Valid Confidence Intervals and Inference in the Presence of Weak Instruments

Charles R. Nelson, Richard Startz, and Eric Zivot [*]

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

February 1996

revised October 1996

Abstract

We investigate confidence intervals and inference for the instrumental variables model with weak instruments. Wald-based confidence intervals 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 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 leads to extremely poor results when the instruments are very weak.

Comments welcomed.
e-mail: cnelson@u.washington.edu.

[*] 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. Financial support from the Royalty Research Fund of the University of Washington is gratefully acknowledged. Responsibility for errors is entirely the authors'.

1. Introduction

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 the Lagrange multiplier, likelihood ratio, and Anderson-Rubin statistics are well-behaved and easy to compute.

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), which always has the correct size, and confidence regions formed by inverting Lagrange multiplier (LM) and likelihood ratio (LR) statistics are unbounded when the first stage regression is not significant. However, while the AR test is directly obtained only for the full set of structural coefficients, the LM and LR statistics are defined for individual coefficients.

3. 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.

4. Confidence regions based on inverting LR and LM statistics have greater power than the AR confidence region under some circumstances, but the degrees of freedom must be adjusted to correct the size in overidentified models. In the good instrument case, the AR confidence regions are wider than those based on LM, LR, or Wald.

5. Non-Wald confidence regions may be empty, cover open regions on the real line or cover the entire real line. While unfamiliar, such confidence regions are appropriate in the case of near non-identification.

6. 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.

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. Nelson and Startz (1990a, b), Maddala and Jeong (1992), Hall, Rudebusch, and Wilcox (1994), Bound, Jaeger and Baker (1995), and Staiger and Stock (1994) document these phenomena. Since weak instruments abound in economic data sets, (see Angrist and Krueger (1991, 1992), Fuhrer, Moore, and Schuh (1995), Hall (1988), McClellan, McNeil, and Newhouse (1994) and Rotenberg (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.

Perhaps surprisingly, until the recent work of Staiger and Stock (SS), the Anderson-Rubin (1949) (AR) method for constructing valid confidence regions was apparently never used in practice. The AR test statistic is exactly F-distributed in finite samples (under normality) and Anderson (1950) had proposed that it be used to construct a confidence region for the set of structural coefficients. SS discuss the AR confidence regions and show that they are of four types: 1) a closed interval, 2) a disjoint region that consists of the values *outside of* a closed interval, 3) the entire real line, and 4) an empty set. Although SS do not discuss explicitly when the AR confidence regions are of a particular type, they observe that the last case is often associated with models that are misspecified; indeed the AR test is jointly a test of the value of the coefficient of the endogenous variable and of the identifying restrictions.

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 in the context of non-linear regression. The intuition is appealing: a flat likelihood will result in an appropriately wide confidence region. Dufour's (1994) results provide theoretical support for the expectation that approximate correct probability levels can be obtained in this way. Indeed, we are able to show that there is a close relationship between the AR statistic, whose distribution we know exactly, and the LR statistic.

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 (1993) has studied the multiple variable case. Hall, Rudebusch, and Wilcox (1994), however caution against choosing among instruments on the basis of their first stage significance, finding that screening worsens small sample bias. In this paper we find that decision rule to be very misleading even if there is only one available instrument and we are obliged to judge its relevance on the basis of the single sample at hand.

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 gives examples of each type. Section 5 discusses results of a Monte Carlo investigation of the actual coverage probabilities and relative power of alternative confidence regions. Section 6 investigates why the Wald statistic performs poorly.

Section 7 recomputes confidence regions from the well-known Campbell-Mankiw (1989) paper on permanent income consumption and finds the evidence shifted away from the permanent income hypothesis. Section 8 concludes the paper.

2. The Limited Information Simultaneous Equation Model and Its Likelihood Function

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 two equations are joint normal and contemporaneously correlated. Our specific results are limited to the case of a single endogenous explanatory variable. For expository purposes, we study the case where no additional exogenous explanatory variables appear in the structural equation. The model may be written as:

$$\underset{(T\times1)}{y} = \underset{(1\times1)}{\beta}\underset{(T\times1)}{x} + \underset{(T\times1)}{u} \quad (1)$$

$$\underset{(T\times1)}{x} = \underset{(T\times k)}{Z}\underset{(k\times1)}{\pi} + \underset{(T\times1)}{v} \quad (2)$$

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \text{iid } N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right] = N(0,\Sigma);\ \ i = 1,..,T \quad (3)$$

While asymptotic distribution theory requires *Z* and *u* to be asymptotically uncorrelated, where necessary we make the slightly stronger assumption that *Z* is fixed in repeated samples and that *u* and *v* are drawn independently of *Z*.

The coefficient β 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 $\pi \neq 0$ and overidentified if k>1 and the number of nonzero elements of π is greater than one.

We now review the instrumental variable and maximum likelihood approaches to estimating β. Define $\rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v}$, $\underset{(T\times 2)}{Y} = [y \quad x]$, $\underset{(T\times T)}{P_Z} = Z(Z'Z)^{-1}Z'$ for any full rank matrix Z, and $M_Z = I - P_Z$.

The two stage least squares estimator (2SLS) for β is $\hat{\beta}_{2SLS} = (x' P_Z x)^{-1} x' P_Z y$. This is also called the instrumental variable (IV) and generalized method of moments (GMM) estimator. Under the GMM framework, the 2SLS estimator solves $\min_{\beta} J_T(\beta) = T^{-1}(y - x\beta)' P_Z (y - x\beta)/\hat{\sigma}^2$, where $\hat{\sigma}^2 = T^{-1}(y - x\hat{\beta}_{2SLS})'(y - x\hat{\beta}_{2SLS})$. Under standard regularity conditions $\sqrt{T}\left(\hat{\beta}_{2SLS} - \beta\right) \longrightarrow N\left(0, \sigma_u^2 (\pi' M_{ZZ} \pi)^{-1}\right)$, where $M_{ZZ} = \operatorname*{plim}_{T\to\infty} T^{-1} Z'Z > 0$.

The maximum likelihood estimator for this model 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 β is given by (cf. Davidson and MacKinnon 1993, p. 647)

$$ {}^{c}(\beta) = -T\ln(2\pi) - \frac{T}{2}\ln(\kappa(\beta)) - \frac{T}{2}\ln|Y'M_Z Y| \tag{4} $$

where

$$ \kappa(\beta) = \frac{(y - \beta x)'(y - \beta x)}{(y - \beta x)' M_Z (y - \beta x)} \tag{5} $$

The LIML estimator of β is obtained by minimizing $\kappa(\beta)$, a result first demonstrated by Rubin(1948); see also Koopmans and Hood(1953). Thus $\hat{\beta}_{LIML} = \operatorname*{argmin}_{\beta} \kappa(\beta)$. Operationally, $\kappa(\hat{\beta}_{LIML}) = \hat{\kappa}$ is the smallest eigenvalue of the matrix $(Y' M_Z Y)^{-1/2} Y'Y (Y' M_Z Y)^{-1/2}$ and $\hat{\beta}_{LIML}$ is given by the k-class estimator formula

$$\hat{\beta}_{\text{LIML}} = \left(x'(I - \hat{\kappa} M_z)x\right)^{-1}\left(x'(I - \hat{\kappa} M_z)y\right). \quad (6)$$

Notice that when $\hat{\kappa} = 1$, which is true in a just identified model, we have $I - \hat{\kappa} M_z = I - M_z = P_z$ and thus $\hat{\beta}_{\text{LIML}} = \hat{\beta}_{2\text{SLS}} = (x' P_z x)^{-1} x' P_z y$. Under standard regularity conditions $\sqrt{T}\left(\hat{\beta}_{\text{LIML}} - \beta\right) \xrightarrow{d} N\left(0, \sigma_u^2 (\pi' M_{zz} \pi)^{-1}\right)$. Thus, note that the asymptotic distribution of $\sqrt{T}\left(\hat{\beta}_{\text{LIML}} - \beta\right)$ is the same as the asymptotic distribution of $\sqrt{T}\left(\hat{\beta}_{2\text{SLS}} - \beta\right)$.

3. Construction of Confidence Sets by Inverting Test Statistics

We are interested in constructing confidence sets for the structural parameter β in (1). Given a test statistic $\psi(\beta_0)$ for the testing the hypothesis $H_0 : \beta = \beta_0$ at the α significance level, the $(1 - \alpha) \cdot 100\%$ confidence set associated with this statistic is defined as

$$C_\psi(\beta; 1 - \alpha) = \left\{ \beta_0 : \psi(\beta_0) \leq cv_{1-\alpha} \right\}$$

where $cv_{1-\alpha}$ is the $1 - \alpha$ quantile from the (asymptotically valid) distribution of the test statistic $\psi(\beta_0)$; i.e., C_ψ contains all of the "acceptable" values of β_0 at level α for the null hypothesis $H_0 : \beta = \beta_0$ using the test statistic $\psi(\beta_0)$. Confidence sets formed this way are said to be determined by "inverting" the test statistic $\psi(\beta_0)$.

We are interested in confidence regions corresponding to seven test statistics: the Wald, LM, and LR statistics based on maximum likelihood estimation, the three analogous statistics based on the GMM framework, and the Anderson-Rubin statistic. Due to the simple form of the hypothesis test there is considerable redundancy among the seven. In fact, there are only two versions of the GMM based statistics and one of these is identical to the version of the MLE LM test we employ. In the just-identified case, the MLE and GMM Wald statistics are the same. Therefore, we need to consider four, or at most five, different ways to compute confidence regions.

The Wald, LM (see Engle (1984)), and LR statistics are given respectively by

$$\text{Wald}\left(\beta_0\right) = \frac{\left(\hat{\beta} - \beta_0\right)^2}{\text{EAVAR}\left(\hat{\beta}\right)} \tag{7}$$

$$\text{LM}\left(\beta_0\right) = \frac{T^{-1} g\left(\beta_0\right)^2}{\text{EAVAR}\left(\beta_0\right)} \tag{8}$$

$$\text{LR}\left(\beta_0\right) = -2\left[{}^{c}\left(\beta_0\right) - {}^{c}\left(\hat{\beta}_{\text{LIML}}\right)\right] \tag{9}$$

where $\hat{\beta}$ denotes a consistent estimate of β, $\text{EAVAR}(\beta)$ denotes an estimate of the asymptotic variance of $\hat{\beta}_{\text{LIML}}$ evaluated at β, and $g(\beta) = \frac{d}{d\beta}\, {}^{c}(\beta)$ is the gradient of the concentrated log likelihood for β. Under standard assumptions, the three statistics are asymptotically $\chi^2(1)$.

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

$$\text{Wald}_{GMM}\left(\beta_0\right) = \frac{\left(\hat{\beta}_{2SLS} - \beta_0\right)' x' \text{P}_Z\, x\left(\hat{\beta}_{2SLS} - \beta_0\right)}{\hat{\sigma}^2}$$

$$\text{LM}_{GMM}\left(\beta_0\right) = \frac{\left(y - x\beta_0\right)' \text{P}_{\hat{x}}\left(y - x\beta_0\right)}{\hat{\sigma}^2}$$

$$\text{LR}_{GMM}\left(\beta_0\right) = \frac{\left[\left(y - x\beta_0\right)' \text{P}_Z\left(y - x\beta_0\right) - \left(y - x\hat{\beta}_{2SLS}\right)' \text{P}_Z\left(y - x\hat{\beta}_{2SLS}\right)\right]}{\hat{\sigma}^2}$$

where $\hat{\sigma}^2 \to \sigma_u^2$.

Several facts are worth noting. First, because of the quadratic nature of the GMM minimization problem and the linearity of the restriction $\beta = \beta_0$, the three GMM statistics are numerically identical so long as the same estimate is used for $\hat{\sigma}^2$. (See appendix, Proposition 1 and also Newey and West (1987).) Second, when using $\hat{\sigma}^2 = \hat{\sigma}^2_{\text{2SLS}} = T^{-1}\left(y - x\hat{\beta}_{\text{2SLS}}\right)'\left(y - x\hat{\beta}_{\text{2SLS}}\right)$, the Wald_{GMM} is simply the (square of) "asymptotic *t*," which is the statistic used essentially always for inference in applied work. Further, the LR_{GMM} is the statistic calculated in the Hansen (1982) framework as the "difference in the J-statistic." Third, use of $\hat{\sigma}^2 = \sigma_0^2 = T^{-1}(y - x\beta_0)'(y - x\beta_0)$ instead of $\hat{\sigma}^2_{2\text{SLS}}$ (σ_0^2 is the natural choice when thinking of using LM_{GMM}) is shown below to make a critical difference in inference.

Other than the choice of $\hat{\sigma}^2$, the GMM statistics are straightforward. In contrast, evaluation of the MLE statistics requires a choice of EAVAR. A variety of specifications are available. For the MLE-Wald statistic, one usually sees the k-class formula which is (7) with $\hat{\beta} = \hat{\beta}_{LIML}$, $\text{EAVAR} = \hat{\sigma}^2_{\text{LIML}} \left(x'(I - \hat{\kappa} M_z)x\right)^{-1}$ and $\hat{\sigma}^2_{\text{LIML}} = T^{-1}\left(y - x\hat{\beta}_{\text{LIML}}\right)'\left(y - x\hat{\beta}_{\text{LIML}}\right)$ which we refer to as $\text{Wald}_{\text{k-class}}$.

For the MLE-LM statistic, it is useful to base EAVAR on the information matrix (see Bowden and Turkington (1984)).

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

$$\text{A}(\beta) = T\left[\frac{x'M_z x}{(y - x\beta)'M_z(y - x\beta)} - \frac{x'x}{(y - x\beta)'(y - x\beta)}\right]$$

$$\text{B}(\beta) = 2T\left[\frac{\left[x'(y - x\beta)\right]^2}{\left[(y - x\beta)'(y - x\beta)\right]^2} - \frac{\left[x'M_z(y - x\beta)\right]^2}{\left[(y - x\beta)M_z'(y - x\beta)\right]^2}\right].$$

The LM statistic can then be written as

$$\text{LM}(\beta_0) = T^{-1}\left\{T\left[\frac{x'u_0}{u_0'u_0} - \frac{x'M_z u_0}{u_0'M_z u_0}\right]\right\}'\left\{\text{A}(\beta_0) + \text{B}(\beta_0)\right\}^{-1}\left\{T\left[\frac{x'u_0}{u_0'u_0} - \frac{x'M_z u_0}{u_0'M_z u_0}\right]\right\} \quad (10)$$

where $u_0 = y - x\beta_0$. The LM statistic as given in (10) is not easily written as a quadratic in β_0. However, using the following approximation results:

$$T\left[\frac{x'u_0}{u_0'u_0} - \frac{x'M_Z u_0}{u_0'M_Z u_0}\right] \overset{A}{\approx} T\,\frac{\hat{x}'u_0}{u_0'u_0} \quad \text{and} \quad A(\beta_0) + B(\beta_0) \overset{A}{\approx} \frac{u_0' P_{\hat{x}} u_0}{u_0'u_0}$$

we obtain a simpler version of the LM statistic

$$\mathrm{LM}_{LIML} = \frac{(y - x\beta_0)' P_{\hat{x}} (y - x\beta_0)}{(y - x\beta_0)'(y - x\beta_0)/T} \tag{11}$$

where $\hat{x} = P_Z x$. Here LM_{LIML} is equal to T times the uncentered R^2 from the regression of $y - x\beta_0$ on $\hat{x}$. We note that this approximation will make LM_{LIML} identical to the corresponding LM statistic in the GMM framework.

Finally, consider the statistic proposed by Anderson and Rubin (1949) and Anderson (1950). Rewrite (1) by adding and subtracting $x\beta_0$ from both sides and substituting in for x using equation (2) to give

$$y^* = Z\psi + u^* \tag{12}$$

where $y^* = y - x\beta_0$, $\psi = \pi(\beta - \beta_0)$, and $u^* = v(\beta - \beta_0) + u$. Then the hypothesis $H_0{:}\ \beta = \beta_0$ in (1) corresponds to the hypothesis $H_0{:}\ \psi = 0$ in (12). The latter hypothesis can be tested with the standard F-statistic

$$\mathrm{AR} = F_{\psi=0} = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(T-k)} = \frac{(y - x\beta_0)' P_Z (y - x\beta_0)/k}{(y - x\beta_0)' M_Z (y - x\beta_0)/(T-k)}$$

If indeed $(u_i\ v_i)$ is distributed iid $N(0,\Sigma)$, 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 $F(k, T-k)$.[1]

Creating confidence regions by "inverting" the corresponding test statistics means to solve for the range of values for β_0 for which the test statistic is less than the appropriate

[1] Given the linear structure of (12), the AR statistic is a monotone transformation of the LR statistic for testing the hypothesis $H_0: \psi = 0$. SS show that $AR \to \chi^2(k)/k$ under fairly general assumptions about the disturbances.

critical value. The results are most easily seen graphically. Figure 2 shows the value of the test statistics for a particular Monte Carlo run for a just identified model with a fairly 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. For the u-shaped Wald statistic, this region is always a closed set. For the other statistics, three additional patterns are possible. These can be seen varying the confidence level in Figure 2; that is, by raising or lowering the horizontal critical value line. 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. In contrast to the familiar Wald-based confidence regions, at high confidence levels no values of the parameters can be ruled out. Consider next what happens at a somewhat lower confidence level. The horizontal line will "cut-off" the peak of the test statistic. 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, the AR statistic in overidentified models at low confidence levels can lead to an empty confidence region.

Fortunately, for each statistic we can give simple closed form solutions for the cut-points. For the Wald statistics, both MLE and GMM, we have the completely familiar closed intervals

$$\hat{\beta} \pm z(\alpha / 2) \cdot \text{EAVAR}\left(\hat{\beta}\right)^{\frac{1}{2}},$$

where $z(\alpha/2)$ is the $\alpha/2$ quantile of the N(0,1) distribution.

The confidence sets for β formed by inverting the LM, LR, and AR statistics are each determined by solutions to an inequality of the form

$$a\beta_0{}^2 + b\beta_0 + c \leq 0 \tag{13}$$

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

[2] The figure is drawn for data generated with $\beta = 1, \pi = 0.1 \quad \sigma_u^2 = \sigma_v^2 = 1, \quad \rho = .99, \quad T = 100, \quad Z \sim N(0,1)$.

$$\beta_{0,i} = \frac{-b \pm \sqrt{d}}{2a}; \; i = 1,2 \tag{14}$$

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.

Finding the confidence set $C_{LM}(\beta;1-\alpha)$ requires finding all values of β_0 satisfying the condition

$$\frac{(y - x\beta_0)' P_{\tilde{x}} (y - x\beta_0)}{(y - x\beta_0)'(y - x\beta_0)} \leq \frac{\chi^2_{1-\alpha}(1)}{T} \equiv \phi_{LM}$$

which can be rearranged as a quadratic of the form of (13). Defining the 2x2 matrix $Q_{LM} = Y'\left(\left[1 - \phi_{LM}\right]I - \left[\left(I - P_{\tilde{x}}\right)\right]\right)Y$ then $a = Q_{22}$, $b = -2 \cdot Q_{12}$, 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 $H_0: \beta = \beta_0$ is accepted if

$$T \cdot \ln\left[\frac{(y - x\beta_0)' M_Z (y - x\beta_0)}{(y - x\beta_0)'(y - x\beta_0)}\right] - T \cdot \ln(\hat{\kappa}) \leq \chi^2_{1-\alpha}(1)$$

which can be rewritten as

$$\frac{(y - x\beta_0)'(y - x\beta_0)}{(y - x\beta_0)' M_Z (y - x\beta_0)} \leq \exp\left\{\frac{\chi^2_{1-\alpha}}{T}\right\} \cdot \hat{\kappa} \equiv \phi_{LR}$$

[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.

This inequality can also be expressed in the form of (13). The corresponding matrix Q is given by $Q_{LR} = Y'\left[I - \phi_{LR} \cdot M_Z\right]Y$.

Finally, the AR confidence set consists of all values of β_0 that satisfy the inequality

$$\frac{(y - x\beta_0)'(y - x\beta_0)}{(y - x\beta_0)' M_z (y - x\beta_0)} \leq 1 + \mathrm{F}_{1-\alpha}(k, T-k)\frac{k}{T-k} \equiv \phi_{AR}$$

where $\mathrm{F}_{1-\alpha}(k, T-k)$ denotes the (1-α) 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 (13) with the corresponding matrix Q given by $Q_{AR} = Y'\left[I - \phi_{AR} \cdot M_Z\right]Y$.

Each of the LM, LR, and AR statistics is a ratio of two quadratic forms. Such functions have the characteristic shape that was seen in Figure 2. Recall that when the critical value line is everywhere above the function, the confidence region is the entire real line. When the critical value is lower such that part of the real line is excluded, a disjoint confidence region results. When the critical value line lies below the asymptote of the statistic, then the confidence region is the familiar closed interval. Finally, in cases where the function does not touch zero and lies above the low critical value, an empty confidence region results. Whether the confidence region is bounded, empty, external, or covers the real line is determined by the signs of a and d as follows.

If a>0, then the inequality may be rewritten as $\beta_0{}^2 + (b/a)\bullet\beta_0 + (c/a) \leq 0$ 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 $(\beta_{LOW}, \beta_{HIGH})$, within which the inequality is satisfied. Alternatively, if a>0 but d<0, the roots are complex so there is no value of β_0 which satisfies the inequality and thus the confidence set is empty.

It is straightforward to show that LM and LR confidence sets cannot be empty because at $\beta_0 = \hat{\beta}$ the statistics equal zero. When $\beta_0 = \hat{\beta}$, the AR statistic tests the significance by regressing the residuals on the instruments. In the just identified case the statistic is zero so the AR confidence set cannot be empty. The AR confidence set 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 $\beta_0{}^2 + (b/a) \bullet \beta_0 + (c/a) \geq 0$ which is again convex from below. If d>0 there are again real solutions to (13), but now it is values of β_0 *outside* the interval $(\beta_{LOW}, \beta_{HIGH})$ which satisfy the inequality (because the sign of the inequality is reversed relative to the previous case) so the confidence set is the disconnected region $[-\infty, \beta_{LOW}]$ $[\beta_{HIGH}, \infty]$. 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 β_0.

The preceding discussion demonstrates that the confidence sets for β constructed by inverting the LM, LR, or AR test statistics can be unbounded. Indeed, Dufour (1994) shows that any valid (1-α)•100% confidence set for β must be unbounded with probability 1-α for nearly nonidentified models. We can show that the AR statistic has this property — and we know, of course, that the nominal size of the AR statistic is the exact size, since the AR is just a statistic from an auxiliary regression. 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.

Unbounded confidence sets occur when the coefficient a in (13) is less than zero. Consequently, the probability of an unbounded confidence set is P{a<0}. Following from that are relationships linking the unboundedness of LM, LR, and AR confidence sets with the usual goodness-of-fit statistics from the first stage regression. (Note that for each of the three statistics, *a* is a weighted difference between the actual and residual first-stage sum of squares.) These relationships can be summarized in the following:

(a) $C_{AR}(\beta;1-\alpha)$ is unbounded if $F_{\pi=0} < F_{1-\alpha}(k, T-k)$

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

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

(c) $C_{LM}(\beta;1-\alpha)$ is unbounded if $T \cdot R^2_{UC} < \chi^2_{1-\alpha}(1)$

where $F_{\pi=0}$ is the F statistic for testing $\pi=0$ in (2) and R^2_{UC} is the uncentered (no intercept) R^2 from (2). See appendix (proposition 2) for proofs.

Thus the AR confidence set, $C_{AR}(\beta;1-\alpha)$, has the very interesting property that it is unbounded whenever the F statistic for testing $\pi=0$ in (2) is insignificant at level α. If (u_i, v_i) are i.i.d. $N(0,\Sigma)$ then the probability that $C_{AR}(\beta;1-\alpha)$ is unbounded when $\pi=0$ is given by $P\{F_{\pi=0} < F_{1-\alpha}(k,T-k)\} = 1-\alpha$. Hence, the AR confidence set satisfies Dufour's condition for a valid confidence set in an unidentified model.

The condition for $C_{LR}(\beta;1\text{-}\alpha)$ 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 $\exp\{T^{-1} \cdot \chi^2_{1-\alpha}(1)\} \approx 1 + T^{-1} \cdot \chi^2_{1-\alpha}(1)$ so that (b) above becomes $F_{\pi=0} < k^{-1} \cdot \chi^2_{1-\alpha}(1)$. Notice that this condition is similar to the condition in (a) above for the AR confidence set since, for large T, $F_{1-a}(k,T-k) \approx k^{-1} \cdot \chi^2_{1-\alpha}(k)$. However, the condition for the LR confidence set uses $\chi^2_{1-\alpha}(1)$ whereas the condition for the AR set uses $\chi^2_{1-\alpha}(k)$. Since $\chi^2_{1-\alpha}(k) > \chi^2_{1-\alpha}(1)$ it follows that when $\pi=0$, P{$C_{LR}(\beta;1-\alpha)$ is unbounded} = P{$F_{\pi=0} < k^{-1} \cdot \chi^2_{1-\alpha}(1)$} < P{$F_{\pi=0} < k^{-1} \cdot \chi^2_{1-\alpha}(k)$} = P{$C_{AR}(\beta;1-\alpha)$ is unbounded} = $1-\alpha$. Hence, in the unidentified case, $C_{LR}(\beta;1-\alpha)$ is unbounded with probability less than $1-\alpha$ and so is not a valid confidence set according to the results of Dufour.

A similar result holds for the LM confidence set. The statistic $T \cdot R^2_{UC}$ is the Lagrange multiplier statistic for testing $H_0: \pi=0$ in (2), and under this null $T \cdot R^2_{UC} \overset{A}{\sim} \chi^2_{1-\alpha}(k)$. Hence, when $\pi=0$, P{$C_{LM}(\beta;1-\alpha)$ is unbounded} = P{$T \cdot R^2_{UC} < \chi^2_{1-\alpha}(1)$} < P{$T \cdot R^2_{UC} < \chi^2_{1-\alpha}(k)$} = $1-\alpha$, implying that $C_{LM}(\beta;1-\alpha)$ is not a valid confidence set according to Dufour's result.

The above remarks regarding $C_{LR}(\beta;1-\alpha)$ and $C_{LM}(\beta;1-\alpha)$ suggest that in the very weak instrument case the LR and LM statistics for testing $H_0:\beta=\beta_0$ are not asymptotically distributed as $\chi^2(1)$ and that a better approximation to limiting distributions is given by $\chi^2(k)$, which the results of Dufour suggest as a bounding distribution for the statistics. (Dufour suggests a bound for the LR statistic based on a transformation of the distribution of the Wilks Λ statistic. Wang and Zivot (1996) show that the $\chi^2(k)$ bound is tighter.) To see why, if the critical value from $\chi^2(k)$ is used to compute $C_{LR}(\beta;1-\alpha)$ and $C_{LM}(\beta;1-\alpha)$ then (for large T relative to k) P{$C_{LR}(\beta;1-\alpha)$ is unbounded} = P{$F_{\pi=0} < k^{-1} \chi^2_{1-\alpha}(k)$} = P{$C_{LM}(\beta;1-\alpha)$ is unbounded} = P{$T R^2_{UC} < \chi^2_{1-\alpha}(k)$} = 1-$\alpha$, whenever $\pi=0$, and therefore $C_{LR}(\beta;1-\alpha)$ and $C_{LM}(\beta;1-\alpha)$ are valid confidence sets.

Thus, the statistical significance of goodness-of-fit statistics from the first stage regression, $F_{\pi=0}$ and $T R^2_{UC}$, has implications for the construction of valid confidence sets obtained by inverting the LR and LM statistics. If $F_{\pi=0} < F_{1-a}(k, T-k)$, or $F_{\pi=0} < k^{-1} \chi^2_{1-\alpha}(k)$, then the LR statistic should be inverted using critical values from $\chi^2(k)$ instead of $\chi^2(1)$. Similarly, if $T R^2_{UC} < \chi^2_{1-\alpha}(k)$ then the LM statistic should be inverted using $\chi^2(k)$ instead of $\chi^2(1)$ critical values. We call the test statistics which switch degrees of freedom based on the first stage statistic LM_{sw} and LR_{sw}.

In light of the above results, it appears that the asymptotic distributions of the ML and GMM test statistics for $H_0: \beta = \beta_0$ in the weak instrument case are poorly approximated by the $\chi^2(1)$ distribution. The local-to-zero framework of SS provides a convenient way to obtain analytical results in the weak instrument case. In this framework, the coefficients π in (2) are modeled as being in a $T^{-1/2}$ neighborhood of zero. This device keeps the statistic $F_{\pi=0}$ roughly constant as the sample size increases. Wang and Zivot (1996) (hereafter, WZ) derive the asymptotic distributions of the Wald, LM, and LR statistics using SS's local-to-zero framework. WZ show that in overidentified models these distributions do not converge to a $\chi^2(1)$ random variable but rather to random variables that depend on the nuisance parameters ρ and k and the noncentrality parameter of the asymptotic distribution of $F_{\pi=0}$. In addition, WZ show that the asymptotic distributions of LM and LR are bounded by the $\chi^2(k)$ distribution, whereas the asymptotic distribution of the Wald statistic is not. Further, in just identified models, WZ show that the LM and LR statistics converge in distribution to the same $\chi^2(1)$ random variable and the AR converges to a $\chi^2(1)$ that is independent of the LM and LR.

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. Later, Bound, Jaeger, and Baker (1995) advocate checking for first-stage significance with a standard F-test. These pre-test rules fail for two reasons. First, there is a flaw in logic in assuming that because a model is identified, asymptotic distribution theory gives a good guide to small sample distributions. Distributions from a data generating process with π very small, but not equal to zero, look pretty much like distributions when π does equal zero. It happens to be true that as the first-stage F rises the asymptotic distribution becomes a good approximation to the true distribution. However, an α significance level in the first-stage does not imply accuracy of structural inference at the α significance level - a much higher level is needed at the first-stage. Second, in a weakly identified DGP, a significant first-stage generally signals a spuriously good fit between the endogenous variable and instrument and this is precisely the case when instrumental variable estimation is worst. It is this phenomenon which led Hall, Rudebusch and Wilcox (1994) to recommend against screening potential instruments.

To illustrate, we ran 10,000 Monte Carlo trials with the parameters given above and with both an unidentified, $\pi = 0$, and a weakly identified, $\pi = 0.1$, model. In Table 1 we report empirical sizes for a nominal five percent Wald test both with and without a five percent first-stage pretest.

Note two facts from the table. First, in the unidentified model the result of pretesting is to draw no conclusion approximately 95 percent of the time and to be *always* wrong when one does draw a conclusion. Second, in the weakly identified model, using the pretest still leaves one wrong 87 percent of the time![5]

4. Examples of Confidence Sets for β

To illustrate the typical shapes of confidence sets for instruments of various quality we generated data from (1)-(3) with $\beta = 1, \sigma_u^2 = \sigma_v^2 = 1, \rho = .99, T = 100,\ Z \sim N(0, I_k)$ for just identified (k=1) and nominally overidentified (k=4) models. For the just identified model, we set π=1(good instrument case), π=0.1(weak instrument case) and π=0(unidentified case). For the overidentified model, we set π=(1,0,0,0)'(good instrument case), π=(0.1,0,0,0)'(weak instrument case) and π=(0,0,0,0)'(unidentified case). For each set of generated data we computed the OLS, 2SLS and LIML estimates of β, the reduced form estimate of π, the reduced form R^2_{UC} and $F_{\pi=0}$ statistics and the confidence sets $C_{Wald}(\beta;0.95)$, $C_{LR}(\beta;0.95)$, $C_{LM}(\beta;0.95)$ and $C_{AR}(\beta;0.95)$. These statistics are summarized in Tables 2 and 3 and the confidence sets are displayed graphically in Figures 1 - 6.

Consider first the results for the just identified models. Figure 1 shows the confidence sets for the good instrument case. The OLS estimate of β is biased and has a very small standard error. The reduced form statistics indicate that Z is a good instrument. Indeed, the 2SLS and LIML estimates of β (identical when k=1) are equal to 1.148 and the standard Wald confidence region $C_{Wald}(\beta;0.95)$, [0.989, 1.307], is fairly small and contains the true value $\beta = 1$. 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. Here $\hat{\beta}_{OLS}$ = 1.975 which is very close to the theoretical point of concentration in an unidentified model (Basmann, 1963 and Phillips, 1983). The reduced form regression statistics indicate that Z is a questionable instrument, although the value of $F_{\pi=0}$ is large enough to reject the hypothesis

[5] Do not note the fact that reversing the use of the pre-test would have led to the correct size in the second column. While amusing, it is only a coincidence.

that $\pi=0$ at the 5% level. However, the weak instrument Z induces a noticeable bias in $\hat{\beta}_{2SLS}$ and, counter to intuition, the Wald confidence interval is fairly short and does not cover the true value $\beta=1$. Since $F_{\pi=0}$ is 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. The length of these intervals reflects much more uncertainty about the value of β than does the length of the Wald interval.

Finally, in the unidentified case of Figure 3, the OLS and 2SLS estimates of β 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 sets in this case are equivalent and contain all possible values of β. This is what we should expect when β is unidentified since the likelihood function is flat. This flatness of the likelihood function is seen very clearly in Figure 3.

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 statistical results are summarized in Table 3 and the test statistics, as functions of β_0, are illustrated in Figures 4-6. For the good instrument case, $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{LIML}$ are very close to the value $\beta=0$ and $C_{\text{Wald}}(\beta, 0.95)$ is quite tight. The reduced form goodness-of-fit statistics, $F_{\pi=0}$ and $T\ \ R^2_{UC}$, are large and indicate that the instruments are of good quality. The LM and LR confidence sets based on $\chi^2(1)$ 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. We note that C_{LR} and C_{LM} based on $\chi^2(k)$ critical values are very close to C_{AR}.

Turning next to the case of one valid, but weak instrument, we see that $\hat{\beta}_{OLS}$ is quite biased and that $\hat{\beta}_{LIML}$ is less biased than $\hat{\beta}_{2SLS}$. C_{Wald} is fairly wide, but does not cover $\beta=1$. Here the reduced form statistics $F_{\pi=0}$ and $T\ \ R^2_{UC}$ are not significant at the 5% level, which raises a red flag indicating that the instruments are poor and β is nearly unidentified. From the previous section we know that C_{AR} will be unbounded and indeed C_{AR} is the disjoint region $[-\infty, 1.816]$ $[2.506, \infty]$. Notice that C_{LM} and C_{LR} based on $\chi^2_{1-\alpha}(k)$ are very close to C_{AR} whereas these sets based on $\chi^2_{1-\alpha}(1)$ have larger right endpoints and are thus "smaller" unbounded regions.

Finally, in the nonidentified case $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{LIML}$ are close to $\hat{\beta}_{OLS}$. C_{Wald} is short and does not cover $\beta = 1$. 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 β.

5. A Monte Carlo Investigation of Size and Power

In this section we analyze the finite sample properties of the 95% confidence regions for β formed by inverting the level 0.05 Wald, LM, LR and AR test statistics for $H_0: \beta = \beta_0$. 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 $H_a: \beta = \beta_a$. Our Monte Carlo design is the same as in section 5 except that we consider $\rho = \{0.99, 0.5, 0\}$. For the power analysis we generate data under the alternatives $\beta_a = \beta_0 + \delta_i$ where δ_i ranges from -2 to 2 in increments of 0.25. The empirical probabilities of the confidence sets under the null are summarized in tables 4 - 9 and results on power are given in Tables 10-15.[6]

Consider first the size results for the just identified models. Since, as shown in Section 3, the LM, LR, and AR statistics are approximately $\chi^2(1)$ regardless of the values of ρ and π, 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 the distribution of the Wald statistic depends on ρ and π in the weak instrument case. In the unidentified case $(\pi = 0)$, C_{Wald} covers the true value $\beta_0 = 1$ less than 37% of the time when $\rho = 0.99$ and 100% of the time when $\rho = 0$. The sets C_{LM}, C_{LR} and C_{AR} are unbounded with

[6] Note that these tables give power, not size-adjusted power.

frequency 0.95, as they should be in an unidentified model, and the set $[-\infty,\infty]$ occurs roughly 85% of the time. The results for the weak instrument case $(\pi = 0.1)$ are similar to the unidentified case. The size distortion of the Wald statistic is not as severe; the sets C_{LM} , C_{LR} and C_{AR} have correct coverage frequencies and are unbounded only 84% of the time. In the good instrument case $(\pi = 1)$, 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 and C_{Wald} is slightly shorter.

Next consider the size results for the nominally over identified, k=4, model. In the unidentified and weak instrument cases, C_{Wald} has actual coverage frequencies much smaller than .95 when $\rho = 0.99$ or 0.5. For example, in the unidentified case with $\rho = 0.99$ the actual frequency is only 1.3%. By contrast, C_{AR} has the correct coverage frequency in all cases and is unbounded with frequencies .95 and .90 in the unidentified and weak instrument cases respectively. In the good instrument case, C_{AR} is always bounded but is about 50% larger, on average, than C_{Wald} . 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 $\chi^2_{.95}(1)$ have actual coverage frequencies less than .95 for all values of ρ, although C_{LM} has nearly the correct frequency when $\rho = 0$. The unbounded confidence sets occur less frequently than 95% in the unidentified case. The sets C_{LM} and C_{LR} based on $\chi^2_{.95}(4)$ have actual frequencies of at least .95 in all cases and this supports the use of $\chi^2(4)$ as a bounding distribution for the LM and LR statistics. These confidence sets are very close to C_{AR} but appear to be slightly larger than C_{AR} in the good instrument case.

In the good instrument case, however, the actual coverage frequencies of C_{LM} and C_{LR} are very close to 1. The sets C_{LM} and C_{LR} based on $\chi^2_{.95}(1)$ or $\chi^2_{.95}(4)$ (based on the significance of $F_{\pi=0}$) perform much better than the sets based solely on $\chi^2_{.95}(4)$. They have approximately 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 $\chi^2_{.95}(1)$ and C_{Wald} .

To summarize our results on empirical size, the Wald confidence intervals are very misleading when there is a poor instrument and strong endogeneity. In the just identified case, LM, AR, and LR statistics all perform well. In the overidentified case, degrees of freedom-switched versions LM_{sw} and LR_{sw} perform well. The AR confidence region always has the correct size, but in the overidentified case is somewhat wider than LM_{sw} and LR_{sw}.

Now consider the issue of power. We present only results for the k=1 model since those for the k=4 case are similar. Regardless of instrument quality, the powers of the LM, LR, and AR statistics are very similar. In the good instrument case, they are also nearly identical to the power of the Wald test, and all four converge to unity at $\delta = 0.5$. For the weak instrument case, the power curves vary considerably depending on the value of ρ. For $\rho = 0$ and $\rho = 0.5$, the power of LM, LR, and AR is roughly symmetric about ß=1 and are fairly flat over the range of β . When $\rho = 0.99$, their power is relatively flat at about 5% except for a spike at $\beta_a = 0$ due to the fact that there is a local minimum in the likelihood function near $\beta = 1$ (see Figure 6) and a global maximum near $\beta = 0$, making the LR statistic for testing the null hypothesis $\beta = 1$ very large. 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 the presence of good instruments.

The power of the Wald statistic is roughly U shaped in the weak instrument case, and the location of minimum power is influenced by the value of ρ, reflecting the concentration phenomenon. In the unidentified case, the power of LM, LR, and AR is flat at 5% for all β whereas the power of the Wald is rather sharply U-shaped and strongly influenced by the value of ρ. In particular, the one case in which the Wald confidence region is notably better than the others is when $\rho = 0$, that is when there is no endogeneity.[7]

[7] Of course, in this case one can do even better by doing least squares instead.

6. 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 $N\left(\hat{\beta}_{2SLS}, \hat{\sigma}^2 \left(x'P_z x\right)^{-1}\right)$ does a poor job approximating the true sampling distribution. Curiously, it's just the other way around. The reported distribution fairly accurate represents the sampling distribution, but with weak instruments and significant endogeneity the sampling distribution isn't located particularly near the true parameter. (See Phillips (1989).) 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.[8] The difference is apparent, but only partially real. Figure 7 shows the same plot magnified by truncating the horizontal scale. The apparently flat likelihood function actually has a very sharp peak around $\hat{\beta}_{2SLS}$. 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 sampling distribution is approximated by $N\left(\hat{\beta}_{2SLS}, \hat{\sigma}^2 \left(x'P_z x\right)^{-1}\right)$. There are both series and closed form expressions for the density of $\hat{\beta}$ in quite general situations. (See Sawa (1969) and Phillips (1983).) These expressions do not lend themselves to easy interpretation. However, Phillips (1989) and Staiger and Stock give the following expression for the exact distribution $\hat{\beta}_{2SLS}$ in the completely unidentified case of (1)-(3).[9]

$$\hat{\beta}_{2SLS} \equiv \beta + \theta + \frac{\eta}{\sqrt{k}} \, t_k \qquad (15)$$

---

[8] If the LR is a little hard to see, look underneath the LM line.

[9] The general expression for the just identified case is given in the appendix.

where $\equiv$ denotes equivalence in distribution, $\theta = \rho\sigma_u/\sigma_v$, $\eta = (1-\rho^2)^{0.5}\sigma_u/\sigma_v$, and t_k denotes a Student-t random variable with k degrees of freedom.

Figure 8 shows the both the exact distribution and the normal approximation evaluated at the median values of $\hat{\beta}_{2SLS}$ and its associated asymptotic standard error from two unidentified models from the Monte Carlo experiments shown in Table 3.

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 ρ=0 there is no endogeneity and the distribution is approximately median unbiased, which is consistent with the result reported in Nelson and Startz (1990a).

Consider the density in (15) as $\rho \to 1$, the worst possible case. Here $\theta \to \sigma_u/\sigma_v$ and $\eta \to 0$ so that the density is zero except for a spike at $\hat{\beta}_{2SLS} = \beta + \sigma_u/\sigma_v$.[10] Given that the estimator collapses to the point of concentration, one can write the instrumental variable residuals as $y - \hat{\beta}_{2SLS}x = y - (\beta + \sigma_u/\sigma_v)x = u - \sigma_u/\sigma_v\, x$. But in this case $x = v$ and $u = \sigma_u/\sigma_v\, v$, so the residuals collapse to zero. Since $\hat{\sigma}^2$ is just the mean sum squared residuals, it too collapses to zero. Thus the Wald confidence intervals, based on $\hat{\sigma}^2$, are far too small. In contrast, the LM, based on $\hat{\sigma}_0^2$, is immune to this problem.

7. Example: Campbell and Mankiw's Estimate of the Fraction of Current Income Consumers

In a classic paper, Campbell and Mankiw (1989) suggested that the slope in the regression of the change in the log of consumption on the change in the log of current income may be interpreted as the fraction of consumers that are current income consumers rather than being permanent income consumers. To deal with the endogeneity of current income, they employ IV, using as instruments various combinations of lags of the change in income,

[10] In contrast, when there is no endogeneity, $\rho = 0$, $\theta = 0$. The exact density is t_k, which other than having fat tails is not too badly represented by a normal.

consumption, T bill yield, and also the lagged error correction term. They find that the asymptotic IV standard errors imply a striking rejection of the permanent income hypothesis, while at the same time rejecting the hypothesis that the fraction of permanent income consumers is zero.

In Table 16 we have estimated the same 9 models presented by Campbell and Mankiw in their Table 1, but for a later time period having the same number of observation for which we could obtain data from the DRI/ McGraw-Hill database. The OLS slope in model is .278, not dissimilar to the .316 reported by CM. We find that all of the 95% Wald IV intervals except one exclude zero, and all except two exclude unity. The tightest intervals are provided by the models with the greatest number of instruments. The impression is that the fraction of current income consumers is not less than about .25, but also is not more than about .90, the remaining being permanent income consumers.

The LM and LR confidence regions presented in Table 16 give a qualitatively different message. Three of the LM confidence regions cover the whole real line. The five closed LM confidence regions are wider than the Wald intervals, but are not symmetric around the IV point estimate. All are shifted and skewed in the positive direction relative to the Wald so that the upper bound increases more than the lower. The upper bound is above unity in one case. The LR results include one external confidence region that has a “hole” of rejection that is narrow and not in the (0, 1) interval. Two more LR confidence regions cover the whole real line. The five closed intervals are again shifted and skewed in the positive direction, away from the permanent income hypothesis, and in only two cases exclude unity. Again, the tightest intervals are provided with the models with the most instruments, and those exclude unity, but the overall impression now is that unity cannot be ruled out, while fractions as low as the OLS estimate are strongly rejected. The effect of considering the non-Wald intervals is to shift the evidence markedly away from the permanent income hypothesis.

8. Conclusions

This paper is motivated by the poor performance of confidence intervals based on Wald test statistics in the context of the estimation of a structural equation using weak instruments.

Traditional Wald confidence regions are much too narrow and actually cover the true parameter value with far lower probability than the nominal level. Here we have investigated alternatives to Wald confidence intervals, in particular those based on inversion of the test statistic of Anderson and Rubin (1949), the likelihood ratio statistic, and the Lagrange multiplier statistic. Counterparts of the AR, LR, and LM confidence regions in the GMM framework are also discussed.

Rather little attention has been paid to construction of confidence regions by inversion of LR and LM test statistics in econometrics. We find that the LR, LM, and AR confidence regions have a similar quadratic structure, implying that they may be closed, unbounded, disjoint. While Wald confidence regions are always bounded, LR, LM, and AR confidence regions are often unbounded when the instrumental variable is of poor quality. When the F in the first stage regression is not significant then unbounded confidence regions are likely to occur, reflecting appropriately the lack of information in the data. Further, the first stage F turns out to serve as a convenient indicator of how to choose appropriate degrees of freedom for constructing the confidence regions. These phenomena are observed in a Monte Carlo experiment that compares the unidentified (irrelevant instrument), weak instrument, and good instrument cases. While the AR confidence region always has correct coverage probability because it is based on an exact distribution, the empirical coverage frequencies for the LR and LM confidence regions are closest to the nominal level if the degrees of freedom is adjusted according to whether the first stage F statistic is significant or not.

Comparisons of power suggest that the LM and LR offer some advantage over AR, but the results are sensitive to adjustment of degrees of freedom base on the first stage F statistic. In comparison to the modified statistics proposed by Staiger and Stock, all three statistics have the considerable advantage of being free of nuisance parameters

In summary, in the instrumental variable framework inference and confidence regions should be based on the LM or LR or on our degrees of freedom adjusted LM_{sw} or LR_{sw} statistics. The AR statistic is also appropriate, having slightly better size and slightly worse power properties.

Appendix

Proof of proposition 1:

It suffices to show that the numerators of the Wald_{GMM}, LM_{GMM}, and LR_{GMM} are numerically equivalent. Write $y - x\hat{\beta} = y - x\beta_0 - x\left(\hat{\beta} - \beta_0\right)$. Then

$\left(y - x\hat{\beta}\right)' P_z \left(y - x\hat{\beta}\right) = \left(y - x\beta_0\right)' P_z \left(y - x\beta_0\right) - 2\left(\hat{\beta} - \beta_0\right)' x' P_z \left(y - x\beta_0\right) + \left(\hat{\beta} - \beta_0\right)' x' P_z x \left(\hat{\beta} - \beta_0\right)$ and the numerator of LR_{GMM} becomes $2\left[P_z x\left(\hat{\beta} - \beta_0\right)\right]' \left(y - x\beta_0\right) - \left[P_z x\left(\hat{\beta} - \beta_0\right)\right]' P_z \left(\hat{\beta} - \beta_0\right)$. Next, observe that $P_z x\left(\hat{\beta} - \beta_0\right) = P_{\hat{x}} y - \hat{x}\beta_0 = P_{\hat{x}} - P_{\hat{x}}\hat{x}\beta_0 = P_{\hat{x}}\left(y - x\beta_0\right)$ since $P_{\hat{x}} x = P_{\hat{x}}\hat{x} = \hat{x}$ Therefore, the numerator of LR_{GMM} simplifies to

$2\left(y - x\beta_0\right)' P_{\hat{x}} \left(y - x\beta_0\right) - \left(y - x\beta_0\right)' P_{\hat{x}} \left(y - x\beta_0\right) = \left(y - x\beta_0\right)' P_{\hat{x}} \left(y - x\beta_0\right)$ which is the numerator for LM_{GMM}.

Next, Consider the numerator for Wald_{GMM}. Simple manipulations yield

$\left(y - x\beta_0\right)' x' P_z x\left(\hat{\beta} - \beta_0\right) = \left[P_{\hat{x}}\left(y - x\beta_0\right)\right]' P_{\hat{x}} \left(y - x\beta_0\right) = \left(y - x\beta_0\right)' P_{\hat{x}} \left(y - x\beta_0\right)$ which is the numerator for LM_{GMM}.

Proof of Proposition 2:

The AR, LR and LM confidence sets are determined by finding all values of β_0 that satisfy (16), and the set will be unbounded if the coefficient a in (16) is less than zero.

Part (a): Here $a = x'[I - \phi_{AR} \cdot M_z]x$ where $\phi_{AR} = 1 + F(k, T-k; 1-\alpha) \cdot \left(k/T-k\right)$. Now a<0 if $(x'x)/(x'M_z x) < \phi_{AR}$ which can be rearranged to give the condition

$$F_{\pi=0} = \frac{(x'x - x'M_z x)/k}{x'M_z x / T - k} < F(k, T-k; 1-\alpha).$$

Part (b): Here $a = x'[I - \phi_{LR} \cdot M_Z]x$ where $\phi_{LR} = \exp\left\{\frac{\chi^2(1;1-\alpha)}{T}\right\} \cdot \hat{\kappa}$. Now a<0 if $(x'x)/(x'M_Z x) < \phi_{LR}$. After some simple manipulations, we obtain the equivalent condition

$$F_{\pi=0} = \frac{(x'x - x'M_Z x)/k}{x'M_Z x / T-k} < \left(\frac{T-k}{k}\right) \cdot \left(\exp\left\{\frac{\chi^2(1;1-\alpha)}{T}\right\} \cdot \hat{\kappa} - 1\right).$$

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

Proposition 3

The general statement of the density of the instrumental variable estimator for model (1)-(3) in the just identified case follows directly from Hinkley (1969) who cites Fieller (1932). The numerator and denominator of the IV estimator for (1)-(3) are distributed bivariate normal with means $\theta_1 = \beta\gamma M_{zz}$ and $\theta_2 = \gamma M_{zz}$, variances $\sigma_1^2 = (\beta^2\sigma_v^2 + \sigma_u^2 + 2\beta\sigma_{uv})M_{zz}$ and $\sigma_2^2 = \sigma_v^2 M_{zz}$, covariance $\sigma_{12} = (\beta\sigma_v^2 + \sigma_{uv})M_{zz}$, and correlation $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$, where we have used γ rather than π for the coefficient in (2) to avoid confusion in the pdf with $\pi \approx 3.1415$. Letting $\Phi(\ )$ represent the standard normal cdf, the pdf for the instrumental variable estimator is

$$f(\beta_{IV}) = \frac{b(\beta_{IV})d(\beta_{IV})}{\sqrt{2\pi}\sigma_1\sigma_2 a^3(\beta_{IV})}\left[\Phi(j(\beta_{IV}))-\Phi(-j(\beta_{IV}))\right] + \frac{\sqrt{1-\rho^2}}{\pi\sigma_1\sigma_2 a^2(\beta_{IV})}\exp\left(-\frac{c}{2(1-\rho^2)}\right)$$

where

$$a(\beta_{IV}) = \left(\frac{\beta_{IV}^2}{\sigma_1^2} - \frac{2\rho\beta_{IV}}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right)^{0.5}$$

$$b(\beta_{IV}) = \frac{\theta_1\beta_{IV}}{\sigma_1^2} - \frac{\rho(\theta_1 + \theta_2\beta_{IV})}{\sigma_1\sigma_2} + \frac{\theta_2}{\sigma_2^2}$$

$$c = \frac{\theta_1^2}{\sigma_1^2} - \frac{2\rho\theta_1\theta_2}{\sigma_1\sigma_2} + \frac{\theta_2^2}{\sigma_2^2}$$

$$d(\beta_{IV}) = \exp\left\{\frac{b^2(\beta_{IV}) - ca^2(\beta_{IV})}{2(1-\rho^2)a^2(\beta_{IV})}\right\}$$

$$j(\beta_{IV}) = \frac{b(\beta_{IV})}{\sqrt{1-\rho^2}\,a(\beta_{IV})}$$

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 (1994), "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models, and Cointegration," Université de Montréal, unpublished paper.

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

Fieller, E.C. (1932), "The distribution of the index in a normal bivariate population," *Biometrika* 24, 428-40.

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.

Geary, R. C. (1930), The Frequency Distribution of the Quotient of Two Normal Variates," *J. of the Royal Statistical Society, Series A*, 93, 442-446.

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1994), "Judging Instrument Relevance in Instrumental Variables Estimation," unpublished paper, Federal Reserve Board.

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, vol. 50, 1029-54.

Hinckley, D. V. (1969), "On the Ratio of Two Correlated Normal Random Variables," *Biometrika*, 56, 635-639.

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*, Vol. 42, No. 5, 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, pp. 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. (1990), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967-976.

__________ and ________ (1990), "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.

Rotenberg, 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, (1993), "Instrument Relevance in Linear Models: A Simple Measure," U Wisc. SSRI.

Staiger, D., and Stock, J. H. (1994), "Asymptotics for Instrumental Variables Regressions with Weakly Correlated Instruments," NBER Technical Working Paper 151, and forthcoming in *Econometrica*.

Venzon, D. J., and Moolgavkar, S. H. (1988), A Method for Computing Profile-Likelihood-Based Confidence Intervals," Applied Statistics, vol. 37, No. 1, 87-94.

Wang, Jiahui and Zivot, Eric, (1996), "Inference on a Structural Parameter in Instrumental Variables Regressions with Weak Instruments," ms in preparation, Dept. of Economics, University of Washington .

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : PQGCVT.OUT
Creator : PGL v-2.00 by AnSoft, Inc.
Pages : 1

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : PQGCVT7.OUT
Creator : PGL v-2.00 by AnSoft, Inc.
Pages : 1

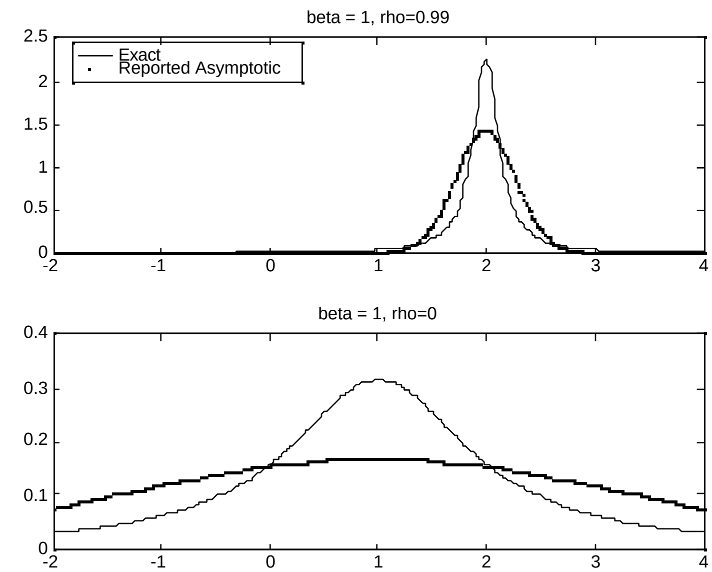
beta = 1, rho=0.99
Exact
Reported Asymptotic
2.5
2
1.5
1
0.5
0
-2
-1
0
1
2
3
4
beta = 1, rho=0
0.4
0.3
0.2
0.1
0
-2
-1
0
1
2
3
4

Figure 8

Table 1

Empirical Rejection Frequencies with and Without First-Stage Pre-test

	$\pi = 0$	$\pi = 0.1$
all	62.92% (6292/10000)	18.72% (1872/10000)
first-stage significant	100% (518/518)	86.88% (1470/1692)
first-stage not significant	60.89% (5774/9482)	4.84% (402/8308)

Table 2: The Just Identified Model; k=1

	$\pi = 1$	$\pi = 0.1$	$\pi = 0.0$
$\hat{\beta}_{OLS}$	0.550 (0.044)	0.975 (0.015)	0.995 (0.013)
$\hat{\beta}_{2SLS}$	0.148 (0.081)	0.657 (0.152)	1.064 (0.101)
$\hat{\pi}$	1.161 (0.010)	0.262 (0.010)	0.162 (0.009)
$F_{\pi=0}$	108.1 (0.000)	5.48 (0.019)	2.092 (0.148)
$T \cdot R^2_{UC}$	52.45 (0.000)	5.297 (0.021)	2.090 (0.148)
C_{Wald}(ß,.95)	[-0.01, 0.31]	[0.38, 0.95]	[0.90, 1.30]
C_{LM}(ß,.95)	[-0.053, 0.285]	[-1.196, 0.832]	[-•, •]
C_{LR}(ß,.95)	[-0.050, 0.284]	[-1.085, 0.831]	[-•, •]
C_{AR}(ß,.95)	[-0.050, 0.284]	[-1.090, 0.831]	[-•, •]

Table 3: The Overidentified Model: k=4

	$\pi = 1$	$\pi = 0.1$	$\pi = 0.0$
$\hat{\beta}_{OLS}$	0.538 (0.050)	0.985 (0.016)	0.995 (0.013)
$\hat{\beta}_{2SLS}$	0.040 (0.101)	0.586 (0.256)	0.898 (0.128)
$\hat{\beta}_{LIML}$	0.027 (0.103)	0.316 (0.530)	0.852 (0.185)
$\hat{\kappa}$	1.012	1.011	1.005
$\hat{\pi}$: 1	1.018 (0.000)	0.118 (0.000)	0.018 (0.010)
2	-0.030 (0.011)	-0.030 (0.011)	-0.030 (0.011)
3	-0.060 (0.010)	-0.060 (0.010)	-0.060 (0.010)
4	0.139 (0.012)	0.139 (0.012)	0.139 (0.012)
$F_{\pi=0}$	21.859 (0.000)	0.681 (0.606)	0.384 (0.820)
$T \cdot R^2_{UC}$	47.927 (0.000)	2.784 (0.595)	1.592 (0.810)
C_{Wald}(ß; .95)	[-0.15, 0.25]	[0.10, 110]	[0.70, 110]
$C^1_{LM}(\beta; .95)$	[-0.214, 0.208]	[-•, 0.837]»[3.232, •]	[-•, •]
$C^4_{LM}(\beta; .95)$	[-0.452, 0.287]	[-•, 0.906]»[1.394, •]	[-•, •]
$C^1_{LR}(\beta; .95)$	[-0.231, 0.197]	[-•, 0.744]»[2.300, •]	[-•, •]
$C^4_{LR}(\beta; .95)$	[-0.460, 0.272]	[-•, 0.828]»[1.446, •]	[-•, •]
C_{AR}(ß; .95)	[-0.412, 0.259]	[-•, 0.816]»[1.506, •]	[-•, •]

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

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table4.ps
Title : PRE1J.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table5.ps
Title : PRE5J.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table6.ps
Title : PRE0J.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table7.ps
Title : PRE1U.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table8.ps
Title : PRE5U.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table9.ps
Title : PRE0U.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table10.ps
Title : POW1J.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table11.ps
Title : POW5J.XLS
Creator : Windows PSCRIPT
Pages : 0

Table 12:
Empirical Power of Various Statistics: k = 1, ρ = 0

π = 0

	Wald	AR	LM	LR
β = −1.00	0.26	0.06	0.06	0.06
β = −0.75	0.21	0.05	0.05	0.06
β = −0.50	0.16	0.05	0.05	0.05
β = −0.25	0.13	0.04	0.04	0.04
β = 0.00	0.08	0.06	0.06	0.07
β = +0.25	0.03	0.05	0.05	0.05
β = +0.50	0.01	0.05	0.05	0.05
β = +0.75	0.00	0.04	0.04	0.05
β = +1.00	0.00	0.06	0.05	0.06
β = +1.25	0.00	0.04	0.04	0.04
β = +1.50	0.01	0.06	0.06	0.07
β = +1.75	0.03	0.05	0.05	0.05
β = +2.00	0.07	0.05	0.05	0.06
β = +2.25	0.12	0.07	0.07	0.07
β = +2.50	0.16	0.05	0.05	0.05
β = +2.75	0.22	0.04	0.04	0.04
β = +3.00	0.24	0.06	0.06	0.06

π = 0.1

	Wald	AR	LM	LR
β = −1.00	0.47	0.14	0.14	0.14
β = −0.75	0.40	0.13	0.13	0.13
β = −0.50	0.34	0.12	0.12	0.12
β = −0.25	0.23	0.11	0.11	0.11
β = 0.00	0.17	0.11	0.11	0.11
β = +0.25	0.09	0.09	0.09	0.09
β = +0.50	0.03	0.07	0.07	0.08
β = +0.75	0.00	0.06	0.05	0.06
β = +1.00	0.00	0.04	0.04	0.04
β = +1.25	0.01	0.04	0.04	0.04
β = +1.50	0.05	0.09	0.09	0.09
β = +1.75	0.08	0.09	0.09	0.09
β = +2.00	0.17	0.11	0.11	0.11
β = +2.25	0.26	0.11	0.11	0.12
β = +2.50	0.31	0.15	0.14	0.15
β = +2.75	0.38	0.12	0.11	0.12
β = +3.00	0.44	0.14	0.13	0.14

π = 1

	Wald	AR	LM	LR
β = −1.00	1.00	1.00	1.00	1.00
β = −0.75	1.00	1.00	1.00	1.00
β = −0.50	1.00	1.00	1.00	1.00
β = −0.25	1.00	1.00	1.00	1.00
β = 0.00	1.00	1.00	1.00	1.00
β = +0.25	1.00	1.00	1.00	1.00
β = +0.50	1.00	1.00	1.00	1.00
β = +0.75	0.68	0.66	0.66	0.66
β = +1.00	0.05	0.05	0.05	0.05
β = +1.25	0.71	0.68	0.68	0.68
β = +1.50	1.00	0.99	0.99	0.99
β = +1.75	1.00	1.00	1.00	1.00
β = +2.00	1.00	1.00	1.00	1.00
β = +2.25	1.00	1.00	1.00	1.00
β = +2.50	1.00	1.00	1.00	1.00
β = +2.75	1.00	1.00	1.00	1.00
β = +3.00	1.00	1.00	1.00	1.00

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table13.ps
Title : POW1U.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table14.ps
Title : POW5U.XLS
Creator : Windows PSCRIPT
Pages : 0

This EPS image does not contain a screen preview.
It will print correctly to a PostScript printer.
File Name : table15.ps
Title : POW0U.XLS
Creator : Windows PSCRIPT
Pages : 0

Table 16: 95% Confidence Intervals for Campbell and Mankiw's Estimate of the Fraction of Current Income Consumers; quarterly data 1961.1:1994.4.

Model: instr.	First stage R-sq	Wald IV	LM switching	LR switching
1: OLS		(.20, .35)		
2: 3 lags Δy	.024	(.04, 1.51)	$(-\infty, +\infty)$	$(-\infty, -.14)$ $(-.01, +\infty)$
3: 5 lags Δy	.060	(.19, .90)	$(-\infty, +\infty)$	$(-\infty, +\infty)$
4: 3 lags Δc	.076*	(.24, .89)	(.30, 1.20)	(.38, 2.15)
5: 5 lags Δc	.115*	(.28, .78)	(.31, .92)	(.42, 1.56)
6: 3 lags Δi	.026	(-.00, 1.06)	$(-\infty, +\infty)$	$(-\infty, +\infty)$
7: 5 lags Δi	.070*	(.28, .78)	(.31, .93)	(.42, 1.56)
8: 7 in total	.170*	(.25, .63)	(.26, .69)	(.30, .84)
9: 10 in total	.210*	(.29, .65)	(.31, .69)	(.40, .92)

* First stage regression significant at .05 level.