\bar{t}'\bar{t}$. When $\bar{T} = 0$, $LR_0 = \bar{S}'\bar{S}$, which is a chi-square- k random variable. When $\bar{T}'\bar{T} \rightarrow \infty$, $LR_0 \rightarrow (\bar{S}'\bar{T})^2/\bar{T}'\bar{T}$, which is a chi-square-one random variable.

Q.E.D.