Improved Inference for the Instrumental Variable Estimator

by

Richard Startz

Charles R. Nelson

Eric Zivot*

October 1999

Abstract

It is now well known that standard asymptotic inference techniques for instrumental variable estimation perform very poorly in the presence of weak instruments. Specifically, standard asymptotic techniques give spuriously small standard errors, leading investigators to accept apparently tight confidence regions which unfortunately may be very far from the true parameter of interest. We present an improved technique for inference on structural parameters based on reduced form estimates. The "*S*-statistic" produces confidence regions based on a joint test of the structural hypothesis and the identification condition. The *S*-statistic converges to the standard asymptotic Wald statistic as identification becomes certain, has much better size properties when the instruments are weak, and may be inverted in closed form to conveniently compute confidence regions. In addition to providing improved inference for instrumental variable estimation, the technique suggested here may be useful in other applications where weak identification is important.

* Department of Economics, University of Washington, Seattle, WA 98195-3330, USA, correspondence to the 1st author at startz@u.washington.edu. The authors would like to thank our research assistants, Ines Gutzeit, Idanna Kaplan, Ming Chien Lo, and Kathryn Temple. Support from NSF grant SBR-9711301 is gratefully acknowledged.

1 Introduction

It is now well known that standard asymptotic inference techniques for instrumental variable estimation perform very poorly in the presence of weak instruments. The failure is of the worst kind — false results are accompanied by reported confidence intervals which lend an appearance of great precision. That point estimates of coefficients do a poor job of telling us the true values of those coefficients is probably irremediable, after all if an equation is poorly identified then the data do not tell us much about the parameters of the system. However, it is possible to create test statistics and confidence intervals that work quite well in the sense that they lead to reasonably accurate inference when instruments are poor and that are essentially identical to the asymptotic test statistics and confidence intervals when the instruments are good.

The usual asymptotic Wald statistic asks how far an estimated structural parameter is from some hypothesized value, where the metric for "how far" is based on an estimated asymptotic standard error which is itself calculated from the estimated structural parameters. When the structural parameters are poorly identified, this leads to misleading inference. The *S*-statistic we introduce here is based on three principles (the first two of which may also be useful in other cases of weak identification). First, we wish to reject the structural hypothesis only if the estimated parameter is far from its hypothesized value *and* the model is identified. Equivalently we should fail to reject if the parameter is close to the specified value *or* the model is unidentified. We accomplish this task by multiplying the distance between the estimated and hypothesized value of the parameter by an "identification statistic" which equals zero if and only if the parameter is unidentified. In this way we have a joint test statistic which is close to zero if either the estimated parameter is close to the specified value *or* if the data does not support

identification. The converse is that if the parameter is identified with (near) certainty, we have multiplied the usual asymptotic statistic by a (near) constant, leaving inference unchanged. Second, we "studentize" the joint statistic by an estimated standard deviation that can be computed from reduced form parameters, which are consistent whether or not the structural parameter is identified. The third principle, specific really to instrumental variables, is that the inaccuracy of IV under weak identification is due in part to a "dividing by zero" phenomenon which multiplication by the identification statistic substantially mitigates.

Having introduced the *S*-statistic, we show how the corresponding confidence regions, which we call *S*-intervals, can be computed in closed form. When an equation is well-identified, *S*-intervals are quite close to traditional asymptotic confidence intervals. When an equation is weakly identified *S*-intervals are appropriately wide, indeed they may be unbounded. Where much of the existing literature on weak instrument instrumental variables has focused on the case of a single endogenous right hand side variable, the method we introduce provides for testing each individual coefficient in the general k-variable model. We further allow for the situation where some coefficients are well-identified while others are not.

The principal goal of the paper is to provide practitioners with improved tools for inference in the presence of weak instruments – tools which can augment or replace the traditional asymptotic tests. Specifically, the *S*-statistic and corresponding confidence regions are computed for each coefficient and can be used in place of the traditional t-statistics and associated confidence intervals. The asymptotic

distribution theory for our new statistic is completely standard.[1] The advantage of the *S*-statistic lies in its better finite sample properties under weak identification. We provide analytic results for a special case and then present further evidence through Monte Carlo simulations.

To set out notation, we begin with a review of the standard instrumental variable model and the corresponding reduced form. We then give a brief literature review. Next we present the *S*-statistic. We investigate the one endogenous variable/one instrument case, where we can give exact finite sample results. For the multiple endogenous variable/multiple instrument case we present Monte Carlo simulations to appraise the performance of the *S*-statistic. We conclude with a brief summary.

2 The General Model

2.1 The Structural Model

We begin with the classic statements about instrumental variables, in the process defining notation for the paper. Consider the *structural* linear equation with k right-hand side variables

$$\underset{(n\times 1)}{y} = \underset{(n\times k)}{X}\ \underset{(k\times 1)}{b} + \underset{(n\times 1)}{u}\ . \tag{1}$$

Suppose that the right-hand side variables are correlated with the errors, $\operatorname{plim}\left(\frac{1}{n}X'u\right)\neq 0$, but there exists a set of q instruments Z, $q\geq k$, $\operatorname{plim}\left(\frac{1}{n}Z'u\right)=0$, $\operatorname{plim}\left(\frac{1}{n}Z'X\right)\neq 0$, where $\operatorname{plim}\left(\frac{1}{n}Z'Z\right)$ is of full rank q and $\operatorname{plim}\left(\frac{1}{n}Z'X\right)$ is of rank k. We further assume that the errors are homoskedastic, $\frac{1}{\sqrt{n}}Z'u\xrightarrow{d}N\left(0,\operatorname{plim}\left(\frac{s_u^2}{n}Z'Z\right)\right)$. b is commonly estimated by instrumental variables (equivalently two stage least squares or generalized method of moments). The instrumental variable estimator is

[1] The statistical theory pretty much amounts to nonlinear tests on least squares coefficients, which is why

$$b_{IV} = \left(X'P_Z X \right)^{-1} X'P_Z y, \text{ where } P_Z \equiv Z\left(Z'Z \right)^{-1} Z'. \tag{2}$$

The asymptotic distribution of b_{IV} is given by $b_{IV} \overset{A}{\sim} N\left(b, V_A \right)$. It is useful to define $\hat{Q} \equiv X'P_Z X$, so the asymptotic variance is given by $V_A = \frac{s_u^2}{n}\left[\text{plim}\left(\frac{1}{n}\hat{Q} \right) \right]^{-1}$.

The *reduced form* of the model consists of the regression of *y* and each column of *X* on all the instruments,

$$\underset{(n\times 1)}{y} = \underset{(n\times q)}{Z}\ \underset{(q\times 1)}{q} + \underset{(n\times 1)}{v}, \tag{3}$$

$$\underset{(n\times k)}{X} = \underset{(n\times q)}{Z}\ \underset{(q\times k)}{\Gamma} + \underset{(n\times k)}{e}, \tag{4}$$

where q is $k\times 1$, Γ is $q\times k$, and e is $n\times k$. Substitute equation (4) into equation (1), deriving

$$y = \left(Z\Gamma + e \right) b + u = Z\Gamma b + \left(eb + u \right). \tag{5}$$

Comparison of equations (3) and (5) highlight the restrictions imposed by identification:

$$q = \Gamma b\ . \tag{6}$$

It is useful to note that $\hat{Q}$ can also be written in terms of the reduced form parameters, $\frac{1}{n}\hat{Q} \equiv \frac{1}{n} X'P_Z X = \frac{1}{n}\hat{\Gamma}'\left(Z'Z \right)\hat{\Gamma}$. The asymptotic variance may be expressed in terms of the reduced form parameters as $V_A = \frac{s_u^2}{n}\left(\Gamma' \text{plim}\left(\frac{1}{n} Z'Z \right)\Gamma \right)^{-1}$.

---

the tests work pretty well.

2.1 Estimation of the Reduced Form

The reduced form coefficients are, of course, estimated by least squares, $\hat{q} = (Z'Z)^{-1} Z'y$ and $\hat{\Gamma} = (Z'Z)^{-1} Z'X$. Because we need the covariance matrix of the estimated coefficients, it is convenient to think of the reduced form as a system of seemingly unrelated regressions

$$\text{vec}(y, X) = (Z \otimes I)\,\text{vec}(q, \Gamma) + \text{vec}(v, e). \tag{7}$$

Define l to be the $q \cdot (1+k) \times 1$ column vector of reduced form coefficients in equation (7) and $\hat{l}$ to be the corresponding estimated least squares coefficients, $\hat{l} = \left[\hat{q}' \quad \hat{\Gamma}_1' \quad \cdots \quad \hat{\Gamma}_k'\right]'$ and $\Sigma_r = E\left(\left[v \quad e_1 \quad e_2 \quad \cdots \quad e_k\right]'\left[v \quad e_1 \quad e_2 \quad \cdots \quad e_k\right]\right)$, then $\text{cov}(\hat{l}) = \frac{1}{n}\Sigma_r \otimes \left(\text{plim}\frac{1}{n}(Z'Z)\right)^{-1}$. If we condition on *Z* and the reduced form errors are normal, or if the sample is reasonably large, then $\hat{l}$ will be normally distributed.

Note that the 2SLS interpretation of the instrumental variable estimator is $b_{IV} = b_{2SLS} = \left(\left(Z\hat{\Gamma}\right)'\left(Z\hat{\Gamma}\right)\right)^{-1}\left(Z\hat{\Gamma}\right)'\left(Z\hat{q}\right)$. The order condition for identification is $q \geq k$. The rank condition is usually written $\text{rank}(\Gamma) = k$. For our purposes it is more useful to write $\text{rank}\left(\text{plim}\frac{1}{n}\Gamma'Z'Z\Gamma\right) = k$, which implies $\left(\Gamma'\,\text{plim}\frac{1}{n}(Z'Z)\Gamma\right)_{ii} > 0\ \forall i$.[2] When the rank condition is satisfied and $q = k$, equation (1) is just identified and the indirect least squares interpretation of the instrumental variable estimator is $b_{IV} = b_{ILS} = \hat{\Gamma}^{-1}\hat{q}$.

[2] If any diagonal element of a symmetric matrix *A* is zero, the matrix cannot be positive definite since the quadratic form $d'Ad = 0$ for the vector $d = 0$ except $d_i = 1$.

3 A Brief Review of the Literature

A series of recent papers have examined the distribution of the instrumental variable estimator under weak identification and the related issue of the performance of the traditional asymptotic tests. Papers include Bekker (1994), Bound, Jaeger, and Baker (1995), Hahn and Hausman (1999), Maddala and Jeong (1992), Nelson and Startz (1990a, b), and Staiger and Stock (1997). Dufour (1997) gives general results for obtaining correct probability levels with weak identification. In particular, Dufour shows that for a statistic of nominal size a to be valid under weak identification, the confidence intervals implied by the statistic must be unbounded at least $1-a$ percent of the time. We return to this point in section 4.2.

Half a century ago, Anderson and Rubin (1949) and Anderson (1950) described the Anderson-Rubin (AR) statistic, which under normality provides an exact small sample test of a hypothesis which specifies values for every element of the b vector. Zivot, Startz, and Nelson (1998) (ZSN) and Dufour and Jasaic (1996) show how to use the AR-statistic to construct confidence regions in the case of a single endogenous variable. ZSN also provide improved statistics for maximum likelihood estimates based on degrees-of-freedom-corrected LR and LM tests. Wang and Zivot (1998) provide an asymptotic justification using the Staiger and Stock local-to-zero asymptotics for these results. Note that these papers are limited to inference in the case of a single endogenous right hand side variable or to hypotheses specifying values for the entire vector of coefficients; here we deal with inference on individual coefficients in the general k-right hand side variable case. Note additionally that while these papers provide confidence intervals for b, the confidence intervals are not in general based

on the instrumental variable estimator. The computed confidence intervals may reject the estimated value of b_{IV}, which is at least a nuisance although probably not a fatal flaw.

Stock and Wright (forthcoming) provides a general procedure for inference for gmm with weak instruments which for the linear single equations model is based on LIML estimates. Stock and Wright point out that using their method "construction of asymptotically valid confidence intervals for subvectors … is somewhat … difficult," but that an asymptotically conservative confidence interval can be found by projecting out parameters as suggested in Dufour (1997). As a practical matter using the procedure to test a hypothesis is relatively straightforward but inverting the test to find confidence intervals requires a numerical search.

Hall, Rudebusch, and Wilcox (1996) examine direct tests of the rank condition based on the size of the smallest canonical correlation. They also point out that for the usual asymptotic Wald statistic screening based on such a pre-test, of which the first-stage R^2 is a special case, can introduce an intended selection bias that worsens rather than mitigates finite sample bias.

Shea (1997) presents a diagnostic for identification. (See also Godfrey (1999).) We examine Shea's measure in detail because it provides insight on the statistic we propose here. One normally thinks of "weak instruments" as meaning that the instruments and right hand side endogenous variables are poorly correlated. Shea presents a useful extension which we discuss after defining a few symbols. Let X_i be the ith column of X and $X_{\sim i}$ be the remainder of X. Define

$\breve{X}_i \equiv$ the residual from regressing X_i on $X_{\sim i}$, or if $k=1$ simply $\breve{X}_i \equiv X_i$. Similarly, where

$\hat{X} = Z\hat{\Gamma} = P_Z X$ are the fitted values from the first-stage, let $\hat{X}_i$ be the ith column of $\hat{X}$ and $\hat{X}_{\sim i}$ be the

remainder of $\hat{X}$. Define $\breve{\hat{X}}_i \equiv$ the residual from regressing $\hat{X}_i$ on $\hat{X}_{\sim i}$, or if $k=1$ simply $\breve{\hat{X}}_i \equiv \hat{X}_i$. Shea calls the squared correlation between $\breve{X}_i$ and $\breve{\hat{X}}_i$ the "partial R^2" for the i[th] endogenous variable. (Note that for *k*=1 the partial R^2 is simply the first-stage R^2.) While the exact distribution for partial R^2 is unknown, Shea uses Monte Carlo experiments to show that partial R^2 is close to zero when weak instruments force the actual distribution of b_{IV} far from its asymptotic approximation.

Shea's simulations show that his partial R^2 does a good job of signaling weak instruments. We pick up this idea but use only the numerator of Shea's statistic, noting that the population value of partial R^2 equals zero if and only if the numerator equals zero. A little algebra, which is implicit in Shea's paper, shows that this numerator equals $\hat{X}_i'\left(I_n - \hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\right)\hat{X}_i$ which equals $1\Big/\left(\hat{Q}^{-1}\right)_{ii}$. Therefore the population value of the numerator of Shea's statistic for the *i*[th] coefficient equals zero if and only if the asymptotic variance of $b_{IV,i}$ is infinite (n.b. $1\Big/\mathrm{AVAR}\left(b_{IV,i}\right) = 1\Big/\left(\frac{s_u^2}{n}\left[\mathrm{plim}\left(\frac{1}{n}Q\right)\right]^{-1}\right)_{ii}$ – so that the *i*[th] coefficient is not identified.)

4 The S-Statistic

4.1 Combining tests of the structural hypothesis and identification

We introduce the *S*-statistic in this section as a joint test of a structural hypothesis and identification and give its asymptotic distribution. In section 5 we give finite sample results for the one-endogenous variable/one instrument case and present intuition for thinking of *S* as a scaling correction to the asymptotic *t*. Define

$$\hat{\Delta}_i \equiv \sqrt{1 \Big/ \left(\hat{Q}^{-1}\right)_{ii}} \,. \tag{8}$$

Computationally $\hat{\Delta}_i$ is the square root of the reciprocal of the i^{th} diagonal element of $\left(X' P_Z X\right)^{-1}$ or, equivalently, the reciprocal of the reported standard error of $b_{IV,i}$ divided by $\sqrt{s^2}$, the standard error of the regression from the IV estimation. Note that $\Delta_i > 0 \,\forall i$ is necessary for the rank condition to hold.

Suppose we wish to test that the i^{th} coefficient equals a hypothesized value, for example that $b_i = b_i^0$. Standard practice is to compare b_i^0 to $b_{IV,i}$ using an asymptotic *t*-test. We augment this comparison so that the test statistic will be close to zero either if the estimated deviation is small or if the evidence for identification is weak by forming

$$\Psi_i \equiv \hat{\Delta}_i \left(b_i^0 - b_{IV,i}\right). \tag{9}$$

It is useful to re-write Ψ_i as a function of the instruments and the estimated reduced form parameters, $\hat{\Gamma}$ and $\hat{q}$:

$$\Psi_i\left(\hat{\Gamma}, \hat{q}; Z; b_i^0\right) = \frac{1}{\sqrt{\left(\left(\hat{\Gamma}' Z' Z \hat{\Gamma}\right)^{-1}\right)_{ii}}} \left(b_i^0 - \left(\left(\hat{\Gamma}' Z' Z \hat{\Gamma}\right)^{-1} \left(\hat{\Gamma}' Z' Z \hat{q}\right)\right)_i \right). \tag{10}$$

In order to studentize Ψ_i we require an estimate of $\text{var}\left(\Psi_i\right)$. Since the estimated reduced form parameters are asymptotically normal, we can estimate $\text{var}\left(\Psi_i\right)$ by the usual Taylor series

approximation by conditioning the distribution of Ψ_i on the list of instruments. With this addendum the *S*-statistic is defined as follows

$$S \equiv \frac{\Psi_i}{\sqrt{\hat{s}^2_{\Psi_i}}}, \text{ where } \hat{s}^2_{\Psi_i} \equiv \frac{\partial \Psi_i}{\partial \hat{l}}' \operatorname{cov}\left(\hat{l}\right) \frac{\partial \Psi_i}{\partial \hat{l}}. \tag{11}$$

Note that *S* is a function of the reduced form parameters, which can be consistently estimated even when the structural parameters are not identified, and *Z*. The partial derivatives $\partial \hat{\Psi}_i / \partial \hat{l}$ are conveniently calculated by the numerical delta method and $\operatorname{cov}\left(\hat{l}\right)$ follows immediately from the reduced form estimates.[3]

Under standard regularity conditions[4], the *S*-statistic is approximately standard normal, or equivalently S^2 is approximately $c^2(1)$. Inference may be made by comparing computed $|S|$ to the usual normal critical values.[5] In a sufficiently well-identified model, $\operatorname{var}\left(\hat{\Delta}_i\right)$ and $\operatorname{cov}\left(\hat{\Delta}_i, b_{IV,i}\right)$ are both close to zero, so $S \approx \left(b_i^0 - b_{IV,i}\right) \Big/ \sqrt{\operatorname{var}\left(b_{IV,i}\right)}$ is close to the usual asymptotic *t*. Below, we show that in the one endogenous variable/one instrument case under normality S has an exact *t*-distribution in small samples regardless of the degree of identification.

[3] Matlab® code for computing the *S*-statistic and associated confidence regions is contained in an appendix available from the authors.

[4] Regularity conditions may include that the parameters are identified. In section 6 we consider improved approximations for cases where b is unidentified rather than merely "weakly" identified.

[5] Because the sign of the square root in equation (8) is arbitrary one-tailed tests aren't possible in most cases. Additionally, $S^2 \sim c^2(1)$ is a better finite sample approximation than is $S \sim N(0,1)$.

Our interest is in inference for weak instruments. There is, however, an interpretation with more general applicability. Ψ_i augments the structural hypothesis of interest by requiring the parameters of interest be identified. The estimated value of Ψ_i will be close to zero either if the hypothesized value of the structural coefficient is correct or if the model is not well-identified. This sort of joint test, which will be rejected only if the structural hypothesis is false *and* the model is well-identified, may be useful in other contexts as well.

4.2 Confidence Regions Based on the S-Statistic

S-confidence regions are computed by "inverting" the *S*-statistic; that is by finding the values of b_i^0 such that $|S| < c$, where c is the critical value for the desired size. In this section we show how to compute *S*-confidence regions in closed form and show that such regions take one of three shapes. We show that the confidence regions, and therefore the underlying *S*-statistics, converge on the usual asymptotic statistic as uncertainty about identification goes to zero. Finally, we demonstrate that the *S*-confidence region asymptotically satisfies Dufour's (1997) condition for unboundedness.

It is convenient to think of the *S*-confidence region as defined by $\Psi^2 < \hat{s}_\Psi^2 \cdot c^2$ (equivalently $S^2 < c^2$). The region takes one of three shapes: a familiar connected interval of the form $\left(b^L, b^H\right)$; the union of two rays $\left(-\infty, b^L\right) \cup \left(b^H, \infty\right)$; or the entire real line. Note that the latter two forms are unbounded.

Figure 1 shows a plot of the (square of the) *S*-statistic and corresponding asymptotic Wald statistic for the first, of two, coefficients for a particular Monte Carlo realization generated with the true $b_1 = 1$. Confidence intervals are the values of b_1^0 such that the test statistic lies below the appropriate

critical value, say $c^2 = 1.96^2$. The Wald statistic plots as a parabola and the corresponding 95 percent confidence region is a closed interval around b_{IV}, $(90.13, 156.43)$, a region which excludes the true value. Indeed, the *t*-statistic against $b = 1$ is 7.23. The *S*-region is the union $(-\infty, 86.29) \cup (97.15, \infty)$, which includes the true b. The *S*-statistic against $b = 1$ is 1.40.

Figure 1 illustrates how the shape of the *S*-region depends on the size of the test. At a larger size, and correspondingly smaller value of *c*, the *S*-region would be a connected interval around b_{IV}. Conversely, at a small size and high value of *c*, the *S*-region will be the entire real line. The confidence region always includes the estimated value b_{IV} in contrast to the AR statistic which may reject b_{IV}, the latter presumably reflecting a failure of the overidentifying restrictions.

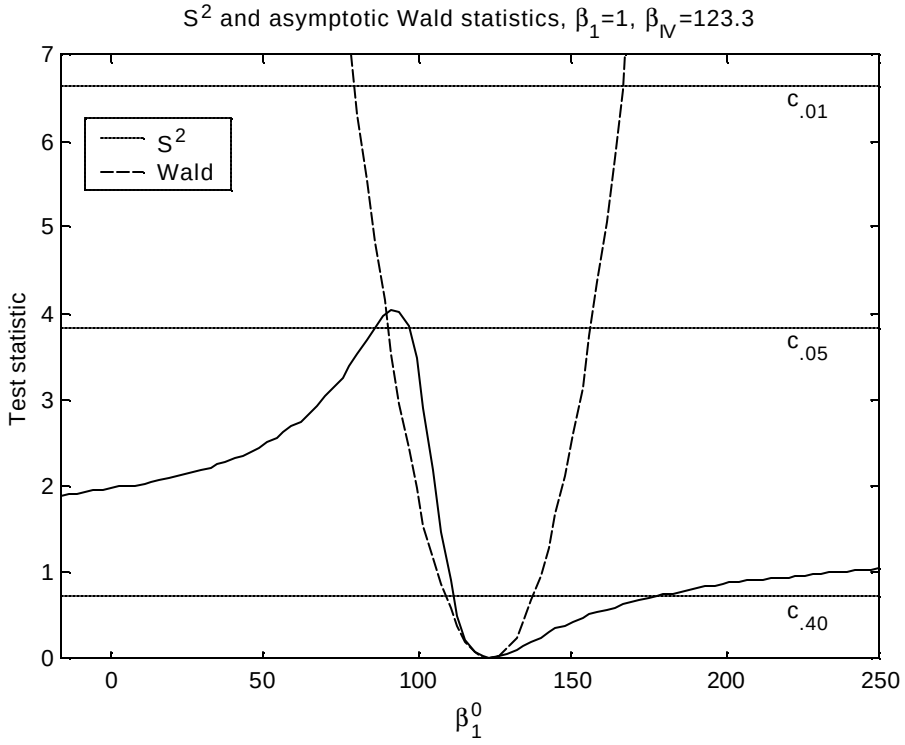

Figure 1

In order to compute confidence intervals it is useful to define $\hat{\jmath} \equiv \hat{\Delta} b_{IV}$ and then note[6]

$$\Psi^2 \equiv \hat{\Delta}^2 b^{0^2} - 2\hat{\Delta}\hat{\jmath} b^0 + \hat{\jmath}^2, \tag{12}$$

$$\hat{s}_\Psi^2 = b^{0^2} \operatorname{var}\left(\hat{\Delta}\right) - 2b^0 \operatorname{cov}\left(\hat{\Delta}, \hat{\jmath}\right) + \operatorname{var}\left(\hat{\jmath}\right). \tag{13}$$

Evaluating the ratio of Ψ^2 to $\hat{s}_\Psi^2$ as $b^0 \to \pm\infty$, it follows immediately that $\lim_{b^0 \to \pm\infty} |S| = \dfrac{\hat{\Delta}}{\sqrt{\operatorname{var}\left(\hat{\Delta}\right)}}$.

Therefore, the *S*-interval is unbounded iff the "*Z*-score" for identification is not significantly different from zero, as in this case *S* asymptotes to a value less than *c*.[7] Said differently, if $\dfrac{\hat{\Delta}}{\sqrt{\operatorname{var}\left(\hat{\Delta}\right)}} < c$, then $|S| < |c|$ for large values of $|b^0|$ so extreme values of b^0 are not rejected. It follows that the *S*-statistic asymptotically satisfies Dufour's (1997) condition requiring, in the case of a near non-identification, that a statistic be unbounded at least $1-a$ percent of the time for the statistic to attain size a.

From equations (12) and (13), the condition $\Psi^2 < \hat{s}_\Psi^2 \cdot c^2$ gives the confidence region defined by the quadratic inequality[8]

$$\left(\hat{\Delta}^2 - c^2 \operatorname{var}\left(\hat{\Delta}\right)\right) b^{0^2} + 2\left(-\hat{\Delta}\hat{\jmath} + c^2 \operatorname{cov}\left(\hat{\Delta}, \hat{\jmath}\right)\right) b^0 + \left(\hat{\jmath}^2 - c^2 \operatorname{var}\left(\hat{\jmath}\right)\right) < 0. \tag{14}$$

---

[6] We drop the subscripts indicating coefficient number where there is no danger of confusion, i.e. $\hat{\Delta}$ in place of $\hat{\Delta}_i$.

[7] Note the parallel to the unbounded confidence intervals corresponding to first stage *F*-statistics in Zivot, Startz, and Nelson.

[8] Computationally we estimate $\operatorname{var}\left(\hat{\Delta}\right) = \dfrac{\partial \hat{\Delta}'}{\partial \hat{l}} \operatorname{cov}\left(\hat{l}\right) \dfrac{\partial \hat{\Delta}}{\partial \hat{l}}$ and $\operatorname{cov}\left(\hat{\Delta}, \hat{\jmath}\right) = \dfrac{\partial \hat{\Delta}'}{\partial \hat{l}} \operatorname{cov}\left(\hat{l}\right) \dfrac{\partial \hat{\jmath}}{\partial \hat{l}}$, where the vectors of partial derivatives can be computed numerically.

The confidence region is defined by the roots of equation (14).[9] Let

$$R = \sqrt{\left(-\hat{\Delta}\hat{\jmath} + c^2 \operatorname{cov}\left(\hat{\Delta}, \hat{\jmath}\right)\right)^2 - \left(\hat{\Delta}^2 - c^2 \operatorname{var}\left(\hat{\Delta}\right)\right)\left(\hat{\jmath}^2 - c^2 \operatorname{var}\left(\hat{\jmath}\right)\right)}, \tag{15}$$

$$\left\{b^L, b^U\right\} = \frac{\left(\hat{\Delta}\hat{\jmath} - c^2 \operatorname{cov}\left(\hat{\Delta}, \hat{\jmath}\right)\right) \pm R}{\hat{\Delta}^2 - c^2 \operatorname{var}\left(\hat{\Delta}\right)}. \tag{16}$$

As can be seen in Figure 1, for small critical values the *S*-statistic cuts the horizontal critical value in two places and the confidence region lies between the two cut points. Specifically, if $\hat{\Delta} \Big/ \sqrt{\operatorname{var}\left(\hat{\Delta}\right)} > c$, then the confidence region from inverting the *S*-statistic is the connected interval $\left(b^L, b^U\right)$. In Figure 1, the 60 percent confidence region is $(111.44, 175.54)$. In the particular example, the identification *Z*-score is 1.187, so the *S*-statistic asymptotes to 1.187 and the *S*-region is connected for sizes less than $2 \cdot \left(1 - \Phi\left(1.187\right)\right) = 0.235$.

For higher critical values, if $\Delta \Big/ \sqrt{\operatorname{var}\left(\Delta\right)} < c$, the confidence region is the union of two rays defined by $\left(-\infty, b^L\right) \cup \left(b^U, \infty\right)$ if *R* is real, and the entire real line otherwise. The corresponding confidence region is the entire real line when the argument to the root in (15) is negative, which occurs for critical values above c^*, where

[9] It may be helpful to note that $\Psi^2 \Big/ \hat{s}_\Psi^2$ is the ratio of two quadratics in b^0, accounting for the characteristic shape illustrated in Figure 1. Zivot, Startz, and Nelson show that the Anderson-Rubin statistic can also be written as a ratio of two quadratics in b^0, so the *AR*-confidence regions can also

$$c^* = \sqrt{\frac{\hat{\Delta}^2 \operatorname{var}(\hat{\jmath}) + \hat{\jmath}^2 \operatorname{var}(\hat{\Delta}) - 2\hat{\Delta}\hat{\jmath} \operatorname{cov}(\hat{\Delta}, \hat{\jmath})}{\operatorname{var}(\hat{\Delta}) \operatorname{var}(\hat{\jmath}) - \operatorname{cov}^2(\hat{\Delta}, \hat{\jmath})}}\,. \tag{17}$$

In Figure 1, the 95 percent confidence region is $(-\infty, 86.29) \cup (97.15, \infty)$. In this example, $c^* = 2.01$; therefore no values of b can be rejected with size less than $2 \cdot (1 - \Phi(2.01)) = 0.044$.

In sufficiently well-identified models the uncertainty about $\hat{\Delta}$ is negligible, so $\operatorname{var}(\hat{\Delta}) << \hat{\Delta}^2$, $\operatorname{cov}(\hat{\Delta}, \hat{\jmath}) << \hat{\Delta}\hat{\jmath}$, and $\operatorname{var}(\hat{\jmath}) \approx \hat{\Delta}^2 \operatorname{var}(b_{IV})$. Equations (14) through (16) reduce to

$$(\hat{\Delta}^2) b^{0^2} + 2(-\hat{\Delta}^2 b_{IV}) b^0 + \hat{\Delta}^2 (b_{IV}^2 - c^2 \operatorname{var}(b_{IV})) < 0\,, \tag{14'}$$

$$R = \sqrt{(-\hat{\Delta}^2 b_{IV})^2 - (\hat{\Delta}^2)\hat{\Delta}^2 (b_{IV}^2 - c^2 \operatorname{var}(b_{IV}))} = \hat{\Delta}^2 c \sqrt{\operatorname{var}(b_{IV})}\,, \tag{15'}$$

$$\{b^L, b^U\} = b_{IV} \pm c\sqrt{\operatorname{var}(b_{IV})}\,. \tag{16'}$$

This establishes that *S*-intervals approach the intervals based on the asymptotic *t*-statistic as identification becomes certain.

The *S*-test and the corresponding confidence region is asymptotically valid and easy to compute. It is much more accurate than the traditional asymptotic statistic in the sense that it gives much wider confidence regions when the parameters are not well identified.

---

take on the same three shapes and a plot of $AR(b^0)$ exhibits the same basic functional form as does the plot of $S^2(b^0)$.

5 The One RHS Variable Model

The case of one right hand side variable is particularly amenable to theoretical analysis and provides much intuition. In the first three parts of this section we consider the one right hand side variable/one instrument model. We show that for this case that the *S*-statistic is exactly, rather than approximately, normal (conditional on *z* and assuming normal errors.) In the fourth subsection we derive the exact distribution of $\hat{\Delta}^2$ for the general $k = 1, q \geq 1$ case.

5.1 Distribution of the structural coefficient

For the one RHS-variable/just identified model we can write the structural equation as

$$y = b\,x + u\,, \tag{18}$$

and the two reduced form equations

$$\begin{aligned} y &= q\,z + v, \\ x &= g\,z + e\,, \end{aligned} \tag{19}$$

Substituting the reduced form equation for x into the structural equation leads to an alternative representation of the first reduced form equation

$$y = bg\,z + (be + u)\,. \tag{20}$$

Thus the identifying restriction is $q = bg$. Note that the instrumental variable estimator of b is given by $b_{IV} = \hat{q}_{OLS} / \hat{g}_{OLS}$, where $\hat{q}_{OLS}$ and $\hat{g}_{OLS}$ are the reduced form coefficients, which simply says that the IV estimator can be thought of as the indirect least squares estimator.

Traditional asymptotic distribution theory states

$b_{IV} \overset{A}{\sim} N\left(b, V_A\right)$, $V_A = \frac{s_u^2}{n}\frac{\text{plim}\left(\frac{1}{n}\sum z^2\right)}{\text{plim}\left(\frac{1}{n}\sum xz\right)^2} = \frac{s_u^2}{n}\left(g^2 \text{ plim}\left(\frac{1}{n}\sum z^2\right)\right)^{-1}$. Inference usually proceeds using the Wald statistic, that is the square of the "asymptotic *t*." The IV, 2SLS, and GMM estimators are all the same here (as is, for that matter, the LIML estimator). We are interested in the case of weak identification. It is illuminating to consider what happens as $g \to 0$. The answer depends on the correlation between u and e. Write $u = te + e^{\perp}$, where $t = \frac{s_{ue}}{s_e^2}$ and $E\left(ee^{\perp}\right) = 0$. The IV and OLS estimators are

$$
\begin{aligned}
b_{IV} &= \frac{\frac{1}{n}\sum zy}{\frac{1}{n}\sum zx} = \frac{\frac{1}{n}\sum z\left(bgz + \left(be + u\right)\right)}{\frac{1}{n}\sum z\left(gz + e\right)} = \frac{\frac{1}{n}\sum z\left(bgz + \left(be + \left(te + e^{\perp}\right)\right)\right)}{\frac{1}{n}\sum z\left(gz + e\right)}, \\
b_{ols} &= \frac{\frac{1}{n}\sum yx}{\frac{1}{n}\sum x^2} = \frac{\frac{1}{n}\sum\left(bgz + \left(be + u\right)\right)\left(gz + e\right)}{\frac{1}{n}\sum\left(gz + e\right)^2} = \frac{\frac{1}{n}\sum\left(bgz + \left(be + \left(te + e^{\perp}\right)\right)\right)\left(gz + e\right)}{\frac{1}{n}\sum\left(gz + e\right)^2}.
\end{aligned}
\tag{21}
$$

The extreme case is when $g \to 0$ so

$$
\begin{aligned}
b_{IV} &= \frac{\frac{1}{n}\sum z\left(be + \left(te + e^{\perp}\right)\right)}{\frac{1}{n}\sum ze} = b + t + \frac{\frac{1}{n}\sum ze^{\perp}}{\frac{1}{n}\sum ze} = b + t + \frac{s_{e^{\perp}}}{s_e} \cdot Cauchy, \\
b_{ols} &= \frac{\frac{1}{n}\sum\left(\left(be + \left(te + e^{\perp}\right)\right)\right)e}{\frac{1}{n}\sum e^2} = b + t + \frac{\frac{1}{n}\sum ee^{\perp}}{\frac{1}{n}\sum e^2} \approx b + t.
\end{aligned}
\tag{22}
$$

The IV estimator is $b + t$ plus a multiple of a Cauchy random variable since the fraction in equation (22) is the ratio of uncorrelated mean zero normals.[10] The OLS estimator is also $b + t$ plus a random

[10] Phillips (1989) and Staiger and Stock (1997) show this result for the general k-RHS variable, completely unidentified case.

variable which converges quickly to zero. In particular, in this unidentified case the median of IV and the median of OLS are approximately the same, $b + t$. For this reason Hausman-Wu endogeneity tests, which look at the difference between IV and OLS parameter estimates, are likely to be of low power (see Staiger and Stock (1997)).

In the absence of endogeneity, $t = 0$, both IV and OLS are approximately median unbiased, although OLS would obviously be the preferred estimator. It is more interesting to study IV when there is endogeneity. Consider the case of maximum endogeneity, $\text{var}\left(e^{\perp}\right) = 0$. Here, both IV and OLS collapse on $b + t$, so b_{IV} and b_{OLS} are both very tightly distributed around the biased estimate. When this happens, the residuals also converge to zero as $y = v = be + u = be + te$, $x = e$, and $y - b_{IV}x = \left(be + te\right) - \left(b + t\right)e = 0$. Thus the reported asymptotic standard errors are spuriously small.

5.2 Test statistics in the simple case

When $q = 1$, $k = 1$ the various test statistics are amenable to analysis in closed form. In this section we make three points. First, the (squares of the) *S*-statistic and the *t*-statistic can both be written as nonlinear Wald tests of the same hypothesis on the reduced form coefficients. The two differ by a scaling factor, in that computation of the *S*-statistic essentially undoes a division-by-zero problem, and differ further by the fact that the *S*-statistic satisfies Dufour's (1997) requirement for unbounded confidence intervals where the *t*-statistic does not. Second, the *S*-statistic is closely related to the Anderson-Rubin statistic. Third, we show that the *S*-statistic has an exact finite sample *t*- distribution (conditional on *z* and assuming normal errors).

First some intuition on the division-by-zero issue. In the simple case the IV estimator is simply the ratio of $\hat{q}$ to $\hat{g}$; and other relevant computations are $\hat{\Delta} = \hat{g}\sqrt{\sum z^2}$, $\hat{\Delta} \Big/ \sqrt{\text{var}\left(\hat{\Delta}\right)} = \hat{g} \Big/ \sqrt{\text{var}\left(\hat{g}\right)}$, and $\Psi = \left(b^0\hat{g} - \hat{q}\right)\sqrt{\sum z^2}$.[11] When g is close to zero, or more to the point when much of the probability mass of $\hat{g}$ is close to zero, then b_{IV} is a fraction, $\hat{q}/\hat{g}$, whose denominator is close to zero. Division by zero has undesirable side effects, including violation of the usually innocuous regularity conditions needed for asymptotic theory. In essence Ψ is computed by multiplying the numerator of the asymptotic *t*, $b^0 - \hat{q} \Big/ \hat{g}$, by Δ to achieve just the desired scaling, $b^0\hat{g} - \hat{q}$, to eliminate the division-by-zero problem and give a well-behaved statistic. Note further that the statistic for testing for identification, $\hat{\Delta} \Big/ \sqrt{\text{var}\left(\hat{\Delta}\right)}$, is also the statistic for how far the denominator of the IV estimator is from zero, $\hat{g} \Big/ \sqrt{\text{var}\left(\hat{g}\right)}$, and is also the standard test for the significance of the first stage regression. The *S*-statistic is unbounded when $\hat{\Delta} \Big/ \sqrt{\text{var}\left(\hat{\Delta}\right)}$ is small, meeting Dufour's criteria for unbounded Wald tests exactly when much of the mass of the denominator of the IV estimator is close to zero.

The *S*-statistic and the asymptotic *t* test alternative normalizations, $b^0 g - q = 0$ versus $b^0 - q \Big/ g = 0$, of the same hypothesis. It may be useful to see that these differing normalizations account for the different behavior of the two statistics.[12] In what follows it will be convenient to note two different computations of residuals and residual variances, based on b_{IV} and on the Anderson-Rubin

[11] Note that we have passed $\hat{g}$ implicitly through the root so that unlike in the general case the sign of $\hat{\Delta}$ is determinate.

regression respectively. For the latter note that the Anderson-Rubin statistic tests $a = 0$ in the regression $y - b_0 x = az + u_{ar}$. The estimated AR coefficient is $\hat{a} = \frac{\sum (y - b_0 x) z}{\sum z^2} = \hat{q} - b_0 \hat{g}$. Define

$$\begin{aligned} \hat{u}_{IV} &\equiv y - b_{IV} x = \left(\hat{q} z + \hat{v}\right) - b_{IV} \left(\hat{g} z + \hat{e}\right) = \hat{v} - b_{IV} \hat{e} \\ \hat{s}^2_{u_{IV}} &= b_{IV}{}^2 \hat{s}^2_e - 2 b_{IV} \hat{s}_{ve} + \hat{s}^2_v \end{aligned}, \tag{23}$$

$$\begin{aligned} \hat{u}_{ar} &\equiv y - b_0 x - \hat{a} z = y - b_0 x - \left(\hat{q} - b_0 \hat{g}\right) z = y - \hat{q} z - b_0 \left(x - \hat{g} z\right) = \hat{v} - b_0 \hat{e} \\ \hat{s}^2_{u_{ar}} &= b^{0^2} \hat{s}^2_e - 2 b^0 \hat{s}_{ve} + \hat{s}^2_v \end{aligned}. \tag{24}$$

We begin with the *S*-statistic. We have

$$\Psi \equiv \Delta \left(b^0 - b_{IV}\right) = \hat{g} \sqrt{\sum z^2} \left(b^0 - \hat{q} / \hat{g}\right) = \left(b^0 \hat{g} - \hat{q}\right) \sqrt{\sum z^2} \tag{25}$$

To compute the denominator of *S* we need the partial derivatives of Ψ w.r.t. the reduced form coefficients $\hat{l} = \left[\hat{q} \quad \hat{g}\right]'$. The derivative is $\partial \Psi / \partial \hat{l} = \sqrt{\sum z^2} \cdot \left[-1 \quad b^0\right]'$, so

$$\hat{s}^2_\Psi = \sqrt{\sum z^2} \cdot \left[-1 \quad b^0\right] \frac{1}{\sum z^2} \begin{bmatrix} s^2_v & s_{ve} \\ s_{ve} & s^2_e \end{bmatrix} \begin{bmatrix} -1 \\ b^0 \end{bmatrix} \cdot \sqrt{\sum z^2} = b^{0^2} \hat{s}^2_e - 2 b^0 \hat{s}_{ve} + \hat{s}^2_v = \hat{s}^2_{u_{ar}}. \tag{26}$$

Note that the denominator of the *S*-statistic is $\sqrt{s^2_{u_{ar}}}$ so we can rewrite the *S*-statistic as $\left(b^0 \hat{g} - \hat{q}\right) / \sqrt{s^2_{u_{ar}} / \sum z^2}$. The *S*-statistic is simply a Wald test of a linear restriction from a particular least squares regression.

[12] As reminder the formal claims in this section apply when $q = 1$ and $k = 1$, although we hope the reader will find the intuition useful more generally.

It turns out that the asymptotic *t* can be rewritten as the reduced form nonlinear Wald statistic on $b^0 - q/g = 0$. The numerator of the *t* is $b^0 - \hat{q}/\hat{g}$. The denominator comes from the usual Taylor series approximation for the variance of b_{IV},

$$\text{var}(b_{IV}) \approx \begin{bmatrix} \frac{1}{\hat{g}} & \frac{-\hat{q}}{\hat{g}^2} \end{bmatrix} \text{cov}(\hat{q},\hat{g}) \begin{bmatrix} \frac{1}{\hat{g}} & \frac{-\hat{q}}{\hat{g}^2} \end{bmatrix}' = \frac{1}{\hat{g}^2 \sum z^2} \left[\hat{b}^2 \hat{s}_e^2 - 2\hat{b}\hat{s}_{ve} + \hat{s}_v^2 \right] = \frac{\hat{s}_{u_{IV}}^2}{\hat{g}^2 \sum z^2}. \quad (27)$$

The last term in (27) is the reported value of V_A for the usual asymptotic *t*-test, so in this special case the asymptotic statistic is exactly the reduced form nonlinear Wald statistic on $b^0 - q/g = 0$. The performance difference between the *S*-statistic and the *t*-statistic arises out of different normalizations for the nonlinear Wald test.

Finally, note that in this special case the *S*-statistic and AR statistic are identical. The numerator of *S* and the AR statistics are equal except for a constant factor, as $\hat{a}\sqrt{\sum z^2} = -\Psi$. From the standard least squares regression results the standard error of $\hat{a}$ is $\sqrt{s_{u_{ar}}^2 / \sum z^2}$. In comparing the *S* and AR the constant $\sqrt{\sum z^2}$ cancels between the numerator and denominator proving the equality of *S* and AR.

Since the *S*-statistic is a Wald test of a linear restriction from a particular least squares regression, it follows immediately that the *S*-statistic is distributed t_{n-1} (conditional on *z* and assuming normal errors). Alternatively, the *S*-statistic is distributed t_{n-1} because it equals the AR statistic.

5.3 The exact distribution of the structural coefficient and the role of uncertainty about identification

Because the $q=1,\ k=1$ model is analytically tractable, we can demonstrate just how badly the asymptotic distribution approximates the true distribution and make precise the sense in which the reported asymptotic *t* reflects a limiting distribution in which uncertainty about the rank condition disappears. In the one-by-one case, the IV estimator is the ratio of the reduced form coefficients $\hat{q}$ and $\hat{g}$. The numerator and denominator are distributed bivariate normal, assuming *z* is fixed and the errors are normal:

$$\begin{bmatrix}\hat{q}\\ \hat{g}\end{bmatrix} \sim N\left(\begin{bmatrix}bg\\ g\end{bmatrix}, \begin{bmatrix}s_q^2 & r_{qg}s_q s_g\\ r_{qg}s_q s_g & s_g^2\end{bmatrix}\right), \tag{28}$$

where

$$\begin{bmatrix}s_q^2 & r_{qg}s_q s_g\\ r_{qg}s_q s_g & s_g^2\end{bmatrix} = \frac{1}{\sum z^2}\begin{bmatrix}b^2 s_e^2 + 2b s_{eu} + s_u^2 & b s_e^2 + s_{eu}\\ b s_e^2 + s_{eu} & s_e^2\end{bmatrix} = \frac{1}{\sum z^2}\begin{bmatrix}s_v^2 & s_{ve}\\ s_{ve} & s_e^2\end{bmatrix}. \tag{29}$$

Since b_{IV} is the ratio of two normal variables, its density follows directly from Hinckley (1969) who cites Fieller (1932). The pdf for b_{IV} is

$$f(b_{IV}) = \frac{b(b_{IV})\,d(b_{IV})}{\sqrt{2p}\,s_q s_g\,a(b_{IV})^3}\left\{\Phi\left(j(b_{IV})\right) - \Phi\left(-j(b_{IV})\right)\right\} + \frac{\sqrt{1-r_{qg}^2}}{p\,s_q s_g\,a(b_{IV})^2}\exp\left(\frac{-c}{2\left(1-r_{qg}^2\right)}\right), \tag{30}$$

where

$$a(b_{IV}) = \left(\frac{b_{IV}^2}{s_q^2} - \frac{2r_{qg}b_{IV}}{s_q s_g} + \frac{1}{s_g^2}\right)^{0.5}, \qquad b(b_{IV}) = \frac{bgb_{IV}}{s_q^2} - \frac{r_{qg}\left(bg + gb_{IV}\right)}{s_q s_g} + \frac{g}{s_g^2}.$$

$$c = \frac{(bg)^2}{s_q^2} - \frac{2r_{qg}bg^2}{s_q s_g} + \frac{g^2}{s_g^2}, \qquad d(b_{IV}) = \exp\left\{\frac{b(b_{IV})^2 - ca(b_{IV})^2}{2(1-r_{qg}^2)a(b_{IV})^2}\right\}, \quad (31)$$

$$j(b_{IV}) = \frac{b(b_{IV})}{a(b_{IV})\sqrt{1-r_{qg}^2}}.$$

One might remark that, in general, equation (30) does not look much like a bell curve.[13]

We can now make more precise the sense in which the asymptotic *t*- describes the limiting distribution of b_{IV} as identification becomes certain. Hinckley (1969) looks at the cdf corresponding to the pdf in (30) and shows

$$\lim_{g/s_g \to \infty} F(b_{IV}) = \Phi\left(\frac{gb_{IV} - gb}{s_q s_g a(b_{IV})}\right). \quad (32)$$

The denominator on the right of (32) is

$$\left(b_{IV}^2 s_g^2 - 2r_{qg}s_q s_g b_{IV} + s_q^2\right)^{0.5} = \left(b_{IV}^2 s_e^2 - 2s_{ve}b_{IV} + s_v^2\right)^{0.5}\left(\sum z^2\right)^{-0.5}. \quad (33)$$

Evaluating this denominator at the estimated parameters, see equation (23), gives $\left(\hat{s}^2_{u_{b_{IV}}} / \sum z^2\right)^{0.5}$. So evaluating the right-hand side of (32) at the estimated parameters gives

$$\Phi\left(\frac{b_{IV} - b}{\left(\hat{s}^2_{u_{b_{IV}}} / \hat{g}^2 \sum z^2\right)^{0.5}}\right) \quad (34)$$

[13] The pdf described by equation (30) can be bimodal. With very weak instruments and high endogeneity the density has two modes which collapse around the point of concentration of the IV estimator.

which is the standard asymptotic result. Thus the standard asymptotic result is equivalent to evaluating the true distribution at the estimated parameters under the assumption that identification is known to hold with certainty.

5.4 The exact distribution of $\hat{\Delta}^2$

Computation of the confidence region for the *S*-statistic depends on the *Z*-score for the identification statistic, $\hat{\Delta}\Big/\sqrt{\text{var}\left(\hat{\Delta}\right)}$. Furthermore, when identification fails completely the distribution of b_{IV} collapses to a point so the distribution of *S* is essentially the same as the distribution of $\hat{\Delta}$. We show here that $\hat{\Delta}^2$ obeys a noncentral c^2 distribution. When the model is not identified the distribution of $\hat{\Delta}^2$ simplifies to a central $c^2(q)$. When the model is identified, in contrast, $\hat{\Delta}^2$ is asymptotically normal.

The precise statement is about the distribution of

$$\frac{\hat{\Delta}^2}{s_e^2} = \hat{\Gamma}'\left(\frac{Z'Z}{s_e^2}\right)\hat{\Gamma}. \tag{35}$$

A quadratic form $w'Aw$ in normal variates $w \sim N(m,V)$ is noncentral $c^{2'}\left(rank(AV), m'Am\right)$ if and only if the product of the weighting matrix and the variance-covariance matrix is idempotent.[14] Here $\hat{\Gamma} \sim N\left(\Gamma, s_e^2\left(Z'Z\right)^{-1}\right)$ and the product $\left(\frac{Z'Z}{s_e^2}\right)\cdot s_e^2\left(Z'Z\right)^{-1} = I_q$ is idempotent so, assuming *Z* is of rank *q*

[14] Searle (1971), p. 57.

$$\frac{\hat{\Delta}^2}{s_e^2} \sim c^{2'}\left(q, \Gamma'\left(\frac{Z'Z}{s_e^2}\right)\Gamma\right). \tag{36}$$

If $\Gamma = 0$ - so the model is not identified - the noncentrality parameter is zero and $\hat{\Delta}^2$ is (proportional to) a central $c^2(q)$. If $\Gamma \neq 0$ - so the model is identified - the noncentrality parameter grows without limit as the sample size increases.[15] The limit of a noncentral c^2 as the noncentrality parameter goes to infinity is a normal distribution,[16] so

$$\hat{\Delta}^2 \xrightarrow{d} N\left(qs_e^2 + n\Gamma' \operatorname{plim}\left(\tfrac{1}{n}Z'Z\right)\Gamma, 2s_e^2\left(qs_e^2 + n2\Gamma' \operatorname{plim}\left(\tfrac{1}{n}Z'Z\right)\Gamma\right)\right). \tag{37}$$

6 Monte Carlo Results and An Alternative Distributional Assumption

If the model is identified, then the square of the *S*-statistic is asymptotically $c^2(1)$. In the previous section we showed that when the model is not identified $\hat{\Delta}^2$ is $c^2(q)$ for $k = 1$. In this section we argue that $c^2(q-k+1)$ is a good approximation to the distribution of S^2 for the general model with *k* right hand side endogenous variables in the unidentified case. We then present Monte Carlo results using both the normal and the c^2 distributions.

6.1 An Alternative Distribution

Returning to Shea's results, it is useful to write $\hat{\Delta}^2$ as

$$\hat{\Delta}^2 = 1\Big/\left(\hat{Q}^{-1}\right)_{ii} = \hat{X}_i'\left(I_n - \hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\right)\hat{X}_i. \tag{38}$$

[15] Note that there are several conventions for describing the noncentrality parameter. We follow Johnson and Kotz (1970) with the notation that if $u \sim N(m, I_q)$, then $u'u \sim c^{2'}(q, m'm)$.

Equation (38) is a quadratic form, $\hat{X}_i \sim N\left(Z\Gamma_i, s_{e_i}^2 P_Z\right)$, and if the system is unidentified $\hat{X}_i$ has mean zero. The arguments used in section 5.4 do not hold precisely because the weighting matrix is stochastic. However, $rank\left(\left[\, I_n - \hat{X}_{\sim i}\left(\hat{X}'_{\sim i}\hat{X}_{\sim i}\right)^{-1}\hat{X}'_{\sim i}\,\right][P_Z]\right) \le q-k+1$ and Monte Carlo experiments suggest that in the unidentified case $c^2\left(q-k+1\right)$ is a reasonable, albeit slightly conservative, distribution from which to draw critical values for the *S*-statistic.

6.2 Monte Carlo Results

We present the results of two sets of Monte Carlo experiments here.[17] The first set of experiments show the relative performance of the *S*-statistic and the asymptotic *t* for a relatively good instrument. The simulations show that in a well identified model the *S*- and *t*-statistics essentially lead to the same conclusions. The second set of experiments use a weakly identified model with very strong endogeneity. Here the *S*-statistic works much better than does the *t*. We also show the results of using critical values from the $c^2\left(q-k+1\right)$ in place of $c^2\left(1\right)$.

6.2.1 Empirical performance of the *S*- and *t*-statistics for a well identified model

The Monte Carlo results presented here shows the relative performance of the *S*- and *t*-statistics over a range of sample sizes for a well identified model. Figure 2 shows the empirical rejection frequencies as the sample size varies from 100 to 10,000.

---

[16] Ibid., p. 135.

[17] The Monte Carlo designs have two RHS endogenous variables and four instruments. In each case we report results for b_1. Details of the design and further results based on a number of designs created by Dufour and Khalaf (1997) are in an appendix available from the authors.

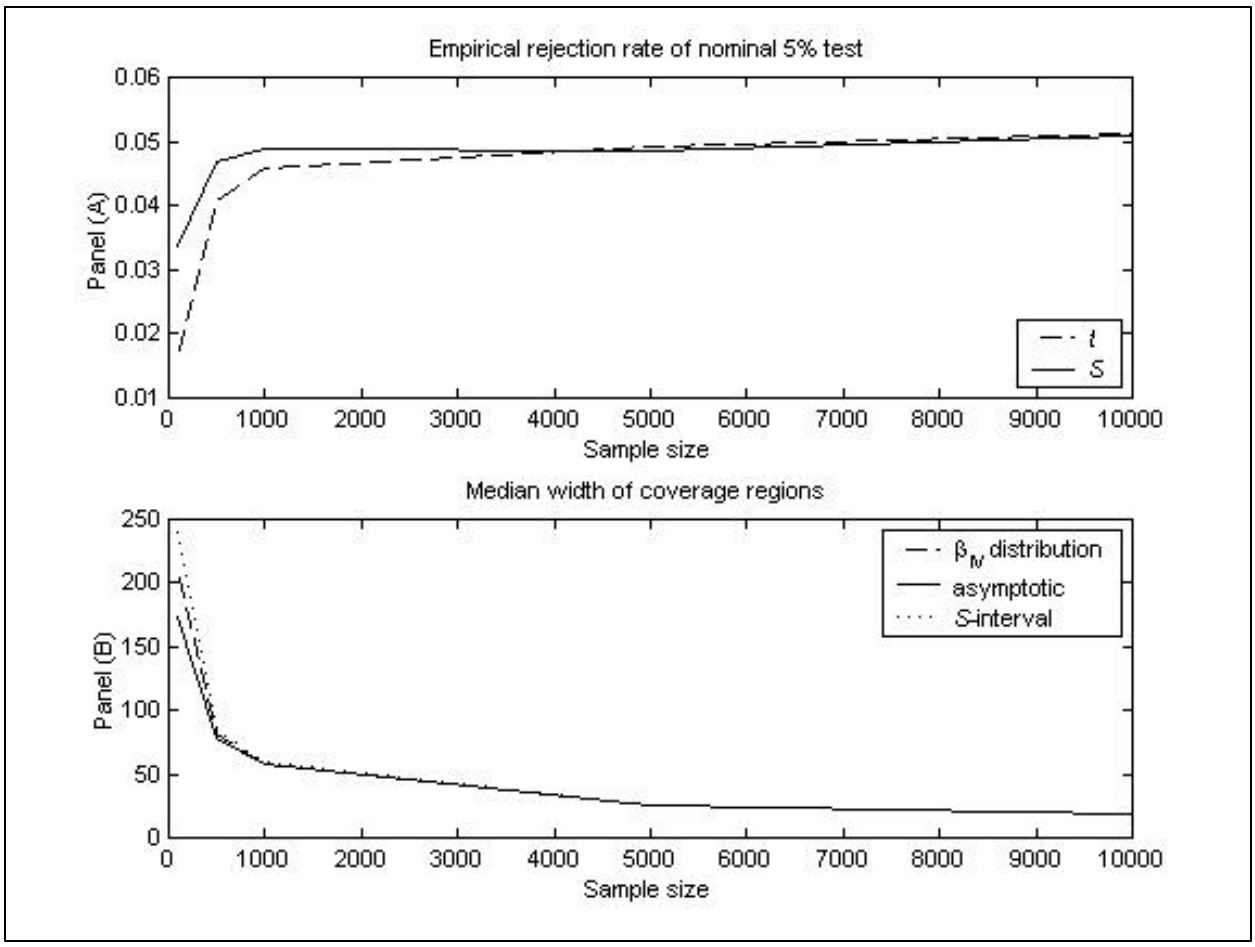

Figure 2

Panel (A) of Figure 2 shows the empirical rejection rates for the *S*- and *t*-statistics as the sample size rises from 100 to 10,000. For the design we used both statistics are a little undersized at small sample sizes and both rise to the nominal 0.05 level when the sample size reaches 1,000. Panel (B) shows three measures of dispersion. The dashed line is the distance between the 2.5th and 97.5th percentiles of b_{IV}. The solid line reports the median value of 2×1.96 times the reported asymptotic standard error. The dotted line gives the median distance between b^L and b^U for the set of closed *S*-intervals. (For $n=100$ 78 percent of the *S*-intervals are closed. All simulated *S*-intervals are closed for $n>100$.) At all sample sizes the width of *S*-intervals gives a better approximation to the width of the

actual distribution than does the width of the asymptotic confidence intervals, although by $n = 500$ the differences are quite small.

In summary, for a well-identified model the results are about the same whether one uses the *S*-statistic or the asymptotic *t*-.

6.2.2 Empirical performance of the *S*- and *t*-statistics for a weakly identified model

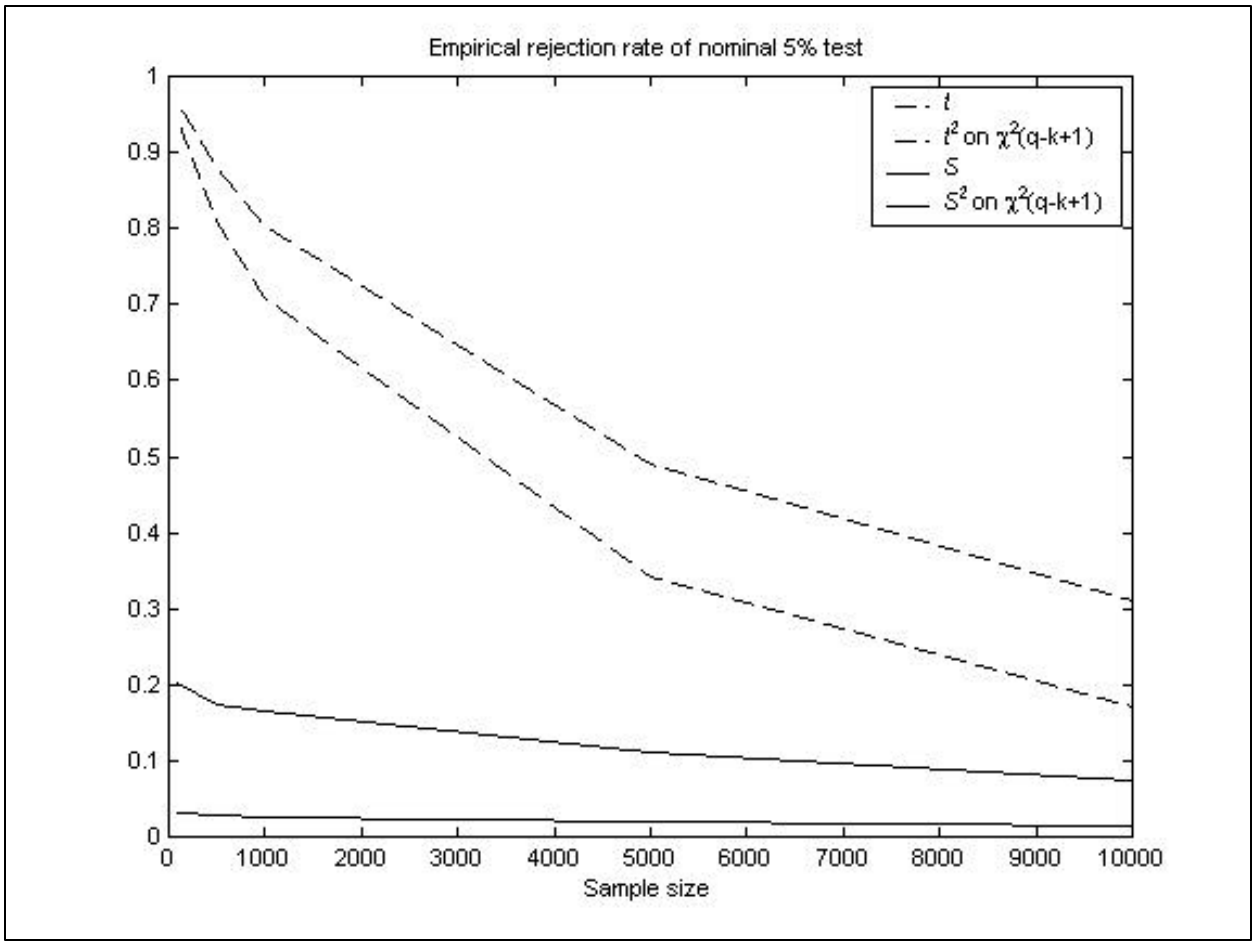

Figure 3

The solid lines in Figure 3 show empirical rejection rates for the *S*-statistic; the dashed lines for the *t*-statistic. In each case the upper line gives the empirical size with critical values drawn from the normal distribution and the lower line shows the size with critical values drawn from the more

conservative $c^2(q-k+1)$. The *t*-statistic performs abysmally, especially at "small" sample sizes, typically leading to rejection of the true value. The performance of the *S*-statistic is far superior, although it too rejects too often. Using the more conservative $c^2(q-k+1)$ critical values the actual size of the S^2-statistic is actually a little under the nominal five percent level. Using the same critical values for the square of the *t*- still leaves far too large a rejection rate.

(1)	*n*	100	500	1000	5000	10000
(2)	size *t* on normal	0.966	0.880	0.805	0.490	0.311
(3)	size t^2 on χ^2(q-k+1)	0.942	0.810	0.711	0.341	0.170
(4)	size *S* on normal	0.201	0.175	0.165	0.110	0.075
(5)	size S^2 on χ^2(q-k+1)	0.031	0.028	0.026	0.020	0.015
(6)	median β	100.251	96.892	93.216	74.293	59.720
(7)	95% coverage width	42.124	94.283	127.559	263.903	315.569
(8)	median asymp conf.	25.852	55.974	78.137	157.961	206.143
(9)	median closed interval	40.764	92.332	124.236	253.121	327.172
(10)	percent closed	0.202	0.195	0.200	0.240	0.260
(11)	size|closed	0.907	0.751	0.648	0.318	0.177
(12)	percent two rays	0.176	0.181	0.182	0.177	0.168
(13)	size|two rays	0.103	0.160	0.195	0.192	0.170
(14)	percent whole line	0.622	0.625	0.618	0.583	0.572
(15)	size|whole line	0.000	0.000	0.000	0.000	0.000
(16)	median β^L|closed	85.207	61.265	39.983	-56.745	-118.921
(17)	median β^U|closed	114.749	123.484	128.062	133.989	134.694

Table 1

The upper panel of Table 1 presents the details behind Figure 3. The lower panel provides further insight. The central tendency of the instrumental variable estimator falls from 100 to 60 as the sample size rises from 100 to 10,000 – as compared to the true value $b=1$. The next row in Table 1 shows the distance between the 2.5th and 97.5th percentiles of b_{IV}. At $n=100$, b_{IV} is tightly concentrated around a value far from the true value. With a sample size of 10,000, b_{IV} remains centered far from the true b but the distribution has spread out considerably. Line (8) reports the median value of 1.96 times the reported standard error. Comparing lines (8) and (7) shows that

reported 95 percent confidence intervals are about a third smaller than the actual distribution. The combination of the miscentering of b_{IV} with the too small intervals is responsible for the high rate of false rejections of the traditional asymptotic statistics.

S-intervals come in one of three forms: closed, the union of two rays, or the entire real line. The latter two are unbounded, which Dufour (1997) shows to be a desirable characteristic. Line (9) of Table 1 gives the median distance between b^L and b^U for the set of closed intervals. Note that lines (9) and (7) are essentially equal, so the width of closed *S*-intervals does a good job at matching the true width of the distribution of b_{IV} for this particular Monte Carlo design. Miscentering nonetheless leads to considerably too many rejections for closed intervals, albeit fewer than from the asymptotic statistics.

Lines (10) through (15) show the division of *S*-intervals into the three types and rejection rates within each type. As the sample size grows there is, as one would expect, an increase in the fraction of closed intervals and a decrease in the rejection rate among those closed intervals. In this very weakly identified model the majority of intervals are unbounded. Even at a sample size of 10,000, the majority of 95 percent *S*-intervals cover the entire real line – which, in light of the absence of much connection between b_{IV} and b, is probably the sensible conclusion.

7 Conclusion

We offer the *S*-statistic as an alternative to the traditional asymptotic *t* for tests and confidence intervals for individual coefficients estimated by instrumental variables. The "*S*-statistic" produces confidence regions based on a joint test of the structural hypothesis and the identification condition. The *S*-statistic converges to the usual asymptotic statistic as identification becomes certain, has much better

size properties when the instruments are weak, and may be inverted in closed form to conveniently compute confidence intervals. We recommend that confidence regions based on the *S*-statistic be reported in addition to or in place of the traditional statistics.

8 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.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of the Instrumental Variables Estimators," *Econometrica*, Vol. 62, No. 3, pp. 657-681.

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. of American Statistical Association*, vol. 90, 443-450.

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

_____ and Jasiac, J. (1996): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," working paper, Université de Montréal.

_____ and Khalaf, L. (1997): "Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations," working paper, Université de Montréal.

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

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.

Godfrey, L.G. (1999), "Instrument Relevance in Multivariate Linear Models," *Review of Economics and Statistics*, LXXXI, No. 3, 550-52.

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

Hahn, Jinyong and Hausman, J.A. (1999), "A New Specification Test for the Validity of Instrumental Variables," working paper, M.I.T.

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

Johnson, N.L., and Kotz, S. (1970): *Distributions in Statistics: Continuous Univariate Statistics – 2*, John Wiley & Sons, New York.

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

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.

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.

Searle, S.R. (1971): *Linear Models,* John Wiley & Sons, New York.

Shea, J., (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.

Stock, J. H. and Wright, J. (1999), "GMM with Weak Identification," *Econometrica,* forthcoming.

Wang, J. and Zivot, E., (1998): "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica* , Vol. 66, No. 6, 1389-1404.

Zivot, E., Startz, R., and Nelson, C.R. (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, Vol. 39, No. 4, pp. 1119-1144.

Appendix – Not for publication

Here we collect some results that, while memorable, are probably not suited for a journal which faces page constraints.

A1. Relating Shea's statistic to ours

The numerator of Shea's partial R^2 is $\breve{X}_i'\breve{\hat{X}}_i$. Following the argument in Shea, particularly his equation (6), we can write

$$\breve{X}_i = \left[I_n - X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1} X_{\sim i}' \right] X_i$$

$$\breve{\hat{X}}_i = \left[I_n - \hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1} \hat{X}_{\sim i}' \right] \hat{X}_i$$

Thus the numerator of Shea's partial R^2 is

$$\begin{aligned}
\breve{X}_i'\breve{\hat{X}}_i &= X_i'\left[I_n - X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1} X_{\sim i}' \right]\left[I_n - \hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1} \hat{X}_{\sim i}' \right]\hat{X}_i \\
&= X_i'\hat{X}_i - X_i'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i - X_i'X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1}X_{\sim i}'\hat{X}_i \\
&\quad + X_i'X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1}X_{\sim i}'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i
\end{aligned}$$

For any subscript j, $X_j'\hat{X}_k = X_j'P_Z X_k = X_j'P_Z'P_Z X_k = \hat{X}_j'\hat{X}_k$, so we can rewrite the numerator as

$$\begin{aligned}
\breve{X}_i'\breve{\hat{X}}_i &= \hat{X}_i'\hat{X}_i - \hat{X}_i'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i - X_i'X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i \\
&\quad + X_i'X_{\sim i}\left(X_{\sim i}'X_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i \\
&= \hat{X}_i'\hat{X}_i - \hat{X}_i'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i
\end{aligned}$$

Now w.o.l.g. partition $\hat{X} = \begin{bmatrix} \hat{X}_i & \hat{X}_{\sim i} \end{bmatrix}$ so that $Q = \hat{X}'\hat{X} = \begin{bmatrix} \hat{X}_i'\hat{X}_i & \hat{X}_i'\hat{X}_{\sim i} \\ \hat{X}_{\sim i}'\hat{X}_i & \hat{X}_{\sim i}'\hat{X}_{\sim i} \end{bmatrix}$. By the usual formula for a partitioned inverse $\left(Q^{-1}\right)_{ii} = \left(\hat{X}_i'\hat{X}_i - \hat{X}_i'\hat{X}_{\sim i}\left(\hat{X}_{\sim i}'\hat{X}_{\sim i}\right)^{-1}\hat{X}_{\sim i}'\hat{X}_i\right)^{-1}$. Thus the reciprocal of the numerator of Shea's statistic is exactly $\left(Q^{-1}\right)_{ii}$.

A2. Documentation of the Monte Carlo behind Figures 2 and 3

The Monte Carlo behind Figures 2 and 3 is for a model with very strong endogeneity. The model has four instruments and two right-hand-side variables. The reduced form coefficients equal zero except that for Figure 2 $\Gamma_{11} = \Gamma_{22} = 1$ and for Figure 3 $\Gamma_{11} = \Gamma_{22} = .01$. Results are reported for 10,000 Monte Carlo trials.

```
function Figure2_IV_99_Gen(n,nTests);
%  n is length of data series
%  nTests is number of Monte Carlo trials

randn('state',0);

q = 4 %% number of instruments
k = 2 %% two RHS for the moment, not everything works for general k
cZ = eye(q);
for i =1:q
for j=1:q
cZ(i,j)=7*(.9^abs(i-j));
end
end
cEps = eye(k);
for i =1:q
for j=1:q
cZ(i,j)=3*(.8^abs(i-j));
end
end
cEps = chol(cEps);   % used below for generating random numbers


Z = rand(n,q)*chol(cZ); %% be sure nothing magic about Z

b = ones(k,1);    %% true beta

gamma = zeros(q,k);
gamma(1,1) = .01;
gamma(2,2) = .01;

r = 100     %% used for correlation between errors
for i = 1:nTests
epsilon = randn(n,k)*cEps; % be sure nothing funny about epsilon
x = Z*gamma + epsilon;
y = x*b + randn(n,1) + r*epsilon*b;
%% Compute the statistics here
end
```

A3. Monte Carlos using the Designs of Dufour and Khalaf

Dufour and Khalaf (1997) present a series of Monte Carlo designs. We used versions of their designs which compare results with good instruments $rank(\Gamma) = k$, very weak instruments $rank(\Gamma) \to 0$, and instruments which do a good job in an individual equation but result in highly correlated fitted values from the first stage, $rank(\Gamma) \to k-1$. The designs specify two endogenous left

hand side variables, a constant, and a varying number of excluded exogenous variables In all cases only two instruments are relevant; where $q > 3$ the extra instruments have $g = 0$. We simulated the versions of the Dufour and Khalaf designs for both the asymptotic *t*- and for the *S*-statistic. Both statistics are compared to five percent critical values from the normal distribution. In all the simulations $b_1 = 10$, $n = 100$, we report statistics for the coefficient of the first endogenous variables. Results are reported for 1,000 simulations.

Design Number	median β_{IV}	std. dev. β_{IV}	empirical size *t*	empirical size *S*	# excluded exogenous	instrument quality
1	10.006	0.061	0.047	0.053	2	good
2	10.122	0.573	0.020	0.036	2	good, collinear
3	10.130	0.183	0.031	0.027	2	good, collinear
4	10.152	0.579	0.040	0.017	2	good, collinear
5	10.282	3.127	0.066	0.021	2	medium, collinear
6	11.124	0.252	0.493	0.020	2	poor, collinear
7	11.148	0.106	0.616	0.021	2	poor, collinear
8	10.127	0.138	0.066	0.047	3	good, collinear
9	10.156	0.174	0.071	0.044	3	good, collinear
10	10.203	0.502	0.107	0.032	3	good, collinear
11	10.395	0.573	0.196	0.052	3	medium, collinear
12	11.127	0.138	0.838	0.088	3	poor, collinear
13	11.155	0.078	0.896	0.088	3	poor, collinear
14	10.146	0.123	0.153	0.075	5	good, collinear
15	10.199	0.143	0.190	0.078	5	good, collinear
16	10.288	0.165	0.260	0.087	5	good, collinear
17	10.569	0.205	0.541	0.141	5	medium, collinear
18	11.134	0.149	0.983	0.322	5	poor, collinear
19	11.150	0.053	0.990	0.329	5	poor, collinear

Table A1

Table A1 confirms the expected results. For models where the asymptotic statistics work well (designs 1-5, 8, and 9), the *S*-statistic also works relatively well. For models where the asymptotic statistics perform poorly, the *S*-statistic continues to work well. The exception is models 18 and 19. Here the performance of the *S*-statistic is mediocre, although far superior to the performance of the *t*.

This is consistent with the suggestion in section 6.1 that the normal distribution is insufficiently conservative when $\Gamma = 0$.

A4 A Matlab® Program to Compute the S-Statistic

```
function [S,betaL,betaU,intervalType,deltaStat,betaHat,asympSE] = Sstat(y,X,Z,beta0,c);
%% function [S,betaL,betaU,intervalType,deltaStat,betaHat,asympSE] = Sstat(y,X,Z,beta0,c);
%% Compute S-statistic and related statistics

%% from "Improved Inference for the Instrumental Variable Estimator"
%% Startz, Nelson, and Zivot
%% April 1999

%% Input arguments:
%% y              -  LHS endogenous variable, n by 1
%% X              -  RHS variables, n by k
%% Z              -  instruments, n by q, q>=k
%% beta0          -  hypothesized values of beta (defaults to 0)
%% c              -  critical value used computing for confidence regions (defaults to 1.96)

%% Output arguments:

%% S              -  S statistic for each coefficient
%% betaL          -  lower limit of S-interval
%% betaU          -  upper limit of S-interval
%% intervalType   -  1 for closed interval, 2 for union of rays, 0 for whole line
%%                   if intervalType==0, betaL and betaU are not defined
%% deltaStat      -  delta/sqrt(var(delta)) for each coefficient
%% betaHat        -  TSLS coefficients
%% asympSE        -  asymptotic standard errors


nArgs = nargin;

if nArgs < 3
   error('At least three input arguments are required')
end

[n ky] = size(y);

if ky ~= 1
```

```
   error('y should have 1 column')
end

[nx k] = size(X);

[nz q] = size(Z);

if (n ~= nx) | (n ~= nz) | (n == 1)
   error('y, X, Z must have the same number of rows and there must be more than one data point')
end

if (q<k)
   error('Order condition fails');
end

if nArgs < 5
   c = 1.96;   %% default 5 percent size
end
if nArgs == 3
      beta0 = zeros(k,1);
end
if max(size(c) > 1
      error('c must be a scalar');
end
      [bRows bCols] = size(beta0);


if (bRows ~= k) | (bCols ~= 1)
      error('beta0 is the wrong size');
end

%% Computational note: Most time is spent in computing the product-moment
%% matrix. This should be precomputed where possible, in particular to avoid
%% order-n operations in the derivative routines

zPz = Z'*Z;
zPzInv = inv(Z'*Z);
ZPX = Z'*X;
ZPy = Z'*y;
```

```
gammaHat = zPzInv*ZPX;  %% reduced form coefficients;
      thetaHat = zPzInv*ZPy;


      %% in what follows we're going to need to Vec all the coefficients


      lambdaHat = [reshape(gammaHat,1,q*k) , thetaHat']';


Q = gammaHat'*zPz*gammaHat;
V = inv(Q);

betaHat = V*(gammaHat'*zPz*thetaHat);  %%2sls coefficients
uHat = y-X*betaHat;           %% compute residuals to get std err.
sig2U = uHat'*uHat/(n-k);
asympSE = diag(sqrt(sig2U*V));


%% compute reduced form vcov (with a little work, the order-n operations
%%    could be avoided
eps = X - Z*gammaHat;
v =   y - Z*thetaHat;
errTemp = [eps v];
vc = errTemp'*errTemp/(n-q);
jointErrCov =  kron(vc,zPzInv);

psi = PSI(lambdaHat,zPz,q,k,beta0); %% compute psi function
deriv = dFdlVec('PSI',lambdaHat,zPz,q,k,beta0);
vcov = deriv'*jointErrCov*deriv;
denom = sqrt(diag(vcov));
S = psi./denom;

delta = DELTA(lambdaHat,zPz,q,k,beta0);   %% compute delta function
derivd = dFdlVec('DELTA',lambdaHat,zPz,q,k,beta0);
vd = diag(derivd'*jointErrCov*derivd);
denomd = sqrt(vd);
deltaStat = delta./denomd;
```

```
%% Now get things for confidence intervals

phi = PHI(lambdaHat,zPz,q,k,beta0); %% compute phi function for conf. ints.
derivp = dFdlVec('PHI',lambdaHat,zPz,q,k,beta0);
vph = diag(derivp'*jointErrCov*derivp);

covdph = diag(derivp'*jointErrCov*derivd);
rHart = (-delta.*phi + c^2*covdph).^2 - (delta.^2-c^2*vd).*(phi.^2-c^2*vph);
R = sqrt(rHart);
betaL = ((delta.*phi - c^2*covdph) - R)./(delta.^2-c^2*vd);
betaU = ((delta.*phi - c^2*covdph) + R)./(delta.^2-c^2*vd);
intervalType = zeros(k,1);
intervalType = 1*(deltaStat > c) + 2*((deltaStat < c) & rHart>0);
%% end function Sstat


function p = PSI(lambdaHat,zPz,q,k,beta0);

%% return numerator of S-stat

gammaHat = reshape(lambdaHat(1:q*k),q,k);
thetaHat = reshape(lambdaHat(q*k+1:(k+1)*q),q,1);

Qinv = inv(gammaHat'*zPz*gammaHat);
bHat = Qinv*(gammaHat'*zPz*thetaHat);

p = (1./sqrt(diag(Qinv))).*(beta0-bHat);

function d = DELTA(lambdaHat,zPz,q,k,beta0);

%% return delta
gammaHat = reshape(lambdaHat(1:q*k),q,k);
Qinv = inv(gammaHat'*zPz*gammaHat);
d = (1./sqrt(diag(Qinv)));


function p = PHI(lambdaHat,zPz,q,k,beta0);

%% return phi
```

```
gammaHat = reshape(lambdaHat(1:q*k),q,k);
thetaHat = reshape(lambdaHat(q*k+1:(k+1)*q),q,1);

Qinv = inv(gammaHat'*zPz*gammaHat);
bHat = Qinv*(gammaHat'*zPz*thetaHat);

p = (1./sqrt(diag(Qinv))).*bHat;

function d = dFdlVec(F,lambdaHat,zPz,q,k,beta0);

% return numerical derivatives of vector function F

nCoefs = length(lambdaHat);   %% note that nCoefs might equal q*(k+1)
d = zeros(nCoefs,k);
F0 = feval(F,lambdaHat,zPz,q,k,beta0);
for iCoef = 1:nCoefs
   dL = .0001*lambdaHat(iCoef);
   if dL == 0
dL = .0001; %%  to handle coefficients being exactly zero
end
   newLambda = lambdaHat;
   newLambda(iCoef) = newLambda(iCoef) + dL;
d(iCoef,:) = ((feval(F,newLambda,zPz,q,k,beta0) - F0)/dL)';
end
```