Testing the Adequacy of Conventional Asymptotics in GMM

Jonathan H. Wright*

*Department of Economics, Johns Hopkins University, Baltimore MD 21202, email: wrightj@jhu.edu. I am grateful to Rob Engle, Jon Faust, Jinyong Hahn, Eric Leeper, Nour Meddahi, Matt Pritsker, David Smith, Jim Stock, the editor, and two anonymous referees for their helpful comments on earlier versions of this manuscript. All remaining errors are the sole responsibility of the author.

Abstract: This paper proposes a new test of the null hypothesis that a generalized method of moments model is sufficiently well identified for conventional asymptotics to be reliable. The idea of the test is to compare the volume of two confidence sets - one that is robust to lack of identification and one that is not. Under the null hypothesis, the relative volume of these two sets is $O_p(1)$, but under the alternative, the robust confidence set has high probability of being unbounded.

1. Introduction.

A key assumption of conventional asymptotic theory in the generalized method of moments (GMM) model (Hansen (1982)) is identification, which requires the moment condition to have a unique zero at the true parameter value and to have a gradient of full rank. Where this assumption fails, or nearly fails, conventional Gaussian asymptotic theory can provide a very poor approximation to the actual sampling distribution of estimators and test statistics. An important special case is the linear instrumental variable (IV) model, in which the identification condition requires the instruments and endogenous variables to be correlated. The many papers showing the inadequacy of conventional asymptotics when identification is weak and/or suggesting alternative asymptotic approximations include Bound, Jaeger and Baker (1995), Hansen, Heaton and Yaron (1996), Bekker (1994), Staiger and Stock (1997), Stock and Wright (2000) and Newey and Windmeijer (2009).

Fortunately, much progress has been made in the last ten years or so on developing approaches to inference in GMM that are robust to failure or near-failure of the identification condition. These approaches all give up on point estimation–consistent point estimation is of course impossible without identification. These robust methods are all instead based on constructing tests of hypotheses concerning the structural coefficient where the asymptotic distribution of the test statistic (or in some cases its exact distribution) is the same regardless of whether the model is identified or not. Confidence sets can be formed by inverting the acceptance regions of these tests. Such identification-robust tests/confidence sets have been proposed by Anderson and Rubin (1949), Stock and Wright (2000), Kleibergen (2002, 2005), Kleibergen and Mavroeidis (2008), Moreira (2003), Andrews, Moreira and Stock (2006), Andrews and Marmer (2008), Guggenberger and Smith (2005, 2008) among others. Andrews and Stock (2006) provide a recent review.

There is a strong case to be made for saying that researchers should always report only identification-robust confidence sets, giving up on point estimation. However, empirical researchers evidently prefer to report point estimates and standard errors. One reason why is because reporting the results of a confidence set for the vector of parameters formed from the inverse of the acceptance region of an identification-robust test statistic is impractical when there are more than or 2 or 3 parameters. Another reason is because conventional Gaussian inference allows us to form confidence sets for subsets of parameters–identification-robust methods do too, but robust confidence sets for subsets of parameters are asymptotically conservative. Under these circumstances, it seems helpful to have a diagnostic so as to indicate whether the identification is sufficiently strong that conventional Wald confidence sets are likely to be adequate. If the diagnostic indicates identification difficulties, the researcher should be warned to use only robust confidence sets. Otherwise, conventional point estimates and confidence intervals can be used.

A number of tests of identification have been proposed. The simplest is the first-stage F-test in the linear IV model. The null hypothesis is one of a lack of identification. Although an important and useful diagnostic, a significant first-stage F-statistic by no means implies that issues of weak instruments can be ignored (see, for example, Hall, Rudebusch and Wilcox (1996), Staiger and Stock (1997) and Stock and Yogo (2005)).[1] Stock and Yogo constructed critical values for a version of the first-stage F-test in which the null hypothesis is instead that identification is too weak for conventional asymptotics to work well. Hahn and Hausman (2002) proposed a test of the hypothesis that the linear IV model *is* identified, based on comparing forward and reverse TSLS regressions That test has however low power against the alternative of weak identification (see, for example, Hausman, Stock and Yogo (2005)). Moreover, it only applies in the linear IV model.

In this paper, I propose a new test of the hypothesis that the model is well identified, applicable in the general GMM model provided that the model has more moment conditions than parameters. The idea is to compare the volume of a Wald confidence set (not robust to identification difficulties) with the volume of a robust confidence set. Under the null that the model is identified, this ratio is $O_p(1)$. Under the alternative, the robust confidence set has high probability of being unbounded. The proposed test has non-trivial power both against the null that the model is completely unidentified and against the alternative that the identification is so weak that conventional Gaussian asymptotics works very poorly. Thus it is a test of the adequacy of conventional Gaussian asymptotics. In this regard, the motivation is similar to that of Stock and Yogo (2005). But the test proposed here is different from

[1] A computationally intensive and asymptotically conservative analog of the first-stage F-test for the GMM model was developed by Wright (2003): this is the only extant test for identification or lack of identification in the nonlinear-in-parameters context that I am aware of.

that of Stock and Yogo in that it flips the null and alternative hypotheses, and is valid in a general GMM context, not just the linear IV model.

The plan for the remainder of this paper is as follows. The GMM model is introduced in section 2. Section 3 describes the proposed test and derives its asymptotic distribution. Section 4 contains a Monte-Carlo simulation evaluating the test. Section 5 concludes.

2. The GMM Model.

The GMM model specifies that $\{Y_t\}_{t=1}^T$ is an observed time series and θ is an nx1 parameter vector with a true value θ_0, in the interior of a compact space Θ, such that

$$E(\phi(Y_t, \theta_0)) = 0$$

where $\phi(.,.)$ is a k-dimensional function, $k \geq n$. The GMM estimator of θ is

$$\hat{\theta} = \arg\min_\theta S(\theta)$$

where

$$S(\theta) = \phi^*(\theta)' W_T \phi^*(\theta),$$

$\phi^*(\theta) = [T^{-1/2}\Sigma_{t=1}^T \phi(Y_t, \theta)]$ and W_T is a symmetric positive definite kxk weighting matrix which converges almost surely to a symmetric nonstochastic $O(1)$ positive definite matrix W. Here are the standard assumptions for the GMM model:

A1: $\phi^*(\theta)$ is twice continuously differentiable, for all θ in Θ.

A2: $T^{-1}\Sigma_{t=1}^T \phi(Y_t, \theta) \rightarrow_{as} E(\phi(Y_t, \theta))$ and $T^{-1}\Sigma_{t=1}^T \frac{d\phi(Y_t,\theta)}{d\theta} \rightarrow_{as} E[\frac{d\phi(Y_t,\theta)}{d\theta}]$, uniformly in θ.

A3: $T^{-1/2}\Sigma_{t=1}^T [\phi(Y_t, \theta) - E(\phi(Y_t, \theta))] \rightarrow_d N(0, A(\theta))$, uniformly in θ, where $A(\theta)$ is 2π-times the zero-frequency spectral density matrix of $\phi(Y_t, \theta)$.

A4: The kxn matrix $B = E[\frac{d\phi(Y_t,\theta_0)}{d\theta}]$ has rank n.

A5: $E(\phi(Y_t, \theta))$ has a unique zero at $\theta = \theta_0$.

A6: $V_T(\theta)$ is an estimator of $A(\theta)$ that is consistent, uniformly in θ.

Assumptions A2 and A3 are high level convergence assumptions. Assumption A4 is the local identification assumption. Assumption A5 is the global identification assumption (Hsiao (1983)). Under these assumptions, $\hat{\theta} \rightarrow_p \theta_0$ and

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow_d N(0, (B'WB)^{-1}B'WAWB(B'WB)^{-1})$$

where $A = A(\theta_0)$. The asymptotically efficient estimator is obtained by choosing a weighting matrix such that $W = A^{-1}$; the variance of this asymptotic distribution is then $(B'A^{-1}B)^{-1}$.

One possible choice of the weighting matrix is the identity matrix. This yields the objective function

$$S_{OS}(\theta) = [T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)]'[T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)]$$

Denote the resulting estimator by $\hat{\theta}_{OS} = \arg\min_\theta S_{OS}(\theta)$. This estimator is not asymptotically efficient. A feasible asymptotically efficient estimator can be obtained by setting the weighting matrix equal to $V_T(\hat{\theta}_{OS})^{-1}$, yielding the objective function

$$S_{TS}(\theta) = [T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)]'V_T(\hat{\theta}_{OS})^{-1}[T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)].$$

Denote the resulting estimator, called the two-step estimator, by $\hat{\theta}_{TS} = \arg\min_\theta S_{TS}(\theta)$. Another feasible asymptotically efficient estimator can be obtained by setting the weighting matrix equal to $V_T(\theta)$, yielding the objective function

$$S_{CU}(\theta) = [T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)]'V_T(\theta)^{-1}[T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta)].$$

Denote the resulting estimator, called the continuous-updating estimator, by $\hat{\theta}_{CU} = \arg\min_\theta S_{CU}(\theta)$. This estimator was proposed by Hansen, Heaton and Yaron (1996). If $k = n$, the two-step and continuous-updating estimators are numerically equivalent.

In empirically relevant sample sizes, the above asymptotic theory often works poorly, as $\hat{\theta}_{TS}$ and $\hat{\theta}_{CU}$ are frequently biased and have sampling distributions far from those predicted by this asymptotic theory. These problems, documented in numerous Monte-Carlo studies, could arise because $E(\phi(Y_t, \theta))$ is zero, or close to zero, even for $\theta \neq \theta_0$ - in violation of assumptions A4 and A5.

However, a simple remedy for this problem is to use the fact that if θ_0 is the true parameter value then assumptions A3 and A6 alone are sufficient to ensure that $S_{CU}(\theta_0)$ converges to a $\chi^2(k)$ distribution. No identification assumption is required. A confidence set for θ can then be formed as the inverse of the acceptance region of this test, i.e. the confidence set of coverage 1-α is $S^*_\theta(\alpha) = \{\theta : S_{CU}(\theta) \leq F_{\chi^2}(k, \alpha)\}$ where $F_{\chi^2}(a, b)$ is the $100b$ percentile of a $\chi^2(a)$ distribution. In a completely unidentified model ($E(\phi(Y_t, \theta)) = 0$, uniformly in θ) or a locally asymptotically underidentified model ($E(\phi(Y_t, \theta)) = O(T^{-1/2})$, uniformly in θ), such a confidence set will have infinite expected volume. But this is the correct statement of our uncertainty about θ in the presence of weak identification (Dufour (1997)). This confidence set is known as the S-set and was proposed by Stock and Wright (2000). In the homoskedastic linear IV model, it reduces to the confidence set of Anderson and Rubin (1949).[2] Many other identification-robust confidence sets exist, some of which were listed in the introduction. They are all based on asymptotically pivotal test statistics where the limiting distribution is either $\chi^2(k)$ or $\chi^2(n)$.

3. The Proposed Test.

This paper proposes a test of the null hypothesis that the model is well identified in the GMM context so long as $k > n$.

Define W_1 as the maximum distance between any two points in the robust S-set, i.e. $W_1 = \sup_{\theta_1,\theta_2} ||\theta_1 - \theta_2||$ such that $S_{CU}(\theta_i) \leq F_{\chi^2}(k, \alpha)$ for $i = 1, 2$, where $||.||$ denotes the L_2-norm. If the S-set is empty, define W_1 to be zero. If it is unbounded, define W_1 to be infinity. The computation of W_1 simplifies in the linear IV model because there is an analytical expression for the Anderson-Rubin (AR) confidence set in this case (Dufour and Taamouti (2005)). Indeed, the AR confidence set reduces to the solution to a quadratic equation in the case $n = 1$.

Likewise define W_2 as the maximum distance between any two points in the usual two-step GMM Wald confidence set for θ, i.e $W_2 = \sup_{\theta_1,\theta_2} ||\theta_1 - \theta_2||$ such that $T(\hat{\theta}_{TS} - \theta_i)'\hat{B}'\hat{A}^{-1}\hat{B}(\hat{\theta}_{TS} - \theta_i) \leq F_{\chi^2}(n, \alpha)$ for $i = 1, 2$, where $\hat{A}$ and $\hat{B}$ are consistent estimators of A and B, respectively. The numerical computation of W_2 is simple, as $W_2 = \frac{2}{\sqrt{T}}\sqrt{\frac{F_{\chi^2}(n,\alpha)}{\hat{\lambda}}}$, where $\hat{\lambda}$ denotes the smallest eigenvalue of $\hat{B}'\hat{A}^{-1}\hat{B}$.

---

[2] Anderson and Rubin assumed normality. Making this additional assumption, their test statistic has an exact F distribution. However, in this paper, we do not assume normality, and so view the Anderson-Rubin test as an asymptotic test.

I refer to W_1 and W_2 as the volumes of the robust S-set and Wald confidence set, respectively. The test statistic that I propose is the ratio of these two volumes,

$$L = W_1/W_2$$

The limiting distributions of L under assumptions A1-A6 is provided in Theorem 1. Proofs of the Theorems are in the appendix.

Theorem 1: Under assumptions A1-A6, if $k > n$,

$$L \rightarrow_d L^* = \sqrt{\frac{F_{\chi^2}(k,\alpha)-\omega}{F_{\chi^2}(n,\alpha)}} 1(F_{\chi^2}(k,\alpha) - \omega \geq 0)$$

where ω is a $\chi^2(k-n)$ random variable.

The null limiting distribution L^* has point mass at zero and nonnegative support as

$$P(L^* = 0) = 1 - P(\omega \leq F_{\chi^2}(k,\alpha))$$

$$P(L^* \leq x) = 1 - P(\omega \leq F_{\chi^2}(k,\alpha) - F_{\chi^2}(n,\alpha)x^2),\ 0 < x < \sqrt{F_{\chi^2}(k,\alpha)/F_{\chi^2}(n,\alpha)}$$

and

$$P(L^* \leq \sqrt{F_{\chi^2}(k,\alpha)/F_{\chi^2}(n,\alpha)}) = 1$$

The proposed test is a one-sided test which rejects the null of adequate identification for large values of L. Critical values for a 5 percent test (the 95th percentile of L^*) are tabulated in Table 1 for various values of k and n. The null limiting distribution of L is degenerate (equal to 1) if $k = n$, because the S-set and Wald confidence sets are then asymptotically equivalent. But the statement of Theorem 1 ruled out this case. The proposed test works for any coverage rate of the Wald and S-sets, α. For all numerical work in this paper, I set $\alpha = 0.95$.

Theorem 2 gives a result on the power of the proposed test.

Theorem 2: If the model is completely unidentified (i.e. assumptions A4 and A5 are not satisfied, and instead $E(\phi(Y_t, \theta)) = 0$ for all θ), then in the limit as the sample size goes to infinity, the power of the test is at least $2\alpha - 1$.

While Theorem 2 does not show that the test is consistent, its rejection rate is guaranteed to asymptote above a certain point that depends on the coverage of the confidence sets, α. For example, if $\alpha = 0.95$, i.e. the robust and Wald confidence sets have 95% nominal coverage, then the rejection rate of the test under this alternative is guaranteed to asymptote above 90% (and could of course be higher).

Zivot, Startz and Nelson (1998) prove that, in the linear IV model with $n = 1$, the AR confidence set of nominal coverage α must be unbounded in any sample in which the first-stage F-test statistic is below the α critical value of a $\chi^2(k)/k$ distribution[3]. It follows that the rejection rate of L must be no less than the acceptance rate of the usual first-stage F-test. As shall be seen in Monte-Carlo simulations below, in the linear IV model with $n = 1$, the rejection rate of the proposed test is often much greater than the acceptance rate of the usual first-stage F-test. Kleibergen and Mavroeidis (2008) show that $S_{CU}(\theta)$ at extreme values of θ can be interpreted as an identification test statistic (in the linear IV

[3]In the linear IV model for general n, Dufour and Taamouti (2005) show that a necessary and sufficient condition for the Anderson-Rubin confidence set to be bounded is that a certain matrix is positive definite. Whenever this matrix is not positive definite, the Anderson-Rubin confidence set will be unbounded, and the proposed test statistic will necessarily reject.

model, it is just the first-stage F-statistic).

The proposed test can be thought of as a Hausman specification test (Hausman (1978)) applied to confidence sets rather than point estimates.

Although L uses the Wald confidence set based on the two-step estimator as the non-robust confidence set, the same asymptotic distribution theory would apply if the Wald confidence set associated with the continuous-updating estimator were used instead. Because the continuous-updating estimator is more robust to weak identification than the two-step estimator, this however should give a less powerful test.

4. Monte-Carlo Results.

4.1 *The Linear IV Model with a Single Included Endogenous Regressor.*

In the first set of Monte-Carlo results, I focus on the linear IV model and base the experimental design on Hahn and Hausman (2002), specifying that

$$y = X\beta + u$$

$$X = Z\Pi + v$$

where y and X are Tx1 matrices of endogenous variables, Z is a Txk matrix of instruments that are independent standard normal random variables, and u and v are conformable matrices of errors such that $w_t = (u_t, v_t^{'})'$ is a vector of zero-mean Gaussian errors with variance-covariance matrix $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. I normalize β to zero and set $\Pi = (\pi, ...\pi)$. The population R^2 in the first-stage regression is $\tilde{R}_f^2 = k\pi^2/(k\pi^2 + 1)$ and measures the strength of identification, so $\pi = \sqrt{\tilde{R}_f^2/(k(1-\tilde{R}_f^2))}$. I set $\rho = 0.5$ and 0.9, k=5, 10 and 30, and $\tilde{R}_f^2$=0, 0.01, 0.1 and 0.5, ranging from no identification to quite strong identification.

Results are reported for T=100 and 1,000 in Tables 2 and 3, respectively. In each experiment, I do 1,000 replications. I calculate (i) the coverage of the Wald confidence interval for β using BTSLS (bias-adjusted TSLS which is more robust to weak instruments than TSLS without being fully robust (Donald and Newey (2001)), (ii) the coverage of the AR confidence set for β, (iii) the rejection rate of the proposed test for identification L (based on comparing TSLS-Wald and AR confidence sets) and (iv) the acceptance rate of the first-stage F-tests for the null of a lack of identification using standard critical values and the critical values of Stock and Yogo (2005)[4], which I call F_1 and F_2, respectively. All confidence sets for β have 95% nominal coverage and the tests F_1, F_2 and L all have 5% nominal size.

The model is formally identified in all the experiments except those for which $\tilde{R}_f^2 = 0$. But one would want the test L to reject if the identification is so weak that the t-statistics exhibit severe size distortions.

The effective coverage rate of the Wald confidence interval can be far below the nominal level, as is well known. Its simulated coverage can be below 20 percent, even if $\tilde{R}_f^2$ is as large as 0.1. The AR confidence set effectively circumvents this problem.

The proposed test, L, has a low rejection rate when conventional asymptotic theory

[4]Specifically these are the critical values to ensure that the effective size of a TSLS Wald test is no greater than 25% under the weak identification asymptotics of Staiger and Stock (1997).

works well, but a high rejection rate when it works poorly. The rejection rate of the proposed test is above 99% in all cases where $\tilde{R}_f^2 = 0$. So the test rejects with probability very close to 1 when instruments are totally irrelevant. The rejection rate of L is consistently above 93% in the case $\tilde{R}_f^2 = 0.01$. In contrast, the test of Hahn and Hausman (2002) has rejection rates of around 10% in many of these simulations (their Table 3). While it is true that the model is formally identified with $\tilde{R}_f^2 = 0.01$, the model cannot be said to be well identified with such a low theoretical first-stage R-squared without an enormous sample size. Viewing the test as testing for the adequacy of conventional asymptotic theory, it is a useful feature of the proposed test that it rejects in such cases.[5] The test has a much lower rejection rate if $\tilde{R}_f^2 = 0.5$. When the identification is strong, the rejection rate of L is increasing in the number of instruments. This makes sense because the limiting distribution of the Wald test is affected by a large number of strong instruments (Bekker (1994)).

The simulation results indicate that the test is useful as an indicator of the adequacy of the standard asymptotic approximation for the Wald test statistic. This motivates considering the coverage of the confidence set for that is either the BTSLS-Wald or the AR confidence set depending on the result of L, F_1 or F_2. Coverage rates for these pretest confidence sets are also reported in Tables 2 and 3. The confidence set using L as the pretest has coverage that is consistently close to 95% (never below 89%), meaning that the proposed test does

[5]Indeed, the problem of testing the composite null that $\tilde{R}_f^2 > 0$ against the point alternative $\tilde{R}_f^2 = 0$ is not well defined, in the sense that such a test must either have power equal to the size, or must fail to control the size uniformly in the parameter space under the null.

a reasonable job of assessing the adequacy of conventional asymptotic theory. The same is true for the coverage rate of the pretest confidence set using F_2. But the coverage rate of the pretest confidence set using the usual first-stage F-test as the pretest can be as low as 62%. I conclude that a researcher working with the linear IV model who has a preference for conventional point estimates and standard errors should use either L or F_2 as a pretest, using a robust confidence set if the model is found to be underidentified, and the BTSLS-Wald confidence set otherwise. An advantage of using L rather than F_2 as the pretest, is that while they both result in similar effective coverage, there are some simulations in which L gives the researcher a much higher chance of using conventional inference.

4.2 *The Consumption CAPM with CRRA Preferences.*

An important feature of the proposed test, L, is that it is applicable in all GMM settings, not just in the homoskedastic linear IV model, unlike F_1, or F_2, or the test of a null of identification proposed by Hahn and Hausman (2002). The second set of Monte-Carlo results evaluates the proposed test in the consumption CAPM with CRRA preferences.

I simulated data from the consumption CAPM with a discount factor, δ, of 0.97 and a coefficient of risk-aversion, γ, of 1.3, following the approach of Tauchen and Hussey (1991)[6]. I then consider GMM estimation of the parameters δ and γ, using both stock and bond returns

[6]This involves fitting a 16-state Markov chain to consumption and stock-market dividend growth caliabrated to approximate the VAR in Kocherlakota (1990). Taking random draws from this Markov chain, numerical quadrature is then used to calculate the prices of a stock and a riskfree asset implied by the consumption CAPM. I am grateful to George Tauchen for his Gauss code for implementing this.

as the test asset returns, and using as instruments either instrument set A: a constant, one lag of stock and bond returns and one lag of consumption growth, or instrument set B: a constant and one lag of consumption growth[7]. These instrument sets were used by Hansen, Heaton and Yaron (1996). The sample sizes considered are 50, 100, 250 and 1000.

For these simulations, I report the coverage rate of two confidence sets for $\theta = (\delta, \gamma)'$: a Wald set centered around the continuous-updating GMM estimator and the S-set. I also report the rejection rate of the proposed test, L. Lastly, I report the coverage rate of the confidence set for θ that is the S-set if L rejects and the Wald set otherwise. Computing the numerator of the proposed test statistic, W_1, is harder than in the linear IV model but can be done by grid search.[8]

The results are given in Table 4. In sample sizes of 50 and 100, the effective coverage rates of the Wald confidence sets are far below the nominal level for both instrument sets. The effective coverage rate of the Wald confidence sets rises with the sample size. The L test has a rejection rates over 95% in sample sizes of 50 and 100, but its rejection rate falls to below 20% in the sample size of 1,000 when the Wald confidence set fares quite well. The confidence set that is the S-set if L rejects and the Wald set otherwise yields coverage of at least 85% in a sample size of 50, and a least 90% with the larger sample sizes.

---

[7] In GMM estimation of the consumption CAPM, I set $V_T(\theta) = T^{-1}\Sigma_{t=1}^{T}\phi(Y_t, \theta)\phi(Y_t, \theta)'$.

[8] The parameter space in each simulation is bounded between the wider of two possible limits: (i) δ between 0.5 and 1.5 and γ between -5 and 60, and (ii) the two-step estimates +/- 30 standard errors. Taking wider bounds could only ever increase W_1, and so make the test statistic $L = W_1/W_2$ more likely to reject.

5. Conclusion.

In this paper, I have proposed a test of the null hypothesis of identification, or more precisely of the adequacy of conventional asymptotic confidence sets. It applies in any GMM model with more moment conditions than parameters. The test is conceptually simple, working by comparing the volume of confidence sets that are robust to underidentification with the volume of the non-robust Wald confidence set. In Monte-Carlo simulations, I evaluated a pretesting strategy of using a Wald confidence set if the proposed test of identification accepts and a fully robust confidence set that gives up on point estimation otherwise. This pretesting strategy has good overall coverage properties, and allows the researcher to use conventional point estimates and standard errors in circumstances where this would not be misleading.

Appendix: Proof of Theorems.

Proof of Theorem 1: Using a second order Taylor series expansion, the continuous-updating objective function can be decomposed into the sum of the Hansen J-test and a Wald-type test statistic as:

$$\begin{aligned} S_{CU}(\theta) &= S_{CU}(\hat{\theta}_{CU}) + (\theta-\hat{\theta}_{CU})'\tfrac{dS_{CU}(\hat{\theta}_{CU})}{d\theta} + \tfrac{1}{2}(\theta-\hat{\theta}_{CU})'\tfrac{d^2S_{CU}(\theta^*)}{d\theta d\theta'}(\theta-\hat{\theta}_{CU}) \\ &= S_{CU}(\hat{\theta}_{CU}) + \tfrac{1}{2}(\theta-\hat{\theta}_{CU})'\tfrac{d^2S_{CU}(\theta^*)}{d\theta d\theta'}(\theta-\hat{\theta}_{CU}) \end{aligned}$$

where θ^* is on the line segment between θ and $\hat{\theta}_{CU}$. Hence $W_1 = \sup_{\theta_1,\theta_2} ||\theta_1-\theta_2||$ such that

$$\begin{aligned} &\tfrac{1}{2}(\theta_i-\hat{\theta}_{CU})'\tfrac{d^2S_{CU}(\theta^*)}{d\theta d\theta'}(\theta_i-\hat{\theta}_{CU}) \leq F_{\chi^2}(k,\alpha) - S_{CU}(\hat{\theta}_{CU}) \text{ for } i=1,2 \\ &\therefore W_1 = 2\sqrt{\tfrac{F_{\chi^2}(k,\alpha)-S_{CU}(\hat{\theta}_{CU})}{\mu/2}}1[F_{\chi^2}(k,\alpha) - S_{CU}(\hat{\theta}_{CU}) \geq 0] \end{aligned}$$

where μ is the smallest eigenvalue of $\frac{d^2 S_{CU}(\theta^*)}{d\theta d\theta'}$. Since $\frac{1}{T}\frac{d^2 S_{CU}(\theta^*)}{d\theta d\theta'} \rightarrow_p 2B'A^{-1}B$, $\frac{1}{2T}\mu \rightarrow_p \lambda$, the smallest eigenvalue of $B'A^{-1}B$. Also, $S_{CU}(\hat{\theta}_{CU}) \rightarrow_d \omega$, which is $\chi^2(k-n)$ distributed (Hansen (1982)). Putting these together, gives

$$T^{1/2}W_1 \rightarrow_d 2\sqrt{\frac{F_{\chi^2}(k,\alpha)-\omega}{\lambda}}1[F_{\chi^2}(k,\alpha)-\omega \geq 0].$$

Meanwhile, $W_2 = \frac{2}{\sqrt{T}}\sqrt{\frac{F_{\chi^2}(n,\alpha)}{\hat{\lambda}}}$, where $\hat{\lambda}$ is the smallest eigenvalue of $\hat{B}'\hat{A}^{-1}\hat{B}$, which is consistent for λ, so that $T^{1/2}W_2 \rightarrow_p 2\sqrt{\frac{F_{\chi^2}(n,\alpha)}{\lambda}}$. Combining these, using Slutsky's Theorem, under assumptions A1-A6,

$$L = \frac{T^{1/2}W_1}{T^{1/2}W_2} \rightarrow_d \sqrt{\frac{F_{\chi^2}(k,\alpha)-\omega}{F_{\chi^2}(n,\alpha)}}1[F_{\chi^2}(k,\alpha)-\omega \geq 0],$$

as required.

Proof of Theorem 2: The argument is adapted from Dufour (1997). If $E(\phi(Y_t,\theta)) = 0$ uniformly in θ, then from assumptions A3 and A6, $S_{CU}(\theta)$ has a marginal $\chi^2(k)$ distribution for any θ. Take any pair $\{\theta_1, \theta_2\}$. By the Bonferroni inequality, the probability that both will satisfy $S_{CU}(\theta_i) \leq F_{\chi^2}(k,\alpha)$ is at least $2\alpha - 1$. Therefore the S-set will be unbounded with at least this probability, and so the test will reject with probability at least $2\alpha - 1$ in the limit as the sample size goes to infinity.

References.

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

Andrews, D.W.K. and V. Marmer (2008): Exactly Distribution-free Inference in Instrumental Variables Regression with Possibly Weak Instruments, *Journal of Econometrics*, 142, pp.183-200.

Andrews, D.W.K. and M.J. Moreira and J.H. Stock (2006): Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression, *Econometrica*, 74, pp.715-752.

Andrews, D.W.K. and J. H. Stock (2006): "Inference with Weak Instruments," in R. Blundell, W.K. Newey and T. Persson, eds., Advances in Economics and Econometrics, Theory and Applications, 9th Congress of the Econometric Society, Cambridge University Press, Cambridge.

Bekker, P.A. (1994): Alternative Approximations to the Distributions of Instrumental Variable Estimators, *Econometrica*, 62, pp.657-681.

Bound, J., D.A. Jaeger and R. Baker (1995): Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak, *Journal of the American Statistical Association*, 90, pp.443-450.

Donald, S.G. and W.K. Newey (2001): Choosing the Number of Instruments, *Econometrica*, 69, pp.1161-1191.

Dufour, J.M. (1997): Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, *Econometrica*, 65, pp.1365-1387.

Dufour, J.M. and M. Taamouti (2005): Projection-based Statistical Inference in Linear Structural Models With Possibly Weak Instruments, *Econometrica*, 73, pp.1351-1365.

Guggenberger, P. and R.J. Smith (2005): Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification, *Econometric Theory*, 21, pp.667-709.

Guggenberger, P. and R.J. Smith (2008): Generalized Empirical Likelihood Tests in Time Series Models With Potential Identification Failure, *Journal of Econometrics*, 142, pp.134-161.

Hahn, J. and J.A. Hausman (2002): A New Specification Test for the Validity of Instrumental Variables, *Econometrica*, 70, pp.163-189.

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

Hausman, J.A. (1978): Specification Tests in Econometrics, *Econometrica*, 46, pp.1251-1271.

Hausman, J.A.. J.H. Stock and M. Yogo (2005): Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments, *Economics Letters*, 89, pp.332-342.

Hansen, L.P. (1982): Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, pp.1029-1054.

Hansen, L.P., J. Heaton and A. Yaron (1996): Finite Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics*, 14, pp.262-280.

Hsiao, C. (1983): "Identification", in *Handbook of Econometrics*, Vol. 1, eds. Z. Griliches and M.D. Intriligator, North-Holland, Amsterdam.

Kleibergen, F. (2002): Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, *Econometrica*, 70, pp.1781-1803.

Kleibergen, F. (2005): Testing Parameters in GMM Without Assuming That They are Identified, *Econometrica*, 73, pp.1103-1124.

Kleibergen, F. and S. Mavroeidis (2008): Inference on subsets of parameters in GMM without assuming identification, working paper.

Kocherlakota, N. (1990): On Tests of Representative Consumer Asset Pricing Models, *Journal of Monetary Economics*, 26, pp.285-304.

Moreira, M.J. (2003): A Conditional Likelihood Ratio Test for Structural Models, *Econometrica*, 71, pp.1027-1048.

Newey, W.K. and F. Windmeijer (2009): GMM Estimation with Many Weak Moment Conditions, *Econometrica*, forthcoming.

Staiger, D. and J.H. Stock (1997): Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, pp.557-586.

Stock, J.H. and J.H. Wright (2000): GMM with Weak Identification, *Econometrica*, 68, pp.1055-1096.

Stock, J.H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in D.W.K. Andrews and J.H. Stock, eds., Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, Cambridge University Press, Cambridge.

Tauchen, G. and R. Hussey (1991): Quadrature Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models, *Econometrica*, 59, pp.371-396.

Wright, J.H. (2003): Detecting Lack of Identification in GMM, *Econometric Theory*, 19, pp.322-330.

Zivot, E. R. Startz and C.R. Nelson (1998): Valid Confidence Intervals and Inference in the Presence of Weak Instruments, *International Economic Review*, 39, pp.1119-1146.

Table 1: Critical Values of L^*

k	n=1	n=2	n=3	n=4	n=5
2	1.248				
3	1.417	1.142			
4	1.542	1.252	1.102		
5	1.642	1.338	1.185	1.080	
6	1.726	1.408	1.251	1.147	1.066
7	1.799	1.469	1.307	1.202	1.123
8	1.863	1.522	1.356	1.249	1.170
9	1.922	1.569	1.398	1.289	1.210
10	1.975	1.612	1.437	1.326	1.245
11	2.024	1.652	1.472	1.358	1.277
12	2.069	1.689	1.505	1.389	1.305
13	2.112	1.723	1.535	1.417	1.332
14	2.152	1.755	1.564	1.443	1.356
15	2.190	1.786	1.591	1.467	1.379
16	2.226	1.814	1.616	1.490	1.401
17	2.260	1.842	1.640	1.512	1.421
18	2.293	1.868	1.663	1.533	1.441
19	2.324	1.893	1.685	1.553	1.459
20	2.354	1.917	1.706	1.572	1.477
21	2.383	1.940	1.726	1.590	1.494
22	2.411	1.962	1.745	1.608	1.510
23	2.438	1.984	1.764	1.625	1.526
24	2.464	2.005	1.782	1.641	1.541
25	2.489	2.025	1.800	1.657	1.556
26	2.514	2.044	1.817	1.673	1.570
27	2.537	2.063	1.833	1.688	1.584
28	2.561	2.081	1.849	1.702	1.597
29	2.583	2.099	1.865	1.716	1.610
30	2.605	2.117	1.880	1.730	1.623

Notes: This table gives the 95th percentile of L^*, the null limiting distribution of L, allowing a one-sided 5% test to be constructed, as described in the text. Entries are only given for $k > n$, as the distribution in Theorem 1 applies in this case only.

Table 2: Monte-Carlo Results: Linear IV Model

$T = 100$

$\tilde{R}^2_f$	ρ	k	Coverage		Rej Rate	Accept Rate		Pretest Coverage		
			Wald	AR	L	F_1	F_2	L	F_1	F_2
0	0.5	5	54.5	95.5	99.8	93.6	100.0	95.5	92.9	95.5
0	0.5	10	37.3	95.9	100.0	94.2	100.0	95.9	92.7	95.9
0	0.5	30	17.7	97.4	100.0	88.3	100.0	97.4	89.3	97.4
0	0.9	5	17.7	95.5	100.0	94.6	100.0	95.5	92.6	95.5
0	0.9	10	9.4	95.9	100.0	92.8	100.0	95.9	91.5	95.9
0	0.9	30	3.5	97.4	100.0	90.0	100.0	97.4	89.4	97.4
0.01	0.5	5	58.9	95.5	99.8	89.4	100.0	95.5	92.3	95.5
0.01	0.5	10	40.0	95.9	99.9	91.5	100.0	95.9	92.2	95.9
0.01	0.5	30	17.9	97.4	100.0	85.2	100.0	97.4	87.4	97.4
0.01	0.9	5	24.1	95.5	99.3	90.3	100.0	95.5	88.6	95.5
0.01	0.9	10	8.5	95.9	99.6	89.7	100.0	95.9	88.4	95.9
0.01	0.9	30	3.9	97.4	100.0	86.7	100.0	97.4	86.4	97.4
0.1	0.5	5	81.8	95.5	93.9	28.6	99.4	95.6	89.4	95.4
0.1	0.5	10	62.4	95.9	95.5	41.4	100.0	95.5	83.2	95.9
0.1	0.5	30	32.5	97.4	99.7	57.6	100.0	97.4	76.3	97.4
0.1	0.9	5	69.6	95.5	95.7	27.9	99.6	94.2	78.1	95.4
0.1	0.9	10	45.2	95.9	95.5	40.2	100.0	94.7	69.4	95.9
0.1	0.9	30	16.4	97.4	98.3	56.7	100.0	97.3	62.4	97.4
0.5	0.5	5	94.3	95.5	47.5	0.0	0.1	96.7	94.3	94.3
0.5	0.5	10	92.1	95.9	55.3	0.0	59.0	95.1	92.1	93.8
0.5	0.5	30	80.5	97.4	86.1	0.0	100.0	96.0	80.5	97.4
0.5	0.9	5	92.7	95.5	57.0	0.0	0.5	95.4	92.7	92.8
0.5	0.9	10	88.8	95.9	65.3	0.0	60.3	93.4	88.8	91.0
0.5	0.9	30	71.3	97.4	90.6	0.0	100.0	94.3	71.3	97.4

Notes: The coverage columns give the effective coverage rates of the bias-adjusted TSLS Wald and AR confidence sets (nominal coverage: 95%). The reject rate column gives the rejection rate of the proposed test of the null of identification, L, which compares TSLS Wald and AR confidence set volumes. The accept rate columns give the acceptance rates of F_1 and F_2, the tests of the null of underidentification comparing the first-stage F statistic with $\chi^2(k)/k$ critical values and the critical values of Stock and Yogo (2005) which ensure that the TSLS Wald test size is no larger than 25%. The pretest coverage columns report the effective coverage rate of the confidence set that is either the Wald or the AR confidence set depending on the tests L, F_1 and F_2. The sample size is $T = 100$ in all cases.

Table 3: Monte-Carlo Results: Linear IV Model

$T = 1000$

$\tilde{R}_f^2$	ρ	k	Coverage		Rej Rate	Accept Rate		Pretest Coverage		
			Wald	AR	L	F_1	F_2	L	F_1	F_2
0	0.5	5	53.9	95.0	100.0	94.2	100.0	95.0	92.6	95.0
0	0.5	10	37.0	94.3	99.9	94.6	100.0	94.3	90.9	94.3
0	0.5	30	20.4	95.5	99.6	95.2	100.0	95.5	92.5	95.5
0	0.9	5	19.0	95.0	100.0	94.5	100.0	95.0	92.6	95.0
0	0.9	10	8.2	94.3	99.8	94.8	100.0	94.3	92.2	94.3
0	0.9	30	3.3	95.5	99.5	94.6	100.0	95.5	93.4	95.5
0.01	0.5	5	82.0	95.0	92.9	31.2	99.9	94.8	88.6	95.0
0.01	0.5	10	62.6	94.3	95.8	45.6	100.0	94.3	82.4	94.3
0.01	0.5	30	33.4	95.5	97.6	68.7	100.0	95.0	80.3	95.5
0.01	0.9	5	69.4	95.0	93.5	31.0	99.8	92.5	78.7	95.0
0.01	0.9	10	44.2	94.3	94.4	46.1	100.0	92.8	70.2	94.3
0.01	0.9	30	16.3	95.5	96.1	68.2	100.0	95.2	70.6	95.5
0.1	0.5	5	94.6	95.0	28.6	0.0	0.0	96.2	94.6	94.6
0.1	0.5	10	93.1	94.3	32.8	0.0	42.7	95.6	93.1	93.0
0.1	0.5	30	84.2	95.5	54.6	0.0	100.0	92.5	84.2	95.5
0.1	0.9	5	92.9	95.0	39.6	0.0	0.0	94.3	92.9	92.9
0.1	0.9	10	91.3	94.3	50.0	0.0	42.4	94.4	91.3	90.3
0.1	0.9	30	75.4	95.5	73.1	0.0	100.0	88.8	75.4	95.5
0.5	0.5	5	94.8	95.0	8.4	0.0	0.0	95.3	94.8	94.8
0.5	0.5	10	95.1	94.3	8.0	0.0	0.0	95.5	95.1	95.1
0.5	0.5	30	93.9	95.5	10.0	0.0	0.0	94.7	93.9	93.9
0.5	0.9	5	94.7	95.0	9.3	0.0	0.0	95.2	94.7	94.7
0.5	0.9	10	95.2	94.3	9.0	0.0	0.0	95.5	95.2	95.2
0.5	0.9	30	93.0	95.5	13.8	0.0	0.0	94.0	93.0	93.0

Notes: As for Table 2, except that the sample size is $T = 1000$.

Table 4: Monte-Carlo Results: Consumption CAPM

		Coverage		Rej Rate	Pretest Coverage
Insts.	T	Wald	Robust S Set	L	
A	50	60.2	86.2	95.8	85.7
A	100	74.4	91.3	95.9	90.1
A	250	83.7	93.1	84.0	90.0
A	1000	91.3	95.1	16.7	91.6
B	50	73.3	93.8	98.8	93.4
B	100	79.8	93.8	97.2	93.1
B	250	88.4	94.7	85.8	90.5
B	1000	92.7	95.6	7.5	92.8

Notes: The coverage columns give the coverage rates of the robust S set and the Wald confidence sets centered around the continuous-updating (CU) GMM point estimates. The rejection rate column gives the rejection rate of the proposed test of a null of identification, L, which compares two-step Wald and S-set volumes. The "pretest coverage" column reports the effective coverage rate of the confidence set that is either the CU-Wald or S-set depending on the result of the proposed identification test.