Valid Confidence Intervals and Inference in the Presence of Weak Instruments

Eric Zivot, Richard Startz and Charles R. Nelson*

Dept. of Economics, Box 353330
University of Washington
Seattle, WA 98195-3330

February 1996

revised December 1997

Abstract

We investigate confidence intervals and inference for the instrumental variables model with weak instruments. Wald-based confidence intervals for a structural parameter perform poorly in that the probability they reject the null is far greater than their nominal size. In the worst case, Wald-based confidence intervals always exclude the true structural parameter value. Confidence intervals based on the LM, LR, and Anderson-Rubin statistics perform far better than the Wald. The Anderson-Rubin statistic always has the correct size, but LM and LR statistics have somewhat greater power. Performance of the LM and LR statistics is improved by a degrees-of-freedom correction in the overidentified case. We show that the practice of "pre-testing" by looking at the significance of the first-stage regression and then making inference based on the Wald statistic leads to extremely poor results when the instruments are very weak. We show pre-testing leads to much better results if inference instead is based on the LM or LR statistics.

* Thanks to Jiahui Wang for excellent research assistance and computer programming, to Andrew Siegel for numerous long talks, and to Paul Ruud for his discussion at the 1996 ASSA meetings. Computations were carried out using GAUSS and Matlab. GAUSS code to compute the statistics in the paper can be found on the first author's web page at http://weber.u.washington.edu/~ezivot/wpapers.htm. Financial support from the Royalty Research Fund of the University of Washington and from the National Science Foundation under grant SBR-9711301 is gratefully acknowledged. Responsibility for errors is entirely the authors'.

1. *Introduction*

A series of recent papers has shown that inference based on instrumental variables (IV) estimation and asymptotic standard errors is generally misleading in finite samples when the instruments are weak. In particular, the IV estimate is strongly biased in the same direction as OLS and the estimated standard error is too small, the result being that the true null hypothesis is rejected much too often. Bound, Jaeger and Baker (1995), Hall, Rudebusch, and Wilcox (1996), Maddala and Jeong (1992), Nelson and Startz (1990a, b) and Staiger and Stock (1997) document these phenomena. Since weak instruments abound in economic data sets, (see Angrist and Krueger (1991, 1992), Campbell and Mankiw (1989), Fuhrer, Moore, and Schuh (1995), Hall (1988), McClellan, McNeil, and Newhouse (1994) and Rotemberg (1984) for some examples) there is clearly the need for procedures which produce test statistics that have the correct size in finite samples and so can be used to construct confidence regions that are valid in the sense of having the stated probability of covering the true value.

Traditionally in instrumental variable estimation, confidence regions are calculated and inferences are drawn based on the normal distribution with mean and variance taken from the sample estimated values of the parameters suggested by asymptotic distribution theory. Which is to say, a confidence region covers the parameter estimate plus or minus a multiple of the "asymptotic standard error." With a well identified system and enough observations this is a valid approach in the sense of producing confidence regions that cover the true value with the stated probability. Unfortunately, when instruments are weak and there is strong endogeneity, this traditional approach produces confidence regions that are highly misleading. Below, we construct examples in which traditional confidence regions *always* exclude the true parameter, or equivalently, in which the size of the Wald test is 100 percent. Fortunately, we are able to show that alternative confidence regions based on inverting the Lagrange multiplier, likelihood ratio, and Anderson-Rubin statistics are well-behaved, in that they have correct coverage probabilities, and easy to compute.

Little attention has been given in the econometrics literature to the possibility of inverting the likelihood ratio (LR) or Lagrange multiplier (LM) statistics to obtain a confidence region. In exception, Gallant(1987, pp. 107 ff.) suggests inverting the LR statistic in the context of non-linear regression[1]. In the weak instrument context structural parameters are nearly nonidentified and the likelihood function should be relatively flat. The intuition is

[1] Inverting LR statistics to form confidence intervals is more common in the nonlinear regression literature in statistics and is discussed in Bates and Watts (1988), Cook and Weisberg (1990), Meeker and Escobar (1995) and Venzon and Moolgavkar (1988).

appealing: a flat likelihood will result in an appropriately wide confidence region when the LR statistic is inverted. Indeed, we show that constructing confidence intervals for structural parameters obtained by inverting the AR, LR and LM statistics gives rise to unbounded intervals when instruments are very weak. Dufour's (1997) results provide theoretical support for the expectation that approximately correct probability levels can be obtained for these intervals. We verify the validity of our proposed confidence sets in finite samples through a series of Monte Carlo experiments. The asymptotic justification for our procedures is given in Wang and Zivot (1997).

There has been considerable interest in the recent literature in diagnostics for knowing when instruments are too weak for asymptotic theory to be valid. Nelson and Startz (1990b) suggested using the significance of the first stage regression, and Bound, Jaeger, and Baker (1995) have reiterated this advice. Shea (1997) has studied the multiple variable case. Hall, Rudebusch, and Wilcox (1996), however caution against choosing among instruments on the basis of their first stage significance, finding that screening worsens small sample bias. They find some merit in the practice of pretesting conditional on a given set of instruments. In this paper we find that Wald-based inference using decision rules of the type suggested by Nelson and Startz can be very misleading even if there is only one available instrument and the econometrician is obliged to judge its relevance on the basis of the single sample at hand. However, conditional on a given set of instruments, we find that pre-testing and using the LM or LR statistics for structural inference gives substantially better results.

The structure of the paper is as follows: Section 2 defines the Limited Information Simultaneous Equation Model studied in this paper, its likelihood function, IV and ML estimator. Section 3 discusses how Wald, LM, and LR statistics can be inverted to obtain confidence regions both within the maximum likelihood and instrumental variable (generalized method of moments) frameworks and shows that empirical confidence regions fall into one of four shapes. Section 4 discusses the properties of valid confidence sets in the presence of weak instruments and Section 5 gives examples of such valid intervals. Section 6 presents results of a Monte Carlo investigation of the actual coverage probabilities and relative power of alternative confidence regions. Section 7 investigates why the Wald statistic performs poorly. Section 8 concludes the paper.

2. The Limited Information Simultaneous Equation Model

The Limited Information Simultaneous Equation Model (LISEM) consists of a single structural equation which can be thought of as being selected from a simultaneous system. The equation relates a dependent endogenous variable, y, to explanatory variables, x, some of which are endogenous in the sense of being correlated with the disturbance in that equation, either because there is feedback in the complete system, or because variables correlated with the explanatory variable have been omitted. An accompanying “first stage regression” equation then relates the explanatory variable to a vector of *k* exogenous variables, *Z*, called instruments. Finally, the disturbances in the structure and first stage are contemporaneously (but not serially) correlated. The specific results in this paper are limited to the case of a single endogenous explanatory variable and, for expository purposes, we study the case where no additional exogenous explanatory variables appear in the structural equation. The results of the paper hold if we have additional exogenous explanatory variables but the algebra becomes unnecessarily cumbersome; see Wang and Zivot (1997) for details.

The LISEM may be written as:

$$\underset{(T\times 1)}{y} = \underset{(1\times 1)(T\times 1)}{bx} + \underset{(T\times 1)}{u} \tag{1}$$

$$\underset{(T\times 1)}{x} = \underset{(T\times k)(k\times 1)}{Zp} + \underset{(T\times 1)}{v} \tag{2}$$

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim iid\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} s_u^2 & s_{uv} \\ s_{uv} & s_v^2 \end{bmatrix}\right) = iid(0,\Sigma)\,. \tag{3}$$

The coefficient *b* in the structural equation (1) is the parameter of interest for inference, while the *k* coefficients in the vector π in the first stage regression (2) are not of direct interest. The model is said to be just identified if k=1 and $p \neq 0$ and overidentified if k>1 and the number of nonzero elements of *p* is greater than one. Define $r = s_{uv}/s_u s_v$, $Y = [y\ x]$, $P_Z = Z(Z'Z)^{-1}Z'$, for any full rank matrix *Z*, and $M_Z = I - P_Z$.

We now review the instrumental variable and maximum likelihood approaches to estimating *b*. The two stage least squares estimator (2SLS) for *b* is $\hat{b}_{2SLS} = (x'P_Z x)^{-1}x'P_Z y$. This is also equal to the efficient instrumental variable (IV) and efficient generalized method of moments (GMM) estimator. The GMM estimator solves $\min_b J_T(b) = T^{-1}(y - xb)'P_Z(y - xb)/\hat{s}^2$, where $\hat{s}^2$ is a consistent estimator of s^2. Under standard regularity conditions $\sqrt{T}\left(\hat{b}_{2SLS} - b\right) \xrightarrow{d} N\left(0, s_u^2(p'V_{ZZ}p)^{-1}\right)$, where $V_{ZZ} = \operatorname{plim}_{T\to\infty} T^{-1}Z'Z > 0$. Under weak instruments, Staiger and Stock (1997), hereafter SS, show that $\hat{b}_{2SLS}$ is not consistent but converges to a ratio of quadratic forms in correlated normal random variables.

The maximum likelihood estimator for this model under the assumption of normality was first derived by Anderson and Rubin (1949) and is referred to as the Limited Information Maximum Likelihood (LIML) estimator. The concentrated log likelihood function for b is given by (cf. Davidson and MacKinnon (1993), p. 647)

$$\ell^c(b) = -T\ln(2p) - \frac{T}{2}\ln(k(b)) - \frac{T}{2}\ln\left|Y'M_Z Y\right| \tag{4}$$

where

$$k(b) = \frac{(y-bx)'(y-bx)}{(y-bx)'M_Z(y-bx)}. \tag{5}$$

The LIML estimator of b is obtained by minimizing $k(b)$, a result first demonstrated by Rubin(1948); see also Koopmans and Hood(1953). Thus $\hat{b}_{\text{LIML}} = \underset{b}{\arg\min}\ k(b)$. Operationally, $k(\hat{b}_{LIML}) = \hat{k}$ is the smallest eigenvalue of the matrix $(Y'M_Z Y)^{-1/2}Y'Y(Y'M_Z Y)^{-1/2}$ and $\hat{b}_{LIML}$ is given by the k-class estimator formula $\hat{b}_{\text{LIML}} = (x'(I-\hat{k}M_Z)x)^{-1}(x'(I-\hat{k}M_Z)y)$. When $\hat{k}=1$, which is true in a just identified model, $\hat{b}_{\text{LIML}} = \hat{b}_{\text{2SLS}} = (x'\text{P}_Z x)^{-1}x'\text{P}_Z y$. Under standard regularity conditions $\sqrt{T}(\hat{b}_{\text{LIML}} - b) \xrightarrow{d} N(0, s_u^2(p'V_{ZZ}p)^{-1})$ for the just identified and overidentified models and is equivalent to the asymptotic distribution of $\sqrt{T}(\hat{b}_{\text{2SLS}} - b)$. Under weak instruments, however, SS show that $\hat{b}_{LIML}$ is not consistent and converges to a distribution different than the one for $\hat{b}_{2SLS}$.

3. Construction of Confidence Sets by Inverting Test Statistics

We are interested in constructing confidence sets for the structural parameter b in (1). Given a test statistic $T(b_0)$ for testing the hypothesis $\text{H}_0: b = b_0$ at the a significance level, the $(1 - a)\cdot 100\%$ confidence set associated with this statistic is defined as

$$C_T(b; 1-a) = \left\{b_0 : T(b_0) \le cv_{1-a}\right\},$$

where cv_{1-a} is the $(1 - a)$ quantile from the (asymptotically valid) distribution of the test statistic $T(b_0)$; i.e., C_T contains all of the "acceptable" values of b_0 at level a for the null hypothesis $\text{H}_0: b = b_0$ using the test statistic $T(b_0)$. Confidence sets formed this way are said to be determined by "inverting" the test statistic $T(b_0)$.

We are interested in confidence regions corresponding to the Wald, LM, and LR statistics based on maximum likelihood estimation, the three analogous statistics based on the GMM/2SLS framework, and the Anderson-Rubin statistic. Due to the simple form of the hypothesis test there is some redundancy among these statistics. Consequently, we only need to consider four, or at most five, different ways to compute confidence regions.

The Wald, LM (see Engle (1984)), and LR statistics based on limited information maximum likelihood are given respectively by

$$\mathrm{Wald}_{\mathrm{LIML}}\left(b_0\right)=\frac{\left(\hat{b}_{LIML}-b_0\right)^2}{\mathrm{EAVAR}\left(\hat{b}_{LIML}\right)},$$

$$\mathrm{LM}_{\mathrm{LIML}}\left(b_0\right)=\frac{g\left(b_0\right)^2}{\mathrm{EAVAR}\left(b_0\right)},$$

$$\mathrm{LR}_{\mathrm{LIML}}\left(b_0\right)=-2\left[\ell^c\left(b_0\right)-\ell^c\left(\hat{b}_{LIML}\right)\right],$$

where $\mathrm{EAVAR}(b)$ denotes an estimate of the asymptotic variance of $\hat{b}_{LIML}$ evaluated at b, and $g(b)=\frac{d}{db}\ell^c(b)$ is the gradient of the concentrated log likelihood for b. Notice that the LM statistics is computed using the value of b under the null hypothesis. Under standard assumptions, the three statistics are asymptotically $c^2(1)$. Under weak instruments, SS show that $Wald_{LIML}$ is not asymptotically $c^2(1)$. They did not investigate the properties of LM_{LIML} or LR_{LIML}.

The analogous 2SLS or efficient GMM based statistics, which also have asymptotic $c^2(1)$ distributions under standard conditions (see Newey and West (1987)), are:

$$\mathrm{Wald}_{2SLS}\left(b_0\right)=\frac{\left(\hat{b}_{2SLS}-b_0\right)^2 x'\,\mathrm{P}_Z\, x}{\hat{s}^2},$$

$$\mathrm{LM}_{2SLS}\left(b_0\right)=\frac{\left(y-xb_0\right)'\,\mathrm{P}_{\hat{x}}\left(y-xb_0\right)}{\hat{s}^2},$$

$$\mathrm{LR}_{2SLS}\left(b_0\right)=\frac{\left[\left(y-xb_0\right)'\,\mathrm{P}_Z\left(y-xb_0\right)-\left(y-x\hat{b}_{2SLS}\right)'\,\mathrm{P}_Z\left(y-x\hat{b}_{2SLS}\right)\right]}{\hat{s}^2},$$

where $\hat{s}^2 \xrightarrow{p} s_u^2$. SS show that *Wald*$_{2SLS}$ is not asymptotically $c^2(1)$ under weak instruments. They did not consider *LM*$_{2SLS}$ or *LR*$_{2SLS}$.

Several facts are worth noting. First, because of the quadratic nature of the 2SLS/GMM minimization problem and the linearity of the restriction $b = b_0$, the three 2SLS statistics are numerically identical so long as the same estimate is used for $\hat{s}^2$. (c.f. Newey and West (1987).) Second, when using $\hat{s}^2 = \hat{s}^2_{\text{2SLS}} = T^{-1}\left(y - x \cdot \hat{b}_{\text{2SLS}}\right)'\left(y - x \cdot \hat{b}_{\text{2SLS}}\right)$, *Wald*$_{2SLS}$ is simply the (square of) "asymptotic *t*," which is the statistic used essentially always for inference in applied work. Further, the *LR*$_{2SLS}$ is the statistic calculated in the Hansen (1982) GMM framework as the "difference in the *J*-statistic." Third, use of $\hat{s}^2 = s_0^2 = T^{-1}(y - xb_0)'(y - xb_0)$ instead of $\hat{s}^2_{2\,\text{SLS}}$ (s_0^2 is the natural choice when thinking of using *LM*$_{2SLS}$) is shown below to make a critical difference in inference when instruments are weak. This is due to the fact that, under the null and weak intruments, $\hat{s}^2_{2\,\text{SLS}}$ is not consistent whereas s_0^2 is.

Other than the choice of $\hat{s}^2$, the 2SLS statistics are straightforward. In contrast, evaluation of the MLE statistics requires a choice of EAVAR. A variety of specifications are available. For *Wald*$_{LIML}$, one usually sees the k-class formula with $\text{EAVAR} = \hat{s}^2_{\text{LIML}} \cdot \left(x'\left(I - \hat{k} \cdot M_Z\right)x\right)^{-1}$ and $\hat{s}^2_{\text{LIML}} = T^{-1}\left(y - x \cdot \hat{b}_{\text{LIML}}\right)'\left(y - x \cdot \hat{b}_{\text{LIML}}\right)$.

For *LM*$_{LIML}$, it is useful to base EAVAR on the information matrix from the concentrated likelihood function (see Bowden and Turkington (1984)).

$$\text{EAVAR}(b) = -\left[\frac{d^2 \ell^c(b)}{db^2}\right]^{-1} = \left[\text{A}(b) + \text{B}(b)\right]^{-1},$$

$$\text{A}(b) = T \cdot \left[\frac{x' M_Z x}{(y - xb)' M_Z (y - xb)} - \frac{x'x}{(y - xb)'(y - xb)}\right]$$

$$\text{B}(b) = 2T \cdot \left[\frac{\left[x'(y - xb)\right]^2}{\left[(y - xb)'(y - xb)\right]^2} - \frac{\left[x' M_Z (y - xb)\right]^2}{\left[(y - xb) M_Z' (y - xb)\right]^2}\right].$$

The LM statistic can then be written as

$$\mathrm{LM}_{\mathrm{LIML}}(b_0) = \left\{ T\left[\frac{x'u_0}{u_0'u_0} - \frac{x'M_Z u_0}{u_0'M_Z u_0} \right] \right\}' \left\{ \mathrm{A}(b_0) + \mathrm{B}(b_0) \right\}^{-1} \left\{ T\left[\frac{x'u_0}{u_0'u_0} - \frac{x'M_Z u_0}{u_0'M_Z u_0} \right] \right\}, \tag{6}$$

where $u_0 = y - xb_0$. The LM statistic as given in (6) is not easily written as a quadratic in b_0. However, using the following approximation results (which are valid under the standard conditions)

$$T\left[\frac{x' u_0}{u_0' u_0} - \frac{x' M_Z u_0}{u_0' M_Z u_0} \right] \overset{A}{\approx} T \frac{\hat{x}' u_0}{u_0' u_0} \text{ and } \mathrm{A}(b_0) + \mathrm{B}(b_0) \overset{A}{\approx} \frac{u_0' \mathrm{P}_{\hat{x}} u_0}{u_0' u_0}$$

we obtain a simpler version of the LM statistic

$$\mathrm{LM}_{LIML} = LM = \frac{(y - x b_0)' \mathrm{P}_{\hat{X}} (y - x b_0)}{(y - x b_0)' (y - x b_0) / T},$$

where $\hat{x} = P_Z x$. Here LM_{LIML} is equal to T times the uncentered R^2 from the regression of $y - xb_0$ on $\hat{\mathrm{x}}$. We note that this approximation will make LM_{LIML} identical to the corresponding LM statistic in the 2SLS framework and henceforth we will refer to these two statistics as LM. To our knowledge the LM statistic is new.

Finally, consider the statistic proposed by Anderson and Rubin (1949) and Anderson (1950) for testing $\mathrm{H}_0\!: b = b_0$. Rewrite (1) by adding and subtracting xb_0 from both sides and substituting in for x using equation (2) to give

$$y^* = Zy + u^*, \tag{7}$$

where $y^* = y - xb_0$, $y = p(b - b_0)$, and $u^* = v(b - b_0) + u$. Then the hypothesis $\mathrm{H}_0\!: b = b_0$ in (1) corresponds to the hypothesis $H_0^*\!: y = 0$ in (7). The latter hypothesis can be tested with the standard F-statistic

$$\mathrm{AR} = F_{y=0} = \frac{(RSS_R^* - RSS_{UR}^*)/k}{RSS_{UR}^*/(T-k)} = \frac{(y - x b_0)' \mathrm{P}_Z (y - x b_0)/k}{(y - x b_0)' M_Z (y - x b_0)/(T-k)}.$$

If indeed (u_i v_i) is distributed iid N(0, S), and the model is correct in the sense that the identifying restrictions that exclude Z from the structural equation are true, then the AR statistic is distributed *exactly* as $F(k, T-k)$. Further, SS show that $AR \overset{d}{\rightarrow} c^2(k)/k$ under fairly general assumptions about the disturbances and the quality of the instruments.

Creating confidence regions by "inverting" the corresponding test statistics means solving for the range of values of b_0 for which the test statistic is less than or equal to the appropriate critical value. The results are most easily seen graphically. Figures 1 through 6 show plots of the test statistics as functions of b_0 for a particular Monte Carlo run for a variety of models with good and weak instrument.[2] For a given statistic, the corresponding confidence region is that region in which the statistic is below the horizontal critical value line. Notice that the Wald statistic, which is a quadratic function, is u-shaped so its confidence region is always a closed set. The other statistics, being ratios of quadratics, can be u-shaped as in Figure 1 or can have both a minimum and a maximum, as seen in Figures 2 and 5. In the latter case, three different types of confidence sets are possible. These can be seen by raising or lowering the critical value line in Figure 2. It is easy to see that at a very high confidence level the test statistics are everywhere below the critical value so that the confidence region includes the entire real line. At a somewhat lower confidence level the line will "cut-off" the peak of the test statistic so the confidence region will consist of the area from the left cut-point to negative infinity and from the right cut-point to positive infinity. Finally, as seen in Figure 4, when $k>1$ the AR statistic can imply an empty confidence region since the statistic is always positive.

For each statistic we can give simple closed form solutions for the cut-points. The confidence sets for b formed by inverting the LM, LR, and AR statistics are each determined by solutions to an inequality of the form

$$ab_0^{\ 2} + bb_0 + c \leq 0 \qquad (8)$$

where values of a, b, and c depend on the data and the critical value for the particular test. The cut-points are the roots of the quadratic equation

$$b_{0,i} = \frac{-b \pm \sqrt{d}}{2a}; \; i = 1, 2$$

where $d = b^2 - 4ac$ denotes the discriminant of the quadratic. We now give the formulas for the quadratic coefficients for the respective statistics and then characterize the shape of the confidence region in terms of the solution to the quadratic.

The confidence set $C_{LM}(b;1-a)$ requires finding all values of b_0 satisfying the condition

[2] The Monte Carlo design is discussed in Section 5.

$$\frac{\left(y-x b_0\right)' \mathrm{P}_{\hat{X}}\left(y-x b_0\right)}{\left(y-x b_0\right)'\left(y-x b_0\right)} \leq \frac{c_{1-a}^2(1)}{T} \equiv f_{LM}$$

which can be rearranged as a quadratic of the form of (8). Defining the 2x2 matrix $Q_{LM} = Y'\left(\left[1-f_{LM}\right]I - \left[\left(I-\mathrm{P}_{\hat{X}}\right)\right]\right)Y$ then $a = Q_{22}$, $b = -2 \bullet Q_{12}$, and $c = Q_{11}$.[3], where Q_{ij} is the $(i,j)^{th}$ element of Q. Note that a is closely related to the significance of the first stage, a topic to which we return below. Turning now to the LR statistic, using the concentrated likelihood function (4), the hypothesis $\mathrm{H}_0\!: b = b_0$ is accepted if

$$\frac{\left(y-x b_0\right)'\left(y-x b_0\right)}{\left(y-x b_0\right)' M_Z\left(y-x b_0\right)} \leq \exp\left\{\frac{c_{1-a}^2}{T}\right\} \bullet \hat{k} \equiv f_{LR}\,.$$

This inequality can also be expressed in the form of (8). The corresponding matrix Q is given by $Q_{LR} = Y'\left[I - f_{LR} \bullet M_Z\right]Y$. Finally, the AR confidence set consists of all values of b_0 that satisfy the inequality

$$\frac{(y-xb_0)'(y-xb_0)}{(y-xb_0)' M_Z (y-xb_0)} \leq 1 + \mathrm{F}_{1-a}(k, T-k)\frac{k}{T-k} \equiv f_{AR}$$

where $\mathrm{F}_{1-a}(k, T-k)$ denotes the (1- a) quantile of the F distribution. Note that this condition is very similar to that given for the LR statistic. As with the LM and LR statistics, the AR confidence set is determined by solving the inequality (8) with the corresponding matrix Q given by $Q_{AR} = Y'\left[I - f_{AR} \bullet M_Z\right]Y$.

Whether the confidence region is bounded, empty, external, or covers the real line is determined by the signs of a and d as follows. First consider the case $a > 0$, so the inequality may be rewritten as $b_0^2 + (b/a) \bullet b_0 + (c/a) \leq 0$ which is convex from below. If the inequality is satisfied at all, it will be for a bounded interval. If also $d > 0$ then the solutions to the quadratic equation are real and there is a bounded interval with end points corresponding to the two solutions, say $\left[b_{LOW}, b_{HIGH}\right]$. We show below that this occurs when the instruments are of good quality. Alternatively, if $d < 0$ the roots are complex so there is no value of b_0 which satisfies the inequality and thus the confidence set is empty. The LM and LR confidence sets cannot be empty because at $b_0 = \hat{b}$ the statistics equal zero. When $b_0 = \hat{b}$, the

[3] The formula given here is for the LISEM model (1)-(3). If additional exogenous regressors are present, then y, x, and Z in the definition of Q should be replaced with the residuals from regressing the y, x, or Z on the exogenous regressors. Modify the formulas for the LR and AR below in the same way.

AR statistic tests significance by regressing the residuals on the instruments. In the just-identified case the AR statistic is therefore zero so the confidence set cannot be empty, but it will be empty in overidentified models when the overidentifying restrictions are rejected.[4]

Next consider $a < 0$. The quadratic inequality may then be rewritten as $\beta_0{}^2 + (b/a) \bullet b_0 + (c/a) \geq 0$ which is again convex from below. If $d > 0$ there are again real solutions to (8), but now it is values of b_0 *outside* the interval $\left[b_{LOW}, b_{HIGH}\right]$ which satisfy the inequality (because the sign of the inequality is reversed) so the confidence set is the disconnected region $\left[-\infty, b_{LOW}\right] \cup \left[b_{HIGH}, \infty\right]$. Finally, if $d < 0$ then there is, again, no real solution to the quadratic equation, but this means that the inequality is satisfied for all values of b_0. We show below that unbounded confidence regions are associated with weak instruments.

4. Valid Confidence Sets in the Presence of Weak Instruments

Dufour (1997) shows that any valid $(1-a) \bullet 100\%$ confidence set for *b* must be unbounded with probability $1-a$ for models in which *b* is nearly nonidentified. Since the AR statistic is an exact test regardless of instrument quality, it must satisfy the Dufour requirement. The LM and LR statistics do not satisfy the Dufour requirement in the overidentified case, but versions with a degrees of freedom modification do, as shown below. Finally, since inverting a Wald statistic always produces a bounded confidence interval, such an interval is not valid in nearly nonidentified models. Indeed, Dufour shows that Wald confidence intervals then have zero coverage probability.

Since an unbounded confidence set occurs when the coefficient *a* in (8) is less than zero, we are able to link unboundedness with goodness-of-fit statistics for the first stage regression as follows:

Proposition

(a) $C_{AR}(b;1-a)$ is unbounded if $F_{p=0} < F_{1-a}\left(k, T-k\right)$

(b) $C_{LR}(b;1-a)$ is unbounded if $F_{\pi=0} < \frac{T-k}{k}\left(\exp\left\{\frac{c_{1-a}^2(1)}{T}\right\}\hat{k}-1\right)$

(c) $C_{LM}(b;1-a)$ is unbounded if $LM_{p=0} = T \cdot R_{UC}^2 < c_{1-a}^2(1)$

---

[4] Note that the AR statistic evaluated at *b* is essentially the *J*-statistic for testing over-identifying restrictions and that the *J*-statistic is minimized at $\hat{b}$.

where $F_{p=0}$ is the F statistic for testing p = 0 in (2), $LM_{p=0}$ is the LM statistic for testing p=0 and R_{UC}^2 is the uncentered (no intercept) R^2 from (2). See appendix for proofs.

The link between the unboundedness of confidence sets and the significance of first stage test statistics allows us to verify the Dufour requirement for a valid confidence set. For example, the AR confidence set, $C_{AR}(b;1-a)$, has the very interesting property that it is unbounded whenever $F_{\pi=0}$ is insignificant at level a. Under Normality and fixed regressors, when the model is not identified (π=0) then P{ $F_{p=0} < F_{1-a}(k,T-k)$ } = 1 − a. Hence, the AR confidence set satisfies Dufour's requirement exactly in finite samples. Under more general conditions, $k \cdot F_{p=0}$ is asymptotically $\chi^2(k)$ and so the Dufour requirement is satisfied asymptotically.

Further, the condition for $C_{LR}(b;1-a)$ to be unbounded can be simplified when T is large relative to k and the overidentifying restrictions are valid. In this case, $\hat{\kappa} \approx 1$ and $\exp\left\{T^{-1} \cdot c_{1-a}^2(1)\right\} \approx 1 + T^{-1} \cdot c_{1-a}^2(1)$ so that (b) above becomes $F_{\pi=0} < k^{-1} \cdot c_{1-a}^2(1)$. Notice that this condition is similar to the condition in (a) above for the AR confidence set since, for large T, $F_{1-a}(k,T-k) \approx k^{-1} \cdot c_{1-a}^2(k)$. However, the condition for the LR confidence set uses $c_{1-a}^2(1)$ whereas the condition for the AR set uses $c_{1-a}^2(k)$. When k=1, these conditions are identical and so the LR confidence set satisfies the Dufour requirement asymptotically. For k>1, $c_{1-a}^2(k) > c_{1-a}^2(1)$ so it follows that when π=0, P{$C_{LR}(b;1-a)$ is unbounded} = P{$F_{\pi=0} < k^{-1} \cdot c_{1-a}^2(1)$} < P{$F_{\pi=0} < k^{-1} \cdot c_{1-a}^2(k)$} = 1-$a$. Hence, in the unidentified case, $C_{LR}(b;1-a)$ is unbounded with probability *less* than *1-a* and so is not a valid confidence set.

Finally, a similar result holds for the LM confidence set. When the model is not identified (p = 0), $LM_{p=0} \overset{A}{\sim} c_{1-a}^2(k)$ and so when k=1 the Dufour requirement is satisified asymptotically. However, when k>1 P{$C_{LM}(b;1-a)$ is unbounded} = P{ $LM_{p=0} < c_{1-a}^2(1)$} < P{ $LM_{p=0} < c_{1-a}^2(k)$} = $1-a$, implying that $C_{LM}(b;1-a)$ is not a valid confidence set.

The above shows that in the very weak instrument case, inverting the LR and LM statistics for testing H_0:$b = b_0$ using $c^2(1)$ critical values is not asymptotically valid when k>1, but is using $c^2(k)$ critical values. It appears, then, that the asymptotic distributions of the LM and LR test statistics in this case are poorly approximated by the $c^2(1)$ distribution and are better approximated by the $c^2(k)$ distribution. This conjecture can be shown rigorously using the local-to-zero framework of SS, which provides a convenient way to obtain analytical results when instruments are weak.[5] Wang and Zivot (1997) derive the asymptotic distributions of the LM and LR statistics using SS's local-to-zero framework and show that when k=1 these statistics converge in distribution to $c^2(1)$ but when $k > 1$ they do not, but rather are bounded by $c^2(k)$.

How, then should we construct confidence intervals in practice when instruments are of doubtful quality? A conservative option is to always use $c^2(k)$ critical values to invert the AR, LM, or LR statistics. In the AR case, the coverage probability is exact anyway, but for the LM and LR the coverage probabilities are *at least* the stated level since their distribution is bounded by $c^2(k)$. However, when k is large and the instruments are good, the resulting confidence intervals for LM and LR are larger than necessary since then we could be using $c^2(1)$ critical values instead. Thus, if we had a reliable rule for switching between critical values based in instrument quality we could achieve, on average, smaller but still valid confidence intervals. The link between the unboundedness of confidence intervals and the significance of the first stage regression suggests that $F_{\pi=0}$ and $LM_{\pi=0}$ are appropriate statistics to use for the switching criterion. In particular, we recommend that if $F_{\pi=0}$ < $F_{1-a}(k, T-k)$, or $F_{\pi=0} < k^{-1} \cdot c^2_{1-a}(k)$, then the LR statistic be inverted using critical values from $c^2(k)$ instead of $c^2(1)$. Similarly, if $LM_{p=0} < c^2_{1-a}(k)$ then the LM statistic should be inverted using $c^2(k)$ instead of $c^2(1)$ critical values. We call the test statistics which switch degrees of freedom based on a pre-test of the significance of the first stage LM_{sw} and LR_{sw}. Hall, Rudebusch, and Wilcox (1996) show that the first stage regression is not very useful in judging the validly of Wald-based confidence intervals, but we find that pre-testing works well as a criterion for selecting degrees of freedom for the LM and LR as will be seen in the following sections.

---

[5] In this framework, the coefficients p in (2) are modeled as being in a $T^{-1/2}$ neighborhood of zero. Specifically, $p = p_T = T^{-1/2}g$ where g is any $k \times 1$ vector of constants. This device keeps the statistic $F_{\pi=0}$ bounded in probability as the sample size increases so that b is asymptotically nearly nonidentified. SS show that the 2SLS and LIML Wald statistics do not converge to $c^2(1)$ random variables but rather to random variables that depend on the nuisance parameters r, k and the noncentrality parameter of the asymptotic distribution of $F_{\pi=0}$. This result is expected since Dufour shows that any statistic that gives rise to a bounded confidence interval for a nearly nonidentified parameter must have a non-pivotal distribution (i.e., a distribution that depends on nuisance parameters). Dufour also shows that if a test statistic produces an unbounded confidence set for a nearly nonidentified parameter then the distribution of the test statistic is pivotal or can be bounded by a pivotal distribution.

5. Examples of Confidence Intervals for *b*

To illustrate the typical shapes of confidence intervals for instruments of various quality we generated data from (1)-(3) with $b = 1, s_u^2 = s_v^2 = 1, r = .99, T = 100$, $Z \sim \mathrm{N}(0, I_k)$ for just identified (k=1) and nominally overidentified (k=4) models. For the just identified model, we set p=1 (good instrument case), p=0.1 (weak instrument case) and p=0 (unidentified case). For the overidentified model, we set p=(1,0,0,0)′ (good instrument case), p=(0.1,0,0,0)′ (weak instrument case) and p=(0,0,0,0)′ (unidentified case). The data represent one Monte Carlo draw for the errors and the same random numbers were used for each plot. The graphs are representative of the typical shapes that occur over many replications. For each set of generated data we computed the OLS, 2SLS and LIML estimates of b, the reduced form estimate of p, the reduced form $LM_{p=0}$ and $F_{\pi=0}$ statistics and the 95% confidence sets C_{Wald}, C_{LR}, C_{LM} and C_{AR}. These statistics are summarized in Tables 1 and 2 and the confidence intervals are displayed graphically in Figures 1 - 6.

Consider first the results for the just identified models. For a good instrument, illustrated in Figure 1, the standard Wald confidence region C_{Wald}, [0.990, 1.310], is fairly small and contains the true value b = 1. The LR, LM and AR regions are all very similar to each other and to the Wald interval in this case. The situation is much different in the weak instrument case seen in Figure 2. The combination of a weak instrument and strong endogeneity induces a noticeable bias in $\hat{b}_{2SLS}$ and, counter to intuition, the Wald confidence interval is fairly short and does not cover the true value b = 1. Since $F_{p=0}$ and $LM_{p=0}$ are significant at the 5% level the LR, LM, and AR confidence sets are all closed intervals but they are considerably larger than the Wald interval and contain the true value. In the unidentified case of Figure 3, the OLS and 2SLS estimates of b are almost identical and the Wald confidence region indicates a very precise estimate even though the goodness-of-fit statistics from the first stage regression suggest a poor instrument. The LR, LM and AR confidence intervals in this case are equivalent and contain all possible values of b. This is what we should expect when b is unidentified.

Now consider a nominally overidentified, four instrument model. We vary the quality of the first instrument, while the other three are always irrelevant (that is, their reduced form coefficients are zero). The good instrument case is very similar to the k = 1 case. The LM and LR confidence sets based on $c^2(1)$ critical values are closed intervals, are very similar to C_{Wald} and have roughly the same length. The AR confidence set, however, is substantially larger that the other sets. Turning next to the case of one valid, but weak instrument, we see that C_{Wald} is fairly wide, but does not cover b =1. Here the reduced form statistics $F_{p=0}$ and $LM_{p=0}$ are not significant at the 5% level using, $c^2(1)$ or $c^2(4)$ critical values, indicating that the instruments are poor and b is nearly unidentified. From the previous section we know that C_{AR}, C_{LM} and C_{LR} will be unbounded and indeed these sets are unbounded disjoint regions. Finally, in the nonidentified case C_{Wald} is short and does not cover b =1. The reduced form goodness-of-fit statistics are small and statistically insignificant at any reasonable level and, consequently, the confidence sets C_{AR}, C_{LM}, and C_{LR} are unbounded, containing all possible values of b.

6. A Monte Carlo Investigation of Size and Power

In this section we analyze the finite sample properties of the 95% confidence regions for b formed by inverting the level 0.05 Wald, LM, LR and AR test statistics for $H_0: \; b = b_0$. We compare empirical coverage probabilities of the confidence sets under the null as well as empirical powers of the test statistics under a range of alternatives $H_a: b = b_a$. Our Monte Carlo design is the same as in section 5 except that we consider r = (0.99,0.5,0}. For the power analysis we generate data under the alternatives $b_a = b_0 + d_i$ where d_i ranges from -2 to 2 in increments of 0.25. The empirical probabilities of the confidence sets under the null for the case r=0.99 are summarized in Tables 3 and 4 and results on power for the cases p = 1 and p = (1,0,0,0)′ are given in Tables 5 and 6.[6]

Consider first the size results for the k = 1 case. Since, as shown in Section 3, the LM, LR, and AR statistics are approximately $c^2(1)$ regardless of the values of r and p, the 95% confidence sets formed by inverting these statistics have empirical coverage frequencies very close to 95% in all cases. However, the situation is different for C_{Wald} since, as SS show, the distribution of the Wald statistic depends on r and p in the weak instrument case. In the unidentified (p = 0) case with r = 0.99, C_{Wald} covers the true value b_0 = 1 in less than 37% of the samples and the sets C_{LM} , C_{LR} and C_{AR} are unbounded with frequency 0.95. The results for the weak instrument case (p = 0.1) are similar to the unidentified case. In the good instrument case (p = 1), all of the 95% confidence sets are bounded intervals with correct coverage frequency. The sets C_{LM} , C_{LR} and C_{AR} are about the same length on average and C_{Wald} is slightly shorter.

[6] The results for the other cases do not add much to the discussion and are therefore omitted. They are reported in the working paper version of the paper.

Next consider the size results for the nominally over identified, k=4, model. In the unidentified and weakly identified cases with r = 0.99 the actual coverage frequencies of C_{Wald} are 1.3% and 14.5%, respectively. By contrast, C_{AR} has the correct coverage frequency regardless of instrument quality but is about 50% larger, on average, than C_{Wald} in the good instrument case. Interestingly, C_{AR} is empty about 2% of the time in the good instrument case and is empty slightly less frequently in the other cases. In the unidentified and weak instrument cases, the sets C_{LM} and C_{LR} computed using $c^2_{.95}(1)$ have actual coverage frequencies less than .95 for all values of r. The sets C_{LM} and C_{LR} based on $c^2_{.95}(4)$ have actual frequencies of at least .95 in all cases and this supports the use of $c^2(4)$ as a bounding distribution for the LM and LR statistics in finite samples. In all cases, C_{LM} and C_{LR} are very close to C_{AR}. Finally, The sets C_{LM} and C_{LR} using $c^2_{.95}(1)$ or $c^2_{.95}(4)$ based on a pretest of the significance of the first stage regression perform much better than the sets based solely on $c^2_{.95}(4)$. They have approximately correct coverage frequencies in all cases and in the good instrument case they are shorter, on average, than C_{AR} and are very close to the sets based on $c^2_{.95}(1)$ and C_{Wald}. Hence, pretesting on the first stage does not adversely affect the size of the LM or LR statistics.

Now consider the issue of power[7]. For the weak instrument cases, the Wald test is severly sized distorted and the tests based on LM, LR and AR are not consistent, which is to be expected since with weak instruments b is nearly nonidentified, and so power comparisons amongst these statistics are not particularly meaningful. With good instruments, however, these tests have approximately the correct size and are consistent so power comparisons are informative. When k=1, the powers of the LM, LR, and AR statistics are very similar and are also nearly identical to the power of the Wald test[8]. Differences in test power occur when k>1 and instruments are good. The powers for the Wald statistic are symmetric in b_a but the powers of the AR, LR and LM tests are asymmetric when $r \neq 0$. The powers of Wald, LR_{SW} and LM_{SW} are very close and are higher than the power of AR. AR has higher power than LR and LM if the $c^2(4)$ bounding distribution is used.

[7] The reported power results are not size adjusted. This mainly affects the Wald statistic since the actual sizes of the AR, LR and LM statistics are close to the nominal size.

[8] We note that Maddala (1974) has previously studied the power of the AR test and shown it to be comparable to the power of the Wald test in just identified models with good instruments.

7. Why Do Traditional Wald Confidence Intervals Perform So Poorly?

With a large enough sample, asymptotic distribution theory approximates actual sampling distributions and should provide a good guide to inference. Having observed the failure of Wald based inference, it is natural to conclude the problem is that the distribution $\mathrm{N}\left(\hat{b}_{2SLS}, \hat{s}^2 (x'P_Z x)^{-1}\right)$ does a poor job of approximating the true sampling distribution. Curiously, it's just the other way around. The reported distribution represents the sampling distribution quite accurately, but with weak instruments and significant endogeneity the sampling distribution is highly concentrated and isn't located particularly near the true parameter (See Phillips (1989) and Nelson and Startz (1990)). We illustrate the problem in two ways, first by looking more closely at the likelihood function and then by comparing the actual and reported sampling distributions.

Return to Figure 2, which shows the Wald statistic and the LR statistic, the latter being the log-likelihood function less a constant. The apparently flat likelihood function actually has a very sharp peak around $\hat{b}_{2SLS}$. The Wald statistic does a *good* job of approximating this peak. Inference doesn't work very well because while the peak in the likelihood function is very sharp, there is very little mass under it.

Turn now to the question of how well the sampling distribution is approximated by $\mathrm{N}\left(\hat{b}_{2SLS}, \hat{s}^2_{2SLS} (x'P_Z x)^{-1}\right)$. There are both series and closed form expressions for the density of $\hat{b}$ in quite general situations. (See Sawa (1969) and Phillips (1983).) These expressions do not lend themselves to easy interpretation. However, Phillips (1989) and SS give the following expression for the exact distribution $\hat{b}_{2SLS}$ in the completely unidentified case of (1)-(3).

$$\hat{b}_{2SLS} \equiv b + q + \frac{h}{\sqrt{k}} \cdot t_k \tag{9}$$

where $\equiv$ denotes equivalence in distribution, $q = rs_u / s_v$, $h = \left(1 - r^2\right)^{0.5} s_u / s_v$, and t_k denotes a Student-*t* random variable with k degrees of freedom. Figure 7 shows both the exact distribution and the normal approximation evaluated at the median values of $\hat{b}_{2SLS}$ and its associated asymptotic standard error from two unidentified models based on the Monte Carlo design used in Sections 5 and 6. In both cases the reported distribution is quite close to the true distribution, differing mostly in that the true distribution (which is somewhat Cauchy-like) has fatter tales. The problem with inference arises in the case of strong endogeneity because the distribution is centered near the point of concentration. When $r = 0$ there is no endogeneity and the distribution is approximately median unbiased, which is consistent with the result reported in Nelson and Startz (1990a).

To understand why Wald-based confidence intervals can be misleadingly small, consider the density in (9) as $r \to 1$, the worst possible case. Here $q \to s_u/s_v$ and $h \to 0$ so that the density is zero except for a spike at $\hat{b}_{2SLS} = b + {}^{s_u}\!/_{s_v}$. Given that the estimator collapses to the point of concentration, one can write the instrumental variable residuals as $y - \hat{b}_{2SLS}x = y - \left(b + {}^{s_u}\!/_{s_v}\right)x = u - {}^{s_u}\!/_{s_v}\, x$. But in this case $x = v$ and $u = {}^{s_u}\!/_{s_v}\, v$, so the residuals collapse to zero. Since $\hat{s}^2_{2SLS}$ is just the mean sum squared residuals, it too collapses to zero. Thus the Wald confidence intervals, based on $\hat{s}^2_{2SLS}$, are far too small. In contrast, the LM, based on $\hat{s}^2_0$, is immune to this problem.

The linkage between the first-stage fit and the sampling distribution of both instrumental estimators and test statistics has led many practitioners to an informal pre-test rule: if the first-stage is "significant" proceed with instrumental variable estimation and Wald-based inference. The logic is that if the first-stage is significant, then it is very unlikely that the model is unidentified. Nelson and Startz (1990b) advise that checking for the first-stage $TR^2 > 2$ is a useful diagnosis. Hall, Rudebusch and Wilcox (1996) performed an extensive Monte Carlo analysis of Wald-based inference using pre-test rules like the one suggested by Nelson and Startz (1990b) and found no clear benefits from pre-testing for the problem of choosing among a possible set of instruments. They also investigated the performance of Wald-based inference conditional on the outcome of a pre-test for instrument relevance with a given instrument set and again found no clear benefits from pre-testing. The poor performance of Wald-based inference in this context is due to the ill behaved distribution of the Wald statistic conditional on the realized value of a pre-test statistic when instruments are weak. The conditional distributions of the LM and LR statistics are better behaved and lead to more accurate inferences. To illustrate we ran a Monte Carlo experiment similar to the one reported in Section 6 of Hall, Rudebusch and Wilcox (1996). Table 7 reports actual coverage probabilities of the Wald, LM and LR 90% confidence intervals with and without regard to the outcome of a pre-test on the significance of the first stage regression. Unconditionally, the Wald, LM and LR intervals have coverage rates of .80, .90 and .89, respectively. Conditional on a significant first stage, however, these rates are .44, .80 and .79.

8. Conclusions

Our principal findings for confidence regions and inference in the presence of weak instruments and strong endogeneity are as follows:

1. Wald-based confidence regions perform poorly in the sense that they lead to the wrong conclusion. The probability they reject the null is far greater than their nominal size. They are too narrow and the probability that they cover the true parameter value is much lower than the stated level.

2. The confidence region proposed by Anderson and Rubin (1949) and confidence regions formed by inverting Lagrange multiplier (LM) and likelihood ratio (LR) statistics may be bounded intervals, empty, cover unbounded regions on the real line or cover the entire real line. These regions are unbounded when the first stage regression is not significant. While unfamiliar, the results of Dufour (1997) indicate such confidence regions are appropriate in the case of near non-identification.

3. While the AR test is defined only for the full set of structural coefficients, the LM and LR statistics can be defined for individual coefficients. Also, the AR statistic is based on an explicit joint test of restrictions on the structural parameters and of the validity of the identifying restriction whereas the LM and LR statistics only explicitly test the restrictions on the structural parameters.

4. The practice of conducting an informal pre-test based on the significance of the first-stage regression and then using the Wald statistic can be worse than not doing a pre-test. Pretesting and then using the LM or LR statistics can increase the power of these tests relative to the AR statistic without greatly distorting the sizes of the tests.

5. The poor performance of Wald-based inference can be understood in part as arising from the bias of the instrumental variable estimator, leading to an underestimate of the variance of the structural parameter. Estimating the variance of the structural errors under the null leads to more accurate inference.

The results in this paper are specific to the case of one endogenous right-hand-side variable. The extension to the multiple endogenous variable case is straightforward in principle but not in practice and is the focus of our future research.

Appendix

The AR, LR and LM confidence sets are determined by finding all values of $ß_O$ that satisfy (8), and the set will be unbounded if the coefficient a in (8) is less than zero.

Part (a): Here $a = x'[I - f_{AR} \bullet M_Z]x$ where $f_{AR} = 1 + F(k, T-k; 1-a) \bullet (k/T-k)$. Now $a < 0$ if $(x'x)/(x'M_Z x) < f_{AR}$ which can be rearranged to give the condition $F_{p=0} = \dfrac{(x'x - x'M_Z x)/k}{x'M_Z x/T-k} < F(k, T-k; 1-a)$.

Part (b): Here $a = x'[I - f_{LR} \bullet M_Z]x$ where $f_{LR} = \exp\left\{\dfrac{c^2(1;1-a)}{T}\right\} \bullet \hat{k}$. Now $a < 0$ if $(x'x)/(x'M_Z x) < f_{LR}$. After some simple manipulations, we obtain the equivalent condition $F_{p=0} = \dfrac{(x'x - x'M_Z x)/k}{x'M_Z x/T-k} < \left(\dfrac{T-k}{k}\right) \bullet \left(\exp\left\{\dfrac{c^2(1;1-a)}{T}\right\} \bullet \hat{k} - 1\right)$.

Part (c): Here $a = x'[P_{\hat{x}} - f_{LM} \bullet I]x$ where $\hat{x} = P_Z x$ and $f_{LM} = c^2(1;1-a)/T$. Then $a < 0$ if $x'P_{\hat{x}}x/x'x < f_{LM}$, which is equivalent to the condition $T \bullet R^2_{UC} = T \bullet \dfrac{x'P_Z x}{x'x} < c^2(1;1-a)$

References

Anderson, T. W. (1950), "Estimation of the Parameters of a Single Equation by the Limited-Information-Maximum-Likelihood Method," *Statistical Inference in Dynamic Economic Models* (Koopmans, T. C. ed.), New York: Wiley.

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

Angrist, J. D., and Krueger, A. B. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, CVI, 979-1014.

______, and _______ (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 87, 328-336.

Basmann, R. L. (1963), "Remarks concerning the application of exact finite sample distribution functions of GCL estimators in econometric statistical inference," *J. American Statistical Association*, 58, 943-976.

Bates, Douglas M. and Watts, Donald G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: John Wiley and Sons, Inc.

Bound, J, Jaeger, D. A. and Baker, R. M. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *J. American Statistical Association*, vol. 90, 443-450.

Bowden, R. J. and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge: Cambridge University Press.

Campbell, John Y. and Mankiw, N. Gregory (1989), "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual 1989.*

Cook, R. Dennis, and Weisberg, Sanford (1990), "Confidence Curves in Nonlinear Regression," *J. Amer. Stat. Assn.*, vol. 85, 545-551.

Davidson, Russell, and MacKinnon, James G. (1993), *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Dufour, J-M (1997), "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica,* 65, 1365-88.

Engle, R. F. (1984), Chapter 13 in *Handbook of Econometrics*, Volume II, Edited by Z. Griliches and M.D. Intriligator, Amsterdam: North Holland.

Fuhrer, Jeffrey C., Moore, George R. and Schuh, Scott D. (1995), Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood versus Generalized Method of Moments," *J. of Monetary Economics*, 35, 115-157.

Gallant, A. Ronald (1987) *Nonlinear statistical models*, New York: Wiley.

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996), "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283-289.

Hall, Robert E. (1988), "Intertemporal Substitution in Consumption," *J. of Political Economy*, 96, 339-357.

Hansen, Lars P. (1982), "Large Sample Properties of Generalized Method of Moments," *Econometrica*, 50, 1029-54.

Klepinger, D., Lundberg, S., and Plotnick, R. (1994), "*Instrument Selection: The Case of Teenage Childbearing and Women's Educational Attainment*," unpublished paper, University of Washington.

Koopmans, Tjalling C, and Hood, William C. (1953) "The Estimation of Simultaneous Linear Economic Relationships," in Hood, William C. and Koopmans, Tjalling C., *Studies in Econometric Method*, New Haven: Yale University Press.

Maddala, G. S. (1974), "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 42, 841-851.

Maddala, G. S. and Jeong, J. (1992), "On the Exact Small Sample Distribution of the IV Estimator," *Econometrica*, 60, pp. 181-83.

McClellan, M, McNeil, B. J., and Newhouse, J. P. (1994), “Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables,” *Journal of the American Medical Association*, 272, 859-866.

Meeker, William Q. and Escobar, Luis A. (1995), “Teaching About Approximate Confidence Regions Based on Maximum Likelihood Estimation,” *The American Statistician*, 49, 48-53.

Nelson, C. R. and Startz, R. (1990a), “Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator,” *Econometrica*, 58, 967-976.

_________ and _______ (1990b), “The Distribution of the Instrumental Variables Estimator and Its t-ratio when the Instrument is a Poor One,” *J. of Business*, 63, S125-S140.

Newey, W. K. and West, K. D. (1987), “Hypothesis testing with efficient method of moments estimators,” *International Economic Review*, 28, 777-87.

Phillips, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equations Model," in M.D. Intriligator & Z. Griliches (eds.) *Handbook of Econometrics*, Chapter 8 and pp. 449-516. Amsterdam: North Holland.

_________ (1989), “Partially Identified Econometric Models.” *Econometric Theory*, 5, 181-240.

Rotemberg, J. J. (1984), “Interpreting the Statistical Failures of Some Rational Expectations Macroeconomic Models,” *American Economic Review*, 74, 188-193.

Rubin, Herman (1948), “Systems of Linear Stochastic Equations,” unpublished Ph.D. dissertation, University of Chicago.

Sawa, T. (1969), “The Exact Finite Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators,” *Journal of the American Statistical Association*, 64, 923-936.

Shea, John, (1997), “Instrument Relevance in Multivariate Linear Models: A Simple Measure,” *The Review of Economics and Statistics*, Vol. LXXIX, No. 2, 348-352.

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

Venzon, D. J., and Moolgavkar, S. H. (1988), A Method for Computing Profile-Likelihood-Based Confidence Intervals,” *Applied Statistics*, 37, 87-94.

Wang, Jiahui and Zivot, Eric, (1997), “Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments,” unpublished manuscript, Dept. of Economics, University of Washington .

Table 1
Data for figures 1-3: k=1, β=1, r=0.99, s_u= s_v=1, z~N(0,1)

	p = 1	p = 0.1	p = 0.0
b_{OLS}	1.550 (0.044)	1.975 (0.015)	1.995 (0.013)
b_{2SLS}	1.148 (0.081)	1.657 (0.152)	2.064 (0.101)
π	1.161 (0.010)	0.262 (0.010)	0.162 (0.009)
$F_{p=0}$	108.1 (0.000)	5.48 (0.019)	2.092 (0.148)
TR^2	52.45 (0.000)	5.297 (0.021)	2.092 (0.148)
$C_{wald}(b\ ,\ 0.95)$	[0.990 1.310]	[1.380, 1.950]	[1.900, 2.300]
$C_{LM}(b\ ,\ 0.95)$	[0.947, 1.285]	[-0.196, 1.832]	[-∞, +∞]
$C_{LR}(b\ ,\ 0.95)$	[0.950, 1.284]	[-0.085, 1.831]	[-∞, +∞]
$C_{AR}(b\ ,\ 0.95)$	[0.950, 1.284]	[-0.090, 1.831]	[-∞, +∞]

Table 2
Data for figures 4-6: k=4, b=1, r=0.99, s_u=s_v=1, z~N(0,I_4)

	p = (1,0,0,0)′	p = (0.1,0,0,0)′	p = (0,0,0,0)′
b_{OLS}	1.538 (0.050)	1.985 (0.016)	1.995 (0.013)
b_{2SLS}	1.040 (0.101)	1.586 (0.256)	1.898 (0.128)
b_{LIML}	1.027 (0.103)	1.316 (0.530)	1.852 (0.185)
k	1.012	1.011	1.005
π_1	1.018 (0.000)	0.118 (0.000)	0.018 (0.010)
π_2	-0.030 (0.011)	-0.030 (0.011)	-0.030 (0.011)
π_3	-0.060 (0.010)	-0.060 (0.010)	-0.060 (0.010)
π_4	0.139 (0.012)	0.139 (0.012)	0.139 (0.012)
$F_{p=0}$	21.859 (0.000)	0.681 (0.606)	0.384 (0.820)
TR^2	47.927 (0.000)	2.784 (0.595)	1.592 (0.810)
$C_{wald}(b\ ,\ 0.95)$	[0.850, 1.250]	[1.100, 2.100]	[1.700, 2.100]
$C_{LM,1}(b\ ,\ 0.95)$	[0.786, 1.208]	[-∞, 1.837] [4.232, +∞]	[-∞, +∞]
$C_{LM,4}(b\ ,\ 0.95)$	[0.548, 1.287]	[-∞, 1.906] [2.394, +∞]	[-∞, +∞]
$C_{LR,1}(b\ ,\ 0.95)$	[0.769, 1.197]	[-∞, 1.744] [3.300, +∞]	[-∞, +∞]
$C_{LR,4}(b\ ,\ 0.95)$	[0.540, 1.272]	[-∞, 1.828] [2.446, +∞]	[-∞, +∞]
$C_{AR}(b\ ,\ 0.95)$	[0.588, 1.259]	[-∞, 1.816] [2.506, +∞]	[-∞, +∞]

Notes: Numbers in parentheses are standard errors for coefficient estimates and are p-values for test statistics.

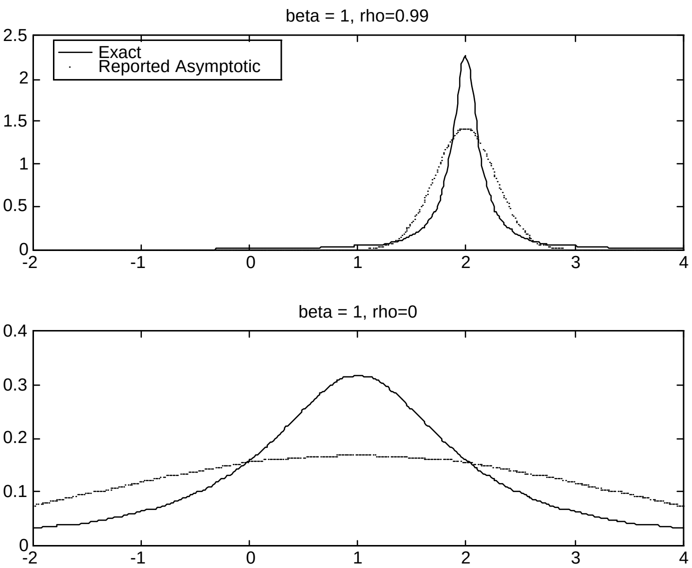
beta = 1, rho=0.99
Exact
Reported Asymptotic
2.5
2
1.5
1
0.5
0
-2
-1
0
1
2
3
4
beta = 1, rho=0
0.4
0.3
0.2
0.1
0
-2
-1
0
1
2
3
4

Figure 7