Econometrica, Vol. 73, No. 4 (July, 2005), 1103–1123

TESTING PARAMETERS IN GMM WITHOUT ASSUMING THAT THEY ARE IDENTIFIED

BY FRANK KLEIBERGEN[1]

We propose a generalized method of moments (GMM) Lagrange multiplier statistic, i.e., the K statistic, that uses a Jacobian estimator based on the continuous updating estimator that is asymptotically uncorrelated with the sample average of the moments. Its asymptotic χ^2 distribution therefore holds under a wider set of circumstances, like weak instruments, than the standard full rank case for the expected Jacobian under which the asymptotic χ^2 distributions of the traditional statistics are valid. The behavior of the K statistic can be spurious around inflection points and maxima of the objective function. This inadequacy is overcome by combining the K statistic with a statistic that tests the validity of the moment equations and by an extension of Moreira's (2003) conditional likelihood ratio statistic toward GMM. We conduct a power comparison to test for the risk aversion parameter in a stochastic discount factor model and construct its confidence set for observed consumption growth and asset return series.

KEYWORDS: Weak instruments, size distortions, covariance matrix estimators, stochastic discount factors.

1. INTRODUCTION

THE GENERALIZED METHOD OF MOMENTS (GMM) of Hansen (1982) offers a flexible estimation framework in which many econometric models can be cast. This alleviates statistical inference in these models because we can use the estimators and test statistics developed for GMM. The limiting distributions of these are constructed under a central limit theorem for the moments and a full rank assumption of the expected Jacobian. The rank assumption is often barely satisfied. Simulation experiments of such instances show that the empirical distributions of GMM estimators and test statistics are then quite different from their normal or χ^2 limiting distributions; see, e.g., Hansen, Heaton, and Yaron (1996). We therefore relax the full rank Jacobian assumption and obtain normal limiting distributions under less stringent regularity conditions.

We analyze the derivative of the objective function of the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996) to obtain a Lagrange multiplier (LM) or score statistic, which we refer to as the K statistic; see Kleibergen (2002). Instead of the sample average of the derivatives of the moments, the K statistic uses a Jacobian estimator based on the CUE. As recognized elsewhere (see, e.g., Brown and Newey (1998) and Donald and Newey (2000)), this Jacobian estimator is asymptotically uncorrelated and,

[1]I thank participants of seminars at Brown, Harvard–MIT, Nuffield College, Princeton, USC, UCSD, Vienna, Yale, UK Econometric Study Group, EC^2 at CORE, ESEMs at Lausanne and Venice, and, especially, the editor and referees for helpful comments and suggestions. This research was partly funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models."

thus, independent of the average moment vector. The usual asymptotic χ^2 distribution of the K statistic therefore holds in a wider set of circumstances than the standard full rank case for the Jacobian. This set includes the weak instrument asymptotics of Staiger and Stock (1997) and Stock and Wright (2000).

The outline of the paper is as follows. In the second section, we discuss GMM and make a novel assumption under which we construct the limiting distributions. In the third section, we obtain the K statistic and construct its limit behavior. In the fourth section, we discuss the specification of the covariance matrix estimators that are to be used in the K statistic. The fifth section deals with a spurious power decline of the K statistic. Because the K statistic is a quadratic form of the derivative of the objective function, it is equal to zero at both the minimum, maximum and the inflection points of the objective function. The latter zero values result while the moment conditions are not satisfied, so the results from the K statistic are then spurious. They can, however, affect the discriminatory power of the K statistic when we do not account for them appropriately. We therefore suggest combinations of the K statistic with statistics that test the moment equation such as a GMM extension of Moreira's (2003) conditional likelihood ratio statistic. In the sixth section, we discuss the construction of confidence sets. The seventh section conducts a power comparison to test the risk aversion parameter in a stochastic discount factor model and constructs its confidence set for observed data. Finally, the eighth section concludes.

We use the following notation throughout the paper: $E(a)$ is the expected value of the random variable a, $\text{vec}(A)$ stands for the column vectorization of the $T \times n$ dimensional matrix A, $\text{vec}(A) = (a_1' \cdots a_n')'$ when $A = (a_1 \cdots a_n)$, $P_A = A(A'A)^{-1}A'$, and $M_A = I_T - P_A$ for a full rank matrix A and the $T \times T$ identity matrix I_T, $|_a$ stands for evaluated in a, $\rightarrow_p$ indicates convergence in probability, and $\rightarrow_d$ indicates convergence in distribution.

2. GENERALIZED METHOD OF MOMENTS

We consider the estimation of the $m \times 1$ parameter vector $\theta = (\theta_1 \cdots \theta_m)'$, whose parameter region is the $\mathbb{R}^m$ and for which the $k_f \times 1$ dimensional moment equation

$$(1) \qquad E(f(\theta_0, Y_t)) = 0$$

holds. The data vector Y_t is observed at time/individual t. The number of equations k_f exceeds or is equal to the number of parameters m. The $k_f \times 1$ dimensional vector function f of θ is finite for finite values of θ, continuous, and twice continuous differentiable. The unique value of θ, at which (1) holds, is equal to θ_0. To estimate θ in (1), we use Hansen's (1982) GMM.

For a data set $(Y_t, t = 1, \ldots, T)$, the objective function for the CUE reads

$$(2) \qquad Q(\theta) = f_T(\theta, Y)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta, Y),$$

with $f_T(\theta, Y) = \sum_{t=1}^{T} f_t(\theta)$, $f_t(\theta) = f(\theta, Y_t)$, and $V_{ff}(\theta)$ is the positive definite covariance matrix of $f_T(\theta, Y)$,

$$V_{ff}(\theta) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}} f_T(\theta, Y)\right), \tag{3}$$

while $\hat{V}_{ff}(\theta)$ is a consistent estimator of $V_{ff}(\theta)$,

$$\hat{V}_{ff}(\theta) \underset{p}{\to} V_{ff}(\theta). \tag{4}$$

ASSUMPTION 1: *The $k_f \times 1$ dimensional derivative of $f_t(\theta_0)$ with respect to θ_i,*

$$q_{i,t}(\theta_0) = \frac{\partial f_t(\theta_0)}{\partial \theta_i} : k_f \times 1, \quad i = 1, \ldots, m, \tag{5}$$

is such that the joint limiting behavior of the sums of the series $\bar{f}_t(\theta_0) = f_t(\theta_0) - E(f_t(\theta_0))$ and $\bar{q}_t(\theta_0) = (\bar{q}_{1t}(\theta_0)' \cdots \bar{q}_{mt}(\theta_0)')'$, with $\bar{q}_{it}(\theta_0) = q_{it}(\theta_0) - E(q_{it}(\theta_0))$, accords with the central limit theorem

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \bar{f}_t(\theta_0) \\ \bar{q}_t(\theta_0) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_f \\ \psi_{\theta_0} \end{pmatrix}, \tag{6}$$

where $\psi_f : k_f \times 1$, $\psi_\theta : k_\theta \times 1$, $k_\theta = m k_f$,

$$\begin{pmatrix} \psi_f \\ \psi_\theta \end{pmatrix} \sim N(0, V(\theta_0)), \tag{7}$$

and $V(\theta)$ is a positive semidefinite symmetric $(k_f + k_\theta) \times (k_f + k_\theta)$ matrix,

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix}, \tag{8}$$

with $V_{\theta f}(\theta) = V_{f\theta}(\theta)' = (V_{\theta f,1}(\theta)' \cdots V_{\theta f,m}(\theta)')'$, $V_{\theta\theta}(\theta) = (V_{\theta\theta,ij}(\theta))$, $i, j = 1, \ldots, m$, and $V_{ff}(\theta)$, $V_{\theta f,i}(\theta)$, $V_{\theta\theta,ij}(\theta)$ are $k_f \times k_f$ dimensional matrices for $i, j = 1, \ldots, m$, and

$$V(\theta) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}} \begin{pmatrix} f_T(\theta, Y) \\ \operatorname{vec}(q_T(\theta, Y)) \end{pmatrix}\right), \tag{9}$$

with $q_T(\theta_0, Y) = \partial f_T(\theta, Y)/\partial\theta'|_{\theta_0} = \sum_{t=1}^{T} (q_{1t}(\theta_0) \cdots q_{mt}(\theta_0))$.

Assumption 1 defines $\bar{q}_t(\theta)$, $t = 1, \ldots, T$, and requests a central limit theorem to hold for it. Assumption 1 is satisfied under weak conditions for

$\bar{f}_t(\theta_0)$ and $\bar{q}_t(\theta_0)$. Sufficient conditions that ensure that Assumption 1 holds are that (i) the rth moment of the absolute value of $\bar{f}_t(\theta_0)$ and $\bar{q}_{i,t}(\theta_0)$, $i = 1, \ldots, m$, is finite for some $r > 2$, (ii) the φ or α-mixing coefficients of $\bar{f}_t(\theta_0)$ and $\bar{q}_{i,t}(\theta_0)$, $i = 1, \ldots, m$, are of size $r/(r-1)$ with $r > 1$, (iii) $V(\theta)$ is well defined, and (iv) the average value of the outer product of $(\bar{f}_t(\theta_0)' \; \bar{q}_t(\theta_0)')'$ converges in probability to $V(\theta)$; see, e.g., Gordin (1969), White and Domowitz (1984), and White (1984, Theorem 5.19).

Assumption 1 is an extension of the assumption that is usually made to obtain the limiting distributions of GMM estimators. Traditionally only a central limit theorem for $f_t(\theta_0)$ is assumed to hold; see, e.g., Hansen (1982), Newey and McFadden (1994), and Stock and Wright (2000). Alongside that assumption, Assumption 1 also imposes a distribution on the limit behavior of $\bar{q}_t(\theta_0)$. It replaces the identification assumption of a full rank value of the expected Jacobian.

The covariance matrix $V(\theta_0)$ in Assumption 1 is assumed to be positive semidefinite and can thus be singular. A singular covariance matrix $V(\theta)$ occurs in several cases, like, for example, when $q_{i,t}(\theta_0) = E(q_{i,t}(\theta_0))$, which results when θ_i is a parameter that belongs to a variable that is part of the set of instruments so that $q_{i,t}(\theta_0)$ is a function of the instruments and when elements of $(f_t(\theta_0)' \; q_t(\theta_0)')'$ are linear combinations of other elements of $(f_t(\theta_0)' \; q_t(\theta_0)')'$, which occurs in panel autoregressive models where $q_t(\theta_0)$ consists of instruments interacted with lagged dependent variables that appear in $f_t(\theta_0)$ as well. The possible singularity of $V(\theta)$ is a weak assumption but important for practical purposes. The singularity of $V(\theta)$ does not result from the quality of the instruments.

3. THE FIRST-ORDER DERIVATIVE AND THE K STATISTIC

3.1. *Testing Hypotheses on All Parameters*

The GMM LM or score statistic of Newey and West (1987b) consists of a quadratic form of the product of the average moment vector, a covariance matrix estimator, and an estimator of the expected Jacobian. Because of the correlation between the average moment vector and the Jacobian estimator, which results from the $V_{\theta f}(\theta)$ covariance matrix in Assumption 1, the limiting behavior of the GMM LM statistic depends on nuisance parameters when, for example, the expected Jacobian equals zero. We therefore prefer to use a Jacobian estimator for the GMM LM statistic that is asymptotically uncorrelated with the average moment vector. The construction of a Jacobian estimator with this appealing property involves the CUE and the first-order derivative of the objective function.

The CUE is obtained by minimizing the objective function $Q(\theta)$ (2) with respect to θ,

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^m} Q(\theta), \tag{10}$$

which leads to the first-order condition (FOC)

$$\text{(11)} \qquad \text{FOC}: \quad \left.\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\hat{\theta}} = 0.$$

Instead of analyzing the first-order derivative of $Q(\theta)$ at the CUE, $\hat{\theta}$, we analyze ($\frac{1}{2}$ times) it at the true value of θ, θ_0. This first-order derivative reads

$$\text{(12)} \qquad \left.\frac{1}{2}\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\theta_0}$$
$$= f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} q_T(\theta_0, Y)$$
$$- \frac{1}{2}\big(f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} \otimes f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\big)\left.\frac{\partial \operatorname{vec}(\hat{V}_{ff}(\theta))}{\partial \theta'}\right|_{\theta_0}.$$

The traditional construction of the limiting distribution of GMM estimators and test statistics only involves the first element of the first-order derivative (12). The second element is left aside because it, under the customary identification assumption of a fixed full rank value of the expected Jacobian

$$\text{(13)} \qquad J_\theta(\theta_0) = \lim_{T\to\infty} E\left\{\frac{1}{T} q_T(\theta_0, Y)\right\},$$

vanishes asymptotically when we scale by the appropriate factor of T; see, e.g., Hansen (1982), Newey and McFadden (1994), and Stock and Wright (2000). When $J_\theta(\theta_0)$ does not have a fixed full rank value, this second element does not vanish asymptotically and influences the limiting distributions. To obtain a statistic whose limiting distribution is insensitive to the value of $J_\theta(\theta_0)$, we therefore use all elements of the first-order derivative (12).

To construct the derivative of $\hat{V}_{ff}(\theta_0)$, we make an assumption about the convergence of the covariance matrix estimator $\hat{V}(\theta_0)$ of the full covariance matrix $V(\theta_0)$ (8). The specification of the covariance matrix estimators is discussed in Section 4.

ASSUMPTION 2: *The convergence behavior of the covariance matrix estimator $\hat{V}(\theta_0)$ toward $V(\theta_0)$ is such that*

$$\text{(14)} \qquad \hat{V}(\theta_0) \underset{p}{\to} V(\theta_0) \quad \textit{and} \quad \frac{\partial \operatorname{vec}(\hat{V}_{ff}(\theta_0))}{\partial \theta'} \underset{p}{\to} \frac{\partial \operatorname{vec}(V_{ff}(\theta_0))}{\partial \theta'}.$$

We use Assumption 2 to construct the first-order derivative (12) from the expression for $V_{ff}(\theta_0)$ (3) (see the Appendix for the construction of the derivative),

$$(15)\qquad \frac{1}{2}\frac{\partial Q(\theta)}{\partial \theta'}\bigg|_{\theta_0} = f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)$$

with

$$(16)\qquad \hat{D}_T(\theta_0, Y) = \big[q_{1,T}(\theta_0, Y) - \hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y) \;\cdots\; q_{m,T}(\theta_0, Y) - \hat{V}_{\theta f,m}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y)\big],$$

and $\hat{V}_{\theta f}(\theta_0) = (\hat{V}_{\theta f,1}(\theta_0)' \;\cdots\; \hat{V}_{\theta f,m}(\theta_0)')'$, $q_T(\theta_0, Y) = (q_{1,T}(\theta_0, Y) \;\cdots\; q_{m,T}(\theta_0, Y))$ with $\hat{V}_{\theta f,i}(\theta_0) : k_f \times k_f$ dimensional matrices and $q_{i,T}(\theta_0, Y) : k_f \times 1$ dimensional vectors for $i = 1, \ldots, m$.

LEMMA 1: *When Assumptions* 1 *and* 2 *hold*,

$$(17)\qquad \sqrt{T}\,\mathrm{vec}\left(\frac{1}{T}\hat{D}_T(\theta_0, Y) - J_\theta(\theta_0)\right) \underset{d}{\to} \psi_{\theta.f},$$
$$\frac{1}{\sqrt{T}} f_T(\theta_0, Y) \underset{d}{\to} \psi_f,$$

where $\psi_{\theta.f} = \psi_\theta - V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f$ *and*

$$(18)\qquad \psi_{\theta.f} \sim N(0, V_{\theta\theta.f}(\theta_0)),$$
$$\psi_f \sim N(0, V_{ff}(\theta_0)),$$

with $V_{\theta\theta.f}(\theta_0) = V_{\theta\theta}(\theta_0) - V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}V_{f\theta}(\theta_0)$, *and* $\psi_{\theta.f}$ *is independent of* ψ_f.

PROOF: See the Appendix.

Lemma 1 shows that $\hat{D}_T(\theta_0, Y)$ is an estimator of the Jacobian $J_\theta(\theta_0)$ whose limiting behavior is independent of the limiting behavior of $f_T(\theta_0, Y)$. This independence holds regardless of the value of the expected Jacobian $J_\theta(\theta_0)$, so the limiting behaviors of $\hat{D}_T(\theta_0, Y)$ and $f_T(\theta_0, Y)$ are independent of one another when (i) $J_\theta(\theta_0)$ has a fixed full rank value, (ii) $J_\theta(\theta_0)$ has a weak value, which we indicate by $J_\theta(\theta_0) = J_{\theta,T}$, $J_{\theta,T} = \frac{1}{\sqrt{T}}C$, and $C : k_f \times m$, (iii) $J_\theta(\theta_0)$ is equal to zero.

THEOREM 1: *Under Assumptions* 1 *and* 2 *and* $H_0 : \theta = \theta_0$, *the limiting behavior of the normalized first-order derivative of* $Q(\theta)$ *reads*

$$\frac{1}{2\sqrt{T}}\left(\left.\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\theta_0}\right)\left(\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)\right)^{-1/2} \underset{d}{\to} \psi'_{Qd\theta}, \tag{19}$$

where $\psi_{Qd\theta} : m \times 1$,

$$\psi_{Qd\theta} \sim N(0, I_m), \tag{20}$$

and $\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)$ *is always nonsingular.*

PROOF: See the Appendix.

DEFINITION 1: The K statistic for testing $H_0 : \theta = \theta_0$ reads

$$\begin{aligned} \mathrm{K}(\theta_0) &= \frac{1}{4T}\left(\left.\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\theta_0}\right)[\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)]^{-1} \\ &\quad \times \left(\left.\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\theta_0}\right)' \\ &= \frac{1}{T} f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1/2} P_{\hat{V}_{ff}(\theta_0)^{-1/2}\hat{D}_T(\theta_0, Y)} \hat{V}_{ff}(\theta_0)^{-1/2} f_T(\theta_0, Y) \end{aligned} \tag{21}$$

and has, under H_0 and Assumptions 1 and 2, a $\chi^2(m)$ limiting distribution.

Definition 1 shows that the only difference between the K statistic and the GMM LM statistic of Newey and West (1987b) concerns the Jacobian estimator. The GMM LM statistic uses the Jacobian estimator $q_T(\theta_0, Y)$, while the K statistic uses $\hat{D}_T(\theta_0, Y)$. The limiting behavior of $q_T(\theta_0, Y)$ is not independent of the limiting behavior of $f_T(\theta_0, Y)$ under weak or zero values of $J_\theta(\theta_0)$ (see Stock and Wright (2000)), so the limiting distribution of the GMM LM statistic then depends on nuisance parameters.

The CUE sets the first-order derivative to zero so the K statistic is equal to zero at the CUE. It thus leads to inference that is centered around the CUE. The CUE involves the Jacobian estimator $\hat{D}_T(\hat{\theta}, Y)$ that is asymptotically independent of the moment equation. It has therefore a smaller asymptotic bias than two-step GMM estimators; see Newey and Smith (2004). Asymptotic biases of semiparametric estimators are reduced in a similar manner by using a Jacobian estimator that is asymptotically independent from the moment equations; see Brown and Newey (1998). The asymptotic independence of the Jacobian estimator $\hat{D}_T(\hat{\theta}, Y)$ and the moment equations can also be explained since $\hat{D}_T(\hat{\theta}, Y)$ acts as a minus one jackknife estimator; see Donald and Newey (2000).

3.2. *Testing Hypotheses on Subsets of the Parameters*

The K statistic from Definition 1 conducts a joint test on all elements of θ. When θ contains several elements, for example, $\theta = (\alpha' \ \beta')'$, with $\alpha : m_\alpha \times 1$ and $\beta : m_\beta \times 1$, $m = m_\alpha + m_\beta$, we can adapt the K statistic to test a hypothesis specified on a subset of the parameters, $H_0^* : \beta = \beta_0$. To construct the limiting distribution for this statistic, we make an additional assumption; see also Kleibergen (2004a).

ASSUMPTION 3: *The $k_f \times m_\alpha$ dimensional Jacobian matrix*

$$J_\alpha(\alpha, \beta) = \lim_{T\to\infty} E\left\{ \frac{1}{T} \sum_{t=1}^{T} \left[\left(\frac{\partial f_t(\alpha, \beta)}{\partial \alpha'} \right) \Big|_{\alpha,\beta} \right] \right\} \tag{22}$$

is a continuous function of (α, β) and has full rank m_α in (α_0, β_0).

Under Assumptions 1–3 and $H_0^* : \beta = \beta_0$, the estimator that solves the FOC with respect to α given β_0, $\hat{\alpha}(\beta_0)$, is a consistent estimator of α_0 and $\sqrt{T}(\hat{\alpha}(\beta_0) - \alpha_0)$ has a normal limiting distribution; see, e.g., Stock and Wright (2000). We can therefore analyze the limiting behavior of the derivative $Q(\theta)$ with respect to θ in $\hat{\theta}_0 = (\hat{\alpha}(\beta_0)' \ \beta_0')'$ that is equal to

$$\begin{aligned} \frac{1}{2} \frac{\partial Q(\theta_0)}{\partial \theta'} \Big|_{\hat{\theta}_0} &= f_T(\hat{\theta}_0, Y)' \hat{V}_{ff}(\hat{\theta}_0)^{-1} \hat{D}_T(\hat{\theta}_0, Y) \\ &= \big(0 \ \ f_T(\hat{\theta}_0, Y)' \hat{V}_{ff}(\hat{\theta}_0)^{-1/2} M_{\hat{V}_{ff}(\hat{\theta}_0)^{-1/2} \hat{D}_{\alpha,T}(\hat{\theta}_0, Y)} \\ &\qquad \times \hat{V}_{ff}(\hat{\theta}_0)^{-1/2} \hat{D}_{\beta,T}(\hat{\theta}_0, Y)\big), \end{aligned} \tag{23}$$

which results since $f_T(\hat{\theta}_0, Y)' \hat{V}_{ff}(\hat{\theta}_0)^{-1} \hat{D}_{\alpha,T}(\hat{\theta}_0, Y) = 0$ and where $\hat{D}_T(\hat{\theta}_0, Y) = (\hat{D}_{\alpha,T}(\hat{\theta}_0, Y) \ \hat{D}_{\beta,T}(\hat{\theta}_0, Y))$ with $\hat{D}_{\alpha,T}(\hat{\theta}_0, Y) : k_f \times m_\alpha$ and $\hat{D}_{\beta,T}(\hat{\theta}_0, Y) : k_f \times m_\beta$ dimensional matrices.

THEOREM 2: *Under Assumptions* 1–3 *and $H_0^* : \beta = \beta_0$, the limiting behavior of the normalized first-order derivative of $Q(\alpha, \beta)$ with respect to β in $\hat{\theta}_0 = (\hat{\alpha}(\beta_0)' \ \beta_0')'$ reads*

$$\begin{aligned} &\frac{1}{2\sqrt{T}} \left(\frac{\partial Q(\alpha, \beta)}{\partial \beta'} \Big|_{\hat{\theta}_0} \right) \\ &\quad \times \big[\hat{D}_{\beta,T}(\hat{\theta}_0, Y)' \hat{V}_{ff}(\hat{\theta}_0)^{-1/2} \\ &\qquad \times M_{\hat{V}_{ff}(\hat{\theta}_0)^{-1/2} \hat{D}_{\alpha,T}(\hat{\theta}_0, Y))} \hat{V}_{ff}(\hat{\theta}_0)^{-1/2} \hat{D}_{\beta,T}(\hat{\theta}_0, Y) \big]^{-1/2} \\ &\underset{d}{\to} \psi'_{Qd\beta}, \end{aligned} \tag{24}$$

where $\psi_{Qd\beta} : m_\beta \times 1$ *and*

$$\psi_{Qd\beta} \sim N(0, I_{m_\beta}). \tag{25}$$

PROOF: The proof results directly from Lemma 1 and Assumptions 1–3.

DEFINITION 2: The K statistic for testing $H_0^* : \beta = \beta_0$ reads

$$\begin{aligned} &\mathrm{K}(\beta_0) \\ &\quad = \frac{1}{4T}\left(\frac{\partial Q(\theta)}{\partial \theta'}\bigg|_{\hat{\theta}_0}\right)[\hat{D}_T(\hat{\theta}_0, Y)'\hat{V}_{ff}(\hat{\theta}_0)^{-1}\hat{D}_T(\hat{\theta}_0, Y)]^{-1}\left(\frac{\partial Q(\theta)}{\partial \theta'}\bigg|_{\hat{\theta}_0}\right)', \end{aligned} \tag{26}$$

where $\hat{\theta}_0 = (\hat{\alpha}(\beta_0)' \ \beta_0')'$, and has, under H_0^* and Assumptions 1–3, a $\chi^2(m_\beta)$ limiting distribution.

Compared to K statistic (21), the expression of the K statistic remains unaltered, but the limiting distribution of (26) has less degrees of freedom. It shows that we can conduct tests on those subsets of the parameters for which Assumption 3 holds for the remaining parameters. When Assumption 3 does not hold, the limiting distribution of (26) depends on nuisance parameters and we should use (21) instead.

It is not always straightforward to determine the parameters for which Assumption 3 is satisfied. Assumption 3 is always satisfied for those parameters θ_i for which $V_{\theta f,i}(\theta_0)$ equals zero. We note that one has to be careful conducting pretests for the rank of $J_\alpha(\alpha, \beta)$ because such pretests affect the size of the statistics computed consecutively.

4. COVARIANCE MATRIX ESTIMATORS

The objective function (2) and the functional expressions of K statistics (21) and (26) all depend on the covariance matrix estimator $\hat{V}_{ff}(\theta_0)$. Because of its use in the K statistic, the covariance matrix estimator $\hat{V}_{ff}(\theta_0)$ has to be a consistent estimator of $V_{ff}(\theta_0)$ and its derivative with respect to θ has to be a consistent estimator of the derivative of $V_{ff}(\theta_0)$ with respect to θ:

$$\frac{\partial \operatorname{vec}(\hat{V}_{ff}(\theta_0))}{\partial \theta'} \underset{q}{\rightarrow} \frac{\partial \operatorname{vec}(V_{ff}(\theta_0))}{\partial \theta'}. \tag{27}$$

The expression of the derivative of $V_{ff}(\theta_0)$ with respect to θ is constructed in the Appendix and consists of elements of $V_{\theta f}(\theta_0)$; see also (16). The covariance estimator $\hat{V}_{ff}(\theta_0)$ is therefore not only a consistent estimator of $V_{ff}(\theta_0)$, but also implies a consistent estimator $\hat{V}_{\theta f}(\theta_0)$ of $V_{\theta f}(\theta_0)$. The second property imposes additional restrictions on $\hat{V}_{ff}(\theta_0)$; see Donald and Newey (2001).

For example, in case of time-series data, $\{\bar{f}_t(\theta_0),\ t = 1, \ldots, T\}$ can be a martingale difference series while $\{\bar{q}_t(\theta_0),\ t = 1, \ldots, T\}$ is not. When we then use heteroscedasticity autocorrelation consistent covariance (HACC) matrix estimators

$$\hat{V}_{ff}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} W(t, j)\hat{f}_t(\theta_0)\hat{f}_j(\theta_0)', \tag{28}$$

$$\hat{V}_{\theta f}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} W(t, j)\hat{q}_t(\theta_0)\hat{f}_j(\theta_0)',$$

where $\hat{f}_t(\theta_0) = f_t(\theta_0) - \frac{1}{T}\sum_{t=1}^{T} f_t(\theta_0)$, $\hat{q}_t(\theta_0) = q_t(\theta_0) - \frac{1}{T}\sum_{t=1}^{T} q_t(\theta_0)$, and $W(t, j)$ is the kernel function of the involved HACC estimator, a White covariance matrix estimator $\hat{V}_{ff}(\theta_0)$ (see White (1980)) is a consistent estimator of $V_{ff}(\theta_0)$, but the specification of $\hat{V}_{\theta f}(\theta_0)$ that it implies is not a consistent estimator of $V_{\theta f}(\theta_0)$. Hence, we have to use higher-order lags in the specification of $\hat{V}_{ff}(\theta_0)$ for $\hat{V}_{ff}(\theta_0)$ to imply a consistent estimator $\hat{V}_{\theta f}(\theta_0)$ of $V_{\theta f}(\theta_0)$, such as, for example, in the Newey–West covariance matrix estimator; see Newey and West (1987a).

5. IMPROVING POWER TOWARD IRRELEVANT ALTERNATIVES

The K statistic equals a quadratic form of the first-order derivative of the GMM objective function. When considered as a function of realized data, it is therefore equal to zero at those values of θ where the GMM objective function attains its minimum, maximum, or has an inflection point. The CUE gives the value of θ where the objective function is minimal. The zero values of the K statistic at the maximal value of the objective function and inflection points are essentially irrelevant because the moment conditions are violated at these values of θ. The sensitivity to these values is avoided when we conduct a pretest of the moment conditions at θ_0, $H_m : E(f_t(\theta_0)) = 0$, using a statistic that is asymptotically independent of the K statistic. The J statistic,[2]

$$\mathrm{J}(\theta_0) = \frac{1}{T} f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1/2} M_{\hat{V}_{ff}(\theta_0)^{-1/2}\hat{D}_T(\theta_0, Y)} \hat{V}_{ff}(\theta_0)^{-1/2} f_T(\theta_0, Y), \tag{29}$$

is well suited for this purpose. It tests $H_m : E(f_t(\theta_0)) = 0$, and converges under H_0 and Assumptions 1 and 2 to a $\chi^2(k_f - m)$ distributed random variable that is independent of the $\chi^2(m)$ distributed random variable to which $\mathrm{K}(\theta_0)$ converges.

[2]We note that this J statistic is not the J statistic of Hansen (1982) that is proportional to the objective function evaluated at a GMM estimator for θ.

Since $\mathrm{J}(\theta_0)$ and $\mathrm{K}(\theta_0)$ add up to Stock and Wright's (2000) S statistic,

$$\mathrm{S}(\theta_0) = \mathrm{J}(\theta_0) + \mathrm{K}(\theta_0) = \frac{1}{T} Q(\theta_0), \tag{30}$$

the J statistic is equal to the S statistic when the K statistic is equal to zero. The K statistic suffers from a decline in power around values of θ where the objective function, i.e., the S statistic, is maximal or has an inflection point. For these values of θ, the J statistic is (approximately) equal to the S statistic, so it has discriminatory power. The J statistic has discriminatory power because it tests the validity of the moment equations, $H_m : E(f_t(\theta_0)) = 0$, while the K statistic tests $H_0 : \theta = \theta_0$ given that the moment equations hold; see Kleibergen (2004b). Around values of θ where the objective function is maximal or has an inflection point, the moment equations do not hold, so the results from the K statistic are spurious.

The limiting distribution of the K statistic is valid under Assumptions 1 and 2 and $H_0 : \theta = \theta_0$. Assumption 1 implies that the moment equation (1), or equivalently H_m, holds. To verify whether the moment equation holds, we conduct a pretest using J statistic (29). Since the J statistic is asymptotically independent of the K statistic, the overall size α of testing the moment equations and H_0 depends on the sizes that we use to test the moment equations using the J statistic, α_{J}, and H_0 using the K statistic, α_{K}, i.e., $\alpha = \alpha_{\mathrm{J}} + \alpha_{\mathrm{K}} - \alpha_{\mathrm{J}}\alpha_{\mathrm{K}} \approx \alpha_{\mathrm{J}} + \alpha_{\mathrm{K}}$. By choosing α_{J} and α_{K} appropriately, we emphasize tests of the moment equations or H_0. For example, when $\alpha = 0.05$, $\alpha_{\mathrm{J}} = 0.01$, and $\alpha_{\mathrm{K}} = 0.04$ implies that we focus on H_0. The pretest using the J statistic restricts the parameter region for θ to values that satisfy the moment equations or, stated differently, it puts an upper bound on the value of the objective function.

5.1. *Conditional GMM Statistic*

Moreira's (2003) conditional likelihood ratio (LR) statistic accounts for the spurious behavior of the K statistic by combining the J and K statistics in a conditional manner using $\hat{D}_T(\theta_0, Y)$ as the conditioning statistic. Moreira's conditional LR statistic is developed for the linear instrumental variables regression with one included endogenous variable, but an adaptation of it can be used in GMM as well (see Kleibergen (2004b)):

$$\begin{aligned} &\text{GMM-M}(\theta_0) \\ &\quad = \frac{1}{2}\Big[\mathrm{K}(\theta_0) + \mathrm{J}(\theta_0) - \mathrm{rk}(\theta_0) \\ &\qquad + \sqrt{\{\mathrm{K}(\theta_0) + \mathrm{J}(\theta_0) + \mathrm{rk}(\theta_0)\}^2 - 4\mathrm{J}(\theta_0)\mathrm{rk}(\theta_0)}\,\Big], \end{aligned} \tag{31}$$

with $\mathrm{rk}(\theta_0)$ a statistic that tests the hypothesis of a lower rank value of $J_\theta(\theta_0)$, $H_r : \mathrm{rank}(J_\theta(\theta_0)) = m - 1$, and is a function of $\hat{D}_T(\theta_0, Y)$ and the (generalized) inverse of $\hat{V}_{\theta\theta.f}(\theta_0) = \hat{V}_{\theta\theta}(\theta_0) - \hat{V}_{\theta f}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}\hat{V}_{f\theta}(\theta_0)$. The rank statistics from Cragg and Donald (1996, 1997), Kleibergen and Paap (2005), and Robin and Smith (2000) can be used for this purpose.[3]

THEOREM 3: *Under Assumptions* 1 *and* 2, *and* $H_0 : \theta = \theta_0$, *the conditional limiting distribution of* GMM-M(θ_0) *given* rk(θ_0), *where* rk(θ_0) *is only a function of* $\hat{D}_T(\theta_0, Y)$ *and the* (*generalized*) *inverse of* $\hat{V}_{\theta\theta.f}(\theta_0)$, *reads*

$$\text{GMM-M}(\theta_0)|\text{rk}(\theta_0) \underset{d}{\to} \frac{1}{2}\Big[\psi_{\mathrm{J}} + \psi_{\mathrm{K}} - \text{rk}(\theta_0) + \sqrt{(\psi_{\mathrm{J}} + \psi_{\mathrm{K}} + \text{rk}(\theta_0))^2 - 4\psi_{\mathrm{J}}\text{rk}(\theta_0)}\Big], \tag{32}$$

where ψ_{J} *and* ψ_{K} *are the independent* $\chi^2(k_f - m)$ *and* $\chi^2(m)$ *distributed random variables to which* J(θ_0) *and* K(θ_0) *converge*.

PROOF: The statistic rk(θ_0) is a function of $\hat{D}_T(\theta_0, Y)$ and $\hat{V}_{\theta\theta.f}(\theta_0)$, so it is asymptotically independent of K(θ_0) and J(θ_0). *Q.E.D.*

Theorem 3 shows that it is straightforward to simulate from the conditional limiting distribution of GMM-M(θ_0) given rk(θ_0): simulate K(θ_0) and J(θ_0) from their independent $\chi^2(m)$ and $\chi^2(k_f - m)$ limiting distributions and use them, alongside rk(θ_0), to construct the realizations of GMM-M(θ_0) from its conditional limiting distribution. This conditional limiting distribution is such that when rk(θ_0) equals zero, it is equal to the $\chi^2(k_f)$ distributed sum of the $\chi^2(m)$ and $\chi^2(k_f - m)$ limiting distributions of K(θ_0) and J(θ_0) while it is equal to the $\chi^2(m)$ limiting distribution of K(θ_0) for large values of rk(θ_0).

The GMM-M(θ_0) leads to inference that is centered around the CUE, $\hat{\theta}$, when GMM-M$(\hat{\theta}) = 0$. GMM-M$(\hat{\theta})$ equals zero when rk$(\hat{\theta})$ exceeds J$(\hat{\theta})$, which, since K$(\hat{\theta})$ equals zero, puts a condition on the rank statistic rk(θ_0) that is used in GMM-M(θ_0).

THEOREM 4: *The specification of the rank statistic* rk(θ), *which results from the Cragg and Donald* (1997) *objective function rank statistic*, *is larger than or equal to the objective function* $Q(\theta)$ *divided by* T *when evaluated at the CUE*.

PROOF: See Kleibergen (2004b).

[3]Since the limiting distribution of rk(θ_0) is not involved in GMM-M(θ_0), singular values of $\hat{V}_{\theta\theta.f}(\theta_0)$ are allowed for the expression of rk(θ_0) that is used in GMM-M(θ_0). When $\hat{V}_{\theta\theta.f}(\theta_0)$ is singular, we obtain its inverse by means of the nonzero eigenvalues and their accompanying eigenvectors that result from a spectral decomposition of $\hat{V}(\theta_0)$. The same eigenvectors are then used to transform $\hat{D}_T(\theta_0, Y)$ in rk(θ_0) as well.

Since $\mathrm{J}(\hat{\theta}) = \frac{1}{T} Q(\hat{\theta})$, Theorem 4 shows that GMM-M(θ_0) with rk(θ_0) resulting from the objective function rank statistic of Cragg and Donald (1997) leads to inference that is centered around the CUE. The rank statistic of Kleibergen and Paap (2005) exceeds the Cragg and Donald (1997) rank statistic, so when we use it for rk(θ_0), we obtain inference using GMM-M(θ_0) that is centered around the CUE as well.

6. CONFIDENCE SETS

To obtain a confidence set for θ, we specify sequences of n increasing values for every element of θ. We then have an m-dimensional grid that contains n^m different values of θ_0. We compute the statistic of interest (i.e., the J, K, GMM-M, or S statistic) for each of these n^m different values of θ_0. All elements in the specified grid for which the asymptotic p-value of the statistic of interest exceeds α are in the $(1-\alpha)100\%$ confidence set of θ. Using an appropriate specification of the grid, we obtain the $(1-\alpha)100\%$ asymptotic confidence set. For the combined J–K test, the $(1-\alpha)100\%$ confidence set that uses α_{K} and α_{J} is obtained as the set of values for which the asymptotic p-values of the J and K statistics exceeds α_{J} and α_{K} resp.

Alongside infinite and finite convex confidence sets, the confidence sets that result from the proposed tests can also be nonconvex; see, e.g., Dufour (1997). The confidence sets that result from the S, J, and combined J–K statistics can be empty, while the confidence sets that result from the K and GMM-M statistics are never empty since these statistics are equal to zero at the CUE.

7. STOCHASTIC DISCOUNT FACTORS

We test the risk aversion parameter in a stochastic discount factor (SDF) model. We compute power curves for some simulated data sets and construct confidence sets using observed asset returns and consumption growth.

7.1. *Stochastic Discount Factor Model*

The specification of $f_t(\theta)$ that results from a SDF model with a constant relative rate of risk aversion utility function reads (see, e.g., Hansen and Singleton (1982))

$$f_t(\delta, \beta) = \left(\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\beta}(\iota_l + R_{t+1}) - \iota_l\right] \otimes X_t\right). \tag{33}$$

The discount factor is given by δ and β is the risk aversion coefficient, so $\theta = (\delta\ \beta)'$. The $l \times 1$ vector R_t contains the returns on l different assets at time t, C_t is consumption at time t, and ι_l is a $l \times 1$ vector of ones. The $k \times 1$

vector X_t contains the instruments that consist of a constant and lagged values of consumption growth and asset returns. Because δ is a scalar, (33) is equivalent to

$$f_t(\alpha, \beta) = (\varphi_t(\alpha, \beta) \otimes X_t), \tag{34}$$

with $\varphi_t(\alpha, \beta) = (\frac{C_{t+1}}{C_t})^{-\beta}(\iota_l + R_{t+1}) - \iota_l \alpha$ and $\alpha = \delta^{-1}$. The derivative of (34) with respect to α equals $-(\iota_l \otimes X_t)$ and is spanned by the instruments so that $V_{\alpha f}(\theta) = 0$. Assumption 3 is thus satisfied for α. We can therefore partial out α and then Assumption 1 implies the central limit theorem

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \bar{f}_t(\hat{\alpha}(\beta_0), \beta_0) \\ \bar{q}_t(\hat{\alpha}(\beta_0), \beta_0) \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_f \\ \psi_\beta \end{pmatrix}, \tag{35}$$

with $\hat{\alpha}(\beta_0)$ the CUE of α given that $\beta = \beta_0$, $q_t(\alpha, \beta) = \frac{\partial}{\partial\beta} f_t(\alpha, \beta) = -(\log(\frac{C_{t+1}}{C_t})(\frac{C_{t+1}}{C_t})^{-\beta}(\iota_l + R_{t+1}) \otimes X_t)$, and $(\psi_f' \; \psi_\gamma') \sim N(0, V(\alpha, \beta))$. The central limit theorem (35) results from a statistical model for $\varphi_t(\alpha, \beta)$ and $\frac{\partial}{\partial\beta}\varphi(\alpha, \beta, Y_t)$ that explains these two series from X_t. This shows that Assumption 1 puts no stringent conditions on the asset return and consumption growth series. The specification of $\hat{D}_{\beta,T}(\hat{\alpha}(\beta), \beta, Y)$ for the stochastic discount factor model that is used in (23) and (26) reads

$$\begin{aligned} &\hat{D}_{\beta,T}(\hat{\alpha}(\beta), \beta) \\ &\quad = \sum_{t=1}^{T} \big[q_t(\hat{\alpha}(\beta), \beta) - \hat{V}_{\theta f}(\hat{\alpha}(\beta), \beta)\hat{V}_{ff}(\hat{\alpha}(\beta), \beta)^{-1} f_t(\hat{\alpha}(\beta), \beta)\big]. \end{aligned} \tag{36}$$

7.2. *Size and Power Comparison*

We generate artificial data using the Monte Carlo design of Tauchen (1986a, 1986b). We calibrate a 10^2-dimensional Markov chain to approximate a Gaussian vector autoregression (VAR) of order 1 fitted to consumption and dividend growth so that $l = 1$. This VAR(1) is also used by Kocherlakota (1990), Hansen, Heaton, and Yaron (1996), and Stock and Wright (2000) it is

$$\begin{pmatrix} c_t \\ d_t \end{pmatrix} = \begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix} + \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{d,t} \end{pmatrix}, \tag{37}$$

with c_t, the log-growth rate of U.S. per capita real annual consumption, and d_t, the log-growth rate of real annual dividends on the Standard & Poors 500 (SP500). The disturbances $(\varepsilon_{c,t} \; \varepsilon_{d,t})'$ are independently normally distributed with mean zero and $\text{var}(\varepsilon_{c,t}) = 0.014$, $\text{var}(\varepsilon_{c,t}) = 0.0012$, $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.43$.

The VAR(1) (37) is used to generate asset return and consumption growth series that satisfy a SDF model. Figure 1 shows the power curves of testing $H_0 : \beta = 1.3$ using Wald statistics based on the CUE and two-step estimator, the standard GMM LM statistic, and the S, K, J, and GMM-M statistics. We use instruments that consist of a constant and up to three period lagged values of consumption growth and asset returns so that $k = 7$. The number of observations equals 250. Each panel in Figure 1 shows the power curves for a different data generating process (DGP). The DGP for panel 1 is a SDF model calibrated to approximate the VAR from (37). The same DGP is used in panel 2 but with $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.95$. The DGP for panels 3 and 4 is a SDF calibrated to the VAR from (37) with the VAR(1) parameter matrix multiplied by 2. In panel 4, $\text{cor}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.95$.

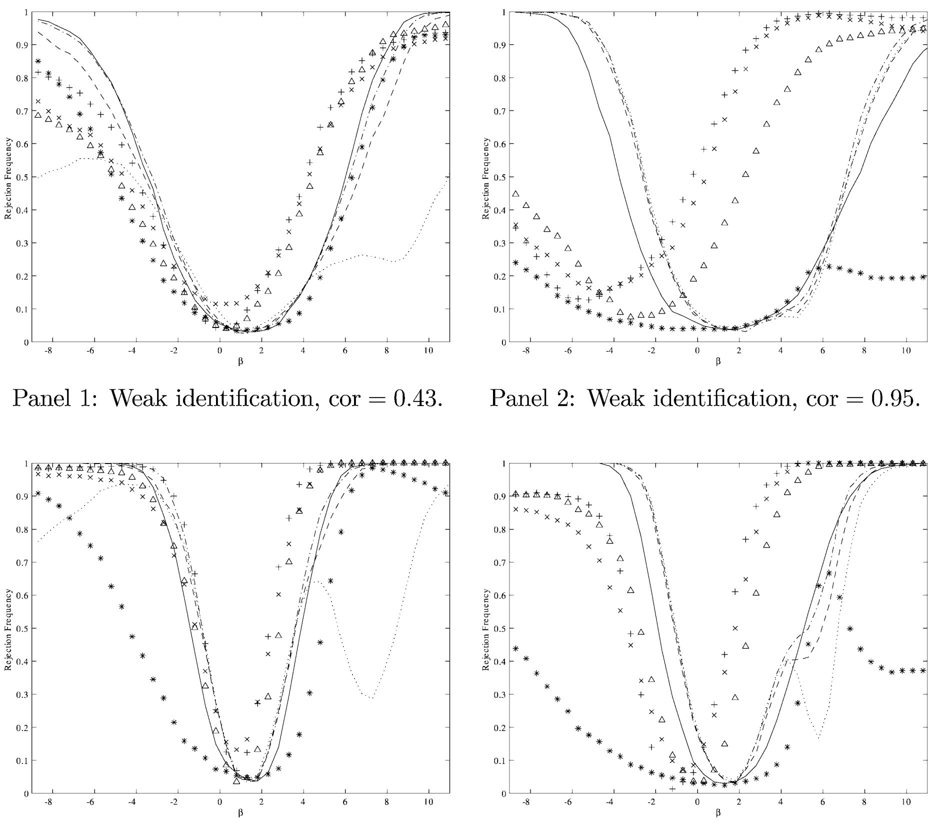

FIGURE 1.—Power curves of statistics that test with 95% (asymptotic) significance $H_0 : \beta = 1.3$. Key: S, solid line; K, dotted, J, stars; J–K with $\alpha_K = 0.04$, $\alpha_J = 0.01$, dashed; GMM-M, dashed-dotted; Wald-CUE, crosses; Wald-two-step, plusses; LM, triangles.

We use Newey–West covariance matrix estimators for $V_{ff}(\alpha, \beta)$, $V_{\theta f}(\alpha, \beta)$, and $V_{\theta\theta}(\alpha, \beta)$ with one lag so that $W(t, j) = 1 - \frac{|t-j|}{2}$ in (28). Since $m_\beta = l = 1$, the specification of $\mathrm{rk}(\theta)$ that we use for the GMM-M statistic (31) and its conditional limiting distribution reads

$$\mathrm{rk}(\hat{\alpha}(\beta), \beta) = \frac{1}{T}\hat{D}_T(\hat{\alpha}(\beta), \beta, Y)'\hat{V}_{\theta\theta.f}(\hat{\alpha}(\beta), \beta)^{-1}\hat{D}_T(\hat{\alpha}(\beta), \beta, Y), \tag{38}$$

where $\hat{V}_{\theta\theta.f}(\hat{\alpha}(\beta), \beta) = \hat{V}_{\theta\theta}(\hat{\alpha}(\beta), \beta) - \hat{V}_{\theta f}(\hat{\alpha}(\beta), \beta)\hat{V}_{ff}(\hat{\alpha}(\beta), \beta)^{-1}\hat{V}_{\theta f}(\hat{\alpha}(\beta), \beta)'$. Theorem 4 shows that the specification of $\mathrm{rk}(\hat{\alpha}(\beta), \beta)$ (38) is such that the GMM-M statistic is equal to zero at the CUE.

Figure 1 shows that the GMM LM and the two-step and CUE Wald statistics are size distorted especially when the correlation parameter is large. The size distortion of the other statistics is always rather small. Figure 1 shows that the power of the J statistic is typically quite low except for those values where the K statistic has its spurious decline of power. The combined J–K test procedure has therefore good power properties. The power of the GMM-M statistic is comparable to the power of the J–K test. In panel 1, which is the DGP where the identification of β is the weakest, the power of the S statistic is comparable to that of the J–K test and the GMM-M statistic. The power of the S statistic is dominated by the J–K test and the GMM-M statistic in most of the other panels. This results from the larger degrees of freedom parameter of the limiting distribution of the S statistic.

7.3. *Stochastic Discount Factor Model for Consumption Growth and Return on SP500*

We construct asymptotic confidence sets of the risk aversion parameter β in a SDF model for yearly observations of the growth of U.S. consumption and an asset return series that consists of the returns on the Cowles Commision index followed by the annual average price of the Standard & Poors monthly composite index. The observations cover the period 1871 to 1993. The data are an extension of the data from Campbell and Shiller (1987) and are also used by Stock and Wright (2000). The instruments consist of a constant and up to three period lagged observations of the two series. The covariance matrices are estimated by means of Newey–West covariance matrix estimators with one lag. We construct asymptotic confidence sets for the risk aversion parameter β using two-step-Wald, Wald-CUE, LM, S, J, GMM-M, and K statistics. The asymptotic critical values result from a $\chi^2(1)$ distribution for the two-step-Wald, Wald-CUE, LM, and K statistics while the asymptotic critical values for the J and S statistics result from a $\chi^2(5)$ and $\chi^2(6)$ distribution resp. For the GMM-M statistic we use its conditional asymptotic distribution given $\mathrm{rk}(\hat{\alpha}(\beta), \beta)$ (38).

Figure 2 shows one minus the (asymptotic) p-value for the different statistics that test the hypothesis $H_0 : \beta = \beta_0$ for a sequence of values of β_0. Figure 2

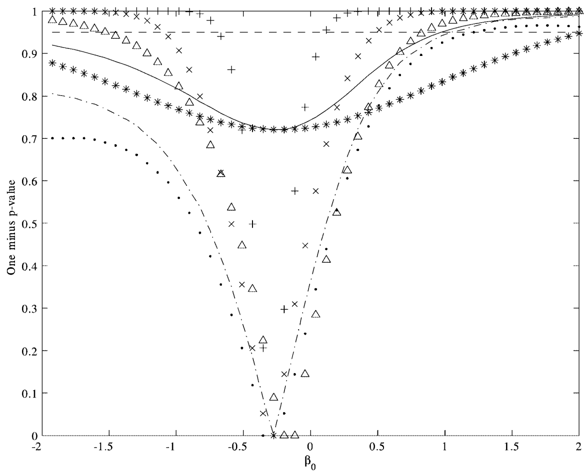

FIGURE 2.—One minus the p-value for statistics that test $H_0 : \beta = \beta_0$ for different values of β_0. Key: S, solid line; K, dotted; J, stars; GMM-M, dashed-dotted; Wald-CUE, crosses; Wald-two-step, plusses; LM, triangles.

contains a line at 0.95 that enables us to construct the 95% asymptotic confidence set in a straightforward manner by using the intersection of the curves of the different statistics with the 95% line.

The one minus p-value plots show that the GMM-M and K statistics are equal to zero at the CUE and that the S and J statistics attain their minimum at the CUE. The one minus p-value plots of the K and GMM-M statistics are rather similar and asymmetric. They differ considerably from the one minus p-value plots of the two-step-Wald, Wald-CUE, and LM statistics, which lead to a considerably smaller 95% confidence set. The one minus p-value plot of the J statistic indicates that the support for the moment equations is limited. The 95% confidence set for β that results from the S statistic is therefore smaller than the 95% confidence sets that result from the K and GMM-M statistics. The 95% confidence set that results from the combined J–K test with $\alpha_{\mathrm{K}} = 0.04$ and $\alpha_{\mathrm{J}} = 0.01$ is also much smaller than the one that results from the K and GMM-M statistics. It is similar to the 95% confidence set that results from the S statistic. When we include the risk-free interest rate in the SDF model, the J and S statistics are always significant at the 99% level, which is in line with Stock and Wright (2000).

8. CONCLUSIONS

The limit behavior of the K statistic, which is a GMM LM statistic that uses a Jacobian estimator that is independent of the moment equations, is shown to be robust to weak instruments. Because it is a quadratic form of the derivative of the objective function, it suffers from a spurious power decline around the maximum and inflection points of the objective function since the moment conditions are then not satisfied. This spurious power decline is overcome by combining the K statistic with a J statistic that tests the validity of the moment equations. A GMM extension of Moreira's conditional likelihood ratio statistic is obtained by conditioning on a statistic that tests the rank of the Jacobian.

Dept. of Economics, P.O. Box B, Brown University, Providence, RI 02912, U.S.A.

and

Dept. of Quantitative Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; Frank_Kleibergen@brown.edu; http://www.econ.brown.edu/fac/Frank_Kleibergen.

Manuscript received July, 2004; final revision received February, 2005.

APPENDIX

DERIVATION OF EQUATION (15): Assumption 2 states that $(\partial \operatorname{vec}(\hat{V}_{ff}(\theta_0)))/\partial\theta' \underset{p}{\rightarrow} (\partial \operatorname{vec}(V_{ff}(\theta_0)))/\partial\theta'$. We therefore construct $(\partial \operatorname{vec}(V_{ff}(\theta_0)))/\partial\theta'$ to show that $(\partial \operatorname{vec}(\hat{V}_{ff}(\theta_0)))/\partial\theta'$ consists of elements of $\hat{V}(\theta_0)$. The expression of $\operatorname{vec}(V_{ff}(\theta_0))$ reads

$$\operatorname{vec}(V_{ff}(\theta_0)) = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} \bar{f}_t(\theta_0) \otimes \bar{f}_j(\theta_0)\right],$$

whose derivative with respect to θ reads

$$\begin{aligned}
&\left.\frac{\partial \operatorname{vec}(V_{ff}(\theta))}{\partial\theta'}\right|_{\theta_0} \\
&\quad = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\frac{\partial \bar{f}_t(\theta_0)}{\partial\theta'} \otimes \bar{f}_j(\theta_0)\right) + \left(\bar{f}_t(\theta_0) \otimes \frac{\partial \bar{f}_j(\theta_0)}{\partial\theta'}\right)\right] \\
&\quad = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\left(\bar{q}_{1,j}(\theta_0) \ \cdots \ \bar{q}_{m,j}(\theta_0)\right) \otimes \bar{f}_t(\theta_0)\right)\right. \\
&\qquad\qquad \left. + \left(\bar{f}_t(\theta_0) \otimes \left(\bar{q}_{1,j}(\theta_0) \ \cdots \ \bar{q}_{m,j}(\theta_0)\right)\right)\right] \\
&\quad = \left(\operatorname{vec}(V_{\theta f,1}(\theta_0)') \ \cdots \ \operatorname{vec}(V_{\theta f,m}(\theta_0)')\right) \\
&\qquad + \left(\operatorname{vec}(V_{\theta f,1}(\theta_0)) \ \cdots \ \operatorname{vec}(V_{\theta f,m}(\theta_0))\right).
\end{aligned}$$

Hence,

$$\left.\frac{\partial \operatorname{vec}(\hat{V}_{ff}(\theta))}{\partial \theta'}\right|_{\theta_0} = \big(\operatorname{vec}(\hat{V}_{\theta f,1}(\theta_0)') \ \cdots \ \operatorname{vec}(\hat{V}_{\theta f,m}(\theta_0)')\big) + \big(\operatorname{vec}(\hat{V}_{\theta f,1}(\theta_0)) \ \cdots \ \operatorname{vec}(\hat{V}_{\theta f,m}(\theta_0))\big),$$

where the $\hat{V}_{\theta f,i}(\theta_0)$'s are defined below (16). For the derivative of the objective function, this implies

$$\begin{aligned}
\left.\frac{1}{2}\frac{\partial Q(\theta)}{\partial \theta'}\right|_{\theta_0} &= f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}q_T(\theta_0, Y) \\
&\quad - \frac{1}{2}\big(f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} \otimes f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\big) \\
&\qquad \times \big[\big(\operatorname{vec}(\hat{V}_{\theta f,1}(\theta_0)') \ \cdots \ \operatorname{vec}(\hat{V}_{\theta f,m}(\theta_0)')\big) \\
&\qquad\quad + \big(\operatorname{vec}(\hat{V}_{\theta f,1}(\theta_0)) \ \cdots \ \operatorname{vec}(\hat{V}_{\theta f,m}(\theta_0))\big)\big] \\
&= f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} \\
&\quad \times \big[q_T(\theta_0, Y) \\
&\qquad - \big(\operatorname{vec}(\hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y)) \ \cdots \\
&\qquad\quad \operatorname{vec}\big(\hat{V}_{\theta f,m}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y)\big)\big)\big] \\
&= f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} \\
&\quad \times \big(q_{1,T}(\theta_0, Y) - \hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y) \ \cdots \\
&\qquad q_{m,T}(\theta_0, Y) - \hat{V}_{\theta f,m}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y)\big),
\end{aligned}$$

with $q_{i,T}(\theta_0, Y) : k_f \times 1$, $i = 1, \ldots, m$, and $q_T(\theta_0, Y) = (q_{1,T}(\theta_0, Y) \ \cdots \ q_{m,T}(\theta_0, Y))$.

PROOF OF LEMMA 1: When we premultiply the central limit theorem from Assumption 1 by

$$\begin{aligned}
\hat{R}(\theta_0) &= \begin{pmatrix} I_{k_f} & 0 \\ -\hat{V}_{\theta f}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} & I_{k_\theta} \end{pmatrix} \\
&\underset{p}{\to} \begin{pmatrix} I_{k_f} & 0 \\ -V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1} & I_{k_\theta} \end{pmatrix} = R(\theta_0),
\end{aligned}$$

which results from Assumption 2, Lemma 1 follows straightforwardly. *Q.E.D.*

PROOF OF THEOREM 1: It results from Lemma 1 that

$$\frac{1}{T^{\nu}}\hat{D}_T(\theta_0, Y) \underset{d}{\to} D,$$

where ν depends on $J_\theta(\theta_0)$, $\nu = 1$ when $J_\theta(\theta_0)$ has a fixed full rank value, and $\nu = 0$ for weak or zero values of $J_\theta(\theta_0)$, and D is a random variable independent of ψ_f. We then obtain

$$\frac{1}{\sqrt{T}}f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\left[\frac{1}{T^{\nu}}\hat{D}_T(\theta_0, Y)\right] \underset{d}{\to} \psi_f' V_{ff}(\theta_0)^{-1}D.$$

The conditional distribution of $\psi_f' V_{ff}(\theta_0)^{-1}D$ given D reads

$$\psi_f' V_{ff}(\theta_0)^{-1}D|D \sim N\big(0, D'V_{ff}(\theta_0)^{-1}D\big).$$

Since D is independent of ψ_f, we obtain an unconditional result by normalizing the expression by $(D'V_{ff}(\theta_0)^{-1}D)^{-1/2}$,

$$\begin{aligned}&\frac{1}{\sqrt{T}}f_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)[\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)]^{-1/2}\\ &\quad \underset{d}{\to} \psi_f' V_{ff}(\theta_0)^{-1}D(D'V_{ff}(\theta_0)^{-1}D)^{-1/2} = \psi_{Qd\theta}',\end{aligned}$$

with $\psi_{Qd\theta}' \sim N(0, I_m)$. The matrix $\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)$ is always nonsingular since for full rank values of $J_\theta(\theta_0)$, $(1/T^2)\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} \times \hat{D}_T(\theta_0, Y) \underset{p}{\to} J_\theta(\theta_0)'V_{ff}(\theta_0)^{-1}J_\theta(\theta_0)$, while for zero or weak values of $J_\theta(\theta_0)$, $\frac{1}{T}\hat{D}_T(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0, Y)$ converges in distribution to a nonsingular random matrix. The same reasoning extends to lower rank values of $J_\theta(\theta_0)$, which result in a combination of the previous two cases. *Q.E.D.*

REFERENCES

BROWN, B. W., AND W. K. NEWEY (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453–464.

CAMPBELL, J. Y., AND R. J. SHILLER (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

CRAGG, J. C., AND S. G. DONALD (1996): "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association*, 91, 1301–1309.

——— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250.

DONALD, S. G., AND W. K. NEWEY (2000): "A Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters*, 67, 239–243.

——— (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1192.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

GORDIN, M. I. (1969): "The Central Limit Theorem for Stationary Processes," *Soviet Mathematics Doklady*, 10, 1174–1176.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., AND K. SINGLETON (1982): "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

——— (2004a): "Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model," *Review of Economics and Statistics*, 86, 418–423.

——— (2004b): "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," Working Paper, Department of Economics, Brown University.

KLEIBERGEN, F., AND R. PAAP (2005): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, forthcoming.

KOCHERLAKOTA, N. R. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: Elsevier Science, Chap. 36, 2113–2148.

NEWEY, W. K., AND K. D. WEST (1987a): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1987b): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–787.

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

ROBIN, J.-M., AND R. J. SMITH (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

TAUCHEN, G. (1986a): "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business & Economic Statistics*, 4, 397–425.

——— (1986b): "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions," *Economic Letters*, 20, 177–181.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48, 817–838.

——— (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–161.