Econometrica, Vol. 71, No. 4 (July, 2003), 1027–1048

A CONDITIONAL LIKELIHOOD RATIO TEST FOR STRUCTURAL MODELS

BY MARCELO J. MOREIRA[1]

This paper develops a general method for constructing exactly similar tests based on the conditional distribution of nonpivotal statistics in a simultaneous equations model with normal errors and known reduced-form covariance matrix. These tests are shown to be similar under weak-instrument asymptotics when the reduced-form covariance matrix is estimated and the errors are non-normal. The conditional test based on the likelihood ratio statistic is particularly simple and has good power properties. Like the score test, it is optimal under the usual local-to-null asymptotics, but it has better power when identification is weak.

KEYWORDS: Instruments, similar tests, Wald test, score test, likelihood ratio test, confidence regions, 2SLS estimator, LIML estimator.

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 Nelson and Startz (1990), 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) shows that Wald-type confidence intervals cannot have correct coverage probability, while Wang and Zivot (1998) show that the standard likelihood ratio test employing chi-square critical values does 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. This procedure yields tests that are exactly similar when the reduced-form errors are normally distributed with known variance. When this assumption is dropped, simple modifications of these tests are shown to have limiting power under weak-instrument asymptotics equal to the exact power when the errors are normal

[1] 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 Donald Andrews, Kenneth Chay, Michael Jansson, James Powell, Paul Ruud, James Stock, a co-editor, and two anonymous referees. Finally, I would like to thank the seminar participants at Berkeley, Brown, Chicago, EPGE/FGV, Harvard, Montreal, Northwestern, Penn, Pittsburgh, Princeton, PUC-Rio, UCLA, UCSD, and Yale.

with known variance. In particular, these modified tests are asymptotically similar even when the structural parameters are unidentified.

The conditional approach is employed to find a critical value function for the likelihood ratio statistic. This conditional likelihood ratio test has good power properties overall. It behaves like the unconditional likelihood ratio test when identification is strong and seems to dominate the test proposed by Anderson and Rubin (1949) and a particular score test. This score-type test was first proposed by Breusch and Pagan (1980) in a general framework, and has been used by Kleibergen (2002) and Moreira (2001) in testing weakly-identified parameters; see Moreira (2002) for a general exposition of the weak-instrument problem.

The conditional approach can also be used to construct valid confidence regions. For example, coefficient values not rejected by the conditional 2SLS Wald test form a confidence region centered around the 2SLS estimate, while the values not rejected by the conditional likelihood ratio test form a confidence region centered around the LIML estimate.[2] These regions have correct coverage probability even when instruments are weak and are informative when instruments are good.

This paper is organized as follows. In Section 2, exact results are developed under the assumption that the reduced-form disturbances are normally distributed with known covariance matrix. Section 3 focuses on the likelihood ratio test. Section 4 extends the results for an unknown error distribution, 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 appendices.

2. NORMAL REDUCED-FORM ERROR DISTRIBUTION WITH KNOWN COVARIANCE MATRIX

2.1. *The Model*

Consider the structural equation

$$y_1 = Y_2\beta + X_1\gamma + u, \tag{1}$$

where y_1 is the $n \times 1$ vector of observations on an endogenous variable, Y_2 is the $n \times l$ matrix of observations on l explanatory endogenous variables, X_1 is the $n \times r$ matrix of observations on r exogenous variables, and u is an $n \times 1$ unobserved disturbance vector having mean zero. This equation is assumed to be part of a larger linear simultaneous equations model in which Y_2 may be correlated with u. The complete system contains k additional exogenous variables (represented

[2] For the Stata .ado file that computes the LIML estimator and constructs confidence regions based on the conditional approach, see Moreira and Poi (2003).

by the matrix X_2) that can be used as instruments for conducting inference on the structural coefficients β. More specifically, we have

$$Y_2 = X_2 \Pi + X_1 \overline{\Gamma} + V_2,$$

where we assume that $k \geq l$ and that the matrix $[X_1, X_2]$ has full column rank $(k+r)$.

For any matrix Q having full column rank, let $N_Q = Q(Q'Q)^{-1}Q'$ and $M_Q = I - N_Q$. It will be convenient to write the reduced form for $Y = [y_1, Y_2]$ in terms of the orthogonal pair $[Z, X_1]$ where $Z = M_{X_1} X_2$. Then the reduced-form system can be written as

$$\begin{aligned} y_1 &= Z\Pi\beta + X_1\delta + v_1, \\ Y_2 &= Z\Pi + X_1\Gamma + V_2, \end{aligned} \tag{2}$$

where $\Gamma = \overline{\Gamma} + (X_1'X_1)^{-1}X_1'X_2\Pi$ and $\delta = \Gamma\beta + \gamma$. The restriction on the coefficients of Z in the reduced form are implied by the identifying assumption that the instruments X_2 do not appear in (1). In this section, we also assume that the n rows of the $n \times (l+1)$ matrix of reduced-form errors $V = [v_1, V_2]$ are i.i.d. normally distributed with mean zero and known nonsingular covariance matrix $\Omega = [\omega_{i,j}]$.

The goal here is to test the null hypothesis $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$, treating Π, Γ, and δ as nuisance parameters. 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 level α if, when the null hypothesis is true,

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

for all admissible values of the nuisance parameters. Since the nuisance parameters 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, making the null rejection probability equal α. In that case, the test is said to be *similar* and $\mathcal{T}$ is said to be *pivotal*. If $\mathcal{T}$ has null distribution dependent on nuisance parameters but can be bounded by a pivotal statistic, then $\mathcal{T}$ is said to be *boundedly pivotal*.

Although structural coefficient tests based on pivotal statistics have been proposed in the literature, the Wald and likelihood ratio statistics most commonly employed in practice are nonpivotal. Under regularity conditions, both statistics are asymptotically chi-square-l and tests using their $1-\alpha$ quantile for c are asymptotically similar with size α. However, when β is almost unidentified,[3] the actual null rejection probability can differ substantially from α since the asymptotic approximation can be very poor when the instruments are weakly correlated with Y_2.

[3] The structural coefficients are unidentified when rank $(\Pi) < l$. The coefficients are almost unidentified when Π is in a small neighborhood around a matrix with rank less than l.

One possible solution to the problem that results from using nonpivotal statistics is to replace the asymptotic chi-square critical value with some larger, conservative value, that guarantees 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 identification is good, these tests have null rejection probabilities much lower than α and reduce power unnecessarily. Moreover, this approach is fruitless for statistics that are not boundedly pivotal. Here we will develop an alternative procedure that allows us to construct tests that are exactly similar.

2.2. *Similar Tests Based on Conditioning*

When Ω is known and the errors are normal, the probability model for Y, given $[Z, X_1]$, is a member of the curved exponential family, and the $k \times (l+1)$ matrix $[Z, X_1]'Y$ is a sufficient statistic for the unknown parameters. Hence, we can restrict attention to tests that depend on Y only through $Z'Y$ and $X_1'Y$. The nuisance parameters δ and Γ can be eliminated by considering tests that depend only on $Z'Y$. This restriction can be justified by requiring the test to be invariant to transformations $g(Y) = Y + X_1 F$ for arbitrary conformable matrices F. For this group of linear transformations of X_1, the maximal invariant in terms of the sufficient statistic is exactly $Z'Y$. Lehmann (1986, Chapter 6) explains the use of invariance in simplifying a hypothesis testing problem.

For any known nonsingular, nonrandom $(l+1) \times (l+1)$ matrix $D, Z'YD$ is also an invariant sufficient statistic. A convenient choice is the matrix

$$D_0 = [b_0, \Omega^{-1} A_0],$$

where b_0 is the $(l+1) \times 1$ vector $[1, -\beta_0']'$ and A_0 is the $(l+1) \times l$ matrix $[\beta_0, I_l]'$. Note that every column of A_0 is orthogonal to b_0. Then the invariant sufficient statistic can be represented by the pair $[S, T]$ where

$$S = Z'Yb_0 = Z'(y_1 - Y_2\beta_0) \quad \text{and} \quad T = Z'Y\Omega^{-1}A_0.$$

The k-dimensional vector S is normally distributed with mean $Z'Z\Pi(\beta - \beta_0)$ and covariance matrix $Z'Zb_0'\Omega b_0$. The $k \times l$ matrix T is independent of S, and $\mathrm{vec}(T)$ is normally distributed with mean $\mathrm{vec}(Z'Z\Pi A'\Omega^{-1}A_0)$ and covariance matrix $A_0'\Omega^{-1}A_0 \otimes Z'Z$, where $A = [\beta, I_l]'$. Thus we have partitioned the invariant sufficient statistic $Z'Y$ into two independent, normally distributed statistics, S having a null distribution not dependent on Π and T having a null distribution dependent on Π. Indeed, when β is known to equal β_0, T is a sufficient statistic for Π and is a one-to-one function of the constrained maximum likelihood estimator $\widehat{\Pi}$:

$$\widehat{\Pi} = (Z'Z)^{-1}T(A_0'\Omega^{-1}A_0)^{-1}.$$

Let $\psi(S, T, \Omega, \beta_0)$ be a statistic for testing the hypothesis that $\beta = \beta_0$ (the statistic may also depend on Z, but that dependency will be ignored in this section). If the null distribution of ψ depends on Π, a test that rejects H_0 when ψ lies in some fixed region will not be similar. Nevertheless, following an approach suggested by the analysis in Lehmann (1986, Chapter 4), it is easy to construct a similar test based on ψ. Although the marginal distribution of ψ may depend on Π, the independence of S and T implies that 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. Thus we have the following result:

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

It is shown in Moreira (2001) that $c_\psi(T, \Omega, \beta_0, \alpha)$ does not depend on T when ψ is pivotal. Thus, the conditional approach for finding a similar test may be thought of as replacing the nonpivotal statistic $\psi(S, T, \Omega, \beta_0)$ by the new statistic $\psi(S, T, \Omega, \beta_0) - c_\psi(T, \Omega, \beta_0, \alpha)$. Alternatively, since conditioning on T is the same as conditioning on $\widehat{\Pi}$, this approach may be interpreted as adjusting the critical value based on a preliminary estimate of Π. Henceforth, $c_\psi(T, \Omega, \beta_0, \alpha)$ will be referred to as the critical value function for the test statistic ψ.

To illustrate the conditional approach, we now consider a number of examples.

EXAMPLE 1: Anderson and Rubin (1949) propose testing the null hypothesis using the statistic S. Since S has zero mean and variance proportional to $Z'Z$ when $\beta = \beta_0$, it is natural to reject the null hypothesis when $S'(Z'Z)^{-1}S$ is large. The Anderson-Rubin statistic for known Ω is

$$AR_0 = S'(Z'Z)^{-1}S/\sigma_0^2,$$

where $\sigma_0^2 = b_0'\Omega b_0$, is the variance of each element of $u_0 \equiv y_1 - y_2\beta_0$. This statistic is distributed chi-square-k under the null hypothesis and it is consequently pivotal. Its conditional distribution given $T = t$ does not depend on t and its critical value function collapses to a constant

$$c_{AR}(t, \Omega, \beta_0, \alpha) = q_\alpha(k),$$

where $q_\alpha(df)$ is the $1-\alpha$ quantile of a chi-square distribution with df degrees of freedom. Moreira (2001) shows that the Anderson-Rubin test is optimal when the model is just-identified. However, this test has poor power properties when the model is over-identified, since the number of degrees of freedom is larger than the number of parameters being tested.

EXAMPLE 2: Consider a particular score statistic

$$LM_0 = S'\widehat{\Pi}\big[\widehat{\Pi}'Z'Z\widehat{\Pi}\big]^{-1}\widehat{\Pi}'S/\sigma_0^2.$$

Breusch and Pagan (1980) propose a score-type statistic in a general framework (including nonlinear models) that reduces to LM_0 in our model. Kleibergen (2002) and Moreira (2001) show that $\widehat{\Pi}$ is independent of S and, consequently, the null distribution of LM_0 is chi-square-l. A Lagrange multiplier test that rejects H_0 for large values of LM_0 has correct null rejection probability as long as the appropriate critical value is used. Again, the score test statistic is pivotal and its critical value function collapses to a constant

$$c_{LM}(t, \Omega, \beta_0, \alpha) = q_\alpha(l).$$

Like the Wald and likelihood ratio tests, this score test is (locally) asymptotically optimal when the structural parameters are identified.

EXAMPLE 3: The Wald statistic centered around the 2SLS estimator is given by

$$W_0 = (b_{2SLS} - \beta_0)' Y_2' N_Z Y_2 (b_{2SLS} - \beta_0)/\hat{\sigma}^2,$$

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. In Appendix B, the critical value function for W_0 is shown to simplify:

$$c_W(T, \Omega, \beta_0, \alpha) = \bar{c}_W(\tau, \Omega, \beta_0, \alpha),$$

where $\tau \equiv (A_0' \Omega^{-1} A_0)^{-1/2} t' (Z'Z)^{-1} t (A_0' \Omega^{-1} A_0)^{-1/2}$.

EXAMPLE 4: The likelihood ratio statistic, for known Ω, is defined as

$$LR_0 = 2\Big[\max_{\beta, \Pi} L(Y; \beta, \Pi, \Omega) - \max_{\Pi} L(Y; \beta_0, \Pi, \Omega)\Big],$$

where L is the log likelihood function after concentrating out δ and Γ. Various authors have noticed that, in curved exponential models, the likelihood ratio test performs well for a wide range of alternatives; see, for example, Van Garderen (2000). In Appendix B we show that the critical value function for the likelihood ratio test has the form

$$c_{LR}(T, \Omega, \beta_0, \alpha) = \bar{c}_{LR}(\tau, \alpha);$$

that is, it is independent of Ω and β_0.

To implement the conditional test based on a nonpivotal statistic ψ, we need to compute the conditional quantile $c_\psi(t, \Omega, \beta_0, \alpha)$. Although in principle the entire critical value function can be derived from the known null distribution of S, for most choices of ψ a simple analytical expression seems out of reach. A Monte Carlo simulation of the null distribution of S is much simpler. Indeed, the applied researcher need only do a simulation for the actual $k \times l$ matrix t

observed in the sample and for the particular β_0 being tested; there is no need to derive the whole critical value function $c_\psi(t, \Omega, \beta_0, \alpha)$.

The critical value function for a given test statistic ψ will generally depend on the $k \times l$ matrix t. However, as noted above, the critical value functions for the Wald and likelihood ratio statistics depend on t only through the $l \times l$ matrix τ. This can be a considerable simplification when $k - l$ is large. In particular, when there is only one endogenous variable on the right-hand side of (1), τ reduces to a scalar. See Appendix B for a more thorough exposition on how to compute $\bar{c}_\psi(\tau, \Omega, \beta_0, \alpha)$.

3. THE CONDITIONAL LIKELIHOOD RATIO TEST

We can now elaborate more detailed expressions for the conditional likelihood ratio test. In Appendix A we show that, when Ω is known, the likelihood ratio statistic is given by

$$\begin{aligned} LR_0 &= \frac{b_0' Y' N_Z Y b_0}{b_0' \Omega b_0} - \min_b \frac{b' Y' N_Z Y b}{b' \Omega b} \\ &= \frac{b_0' Y' N_Z Y b_0}{b_0' \Omega b_0} - \bar{\lambda}^{\min}, \end{aligned}$$

where b is the $(l+1) \times 1$ vector $[1, -\beta']'$ and $\bar{\lambda}^{\min}$ is the smallest eigenvalue of $\Omega^{-1/2} Y' N_Z Y \Omega^{-1/2}$. This expression can be simplified somewhat when written in terms of the standardized statistics

$$\overline{S} = (Z'Z)^{-1/2} S (b_0' \Omega b_0)^{-1/2} \quad \text{and} \quad \overline{T} = (Z'Z)^{-1/2} T (A_0' \Omega^{-1} A_0)^{-1/2}.$$

Under the null hypothesis, $\overline{S}$ has mean zero so $\overline{S}'\overline{S}$ has chi-square distribution with k degrees of freedom. The statistic $\overline{T}'\overline{T}$ is distributed as noncentral Wishart with noncentrality related to $\Pi' Z' Z \Pi$; it can be viewed as a natural statistic for testing the hypothesis that $\Pi = 0$ under the assumption that $\beta = \beta_0$. Then, we find

$$LR_0 = \overline{S}'\overline{S} - \bar{\lambda}^{\min},$$

where $\bar{\lambda}^{\min}$ is also the smallest eigenvalue of $(\overline{S}, \overline{T})'(\overline{S}, \overline{T})$.

A further simplification is possible when $l = 1$. In this case, β is a scalar, the $k \times l$ matrix Π reduces to a k-dimensional vector π, and the matrix A_0 simplifies to the vector $a_0 = [\beta_0, 1]'$. In Appendix C, the likelihood ratio statistic is shown to be given by

$$LR_0 = \frac{1}{2}\left[\overline{S}'\overline{S} - \overline{T}'\overline{T} + \sqrt{[\overline{S}'\overline{S} + \overline{T}'\overline{T}]^2 - 4[\overline{S}'\overline{S} \cdot \overline{T}'\overline{T} - (\overline{S}'\overline{T})^2]}\right]. \tag{3}$$

When $k = 1$, $\overline{S}$ and $\overline{T}$ are scalars, and the LR_0 statistic collapses to the pivotal Anderson-Rubin statistic $\overline{S}'\overline{S}$. In the overidentified case, the LR_0 statistic depends also on $\overline{T}$ and is no longer pivotal.

Even in the special case $l = 1$, an analytic expression of the critical value function for the LR_0 statistic is not available. However, some general properties of the function are known.

PROPOSITION 1: *When $l = 1$, the critical value function for the conditional LR_0 test is a decreasing function of the scalar $\tau = \bar{t}'\bar{t}$, satisfying*

$$\bar{c}_{LR}(\tau, \alpha; k) \to q_\alpha(1) \quad as \quad \tau \to \infty,$$
$$\bar{c}_{LR}(\tau, \alpha; k) \to q_\alpha(k) \quad as \quad \tau \to 0.$$

Table I 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. The slight variation in the first column of Table I is due to simulation error. For each k, the critical value function has approximately an exponential shape. For example, when $k = 4$, $\bar{c}_{LR}(\tau, 0.05; 4)$ is well approximated by the function $3.84 + 5.65 \cdot \exp(-\tau/7)$. When the vector π is far from the origin (and hence identification is strong), τ tends to take on a large value, and the conditional likelihood ratio test behaves like the unconditional likelihood ratio test. When π is near the origin (and hence identification is weak), τ tends to take on a small value and the appropriate critical value is larger.

The conditional method connects and builds on previous work. First, the shape of the critical value function indicates why the method proposed by Wang and Zivot (1998) leads to a test with low power. Their critical value based on the $1 - \alpha$ chi-square-k quantile is the upper bound for the true critical value function $\bar{c}_{LR}(\tau, \alpha; k)$. Second, this critical value function can 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 null rejection probability without unnecessarily wasting

TABLE I
CRITICAL VALUE FUNCTION OF THE LIKELIHOOD RATIO TEST

	k							
τ	1	2	3	4	5	10	20	50
0	3.96	5.88	7.79	9.45	11.19	18.26	31.23	67.42
1	3.80	5.64	7.13	8.67	10.25	17.39	30.72	66.18
5	3.85	4.48	5.45	6.49	7.54	13.90	26.74	62.35
10	3.83	4.27	4.64	5.09	5.79	10.37	21.87	58.35
20	3.88	3.97	4.21	4.38	4.74	6.41	14.07	47.98
50	3.85	3.91	4.01	4.20	4.10	4.73	6.15	21.44
75	3.73	3.85	4.02	3.88	4.04	4.36	5.11	10.13
100	3.86	3.74	3.79	3.93	4.06	4.22	4.76	7.36
50000	3.88	3.90	3.92	3.71	3.88	3.74	3.84	3.88

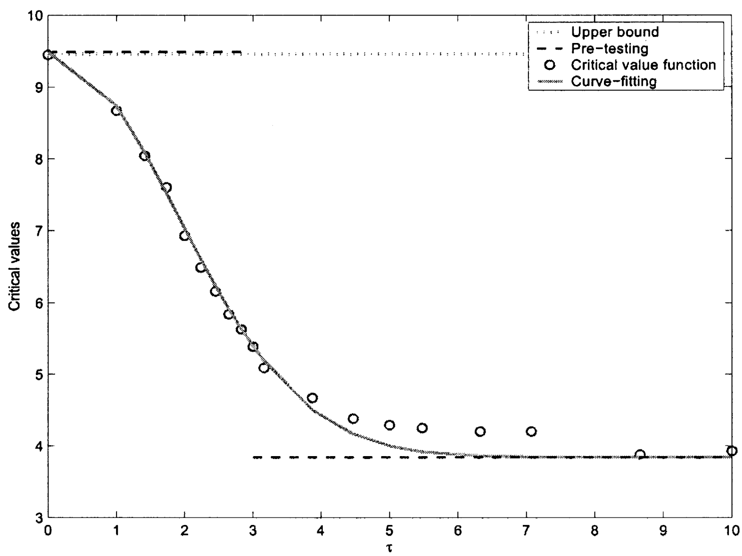

FIGURE 1.—Critical value function for $k = 4$.

power. Figure 1 illustrates each method, sketching its respective critical values[4] for different values of τ when there are four instruments.

4. UNKNOWN REDUCED-FORM ERROR DISTRIBUTION

In practice, of course, the reduced-form covariance matrix is unknown. Furthermore, there is no compelling reason to believe that the errors are exactly normally distributed. However, since Ω can be well estimated even when identification is weak, it is plausible to replace Ω by a consistent estimator in the tests developed in Section 2. A natural choice is the unrestricted least squares estimator $\widehat{\Omega} = Y'M_X Y/(n-k)$. Considering that standardized sums of independent random variables should be approximately normal, the modified tests are expected to behave well in moderately sized samples even with non-normal errors. Thus, for a statistic ψ, one might reject the null hypothesis that $\beta = \beta_0$ when

$$\psi(S, \widehat{T}, \widehat{\Omega}, \beta_0) > c_{\psi}(\widehat{T}, \widehat{\Omega}, \beta_0, \alpha),$$

where $\widehat{T} = Z'Y\widehat{\Omega}^{-1}A_0$. The critical value function could again be obtained by the appropriate quantile of ψ from a simulation where the randomness of $\widehat{\Omega}$ is ignored and S is drawn from a normal distribution with mean zero and covariance matrix $Z'Zb_0'\widehat{\Omega}b_0$.

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

For example, replacing Ω by $\widehat{\Omega}$ in the score statistic in Example 2, we obtain the *LM* test, which is the score test used by Kleibergen (2002) and Moreira (2001). Analogously, we can substitute Ω for $\widehat{\Omega}$ in the likelihood ratio statistic LR_0:

$$LR_1 = \frac{b_0' Y' N_Z Y b_0}{b_0' \widehat{\Omega} b_0} - \hat{\lambda}^{\min},$$

where $\hat{\lambda}^{\min}$ is the smallest eigenvalue of $\widehat{\Omega}^{-1/2} Y' N_Z Y \widehat{\Omega}^{-1/2}$. Alternatively, we can use the actual likelihood ratio statistic for the normal distribution with unknown variance,

$$LR = \frac{n}{2} \ln\left(1 + \frac{b_0' Y' N_Z Y b_0}{b_0' Y' M_X Y b_0}\right) - \frac{n}{2} \ln\left(1 + \frac{\hat{\lambda}^{\min}}{n-k}\right).$$

Simulations suggest that, even for relatively small samples, the LR_1 and LR statistics are close to the LR_0 statistic. Therefore, the critical values in Table I can be used for the conditional LR_1 and LR tests (respectively, LR_1^* and LR^*) by replacing τ by

$$\hat{\tau} = \left(A_0' \widehat{\Omega}^{-1} A_0\right)^{-1/2} \hat{t}' (Z'Z)^{-1} \hat{t} (A_0' \widehat{\Omega}^{-1} A_0)^{-1/2}.$$

In the next section we show that this substitution can be justified asymptotically even when identification is weak and when the assumptions of normal errors and exogenous instruments are relaxed.

4.1. *Weak-Instrument Asymptotics*

To examine the approximate properties of test statistics and estimators in models where identification is weak, Staiger and Stock (1997) consider the "weak-instrument" asymptotics. In these nonconventional asymptotics, the matrix Π converges to the zero matrix as the sample size n increases. Using this approach, we find that, under some regularity conditions, the limiting rejection probabilities of our conditional tests based on an estimated Ω equal the exact rejection probabilities when the errors are normal with known variance. This implies that our tests are asymptotically similar no matter how weak the instruments.

THEOREM 2: *Consider the simultaneous equations model in Section* 2. *Suppose*:

(i) $Z'Z/n \xrightarrow{p} Q$ *where* Q *is positive definite and* $Z'V/\sqrt{n} \xrightarrow{d} \Psi_{zv}$ *where* $\text{vec}(\Psi_{zv}) \sim N(0, \Omega \otimes Q)$.

(ii) $\Pi = C/\sqrt{n}$, *where* C *is a fixed* $l \times k$ *matrix*.

(iii) $\bar{\psi}$ *is continuous function that satisfies the homogeneity condition*

$$\bar{\psi}(S, \widehat{T}, Z'Z, \widehat{\Omega}, \beta_0) = \bar{\psi}(n^{-1/2} S, n^{-1/2} \widehat{T}, n^{-1} Z'Z, \widehat{\Omega}, \beta_0).$$

(iv) *The critical value function* $c_{\bar{\psi}}$ *derived under the assumption of normality and known* Ω *is continuous*.

Then the conditional test based on the statistic $\bar{\psi}(S, \widehat{T}, Z'Z, \widehat{\Omega}, \beta_0)$ *has limiting rejection probability equal to the exact rejection probability derived under the assumption of normal reduced-form disturbances with known variance*.

Assumption (i) is similar to that made in the standard asymptotic theory for instrumental variable estimation. If Z is nonrandom with bounded elements and the errors are i.i.d. with finite second moment, the first part of assumption (i) implies the second part. Theorem 2 also allows for the case of lagged endogenous variables as long as we adapt the convergence rates. Endogenous variables that contain unit roots are already ruled out because of the normality of the limiting distribution of $Z'V/\sqrt{n}$. Of course, approximations based on (i) can be poor in small samples if the error distribution has very thick tails. Assumption (ii) states that the coefficients on the instruments are in the neighborhood of zero. Note that C is allowed to be the zero matrix so that Theorem 2 holds even when the structural parameter is not identified. Assumptions (iii) and (iv) appear to be satisfied for all the commonly proposed test statistics, including the Anderson-Rubin, score, conditional likelihood ratio, and conditional Wald tests.

Theorem 2 asserts that the conditional likelihood ratio test is similar under the weak instrument asymptotics. When identification is not weak, the usual asymptotic arguments can be applied to show that our conditional likelihood ratio test is asymptotically similar when $l = 1$. In this case, Engle (1984) shows that the likelihood ratio statistic is asymptotically chi-square-one and Proposition 1 asserts that the critical value function converges to the usual asymptotic chi-square-one critical value. Furthermore, we can expect the null rejection probability of the conditional likelihood ratio test to converge uniformly under some regularity conditions.[5] Following Andrews (1986, p. 267), we can guarantee uniform convergence if, under the null hypothesis, the finite-sample power functions $\{\zeta_n(\pi, \Omega); n \geq 1\}$ are equicontinuous over the compact set $\mathbf{K} = \mathbf{P} \times \boldsymbol{\Omega}$ in which (π, Ω) takes values. Here, the set $\mathbf{P}$ can include the nonidentification case $(\pi = 0)$ and $\boldsymbol{\Omega}$ is a set of invertible 2×2 matrices. For a more thorough exposition of equicontinuity and uniform convergence, see Parzen (1954).

4.2. *Power Comparison*

To assess the performance of the conditional approach, we examine the performance of the four tests described in Section 2: the conditional likelihood ratio test (denoted as LR^*), the conditional Wald test based on the 2SLS estimator (W^*), the Anderson-Rubin test (AR), and the score test (LM). Using Theorem 2, the asymptotic power for each test is computed following Design I of Staiger and Stock (1997). In this design, $l = 1$ and $r = 0$ so the structural equation has only one explanatory variable. The hypothesized value β_0 for its coefficient is taken to

[5] Rothenberg (1984) and Horowitz and Savin (2000) discuss the problem of size distortions due to asymptotic approximations.

be zero. The elements of the matrix Z are drawn as independent standard normal random variables and then held fixed. Two different values of the π vector are used so that $\lambda'\lambda/k = \pi'Z'Z\pi/(\omega_{22}k)$, the "population" first-stage F-statistic (in the notation of Staiger and Stock), 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. Here, we report only results for $\rho = 0.50$, although we have considered different degrees of endogeneity of y_2.

Figures 2 and 3 graph the rejection probabilities of these four tests as functions of the true value β, respectively, for $k = 4$ and $k = 10$.[6] 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 under the weak-instrument asymptotics. As expected, the asymptotic power curves become steeper as the quality of instruments improves.

The AR test has poor power when the number of instruments is large. Although the LM, W^*, and LR^* tests are optimal under the local-to-null asymptotics, some of these tests do not have good power when instruments are weak. The LM test has relatively low power both for the weak-instrument case and for some values of β for the good-instrument case. The W^* test is biased, reflecting the finite-sample bias of the $2SLS$ estimator. These poor power properties are not shared by the conditional likelihood ratio test. The LR^* test not only seems to dominate the Anderson-Rubin and score tests[7] under the weak-instrument asymptotics, but is optimal under the usual asymptotics.

4.3. *Monte Carlo Simulations*

Theorem 2 shows that, under some regularity assumptions, the conditional approach leads to asymptotically similar tests even when the errors are nonnormal and the reduced-form covariance matrix is estimated. In this section, we present some evidence suggesting that the weak-instrument asymptotics work quite well in moderately sized samples. To evaluate the actual rejection probability under H_0, 1,000 Monte Carlo simulations were performed based on Design I of Staiger and Stock (1997) for 80 observations. Results are reported for ρ taking the values 0.00, 0.50, and 0.99.

Table II presents null rejection probabilities for the following tests: Anderson-Rubin (AR),[8] the Hessian-based score test (LM_H), the score test used by Kleibergen (2002) and Moreira (2001) (LM), 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 critical value functions for the conditional tests at 5% nominal level were based on 1,000 replications.

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

[7] Other tests 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, their conditional counterparts seem to have asymptotic power no larger than the conditional likelihood ratio test.

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

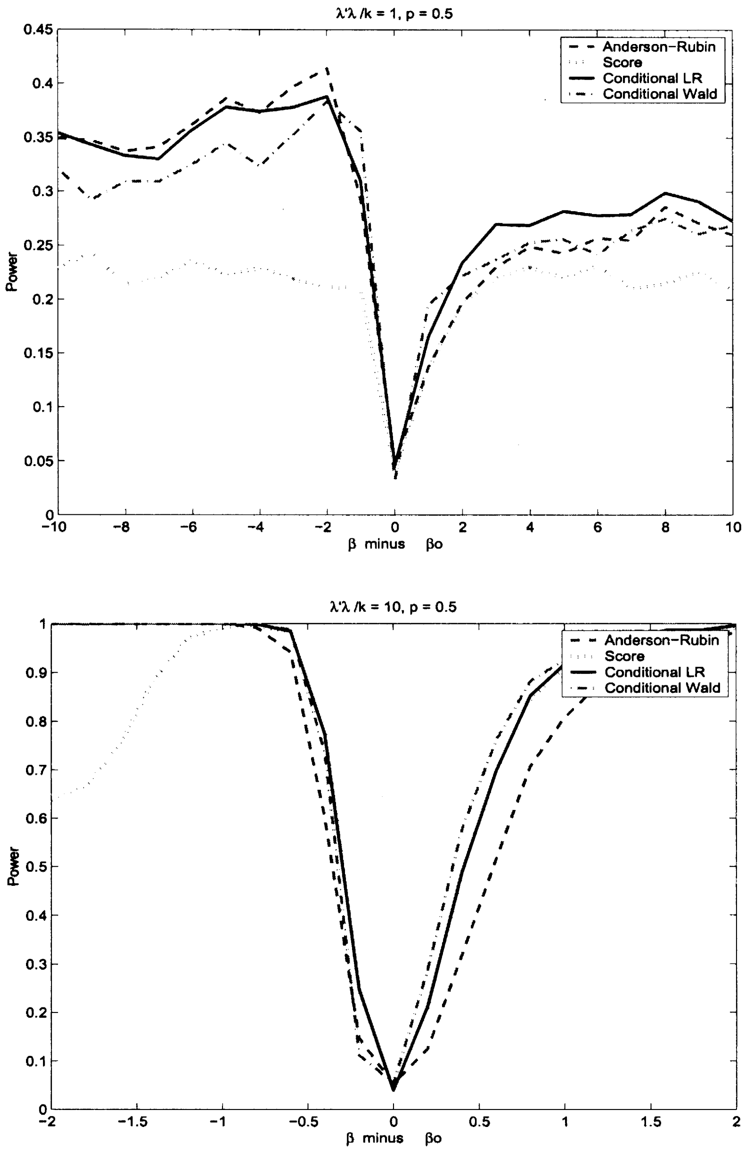

FIGURE 2.—Asymptotic power of tests: 4 instruments.

Recall that the AR, LM, LR^*, and W^* tests are similar under the weak-instrument asymptotics, whereas the LM_H, LR, and W tests are not. Indeed, Table II shows that the LM_H test does not have null rejection probability close to the 5% nominal level, whereas the LM test does. Likewise, the LR and W tests perform more poorly 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

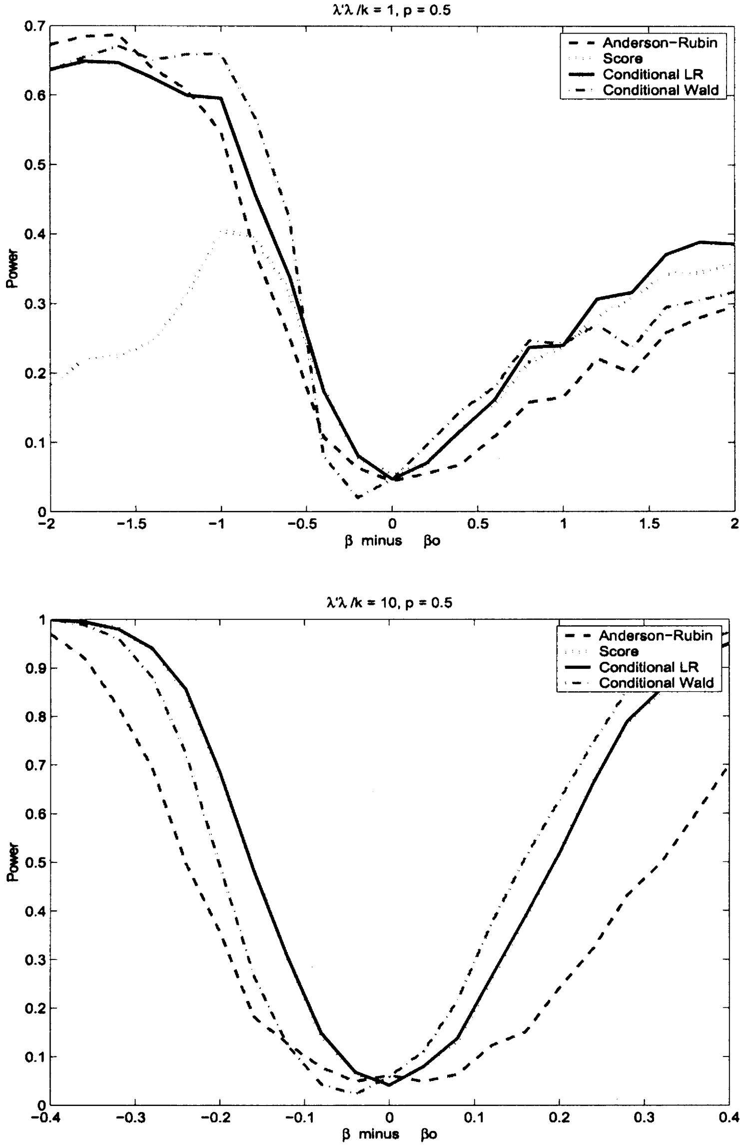

FIGURE 3.—Asymptotic power of tests: 10 instruments.

test range from 0.002–0.992. 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 analogous.[9] Table III shows the null rejection probabilities of some 5% tests when Staiger and Stock's Design II

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

TABLE II
PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5%
NORMAL DISTURBANCES

ρ	$\lambda'\lambda/k$	AR	LM_H	LM	LR	LR^*	W	W^*
0.00	0.00	6.20	4.00	5.40	20.70	5.00	0.20	3.00
0.00	1.00	5.30	4.90	5.80	16.20	6.30	1.00	5.20
0.00	10.00	6.10	4.60	4.70	5.80	4.60	3.30	4.00
0.50	0.00	6.70	13.00	5.70	22.00	5.60	13.00	5.10
0.50	1.00	6.10	9.00	5.50	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	6.50	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	5.90	6.00	6.10	13.40	5.80

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 an equal number of observations in each cell. 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, Table IV compares the asymptotic power with the actual power of the conditional LR^* test when Staiger and Stock's Design I with 80 observations is used for the parameters $\lambda'\lambda/k = 1.00$ and $\rho = 0.50$. The difference between the two power curves is small, which suggests that the weak-instrument asymptotics work quite well. Similar results not reported here were obtained using other tests and other designs.

5. CONFIDENCE REGIONS

Confidence regions for β with approximately correct coverage probability can be constructed by inverting approximately similar tests. Although Dufour (1997), building on work by Gleser and Hwang (1987), shows that Wald-type confidence

TABLE III
PERCENT REJECTED UNDER H_0 AT NOMINAL LEVEL OF 5%
NON-NORMAL DISTURBANCES

ρ	$\lambda'\lambda/k$	AR	LM_H	LM	LR	LR^*	W	W^*
0.00	0.00	6.20	4.40	5.80	23.80	5.90	0.30	3.80
0.00	1.00	6.40	4.00	5.90	22.50	6.50	0.20	3.80
0.00	10.00	5.90	7.30	8.50	12.10	8.10	2.90	7.10
0.50	0.00	7.20	8.60	6.80	23.40	7.90	4.40	5.60
0.50	1.00	6.50	6.70	6.60	21.80	7.50	3.10	5.40
0.50	10.00	6.70	6.70	7.30	10.80	7.40	4.20	5.40
0.99	0.00	7.60	41.30	7.60	24.30	7.90	96.90	7.00
0.99	1.00	6.60	29.20	7.30	8.40	7.00	81.20	5.60
0.99	10.00	5.70	11.10	6.80	6.70	7.20	27.40	3.10

TABLE IV
PERCENT REJECTED AT NOMINAL LEVEL OF 5%
CONDITIONAL LIKELIHOOD RATIO TEST

β	Asymptotic Power	Actual Power
−10.00	35.40	36.70
−8.00	33.30	35.20
−6.00	35.70	36.10
−4.00	37.40	36.80
−2.00	38.80	39.90
0.00	4.60	4.90
2.00	23.40	23.30
4.00	26.90	27.50
6.00	27.80	29.40
8.00	29.90	30.50
10.00	27.30	29.40

intervals are not valid when identification can be arbitrarily weak, the confidence regions based on the conditional Wald test have correct coverage probability no matter how weak the instruments are. Likewise, if the score or conditional likelihood ratio tests are used, the resulting confidence regions have approximately correct levels. Moreover, the regions based on the conditional Wald test necessarily contain the 2SLS estimator of β, while those based on the conditional likelihood ratio or score tests are centered around the LIML estimator of β. Therefore, confidence regions based on these tests can be used as evidence of the accuracy of their respective estimators. For example, Cruz and Moreira (2002) employ the conditional tests to reassess the accuracy of the estimates of returns to schooling by Angrist and Krueger (1991).

To illustrate how informative the confidence regions based on the conditional likelihood ratio test are when compared with those based on the score test, Design I of Staiger and Stock (1997) is once more used. One sample is drawn in which the true value of β is zero and $\rho = 0.50$. Figure 4 plots the likelihood ratio and score statistics and their respective critical value functions at the significance level of 5% against β_0.[10] The region in which each statistic is below its critical value curve is the corresponding confidence set.

Figure 4 suggests that the LR^* confidence regions are considerably smaller than those of LM, as a result of the better power properties of the conditional likelihood ratio test. When $\lambda'\lambda/k = 1$, the conditional likelihood ratio confidence region is the set $[-1.02, 1.37]$, while the score confidence region is the nonconvex set $[-\infty, -6.72] \cup [-0.57, 1.12] \cup [2.58, \infty]$. When $\lambda'\lambda/k = 10$, the conditional likelihood ratio confidence region is the set $[-0.45, 0.18]$ while the score confidence region is the set $[-0.45, -0.18] \cup [1.35, 1.60]$.

[10] Here, we run 10,000 Monte Carlo replications to compute each point of the critical value function.

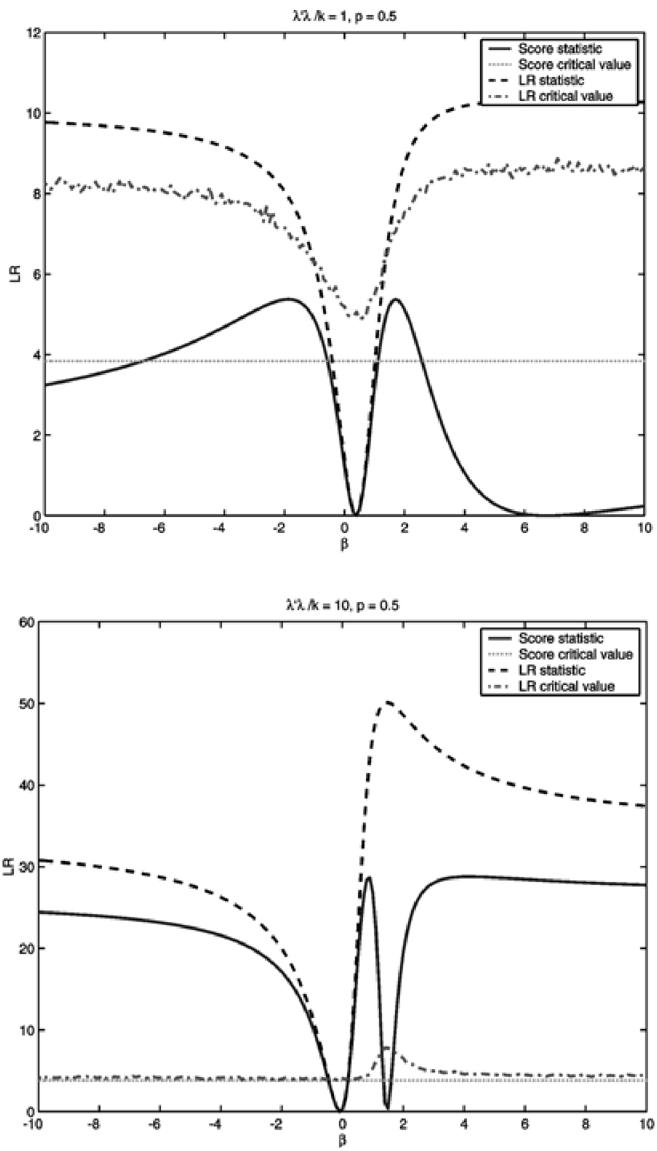

FIGURE 4.—Confidence regions.

In both cases, the score test fails to reject some nonlocal yet relevant alternatives. As noted by Kleibergen (2002), the bad performance of the LM confidence region can be partially explained by the fact that the score statistic equals zero at two points, both satisfying the quadratic (in β_0) expression:

$$a_0'\widehat{\Omega}^{-1}Y'N_Z Y b_0 = 0.$$

6. CONCLUSIONS

Previous authors, e.g. Anderson, Kunitomo, and Sawa (1982), have noted that the simultaneous equations model with known reduced-form covariance matrix has a simpler mathematical structure than the model with unknown covariance matrix, but inference procedures for the two models behave very much alike in moderately sized samples. Based on this fact, Moreira (2001) applies classical statistical theory to characterize the whole class of similar tests with normal errors and known covariance matrix. Exploiting this finding, we develop a general procedure for constructing valid tests of structural coefficients based on the conditional distribution of nonpivotal statistics. Replacing the unknown covariance matrix by a consistent estimator appears to have little effect on the null rejection probability and on power.

Even with non-normal errors, the proposed conditional (pseudo) likelihood ratio test has correct null rejection probability when identification is weak, and good power when identification is strong. This test is equivalent to the usual likelihood ratio test under the usual asymptotics. Moreover, power comparisons using weak-instrument asymptotics suggest that this test dominates other asymptotically similar tests such as the Anderson-Rubin test and a particular score test.

Like the Anderson-Rubin and score approaches, the conditional tests proposed here attain similarity under arbitrarily weak identifiability only when all the unknown endogenous coefficients are tested. Inference on the coefficient of one endogenous variable when the structural equation contains additional endogenous explanatory variables is not allowed. Dufour (1997) shows how this limitation can be overcome in the context of the Anderson-Rubin test, and the same projection approach presumably could be applied here. However, this may entail considerable loss of power.

Finally, 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, Harvard University, Littauer Center M-6, 1875 Cambridge Street, Cambridge, MA 02138 USA; moreira@fas.harvard.edu; http://post.economics.harvard.edu/marcelo/moreira.html

Manuscript received November, 2001; final revision received December, 2002.

APPENDIX A: LIKELIHOOD RATIO DERIVATION

Ignoring an additive constant and assuming normal errors, the log-likelihood function (after concentrating out δ and Γ) can be written as

$$L(Y;\beta,\Pi,\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}[\text{tr}(\Omega^{-1}V'M_{X_1}V)], \tag{A.1}$$

where $V = Y - Z\Pi_* - X_1[\delta, \Gamma]$ and $\Pi_* = \Pi A'$. Using Lagrange multipliers to maximize L with respect to Π_* subject to the constraint that $\Pi_* b = 0$, we find that $\Pi_*(\beta, \Omega) = (Z'Z)^{-1}Z'Y[I - b(b'\Omega b)^{-1}b'\Omega]$. The concentrated log-likelihood function, $L_c(Y; \beta, \Omega)$, defined as $L(Y; \beta, \Pi(\beta, \Omega), \Omega)$, is given by

$$L_c(Y; \beta, \Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}\left[\mathrm{tr}(\Omega^{-1}Y'M_X Y) + \frac{b'Y'N_Z Yb}{b'\Omega b}\right],$$

where $X = (X_1, X_2)$. When evaluated at $\hat{\beta}$, the maximum likelihood estimator of β when Ω is known, this becomes

$$L_c(Y; \hat{\beta}, \Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}[\mathrm{tr}(\Omega^{-1}Y'M_X Y) + \bar{\lambda}^{\min}],$$

where $\bar{\lambda}^{\min}$ is the smallest eigenvalue of $\Omega^{-1/2}Y'N_Z Y\Omega^{-1/2}$. It follows that the likelihood ratio statistic when Ω is known, LR_0, is

$$LR_0 = \overline{S}'\overline{S} - \bar{\lambda}^{\min}. \tag{A.2}$$

To find the likelihood ratio when Ω is unknown, we maximize (A.1) with respect to Ω, obtaining $\Omega(\beta, \Pi) = (Y - Z\Pi_*)'M_{X_1}(Y - Z\Pi_*)/n$. Inserting this into (A.1) and dropping an additive constant, we obtain

$$L^*(Y; \beta, \Pi, \Omega(\beta, \Pi)) = -\frac{n}{2}\ln|V'M_{X_1}V|.$$

Using standard facts about determinants, we find that the maximum value of the log-likelihood function for a fixed β is given by

$$L_c^*(Y; \beta) = -\frac{n}{2}\ln\left(1 + \frac{b'Y'N_Z Yb}{b'Y'M_X Yb}\right) - \frac{n}{2}\ln|Y'MY|.$$

Moreover, the concentrated log-likelihood function evaluated at the maximum likelihood estimator β_{LIML} is then given by

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

where $\lambda^{\min}$ is the smallest eigenvalue of $(Y'M_X Y)^{-1/2}Y'N_Z Y(Y'M_X Y)^{-1/2}$. Since the LR, the likelihood-ratio statistic when Ω is unknown, is defined as $2[L_c^*(Y; \beta_{LIML}) - L_c^*(Y; \beta_0)]$, it follows that

$$LR = n\ln\left(1 + \frac{b_0'Y'N_Z Yb_0}{b_0'Y'M_X Yb_0}\right) - n\ln\left(1 + \frac{\lambda^{\min}}{n-k}\right).$$

APPENDIX B: Critical Value Function

As in Section 3, we define the standardized statistics

$$\overline{S} = (Z'Z)^{-1/2}S(b_0'\Omega b_0)^{-1/2} \quad \text{and} \quad \overline{T} = (Z'Z)^{-1/2}T(A_0'\Omega^{-1}A_0)^{-1/2}.$$

Suppose that a statistic $\bar{\psi}(S, T, Z'Z, \Omega, \beta_0)$ is such that it depends on S and T only through $\overline{S}'\overline{S}$, $\overline{S}'\overline{T}$, and $\overline{T}'\overline{T}$. That is, for a suitable function φ we have:

$$\bar{\psi}(S, T, Z'Z, \Omega, \beta_0) = \varphi(\overline{S}'\overline{S}, \overline{S}'\overline{T}, \overline{T}'\overline{T}, \Omega, \beta_0).$$

Note that $\overline{S}'\overline{S} = \overline{S}'M_{\overline{T}}\overline{S} + \overline{S}'\overline{T}(\overline{T}'\overline{T})^{-1}\overline{T}'\overline{S}$. Thus, there exists a function $\bar{\varphi}$ such that

$$\bar{\psi}(S, T, Z'Z, \Omega, \beta_0) = \bar{\varphi}(\overline{S}'M_{\overline{T}}\overline{S}, \overline{S}'\overline{T}, \overline{T}'\overline{T}, \Omega, \beta_0).$$

Now, conditional on $\overline{T} = \bar{t}$, $\overline{T}'\overline{S}$ and $\overline{S}'M_{\overline{T}}\overline{S}$ are independent with a $N(0, \tau)$ distribution and chi-square distribution with $k-l$ degrees of freedom under the null hypothesis. Therefore, for fixed k and l, the critical value function for the $\bar{\psi}$ statistic depends (at most) on τ, Ω, β_0, and α. This feature holds for all four statistics considered in Section 2.2. This reasoning can also be applied to compute the critical value function $\bar{c}_{\psi}(\tau, \Omega, \beta_0, \alpha)$ by Monte Carlo replications. We only have to simulate the conditional null distribution of $\bar{\psi}$ from the known null distribution of $\overline{T}'\overline{S}$ and $\overline{S}'M_{\overline{T}}\overline{S}$.

Finally, if φ depends on Ω and β_0 only through $\overline{S}'\overline{S}$, $\overline{S}'\overline{T}$, and $\overline{T}'\overline{T}$, then an analogous argument shows that the critical value function for the $\bar{\psi}$ statistic depends (at most) on τ and α (for fixed k and l). This property holds for the likelihood ratio statistic.

APPENDIX C: Proofs

Proof of Theorem 1: In fact, we need only assume that $\psi(S, T, \Omega, \beta_0)$ is a continuous random variable for all t except for a set having T-probability zero. For any t where $\psi(S, T, \Omega, \beta_0)$ 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, $\Pr[\psi(S, T, \Omega, \beta_0) > c_\alpha(T) | T = t] = \alpha$. Since this holds for all t, it follows that $\Pr[\psi(S, T, \Omega, \beta_0) > c_\alpha(T)] = \alpha$ unconditionally. *Q.E.D.*

Proof of Proposition 1: Note that $(\overline{S}, \overline{T}) = (Z'Z)^{-1/2}Z'Y\Omega^{-1/2}J$ where

$$J = \left[\Omega^{1/2}b_0(b_0'\Omega b_0)^{-1/2}, \Omega^{-1/2}A_0(A_0'\Omega A_0)^{-1/2}\right]$$

is an orthogonal matrix. Thus the eigenvalues of $\Omega^{-1/2}Y'N_ZY\Omega^{-1/2}$ are the same as the eigenvalues of $(\overline{S}, \overline{T})'(\overline{S}, \overline{T})$. This shows that the LR_0 statistic indeed depends only on $\overline{S}'\overline{S}$, $\overline{S}'\overline{T}$, and $\overline{T}'\overline{T}$. When $l = 1$, the smallest eigenvalue is then given by

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

Therefore, the LR_0 test statistic is given by expression (3). For $\overline{T}'\overline{T} \neq 0$, LR_0 can be rewritten as

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

where $Q_1 = \overline{S}'\overline{T}(\overline{T}'\overline{T})^{-1}\overline{T}'\overline{S}$ and $Q_{k-1} = \overline{S}'[I - \overline{T}(\overline{T}'\overline{T})^{-1}\overline{T}']\overline{S}$. Conditional on $\overline{T} = \bar{t}$, 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, for fixed k and l, the critical value function for the LR_0 statistic depends only on τ and α. This last argument suggests an easier way to do Monte Carlo simulations to compute the critical value function for the LR_0 statistic than the general method proposed in Appendix B. Here, we have to do replications from variables (Q_1 and Q_{k-1}) whose null distributions do not depend on τ at all.

When $\tau = 0$, $LR_0 = \overline{S}'\overline{S}$, which is a chi-square-$k$ random variable. When $\tau \to \infty$, $LR_0 \to (\overline{S}'\bar{t})^2/\bar{t}'\bar{t}$, which is a chi-square-one random variable. Finally, we find that the critical value function $\bar{c}_{LR}(\tau, \alpha; k)$ is a decreasing function of τ. Our claim is that, for each ω, the derivative of $LR_0(Q_1(\omega), Q_{k-1}(\omega), \tau)$ with respect to τ is negative. We will prove this claim by contradiction. Suppose that the derivative is positive:

$$\text{(A.3)} \qquad \frac{\partial LR_0}{\partial \tau} = -1 + \frac{(Q_1 + Q_{k-1} + \overline{T}'\overline{T}) - 2\cdot Q_{k-1}}{\left[(Q_1 + Q_{k-1} + \tau)^2 - 4Q_{k-1}\cdot\tau\right]^{1/2}} > 0.$$

But (A.3) holds if, and only if,

$$(Q_1+Q_{k-1}+\overline{T}'\overline{T})-2\cdot Q_{k-1}>\sqrt{(Q_1+Q_{k-1}+\overline{T}'\overline{T})^2-4Q_{k-1}\cdot\overline{T}'\overline{T}}.$$

Taking square of both sides (and noting that the right-hand-side is larger than zero), we have

$$\left[(Q_1+Q_{k-1}+\overline{T}'\overline{T})-2\cdot Q_{k-1}\right]^2>(Q_1+Q_{k-1}+\overline{T}'\overline{T})^2-4Q_{k-1}\cdot\overline{T}'\overline{T}.$$

Simplifying this expression, we have

$$-4\cdot Q_1\cdot Q_{k-1}>0,$$

which is a contradiction. Thus, the null rejection probability is a decreasing function of τ for a fixed critical value c. Since the critical value function $\bar{c}_{LR}(\tau,\alpha;k)$ is such that the null rejection probability equals α for each $T=t$, it must be a decreasing function of τ. *Q.E.D.*

PROOF OF THEOREM 2: By definition,

$$\begin{aligned}\frac{1}{\sqrt{n}}[S,T]&=\frac{1}{\sqrt{n}}Z'[Z\Pi A'b_0+Vb_0,Z\Pi A'\Omega^{-1}A_0+V\Omega^{-1}A_0]\\&=\frac{1}{\sqrt{n}}Z'Z\Pi[A'b_0,A'\Omega^{-1}A_0]+\frac{1}{\sqrt{n}}Z'V[b_0,\Omega^{-1}A_0].\end{aligned}$$

Under Assumption (i),

$$\frac{1}{\sqrt{n}}Z'Z\Pi[A'b_0A'\Omega^{-1}A_0]\xrightarrow{p}QC[\beta-\beta_0,A'\Omega^{-1}A_0],$$

using the fact that $A'b_0=\beta-\beta_0$. Let $u\equiv Vb_0$ and $\varepsilon\equiv V\Omega^{-1}A_0$. Then, we have

$$\frac{1}{\sqrt{n}}Z'V[b_0,\Omega^{-1}A_0]\xrightarrow{d}[\Psi_{zu},\Psi_{z\varepsilon}],$$

where $\Psi_{zu}\equiv\Psi_{zv}b_0$ and $\Psi_{z\varepsilon}\equiv\Psi_{zv}\Omega^{-1}A_0$. In particular, Ψ_{zu} is independent of $\Psi_{z\varepsilon}$ since u is uncorrelated with ε. The statistic T is a function of the unknown variance of the disturbances. However,

$$\frac{1}{\sqrt{n}}(\widehat{T}-T)=\frac{1}{\sqrt{n}}Z'Y\{\widehat{\Omega}^{-1}-\Omega^{-1}\}A_0\xrightarrow{p}0,$$

since $Z'Y/\sqrt{n}$ converges in distribution and $\widehat{\Omega}^{-1}-\Omega^{-1}\xrightarrow{p}0$. Therefore, $\bar{\psi}$ has the same limiting distribution as

(B.1) $$\bar{\psi}(\Psi_{zu}+QC(\beta-\beta_0),\Psi_{z\varepsilon}+QCA'\Omega^{-1}A_0,Q,\Omega,\beta_0),$$

using Assumption (iii). Analogously, using Assumptions (iii) and (iv) the critical value function $c_{\bar{\psi}}$ converges in distribution to

(B.2) $$c_{\bar{\psi}}(\Psi_{z\varepsilon}+QCA'\Omega^{-1}A_0,Q,\Omega,\beta_0).$$

Consequently, $\bar{\psi}(S,T,\Omega,\beta_0)-c_{\bar{\psi}}(T,\Omega,\beta_0,\alpha)$ converges in distribution to the difference in expressions (B.1) and (B.2). *Q.E.D.*

REFERENCES

ANDERSON, T. W., N. KUNITOMO, AND T. SAWA (1982): "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 50, 1009–1028.

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. (1986): "Complete Consistency: A Testing Analogue of Estimator Consistency," *The Review of Economic Studies*, 53, 263–269.

ANGRIST, J., AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

BOUND, J., D. A. JAEGER, AND R. M. 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, 443–450.

BREUSCH, T. S., AND A. R. PAGAN (1980): "The Lagrange Multiplier Test and its Applications to Model Specifications in Econometrics," *The Review of Economic Studies*, 47, 239–253.

CRUZ, L. M., AND M. J. MOREIRA (2002): "Recipes for Applied Researchers: Inference When Instruments May Be Weak," Unpublished Manuscript, UC Berkeley.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

ENGLE, R. F. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam: Elsevier Science, Ch. 13, pp. 775–826.

GLESER, L. J., AND J. T. HWANG (1987): "The Non-Existence of $100(1-\alpha)\%$ Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models," *Annals of Statistics*, 15, 1351–1362.

HOROWITZ, J. L., AND N. E. SAVIN (2000): "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, 95, 375–389.

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

LEHMANN, E. L. (1986): *Testing Statistical Hypothesis*, Wiley Series in Probability and Mathematical Analysis, 2nd edn. New York: John Wiley & Sons.

MOREIRA, M. J. (2001): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," Center for Labor Economics Working Paper Series, 37, UC Berkeley.

——— (2002): "Tests with Correct Size in the Simultaneous Equations Model," Ph.D. Thesis, UC Berkeley.

MOREIRA, M. J., AND B. P. POI (2003): "Implementing Tests with Correct Size in the Simultaneous Equations Model," *Stata Journal*, 3, 57–70.

NELSON, C. R., AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

PARZEN, E. (1954): "On Uniform Convergence of Families of Sequences of Random Variables," *University of California Publications in Statistics*, 2, 23–54.

ROTHENBERG, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam: Elsevier Science, Ch. 15, pp. 881–935.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

VAN GARDEREN, K. J. (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, Ch. 8, pp. 230–280.

WANG, J., AND E. ZIVOT (1998): "Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.