



Performance of conditional Wald tests in IV regression with weak instruments

Donald W.K. Andrews^a, Marcelo J. Moreira^b, James H. Stock^{b,c,*}

^a*Cowles Foundation for Research in Economics, Yale University, USA*

^b*Department of Economics, Harvard University, USA*

^c*The National Bureau of Economic Research, USA*

Available online 10 August 2006

Abstract

We compare the powers of five tests of the coefficient on a single endogenous regressor in instrumental variables regression. Following Moreira [2003, A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048], all tests are implemented using critical values that depend on a statistic which is sufficient under the null hypothesis for the (unknown) concentration parameter, so these conditional tests are asymptotically valid under weak instrument asymptotics. Four of the tests are based on k -class Wald statistics (two-stage least squares, LIML, Fuller's [Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–953], and bias-adjusted TSLS); the fifth is Moreira's (2003) conditional likelihood ratio (CLR) test. The heretofore unstudied conditional Wald (CW) tests are found to perform poorly, compared to the CLR test: in many cases, the CW tests have almost no power against a wide range of alternatives. Our analysis is facilitated by a new algorithm, presented here, for the computation of the asymptotic conditional p -value of the CLR test.

© 2006 Elsevier B.V. All rights reserved.

JEL classification: C12; C30

Keywords: Instrumental variables regression; Power envelope; Weak identification; k -class estimators; Conditional likelihood ratio test

*Corresponding author. Department of Economics, Harvard University, Cambridge MA, USA.

Tel.: +1 617 496 0502; fax: +1 617 495 7730.

E-mail address: james_stock@harvard.edu (J.H. Stock).

1. Introduction

There has been considerable recent attention to the problem of hypothesis testing in instrumental variables (IV) regression when the instruments might be weak, that is, when the partial correlation between the instruments and the included endogenous variables is low. When instruments are weak, conventional test statistics, such as the usual t -ratio constructed using the two-stage least squares (TSLS) estimator, have null distributions that are poorly approximated by a standard normal (cf. Nelson and Startz, 1990a). The result is that conventional IV tests can result in large size distortions if the instruments are weak, even in large samples. For reviews of hypothesis testing in the presence of weak instruments, see Stock et al. (2002) and Dufour (2003).

Recently, Moreira (2001, 2003) introduced the idea of implementing tests in IV regression not using a single fixed critical value, but instead using a critical value that is itself a function of a statistic chosen so that the resulting test has the correct size even if the instruments are weak. More specifically, the distribution of a large class of IV test statistics depends on a parameter λ , which is the rescaled concentration parameter (a precise definition of λ is given below). By computing critical values conditional on a statistic that is sufficient for λ under the null hypothesis, the resulting test has the desired rejection rate for all values of λ , including the unidentified case $\lambda = 0$, and thus has the correct size. Moreira (2003) suggested two specific test statistics, the limited information likelihood ratio (LR) statistic and the TSLS Wald statistic, that can be used to construct conditional tests. The power of the resulting conditional likelihood ratio (CLR) test has been examined in detail in Andrews et al. (2006) (hereafter, AMS), however conditional Wald (CW) tests have not been examined numerically. One practical difficulty with studying and using these statistics is that their conditional distributions are not simple, so they require the numerical evaluation of conditional critical values (or conditional p -values).

This paper has two objectives. The first is to compare the power of the (two-sided) CLR test with four two-sided CW tests in the case that there is a single included endogenous regressor. The first CW test is the usual TSLS Wald test statistic, evaluated using conditional critical values. Other k -class estimators, however, have distributions that are more tightly centered around the true coefficient than TSLS (e.g. Rothenberg, 1984). This suggests that these estimators might produce CW tests that have better power than TSLS when instruments are weak. In addition to the CW TSLS test, we therefore consider CW tests based on three other k -class estimators: the limited information maximum likelihood (LIML), a k -class estimator proposed by Fuller (1977), and the so-called bias-adjusted TSLS estimator. While these three other k -class Wald statistics are not fully robust to weak instruments, the size distortions arising from their use with conventional unconditional normal critical values are much less than for the TSLS Wald statistic (Stock and Yogo, 2005); perhaps after conditioning, their performance could be substantially better than the CW TSLS test.

All five of these conditional tests—the CLR test and the four CW tests—are asymptotically equivalent under conventional strong instrument asymptotics. AMS show that the CLR test is numerically approximately uniformly most powerful among the class of locally unbiased invariant tests.¹ Because the CW tests are invariant but not locally

¹Locally unbiased refers to the power function being flat at $\beta = \beta_0$. The invariance referred to here is with respect to full-rank orthogonal transformations of the instruments.

unbiased, in theory it is possible for the CW tests to have higher power than the CLR test against some local alternatives when instruments are weak. Absent theoretical results comparing the CLR and CW tests with weak instruments, the power comparisons must be numerical.

The second objective of the paper is to address a practical problem in the implementation of conditional tests. Previous applications of conditional tests in IV regression compute the conditional critical value by simulation, given a particular observed value of the sufficient statistic for λ . But this approach is cumbersome and, depending on the number of Monte Carlo draws, either slow or inaccurate. Alternatively, if a lookup table of critical values is used, the grid must be fine enough to avoid interpolation error, and no such table has been published. We solve this practical problem by providing a new algorithm for the computation of conditional p -values for a class of conditional tests. This algorithm involves a single-dimensional numerical integration and is fast and accurate. This p -value algorithm eliminates the need for an accurate lookup table of critical values for conditional hypothesis testing.²

So that we may compare the performance of these five tests when instruments are weak, yet provide results that do not hinge on specific distributional assumptions or the sample size, we compare asymptotic power derived using the weak instrument asymptotics of Staiger and Stock (1997). By deriving the limiting distribution of IV statistics along a sequence that holds λ constant, weak instrument asymptotics provide a good approximation to the distributions of the statistics even if the instruments are weak or, in the limit, irrelevant.

We performed a comprehensive numerical analysis of the power of the CLR and CW tests, for strengths of instruments ranging from very weak to strong, for the number of instruments ranging from 2 to 20, and for different values of the correlation between the reduced-form errors. Our main finding is that the CW tests have some very undesirable power properties: they have very low power against a large range of alternative parameter values and their power curves can be non-monotonic. The CW tests occasionally have higher power than the CLR test, but when they do, the power gain is small. We therefore recommend against the use of any of these four k -class CW tests in applied econometric work, even as robustness checks. Because the performance of these tests is poor, we also recommend against constructing confidence intervals by inverting these CW tests. AMS summarize the results of a thorough comparison of the CLR tests and two other tests that are valid under weak instruments, the Anderson-Rubin (1949) test and Kleibergen's (2002) LM test; the CLR test is found to have power that is typically better, and never worse, than these two competitors. Taken together, the results here and in AMS indicate that the CLR test has very good overall power properties, and we recommend its use in applied work.³

The paper is organized as follows. Section 2 presents the model and the test statistics. The weak-instrument asymptotic distributions of these test statistics have been derived elsewhere but are briefly stated in Section 3 for completeness. Section 4 presents the new

²Software implementing this algorithm for computing the CLR p -value is available at <http://ksghome.harvard.edu/~JStock/>

³The tests in this paper apply for homoskedastic errors. Heteroskedasticity- and autocorrelation-robust versions are available for all these tests; see AMS and Andrews and Stock (2005) for formulas and additional references.

method for numerical evaluation of the conditional p -value. The power comparisons are summarized in Section 5.

2. Conditional tests and their weak-instrument asymptotic limits

This section begins by introducing the model and notation. We then summarize the conditional testing approach in the IV regression model and present the specific conditional test statistics of interest in this paper.

2.1. The model and notation

The structural equation and the reduced-form equation for the single included endogenous variable are

$$y_1 = y_2\beta + X\gamma_1 + u, \quad (1)$$

$$y_2 = Z\pi + X\xi + v_2, \quad (2)$$

where y_1 and y_2 are $n \times 1$ vectors of endogenous variables, X is an $n \times p$ matrix of exogenous regressors, and Z is an $n \times k$ matrix of k IV. It is assumed that Z is constructed so that $Z'X = 0$. This orthogonality assumption entails no loss of generality: if \tilde{Z} denotes an original $n \times k$ matrix of observed IV with $\tilde{Z}'X \neq 0$, then set $Z = M_X\tilde{Z}$, where $M_X = I - X(X'X)^{-1}X'$.

Our interest is in two-sided tests of the null hypothesis

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0. \quad (3)$$

The reduced form of (1) and (2) is

$$Y = Z\pi a' + X\eta + V, \quad (4)$$

where $Y = [y_1 \ y_2]$, $V = [v_1 \ v_2]$, $a = [\beta \ 1]'$, and $\eta = [\gamma \ \xi]$, where $v_1 = u + v_2\beta$ and $\gamma = \gamma_1 + \xi\beta$. The reduced-form errors are assumed to be homoskedastic with covariance matrix Ω :

$$E(V_i V_i' | X_i, Z_i) = \Omega = \begin{bmatrix} \omega_{11} & \omega_{21} \\ \omega_{12} & \omega_{22} \end{bmatrix}, \quad i = 1, \dots, n, \quad (5)$$

where the subscript i refers to the i th observation. Let ρ denote the correlation between V_{1i} and V_{2i} . Throughout, it is assumed that $E(u_i | X_i, Z_i) = 0$. Additional distributional assumptions are discussed below.

A key measure of the strength of the instruments is λ , defined as

$$\lambda = \pi' Z' Z \pi. \quad (6)$$

Upon dividing λ by ω_{22} to be unitless, λ/ω_{22} is the so-called concentration parameter that governs the quality of the standard large-sample normal approximations to the distribution of IV estimators (e.g. [Rothenberg, 1984](#)). The expected value of the first-stage F statistic testing the hypothesis that $\pi = 0$ in (2) is asymptotically $1 + \lambda/(\omega_{22}k)$ under weak instrument asymptotics. Small values of $\lambda/(\omega_{22}k)$ correspond to weak instruments.

2.2. The conditional testing approach

The distribution of IV statistics commonly used to test H_0 , such as the TSLS Wald statistic, depends on the nuisance parameter λ . The idea of [Moreira's \(2001, 2003\)](#) conditional testing approach is to evaluate such test statistics conditional on a statistic that is sufficient for λ under the null hypothesis. Because the conditional distribution does not depend on λ , it is possible to control the size of the test regardless of the true value of λ .

2.2.1. The Gaussian model with Ω known

To make this argument precise, suppose for the moment that V_i is i.i.d. $N(0, \Omega)$, that Z is nonrandom, and that Ω is known. Define the statistic Q to be

$$Q \equiv \begin{bmatrix} Q_S & Q_{TS} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & T'S \\ S'T & T'T \end{bmatrix}, \quad (7)$$

where

$$S = (Z'Z)^{-1/2} Z' Y b_0 / (b_0' \Omega b_0)^{1/2}, \quad (8)$$

$$T = (Z'Z)^{-1/2} Z' Y \Omega^{-1} a_0 / (a_0' \Omega^{-1} a_0)^{1/2}, \quad (9)$$

where $b_0 = [1 - \beta_0]'$ and $a_0 = [\beta_0 \quad 1]'$. Let $P_Z = Z(Z'Z)^{-1} Z'$. Note that Q and $Y' P_Z Y$ are related by⁴

$$Q = J_0' \Omega^{-1/2} Y' P_Z Y \Omega^{-1/2} J_0, \quad (10)$$

$$Y' P_Z Y = \Omega^{1/2} J_0^{-1} Q J_0^{-1} \Omega^{1/2}, \quad (11)$$

where

$$J_0 = \begin{bmatrix} \frac{\Omega^{1/2} b_0}{\sqrt{b_0' \Omega b_0}} & \frac{\Omega^{-1/2} a_0}{\sqrt{a_0' \Omega^{-1} a_0}} \end{bmatrix}. \quad (12)$$

Note that $J_0' J_0 = I$.

If the instruments are fixed and the errors are Gaussian, the statistics S and T are jointly normally distributed and the distribution of S does not depend on λ under the null hypothesis. [AMS](#) show that, under these assumptions, Q is the maximal invariant for (β, λ) under the group of full-rank orthogonal transformations of Z . [Moreira \(2003\)](#) shows that Q_T is sufficient for λ under the null hypothesis. Thus, any statistic that is a function of Q can be used to construct a similar test by comparing its value to the $1 - \alpha$ critical value of the conditional null distribution of the statistic, given $Q_T = q_T$; this procedure yields a test with size $1 - \alpha$.

2.2.2. Extension to Ω unknown

In practice, Ω is unknown so tests based on Q are infeasible. However, a feasible counterpart of Q can be constructed by replacing Ω by an estimator. Specifically, let

$$\hat{\Omega} = Y'^{\perp} M_Z Y^{\perp} / (n - k - p), \quad (13)$$

⁴Throughout, we adopt the convention that $A = A^{1/2} A^{1/2}$ and $A^{-1} = A'^{-1/2} A^{-1/2}$ for a positive definite matrix A .

where $M_Z = I - P_Z$ and $Y^\perp = [y_1^\perp \ y_2^\perp] = M_Z Y$. Then Ω in the definition of Q can be replaced by $\hat{\Omega}$, yielding \hat{Q} , the feasible counterpart of Q :

$$\hat{Q} = \hat{J}'_0 \hat{\Omega}^{-1/2} Y' P_Z Y \hat{\Omega}^{-1/2} \hat{J}_0, \tag{14}$$

where

$$\hat{J}_0 = \begin{bmatrix} \frac{\hat{\Omega}'^{1/2} b_0}{\sqrt{b_0' \hat{\Omega} b_0}} & \frac{\hat{\Omega}^{-1/2} a_0}{\sqrt{a_0' \hat{\Omega}^{-1} a_0}} \end{bmatrix}. \tag{15}$$

2.3. Specific conditional tests

The family of test statistics that can be written as functions of $(\hat{Q}, \hat{\Omega})$ is large and includes the tests commonly used in applied IV regression. All such tests can be made robust to weak instruments by evaluating them using conditional critical values, given \hat{Q}_T . Here, we provide explicit expressions for the tests examined in this study.

2.3.1. *k*-class Wald statistics

To avoid notational confusion with the number of IVs, k , we denote the k -class parameter by \underline{k} . The k -class estimator of β is

$$\hat{\beta}(\underline{k}) = [y_2'^\perp (I - \underline{k} M_Z) y_2^\perp]^{-1} [y_2'^\perp (I - \underline{k} M_Z) y_1^\perp]. \tag{16}$$

The Wald statistic based on (16) is

$$\hat{W}(\underline{k}) = \frac{[\hat{\beta}(\underline{k}) - \beta_0]^2}{\hat{\sigma}_u^2(\underline{k})/[y_2'^\perp (I - \underline{k} M_Z) y_2^\perp]}, \tag{17}$$

where $\hat{\sigma}_u^2(\underline{k}) = \hat{u}(\underline{k})' \hat{u}(\underline{k}) / (n - k - p)$, where $\hat{u}(\underline{k}) = y_1^\perp - y_2^\perp \hat{\beta}(\underline{k})$.

We consider four specific k -class estimators: TSLS, the limited information maximum likelihood estimator (LIML), a modified LIML estimator proposed by Fuller (1977), and bias-adjusted TSLS (BTSLS; Nagar, 1959, Rothenberg, 1984). The values of \underline{k} for these estimators are (see Donald and Newey (2001)):

$$\text{TSLS : } \underline{k} = 1, \tag{18}$$

$$\text{LIML : } \underline{k} = \hat{k}_{\text{LIML}} = \text{the smallest root of } \det(Y'^\perp Y^\perp - k Y'^\perp M_Z Y^\perp) = 0, \tag{19}$$

$$\text{Fuller : } \underline{k} = \hat{k}_{\text{LIML}} - c / (n - k - p) \text{ where } c \text{ is a positive constant,} \tag{20}$$

$$\text{BTSLS : } \underline{k} = n / (n - k + 2). \tag{21}$$

In the numerical work, we examine the Fuller estimator with $c = 1$, which is the best unbiased estimator to second-order among estimators with $k = 1 + a(\hat{k}_{\text{LIML}} - 1) - c / (n - k - p)$ for some constants a and c (Rothenberg, 1984). For further discussion, see Donald and Newey (2001), Stock et al. (2002, Section 6.1), and Hahn et al. (2004).

These statistics are functions of the data only through \hat{Q} and $\hat{\Omega}$. Specifically,

$$\hat{\beta}(\underline{k}) = [y_2' P_Z y_2 - \kappa \hat{\omega}_{22}]^{-1} [y_2' P_Z y_1 - \kappa \hat{\omega}_{12}], \tag{22}$$

$$\widehat{W}(k) = \frac{[\widehat{\beta}(k) - \beta_0]^2 (y_2' P_Z y_2 - \kappa \widehat{\omega}_{22})}{\widehat{b}(k)' [\widehat{\Omega} + Y' P_Z Y / (n - k - p)] \widehat{b}(k)}, \quad (23)$$

where $\widehat{b}(k) = [1 - \widehat{\beta}(k)]'$, $\kappa = (n - k - p)(k - 1)$ and $\widehat{k}_{\text{LIML}} = 1 + \kappa_{\text{LIML}} / (n - k - p)$, where κ_{LIML} is the smallest root of $\det[Y' P_Z Y - \kappa \widehat{\Omega}] = 0$. It follows from (15), (23), and $Y' P_Z Y = \widehat{\Omega}^{1/2} \widehat{J}_0^{-1} \widehat{Q} \widehat{J}_0^{-1} \widehat{\Omega}^{1/2}$ that $\widehat{W}(k)$ is a function of the data only through \widehat{Q} and $\widehat{\Omega}$.

2.3.2. The LR statistic

The other test considered in this study is [Moreira's \(2003\)](#) CLR test, which is based on the statistic

$$\widehat{\text{LR}} = \frac{1}{2} \{ \widehat{Q}_S - \widehat{Q}_T + [(\widehat{Q}_S + \widehat{Q}_T)^2 - 4(\widehat{Q}_S \widehat{Q}_T - \widehat{Q}_{\text{ST}})]^{1/2} \}. \quad (24)$$

Evidently, $\widehat{\text{LR}}$ depends on the data only through \widehat{Q} .

3. Weak instrument asymptotic distributions

In most applications, the instruments are not fixed and there is no reason to think that the errors are normally distributed. However, the conditional testing strategy is nonetheless valid in large samples, even if instruments are weak. The asymptotic justification relies on weak instrument asymptotics. Specific assumptions under which weak instrument asymptotics apply are available in the literature (see [Staiger and Stock, 1997](#), and [AMS](#)), so we do not list them here, and instead only provide the results of the calculations.

The weak-instrument asymptotic limits of \widehat{Q} and $\widehat{\Omega}$ are

$$\widehat{Q} \xrightarrow{d} Q_\infty \quad \text{and} \quad \widehat{\Omega} \xrightarrow{p} \Omega, \quad (25)$$

where Q_∞ has a noncentral Wishart distribution with noncentrality matrix $\lambda h_\beta h_\beta'$, where $h_\beta = [c_\beta \ d_\beta]'$, $c_\beta = (\beta - \beta_0) / (b_0' \Omega b_0)^{1/2}$, $d_\beta = a' \Omega^{-1} / (a_0' \Omega^{-1} a_0)^{1/2}$, and $a = [\beta \ 1]'$ (see [AMS](#)).

It follows from (25) and the continuous mapping theorem that the statistics in Section 2.3 have weak-instrument asymptotic distributions that can be characterized in terms of the distribution of Q_∞ . Let $\Psi = \psi(\widehat{Q}, \widehat{\Omega})$ denote a test statistic that depends on the data only through \widehat{Q} and $\widehat{\Omega}$. Then

$$\widehat{\Psi} \xrightarrow{d} \Psi_\infty = \psi(Q_\infty, \Omega). \quad (26)$$

The limiting representation of the k -class Wald statistic (see [Stock and Yogo, 2005](#), for a derivation using different notation) is

$$\widehat{W}(k) \xrightarrow{d} W_\infty(\kappa_\infty) = \frac{[\Xi_{\infty,12} - \kappa_\infty \omega_{12}]^2}{[\Xi_{\infty,22} - \kappa_\infty \omega_{22}] [\widehat{b}_\infty(\kappa_\infty)' \Omega \widehat{b}_\infty(\kappa_\infty)]}, \quad (27)$$

where $\Xi_\infty = \Omega^{1/2} J_0^{-1} Q_\infty J_0^{-1} \Omega^{1/2}$, $\widehat{b}_\infty(\kappa) = [1 - \widehat{\beta}_\infty(\kappa)]'$, $\widehat{\beta}_\infty(\kappa) = [\Xi_{\infty,22} - \kappa \omega_{22}]^{-1} [\Xi_{\infty,21} - \kappa \omega_{21}]$, and κ_∞ depends on which k -class estimator is used:

$$\text{TSLS : } \kappa_\infty = 1, \quad (28)$$

$$\text{LIML} : \kappa_\infty = \kappa_{\text{LIML},\infty} \quad \text{where } \kappa_{\text{LIML},\infty} = \text{the smallest root of } \det(\Xi_\infty - \kappa\Omega) = 0, \quad (29)$$

$$\text{Fuller} : \kappa_\infty = \kappa_{\text{LIML},\infty} - c \quad \text{where } c \text{ is the constant in (20)}, \quad (30)$$

$$\text{BTSLs} : \kappa_\infty = k - 2 \quad (\text{where } k \text{ is the number of instruments}). \quad (31)$$

The weak-instrument asymptotic representation of the $\widehat{\text{LR}}$ statistic follows from (24) and (25):

$$\widehat{\text{LR}} \xrightarrow{d} \text{LR}_\infty = \frac{1}{2} \{ Q_{\infty,S} - Q_{\infty,T} + [(Q_{\infty,S} + Q_{\infty,T})^2 - 4(Q_{\infty,S}Q_{\infty,T} - Q_{\infty,ST}^2)]^{1/2} \}. \quad (32)$$

3.1. Asymptotic power functions

The asymptotic conditional critical value of the test statistic $\hat{\Psi}$, given $\hat{Q}_T = q_T$, is $c(q_T, \alpha)$, which is the $1 - \alpha$ quantile of the conditional distribution of Ψ_∞ given $Q_{\infty,T} = q_T$. The asymptotic power of the test against the alternative β is

$$\text{asymptotic power} = \text{Pr}_{\beta,\lambda}[\Psi_\infty > c(Q_{\infty,T}, \alpha)], \quad (33)$$

where the probability is computed with respect to the distribution of Q_∞ when the true parameter values are (β, λ) . Under H_0 , the probability in (33) equals α and does not depend on λ , but under the alternative the power in general depends on λ as well as β .

4. Numerical evaluation of asymptotic p -values of conditional tests

In this section, distributional results in AMS are used to provide a simple algorithm, involving only one-dimensional integration, for the evaluation of conditional p -values of IV test statistics that are monotone increasing in Q_S (or \hat{Q}_S). Because the distribution of Q_∞ is the same as the distribution of Q under normal errors and fixed instruments, the algorithm is presented using the expositional expedient of the fixed instrument/Gaussian model. As a special case, we provide explicit expressions for the conditional p -value of the CLR test. The section concludes by providing the asymptotic interpretation of these p -values when the IVs are random and the errors non-normal.

4.1. Exact p -values in the known- Ω Gaussian model

The task is to compute the probability under the null hypothesis that $\Psi = \psi(Q, \Omega)$ exceeds a constant m , under the assumptions of the fixed IV/Gaussian model. Following AMS, let $S_2 = Q_{ST}/(Q_S Q_T)^{1/2}$. Accordingly, we can write Ψ as $\Psi = g(Q_S, S_2, Q_T)$ (the mapping from Q to (Q_S, S_2, Q_T) is one-to-one, and the dependence on Ω is subsumed into g). We consider statistics that are invertible functions of q_S . Specifically, suppose there exists the inverse function g_1^{-1} such that

$$g(g_1^{-1}(m; s_2, q_T), s_2, q_T) = m \quad \text{and} \quad g_1^{-1}(m; s_2, q_T) \text{ is monotone increasing in } m. \quad (34)$$

Under the null hypothesis, the conditional probability that Ψ exceeds a constant m , given $Q_T = q_T$, is

$$\begin{aligned} p(m; q_T) &= 1 - \Pr_0[g(Q_S, S_2, Q_T) < m | Q_T = q_T] \\ &= 1 - \int_{-1}^1 \Pr_0[g(Q_S, S_2, Q_T) < m | S_2 = s_2, Q_T = q_T] f_{S_2|Q_T}(s_2 | Q_T = q_T) ds_2, \end{aligned} \quad (35)$$

where $f_{S_2|Q_T}$ is the conditional distribution of S_2 given Q_T under H_0 , $\Pr_0[\cdot]$ denotes the probability evaluated under H_0 , and the limits of integration arise because $|S_2| \leq 1$. Note that $\Pr_0[g(Q_S, S_2, Q_T) < m | Q_T = q_T]$ does not depend on λ because Q_T is sufficient for λ under H_0 .

If the inverse function g_1^{-1} in (34) exists, then the final expression in (35) can be rewritten as an inequality expressed in terms of Q_S :

$$p(m; q_T) = 1 - \int_{-1}^1 \Pr_0[Q_S < g_1^{-1}(m; S_2, Q_T) | S_2 = s_2, Q_T = q_T] f_{S_2|Q_T}(s_2 | Q_T = q_T) ds_2. \quad (36)$$

It is shown in AMS (Lemma 3) that, under H_0 , Q_S , S_2 , and Q_T are mutually independent, Q_S has a χ^2 distribution with k degrees of freedom, and S_2 has the density

$$f_{S_2}(s_2) = K_4(1 - s_2^2)^{(k-3)/2}, \quad (37)$$

where $K_4 = \Gamma(k/2)/[pi^{1/2}\Gamma((k-1)/2)]$, where $pi = 3.1416\dots$ and $\Gamma(\cdot)$ is the gamma function. Accordingly, (36) becomes

$$p(m; q_T) = 1 - K_4 \int_{-1}^1 \Pr[\chi_k^2 < g_1^{-1}(m; s_2, q_T)] (1 - s_2^2)^{(k-3)/2} ds_2. \quad (38)$$

The conditional p -value is $p(\Psi^{\text{obs}}; q_T)$, where Ψ^{obs} is the observed value of the test statistic in the data. Because statistical software packages include functions for the cumulative χ^2 distribution, evaluation of $p(m; q_T)$ based on (38) requires only a single numerical integration (over S_2).

4.2. Specialization to the CLR test

The LR statistic, written in terms of Q , is

$$\begin{aligned} \text{LR} &= \frac{1}{2}\{Q_S - Q_T + [(Q_S + Q_T)^2 - 4(Q_S Q_T - Q_{ST}^2)]^{1/2}\} \\ &= \frac{1}{2}\{Q_S - Q_T + [(Q_S + Q_T)^2 - 4Q_S Q_T(1 - S_2^2)]^{1/2}\}, \end{aligned} \quad (39)$$

where the second line expresses the LR statistic as a function of Q_S , Q_T , and S_2 . Inspection of (39) reveals that LR is monotone increasing in Q_S . For the LR statistic, the inverse function g_1^{-1} is

$$g_1^{-1}(m; s_2, q_T) = \frac{q_T + m}{1 + q_T s_2^2 / m}. \quad (40)$$

Substitution of (40) into (38) yields the following expression for the asymptotic p -value for the CLR statistic:

$$p(m; q_T) = \Pr_0[\text{LR} > m | Q_T = q_T] = 1 - 2K_4 \int_0^1 \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T s_2^2/m} \right] (1 - s_2^2)^{(k-3)/2} ds_2, \tag{41}$$

where the limits of integration have been changed to exploit the symmetry of the LR statistic in S_2 . The conditional p -value is $p(\text{LR}^{\text{obs}}; q_T)$, where LR^{obs} is the observed value of the test statistic.

4.2.1. Numerical considerations

Our experience is that the details of how best to compute the integral in (41) depend on k . For $k = 2$, the integrand in (41) is unbounded at $s_2 = 1$ and direct numerical integration is unreliable. However, in this case the integral can be handled by the change of variables, $u = \sin^{-1}(s_2)$, yielding

$$p(m; q_T) = 1 - 2K_4 \int_0^{\pi/2} \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T \sin^2(u)/m} \right] du, \tag{42}$$

which can be integrated by standard methods, e.g. Simpson’s rule.

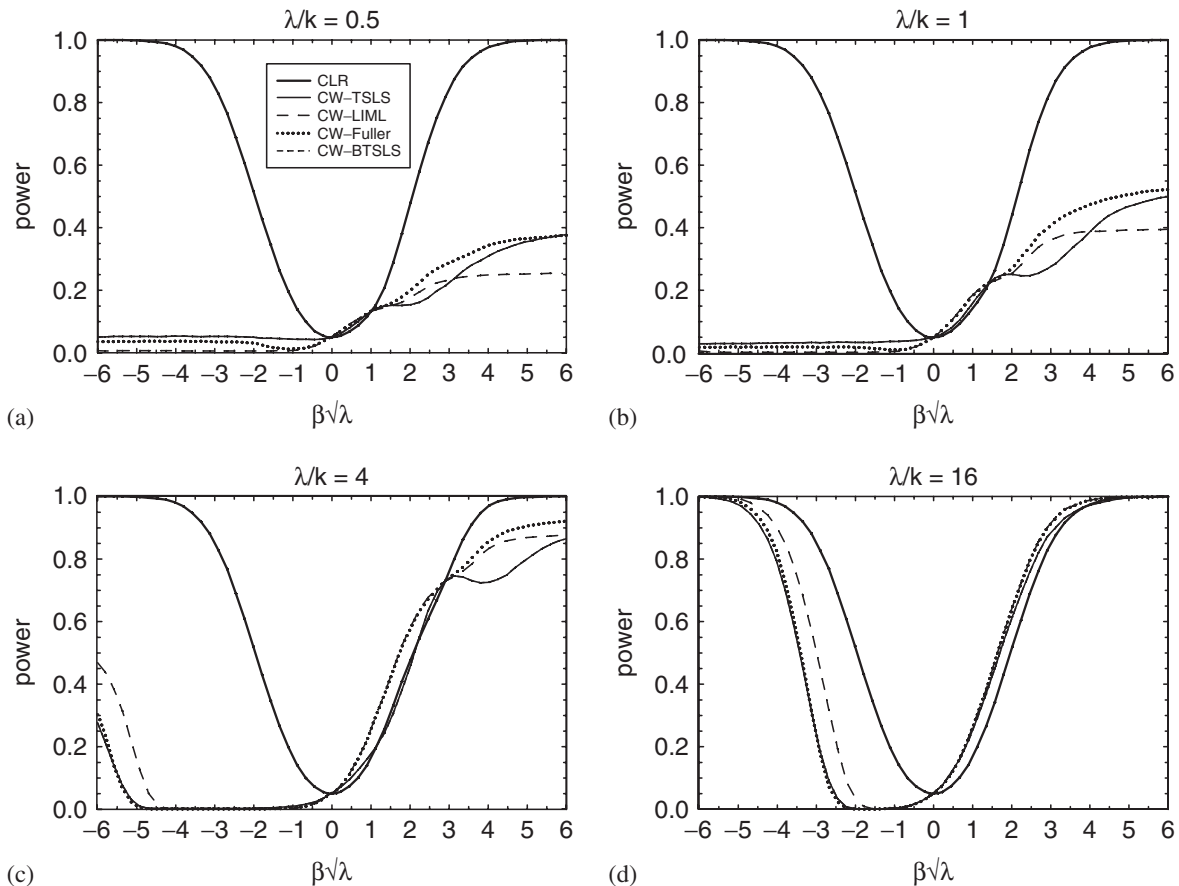


Fig. 1. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 2$, $\rho = 0.95$.

For $k = 3$, the exponent in the integrand is zero and the integral becomes

$$p(m; q_T) = 1 - 2K_4 \int_0^1 \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T s_2^2 / m} \right] ds_2, \tag{43}$$

which is readily evaluated using Simpson’s rule.

For $k = 4$, the term $(1 - s_2^2)^{(k-3)/2}$ is sufficiently nonlinear approaching $s_2 = 1$ that care needs to be taken in performing the integration near this boundary to get acceptable numerical accuracy. However, the term $\Pr[\chi_k^2 < ((q_T + m)/(1 + q_T s_2^2 / m))]$ is insensitive to s_2 in a small neighborhood of 1. Thus, for the region within ε of 1, with $k = 4$ the integral in (41) can be approximated as

$$\begin{aligned} & 2K_4 \int_{1-\varepsilon}^1 \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T s_2^2 / m} \right] (1 - s_2^2)^{(k-3)/2} ds_2 \\ & \cong 2K_4 \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T (1 - \varepsilon/2)^2 / m} \right] \int_{1-\varepsilon}^1 \sqrt{1 - s_2^2} ds_2 \\ & \cong 2K_4 \Pr \left[\chi_k^2 < \frac{q_T + m}{1 + q_T (1 - \varepsilon/2)^2 / m} \right] \end{aligned}$$

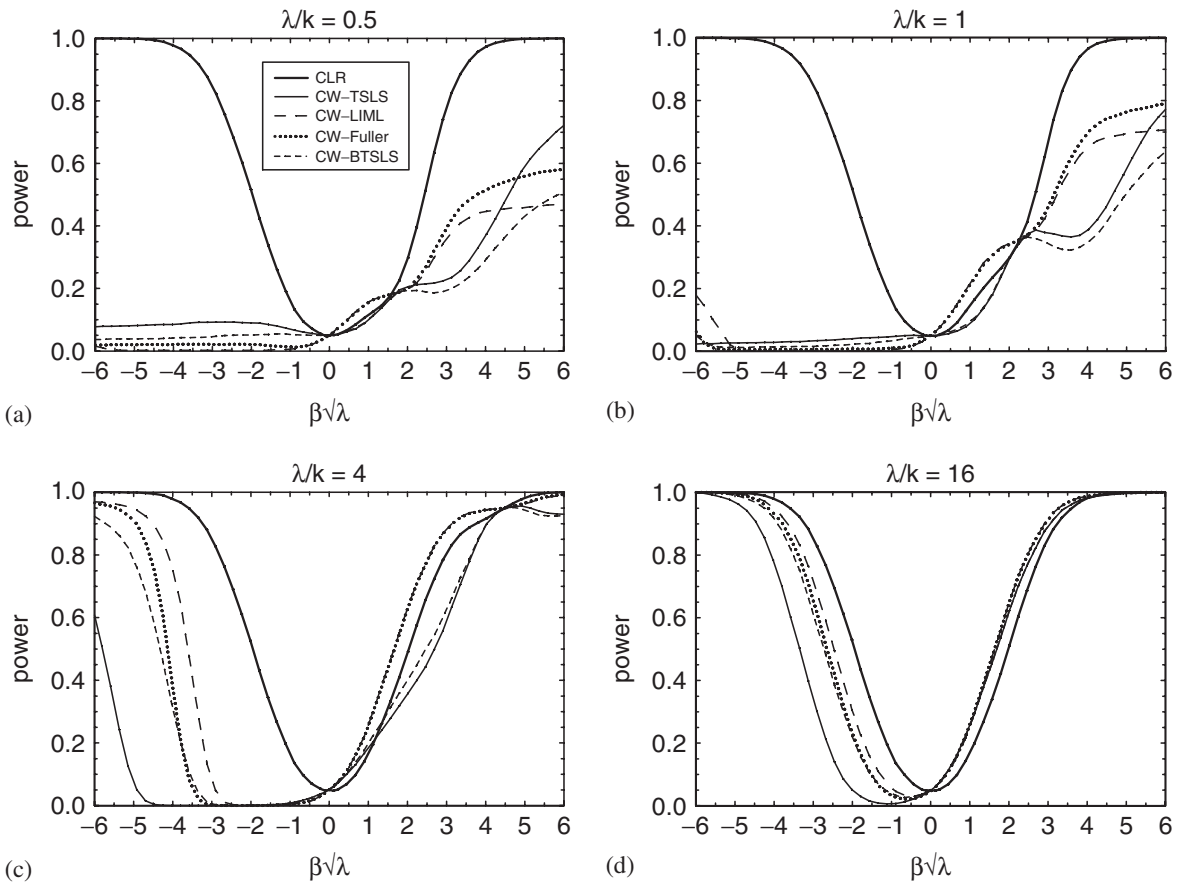


Fig. 2. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 5$, $\rho = 0.95$.

$$\times \left\{ \frac{1}{2} [\sin^{-1}(1) - \sin^{-1}(1 - \varepsilon)] - \left(\frac{1 - \varepsilon}{2} \right) \sqrt{1 - (1 - \varepsilon)^2} \right\}. \tag{44}$$

Standard numerical methods can be used for integration on the range $[0, 1 - \varepsilon]$. Numerical investigation found that setting $\varepsilon = 0.02$ yields good results.

For $k > 4$, the integral does not simplify but the integrand is bounded and well-behaved and can be evaluated accurately using standard numerical methods, e.g. Simpson’s rule.

4.3. Interpretation as asymptotic p-values

These p -values are asymptotic p -values under weak instrument asymptotics. Let $\hat{\Psi}$ denote the test statistic computed using \hat{Q} and $\hat{\Omega}$. The results of Section 3 and AMS imply that

$$\Pr_0[\hat{\Psi} > m | \hat{Q}_T = \hat{q}_T] - p(m; \hat{q}_T) \xrightarrow{p} 0. \tag{45}$$

It follows that rejection using the asymptotic α -level conditional critical values is equivalent to the asymptotic conditional p -value being less than α .

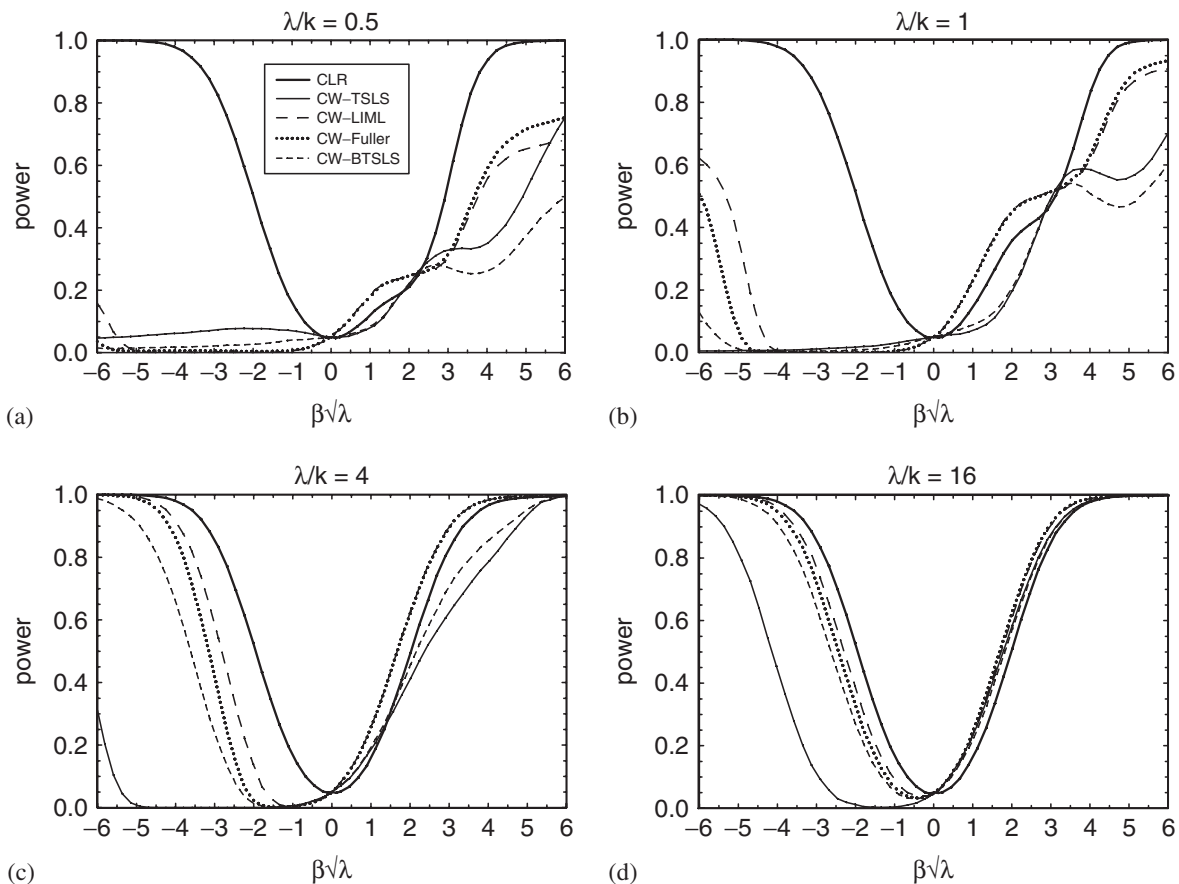


Fig. 3. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 10$, $\rho = 0.95$.

5. Asymptotic power functions of the conditional Wald and CLR tests

We now turn to a comparison of the asymptotic powers of the CW and CLR tests. To save space, a representative subset of the results are presented here; full results can be viewed on the Web (see footnote 2).

5.1. Design

By taking linear transformations and suitably redefining Y , V , β , and π , the model (4) can always be rewritten so that $\beta_0 = 0$ and $\omega_{11} = \omega_{22} = 1$. Therefore, without loss of generality we set $\beta_0 = 0$ and

$$\Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{46}$$

With this standardization, the concentration parameter is $\lambda/\omega_{22} = \lambda$, the expected value of the first-stage F -statistic testing $\pi = 0$ is $1 + \lambda/k$, and the weak-instrument asymptotic powers of the CW and CLR tests depend only on β , λ , ρ , and k .

Asymptotic power functions were computed for the CLR and TSLS, LIML, Fuller, and BTSLS CW tests using the weak-instrument asymptotic representations (27)–(32). All tests have significance level $\alpha = 0.05$. The CLR test was implemented by computing the

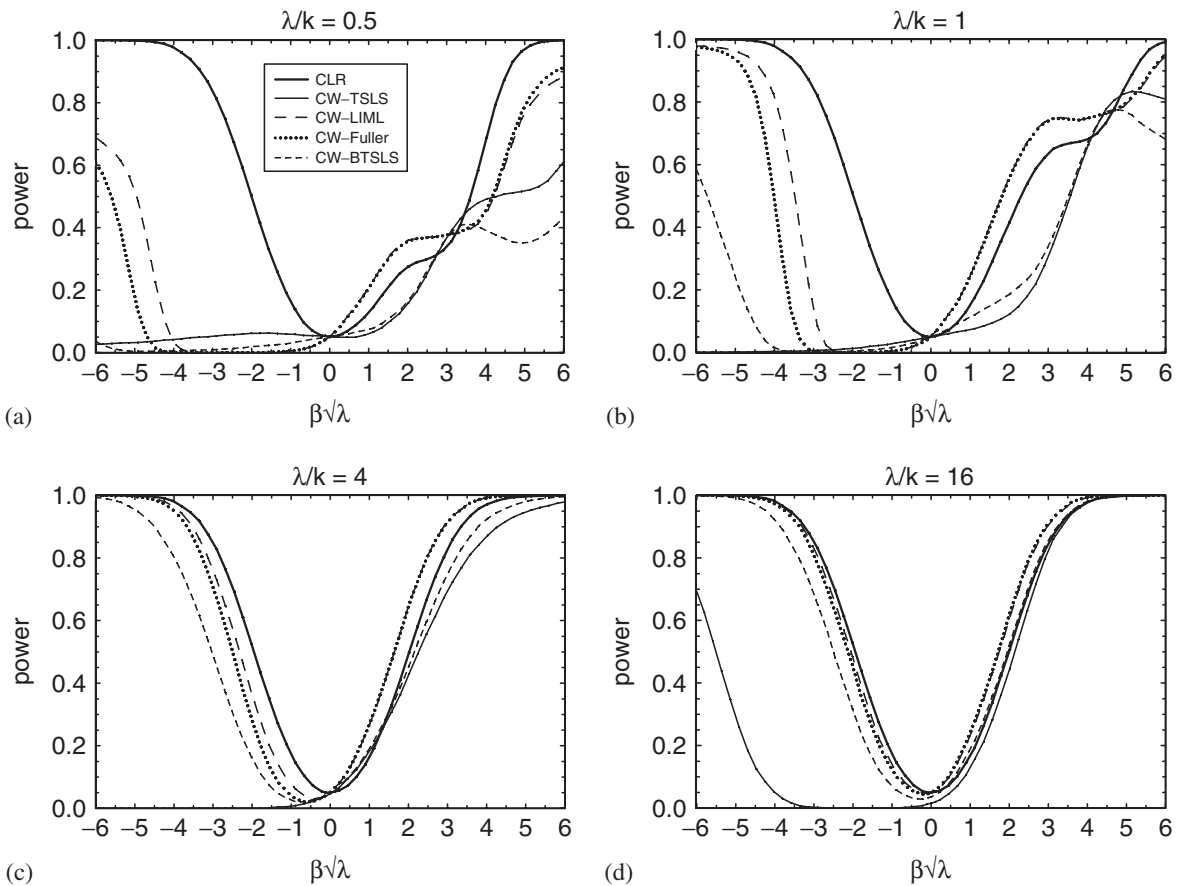


Fig. 4. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 20$, $\rho = 0.95$.

conditional p -value using the algorithm in Section 4.2 and comparing it to α . The CW tests were implemented by first computing a lookup table of critical values on a grid of 150 values of q_T , then interpolating this lookup table for a given realization of q_T . All results are based on 5000 Monte Carlo draws. To assess the numerical accuracy, rejection rates were computed under the null and are reported as the values of the power plots at $\beta = 0$; these null rejection rates are within Monte Carlo error of α .

Results were computed for $k = 2, 5, 10,$ and 20 ; $\rho = 0.95, 0.5,$ and 0.2 ; and $\lambda/k = 0.5$ (very weak instruments), $1, 2, 4, 8,$ and 16 (strong instruments).

5.2. Results

Representative results are reported in Figs. 1–6. Each figure has four panels, corresponding to $\lambda/k = 0.5, 1, 4,$ and 16 . Each curve represents the power function of the indicated test for the value of k and ρ in that figure and of λ/k in that panel. The horizontal axis is scaled to be $\beta\lambda^{1/2}$ so that the results are readily comparable across figures and panels. The scaling $\beta\lambda^{1/2}$ has an intuitive interpretation: λ is a measure of the amount of information in the instruments and at a formal level can be thought of as an effective sample size (Rothenberg, 1984). Thus, $\beta\lambda^{1/2}$ can be thought of as a local alternative, except that the neighborhood of locality is $1/\lambda^{1/2}$ rather than $1/n^{1/2}$ as is usually the case.

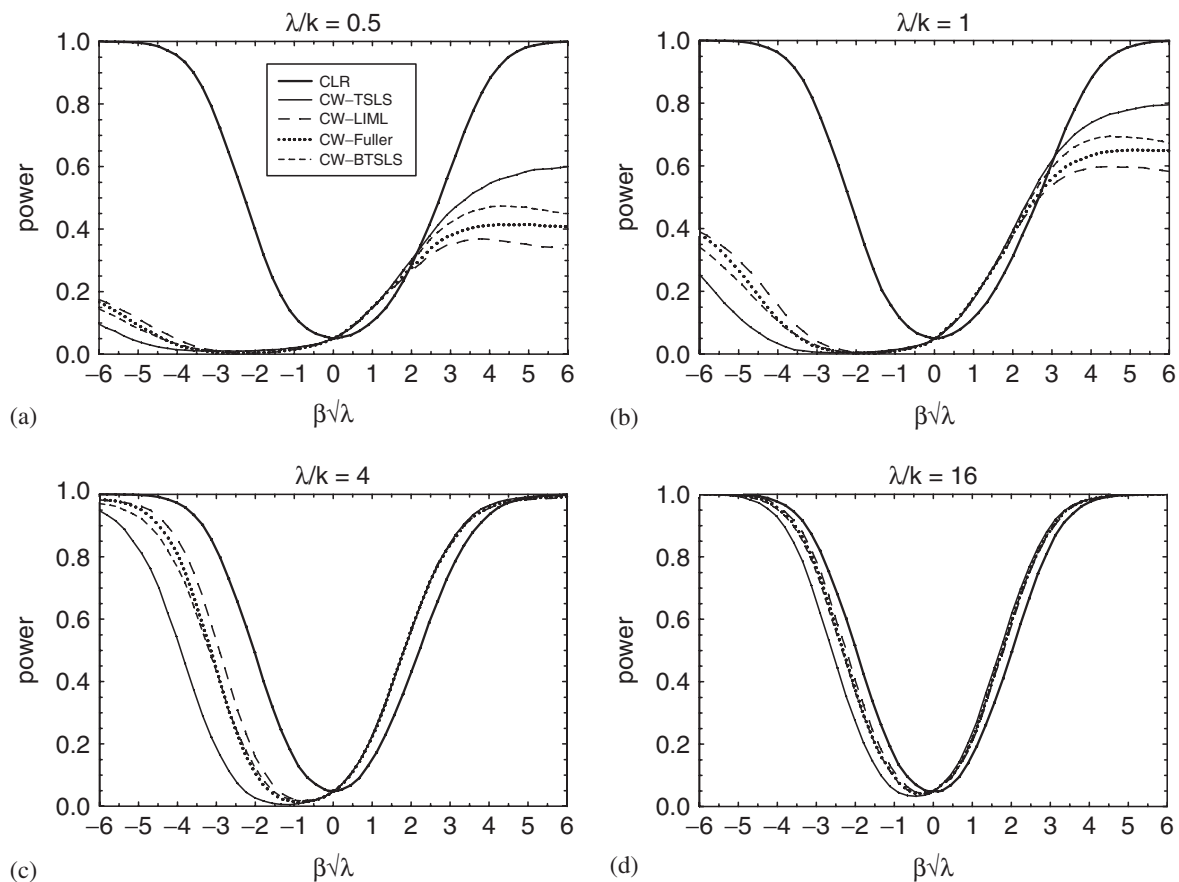


Fig. 5. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 5, \rho = 0.50$.

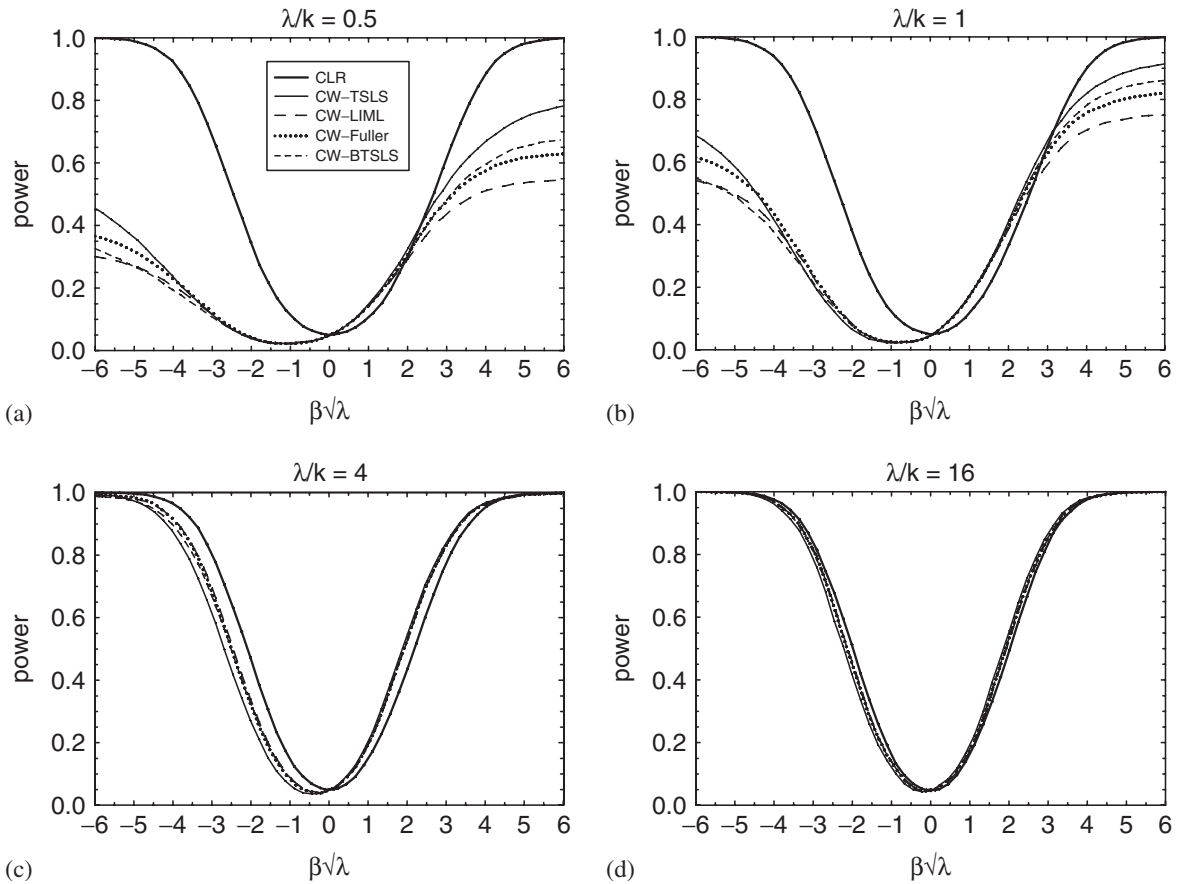


Fig. 6. Asymptotic power functions of conditional LR and conditional Wald tests: $k = 5$, $\rho = 0.20$.

Fig. 1 considers the case $k = 2$ and $\rho = .95$. Generally speaking, conventional strong-instrument asymptotic approximations tend to break down most severely at high values of ρ (cf. Nelson and Startz, 1990a, b), so the case $\rho = 0.95$ is a useful benchmark. As Fig. 1 shows, the contrast between the performance of the CLR test and the CW tests is stark. For all values of λ/k in this figure, including the case of relatively strong instruments, all four CW tests are biased, with rejection rates effectively equal to zero for some values of the alternative. When instruments are weak (panels (a) and (b)), the CW tests reject with frequency well under 5% for negative values of β . This is not because there is insufficient information to perform valid inference: the CLR test has monotonically increasing power for negative β and indeed performs comparably against positive and negative values of β .

Even against positive values of β , the CW tests do not perform as well as the CLR test when instruments are weak and/or alternatives are distant. Only in the strong-instrument case with $\beta > 0$ are the power functions of the five conditional tests comparable. While there are differences among the power functions of the CW tests, these differences are relatively small (for $k = 2$, BTSLs and TSLS are identical, see (21)).

Figs. 2–4 examine the effect of increasing the number of instruments to $k = 5$, 10, and 20, respectively, while keeping ρ fixed at 0.95. The conclusions drawn from Fig. 1 continue to hold for Figs. 2–4: the CW tests are biased, and their power functions are typically below—sometimes far below—that of the CLR test. In addition, it is evident in some of the panels that the power functions of the CW tests are not monotonic. For the larger values of

k , some differences among the CW tests emerge. The test based on TSLS generally has the least desirable power properties; this is particularly striking in Fig. 4, panel (c), in which the other three CW tests have power functions coming close to that of the CLR test but the CW-TSLS test has power of essentially zero against $\beta < 0$. Among the remaining three CW tests, the CW-Fuller test seems to work best overall. Still, the performance of the CW-Fuller test is, in an overall sense, very poor relative to the CLR test.

Figs. 5 and 6 examine the effect of changing ρ to 0.5 and 0.2, respectively, while holding the number of instruments constant at $k = 5$. Again, the same conclusions emerge. These tests are all asymptotically equivalent under strong instrument asymptotics, and this equivalence becomes evident in Fig. 6 in the case $\lambda/k = 16$: in the case that there is very little endogeneity and strong instruments, all the tests perform comparably. However, if there are weak instruments and/or significant amounts of endogeneity, then the CW tests perform worse, often much worse, than the CLR test.

These results are all asymptotic and do not reflect the sampling uncertainty arising from the estimation of Ω . In numerical work not tabulated here, we find that the weak-instrument asymptotic approximation to the finite-sample rejection rates of the feasible versions of these tests, based on estimated Ω , is very good for moderate sample sizes ($n \geq 100$) as long as there are not too many instruments ($k \leq 20$). The weak-instrument asymptotic power comparisons in Figs. 1–6 therefore can be expected to provide reliable guidance for comparing the powers of feasible CW and CLR tests in sample sizes typically encountered in econometric applications.

The evident conclusion for applied work is that researchers choosing among these tests should use the CLR test. The strong asymptotic bias and often-low power of the CW tests indicate that they can yield misleading inferences and are not useful, even as robustness checks.

Acknowledgments

The authors gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-0001706 (Andrews), SES-0418268 (Moreira), and SBR-0214131 (Stock), respectively.

References

- Anderson, T.W., Rubin, H., 1949. Estimators of the parameters of a single equation in a complete set of stochastic equations. *Annals of Mathematical Statistics* 21, 570–582.
- Andrews, D.W.K., Stock, J.H., 2005. Inference with weak instruments. Working Paper.
- Andrews, D.W.K., Moreira, M., Stock, J.H., 2006. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.
- Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. *Econometrica* 69, 1161–1191.
- Dufour, J.-M., 2003. Presidential address: identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767–808.
- Fuller, W.A., 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–953.
- Hahn, J., Hausman, J., Kuersteiner, G., 2004. Estimation with weak instruments: accuracy of higher-order bias and MSE approximations. *Econometrics Journal* 7, 272–306.
- Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1803.

- Moreira, M.J., 2001. Tests with correct size when instruments can be arbitrarily weak. Center for Labor Economics, Working Paper Series, 37, University of California, Berkeley.
- Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.
- Nagar, A.L., 1959. The bias and moment matrix of the general k -class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575–595.
- Nelson, C.R., Startz, R., 1990a. The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one. *Journal of Business* 63, S125–S140.
- Nelson, C.R., Startz, R., 1990b. Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica* 58, 967–976.
- Rothenberg, T.J., 1984. Approximating the distributions of econometric estimators and test statistics. In: Griliches, Z., Intriligator, M.D. (Eds.), *Handbook of Econometrics*, vol. 2. North-Holland, Amsterdam, pp. 881–935.
- Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
- Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20, 518–529.
- Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Stock, J.H., Andrews, D.W.K. (Eds.), *Identification and inference for econometric models: a Festschrift in honor of Thomas Rothenberg*. Cambridge University Press, Cambridge, pp. 80–108.