

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 674

July 2000

DETECTING LACK OF IDENTIFICATION IN GMM

Jonathan H. Wright

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [www.bog.frb.fed.us](http://www.bog.frb.fed.us).

## DETECTING LACK OF IDENTIFICATION IN GMM

Jonathan H. Wright\*

**Abstract:** In the standard linear instrumental variables regression model, it must be assumed that the instruments are correlated with the endogenous variables in order to ensure the consistency and asymptotic normality of the usual instrumental variables estimator. Indeed, if the instruments are only slightly correlated with the endogenous variables, the conventional Gaussian asymptotic theory may still provide a very poor approximation to the finite sample distribution of the usual instrumental variables estimator. Because of the crucial role of this identification condition, it is common to test for instrument relevance by a first-stage F-test. Identification issues also arise in the generalized method of moments model, of which the linear instrumental variables model is a special case. But I know of no means of testing for identification in this model, in the existing literature. This paper proposes a test of the null of underidentification in the generalized method of moments model.

**KEYWORDS:** Generalized Method of Moments, Identification, Asset Pricing, Instrumental Variables.

\* Wright is an economist in the International Finance Division of the Federal Reserve Board. I am grateful to Jon Faust, John Fernald and Jim Stock for their 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, without which the usual IV estimator is neither consistent nor asymptotically normal. If the correlation between the instruments and the endogenous variables is nonzero, but slight, then the conventional Gaussian asymptotic theory for the IV model can nevertheless provide a very poor approximation to the actual sampling distribution of estimators and test statistics. An enormous 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 (see, for example, Richardson (1968), Sawa (1969), Phillips (1980, 1983), Nelson and Startz (1990a, 1990b), Maddala and Jeong (1992) and Bound, Jaeger and Baker (1995)). 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 ordinary least squares 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.

Recognizing the identification assumption on which the IV model relies, it is quite common in the applied literature to test for instrument relevance by a first-stage F-test (Basmann (1960)). This involves running a regression of the endogenous variables on the instruments and testing the null hypothesis of the joint insignificance of the slope coefficients, by means of an F-test. The null hypothesis is one of a total lack of identification. A rejection of this hypothesis by no means implies that issues of weak instruments can be ignored (see, for example, Staiger and Stock (1997)). But a failure to reject this hypothesis is a strong indication of identification difficulties. The first-stage F-test is an important and useful diagnostic in the IV model.

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 here requires a certain matrix to have full column rank). The identification condition is crucial in the GMM model, just as it is in the linear IV special case. But I am aware of no test of the identification condition in the GMM model, in the existing literature. This paper proposes such a test. Essentially, it involves an application of

the general approach to testing the rank of a matrix proposed by Cragg and Donald (1997), using a Bonferroni argument to eliminate a nuisance parameter that is not consistently estimable. The proposed test is, in general, asymptotically conservative.

The first-stage F-test in the linear IV model tests the null hypothesis that all of the instruments are uncorrelated with all of the endogenous regressors. In the presence of multiple endogenous regressors, the linear IV model may still fail to be fully identified even if this null hypothesis is false. Choi and Phillips (1992) discuss this situation and call it partial identification. It could arise, for example, if only one instrument is correlated with the endogenous variables. The first-stage F-test will not detect this kind of identification problem; if the model is partially identified then the first-stage F-test will reject with probability one asymptotically (Nelson and Startz (1990b), Shea (1997)). Cragg and Donald (1993) propose a test of the null of partial identification in the linear IV model which can detect this kind of identification problem<sup>1</sup>. The test that I propose in the GMM framework reduces to Cragg and Donald's test in the linear IV model. If the test that I propose in the GMM framework is used to test the null of a complete lack of identification, then it reduces to the usual first-stage F-test in the linear IV model. In the special case of

---

<sup>1</sup>Shea (1997) proposed a partial R-squared measure of instrument relevance in the linear IV model with multiple regressors, likewise motivated by the fact that a high R-squared (or F-statistic) in the first-stage regression does not imply that the model is fully identified. This measure does not however allow any formal test to be conducted.

the linear IV model, the proposed test is no longer asymptotically conservative.

The plan for the remainder of this paper is as follows. Section 2 discusses the GMM model, the consequences of identification problems and approaches for conducting inference when identification cannot be assumed. Section 3 discusses the proposed test for detecting underidentification. Section 4 contains Monte-Carlo evidence on its performance in the context of GMM estimation of the consumption capital asset pricing model (CAPM) of Lucas (1978). Section 5 concludes.

## 2. The GMM Model.

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

$$E_t(h(Y_t, \theta_0)) = 0 \tag{2.1}$$

where  $h(\cdot, \cdot)$  is a  $G$ -dimensional function and  $E_t(\cdot)$  denotes the expectation conditional on any variable in the information set at time  $t$ . Let  $Z_t$  be a  $k \times 1$  vector of instruments in the information set at time  $t$ . It follows that  $E(\phi(Y_t, \theta_0)) = 0$ , where  $\phi(Y_t, \theta) = h(Y_t, \theta) \otimes Z_t$ , a  $Gk \times 1$  vector. It is assumed that  $Gk \geq p$ .

The GMM estimator of  $\theta$  is

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

where

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

$\phi^*(\theta) = [T^{-1/2}\sum_{t=1}^T\phi(Y_t, \theta)]$  and  $W_T$  is a symmetric positive definite  $GkxGk$  weighting matrix which converges almost surely to a symmetric nonstochastic  $O(1)$  positive definite matrix  $W$ .

From here until the start of subsection 2.1, I adopt all of the following standard assumptions for the GMM model:

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

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

Assumption A3:  $T^{-1/2}\sum_{t=1}^T\phi(Y_t, \theta_0) \rightarrow_d N(0, A)$ , where  $A = E(\phi(Y_t, \theta_0)\phi(Y_t, \theta_0)')$ .

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

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

Assumption A6:  $V_T(\theta) = T^{-1}\sum_{t=1}^T\phi(Y_t, \theta)\phi(Y_t, \theta)'$  is a continuous function of  $\theta$  which converges to  $E(\phi(Y_t, \theta)\phi(Y_t, \theta)')$  uniformly in  $\theta$ .

Assumptions A2 and A3 are high level convergence assumptions. Assumption A4 is the identification assumption and is the focus of this paper. Assumption A5 rules out the possibility that  $\theta$  could be locally, but not globally identified (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  $(B'A^{-1}B)^{-1}$ .

In practice, GMM inference requires a specific weighting matrix to be chosen. One possible choice of the weighting matrix is the identity matrix. This objective function is

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

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

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

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

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

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  $Gk > p$ , then there are surplus instruments, which may be used to test the restriction that  $E\phi(Y_t, \theta_0) = 0$ . This may be done by Hansen's J-test, the statistic for which is  $J_{TS} = S_{TS}(\hat{\theta}_{TS})$  or  $J_{CU} = S_{CU}(\hat{\theta}_{CU})$ , depending on whether the two-step or continuous-updating estimator is used. Under the null,  $J_{TS}$  and  $J_{CU}$  are both asymptotically  $\chi^2$  distributed on  $Gk - p$  degrees of freedom.

### 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 J-statistics have erratic rejection rates. These problems have been documented in Monte-Carlo studies by Tauchen (1986), Kocherlakota (1990) and by the papers in a 1996 special issue of the *Journal of Business and Economic Statistics* on GMM estimation, including Hansen, Heaton and Yaron (1996). Problems with the conventional asymptotic theory in the GMM model could occur for a number of reasons. It could be that  $T^{-1/2}\sum_{t=1}^T\phi(Y_t, \theta_0)$  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  $\theta \neq \theta_0$ .

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 in the consumption CAPM, considered by Tauchen (1986), Kocherlakota (1990) and Hansen, Heaton and Yaron (1996). So, it can account for many of the puzzles found in Monte-Carlo studies of the GMM method. 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 (Anderson and Rubin (1949), Staiger and Stock (1997))<sup>2</sup>. 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  $GK$  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

---

<sup>2</sup>The ordinary Anderson-Rubin confidence set applies in the linear IV model. In this model, other closely related confidence sets have been proposed, which are likewise robust to identification problems (Zivot and Wang (1998)).

test, i.e. the confidence set of coverage  $1-\alpha$  is  $S_{\theta}^*(\alpha) = \{\theta : S_{CU}(\theta) \leq F_{\chi^2}(\alpha, GK)\}$  where  $F_{\chi^2}(a, b)$  is the  $100a$  percentile of a  $\chi^2$  distribution on  $b$  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)). In an identified model, the S-set is asymptotically equivalent to the usual confidence ellipse, based on conventional asymptotic theory.

### *2.3 Local Identification, Global Identification and Partial Identification.*

Assumption A4, specifying that  $B$  is of rank  $p$ , is the local identification assumption (Hsiao (1983)). It implies that  $E(\phi(Y_t, \theta))$  has a zero at  $\theta_0$ , that is unique at least in a neighborhood of  $\theta_0$ . Assumption A5 is the global identification assumption. A model could be locally identified without being globally identified. In this case, we would expect S-sets to be disjoint, in large samples. However, henceforth in this paper, all references to identification refer exclusively to local identification.

If  $B$  is of rank  $p$ , then  $\theta$  is completely identified. If  $B$  is of rank zero, then  $\theta$  is completely unidentified (or completely underidentified). The case in which  $B$  has a nonzero rank, smaller than  $p$ , can be called partial identification (or partial

underidentification). Choi and Phillips (1992) discuss a linear IV model in which some endogenous variables are correlated with the instruments, while others are not. This results in partial identification. The parameters associated with the endogenous variables which are uncorrelated with the instruments are not consistently estimable. The remaining parameters are root- $T$  consistently estimable. What is less intuitive, but also shown by Choi and Phillips is that these latter parameters do not have their usual normal limiting distributions; their limiting distributions are affected by the fact that other parameters are not consistently estimable. Stock and Wright (2000) provide the limiting asymptotic distribution of the estimators in the GMM model, in the presence of partial identification. In the standard consumption CAPM, a canonical application of the GMM method,  $B$  is unlikely to have zero rank, but may well have rank smaller than  $p$  (see Stock and Wright, and section 4 below). So partial identification is an important issue in considering identification in the GMM model.

### **3. Detecting Underidentification.**

The focus of this paper is on tests for detecting a lack of identification. In the linear IV model, it is common to test for instrument relevance by a first-stage F-test (Basmann (1960)). This involves running a regression of the endogenous variables on the instruments and testing the null hypothesis of the joint insignificance of the slope coefficients, by means of an F-test. The null hypothesis is one of a total lack

of identification. A rejection of this hypothesis by no means implies that issues of weak instruments can be ignored. But a failure to reject this hypothesis is a strong indication of identification difficulties. However, I know of no analogous test in the GMM model. The paper provides such a test<sup>3</sup>.

The identification condition which I wish to test is assumption A4, that the matrix  $B$  is of rank  $p$ . I shall do this by testing the null hypothesis of underidentification (partial or complete): that  $\rho(B) = r$ , where the operator  $\rho(\cdot)$  denotes the rank of its argument and  $r < p$ . For this purpose, I define  $\hat{B}(\theta) = T^{-1}\sum_{t=1}^T \frac{d\phi(Y_t, \theta)}{d\theta}$  and

$$\hat{C}(\theta) = T^{-1}\sum_{t=1}^T [vec(\frac{d\phi(Y_t, \theta)}{d\theta}) - vec(\hat{B}(\theta))] [vec(\frac{d\phi(Y_t, \theta)}{d\theta}) - vec(\hat{B}(\theta))]$$

and introduce three more possible assumptions:

Assumption B1:  $T^{1/2}[vec(\hat{B}(\theta_0)) - vec(B)] \rightarrow_d N(0, C)$  where  $C = E([vec(\frac{d\phi(Y_t, \theta_0)}{d\theta}) - B])' [vec(\frac{d\phi(Y_t, \theta_0)}{d\theta}) - B])$  is a finite positive-definite matrix of full rank<sup>4</sup>.

Assumption B2:  $\hat{C}(\theta_0) \rightarrow_p C$ .

Assumption B3:  $V_T(\theta_0) \rightarrow_p A$

---

<sup>3</sup>The S-set works regardless of whether the model is identified or not. However, a means of detecting underidentification is still useful because (a) if the model is identified, consistent point estimates of the parameters are available, (b) if the model is identified, confidence intervals for individual parameters which do not need to be asymptotically conservative are available and (c) the S-set is hard to represent when there are more than 2 parameters.

<sup>4</sup>The requirement that  $C$  be of full rank could be relaxed along the lines proposed by Robin and Smith (2000).

Assumptions B1, B2 and B3 are high level convergence assumptions, as are assumptions A2 and A3 above. In specific models, more primitive conditions may be available that in turn imply that these assumptions hold. Assumptions A3, B1, B2 and B3 all apply only at  $\theta = \theta_0$ . Assumption B3 is a special case of assumption A6.

Let

$$L(\theta, r) = T \min_{P \in \Omega(r)} [\text{vec}(\hat{B}(\theta)) - \text{vec}(P)]' \hat{C}(\theta)^{-1} [\text{vec}(\hat{B}(\theta)) - \text{vec}(P)]$$

where  $\Omega(r)$  is the space of all  $Gk \times p$  matrices of rank  $r$ . If  $\theta_0$  were known,  $L(\theta_0, r)$  would be the statistic for testing the rank of  $B$ , proposed by Cragg and Donald (1997).

My first theorem is a direct consequence of Theorem 1 of Cragg and Donald.

Theorem 1: Under assumptions B1 and B2, if  $B$  is in fact of rank  $r$ , then  $L(\theta_0, r)$  converges to a  $\chi^2$  random variable on  $(Gk - r)(p - r)$  degrees of freedom.

Clearly  $\theta_0$  is not known and is not consistently estimable (under the null hypothesis).

However, under assumptions A3 and B3,  $S_\theta^*(\alpha)$  is a confidence set for  $\theta$  with asymptotic coverage  $1 - \alpha$ . My second theorem follows from this fact, Theorem 1 and the Bonferroni inequality.

Theorem 2: Under assumptions A3, B1, B2 and B3, if  $B$  is in fact of rank  $r$  and

$$L^*(r) = \inf_{\theta \in S_\theta^*(\alpha)} L(\theta, r), \text{ then } \lim_{T \rightarrow \infty} P(L^*(r) \leq F_{\chi^2}(\alpha, (Gk - r)(p - r))) \leq 2\alpha.$$

$L^*(r)$  is a feasible statistic which yields an asymptotically conservative test; the limiting probability of incorrectly rejecting the null is bounded above by  $2\alpha$ . This is the

test statistic proposed in this paper; the test rejects iff  $L^*(r)$  exceeds the  $\chi^2$  critical value. If  $\rho(B)$  is in fact *lower* than the hypothesized rank  $r$ , I conjecture (following Cragg and Donald) that the test will continue to be asymptotically conservative. If  $\rho(B)$  is *higher* than the hypothesized rank  $r$ , the test will reject with probability one, asymptotically. To detect identification difficulties, it is perhaps especially appropriate to test the null that  $\rho(B) = p - 1$ .

### 3.1 Existing Identification Tests and the Linear IV Model.

Consider the simple linear IV model

$$\begin{aligned} y_t &= \beta' x_t + u_t \\ x_t &= \Pi' z_t + v_t, \end{aligned}$$

where  $u_t$  is a scalar error term,  $v_t$  is a  $p \times 1$  error term,  $\zeta_t = (u_t, v_t')'$  is iid with mean zero and finite positive-definite covariance matrix  $\Sigma = \begin{pmatrix} \sigma_{uu} & \Sigma'_{vu} \\ \Sigma_{vu} & \Sigma_{vv} \end{pmatrix}$  (partitioned conformably with  $\zeta_t$ ),  $x_t$  is a  $p \times 1$  vector of endogenous variables,  $z_t$  is a  $k \times 1$  vector of instruments that is uncorrelated with  $\zeta_t$ ,  $k \geq p$  and  $\Pi$  is a  $k \times p$  matrix. The sample size is  $T$ . Also, assume that  $T^{-1} \sum_{t=1}^T \zeta_t \zeta_t' \rightarrow_p \Sigma$ ,  $T^{-1} \sum_{t=1}^T w_t w_t' \rightarrow_p Q$  and  $T^{-1/2} \sum_{t=1}^T \zeta_t \otimes z_t \rightarrow_d N(0, \Sigma \otimes Q)$ , where  $Q = E(z_t z_t')$  is a finite positive-definite matrix of full rank. Lastly, assume that  $\Sigma_{vu} \neq 0$ , so that the regressors in this model are endogenous. The identification condition<sup>5</sup> is that  $\Pi$  is of rank  $p$ .

---

<sup>5</sup>The rank of  $B$  is the same as rank of  $\Pi$ , because I have assumed that  $Q$  is of full rank and

This model is a special case of the GMM model<sup>6</sup>, with  $G = 1$ ,  $\theta = \beta$ ,  $Y_t = (y_t, x_t', z_t')'$ ,  $Z_t = z_t$ ,  $h(Y_t, \theta) = y_t - \beta'x_t$  and  $\phi(Y_t, \theta) = z_t'(y_t - \beta'x_t)$ . Accordingly,  $\frac{d\phi(Y_t, \theta)}{d\theta} = z_t x_t'$ , which does not depend on the parameter vector. This represents an extremely useful simplification, since it in turn means that  $L(\theta, r)$  does not depend on  $\theta$  and has a  $\chi^2$  null limiting distribution as specified in Theorem 1. There is therefore no need to eliminate  $\theta$  by a Bonferroni argument as in Theorem 2; to do so would simply halve the asymptotic size of the test<sup>7</sup>. Accordingly, in the linear IV model, the proposed test is no longer asymptotically conservative and instead has size that is equal to its nominal level, asymptotically. Let  $Z$  be the  $T \times k$  matrix, the  $t$ th row of which is  $z_t'$  and let  $X$  be the  $T \times p$  matrix, the  $t$ th row of which is  $x_t'$ . Since  $C = \text{Var}(\text{vec}(z_t v_t')) = \Sigma_{vv} \otimes Q$ ,  $\hat{C}$  can be set equal to  $\hat{\Sigma}_{vv} \otimes \hat{Q}$ , where  $\hat{Q} = T^{-1} Z' Z$ ,  $\hat{\Sigma}_{vv} = T^{-1} \Sigma_{t=1}^T \hat{v}_t \hat{v}_t'$ ,  $\hat{v}_t = x_t - \hat{\Pi}' z_t$  and  $\hat{\Pi} = (Z' Z)^{-1} Z' X$  is the estimate of  $\Pi$  from a multivariate regression of  $x_t$  on  $z_t$ . So, the proposed test statistic can be written in the form

$$L(r) = T \min_{P \in \Omega(r)} [\text{vec}(T^{-1} Z' X) - \text{vec}(P)]' [\hat{\Sigma}_{vv}^{-1} \otimes \hat{Q}^{-1}] [\text{vec}(T^{-1} Z' X) - \text{vec}(P)].$$

---

because  $E(\frac{d\phi(Y_t, \theta)}{d\theta}) = Q\Pi$ . There is no loss of generality in assuming that  $Q$  is of full rank, since the instruments can always be rotated so that they are mutually orthogonal.

<sup>6</sup>The two-step and continuous updating GMM estimators reduce to the usual TSLS and LIML estimators, respectively.

<sup>7</sup>In the general GMM model, the gradient of the moment condition depends on the parameter vector (which is not consistently estimable without complete identification), necessitating the elimination of the nuisance parameter  $\theta$  by a Bonferroni argument. The linearity of the standard IV model ensures that the gradient of the moment condition does not depend on  $\theta$  and so this difficulty is circumvented.

Here, and for the remainder of this section, I drop the now-degenerate dependence of this statistic on  $\theta$ . In this specific model, there is a closed form expression<sup>8</sup> for  $L(r)$ , since

$$\begin{aligned} L(r) &= T \min_{P \in \Omega(r)} \text{tr}[\hat{\Sigma}_{vv}^{-1}(T^{-1}Z'X - P)\hat{Q}^{-1}(T^{-1}Z'X - P)] \\ &= T \min_{\bar{P} \in \Omega(r)} \text{tr}[(\hat{Q}^{1/2}\hat{\Pi}\hat{\Sigma}_{vv}^{-1/2} - \bar{P})'(\hat{Q}^{1/2}\hat{\Pi}\hat{\Sigma}_{vv}^{-1/2} - \bar{P})] = \sum_{i=1}^{p-r} \lambda_i, \end{aligned}$$

where  $\lambda_i$  is the  $i$ th generalized eigenvalue of  $\hat{\Pi}'Z'Z\hat{\Pi}$  with respect to  $\hat{\Sigma}_{vv}$ , arranged in increasing order ( $\lambda_1 \leq \lambda_2 \leq \dots \lambda_p$ ). This is simply the test statistic which Cragg and Donald (1993) proposed to test the hypothesis that the rank of  $\Pi$  is  $r$ , in the linear IV model<sup>9</sup>. As discussed in the introduction, Cragg and Donald's test has the advantage over the usual first-stage F-test that it can detect partial underidentification, not just complete underidentification. If the researcher wishes to test the hypothesis of complete underidentification in the linear IV model, then the proposed test statistic,  $L(0)$ , is simply  $\text{tr}(\hat{\Sigma}_{vv}^{-1}\hat{\Pi}'Z'Z\hat{\Pi}) = \text{vec}(\hat{\Pi})'(\hat{\Sigma}_{vv}^{-1} \otimes Z'Z)\text{vec}(\hat{\Pi})$ . This is just the usual first-stage F-test statistic<sup>10</sup>.

---

<sup>8</sup>The test statistic in Cragg and Donald (1997) always simplifies to a closed form expression (a solution to an eigenvalue problem), whenever  $\hat{C}$  can be written in the form  $H_1 \otimes H_2$ , where  $H_1$  and  $H_2$  are  $Gk \times Gk$  and  $p \times p$  matrices, respectively. This was pointed out in footnote 4 of Cragg and Donald.

<sup>9</sup>It is also the test statistic which Anderson (1951) proposed for testing the rank of a coefficient matrix in a multivariate least squares regression (here a regression of  $x_t$  on  $z_t$ ).

<sup>10</sup>For  $p > 1$ , this is a multivariate F-statistic (Staiger and Stock (1997)).

## 4. Monte-Carlo Results.

One leading application of the GMM method is to inference in the consumption CAPM of Lucas (1978) with intertemporally separable constant relative risk aversion (CRRA) preferences. In this section, I evaluate the size and power of the proposed test in this model. The Euler equation in this model is as given in equation 2.1, with

$$h(Y_t, \theta) = \delta R_{t+1} \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} - i_G$$

where  $Y_t = (R'_{t+1}, \frac{C_{t+1}}{C_t})'$ ,  $\theta = (\delta, \gamma)'$ ,  $R_t = (R_{1t}, R_{2t}, \dots, R_{Gt})'$  is a  $G \times 1$  vector of gross asset returns,  $C_t$  is nonstoreable consumption,  $i_G$  is a  $G \times 1$  vector of ones,  $\delta$  is the discount factor and  $\gamma$  is the coefficient of relative risk aversion (Hansen and Singleton (1982)). Let  $D_{it}$ ,  $P_{it}$  and  $v_{it}$  denote the dividend, price and price-dividend ratio for asset  $i$  at date  $t$ , so that  $v_{it} = \frac{P_{it}}{D_{it}}$  and  $R_{it} = \frac{P_{it+1} + D_{it+1}}{P_{it}}$ . The  $i$ th element of the Euler equation can be rewritten as

$$v_{it} = E_t \left[ \delta (1 + v_{it+1}) \frac{D_{it+1}}{D_{it}} \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \quad (4.1)$$

Note that if  $(\frac{D_{it+1}}{D_{it}}, \frac{C_{t+1}}{C_t})'$  is iid (independent of anything in the information set at date  $t$ ), then the price-dividend ratio for the  $i$ th asset is constant and is  $v_i^* = [\frac{1}{\delta E[\frac{D_{it+1}}{D_{it}} (\frac{C_{t+1}}{C_t})^{-\gamma}]} - 1]^{-1}$ . The gross return on the  $i$ th asset is  $R_{it} = \frac{D_{it}}{D_{it-1}} \frac{v_i^*}{1+v_i^*}$ .

In GMM estimation of this model,  $\phi(Y_t, \theta) = (\delta R_{t+1} (\frac{C_{t+1}}{C_t})^{-\gamma} - i_G) \otimes Z_t$  and

$$B = \left( \begin{array}{cc} E[(\frac{C_{t+1}}{C_t})^{-\gamma} R_{t+1} \otimes Z_t] & -E[\delta (\frac{C_{t+1}}{C_t})^{-\gamma} \log(\frac{C_{t+1}}{C_t}) R_{t+1} \otimes Z_t] \end{array} \right).$$

Let the vector of instruments consist of an intercept, lagged returns and lagged consumption growth, as is typical in this literature. It is hard to imagine a case in which the matrix  $B$  would have rank zero<sup>11</sup>. But, it is easy to imagine circumstances in which it has rank 1, while the identification condition requires it to have rank 2. Suppose that  $G = 1$  and  $(\frac{D_{it+1}}{D_{it}}, \frac{C_{t+1}}{C_t})'$  is iid and independent of  $Z_t$ . Then  $(R_{t+1}, \frac{C_{t+1}}{C_t})'$  is independent of  $Z_t$ , so  $B$  is a  $k \times 2$  matrix of rank 1, because all of its rows are proportional to the first row<sup>12</sup>. Alternatively, suppose that consumption is constant. Then (regardless of the number of assets  $G$  or the instruments  $Z_t$ ), all the elements of the second column of  $B$  are equal to zero, and the matrix again has rank 1. While consumption is clearly not literally constant,  $\frac{C_{t+1}}{C_t}$  is always very nearly 1 and this is, intuitively, a likely source of the difficulty in estimating  $\gamma$ . Assumptions A3, B1, B2 and B3 will hold in this model, if it is assumed that  $(R_{t+1}, \frac{C_{t+1}}{C_t}, Z_t)'$  is  $m$ -dependent,  $|\gamma| \leq \bar{\gamma}$  and  $R_{t+1}, (\frac{C_{t+1}}{C_t})^{\bar{\gamma}}, \log(\frac{C_{t+1}}{C_t})$  and  $Z_t$  all have  $8 + \omega$  finite moments, for some  $\omega > 0$ .

To simulate data from the consumption CAPM, I follow the approach of Tauchen

---

<sup>11</sup>This is the main reason why this paper does not simply focus on testing the hypothesis that  $B$  has zero rank, like the first-stage F-test in the linear IV model. More generally, the point is that in the linear IV model the constant is an instrument which identifies only the intercept parameter in the structural equation: this parameter is of usually of little interest and its identifiability is overlooked by simply projecting out the constant term. No such simplification is available in the general nonlinear GMM context; so if the constant has identifying power, the matrix  $B$  must have nonzero rank.

<sup>12</sup>If  $G > 1$ , then the model will typically be formally identified even if  $\frac{D_{it+1}}{D_{it}}$  and  $\frac{C_{t+1}}{C_t}$  are iid. I would expect this identification to be very weak.

and Hussey (1991) (also used in Tauchen (1986), Kocherlakota (1990), Hansen, Heaton and Yaron (1996) and Stock and Wright (2000)). This involves fitting a 16-state Markov chain to consumption and stock-market dividend growth calibrated so as to approximate the first-order VAR:

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

where  $D_t$  is the stock-market dividend at date  $t$  and  $(u_{ct}, u_{dt})'$  is iid normal with mean zero and variance  $\Lambda$ . Consumption growth and dividend growth are the state variables. 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 (from equation 4.1). In this way, time series of consumption growth and stock returns may be simulated<sup>13</sup>.

The Monte-Carlo simulation in this subsection is based on the Euler equation in this consumption CAPM for a single asset (stocks,  $G=1$ ) and with instruments  $Z_t = (1, R_t, \frac{C_t}{C_{t-1}})$ . I consider three sample sizes:  $T=100, 500$  and  $1000$ . The sample size of 100 is most commonly used in Monte-Carlo simulations, as it is approximately equal to the available sample size for U.S. annual data. The larger sample sizes are also considered so as to include models with less severe identification problems. To

---

<sup>13</sup>I am grateful to George Tauchen for his Gauss code for implementing this.

complete the specification, I must specify values of  $\delta$ ,  $\gamma$ ,  $\mu$ ,  $\Phi$  and  $\Lambda$ . I consider 5 combinations of these parameters, as listed in Table 1, and refer to the resulting models as models M1-M5, respectively.

Models M1 and M2 set  $\Phi$  equal to a matrix of zeros so as to wipe out all serial dependence in the growth rates of consumption and dividends. Note that in these models, consumption growth and dividend growth are still contemporaneously correlated. The true rank of  $B$  is 1 in these models; the models are not fully identified and so the rejection rate of the proposed test, used to test the hypothesis that  $\rho(B) = 1$ , represents the size of the test.

Kocherlakota (1990) chose values of  $\mu$ ,  $\Phi$  and  $\Lambda$  by fitting a bivariate VAR(1) to historical U.S. annual real dividend growth and real consumption growth data. Models M3 and M4 use his values of  $\mu$ ,  $\Phi$  and  $\Lambda$ . The true rank of  $B$  is 2 in these models; the models are identified and so the rejection rate of the proposed test, used to test the hypothesis that  $\rho(B) = 1$ , represents the power of the test. However,  $B$  is close to being rank-deficient and so weak identification problems may arise. Hansen, Heaton and Yaron and Stock and Wright (2000) show that the conventional Gaussian asymptotic theory works very poorly in small samples in these models (e.g. the estimators are biased and have nonnormal sampling distributions). It is of interest to see how the power of the proposed test is related to the quality of the conventional asymptotic theory.

Model M5 uses values of  $\mu$ ,  $\Phi$  and  $\Lambda$  for which Tauchen (1986) found that the conventional asymptotic theory works well, even in small samples. My model M5 is Tauchen's experiment 1B. The true rank of  $B$  is 2 in this model; the model is identified and so the rejection rate of the proposed test, used to test the hypothesis that  $\rho(B) = 1$ , represents the power of the test.

In models M1, M3 and M5,  $\delta = 0.97$  and  $\gamma = 1.3$ . In models M2 and M4,  $\delta = 1.139$  and  $\gamma = 13.7$ . These latter parameter values were advocated by Kocherlakota (1990) as a solution to the equity premium puzzle.

The simulated rejection rates of  $L^*(1)$  are reported in Tables 2-6 for models M1-M5, respectively. The rejection rates represent the size of the test in Tables 2 and 3 (models M1 and M2) and represent the power of the test in the other tables. In these tables, I also report:

- (a) the 10th, 50th and 90th percentiles of  $\hat{\delta}_{TS}$ ,  $\hat{\gamma}_{TS}$ ,  $\hat{\delta}_{CU}$  and  $\hat{\gamma}_{CU}$  the two-step and continuous-updating GMM estimators of  $\delta$  and  $\gamma$ ,
- (b) the 10th, 50th and 90th percentiles of the associated t statistics ( $t_{\hat{\delta}_{TS}}$ ,  $t_{\hat{\gamma}_{TS}}$ ,  $t_{\hat{\delta}_{CU}}$  and  $t_{\hat{\gamma}_{CU}}$ ), which will be close to their theoretical values (-1.282, 0 and 1.282) if the conventional asymptotic theory is working well,
- (c) the Kolmogorov-Smirnov statistics comparing the sampling distributions of the t statistics to the standard normal distribution and
- (d) the rejection rates of  $J_{TS}$  and  $J_{CU}$ .

All tests have a 5% nominal size.

In the simulations without complete identification (models M1 and M2), the estimators of  $\delta$  and  $\gamma$  can exhibit severe median bias and the sampling distribution of the associated t-statistics can be very far from the standard normal distribution predicted by conventional asymptotic theory. These problems are much worse with the parameters  $\delta = 1.139$ ,  $\gamma = 13.7$  (model M2) than with the parameters  $\delta = 0.97$ ,  $\gamma = 1.3$  (model M1). Indeed, in model M1,  $t_{\hat{\delta}_{TS}}$  and  $t_{\hat{\gamma}_{TS}}$  are both close to being standard normal in samples of size 100 and 500, but then become less normal as the sample size increases to 1000, in a clear sign of a lack of identification. In all cases, the actual size of the proposed test is well below the nominal 5% level (not surprisingly, as the test is asymptotically conservative).

In model M3, the predictions of conventional asymptotic theory work better as the sample size increases. In a sample of size 100, the estimators exhibit some median bias and the sampling distribution of the associated t-statistics is somewhat nonnormal. These problems get less severe as the sample size increases. The proposed test has very little power in the sample of size 100, but its power increases with the sample size. In this model, with a sample size of 1000, conventional asymptotic theory works well and the proposed test has 90.6% power.

In model M4, the predictions of conventional asymptotic theory work less well. In a sample size of 100, the estimators exhibit considerable median bias and the

sampling distribution of the associated t-statistics is far from being standard normal. Again the problems get less severe as the sample size increases. The proposed test has very little power in the sample of size 100. Its power increases with the sample size, but (appropriately) does so more slowly than in model M3.

In model M5, I find (like Tauchen (1986)) that the conventional asymptotic theory works well in all sample sizes. Correspondingly, the proposed test has a very high rejection rate in all sample sizes.

In summary, the proposed test is conservative under the null while, in the simulations where the identification condition is satisfied (models M3-M5), the its power is strongly associated with the quality of the conventional asymptotic theory. The relationship is not a perfect one; no more than it is with the first-stage F-test in the linear IV model (Staiger and Stock (1997))<sup>14</sup>.

Returns on a riskfree asset can also be generated by the algorithm of Tauchen and Hussey (1991). I do not run simulations in which both stocks and a riskfree asset are in the Euler equation ( $G = 2$ ). This is because in such a model, the matrix  $B$  will always have rank 2, whenever there is an intercept in the set of instruments (given that consumption is not constant). This is true even if all the other instruments are random numbers. In this sense, the null hypothesis of underidentification is not

---

<sup>14</sup>The first-stage F-test is an imperfect measure of the quality of the conventional asymptotic theory in the linear IV model. This is true even abstracting from issues of partial identification (e.g. in a model with a single endogenous regressor).

well defined when  $G = 2$ . Of course, such a model is likely to be only very weakly identified.

## 5. Conclusion.

In the linear IV model, the first-stage F-test is an important diagnostic for checking the quality of the instruments. In the existing literature, there is no analog of this test in the GMM model. This paper has proposed a test of the null of partial or complete underidentification in the general GMM model and has argued that it is a useful diagnostic for identification difficulties in this model.

## References.

- Anderson, T.W. (1951): Estimating Linear Restrictions on Regression Coefficients for the Multivariate Normal Distribution, *Annals of Mathematical Statistics*, 22, pp.327-351.
- 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.
- Basman, R.L. (1960): On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics, *Journal of the American Statistical Association*, 55, pp.650-659.

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.

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.

Cragg, J.G. and S.G. Donald (1997): Inferring the Rank of a Matrix, *Journal of Econometrics*, 76, pp.223-250.

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

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.

- Hansen, L.P. and K. Singleton (1982): Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models, *Econometrica*, 50, pp.1269-1286.
- 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.
- Maddala, G.S. and J. Jeong (1992): On the Exact Small Sample Distribution of the Instrumental Variable Estimator, *Econometrica*, 60, pp.181-183.
- Nelson, C.R. and R. Startz (1990a): Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator, *Econometrica*, 58, pp.967-976.
- Nelson, C.R. and R. Startz (1990b): The Distribution of the Instrumental Variable Estimator and Its t-ratio When the Instrument is a Poor One, *Journal of Business*, 63, pp.5125-5140.
- Phillips, P.C.B. (1980): The Exact Finite Sample Density of Instrumental Variables Estimators in an Equation with  $n+1$  Endogenous Variables, *Econometrica*, 48, pp.861-878.
- Phillips, P.C.B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model", in *Handbook of Econometrics*, Vol. 1, eds. Z. Griliches and M.D. Intriligator, North-Holland, Amsterdam.
- Richardson, D.H. (1968): The Exact Distribution of a Structural Coefficient Estima-

- tor, *Journal of the American Statistical Association*, 63, pp.1214-1226.
- Robin, J.M. and R.J. Smith (2000): Tests of Rank, *Econometric Theory*, 16, pp.151-175.
- Sawa, T. (1969): The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators, *Journal of the American Statistical Association*, 64, pp.923-936.
- Shea, J. (1997): Instrument Relevance in Multivariate Linear Models: A Simple Measure, *Review of Economics and Statistics*, 79, pp.348-352.
- 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.000-000.
- 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.
- Zivot, E. and J. Wang (1998): Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments, *Econometrica*, 66, pp.1389-1404.

Table 1: Parameter Values in the Models

Model	$\mu$	$\Phi$	$\Lambda$	$\delta$	$\gamma$
M1	$\begin{pmatrix} 0.018 \\ 0.013 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{pmatrix}$	0.97	1.3
M2	$\begin{pmatrix} 0.018 \\ 0.013 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.0012 & 0.0017 \\ 0.0017 & 0.0146 \end{pmatrix}$	1.139	13.7
M3	$\begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix}$	$\begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}$	$\begin{pmatrix} 0.0012 & 0.00177 \\ 0.00177 & 0.014 \end{pmatrix}$	0.97	1.3
M4	$\begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix}$	$\begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}$	$\begin{pmatrix} 0.0012 & 0.00177 \\ 0.00177 & 0.014 \end{pmatrix}$	1.139	13.7
M5	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -0.5 & 0 \\ 0 & -0.5 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$	0.97	1.3

Note: Models M3 and M4 were considered by Kocherlakota (1990). Models M1 and M2 set  $\Phi$  to a matrix of zeros and adjust  $\mu$  and  $\Lambda$  so that the growth rates of consumption and dividends are the same as in models M3 and M4. Model M5 is experiment 1B of Tauchen (1996). The models are discussed further in the text.

Table 2: Simulation results in model M1.

	$T=100$	$T=500$	$T=1000$
10th %tile of $\hat{\delta}_{TS}$	0.886	0.884	0.894
Median of $\hat{\delta}_{TS}$	0.971	0.971	0.972
90th %tile of $\hat{\delta}_{TS}$	1.036	1.035	1.025
10th %tile of $\hat{\gamma}_{TS}$	-2.968	-3.111	-2.621
Median of $\hat{\gamma}_{TS}$	1.264	1.300	1.424
90th %tile of $\hat{\gamma}_{TS}$	5.331	5.415	4.715
10th %tile of $\hat{\delta}_{CU}$	0.794	0.765	0.790
Median of $\hat{\delta}_{CU}$	0.967	0.965	0.967
90th %tile of $\hat{\delta}_{CU}$	1.068	1.068	1.065
10th %tile of $\hat{\gamma}_{CU}$	-6.674	-7.753	-6.668
Median of $\hat{\gamma}_{CU}$	1.284	1.207	1.281
90th %tile of $\hat{\gamma}_{CU}$	10.367	10.578	9.903
10th %tile of $t_{\hat{\delta}_{TS}}$	-1.379	-1.430	-1.379
Median of $t_{\hat{\delta}_{TS}}$	0.025	0.012	0.025
90th %tile of $t_{\hat{\delta}_{TS}}$	1.122	1.131	1.092
10th %tile of $t_{\hat{\gamma}_{TS}}$	-1.236	-1.258	-1.292
Median of $t_{\hat{\gamma}_{TS}}$	-0.012	0.000	0.030
90th %tile of $t_{\hat{\gamma}_{TS}}$	1.258	1.308	1.218
10th %tile of $t_{\hat{\delta}_{CU}}$	-2.443	-2.525	-2.439
Median of $t_{\hat{\delta}_{CU}}$	-0.079	-0.134	-0.070
90th %tile of $t_{\hat{\delta}_{CU}}$	1.485	1.513	1.484
10th %tile of $t_{\hat{\gamma}_{CU}}$	-2.075	-2.021	-2.029
Median of $t_{\hat{\gamma}_{CU}}$	-0.011	-0.050	-0.008
90th %tile of $t_{\hat{\gamma}_{CU}}$	2.183	2.138	2.228
$KS(t_{\hat{\delta}_{TS}})$	0.037	0.032	0.082
$KS(t_{\hat{\gamma}_{TS}})$	0.017	0.020	0.072
$KS(t_{\hat{\delta}_{CU}})$	0.127	0.150	0.152
$KS(t_{\hat{\gamma}_{CU}})$	0.114	0.129	0.117
Rej Rate of $J_{TS}$ test	1.6	1.6	1.7
Rej Rate of $J_{CU}$ test	0.6	0.2	0.4
Rej Rate of Proposed test	0.7	0.9	1.0

Notes: In this model, the rejection rate of the proposed test represents its size.  $KS(\cdot)$  refers to the Kolmogorov-Smirnov statistic comparing the sampling distribution of the argument with a standard normal distribution. All tests have 5% nominal size. 1,000 replications were conducted in each simulation.

Table 3: Simulation results in model M2.

	$T=100$	$T=500$	$T=1000$
10th %tile of $\hat{\delta}_{TS}$	0.918	0.922	0.919
Median of $\hat{\delta}_{TS}$	1.002	1.005	1.001
90th %tile of $\hat{\delta}_{TS}$	1.076	1.082	1.070
10th %tile of $\hat{\gamma}_{TS}$	-2.653	-2.734	-2.630
Median of $\hat{\gamma}_{TS}$	1.474	1.670	1.496
90th %tile of $\hat{\gamma}_{TS}$	6.445	6.989	6.154
10th %tile of $\hat{\delta}_{CU}$	0.353	0.331	0.294
Median of $\hat{\delta}_{CU}$	1.043	1.036	1.029
90th %tile of $\hat{\delta}_{CU}$	1.191	1.144	1.134
10th %tile of $\hat{\gamma}_{CU}$	1.685	1.642	1.675
Median of $\hat{\gamma}_{CU}$	13.152	14.018	13.852
90th %tile of $\hat{\gamma}_{CU}$	74.698	61.583	64.901
10th %tile of $t_{\hat{\delta}_{TS}}$	-3.220	-5.810	-7.083
Median of $t_{\hat{\delta}_{TS}}$	-1.894	-3.173	-3.644
90th %tile of $t_{\hat{\delta}_{TS}}$	-0.542	-0.654	-0.901
10th %tile of $t_{\hat{\gamma}_{TS}}$	-2.371	-2.336	-2.427
Median of $t_{\hat{\gamma}_{TS}}$	-1.289	-1.244	-1.318
90th %tile of $t_{\hat{\gamma}_{TS}}$	-0.368	-0.327	-0.394
10th %tile of $t_{\hat{\delta}_{CU}}$	-11.536	-20.280	-22.966
Median of $t_{\hat{\delta}_{CU}}$	-1.348	-2.419	-3.024
90th %tile of $t_{\hat{\delta}_{CU}}$	0.410	0.047	-0.085
10th %tile of $t_{\hat{\gamma}_{CU}}$	-1.833	-2.020	-2.045
Median of $t_{\hat{\gamma}_{CU}}$	-0.027	0.020	0.011
90th %tile of $t_{\hat{\gamma}_{CU}}$	7.311	5.511	5.765
$KS(t_{\hat{\delta}_{TS}})$	0.656	0.735	0.780
$KS(t_{\hat{\gamma}_{TS}})$	0.549	0.539	0.557
$KS(t_{\hat{\delta}_{CU}})$	0.427	0.551	0.583
$KS(t_{\hat{\gamma}_{CU}})$	0.210	0.173	0.203
Rej Rate of $J_{TS}$ test	1.1	1.0	0.9
Rej Rate of $J_{CU}$ test	0.9	1.2	2.0
Rej Rate of Proposed test	0.2	0.2	0.6

Notes: As for Table 2.

Table 4: Simulation results in Model M3.

	$T=100$	$T=500$	$T=1000$
10th %tile of $\hat{\delta}_{TS}$	0.917	0.941	0.949
Median of $\hat{\delta}_{TS}$	0.978	0.972	0.970
90th %tile of $\hat{\delta}_{TS}$	1.026	0.995	0.989
10th %tile of $\hat{\gamma}_{TS}$	-1.315	-0.180	0.227
Median of $\hat{\gamma}_{TS}$	1.699	1.385	1.321
90th %tile of $\hat{\gamma}_{TS}$	4.213	2.626	2.262
10th %tile of $\hat{\delta}_{CU}$	0.847	0.934	0.947
Median of $\hat{\delta}_{CU}$	0.972	0.970	0.969
90th %tile of $\hat{\delta}_{CU}$	1.035	0.996	0.989
10th %tile of $\hat{\gamma}_{CU}$	-4.254	-0.501	0.144
Median of $\hat{\gamma}_{CU}$	1.394	1.298	1.258
90th %tile of $\hat{\gamma}_{CU}$	4.843	2.662	2.259
10th %tile of $t_{\hat{\delta}_{TS}}$	-1.100	-1.238	-1.272
Median of $t_{\hat{\delta}_{TS}}$	0.217	0.079	-0.011
90th %tile of $t_{\hat{\delta}_{TS}}$	1.260	1.302	1.319
10th %tile of $t_{\hat{\gamma}_{TS}}$	-1.091	-1.167	-1.269
Median of $t_{\hat{\gamma}_{TS}}$	0.203	0.081	0.032
90th %tile of $t_{\hat{\gamma}_{TS}}$	1.437	1.310	1.340
10th %tile of $t_{\hat{\delta}_{CU}}$	-2.004	-1.488	-1.408
Median of $t_{\hat{\delta}_{CU}}$	0.067	0.000	-0.052
90th %tile of $t_{\hat{\delta}_{CU}}$	1.441	1.354	1.332
10th %tile of $t_{\hat{\gamma}_{CU}}$	-1.847	-1.462	-1.377
Median of $t_{\hat{\gamma}_{CU}}$	0.061	-0.003	-0.056
90th %tile of $t_{\hat{\gamma}_{CU}}$	1.753	1.345	1.368
$KS(t_{\hat{\delta}_{TS}})$	0.092	0.045	0.027
$KS(t_{\hat{\gamma}_{TS}})$	0.086	0.044	0.024
$KS(t_{\hat{\delta}_{CU}})$	0.087	0.037	0.035
$KS(t_{\hat{\gamma}_{CU}})$	0.084	0.038	0.037
% Rej Rate of $J_{TS}$ test	1.3	1.8	3.2
% Rej Rate of $J_{CU}$ test	1.1	2.9	4.7
% Rej Rate of Proposed test	6.2	52.0	90.6

Notes: In this model, the rejection rate of the proposed test represents its power.  $KS(\cdot)$  refers to the Kolmogorov-Smirnov statistic comparing the sampling distribution of the argument with a standard normal distribution. All tests have 5% nominal size. 1,000 replications were conducted in each simulation.

Table 5: Simulation results in Model M4.

	$T=100$	$T=500$	$T=1000$
10th %tile of $\hat{\delta}_{TS}$	1.026	1.096	1.108
Median of $\hat{\delta}_{TS}$	1.094	1.126	1.133
90th %tile of $\hat{\delta}_{TS}$	1.165	1.163	1.158
10th %tile of $\hat{\gamma}_{TS}$	5.167	9.938	10.911
Median of $\hat{\gamma}_{TS}$	9.598	12.502	13.150
90th %tile of $\hat{\gamma}_{TS}$	16.577	17.224	17.060
10th %tile of $\hat{\delta}_{CU}$	0.916	1.093	1.110
Median of $\hat{\delta}_{CU}$	1.111	1.130	1.135
90th %tile of $\hat{\delta}_{CU}$	1.202	1.170	1.163
10th %tile of $\hat{\gamma}_{CU}$	8.456	10.612	11.292
Median of $\hat{\gamma}_{CU}$	13.166	13.452	13.632
90th %tile of $\hat{\gamma}_{CU}$	41.950	22.112	18.323
10th %tile of $t_{\hat{\delta}_{TS}}$	-2.018	-1.751	-1.710
Median of $t_{\hat{\delta}_{TS}}$	-0.768	-0.466	-0.303
90th %tile of $t_{\hat{\delta}_{TS}}$	0.273	0.583	0.739
10th %tile of $t_{\hat{\gamma}_{TS}}$	-2.011	-1.627	-1.550
Median of $t_{\hat{\gamma}_{TS}}$	-0.752	-0.391	-0.238
90th %tile of $t_{\hat{\gamma}_{TS}}$	0.278	0.727	1.028
10th %tile of $t_{\hat{\delta}_{CU}}$	-2.879	-1.849	-1.601
Median of $t_{\hat{\delta}_{CU}}$	-0.473	-0.324	-0.211
90th %tile of $t_{\hat{\delta}_{CU}}$	0.686	0.823	0.886
10th %tile of $t_{\hat{\gamma}_{CU}}$	-1.513	-1.375	-1.341
Median of $t_{\hat{\gamma}_{CU}}$	-0.087	-0.076	-0.031
90th %tile of $t_{\hat{\gamma}_{CU}}$	2.924	1.545	1.414
$KS(t_{\hat{\delta}_{TS}})$	0.314	0.188	0.144
$KS(t_{\hat{\gamma}_{TS}})$	0.330	0.170	0.107
$KS(t_{\hat{\delta}_{CU}})$	0.188	0.129	0.091
$KS(t_{\hat{\gamma}_{CU}})$	0.102	0.043	0.025
% Rej Rate of $J_{TS}$ test	19.1	14.8	10.3
% Rej Rate of $J_{CU}$ test	2.8	4.5	4.6
% Rej Rate of Proposed test	6.0	27.3	52.1

Notes: As for Table 4.

Table 6: Simulation results in Model M5.

	$T=100$	$T=500$	$T=1000$
10th %tile of $\hat{\delta}_{TS}$	0.955	0.963	0.965
Median of $\hat{\delta}_{TS}$	0.970	0.970	0.970
90th %tile of $\hat{\delta}_{TS}$	0.984	0.976	0.974
10th %tile of $\hat{\gamma}_{TS}$	1.084	1.200	1.226
Median of $\hat{\gamma}_{TS}$	1.269	1.299	1.302
90th %tile of $\hat{\gamma}_{TS}$	1.587	1.432	1.386
10th %tile of $\hat{\delta}_{CU}$	0.955	0.963	0.965
Median of $\hat{\delta}_{CU}$	0.970	0.970	0.970
90th %tile of $\hat{\delta}_{CU}$	0.984	0.976	0.976
10th %tile of $\hat{\gamma}_{CU}$	1.099	1.203	1.229
Median of $\hat{\gamma}_{CU}$	1.306	1.304	1.304
90th %tile of $\hat{\gamma}_{CU}$	1.639	1.439	1.390
10th %tile of $t_{\hat{\delta}_{TS}}$	-1.349	-1.381	-1.342
Median of $t_{\hat{\delta}_{TS}}$	-0.012	0.003	-0.017
90th %tile of $t_{\hat{\delta}_{TS}}$	1.315	1.332	1.326
10th %tile of $t_{\hat{\gamma}_{TS}}$	-1.364	-1.322	-1.289
Median of $t_{\hat{\gamma}_{TS}}$	-0.149	-0.006	0.027
90th %tile of $t_{\hat{\gamma}_{TS}}$	1.194	1.368	1.285
10th %tile of $t_{\hat{\delta}_{CU}}$	-1.417	-1.392	-1.342
Median of $t_{\hat{\delta}_{CU}}$	-0.030	-0.003	-0.040
90th %tile of $t_{\hat{\delta}_{CU}}$	1.301	1.335	1.328
10th %tile of $t_{\hat{\gamma}_{CU}}$	-1.260	-1.291	-1.214
Median of $t_{\hat{\gamma}_{CU}}$	0.030	0.044	0.072
90th %tile of $t_{\hat{\gamma}_{CU}}$	1.347	1.406	1.349
$KS(t_{\hat{\delta}_{TS}})$	0.015	0.029	0.025
$KS(t_{\hat{\gamma}_{TS}})$	0.076	0.024	0.026
$KS(t_{\hat{\delta}_{CU}})$	0.024	0.030	0.024
$KS(t_{\hat{\gamma}_{CU}})$	0.035	0.035	0.040
% Rej Rate of $J_{TS}$ test	5.2	4.2	4.8
% Rej Rate of $J_{CU}$ test	5.5	4.3	5.0
% Rej Rate of Proposed test	83.5	100.0	100.0

Notes: As for Table 4.