Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 732

First Version: July 2002

This Version: November 2002

# Testing the Null of Identification in GMM

Jonathan H. Wright

Testing the Null of Identification in GMM

Jonathan H. Wright*

Abstract: This paper proposes a new test of the null hypothesis that a generalized method of moments model is identified. The test can detect local or global underidentification, and underidentification in some or all directions. 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.

Keywords: identification, robust confidence sets, weak instruments, generalized method of moments.

JEL Classifications: C12, C20, C30.

* Board of Governors of the Federal Reserve System. I am grateful to Rob Engle, Jon Faust, Eric Leeper, Nour Meddahi, Matt Pritsker, David Smith and Jim Stock for helpful comments. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any person associated with the Federal Reserve System.

# 1. Introduction.

One of the key assumptions of the standard linear instrumental variables (IV) model is that the instruments and endogenous variables are correlated. This is the identification assumption, on which the conventional asymptotic theory for the IV model depends. Indeed even if the correlation between the instruments and the endogenous variables is nonzero, but slight, then the conventional Gaussian asymptotic theory can nevertheless provide a very poor approximation to the actual sampling distribution of estimators and test statistics (see, for example, Bound, Jaeger and Baker (1995)). A large literature has considered the exact sampling distribution of the two stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators in models with nonstochastic instruments and Gaussian innovations. These exact distributions are far from the limits obtained from conventional asymptotic theory when the instruments are weak. TSLS is severely biased in the direction of the probability limit of OLS and the associated t-statistic is highly nonnormal and can even be bimodal. Recently, alternative asymptotic nestings have been proposed, which provide much better approximations to the actual sampling distribution of estimators and test statistics in the IV model. Bekker (1994) models the number of instruments as being an increasing function of the sample size. Staiger and Stock (1997) model the correlation between the instruments and endogenous variables as being local to zero.

The generalized method of moments (GMM) model (Hansen (1982)) nests the linear IV model as a special case. It is not surprising that analogous issues arise in this model. Many researchers have found that, in a wide variety of contexts, the conventional Gaussian asymptotic theory provides a poor approximation to the sampling distribution of GMM estimators and test statistics. There are many possible reasons why this could happen, but they include identification problems. The identification condition in the GMM model requires the moment condition to have a unique zero at the true parameter value and to have a full rank gradient and is crucial, just as it is in the linear IV special case.

Fortunately, approaches to inference are available that are robust to failure or near-failure of the identification condition. These robust approaches to inference do not permit precise inference on a structural coefficient that is not well identified - that is of course impossible. Robust approaches to inference consist of hypothesis tests and confidence sets that correctly reflect the lack of identification. If the instruments are completely irrelevant then a robust confidence set should contain the whole parameter space (or at least any point in it with probability equal to the coverage of the confidence set). In the context of the linear IV model with fixed instruments and Gaussian errors, Anderson and Rubin (1949) proposed an exact test of the hypothesis that the entire vector of structural coefficients takes on a specified value. An exact confidence set for the entire vector of structural coefficients can be computed as the set of hypothesized coefficient vectors for which this test does not reject. This confidence set has coverage close to the nominal level with stochastic instruments and/or nonnormal errors, even if the instruments are weak (Staiger and Stock (1997)). It is thus a robust confidence set. Stock and Wright (2000) construct a nonlinear analog of the Anderson-Rubin confidence set applicable in a general GMM model with possible identification difficulties. Other robust confidence sets were proposed by Wang and Zivot (1998), Kleibergen (2001, 2002) and Moreira (2002), and are discussed by Stock, Yogo and Wright (2002).

All of these robust confidence sets that control coverage whether the model is identified or not are formed by inverting the acceptance region of a test statistic. None of them give us point estimates - consistent point estimation is impossible without identification. There is a strong case to be made for saying that researchers should always report only robust confidence sets. However, practitioners are fond of reporting point estimates and associated standard errors. Part of this attachment may be force of habit and part of it may be that there is real value in a point estimate. In addition, robust confidence sets do not generally give us confidence sets for individual elements of the whole parameter vector, other than conservative confidence sets obtained by projection methods. Point estimates and conventional standard errors of course yield confidence intervals for individual elements of the parameter vector. For these reasons, it seems important to provide a diagnostic so as to indicate whether there is an identification problem or not. If the diagnostic indicates identification difficulties, the researcher should be warned to use only robust confidence sets. Otherwise, the researcher may rely on conventional point estimates and confidence sets.

In the linear IV model, the first-stage F-test involves running a regression of the endogenous variables on the instruments and testing the null hypothesis of the joint insignificance of the slope coefficients. 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 (2001)). A computationally intensive and asymptotically conservative analog of this test for the GMM model was developed by Wright (2002): this is the only extant test for identification or lack of identification in the nonlinear-in-parameters context that I am aware of.

The first-stage F-test tests the hypothesis that the model is *not* identified. Recently, Hahn and Hausman (2002) proposed a test of the hypothesis that the linear IV model *is* identified. In the case of a single right hand side endogenous variable, the idea is to compare the forward and reverse TSLS regressions. I am not aware of any existing test of the hypothesis that the model *is* identified that is applicable in the nonlinear-in-parameters context.

In this paper, I propose a new test of the hypothesis that the model *is* 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 S-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.

I argue that for a test of either the null of identification or of a lack of identification to be useful, it must indicate a lack of identification not only when the model is completely unidentified but also when the identification is so weak that conventional Gaussian asymptotics works very poorly. Comparing the first-stage F-statistic with $\chi^2_k/k$ critical values, where $k$ denotes the number of instruments, does not satisfy this requirement. In the context of the linear IV model, Stock and Yogo (2001), using the weak instrument asymptotics of Staiger and Stock (1997), show how to solve for the critical value of a first-stage F-statistic that ensures that the weak instrument asymptotic coverage of the TSLS Wald confidence set is no smaller than some bound, or that the asymptotic bias of TSLS is no greater than some bound. These critical values are much higher than the $\chi^2_k/k$ critical values with which the first-stage F-statistic is usually compared. Although the test of a null of identification in GMM that I propose in this paper does not have any weak identification asymptotic motivation along the lines provided by Stock and Yogo in the linear IV model, the proposed test focuses directly on the comparison of robust and non-robust confidence sets and rejects the null of proper identification whenever robust and non-robust confidence sets are very different. As such, it might be hoped that the proposed test will indicate circumstances in which identification is so weak that conventional asymptotics are seriously misleading and that a strategy of using the robust confidence set when the proposed test rejects, but using the regular Wald confidence set otherwise will control coverage quite well (while also indicating circumstances in which a researcher may safely use point estimates and conventional standard errors). This will of course be demonstrated in the Monte-Carlo simulations below.

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 properties of the proposed test. Section 5 contains two empirical applications. Section 6 concludes.

## 2. The GMM Model.

The GMM model specifies that $\{Y_t\}_{t=1}^{T}$ is an observed time series and $\theta$ is an $n$x1 parameter vector with a true value $\theta_0$, in the interior of a compact space $\Theta$, such that

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

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

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

Assumption 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 $\theta$.

Assumption A3: $T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t,\theta_0) \rightarrow_d N(0,A)$, where $A$ is $2\pi$-times the zero-frequency spectral density matrix of $\phi(Y_t,\theta_0)$.

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

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

Assumption A6: $V_T(\theta)$ is an estimator of $2\pi$-times the zero-frequency spectral density matrix of $\phi(Y_t,\theta)$ that is consistent, uniformly in $\theta$.

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, $\hat{\theta} \rightarrow_p \theta$ and

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

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.

2.1 *Problems with Standard Gaussian Asymptotics for GMM.*

The above asymptotic theory often works poorly in practice. Often, in empirically relevant sample sizes, $\hat{\theta}_{TS}$ and $\hat{\theta}_{CU}$ are biased and have sampling distributions far from those predicted by this asymptotic theory, and the associated t- and F-statistics have erratic rejection rates. These problems, documented in numerous Monte-Carlo studies, could arise because $T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t, \theta_0)$ fails to converge to normality, or converges only very slowly. Alternatively, it could be that $E(\phi(Y_t, \theta))$ is zero, or close to zero, even for $\theta \neq \theta_0$ - in violation of assumptions A4 and A5.

The focus of this paper is on problems with the asymptotic theory underlying GMM which arise from this latter source: an identification problem. Stock and Wright (2000) proposed an alternative asymptotic nesting in which $E(\phi(Y_t, \theta)) = O(T^{-1/2})$, uniformly in $\theta$. They derive an alternative asymptotic theory which nests the completely unidentified model ($E(\phi(Y_t, \theta)) = 0$, uniformly in $\theta$) as a special case. This alternative asymptotic theory works much better than the conventional Gaussian asymptotic theory in providing an approximation to the finite sample distributions of GMM estimators and test statistics when identification is weak. In the linear IV model, it reduces to the nesting proposed by Staiger and Stock (1997).

2.2 *S-sets.*

The weak identification problem in GMM may be effectively circumvented by the use of S-sets, as proposed in Stock and Wright (2000). The approach dispenses with point estimation and instead forms a confidence set for $\theta$ directly from an objective function, using a nonlinear analog of the Anderson-Rubin confidence set. If assumption A3 holds, and if $V_T(\theta_0) \rightarrow_p A$, then the continuous-updating objective function evaluated at the true parameter vector, $S_{CU}(\theta_0)$, converges to a $\chi^2$ distribution on $k$ degrees of freedom. No identification assumption (assumption A4 or A5) is required for this to hold. The confidence set for $\theta$ is formed as the inverse of the acceptance region of this test, i.e. the confidence set of coverage 1-$\alpha$ is $S_\theta^*(\alpha) = \{\theta : S_{CU}(\theta) \leq F_{\chi^2}(k,\alpha)\}$ where $F_{\chi^2}(a,b)$ is the 100$b$ percentile of a $\chi^2$ distribution on $a$ degrees of freedom. In a completely unidentified model ($E(\phi(Y_t,\theta)) = 0$, uniformly in $\theta$) or a locally asymptotically underidentified model ($E(\phi(Y_t,\theta)) = O(T^{-1/2})$, uniformly in $\theta$), such a confidence set will have infinite expected volume. But this is the correct statement of our uncertainty about $\theta$ in the presence of weak identification. More formally, under these circumstances, any confidence set that is valid (i.e. controls coverage) must have infinite expected volume (Dufour (1997)).

2.3 *The Homoskedastic Linear IV Model*

The linear IV model with iid errors specifies that

$$y = X\beta + u$$

$$X = Z\Pi + v$$

where $y$ and $X$ are $T$x1 and $T$x$n$ matrices of endogenous variables, $Z$ is a $T$x$k$ matrix of instruments ($k \geq n$), and $u$ and $v$ are conformable matrices of errors such that $w_t = (u_t, v_t^{'})'$ has variance-covariance matrix $\Omega$ (the $t$ subscript on any matrix denotes the $t$th observation for that variable). The first of these equations is the structural equation of interest, and the researcher wants to do inference on the coefficient vector $\beta$. Without loss of generality, there are no included exogenous variables in this equation. If they were present, they would just be projected out. Because $u_t$ and $v_t$ are correlated, OLS is biased.

This model is a special case of the general GMM model with $\phi(Y_t, \theta) = (y_t - X_t^{'}\beta)Z_t$, $\theta = \beta$ and $V_T(\theta) = T^{-1}\Sigma_{t=1}^T Z_t Z_t^{'} T^{-1}\Sigma_{t=1}^T (y_t - X_t^{'}\beta)^2$. Since $B = E[\frac{d\phi(Y_t,\theta_0)}{d\theta}] = E[Z_t X_t^{'}] = E[Z_t Z_t^{'}]\Pi$, the local identification condition reduces to requiring that the matrix $\Pi$ has rank $n$. If this condition is satisfied, assumption A5 is satisfied too. The TSLS estimator is

$$\hat{\beta}_{TSLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$$

which is the two-step GMM estimator. The continuous-updating GMM estimator reduces to LIML. The S-set reduces to

$$\{\beta : \tfrac{(y-X\beta)'P_Z(y-X\beta)}{(y-X\beta)'(y-X\beta)/T} \leq F_{\chi^2}(k, \alpha)\}$$

where $P_Z = Z(Z'Z)^{-1}Z'$. This is the Anderson-Rubin confidence set (Anderson and Rubin (1949)). If $n = 1$ (a single right hand side endogenous variable), the confidence set can be written as

$$\{\beta : (X'P_Z X - \frac{F_{\chi^2(k,\alpha)}}{T} X'X\}\beta^2 + 2\{\frac{F_{\chi^2(k,\alpha)}}{T} y'X - y'P_Z X\}\beta + \{y'P_Z y - \frac{F_{\chi^2(k,\alpha)}}{T} y'y \leq 0\}$$

This is a quadratic inequality the solution to which is a confidence set for $\beta$ that must be an interval, the whole real line, the complement of an interval or an empty confidence set. The Anderson-Rubin confidence set is robust to weak instruments. Staiger and Stock (1997) show that it controls coverage in their weak instrument asymptotics. The test is indeed an exact test, if the instruments are fixed and errors are Gaussian (after replacing $\chi^2$ critical values with $F$ critical values).

The first-stage F-test tests the hypothesis that $\Pi = 0$. A rejection of this hypothesis by no means implies that there are no identification difficulties. The identification may not be strong enough for conventional asymptotic theory to work well, as discussed in the introduction. There is an additional issue if $n > 1$. The identification condition requires $\Pi$ to have rank $n$, which is a stronger requirement than $\Pi \neq 0$ if $n > 1$. The case in which $\Pi$ has nonzero rank, smaller than $n$, can be called partial identification and also leads to failure of conventional asymptotic theory, as discussed by Choi and Phillips (1992). Cragg and Donald (1993) extend the first-stage F-test to test the hypothesis that $\Pi$ has a specified rank smaller than $n$. It is a convenient feature of testing the null hypothesis of identification, as I do in this paper, that the test is consistent against the alternative that $\Pi$ has any rank smaller than $n$ (or that $B$ has any rank smaller than $n$, in the general GMM case).

2.4 *Identification Problems in Other Cases.*

The linear IV model with iid errors is only one special case of the general GMM model. The GMM model can accommodate heteroskedastic and/or dependent errors. Indeed, any maximum likelihood estimator is a special case of GMM in which the moment condition is the score function.

A leading application of the GMM model is the consumption capital asset pricing model (CAPM) of Lucas (1978). With constant relative risk aversion (CRRA) preferences, the consumption CAPM Euler equation implies the GMM moment condition $E(Z_t \otimes [\delta R_{t+1}(\frac{C_{t+1}}{C_t})^{-\gamma} - 1]) = 0$ where $C_{t+1}$ and $R_{t+1}$ denote consumption and a vector of gross asset returns, respectively, $\delta$ is the discount factor, $\gamma$ is the coefficient of relative risk aversion and $Z_t$ is any variable in the information set at time $t$. This can be interpreted as a nonlinear-in-parameters instrumental variables model with $Z_t$ as the instrument vector. However, consumption growth and asset returns are notoriously hard to predict - meanwhile the identification assumptions (A4 and A5) require nonlinear functions of these to be forecastable. Weak identification is thus a prime issue in this context (see Stock and Wright (2000), Wright (2002) and Stock, Wright and Yogo (2002)). The first-stage F-statistic does not apply in the nonlinear-in-parameters GMM model.

Identification in GMM essentially requires that the objective function be locally quadratic (assumption A4) and that it does not have multiple local minima which give the same value of the population objective function (assumption A5). These conditions apply even where there are no instrumental variables. For example, a logit or probit model in which most of the dependent variables are zeros (or in which most of the dependent variables are one) will have a likelihood that is flat around the true parameter value, causing an identification problem. Likewise, the pseudo-Gaussian likelihood function for a ratio of two parameters will be flat if the denominator is close to zero. Pagan and Robertson (1998) argue that weak identification problems can arise in the structural VAR literature.

This motivates the construction of a test of the identification conditions in the general GMM model (assumptions A4 and A5).

## 3. The Proposed Test.

As discussed above, identification is a key requirement for conventional asymptotic theory. I am however only aware of one extant test for the null hypothesis that the model *is* identified. That is the test proposed by Hahn and Hausman (2002) that compares forward and reverse TSLS estimators. It applies only in the linear IV model.

This paper proposes a new test of the null that the model is identified, which applies in the general 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 robust S-set is empty, define $W_1$ to be zero.

Likewise define $W_2$ as the maximum distance between any two points in the usual two-step Wald confidence set for $\theta$, 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 $\hat{B}'\hat{A}^{-1}\hat{B}$. The numerical computation of $W_1$ is harder, but simplifies in the linear IV model with $n = 1$. In this case, there exists a closed form expression for the robust S-set, which is the solution to a quadratic equation, equation (1). If the confidence set is an interval, $W_1$ is the width of this interval. If the confidence set is empty, $W_1 = 0$. If the confidence set is the whole real line or is the complement of an interval, $W_1$ is infinite. Computation of $W_1$ in other cases is discussed in the Monte-Carlo simulations below.

I refer to $W_1$ and $W_2$ as the volumes of the robust S-set and usual two-step 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 the null of identification (assumptions A4 and A5) is provided in Theorem 1, the proof of which is given in the appendix. It involves consideration of the behavior of the two-step and continuous-updating GMM objective functions in a $T^{-1/2}$-neighborhood of the true parameter value.

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

$$L \to_d \sqrt{\frac{\phi' G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha)}{F_{\chi^2}(n,\alpha)}} 1[\phi' G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha) \geq 0]$$

where $\phi$ is a $N(0, I_k)$ $k$x1 vector and $G = A^{-1/2}B$, a matrix of order $k$x$n$.

The null limiting distribution in Theorem 1 can easily be simulated, given a consistent estimate of $G$ (which is consitently estimable under the null of identification). The proposed test is a one-sided test which rejects the null of identification for large values of $L$. In the linear IV model, if the rank of $\Pi$ is less than $n$, (or, more generally, if there exist two distinct zeros of the equation $Ef(Y_t, \theta) = 0$ that are infinitely far apart), then Dufour (1997) shows that $W_1$ is unbounded with probability of at least $2\alpha$-1, asymptotically. Meanwhile, $W_1$ is $O_p(1)$, and so the test rejects with an asymptotic probability of at least $2\alpha$-1. While the test is not necessarily consistent against this alternative, its rejection rate is guaranteed to asymptote above a certain point. 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).

The null limiting distribution of $L$ is degenerate, equal to 1, if $k = n$. But the statement of Theorem 1 ruled out this case. If $k = n$, the robust S-set is asymptotically equivalent to the Wald confidence set. This is not so in the case $k > n$, where the robust S-set wastes power relative to the Wald set, if the model really is identified. Alternative confidence sets have been proposed that are robust to weak identification. Some of these have the additional feature that they *are* asymptotically equivalent to the Wald confidence set if the model really is identified and $k > n$ (Kleibergen (2001) in the linear IV model and Kleibergen (2002) in the general GMM context). A test for a null of identification could be conducted by comparing the volume of a Wald and a robust confidence set in cases where these two are first-order asymptotically equivalent. This would however be harder as the test would be based on the second-order asymptotic difference between the two confidence sets. The fact that the robust S-set wastes power if $k > n$ greatly simplifies the derivation of an expression for the relative volume of the robust and non-robust confidence sets.

The proposed test works for any coverage rate of the Wald and S-sets, $\alpha$. For all numerical work in this paper, I set $\alpha = 0.95$.

It would actually be possible to do two-tailed tests based on $L$ as well. If the robust S-set is empty, or excessively small (relative to the Wald set), this indicates a specification problem. A two-tailed test would be a joint test of identification and specification that would have power against both lack of identification and model misspecification. Many important papers, including Bound, Jaeger and Baker (1995) have been concerned simultaneously about identification and specification. But, in this paper, I focus exclusively on testing for identification problems, and so adopt a one-tailed test.

Although the test as defined here 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[1].

Zivot, Startz and Nelson (1998) prove that, in the linear IV model with $n = 1$, the S-set (which is just the Anderson-Rubin confidence set) of nominal coverage $\alpha$ must be unbounded in any sample in which the first-stage F-test statistic is below the $\alpha$ critical value of a $\chi^2_k/k$ distribution[2]. It follows that the rejection rate of my test of the null of identification must (numerically) be no less than the acceptance rate of the usual first-stage F-test which is testing the null of a lack of identification. As shall be seen in Monte-Carlo simulations below, in the linear IV mdel with $n = 1$, the rejection rate of the proposed test is often much greater than the acceptance rate

---

[1] There are two reasons for focussing on the two-step estimator. Firstly, practitioners using a non-robust estimator use the two-step estimator (TSLS) far more frequently than they use the continuous-updating estimator (LIML). Secondly, the continuous-updating estimator is not robust to lack of identification, but still works somewhat better than the two-step estimator when identification is weak. Accordingly, it might be expected (and it turns out in simulations to be true) that a test for a lack of identification that compares the S-set with the two-step Wald confidence set is more powerful in the sense that it has a higher rejection rate than the test that compares the S-set with the continuous-updating Wald confidence set when identification is weak.

[2] In the linear IV model for general $n$, Dufour and Taamouti (2002) 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.

of the usual first-stage F-test.

The proposed test is similar in spirit to a Hausman specification test. Any test that compares two estimators one of which is consistent under the null and alternative hypotheses and the other of which is consistent under the null hypothesis alone is a Hausman specification test (Hausman (1978)). Consistent estimation is however impossible in a model that is not identified[3]. The test that I am proposing instead compares the volume of two confidence sets, which are of the same order under the null, but not under the alternative.

## 4. Monte-Carlo Results.

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

In my first set of Monte-Carlo results, I focus on the linear IV model with $n = 1$. The experimental design follows Hahn and Hausman (2002). I normalize $\beta$ to zero. The instruments are $k$ independent standard normal random variables. The errors $w_t$ are Gaussian with $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ - the parameter $\rho$ governs the endogeneity of $x_i$. I set $\Pi = (\phi, ...\phi)$. The population $R^2$ in the first-stage regression is $\tilde{R}_f^2 = k\phi^2/(k\phi^2 + 1)$, so $\phi = \sqrt{\tilde{R}_f^2/(k(1 - \tilde{R}_f^2))}$. I use the following parameter combinations:

[3] This does not necessarily prevent one from constructing a test based on the difference between two estimators neither of which is consistent under the underidentified alternative. The test of Hahn and Hausman (2002) is such a test. A test based on the difference between the continuous-updating and two-step GMM estimators would be too.

$T$=100, 250, 1000

$\rho$=-0.9, -0.5, 0.5, 0.9

$k$=5, 10, 30

Results are reported for $\tilde{R}_f^2$=0, 0.01, 0.1, 0.3 and 0.5 in Tables 1-5, respectively. In each experiment, I do 1,000 replications. I calculate (i) the coverage of the usual Wald confidence interval (which is 1 minus the size of the TSLS t-test testing the hypothesis that $\beta = 0$), (ii) the coverage of the Anderson-Rubin confidence set, (iii) the rejection rate of the proposed test $L$, (iv) the coverage of the confidence set that is the Anderson-Rubin confidence set if $L$ rejects and the Wald otherwise, (v) the acceptance rate of the first-stage F-test, (vi) the coverage of the confidence set that is the Wald confidence set if the first-stage F-test rejects and the Anderson-Rubin otherwise, (vii) the acceptance rate of the test comparing the first-stage F-test with the weak identification asymptotic critical values of Stock and Yogo (2001)[4] and (viii) the coverage of the confidence set that is the Wald confidence set if the first-stage F-test rejects using these latter critical values and the Anderson-Rubin otherwise. All confidence intervals have 95% nominal coverage and all tests have 5% nominal size. I report the acceptance rate for the first-stage F-test but the rejection rate for my proposed test. This gives some comparability, since the first-stage F-test tests a

[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).

null hypothesis of underidentification while the proposed test tests a null hypothesis of identification. However, these tests are conceptually quite distinct.

The model is formally identified in all the experiments except those for which $\tilde{R}_f^2 = 0$. But we would want the test $L$ to reject if the identification is so weak that the TSLS t-statistic exhibits severe size distortions. The potential practical usefulness of the proposed test is as a pretest as in (iv). The hope, to be evaluated in these experiments, is that the effective coverage of the confidence set that is robust if the test rejects and non-robust otherwise will generally be reasonably close to the nominal level. The researcher is however better off with this strategy than simply using the robust confidence set always in the sense that the researcher will sometimes be allowed to use point estimates and standard errors, which may be preferable for reasons discussed in the introduction.

The effective coverage rate of the Wald confidence interval can be far below the nominal level when $\rho > 0$, as is well known. In extreme cases, its simulated effective coverage is under 1 percent. The Anderson-Rubin confidence set effectively circumvents this problem. The proposed tests have low rejection rates when conventional asymptotic theory works well, but high rejection rates when it works poorly.

The rejection rate of the proposed test is above 99% in all cases where $\tilde{R}_f^2 = 0$. Although I do not have a formal proof of the test's consistency, this indicates that the test rejects with probability very close to 1 when instruments are totally irrelevant.

The rejection rate of the proposed test is consistently above 90% in the case $\tilde{R}_f^2 = 0.01$. 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 in a sample size of 1,000 or less. Since I am thinking of the test as testing for the adequacy of conventional asymptotic theory, I believe that it is a useful feature of the proposed test that it rejects in such cases[5].

The coverage rate of the confidence set that conditions on the result of the pretest proposed in this paper is never below 73% and is typically quite a bit higher than this. The coverage rate of the confidence set that conditions on the result of the first-stage F-test can be as low as 8%. It seems intuitive that the test proposed in this paper works relatively well as a pretest because it is based on the direct comparison of robust and non-robust confidence sets and only allows the researcher to use the non-robust approach when this gives results that are close to the robust approach.

The coverage rates of the confidence set that conditions on the comparison of the first-stage F-statistic with the critical values of Stock and Yogo (2001) is never

[5] In fact, 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.

below 83%[6]. There are however several cases in which the test of a null of identification accepts the null hypothesis more than 50% of the time (allowing the researcher to use conventional point estimates and standard errors), while the test that compares the first-stage F-statistic with the critical values of Stock and Yogo accepts the null of underidentification in all simulations (preventing the researcher from using conventional point estimates and standard errors).

Of course, the researcher who never wants effective coverage to be appreciably different from the nominal level should just always use the robust confidence sets.

4.2 *The Consumption CAPM with CRRA Preferences.*

An important feature of the proposed test is that it is applicable in all GMM models, not just in the linear IV model, unlike the first-stage F-test or the test of a null of identification proposed by Hahn and Hausman (2002). My second set of Monte-Carlo results evaluate the proposed test in the context of the consumption CAPM Euler equation with CRRA preferences.

To simulate data from the consumption CAPM, I follow the approach of Tauchen and Hussey (1991) (also used in Tauchen (1986), Kocherlakota (1990), Hansen, Heaton

[6] This is not surprising, since these critical values are designed to ensure that the effective coverage of the TSLS Wald confidence set is bounded below by 75% under weak instrument asymptotics. Stock and Yogo provide critical values for different maximal biases or size distortions. I picked the critical values that bound the effective size at 25% because the effective coverage rates of the confidence set for $\beta$ obtained with them are comparable with the effective coverage rates of the confidence sets that condition on the results of the test for identification proposed in this paper.

and Yaron (1996) and Stock and Wright (2000)). This involves fitting a 16-state Markov chain to consumption and stock-market dividend growth (the state variables) calibrated so as to approximate the first-order VAR:

$$\begin{pmatrix} \log(\frac{C_t}{C_{t-1}}) \\ \\ \log(\frac{D_t}{D_{t-1}}) \end{pmatrix} = \mu + \Phi \begin{pmatrix} \log(\frac{C_{t-1}}{C_{t-2}}) \\ \\ \log(\frac{D_{t-1}}{D_{t-2}}) \end{pmatrix} + \begin{pmatrix} u_{ct} \\ \\ u_{dt} \end{pmatrix}$$

where $D_t$ is the stock-market dividend at date $t$ and $(u_{ct}, u_{dt})'$ is iid normal with mean zero and variance $\Lambda$. I use the same values of $\mu$, $\Phi$ and $\Lambda$ as Kocherlakota (1990), who chose these by fitting a bivariate VAR(1) to historical US annual real dividend growth and real consumption growth data[7]. I set the discount factor, $\delta$, to 0.97 and the coefficient of relative risk aversion, $\gamma$, to 1.3. Taking random draws of consumption growth and dividend growth from this Markov chain, numerical quadrature is then used to calculate the prices of a stock and a riskfree asset in each period implied by the consumption CAPM with intertemporally separable CRRA preferences. In this way, time series of consumption growth, stock returns and bond returns may be simulated[8]. I then consider GMM estimation of the parameters $\delta$ and $\gamma$, using both stock and bond returns as the elements of $R_{t+1}$, using as instruments either instrument set A: a constant, one lag of stock and bond returns and one lag of consumption growth or

[7] These are $\mu = \begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix}$, $\Phi = \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0.0012 & 0.00177 \\ 0.00177 & 0.014 \end{pmatrix}$.

[8] I am grateful to George Tauchen for his Gauss code for implementing this.

instrument set B: a constant and one lag of consumption growth[9]. These instrument sets were used by Hansen, Heaton and Yaron (1996). I use sample sizes of 100 (mirroring the available sample sizes for U.S. annual data) and 250 and 1,000 (to see the effects of hypothetical larger sample sizes).

For these simulations, I report the coverage rate of a Wald confidence set and an S-set for $\theta = (\delta, \gamma)'$. Note that computing the coverage rate of the S-set just involves comparing the continuous-updating objective function with the $\chi^2_k$ critical value at the true parameter value. I also report the rejection rate of the proposed test, $L$. Lastly, I report the coverage rate of the confidence set for $\theta$ 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 with a single endogenous right hand side variable. I adopt the following algorithm. I first bound the parameter space to be the wider of two possible bounds: (i) a bound for $\delta$ between 0.5 and 1.5 and $\gamma$ between -5 and 60, and (ii) the two-step estimate +/- 30 standard errors. Next I take random points, uniformly from within this parameter space, evaluate the continuous updating objective function at each of these points, and save the first 1,000 points for which the objective function is below the $\chi^2_k$ critical value. These points are all in the robust S-set[10]. I then compute

---

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

[10] If after taking 100,000 random points from the bounded parameter space, none of them is

the maximum distance ($L_2$-norm) between all pairs of these 1,000 points, and take this as $W_1$. Note that taking wider bounds, or drawing more points, can only ever increase $W_1$, and so make the test statstic $L = W_1/W_2$ more likely to reject. In the case $n = 2$ (such as in this example), this algorithm is not too computationally burdensome, and is likely to provide a good approximation to $W_1$. The algorithm could in principle work for any $n$, but for large $n$ computing an accurate approximation to $W_1$ in this way would be computationally extremely costly.

The results are given in Table 6. In a sample size of 100, the effective coverage rates of the Wald confidence set are far below the nominal level for both instrument sets (beween 40 and 50%), mirroring the inadequacy of conventional asymptotic approximations in this case, as documented by Hansen, Heaton and Yaron (1996). The effective coverage rate rises with the sample size. The proposed tests have rejection rates over 90% in the sample size of 100, but have lower rejection rates in the larger sample sizes. The rejection rate for $L$ can fall to under 15% in the sample size of 1000 when the conventional asymptotic approximation is not too bad. The confidence set that is the S-set if $L$ rejects and the Wald set otherwise yields coverage of above 78% in all cases.

## 5. Empirical Applications.

---

included in the robust S-set, I conclude that it is empty and set $W_1 = 0$.

5.1 *Financial Intermediation and Growth using Legal Origin Dummies as Instruments.* One of the applications of instrumental variables methods that has received considerable attention recently has been the regression of growth rates on measures of financial development. There is a clear problem of endogeneity in this regression. Recognizing this, authors such as Levine, Loayza and Beck (2000) have used legal origin dummies as instruments. The legal system in each country can usually be unambiguously traced to one of four origins: English Common Law, French Napoleonic Law, German law, or Scandanavian law. It is hoped that the legal origin dummies are correlated with financial development, but do not affect growth rates other than through financial development.

Following Levine, Loayza and Beck (2000), I ran a cross-country regression of real GDP per capita growth rates over the years 1960-1995 on an index of financial development and on three other sets of control variables, treating financial development as endogenous and the controls as included exogenous variables. Financial development was measured by the log of the total value of credits issued by financial intermediaries to the private sector, divided by GDP (the measure of financial intermediation preferred by Levine, Loayza and Beck). Following Levine, Loayza and Beck, the three sets of control variables are (i) the simple conditioning information set: schooling and 1960 real GDP per capita, in logs, (ii) the policy conditioning information set: the simple information set plus the government share in GDP, the trade share in GDP, inflation and the black market premium, and (iii) the full information set: the policy information set plus indicators of revolutions, coups and ethnic fractionalization. I used TSLS, with English, French and German legal origin dummies as instruments[11]. The first-stage F-statistic for this regression is 5.63 with the simple conditioning information set, 5.43 with the policy conditioning information set and 5.80 with the full conditioning information set. Meanwhile, the 1% critical value in the test that compares this to a $\chi^2_k/k$ distribution is 3.78, so this test clearly rejects in all three cases. The values of the test statistic $L$, and the 5% critical values are reported in Table 7. For all three sets of controls, the null of identification is rejected. I conclude that even if these instruments are uncorrelated with the error term in the structural equation, their correlation with the endogenous regressor is not high enough to allow a researcher to conduct inference in the conventional way. For the simple conditioning information set, the regular Wald confidence set for the effect of financial innovation on growth is $2.34 \pm 1.20$. The Anderson-Rubin confidence set for this parameter goes from 0.51 to 5.77, which, though quite different, is still an interval containing only strictly positive values. Results with the other sets of controls are similar.

5.2 *The Consumption CAPM with CRRA Preferences.*

[11] I am grateful to Ross Levine for providing me with these data. A more detailed description of the data can be found in Levine, Loayza and Beck.

In this application, I use US annual data from Campbell and Shiller (1987), updated to cover the years 1889-1999. The data consist of stock returns, bond returns and consumption growth, all in real terms, as described by Campbell and Shiller. I then consider GMM estimation of the parameters $\delta$ and $\gamma$, using both stock and bond returns as the elements of $R_{t+1}$, using instrument sets A and B (as defined the Monte-Carlo simulation above). The values of the test statistic $L$ (computed as described in the Monte-Carlo simulation above) and the 5% critical values are reported in Table 8. For both instrument sets, the null of identification is rejected. The two-step estimator indicates precisely identified parameters with small risk aversion but the robust S-sets are large and indicate high risk aversion, a pattern that was found by Stock and Wright (2000) with an earlier dataset.

## 6. Conclusion.

In this paper, I have proposed a test of the null hypothesis of identification, that allows for the detection of a local or global underidentification, and of underidentification in some or all directions. 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. When the test rejects, inference should be conducted only by methods that are robust to underidentification.

# Appendix: Proof of Theorem 1.

Let $b = T^{1/2}(\theta - \theta_0)$ where $\theta_0$ denotes the true parameter value. The continuous-updating GMM objective function can be written as

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

$$[T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t,\theta_0)+T^{-1}\Sigma_{t=1}^{T}\tfrac{d\phi(Y_t,\theta^*)}{d\theta}b]'V_T(\theta)^{-1}[T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t,\theta_0)+T^{-1}\Sigma_{t=1}^{T}\tfrac{d\phi(Y_t,\theta^*)}{d\theta}b]$$

where $\theta^*$ is on the line segment between $\theta$ and $\theta_0$. So

$$S_{CU}(\theta) = [V_T(\theta)^{-1/2'}T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t,\theta_0) + V_T(\theta)^{-1/2'}T^{-1}\Sigma_{t=1}^{T}\tfrac{d\phi(Y_t,\theta^*)}{d\theta}b]'$$

$$[V_T(\theta)^{-1/2}T^{-1/2}\Sigma_{t=1}^{T}\phi(Y_t,\theta_0) + V_T(\theta)^{-1/2}T^{-1}\Sigma_{t=1}^{T}\tfrac{d\phi(Y_t,\theta^*)}{d\theta}b]$$

which converges to

$$(\phi+Gb)'(\phi+Gb) = \phi'\phi + (b+(G'G)^{-1}G'\phi)'G'G(b+(G'G)^{-1}G'\phi) - \phi'G(G'G)^{-1}G'\phi.$$

The value of $b$ that minimizes $(\phi+Gb)'(\phi+Gb)$ is $(G'G)^{-1}G'\phi$. So,

$$T^{1/2}W_1 \rightarrow_d 1[(\phi+G(G'G)^{-1}G'\phi)'(\phi+G(G'G)^{-1}G'\phi) \leq F_{\chi^2}(k,\alpha)]\,\text{sup}_{b_1,b_2}\,||b_1-b_2||$$

such that

$$\phi'\phi + (b_i+(G'G)^{-1}G'\phi)'G'G(b_i+(G'G)^{-1}G'\phi) - \phi'G(G'G)^{-1}G'\phi \leq F_{\chi^2}(k,\alpha),$$

for $i = 1, 2$, which is equal to

$$1[\phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha) \geq 0]\,\text{sup}_{b_1,b_2}\,||b_1-b_2||$$

such that

$$b_i{}'G'Gb_i \leq \phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha),$$

for $i = 1, 2$. So $T^{1/2}W_1 \rightarrow_d 2\sqrt{\frac{\phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha)}{\lambda}} 1[\phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha) \geq 0]$, where $\lambda$ is the smallest eigenvalue of $G'G = B'A^{-1}B$. Meanwhile, $W_2 = \frac{2}{\sqrt{T}}\sqrt{\frac{F_{\chi^2}(n,\alpha)}{\hat{\lambda}}}W_2$, recalling that $\hat{\lambda}$ is defined as the smallest eigenvalue of $\hat{B}'\hat{A}^{-1}\hat{B}$, which is consistent for $\lambda$, so that $T^{1/2}W_2 \rightarrow_d 2\sqrt{\frac{F_{\chi^2}(n,\alpha)}{\lambda}}$.

Combining these, under assumptions $A1 - A6$,

$$L = \frac{T^{1/2}W_1}{T^{1'2}W_2} \rightarrow_d \sqrt{\frac{\phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha)}{F_{\chi^2}(n,\alpha)}} 1[\phi'G(G'G)^{-1}G'\phi - \phi'\phi + F_{\chi^2}(k,\alpha) \geq 0],$$

as required. ■

## 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.

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.

Campbell, J.Y. and R.J. Shiller (1987): Cointegration and Tests of Present Value Models, *Journal of Political Economy*, 95, pp.1062-1088.

Choi, I. and P.C.B. Phillips (1992): Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations, *Journal of Econometrics*, 51, pp.113-150.

Cragg, J.G. and S.G. Donald (1993): Testing Identifiability and Specification in Instrumental Variable Models, *Econometric Theory*, 9, pp.222-240.

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 (2002): Projection-based Statistical Inference in Linear Structural Models With Possibly Weak Instruments, mimeo.

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.

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.

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

Kleibergen, F. (2001): Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, *Econometrica*, forthcoming.

Kleibergen, F. (2002): Testing Parameters in GMM Without Assuming That They are Identified, mimeo.

Levine, R., N. Loayza and T. Beck (2000): Financial Intermediation and Growth: Causality and Causes, *Journal of Monetary Economics*, 46, pp.31-77.

Pagan, A.R. and J.C. Robertson (1998), Structural Models of the Liquidity Effect, *Review of Economics and Statistics*, 80, 202-217.

Moriera, M.J. (2002): A Conditional Likelihood Ratio Test for Structural Models, mimeo.

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., J.H. Wright and M. Yogo (2002): GMM with Weak Identification, *Journal of Business and Economic Statistics*, forthcoming.

Stock, J.H. and M. Yogo (2001): Testing for Weak Instruments in Linear IV Regression, mimeo.

Tauchen, G. (1986): Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data, *Journal of Business and Economic Statistics*, 4, pp.397-425.

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

Wang, J. and E. Zivot (1998): Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments, *Econometrica*, 66, pp.1389-1404.

Wright, J.H. (2002): Detecting Lack of Identification in GMM, *Econometric Theory*, forthcoming.

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: Monte-Carlo Results: Linear IV Model $\tilde{R}_f^2 = 0$

| | | | Coverage | | Rej Rate | Accept Rate | | Pretest Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\rho$ | $k$ | Wald | AR | $L$ | $F_1$ | $F_2$ | $L$ | $F_1$ | $F_2$ |
| 100 | -0.9 | 5 | 100.0 | 96.1 | 100.0 | 96.0 | 100.0 | 96.1 | 98.3 | 96.1 |
| 100 | -0.9 | 10 | 100.0 | 96.1 | 99.9 | 94.1 | 100.0 | 96.2 | 98.7 | 96.1 |
| 100 | -0.9 | 30 | 100.0 | 97.2 | 100.0 | 89.1 | 100.0 | 97.2 | 99.6 | 97.2 |
| 100 | -0.5 | 5 | 100.0 | 96.1 | 100.0 | 94.3 | 100.0 | 96.1 | 96.7 | 96.1 |
| 100 | -0.5 | 10 | 100.0 | 96.1 | 99.8 | 93.9 | 100.0 | 96.2 | 96.8 | 96.1 |
| 100 | -0.5 | 30 | 100.0 | 97.2 | 100.0 | 88.9 | 100.0 | 97.2 | 97.9 | 97.2 |
| 100 | 0.5 | 5 | 82.3 | 96.1 | 99.8 | 94.1 | 100.0 | 96.1 | 93.4 | 96.1 |
| 100 | 0.5 | 10 | 56.2 | 96.1 | 100.0 | 91.1 | 100.0 | 96.1 | 91.3 | 96.1 |
| 100 | 0.5 | 30 | 15.7 | 97.2 | 100.0 | 90.9 | 100.0 | 97.2 | 89.4 | 97.2 |
| 100 | 0.9 | 5 | 12.1 | 96.1 | 99.9 | 93.8 | 100.0 | 96.1 | 92.4 | 96.1 |
| 100 | 0.9 | 10 | 0.8 | 96.1 | 100.0 | 93.1 | 100.0 | 96.1 | 92.1 | 96.1 |
| 100 | 0.9 | 30 | 0.0 | 97.2 | 100.0 | 89.0 | 100.0 | 97.2 | 88.4 | 97.2 |
| 250 | -0.9 | 5 | 100.0 | 95.6 | 99.9 | 95.8 | 100.0 | 95.6 | 97.8 | 95.6 |
| 250 | -0.9 | 10 | 100.0 | 94.6 | 99.7 | 94.4 | 100.0 | 94.9 | 97.8 | 94.6 |
| 250 | -0.9 | 30 | 100.0 | 96.3 | 100.0 | 93.0 | 100.0 | 96.3 | 99.2 | 96.3 |
| 250 | -0.5 | 5 | 100.0 | 95.6 | 99.9 | 94.3 | 100.0 | 95.6 | 96.1 | 95.6 |
| 250 | -0.5 | 10 | 100.0 | 94.6 | 99.9 | 94.9 | 100.0 | 94.7 | 95.4 | 94.6 |
| 250 | -0.5 | 30 | 100.0 | 96.3 | 100.0 | 94.7 | 100.0 | 96.3 | 96.6 | 96.3 |
| 250 | 0.5 | 5 | 81.7 | 95.6 | 99.9 | 94.2 | 100.0 | 95.5 | 93.4 | 95.6 |
| 250 | 0.5 | 10 | 53.1 | 94.6 | 100.0 | 92.8 | 100.0 | 94.6 | 89.8 | 94.6 |
| 250 | 0.5 | 30 | 15.6 | 96.3 | 100.0 | 92.7 | 100.0 | 96.3 | 90.0 | 96.3 |
| 250 | 0.9 | 5 | 10.4 | 95.6 | 100.0 | 95.0 | 100.0 | 95.6 | 92.7 | 95.6 |
| 250 | 0.9 | 10 | 0.7 | 94.6 | 100.0 | 92.5 | 100.0 | 94.6 | 90.5 | 94.6 |
| 250 | 0.9 | 30 | 0.0 | 96.3 | 100.0 | 94.4 | 100.0 | 96.3 | 93.0 | 96.3 |
| 1000 | -0.9 | 5 | 100.0 | 94.9 | 99.8 | 94.4 | 100.0 | 95.1 | 98.3 | 94.9 |
| 1000 | -0.9 | 10 | 100.0 | 95.7 | 99.9 | 95.8 | 100.0 | 95.8 | 97.9 | 95.7 |
| 1000 | -0.9 | 30 | 100.0 | 95.1 | 99.8 | 94.7 | 100.0 | 95.3 | 97.6 | 95.1 |
| 1000 | -0.5 | 5 | 100.0 | 94.9 | 99.9 | 93.6 | 100.0 | 95.0 | 95.7 | 94.9 |
| 1000 | -0.5 | 10 | 100.0 | 95.7 | 99.8 | 95.7 | 100.0 | 95.8 | 96.0 | 95.7 |
| 1000 | -0.5 | 30 | 100.0 | 95.1 | 99.8 | 93.9 | 100.0 | 95.3 | 95.8 | 95.1 |
| 1000 | 0.5 | 5 | 83.1 | 94.9 | 99.6 | 94.2 | 100.0 | 94.9 | 92.5 | 94.9 |
| 1000 | 0.5 | 10 | 55.4 | 95.7 | 99.9 | 94.8 | 100.0 | 95.7 | 92.5 | 95.7 |
| 1000 | 0.5 | 30 | 13.8 | 95.1 | 99.6 | 94.7 | 100.0 | 95.1 | 91.2 | 95.1 |
| 1000 | 0.9 | 5 | 11.7 | 94.9 | 99.6 | 94.9 | 100.0 | 94.9 | 92.1 | 94.9 |
| 1000 | 0.9 | 10 | 1.2 | 95.7 | 99.8 | 95.2 | 100.0 | 95.7 | 93.4 | 95.7 |
| 1000 | 0.9 | 30 | 0.0 | 95.1 | 99.6 | 94.3 | 100.0 | 95.1 | 92.6 | 95.1 |

Notes: The coverage columns give the coverage rates of Wald and AR confidence sets. The reject rate column gives the rejection rate of the proposed test. The accept rate columns give the rejections rates of the tests based on comparing the first-stage F-statistic with $\chi_k^2/k$ critical values and the critical values of Stock and Yogo (2001), designed to ensure that the TSLS Wald test size is no larger than 25%. For these columns, I report an acceptance rate, rather than a rejection rate because the null hypothesis is that the model is *not* identified, whereas for my test the null hypothesis is that the model *is* identified. The last four columns report the effective coverage rate of the confidence set that is the Wald or AR confidence set depending on the results of each identification/underidentification test.

Table 2: Monte-Carlo Results: Linear IV Model $\tilde{R}_f^2 = 0.01$

| | | | Coverage | | Rej Rate | Accept Rate | | Pretest Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\rho$ | $k$ | Wald | AR | $L$ | $F_1$ | $F_2$ | $L$ | $F_1$ | $F_2$ |
| 100 | -0.9 | 5 | 100.0 | 96.1 | 99.4 | 91.4 | 100.0 | 96.7 | 98.2 | 96.1 |
| 100 | -0.9 | 10 | 100.0 | 96.1 | 99.1 | 89.2 | 100.0 | 97.0 | 99.0 | 96.1 |
| 100 | -0.9 | 30 | 100.0 | 97.2 | 100.0 | 86.3 | 100.0 | 97.2 | 99.7 | 97.2 |
| 100 | -0.5 | 5 | 100.0 | 96.1 | 99.8 | 89.1 | 100.0 | 96.2 | 96.6 | 96.1 |
| 100 | -0.5 | 10 | 100.0 | 96.1 | 99.6 | 89.0 | 100.0 | 96.4 | 97.4 | 96.1 |
| 100 | -0.5 | 30 | 100.0 | 97.2 | 100.0 | 86.3 | 100.0 | 97.2 | 97.8 | 97.2 |
| 100 | 0.5 | 5 | 82.3 | 96.1 | 99.8 | 88.4 | 100.0 | 96.1 | 92.4 | 96.1 |
| 100 | 0.5 | 10 | 57.4 | 96.1 | 99.6 | 87.9 | 100.0 | 96.0 | 89.4 | 96.1 |
| 100 | 0.5 | 30 | 17.6 | 97.2 | 100.0 | 87.3 | 100.0 | 97.2 | 86.3 | 97.2 |
| 100 | 0.9 | 5 | 24.1 | 96.1 | 99.3 | 89.2 | 100.0 | 96.1 | 88.4 | 96.1 |
| 100 | 0.9 | 10 | 2.2 | 96.1 | 99.5 | 88.2 | 100.0 | 96.1 | 87.1 | 96.1 |
| 100 | 0.9 | 30 | 0.0 | 97.2 | 100.0 | 85.5 | 100.0 | 97.2 | 85.1 | 97.2 |
| 250 | -0.9 | 5 | 100.0 | 95.6 | 98.3 | 80.2 | 100.0 | 97.2 | 99.0 | 95.6 |
| 250 | -0.9 | 10 | 100.0 | 94.6 | 97.8 | 85.2 | 100.0 | 96.8 | 98.0 | 94.6 |
| 250 | -0.9 | 30 | 100.0 | 96.3 | 99.4 | 88.2 | 100.0 | 96.9 | 99.0 | 96.3 |
| 250 | -0.5 | 5 | 100.0 | 95.6 | 99.2 | 78.5 | 100.0 | 96.0 | 97.3 | 95.6 |
| 250 | -0.5 | 10 | 100.0 | 94.6 | 99.3 | 85.4 | 100.0 | 95.2 | 96.4 | 94.6 |
| 250 | -0.5 | 30 | 100.0 | 96.3 | 100.0 | 88.7 | 100.0 | 96.3 | 97.0 | 96.3 |
| 250 | 0.5 | 5 | 84.8 | 95.6 | 99.0 | 79.9 | 100.0 | 95.8 | 91.2 | 95.6 |
| 250 | 0.5 | 10 | 61.6 | 94.6 | 99.4 | 81.4 | 100.0 | 94.3 | 84.9 | 94.6 |
| 250 | 0.5 | 30 | 18.9 | 96.3 | 100.0 | 87.3 | 100.0 | 96.3 | 85.1 | 96.3 |
| 250 | 0.9 | 5 | 32.8 | 95.6 | 97.9 | 80.7 | 100.0 | 95.4 | 79.6 | 95.6 |
| 250 | 0.9 | 10 | 5.5 | 94.6 | 97.9 | 81.8 | 100.0 | 94.5 | 80.8 | 94.6 |
| 250 | 0.9 | 30 | 0.0 | 96.3 | 99.6 | 90.2 | 100.0 | 96.3 | 89.3 | 96.3 |
| 1000 | -0.9 | 5 | 100.0 | 94.9 | 94.4 | 31.8 | 99.9 | 98.4 | 99.2 | 95.0 |
| 1000 | -0.9 | 10 | 100.0 | 95.7 | 95.9 | 45.9 | 100.0 | 98.8 | 99.3 | 95.7 |
| 1000 | -0.9 | 30 | 100.0 | 95.1 | 95.6 | 68.4 | 100.0 | 99.0 | 99.5 | 95.1 |
| 1000 | -0.5 | 5 | 100.0 | 94.9 | 92.4 | 30.5 | 100.0 | 97.4 | 98.8 | 94.9 |
| 1000 | -0.5 | 10 | 100.0 | 95.7 | 95.8 | 47.0 | 100.0 | 97.6 | 98.5 | 95.7 |
| 1000 | -0.5 | 30 | 100.0 | 95.1 | 97.6 | 66.2 | 100.0 | 96.8 | 98.1 | 95.1 |
| 1000 | 0.5 | 5 | 86.8 | 94.9 | 93.5 | 30.8 | 100.0 | 94.9 | 86.5 | 94.9 |
| 1000 | 0.5 | 10 | 74.4 | 95.7 | 94.8 | 43.7 | 100.0 | 94.0 | 80.0 | 95.7 |
| 1000 | 0.5 | 30 | 27.2 | 95.1 | 97.3 | 67.1 | 100.0 | 94.6 | 69.6 | 95.1 |
| 1000 | 0.9 | 5 | 63.7 | 94.9 | 94.4 | 30.7 | 99.7 | 93.4 | 65.5 | 94.8 |
| 1000 | 0.9 | 10 | 26.7 | 95.7 | 96.0 | 43.2 | 100.0 | 94.7 | 48.2 | 95.7 |
| 1000 | 0.9 | 30 | 0.0 | 95.1 | 95.9 | 65.6 | 100.0 | 94.9 | 65.2 | 95.1 |

See notes to Table 1.

Table 3: Monte-Carlo Results: Linear IV Model $\tilde{R}_f^2 = 0.1$

| | | | Coverage | | Rej Rate | Accept Rate | | Pretest Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\rho$ | $k$ | Wald | AR | $L$ | $F_1$ | $F_2$ | $L$ | $F_1$ | $F_2$ |
| 100 | -0.9 | 5 | 100.0 | 96.1 | 96.1 | 30.8 | 99.5 | 98.8 | 99.3 | 96.2 |
| 100 | -0.9 | 10 | 100.0 | 96.1 | 96.0 | 43.4 | 100.0 | 99.0 | 99.4 | 96.1 |
| 100 | -0.9 | 30 | 100.0 | 97.2 | 97.8 | 57.8 | 100.0 | 99.4 | 99.9 | 97.2 |
| 100 | -0.5 | 5 | 100.0 | 96.1 | 95.9 | 31.5 | 99.8 | 97.9 | 98.7 | 96.1 |
| 100 | -0.5 | 10 | 100.0 | 96.1 | 95.0 | 39.9 | 100.0 | 98.3 | 98.8 | 96.1 |
| 100 | -0.5 | 30 | 100.0 | 97.2 | 99.8 | 56.6 | 100.0 | 97.4 | 98.7 | 97.2 |
| 100 | 0.5 | 5 | 89.0 | 96.1 | 95.1 | 29.2 | 99.4 | 95.8 | 89.1 | 95.9 |
| 100 | 0.5 | 10 | 73.3 | 96.1 | 95.1 | 40.0 | 100.0 | 95.2 | 79.5 | 96.1 |
| 100 | 0.5 | 30 | 29.5 | 97.2 | 99.8 | 56.7 | 100.0 | 97.2 | 63.6 | 97.2 |
| 100 | 0.9 | 5 | 70.0 | 96.1 | 94.9 | 28.5 | 99.3 | 93.9 | 71.8 | 95.6 |
| 100 | 0.9 | 10 | 26.9 | 96.1 | 95.4 | 37.5 | 100.0 | 94.4 | 44.5 | 96.1 |
| 100 | 0.9 | 30 | 0.1 | 97.2 | 98.1 | 57.9 | 100.0 | 97.1 | 57.9 | 97.2 |
| 250 | -0.9 | 5 | 100.0 | 95.6 | 84.2 | 0.3 | 82.8 | 99.5 | 100.0 | 97.6 |
| 250 | -0.9 | 10 | 100.0 | 94.6 | 86.1 | 3.5 | 100.0 | 99.0 | 99.7 | 94.6 |
| 250 | -0.9 | 30 | 100.0 | 96.3 | 95.2 | 15.2 | 100.0 | 99.0 | 100.0 | 96.3 |
| 250 | -0.5 | 5 | 100.0 | 95.6 | 76.2 | 1.1 | 80.0 | 99.7 | 100.0 | 96.8 |
| 250 | -0.5 | 10 | 100.0 | 94.6 | 78.6 | 2.7 | 99.9 | 98.8 | 99.2 | 94.6 |
| 250 | -0.5 | 30 | 100.0 | 96.3 | 94.5 | 14.1 | 100.0 | 98.3 | 99.5 | 96.3 |
| 250 | 0.5 | 5 | 93.7 | 95.6 | 75.8 | 0.6 | 80.7 | 95.8 | 93.7 | 93.8 |
| 250 | 0.5 | 10 | 82.6 | 94.6 | 78.6 | 2.1 | 100.0 | 90.7 | 82.5 | 94.6 |
| 250 | 0.5 | 30 | 43.7 | 96.3 | 94.3 | 15.5 | 100.0 | 93.4 | 49.4 | 96.3 |
| 250 | 0.9 | 5 | 83.0 | 95.6 | 84.5 | 0.8 | 83.0 | 90.2 | 82.9 | 86.6 |
| 250 | 0.9 | 10 | 54.8 | 94.6 | 87.9 | 2.2 | 100.0 | 86.9 | 54.9 | 94.6 |
| 250 | 0.9 | 30 | 1.2 | 96.3 | 95.5 | 16.5 | 100.0 | 95.0 | 16.4 | 96.3 |
| 1000 | -0.9 | 5 | 99.8 | 94.9 | 40.2 | 0.0 | 0.0 | 100.0 | 99.8 | 99.8 |
| 1000 | -0.9 | 10 | 100.0 | 95.7 | 48.8 | 0.0 | 41.3 | 99.9 | 100.0 | 98.6 |
| 1000 | -0.9 | 30 | 100.0 | 95.1 | 72.4 | 0.0 | 100.0 | 100.0 | 100.0 | 95.1 |
| 1000 | -0.5 | 5 | 98.9 | 94.9 | 29.3 | 0.0 | 0.0 | 99.5 | 98.9 | 98.9 |
| 1000 | -0.5 | 10 | 99.6 | 95.7 | 33.5 | 0.0 | 41.1 | 99.9 | 99.6 | 98.6 |
| 1000 | -0.5 | 30 | 100.0 | 95.1 | 55.5 | 0.0 | 100.0 | 100.0 | 100.0 | 95.1 |
| 1000 | 0.5 | 5 | 93.6 | 94.9 | 29.3 | 0.0 | 0.0 | 94.8 | 93.6 | 93.6 |
| 1000 | 0.5 | 10 | 92.4 | 95.7 | 33.4 | 0.0 | 39.3 | 94.0 | 92.4 | 92.2 |
| 1000 | 0.5 | 30 | 71.1 | 95.1 | 55.2 | 0.0 | 100.0 | 81.0 | 71.1 | 95.1 |
| 1000 | 0.9 | 5 | 90.0 | 94.9 | 38.2 | 0.0 | 0.0 | 92.1 | 90.0 | 90.0 |
| 1000 | 0.9 | 10 | 84.5 | 95.7 | 46.9 | 0.0 | 41.1 | 88.0 | 84.5 | 83.3 |
| 1000 | 0.9 | 30 | 29.6 | 95.1 | 73.0 | 0.0 | 100.0 | 73.9 | 29.6 | 95.1 |

See notes to Table 1.

Table 4: Monte-Carlo Results: Linear IV Model $\tilde{R}_f^2 = 0.3$

| | | | Coverage | | Rej Rate | Accept Rate | | Pretest Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\rho$ | $k$ | Wald | AR | $L$ | $F_1$ | $F_2$ | $L$ | $F_1$ | $F_2$ |
| 100 | -0.9 | 5 | 100.0 | 96.1 | 82.4 | 0.0 | 44.5 | 99.7 | 100.0 | 98.8 |
| 100 | -0.9 | 10 | 100.0 | 96.1 | 84.4 | 0.6 | 99.7 | 99.4 | 99.9 | 96.4 |
| 100 | -0.9 | 30 | 100.0 | 97.2 | 95.3 | 4.3 | 100.0 | 99.9 | 100.0 | 97.2 |
| 100 | -0.5 | 5 | 99.7 | 96.1 | 73.8 | 0.1 | 44.7 | 99.6 | 99.7 | 98.7 |
| 100 | -0.5 | 10 | 99.8 | 96.1 | 76.2 | 0.6 | 99.7 | 99.7 | 99.8 | 96.1 |
| 100 | -0.5 | 30 | 100.0 | 97.2 | 94.4 | 4.6 | 100.0 | 99.4 | 100.0 | 97.2 |
| 100 | 0.5 | 5 | 93.2 | 96.1 | 74.1 | 0.3 | 42.1 | 95.7 | 93.2 | 92.8 |
| 100 | 0.5 | 10 | 86.3 | 96.1 | 74.3 | 0.5 | 99.7 | 94.2 | 86.3 | 95.9 |
| 100 | 0.5 | 30 | 52.8 | 97.2 | 94.3 | 4.9 | 100.0 | 94.1 | 53.9 | 97.2 |
| 100 | 0.9 | 5 | 87.0 | 96.1 | 83.4 | 0.2 | 41.5 | 93.4 | 87.0 | 86.5 |
| 100 | 0.9 | 10 | 66.7 | 96.1 | 83.2 | 0.3 | 99.6 | 87.3 | 66.7 | 95.9 |
| 100 | 0.9 | 30 | 5.1 | 97.2 | 95.2 | 4.2 | 100.0 | 95.1 | 8.1 | 97.2 |
| 250 | -0.9 | 5 | 99.9 | 95.6 | 43.2 | 0.0 | 0.0 | 100.0 | 99.9 | 99.9 |
| 250 | -0.9 | 10 | 100.0 | 94.6 | 52.7 | 0.0 | 47.9 | 99.7 | 100.0 | 97.8 |
| 250 | -0.9 | 30 | 100.0 | 96.3 | 80.3 | 0.0 | 100.0 | 100.0 | 100.0 | 96.3 |
| 250 | -0.5 | 5 | 99.7 | 95.6 | 32.8 | 0.0 | 0.0 | 99.9 | 99.7 | 99.7 |
| 250 | -0.5 | 10 | 99.4 | 94.6 | 39.2 | 0.0 | 46.1 | 99.6 | 99.4 | 97.2 |
| 250 | -0.5 | 30 | 100.0 | 96.3 | 67.7 | 0.0 | 100.0 | 100.0 | 100.0 | 96.3 |
| 250 | 0.5 | 5 | 96.9 | 95.6 | 34.3 | 0.0 | 0.1 | 97.6 | 96.9 | 96.9 |
| 250 | 0.5 | 10 | 90.6 | 94.6 | 39.1 | 0.0 | 44.2 | 92.4 | 90.6 | 90.5 |
| 250 | 0.5 | 30 | 72.9 | 96.3 | 68.6 | 0.0 | 100.0 | 86.3 | 72.9 | 96.3 |
| 250 | 0.9 | 5 | 93.6 | 95.6 | 45.8 | 0.0 | 0.0 | 94.5 | 93.6 | 93.6 |
| 250 | 0.9 | 10 | 81.7 | 94.6 | 49.5 | 0.0 | 43.1 | 86.1 | 81.7 | 81.1 |
| 250 | 0.9 | 30 | 27.6 | 96.3 | 80.7 | 0.0 | 100.0 | 80.9 | 27.6 | 96.3 |
| 1000 | -0.9 | 5 | 98.8 | 94.9 | 14.9 | 0.0 | 0.0 | 99.1 | 98.8 | 98.8 |
| 1000 | -0.9 | 10 | 99.5 | 95.7 | 15.2 | 0.0 | 0.0 | 99.6 | 99.5 | 99.5 |
| 1000 | -0.9 | 30 | 100.0 | 95.1 | 26.1 | 0.0 | 100.0 | 100.0 | 100.0 | 95.1 |
| 1000 | -0.5 | 5 | 98.4 | 94.9 | 12.5 | 0.0 | 0.0 | 98.6 | 98.4 | 98.4 |
| 1000 | -0.5 | 10 | 99.1 | 95.7 | 12.2 | 0.0 | 0.0 | 99.4 | 99.1 | 99.1 |
| 1000 | -0.5 | 30 | 99.5 | 95.1 | 18.2 | 0.0 | 100.0 | 99.8 | 99.5 | 95.1 |
| 1000 | 0.5 | 5 | 95.3 | 94.9 | 12.1 | 0.0 | 0.0 | 95.6 | 95.3 | 95.3 |
| 1000 | 0.5 | 10 | 95.8 | 95.7 | 12.3 | 0.0 | 0.0 | 96.2 | 95.8 | 95.8 |
| 1000 | 0.5 | 30 | 88.9 | 95.1 | 18.8 | 0.0 | 100.0 | 90.4 | 88.9 | 95.1 |
| 1000 | 0.9 | 5 | 93.6 | 94.9 | 14.8 | 0.0 | 0.0 | 94.2 | 93.6 | 93.6 |
| 1000 | 0.9 | 10 | 93.7 | 95.7 | 15.0 | 0.0 | 0.0 | 94.2 | 93.7 | 93.7 |
| 1000 | 0.9 | 30 | 71.3 | 95.1 | 26.5 | 0.0 | 99.9 | 75.1 | 71.3 | 95.1 |

See notes to Table 1.

Table 5: Monte-Carlo Results: Linear IV Model $\tilde{R}_f^2 = 0.5$

| | | | Coverage | | Rej Rate | Accept Rate | | Pretest Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\rho$ | $k$ | Wald | AR | $L$ | $F_1$ | $F_2$ | $L$ | $F_1$ | $F_2$ |
| 100 | -0.9 | 5 | 99.8 | 96.1 | 62.9 | 0.0 | 0.2 | 100.0 | 99.8 | 99.8 |
| 100 | -0.9 | 10 | 99.9 | 96.1 | 66.5 | 0.0 | 58.2 | 99.9 | 99.9 | 98.7 |
| 100 | -0.9 | 30 | 100.0 | 97.2 | 89.0 | 0.0 | 100.0 | 100.0 | 100.0 | 97.2 |
| 100 | -0.5 | 5 | 99.2 | 96.1 | 52.3 | 0.0 | 0.0 | 99.8 | 99.2 | 99.2 |
| 100 | -0.5 | 10 | 99.0 | 96.1 | 57.2 | 0.0 | 57.4 | 99.9 | 99.0 | 98.3 |
| 100 | -0.5 | 30 | 100.0 | 97.2 | 85.6 | 0.0 | 100.0 | 100.0 | 100.0 | 97.2 |
| 100 | 0.5 | 5 | 95.2 | 96.1 | 52.8 | 0.0 | 0.6 | 96.8 | 95.2 | 95.2 |
| 100 | 0.5 | 10 | 90.9 | 96.1 | 55.8 | 0.0 | 56.6 | 94.9 | 90.9 | 91.7 |
| 100 | 0.5 | 30 | 71.9 | 97.2 | 85.1 | 0.0 | 100.0 | 92.7 | 71.9 | 97.2 |
| 100 | 0.9 | 5 | 91.9 | 96.1 | 63.3 | 0.0 | 0.4 | 94.5 | 91.9 | 91.9 |
| 100 | 0.9 | 10 | 81.4 | 96.1 | 65.0 | 0.0 | 56.8 | 89.6 | 81.4 | 84.6 |
| 100 | 0.9 | 30 | 25.0 | 97.2 | 88.7 | 0.0 | 100.0 | 88.9 | 25.0 | 97.2 |
| 250 | -0.9 | 5 | 99.9 | 95.6 | 24.6 | 0.0 | 0.0 | 100.0 | 99.9 | 99.9 |
| 250 | -0.9 | 10 | 99.7 | 94.6 | 31.1 | 0.0 | 0.0 | 99.9 | 99.7 | 99.7 |
| 250 | -0.9 | 30 | 100.0 | 96.3 | 56.4 | 0.0 | 100.0 | 100.0 | 100.0 | 96.3 |
| 250 | -0.5 | 5 | 99.3 | 95.6 | 19.0 | 0.0 | 0.0 | 99.5 | 99.3 | 99.3 |
| 250 | -0.5 | 10 | 98.4 | 94.6 | 24.6 | 0.0 | 0.0 | 98.7 | 98.4 | 98.4 |
| 250 | -0.5 | 30 | 99.9 | 96.3 | 44.8 | 0.0 | 100.0 | 100.0 | 99.9 | 96.3 |
| 250 | 0.5 | 5 | 97.2 | 95.6 | 19.3 | 0.0 | 0.0 | 97.6 | 97.2 | 97.2 |
| 250 | 0.5 | 10 | 93.4 | 94.6 | 24.5 | 0.0 | 0.0 | 93.9 | 93.4 | 93.4 |
| 250 | 0.5 | 30 | 83.4 | 96.3 | 45.4 | 0.0 | 100.0 | 88.7 | 83.4 | 96.3 |
| 250 | 0.9 | 5 | 96.5 | 95.6 | 25.1 | 0.0 | 0.0 | 97.1 | 96.5 | 96.5 |
| 250 | 0.9 | 10 | 89.0 | 94.6 | 29.9 | 0.0 | 0.0 | 90.5 | 89.0 | 89.0 |
| 250 | 0.9 | 30 | 58.8 | 96.3 | 57.5 | 0.0 | 100.0 | 76.9 | 58.8 | 96.3 |
| 1000 | -0.9 | 5 | 98.4 | 94.9 | 9.9 | 0.0 | 0.0 | 98.5 | 98.4 | 98.4 |
| 1000 | -0.9 | 10 | 99.0 | 95.7 | 10.0 | 0.0 | 0.0 | 99.4 | 99.0 | 99.0 |
| 1000 | -0.9 | 30 | 99.8 | 95.1 | 14.2 | 0.0 | 0.0 | 100.0 | 99.8 | 99.8 |
| 1000 | -0.5 | 5 | 98.1 | 94.9 | 8.7 | 0.0 | 0.0 | 98.2 | 98.1 | 98.1 |
| 1000 | -0.5 | 10 | 98.7 | 95.7 | 9.0 | 0.0 | 0.0 | 99.1 | 98.7 | 98.7 |
| 1000 | -0.5 | 30 | 99.3 | 95.1 | 11.8 | 0.0 | 0.0 | 99.6 | 99.3 | 99.3 |
| 1000 | 0.5 | 5 | 96.1 | 94.9 | 8.5 | 0.0 | 0.0 | 96.4 | 96.1 | 96.1 |
| 1000 | 0.5 | 10 | 96.3 | 95.7 | 9.2 | 0.0 | 0.0 | 96.7 | 96.3 | 96.3 |
| 1000 | 0.5 | 30 | 93.5 | 95.1 | 11.9 | 0.0 | 0.0 | 94.2 | 93.5 | 93.5 |
| 1000 | 0.9 | 5 | 94.5 | 94.9 | 10.1 | 0.0 | 0.0 | 95.0 | 94.5 | 94.5 |
| 1000 | 0.9 | 10 | 95.2 | 95.7 | 10.2 | 0.0 | 0.0 | 95.6 | 95.2 | 95.2 |
| 1000 | 0.9 | 30 | 85.0 | 95.1 | 14.2 | 0.0 | 0.0 | 86.4 | 85.0 | 85.0 |

See notes to Table 1.

Table 6: Monte-Carlo Results: Consumption CAPM

| | | Coverage | | Rej Rate | Pretest Coverage |
|---|---|---|---|---|---|
| Inst. Set | $T$ | Wald | $S$ | $L$ | $L$ |
| A | 100 | 41.8 | 90.9 | 93.7 | 89.2 |
| A | 250 | 52.1 | 93.2 | 83.2 | 85.6 |
| A | 1000 | 74.3 | 95.1 | 19.0 | 78.4 |
| B | 100 | 45.6 | 93.3 | 97.1 | 91.7 |
| B | 250 | 59.2 | 94.6 | 79.9 | 86.8 |
| B | 1000 | 75.8 | 95.6 | 13.1 | 78.4 |

Notes: The coverage columns give the coverage rates of Wald and S sets. The rejection rate column give the rejection rate of the proposed test. The pretest coverage column reporta the effective coverage rate of the confidence set that is the Wald or S set depending on the result of the proposed identification test.

Table 7: Growth Regression

| Controls | Test Statistic $L$ | 5 % Critical Values |
|---|---|---|
| Simple | 2.20 | 1.42 |
| Policy | 2.29 | 1.42 |
| Full | 2.12 | 1.42 |

Notes: This table reports the proposed identification test statistic and associated critical values in three specifications of the cross-country growth regression using legal origin dummies as instruments, as discussed in the text.

Table 8: Consumption CAPM with Annual US Data

| Inst. Set | Test Statistic $L$ | 5 % Critical Values |
|---|---|---|
| A | 5.58 | 1.52 |
| B | 2.16 | 1.25 |

Notes: This table reports the proposed identification test statistic and associated critical values in nonlinear GMM Euler equation estimation of the consumption CAPM for stock returns and bond returns with annual US data, using two alternative sets of instruments, as discussed in the text.