Econometrica, Vol. 65, No. 5 (September, 1997), 1129–1151

ASYMPTOTIC THEORY OF INTEGRATED CONDITIONAL MOMENT TESTS

BY HERMAN J. BIERENS AND WERNER PLOBERGER[1]

In this paper we derive the asymptotic distribution of the test statistic of a generalized version of the integrated conditional moment (ICM) test of Bierens (1982, 1984), under a class of $\sqrt{n}$ -local alternatives, where n is the sample size. The generalized version involved includes neural network tests as a special case, and allows for testing misspecification of dynamic models.

It appears that the ICM test has nontrivial local power. Moreover, for a class of "large" local alternatives the consistent ICM test is more powerful than the parametric t test in a neighborhood of the parametric alternative involved. Furthermore, under the assumption of normal errors the ICM test is asymptotically admissible, in the sense that there does not exist a test that is uniformly more powerful.

The asymptotic size of the test is case-dependent: the critical values of the test depend on the data-generating process. In this paper we derive case-independent upperbounds of the critical values.

KEYWORDS: Nonlinear regression models, time-series models, conditional moment test, test of functional form, local power, admissibility, neural networks.

1. INTRODUCTION

CONDITIONAL MOMENT (CM) TESTS have been proposed by Newey (1985) and Tauchen (1985) in the context of maximum likelihood models, but as these authors show, most misspecification tests of functional form are special forms of CM tests. A typical CM test takes the form of a quadratic form of finitely many weighted means of the residuals, where the weights are functions of the regressors. These CM tests are in general not consistent. In order to achieve consistency, Bierens' (1982, 1990) consistent conditional moment tests employ a class of weight functions indexed by a continuous nuisance parameter, so that actually uncountable many weight functions are employed. In order to obtain a single test statistic, Bierens (1982) proposes to integrate these nuisance parameters out. Therefore we shall call the test of Bierens (1982) the Integrated Conditional Moment (ICM) test.

In Section 2 we review the ICM test and discuss the choice of the weight functions. In Section 3 we derive the asymptotic distribution of the ICM test

[1] The thoughtful comments of three anonymous referees, leading to significant improvements of this paper, are gratefully acknowledged. Previous versions of this paper have been presented by the first author at ESEM 1993, Cornell University, Texas A&M University, and the University of Pennsylvania, and by the second author at ESEM 1994. The working paper version of Section 4, Ploberger and Bierens (1996), was presented by the second author at ESEM 1996. Most of the research in this paper was done while the first author was affiliated with Southern Methodist University and the second author with the Technische Universität Wien. The financial support from the "Ausseninstitut der Technische Universität Wien" is gratefully acknowledged.

under a general class of $\sqrt{n}$-local alternatives, where we allow the data-generating process to be dependent. In Section 4 we focus on the class of "large" local alternatives, and compare the asymptotic power of the consistent ICM test against these alternatives with the asymptotic power of the parametric t test. It appears that for sufficiently large local alternatives the consistent ICM test is in general more powerful than the t test. The rate of convergence to 1 of the asymptotic power functions of the consistent ICM test and the t test is the same only if the t test is conducted on the basis of the local alternative itself, thus assuming that the alternative is completely parametric and all the variables involved are observable. In Section 5 we prove the admissibility of the ICM test under the assumption of normal errors, i.e., we show that there does not exist a uniformly more powerful test. Finally, in Section 6 we derive case-independent upperbounds of the asymptotic critical values of the ICM test.

Next to the conditional moment testing approaches of Bierens (1982, 1984, 1987, 1990), Bierens and Hartog (1988), De Jong (1996), De Jong and Bierens (1994), White (1989), and Stinchcombe and White (1993), there is also a competing line of recent literature on conditional moment tests based on comparison of parametric and (semi-)nonparametric models. See, e.g., Wooldridge (1992), Yatchew (1992), Gozalo (1993), Hardle and Mammen (1993), Horowitz and Hardle (1994), Fan and Li (1996), and Hong and White (1996), for published papers in this area. Although not all of these authors derive local power results, the ones who do, find local alternatives that shrink to the null at a slower rate than $1/\sqrt{n}$. Only Hardle and Mammen (1993) manage to achieve $\sqrt{n}$-local power, but only in one direction. In contrast, we will show in this paper that our ICM test has nontrivial $\sqrt{n}$-local power in all directions.

The proofs of theorems and lemmas are given in the Appendix, except in cases where these proofs are also helpful in understanding the main argument. Also the assumptions (A, B, and C) are stated in the Appendix. Convergence results and conditions indicated by "$\to$" that involve random variables refer to convergence in probability, unless otherwise stated. The indicator function is denoted by $I(\cdot)$, and indexed expectations signs, e.g. E_g, indicate that the expectation is taken under a certain hypothesis "g."

2. THE INTEGRATED CONDITIONAL MOMENT TEST

2.1. *Introduction*

Consider a stationary vector time series process $(y_t, x_t) \in \mathbb{R} \times \mathbb{R}^k$, which is observable for $t = 1, \ldots, n$. In parametric time series regression analysis (including ARMA and ARMAX models) we usually specify the conditional expectation of y_t relative to the σ-algebra $\mathscr{F}_{t-1}$ generated by the entire past of the process (y_t, x_t) as a known function $f_t(\theta)$ of lagged y_t's and x_t's and a parameter vector θ:

$$(1) \qquad H_0: \exists \theta_0 \in \Theta \subset \mathbb{R}^m: y_t = f_t(\theta_0) + u_t, \quad \text{where} \quad u_t = y_t - E[y_t | \mathscr{F}_{t-1}]$$

and Θ is the parameter space. In the case of independent data with x_t the vector of dependent variables one should interpret $\mathscr{F}_{t-1}$ as the σ-algebra generated by x_t and $f_t(\theta)$ as $f(x_t, \theta)$ for some given function f. The consistent tests of Bierens (1982, 1990) in the i.i.d. case and Bierens (1984) and De Jong (1996) in the time series case test the null hypothesis (1) against the general fixed alternative:

$$(2) \qquad H_1: \sup_{\theta \in \Theta} P[E(y_t | \mathscr{F}_{t-1}) = f_t(\theta)] < 1.$$

Note that the stationarity assumption implies that $E[y_t | \mathscr{F}_{t-1}]$ is stationary, and that therefore $f_t(\theta)$ should be specified stationary. If so, either (1) or (2) holds for all t.

Now consider the random function $\hat{z}(\xi) = (1/\sqrt{n})\Sigma_{t=1}^{n}[y_t - f_t(\hat{\theta})]w_t(\xi)$, $\xi \in \Xi$, where $\hat{\theta}$ is the nonlinear least squares estimator and $\{w_t(\xi)\}$ is an infinite set of weights indexed by $\xi \in \Xi$. As is shown in Bierens (1990) for the i.i.d. case, under the null hypothesis this random function converges weakly to a Gaussian random function $z(\xi)$, while under the alternative, $\hat{z}(\xi)/\sqrt{n}$ converges to a nonstochastic nonzero limit function, for weight functions $w_t(\xi) = \exp(\xi^T\Phi(x_t))$, with Φ bounded one-to-one mapping. De Jong (1996) proves a similar result for time series models for the case where Ξ grows in dimension to infinity with the sample size. However, in this paper we focus on the asymptotic theory of ICM tests under local alternatives, where the dimension of the set Ξ remains fixed.

The test statistic of the ICM test takes the form

$$(3) \qquad \hat{T} = \int |\hat{z}(\xi)|^2 \, d\mu(\xi)$$

where $\mu(\xi)$ is a probability measure on Ξ that is absolutely continuous with respect to Lebesgue measure on Ξ. This is (in essence) the form of the integrated consistent conditional moment test proposed by Bierens (1982).

2.2. *The Weight Function*

Stinchcombe and White (1993) have shown that there exists a wide class of weight functions $w_t(\xi)$, including the exponential weight functions employed by Bierens (1982, 1984, 1990), that generate consistent CM tests. For example the logistic weight function will also work, which then gives rise to White's (1989) neural network[2] version of the randomized CM tests of Bierens (1987; 1988; 1994b, Ch. 5). See also Lee, White, and Granger (1993). For the purpose of the ICM test, however, the following straightforward extension of Theorem 1 of Bierens (1982, 1990) is sufficiently general:

THEOREM 1: *Let u be a random variable satisfying $E|u| < \infty$, and let x be a bounded k-variate random vector such that $P[E(u|x) = 0] < 1$. If $w(u)$ is a*

[2] Corollary 1 of Bierens (1990) with $\exp(u)$ replaced by $1/(1 + \exp(c - u))$, with $c \neq 0$, provides a proof of why neural network methods work. See Bierens (1994a).

complex or real valued function that is infinitely many times continuously differentiable in $u = 0$ *and satisfies the condition*

$$\{s \in \mathbb{N}: (d/du)^s w(u)\,|_{u=0} = 0\} \text{ is finite}, \tag{4}$$

then $\forall \varepsilon > 0\ \exists \xi \in \mathbb{R}^k$: $E[u \cdot w(\xi^T x)] \neq 0$ *and* $\|\xi\| < \varepsilon$. *If in addition to* (4), $w(u)$ *is a power series on an open interval* R_0 *of the real line with closure containing* 0: $\forall u \in R_0 \subset \mathbb{R}$: $w(u) = \sum_{s=0}^{\infty}(\gamma_s/s!)u^s$, *where* $\gamma_s = (d/du)^s w(u)\,|_{u=0}$, *then the set* $S = \{\xi \in \mathbb{R}^k$: $E(u \cdot w(x^T\xi)) = 0$ *and* $P[x^T\xi \in R_0] = 1\}$ *has Lebesgue measure zero and is nowhere dense.*

Note that condition (4) applies to the logistic function $w(u) = 1/[1 + \exp(c - u)]$ only if the constant $c \neq 0$, which is just how the logistic function is employed in neural nets. Another example of such a weight function is $w(u) = \cos(u) + \sin(u)$. The result in Theorem 1 implies that $P[E(u\,|\,x) = 0] < 1$ if and only if $\int\{E[u \cdot w(\xi^T x)]\}^2\, d\mu(\xi) > 0$, provided that μ is chosen absolutely continuous with respect to Lebesgue measure, with compact support Ξ having positive Lebesgue measure. The latter condition is assumed throughout this paper. Cf. Assumption A.2. If only condition (4) holds, then the origin should be contained in the interior of Ξ. If the vector x is not bounded, we can without loss of generality replace x in Theorem 1 by $\Phi(x)$, with Φ a bounded one-to-one mapping.

3. THE LIMITING DISTRIBUTION OF THE ICM TEST UNDER LOCAL ALTERNATIVES AND DATA-DEPENDENCE

3.1. *The Local Alternative*

Consider local alternatives of the form

$$H_1^L:\ y_{t,n} = f_t(\theta_0) + g_t/\sqrt{n} + u_t, \tag{5}$$

where the error u_t is the same as under the null hypothesis (1). The detailed maintained hypotheses regarding the f_t, g_t and the weight functions $w_t(\xi)$ are given in the Appendix, as Assumption A. These assumptions allow the g_t's to depend on lagged dependent variables as well.[3]

[3] In the presence of lagged dependent variables in $f_t(\theta)$ and/or g_t there are two, possibly different, interpretations of the local alternative involved. The first interpretation is that the lagged dependent variables in $f_t(\theta)$ and g_t are generated by the null model. Thus, the local alternative is then actually of the form $y_{t,n} = y_t + g_t/\sqrt{n}$, where the y_t's are generated by the null model. The second interpretation is that the lagged dependent variables in f_t and g_t are now the lagged $y_{t,n}$ generated by local alternative. The latter interpretation makes the random variables $f_t(\theta_0)$ and g_t triangular arrays. Although all our assumptions and proofs are stated in terms of single arrays, our results straightforwardly carry over to triangular arrays. The same applies to the weight functions $w_t(\xi)$.

Under the local alternative (5) the process $\hat{z}(\xi)$ now becomes

(6) $$\hat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[u_t + g_t/\sqrt{n} + f_t(\theta_0) - f_t(\hat{\theta})\right] w_t(\xi)$$

where $\hat{\theta}$ is the nonlinear least squares estimator of θ_0. Then it follows from Assumption A, similarly to Bierens (1990),

(7) $$\hat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \phi_t(\xi) + \frac{1}{n} \sum_{t=1}^{n} g_t \phi_t(\xi) + o_p(1),$$

uniformly over ξ in Ξ, where

(8) $$\phi_t(\xi) = w_t(\xi) - b(\theta_0, \xi)^T A(\theta_0)^{-1} (\partial/\partial\theta^T) f_t(\theta_0),$$

with

$$A(\theta) = \text{plim}_{n\to\infty}(1/n) \sum_{t=1}^{n} \left\{(\partial/\partial\theta^T) f_t(\theta)\right\}\{(\partial/\partial\theta) f_t(\theta)\},$$

$$b(\theta, \xi) = \text{plim}_{n\to\infty}(1/n) \sum_{t=1}^{n} (\partial/\partial\theta^T) f_t(\theta) w_t(\xi).$$

3.2. *The Limiting Distribution of the ICM Test under Local Alternatives*

Let

(9) $$z_n(\xi) = (1/\sqrt{n}) \sum_{t=1}^{n} u_t \phi_t(\xi) + (1/n) \sum_{t=1}^{n} g_t \phi_t(\xi).$$

Assumption A guarantees the tightness of the process $z_n(\ )$ and the asymptotic normality of the finite distributions of $z_n(\ )$. See the Appendix. Consequently, z_n converges weakly to a Gaussian process z. Cf. Billingsley (1968). Using (7) Theorem 2 now follows.

THEOREM 2: *Let Assumption A hold. If H_1^L is true, then $\hat{z} \Rightarrow z$, where z is a Gaussian process on Ξ with mean function $\eta(\xi) = \text{plim}_{n\to\infty}(1/n)\Sigma_{t=1}^n g_t \phi_t(\xi)$ and covariance function $\Gamma(\xi_1, \xi_2) = \text{plim}_{n\to\infty}(1/n)\Sigma_{t=1}^n u_t^2 \phi_t(\xi_1)\phi_t(\xi_2)$. Moreover,*

(10) $$\hat{T} \to T = \int z^2(\xi)\, d\mu(\xi) \qquad \textit{in distribution.}$$

In order to analyze the nature of the limiting distribution T in (10), we need the following further elaboration of Mercer's theorem:

LEMMA 1: *Let $\Gamma(\xi_1, \xi_2)$ be a real valued positive semi-definite continuous function on $\Xi \times \Xi$, where Ξ is a compact space, and let μ be a probability* *measure on Ξ. The solutions λ_i and $\psi_i(\cdot)$, $i=1,2,3,\ldots$, of the eigenvalue problem $\int \Gamma(\xi_1,\xi_2)\psi_i(\xi_2)\,d\mu(\xi_2)=\lambda_i\psi_i(\xi_1)$ are real valued and the function Γ has the series representation $\Gamma(\xi_1,\xi_2)=\sum_{i=1}^{\infty}\lambda_i\psi_i(\xi_1)\psi_i(\xi_2)$, where the series involved converges uniformly on $\Xi\times\Xi$. The eigenvalues λ_i are nonnegative and satisfy $\sum_{i=1}^{\infty}\lambda_i<\infty$. Moreover, the eigenfunctions $\psi_i(\cdot)$ are continuous and can be chosen orthonormal and complete in the space $C(\Xi)$ of continuous real functions on Ξ as well as on the space $L_2(\mu)$ of squared integrable functions with respect to μ, i.e.: $\int\psi_i(\xi)\psi_j(\xi)\,d\mu(\xi)=I(i=j)$, and every function ϕ in $C(\Xi)$ or $L_2(\mu)$ can be written as $\phi(\xi)=\sum_{i=1}^{\infty}g_i\psi_i(\xi)$ a.s. $L_2(\mu)$ with Fourier coefficients $g_i=\int\phi(\xi)\psi_i(\xi)\,d\mu(\xi)$ satisfying $\sum_{i=1}^{\infty}g_i^2<\infty$.*

Now let the function Γ in Lemma 1 be equal to the covariance function in Theorem 2. Note that the continuity of $z(\ )$ and the compactness of Ξ imply that $z(\ )$ is square-integrable: $z\in L_2(\mu)$ a.s. Since the set $\{\psi_i(\xi), i=1,2,3,\ldots\}$ of eigenfunctions is complete, we can therefore apply Parseval's equality and conclude from Lemma 1, with ϕ replaced by z, that $T=\sum_{i=1}^{\infty}[\int z(\xi)\psi_i(\xi)\,d\mu(\xi)]^2$. Moreover, the Gaussianity of $z(\ )$ implies that the Fourier coefficients

$$\int z(\xi)\psi_i(\xi)\,d\mu(\xi) \qquad (i=1,2,3,\ldots) \tag{11}$$

are Gaussian too. Therefore, for the characterization of their joint distribution we only need to compute covariances and means. The covariances are:

$$E\left\{\int[z(\xi)-\eta(\xi)]\psi_i(\xi)\,d\mu(\xi)\int[z(\xi)-\eta(\xi)]\psi_j(\xi)\,d\mu(\xi)\right\}$$
$$=\iint\Gamma(\xi_1,\xi_2)\psi_i(\xi_1)\psi_j(\xi_2)\,d\mu(\xi_1)\,d\mu(\xi_2)=\lambda_i I(i=j),$$

so that the sequence (11) is independent. Moreover, it is easy to see that the mean of the ith element of the sequence (11) is just the ith Fourier coefficient of $\eta(\ )$:

$$\eta_i=\int\eta(\xi)\psi_i(\xi)\,d\mu(\xi). \tag{12}$$

Therefore, the asymptotic distribution of the ICM test under the local alternative (5) can be described as follows:

THEOREM 3: *Under the local alternative* (5) *and Assumption A*, $T=\int z(\xi)^2\,d\mu(\xi)\sim\sum_{i=1}^{\infty}(\eta_i+\varepsilon_i\sqrt{\lambda_i})^2$, *where the ε_i are i.i.d. $N(0,1)$, and the η_i are defined by* (12).

Note that the eigenvalues λ_i depend on the covariance function Γ, which in its turn depends on the data-generating process under the null. Cf. Bierens (1990). Therefore, the asymptotic null distribution

$$T_0=\sum_{i=1}^{\infty}\varepsilon_i^2\lambda_i, \tag{13}$$

where ε_i is i.i.d. $N(0,1)$, is case-dependent.

3.3. *Local and Global Power of the ICM Test*

The result of Theorem 3 implies that in general the ICM test has nontrivial $\sqrt{n}$-local power:

COROLLARY 1: *If and only if the mean function* $\eta(\xi)$ *is such that*

$$\int \eta(\xi)^2 \, d\mu(\xi) > 0, \tag{14}$$

then for every $K>0$, $P(T>K) > P(T_0 > K)$.

There is a direct link between local and global power of the ICM test. Consider the global alternative

$$H_1^G\colon\ y_t = f_t(\theta_0) + g_t + u_t, \tag{15}$$

where f_t, g_t, and u_t are the same as before and $\theta_0 = \text{plim}_{n\to\infty} \hat{\theta}$, with $\hat{\theta}$ the nonlinear least squares estimator of the parameter vector of the null model. Note that this probability limit may be different under the null (1) and the global alternative (15). Then it follows from (7) through (9) that under (15), $\eta(\xi) = \text{plim}_{n\to\infty} \hat{z}(\xi)/\sqrt{n}$; hence condition (14) is then equivalent to $\text{plim}_{n\to\infty} \hat{T}/n > 0$. As shown by Bierens (1990), in the cross-section case the latter can be achieved by a suitable choice of the weight functions $w_t(\xi)$ and the measure $\mu(\xi)$. Cf. Section 2. In the time series case, however, one may need to define μ and w_t as functions of infinite dimensional vectors ξ as in De Jong (1996) in order to achieve condition (14). Although all our proofs are based on the condition that the space Ξ has a fixed finite dimension, the key results in this paper will carry over to the general case considered by De Jong (1996).

Note that condition (14) fails to hold for local alternatives with $g_t = \beta^T(\partial/\partial\theta^T) f_t(\theta_0)$ for a fixed parameter vector β and $\theta_0 = \text{plim}_{n\to\infty} \hat{\theta}$, because then it follows easily from (8) and (9) that $\eta(\xi) = 0$. For linear models $f_t(\theta) = x_t^T\theta$ this is not surprising because the local alternative then becomes $y_{t,n} = f_t(\theta_0 + \beta/\sqrt{n}) + u_t$, which is a correctly specified model. For the global alternative (15) the first-order conditions for nonlinear least squares guarantees that $E[g_t(\partial/\partial\theta^T) f_t(\theta_0)] = 0$, hence $g_t = \beta^T(\partial/\partial\theta^T) f_t(\theta_0)$ then implies $\beta = 0$ and thus $g_t = 0$.

4. ASYMPTOTIC POWER OF THE ICM TEST AGAINST LARGE LOCAL ALTERNATIVES

Consider the following class of "large" local alternatives:

$$\begin{aligned} &H_1^L(c)\colon\ y_{t,n} = f_t(\theta_0) + c\sigma g_t/\sqrt{n} + u_t \qquad (t = 1,\dots,n),\\ &\text{with } E(g_t^2) = 1 \text{ and } E(u_t^2 \mid \mathscr{F}_{t-1}) = \sigma^2 \text{ a.s.,} \end{aligned} \tag{16}$$

where c is a "large" positive constant. Clearly, the standardization $E(g_t^2)=1$ does not cause any loss of generality, and the same applies to the factor σ in front of g_t. The present form of the large local alternative involved has been chosen for convenience. The rate of convergence of the asymptotic power function $\Pi_{\text{ICM}}(c)=\lim_{n\to\infty} P[\text{ICM-test rejects } H_0 \mid H_1^L(c)]$ of the ICM test for $c\to\infty$ is given by Lemma 2.

LEMMA 2: *Let K be an arbitrary positive constant. Under Assumption A and the local alternative* (16),

$$(17)\qquad \lim_{c\to\infty} c^{-2}\ln[P(T(c)\le K)] = -\frac{1}{2}\sigma^2\sum_{i=1}^{\infty}\eta_i^2/\lambda_i \ge -\frac{1}{2},$$

where $\eta_i^2/\lambda_i=0$ *if* $\lambda_i=0$.

If the ICM test has nontrivial local power, then by Corollary 1, $\eta_i^2/\lambda_i>0$ for at least one i, hence $\Pi_{\text{ICM}}(c)$ then approaches 1 at an exponential rate as $c\to\infty$. This result establishes once more the relation between local and global power.

Next, we set forth conditions under which the inequality in (17) becomes an equality. First, we need to restrict the local alternative involved to the class of orthogonal alternatives:

$$(18)\qquad E\left[g_t(\partial/\partial\theta^T)f_t(\theta_0)\right]=0.$$

In view of the discussion in Section 3.3 the orthogonality condition (18) is hardly a restriction. Also, we need the stationarity and mild regularity conditions stated in Assumption B in the Appendix. Moreover, it is convenient (but not strictly necessary) to require that the ICM test be consistent. As argued before, in the case of independent data consistency of the ICM test is no restriction, but since we confine our analysis to ICM tests with weight functions $w_t(\xi)$ and probability measure $\mu(\xi)$ on a *fixed* dimensional index set Ξ, consistency is not guaranteed in the time series case. See De Jong (1996). However, even then we can define a class of global alternatives against which the ICM test is consistent, as follows. Let $\Im_{t-1}$ be the σ-algebra generated by $\{w_t(\xi),\xi\in\Xi\}$, where w_t is chosen such that for any $\mathscr{F}_{t-1}$-measurable random variable v_t satisfying $P[E(v_t\mid\Im_{t-1})=0]<1$ the set $\{\xi\in\Xi: E[v_t w_t(\xi)]=0\}$ has Lebesgue measure zero and is nowhere dense. Cf. Theorem 1. Then the ICM test is consistent against all global alternatives (15) for which g_t is measurable with respect to $\Im_{t-1}$. However, then we also need to require that $(\partial/\partial\theta^T)f_t(\theta_0)$ is measurable with respect to $\Im_{t-1}$.

THEOREM 4: *Under Assumptions A and B, the orthogonality condition* (18), *and the condition that either the ICM test is consistent or* g_t *and* $(\partial/\partial\theta^T)f_t(\theta_0)$ *are measurable with respect to* $\Im_{t-1}$, *we have* $\lim_{c\to\infty}c^{-2}\ln[1-\Pi_{\text{ICM}}(c)]=-1/2$.

This result implies that for each δ in the interval (0, 1) we can find a c_δ such that for all $c > c_\delta$,

$$(19)\qquad \exp\left[-\frac{1}{2}(1+\delta)c^2\right] \le 1 - \Pi_{\mathrm{ICM}}(c) \le \exp\left[-\frac{1}{2}(1-\delta)c^2\right].$$

Moreover, note that for the consistent ICM test the result in Theorem 4 is remarkable in that it neither depends on the choice of the weight function w_t and the probability measure μ, nor on the significance level, as long as these choices preserve consistency.

The result in Theorem 4 is even more remarkable if we compare it with the asymptotic t test of the null hypothesis $\delta_0 = 0$ in the auxiliary regression model

$$(20)\qquad \begin{aligned} &y_t = f_t(\theta_0) + \delta_0 g_t^* + u_t \qquad (t = 1, \dots, n),\\ &\text{with}\quad E(u_t^2 \mid \mathscr{F}_{t-1}) = \sigma^2, \quad g_t^* \text{ is stationary and measurable } \mathscr{F}_{t-1},\\ &E\left(g_t^{*2}\right) = 1, \quad E\left[g_t^*(\partial/\partial\theta^T) f_t(\theta_0)\right] = 0, \end{aligned}$$

where g_t^* is some "guess" of the g_t in (16), which for the sake of a fair comparison is assumed to satisfy the same conditions as for g_t. Denoting the least squares estimator of δ_0 by $\hat{\delta}$, it is a standard exercise to show that under Assumptions A and B and the local alternative (16), $\sqrt{n}\,\hat{\delta} \to N(c\sigma\rho, \sigma^2)$ in distribution, where $\rho = \mathrm{corr}(g_t, g_t^*)$. Therefore, under the local alternative (16) the t statistic $\hat{t}_\delta(c)$ of $\hat{\delta}$ satisfies $\hat{t}_\delta(c) \to N(c\rho, 1)$ in distribution. Similarly to the proof of Theorem 4 it is now easy to show the following theorem.

THEOREM 5: *Let Assumptions A and B hold, and denote the asymptotic power function of the t test of the hypothesis* $\delta_0 = 0$ *in the auxiliary regression* (20) *by* $\Pi_t(c) = \lim_{n\to\infty} P[t \text{ test rejects } H_0 \mid H_1^L(c)]$. *Then* $\lim_{c\to\infty} c^{-2} \ln[1 - \Pi_t(c)] = -\rho^2/2$, *where* ρ *is the correlation coefficient of* g_t^* *and* g_t.

Similarly to (19) this result implies that for each δ in the interval (0, 1) we can find a c_δ such that for all $c > c_\delta$,

$$(21)\qquad \exp\left[-\frac{1}{2}(1+\delta)c^2\rho^2\right] \le 1 - \Pi_t(c) \le \exp\left[-\frac{1}{2}(1-\delta)c^2\rho^2\right].$$

Comparing (19) and (21) we see that if the correlation coefficient ρ involved is *not* equal to -1 or $+1$, then there exists a c_0 such that $\Pi_{\mathrm{ICM}}(c) > \Pi_t(c)$ for $c > c_0$. Thus the asymptotic power function of the consistent ICM test converges to 1 at the same rate as the asymptotic power function of the t test only if $g_t^* = g_t$. Theorem 5, though, implies that the t test is consistent against all global alternatives for which g_t^* and g_t have nonzero correlation, but as long as the correlation between g_t^* and g_t is not perfect the ICM test is more powerful than the t test, uniformly for large c's. This is surprising. The common intuition is that a consistent test spreads its power thinly over all possible alternatives, and that therefore a test that is designed to have optimal power against a particular alternative is in general more powerful in a neighborhood of this alternative than a consistent test. Our results refute this.

5. ADMISSIBILITY OF THE ICM TEST

5.1. *Introduction*

We show now, by adapting the approach of Andrews and Ploberger (1993, 1994), that the ICM test in the form of an asymptotic α-level test

(22) $\tau_n = I(\tilde{p} < \alpha),$

where α is the significance level and $\tilde{p}$ is the estimated p-value, is asymptotically admissible; i.e., we show that there does not exist a test that is uniformly more powerful than the ICM test, provided the errors u_t are conditionally normally distributed and some regularity conditions hold. See Assumption C in the Appendix. Note that the asymptotic p-values can be consistently estimated, using the conditional Monte Carlo approach of Hansen (1996) and De Jong (1996).

Consider probability measures $P_{0,n}$, the probability measures which generate the data under the null hypothesis, and a family of probability measures $P_{g,n}$, $g \in G$, representing alternatives. One may interpret the index g as the functional form of the random variables g_t in model (5). In particular, we confine the index set G of alternatives to local alternatives (5) for which Assumption C holds. Note that for such an alternative g we can define $P_{g,n}$ indirectly by the likelihood ratio $dP_{g,n}/dP_{0,n}$, which under Assumption C is well-defined, so that both $P_{0,n}$ and $P_{g,n}$ are defined on the same probability space.

Next, consider weighted alternatives $P_{1,n} = \int P_{g,n} dQ_n(g)$, where the Q_n are probability measures on G. The α-level likelihood ratio test for testing $P_{0,n}$ against $P_{1,n}$ takes the form $\rho_n = I(dP_{1,n}/dP_{0,n} > K_{\alpha,n})$, where $K_{\alpha,n}$ is the corresponding α-fractile of the likelihood ratio involved. We shall show that under the null our ICM test τ_n in the form (22) is asymptotically equivalent to the LR test for a particular measure Q_n, i.e., $P_{0,n}(\tau_n = \rho_n) \to 1$. Now consider an arbitrary sequence γ_n of asymptotic α-level tests competing with τ_n. We distinguish three cases. The first case is where γ_n and τ_n are asymptotically equivalent under the null, i.e.,

(23) $P_{0,n}(\tau_n = \gamma_n) \to 1.$

Then we can show that in the case of the ICM test these two tests are also equivalent under all alternatives $P_{g,n}$, $g \in G$, i.e.,

(24) $P_{g,n}(\tau_n = \gamma_n) \to 1, \quad \text{for each} \quad g \in G.$

The second case is where γ_n and τ_n are essentially different under the null, in the sense that

(25) $\liminf_{n \to \infty} P_{0,n}(\tau_n \neq \gamma_n) > 0.$

Then we can show that

$$(26)\qquad \liminf_{n\to\infty}\left(\int (E_g\tau_n)\,dQ_n(g) - \int (E_g\gamma_n)\,dQ_n(g)\right) > 0.$$

Thus, in this case the tests τ_n have the highest "average" (with respect to Q_n) asymptotic power. The third case is where neither (23) nor (25) are true. Then $\liminf_{n\to\infty} P_{0,n}(\tau_n \neq \gamma_n) = 0$ and $\limsup_{n\to\infty} P_{0,n}(\tau_n \neq \gamma_n) > 0$. Therefore there exists a subsequence n_j such that $\lim_{j\to\infty} P_{0,n_j}(\tau_{n_j} \neq \gamma_{n_j}) = 0$, hence $\lim_{j\to\infty} P_{0,n_j}(\tau_{n_j} = \gamma_{n_j}) = 1$. Also, there exists a subsequence n_j such that $\liminf_{j\to\infty} P_{0,n_j}(\tau_{n_j} \neq \gamma_{n_j}) > 0$, and any subsequence of (n) contains a further subsequence for which one of these results hold. Consequently, the left-hand side of (26) is nonnegative. It is easy to see that the results for these three cases together exclude the possibility that asymptotically the test γ_n is uniformly more powerful than τ_n.

For proving the result (24), we need the following lemma.

LEMMA 3: *Let Assumption C hold. If the tests γ_n and τ_n are asymptotically equivalent under $P_{0,n}$, then so are they under $P_{g,n}$.*

Moreover, for proving (26) we need this lemma:

LEMMA 4: *Let $L_n = dP_{1,n}/dP_{0,n}$ be the likelihood ratio. Assume that under the null $P_{0,n}$, L_n converges in distribution to a continuously distributed random variable L with $E(L) = 1$. If under $P_{0,n}$ the asymptotic α-level test τ_n is asymptotically equivalent to the α-level LR test ρ_n, and γ_n is a competing asymptotic α-level test that is essentially different from τ_n, i.e., (25) holds, then the asymptotic power of the test τ_n is higher than the asymptotic power of the test γ_n (i.e., (26) holds).*

Lemmas 3 and 4 are concerned with tests of simple hypotheses, whereas in the case of the ICM test we have composite hypotheses, because the null distribution as well as the alternative distribution depend on the parameter θ_0. Thus, loosely speaking, the actual index set of alternatives is of the form $G \times \Theta$. However, this is no problem. If for all fixed θ in the interior of Θ there does not exist a test that is, uniformly on G, asymptotically more powerful than the ICM test, then there also cannot exist a test that is uniformly on $G \times \Theta$ asymptotically more powerful than the ICM test. Therefore, we can now merge and extend Lemmas 3 and 4 to the following lemma:

LEMMA 5: *Let Assumption C hold, and let τ_n be the ICM test in the form (22). Let $L_{0,n}(\theta)$ be the likelihood of the data under the null hypothesis for a particular parameter vector θ in Θ. Similarly, let $L_{1,n}(\theta, g)$ be the likelihood of the data under a particular alternative $g \in G$ and a parameter vector $\theta \in \Theta$. Suppose that for any θ*

in Θ *it is possible to construct probability measures* $Q_{\theta,n}$ *which, with* $L_{1,n}(\theta) = \int L_{1,n}(\theta, g)\, dQ_{\theta,n}(g)$, *have the properties that under the null hypothesis,*

$$\ln(L_{1,n}(\theta)/L_{0,n}(\theta)) - \hat{T}/c \to d(\theta) \tag{27}$$

and

$$L_{1,n}(\theta)/L_{0,n}(\theta) \to V_\theta \text{ in distribution}, \quad \text{where} \quad E(V_\theta) = 1, \tag{28}$$

where c is a constant and $d(\theta)$ *a nonrandom function. Then the ICM test* τ_n *is admissible.*

Note that condition (27) ensures that the ICM test is asymptotically equivalent to a LR test, and that, since $\hat{T}$ is asymptotically continuously distributed under the null, so is the likelihood ratio involved. Moreover, the conditions (27) and (28) ensure that the conclusion of Lemma 3 also holds for $P_{1,n}$.

5.2. *Asymptotic Admissibility*

For proving the asymptotic admissibility of the ICM test it suffices now to construct probability measures P_g and $Q_{\theta,n}$ that satisfy the conditions of Lemma 5, as follows. Denote

$$g_{t,i} = \int \phi_t(\xi)\psi_i(\xi)\, d\mu(\xi) \quad \text{if } t \geq 1, \quad g_{t,t,n} = 0 \quad \text{if } t < 1; \tag{29}$$

cf. (8) and Lemma 1. Then it follows from (9) that under the null hypothesis (1), with $\theta(=\theta_0)$ any point in the parameter space Θ satisfying Assumption A, that

$$\hat{T} = \int z_n(\xi)^2\, d\mu(\xi) + o_p(1) = \sum_{i=1}^{\infty} \left(\int z_n(\xi)\psi_i(\xi)\, d\mu(\xi) \right)^2 + o_p(1)$$

$$= \sum_{i=1}^{\infty} \left[(1/\sqrt{n}) \sum_{t=1}^{n} (y_t - f_t(\theta)) \int \phi_t(\xi)\psi_i(\xi)\, d\mu(\xi) \right]^2 + o_p(1);$$

hence under Assumption A and the null hypothesis,

$$\hat{T} = \sum_{i=1}^{\infty} \left[(1/\sqrt{n}) \sum_{t=1}^{n} (y_t - f_t(\theta)) g_{t,i} \right]^2 + o_p(1). \tag{30}$$

However, the random variables $g_{t,i}$ in (29) also form the basis for the following class of alternative hypotheses:

$$H_1^L\colon\ y_{t,n} = f_t(\theta) + (\sigma/\sqrt{n}) \sum_{i=1}^{N_n} \upsilon_i g_{t,i} + u_t, \tag{31}$$

where the υ_i's are random coefficients and N_n converges to infinity with n at a sufficiently slow rate. We can associate these alternatives to a subset G_n of the set G of alternatives considered in Lemmas 3, 4, and 5. Thus, each alternative g in G_n corresponds to a sequence υ_i of coefficients, with $\upsilon_i = 0$ for $i > N_n$, and the null hypothesis corresponds to the case $\upsilon_i = 0$ for $i = 1, 2, \ldots$.

Given a g in G_n, we can now write the log likelihood ratio under the alternative g as:

$$\ln\left(\frac{L_{1,n}(\theta, g)}{L_{0,n}(\theta)}\right) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{N_n} \upsilon_i \sum_{t=1}^{n} g_{t,i}(y_t - f_t(\theta)) - \frac{1}{2}(1/n) \sum_{t=1}^{n} \left(\sum_{i=1}^{N_n} \upsilon_i g_{t,i} \right)^2 .$$

Let g_N be an alternative for which $\upsilon_i = 0$ for $i > N$, where N may possibly depend on n. Denoting

$$(32)\qquad \begin{aligned} a_N &= \frac{1}{\sigma}\left((1/\sqrt{n}) \sum_{t=1}^{n} g_{t,1}(y_t - f_t(\theta)), \ldots, (1/\sqrt{n}) \sum_{t=1}^{n} g_{t,N}(y_t - f_t(\theta))\right)^T, \\ B_N &= \left((1/n) \sum_{t=1}^{n} g_{t,i_1} g_{t,i_2}\right) \qquad (i_1, i_2 = 1, 2, \ldots, N), \end{aligned}$$

and $V_N = (\upsilon_1, \ldots, \upsilon_N)^T$, we can now write the log likelihood ratio as

$$\ln\left(\frac{L_{1,n}(\theta, g_N)}{L_{0,n}(\theta)}\right) = a_N^T V_N - \frac{1}{2} V_N^T B_N V_N .$$

A suitable measure Q_n on G can now be constructed implicitly by letting $V_N \sim N(0, (cI_N - \Lambda_N/\sigma^2)^{-1})$, independently of the data, where $\Lambda_N = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ and $c > \max\{\lambda_i/\sigma^2, i = 1, 2, 3, \ldots\}$, with N depending on n. Then

$$(33)\qquad \int L_{1,n}(\theta, g)\, dQ_n(g)/L_{0,n}(\theta)$$

$$= \frac{\sqrt{\det(cI_N - \Lambda_N/\sigma^2)}\, \exp\left(\frac{1}{2} a_N^T \left(B_N + (cI_N - \Lambda_N/\sigma^2)\right)^{-1} a_N\right)}{\sqrt{\det\left(B_N + (cI_N - \Lambda_N/\sigma^2)\right)}}$$

$$= \frac{L_{1,n}(\theta)}{L_{0,n}(\theta)},$$

say. We show now that condition (27) of Lemma 5 holds. Observe from (8) and Lemma 1 that

$$(34)\qquad (1/n)\sum_{t=1}^{n} g_{t,i_1}g_{t,i_2}$$

$$= (1/n)\sum_{t=1}^{n}\iint \phi_t(\xi_1)\phi_t(\xi_2)\psi_{i_1}(\xi_1)\psi_{i_2}(\xi_2)\,d\mu(\xi_1)\,d\mu(\xi_2)$$

$$= \frac{1}{\sigma^2}\iint \tilde{\Gamma}(\xi_1,\xi_2)\psi_{i_1}(\xi_1)\psi_{i_1}(\xi_2)\,d\mu(\xi_1)\,d\mu(\xi_2)$$

$$\to \frac{1}{\sigma^2}\iint \Gamma(\xi_1,\xi_2)\psi_{i_1}(\xi_1)\psi_{i_1}(\xi_2)\,d\mu(\xi_1)\,d\mu(\xi_2)$$

$$= \frac{\lambda_{i_1} I(i_1 = i_2)}{\sigma^2} \quad \text{where}$$

$$\tilde{\Gamma}(\xi_1,\xi_2) = (1/n)\sum_{t=1}^{n} E(u_t^2 \mid \mathscr{F}_{t-1})\phi_t(\xi_1)\phi_t(\xi_2)$$

$$= \sigma^2(1/n)\sum_{t=1}^{n}\phi_t(\xi_1)\phi_t(\xi_2)$$

and the convergence result involved follows from Assumptions A and C. Thus for fixed N, $B_N \to (1/\sigma^2)\Lambda_N$. Consequently, it follows that under the null hypothesis and Assumptions A and C, $a_N^T(B_N + cI_N - \Lambda_N/\sigma^2)^{-1}a_N - c^{-1}a_N^T a_N \to 0$ ($n \to \infty, N$ fixed). Moreover, it follows from (30) and (32) that $\hat{T} - \sigma^2 a_N^T a_N = \sum_{i=N+1}^{\infty}(\int z_n(\xi)\psi_i(\xi)\,d\mu(\xi))^2 + o_p(1)$. Since for fixed N, $\lim_{n\to\infty} E[\sum_{i=N+1}^{\infty}(\int z_n(\xi)\psi_i(\xi)\,d\mu(\xi))^2] = \sum_{i=N+1}^{\infty}\lambda_i$, it follows now from Lemma 6 below that there exists a sequence N_n converging slowly to infinity with n such that

$$(35)\qquad a_{N_n}^T\left(B_{N_n} + cI_{N_n} - \Lambda_{N_n}/\sigma^2\right)^{-1} a_{N_n} - c^{-1}\hat{T}/\sigma^2 \to 0.$$

LEMMA 6: *If $A_{n,N}$ and B_N are random variables such that $A_{n,N} \to B_N$ for fixed N and $n \to \infty$, and $B_N \to 0$ for $N \to \infty$, then there exists a subsequence N_n^* converging to infinity with n such that $A_{n,N_n} \to 0$ for all subsequences N_n satisfying $N_n \le N_n^*$ and $N_n \to \infty$.*

Moreover, we have

$$(36)\qquad \ln\left(\frac{\sqrt{\det(cI_N - \Lambda_N/\sigma^2)}}{\sqrt{\det(B_N + cI_N - \Lambda_N/\sigma^2)}}\right) \to \ln\left(\sqrt{\det(I_N - c^{-1}\Lambda_N/\sigma^2}\right)$$

for $n \to \infty$ and N fixed, and

$$(37)\qquad \ln\left(\sqrt{\det(I_N - c^{-1}\Lambda_N/\sigma^2)}\right) = \frac{1}{2}\sum_{i=1}^{N}\ln\left(1 - \frac{\lambda_i}{c\sigma^2}\right) \to \frac{1}{2}\sum_{i=1}^{\infty}\ln\left(1 - \frac{\lambda_i}{c\sigma^2}\right)$$

as $N \to \infty$. Again applying Lemma 6, we can replace N in (36) and (37) by the same sequence N_n as before. Combining (35), (36), and (37) then yields

$$\ln\left(\frac{L_{1,n}(\theta)}{L_{0,n}(\theta)}\right) - \left(\frac{\hat{T}}{2c\sigma^2} + \frac{1}{2}\sum_{i=1}^{\infty} \ln\left(1 - \frac{\lambda_i}{c\sigma^2}\right)\right) \to 0 \tag{38}$$

under the null hypothesis, where the likelihood ratio involved is defined by (33) with N replaced by N_n. This proves part (27) of Lemma 5, with $d(\theta) = (1/2)\sum_{i=1}^{\infty} \ln[1 - (\lambda_i/c\sigma^2)]$. Part (28) of Lemma 5 follows from Theorem 3 and (38), i.e., under the null hypothesis

$$\frac{\hat{T}}{2c\sigma^2} + \frac{1}{2}\sum_{i=1}^{\infty} \ln\left(1 - \frac{\lambda_i}{c\sigma^2}\right)$$
$$\to \frac{1}{2}\sum_{i=1}^{\infty} \varepsilon_i^2 \frac{\lambda_i}{c\sigma^2} + \frac{1}{2}\sum_{i=1}^{\infty} \ln\left(1 - \frac{\lambda_i}{c\sigma^2}\right) = \ln(V_\theta),$$

say, in distribution, where the ε_i's are i.i.d. $N(0,1)$. Clearly, $E(V_\theta) = 1$. This completes the admissibility proof:

THEOREM 6: *Under Assumptions A and C, the ICM test in the form* (22) *is admissible.*

6. THE SIZE OF THE ICM TEST

As mentioned before, the practical applicability of the ICM test is hampered by the fact that the limiting distribution of the test statistic under the null hypothesis is case-dependent and can therefore not be tabulated. A possible way to get around this problem is the conditional Monte Carlo approach of Hansen (1996) and De Jong (1996), which however requires a substantial amount of computer time. Therefore, we shall derive case-independent upperbounds of the asymptotic critical values of the ICM test, on the basis of the following lemma:

LEMMA 7: *Let $c_1, \ldots, c_N$ be positive constants such that the equality $(1/k)\sum_{i=1}^{k} c_i = (1/m)\sum_{j=1}^{m} c_j$ implies $k = m$, and let $x_1, \ldots, x_N$ be variables. Then the solution of the LP problem $\max \sum_{j=1}^{N} c_j x_j$ subject to $x_1 \geq x_2 \geq \cdots \geq x_N \geq 0$, $\sum_{j=1}^{N} x_j = 1$, is of the form $x_j = 1/m$ for $j = 1, \ldots, m$, $x_j = 0$ for $j = m+1, \ldots, N$.*

Substituting $c_j = \varepsilon_j^2$ and $x_j = \lambda_j / \sum_{i=1}^{N} \lambda_i$, letting $N \to \infty$, and using the fact that $\int \Gamma(\xi, \xi)\, d\mu(\xi) = \sum_{i=1}^{\infty} \lambda_i = E(T_0)$, Theorem 7 follows now from (13) and Lemma 7:

THEOREM 7: *Let ε_j be NID$(0,1)$ and let*

$$\overline{W} = \sup_{m \geq 1} (1/m) \sum_{j=1}^{m} \varepsilon_j^2. \tag{39}$$

For $\eta > 0$, $P[T_0 > \eta E(T_0)] \le P[\overline{W} > \eta]$, *where* T_0 *is the random variable defined in* (13). *Consequently, under Assumption A and the null hypothesis* (1),

$$\lim_{n \to \infty} P\left[\hat{T} > \eta \int \hat{\Gamma}(\xi, \xi)\, d\mu(\xi)\right] \le P[\overline{W} > \eta].$$

Using 10,000 replications, we have derived the 10%, 5%, and 1% quantiles of the random variable (39) by Monte Carlo simulation, i.e.,

$$(40) \qquad P(\overline{W} > 3.23) = 0.10; \quad P(\overline{W} > 4.26) = 0.05; \quad P(\overline{W} > 6.81) = 0.01.$$

Thus, conducting the ICM test at say the 5% significance level, we reject the null hypothesis of model correctness if

$$(41) \qquad \hat{T}_n > 4.26 \int \hat{\Gamma}(\xi, \xi)\, d\mu(\xi).$$

Note that Bierens (1982) proposed to derive critical values of the ICM test on the basis of Chebyshev's inequality for first moments; e.g., under H_0,

$$(42) \qquad \lim_{n \to \infty} P\left[\hat{T}_n > 20 \int \hat{\Gamma}(\xi, \xi)\, d\mu(\xi)\right] \le 0.05.$$

Comparing (41) and (42) we see that the new upperbounds of the critical values in (40) are much sharper than the ones based on Chebyshev's inequality.

Dept. of Economics, The Pennsylvania State University, 608 Kern Graduate Bldg., University Park, PA 16802, USA, and Tilburg University, the Netherlands,

and

Dept. of Economics, University of St. Andrews, St. Andrews K4 169 AJ, Scotland

Manuscript received March, 1994; final revision received February, 1997.

APPENDIX

Assumptions

ASSUMPTION A.1: *The parameter space* Θ *is a compact subset of* $\mathbb{R}^m$. *The true parameter vector* θ_0 *is contained in the interior of* Θ. *The response function* $f_t(\theta)$ *is twice continuously differentiable on* Θ, *and* u_t *and* $f_{t+1}(\theta)$ *are measurable with respect to* $\mathscr{F}_t$, *where* $\mathscr{F}_t$ *is the sequence of* σ*-algebras generated by* (y_{t-j}, x_{t-j}), $j = 0, 1, 2, \ldots$ *Moreover,* $E(u_t | \mathscr{F}_{t-1}) = 0$ *a.s. Furthermore,* g_t *is measurable with respect to* $\mathscr{F}_{t-1}$.

ASSUMPTION A.2: $w_t(\xi)$ *is a sequence of real valued random functions on* Ξ, *where* Ξ *is a compact subset of a Euclidean space with positive Lebesgue measure, such that* $w_t(\xi)$ *is measurable with respect to* $\mathscr{F}_{t-1}$. *Moreover, the probability measure* μ *is chosen absolutely continuous with respect to Lebesgue measure, with support* Ξ.

ASSUMPTION A.3: *Let* $A_n(\theta) = (1/n)\sum_{t=1}^{n}\{(\partial/\partial\theta^T)f_t(\theta)\}\{(\partial/\partial\theta)f_t(\theta)\}$. *Then* $A_n(\theta) \to A(\theta)$ *uniformly on* Θ, *where* $A(\theta)$ *is a nonstochastic matrix function such that* $A(\theta_0)$ *is positive definite. Moreover, the least squares estimator* $\hat{\theta}$ *satisfies*

$$\sqrt{n}(\hat{\theta} - \theta_0) = A(\theta_0)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \frac{\partial}{\partial\theta^T} f_t(\theta_0) + \frac{1}{n}\sum_{t=1}^{n} g_t \frac{\partial}{\partial\theta^T} f_t(\theta_0)\right) + o_p(1).$$

ASSUMPTION A.4: *Let* $\hat{b}(\theta,\xi) = (1/n)\sum_{t=1}^{n}(\partial/\partial\theta^T)f_t(\theta)w_t(\xi)$. *Then* $\hat{b}(\theta,\xi) \to b(\theta,\xi)$ *uniformly on* $\Theta \times \Xi$, *where* $b(\theta,\xi)$ *is a nonstochastic function satisfying* $\sup_{\theta\in\Theta,\xi\in\Xi}\|b(\theta,\xi)\| < \infty$.

ASSUMPTION A.5: *The weight functions* $w_t(\xi)$ *are differentiable on* Ξ, *and* $\limsup_{n\to\infty}(1/n)\sum_{t=1}^{n} E[u_t^2 \sup_{\xi\in\Xi}\|(\partial/\partial\xi^T)w_t(\xi)\|^2] < \infty$; $(\partial/\partial\xi^T)\hat{b}(\theta,\xi) \to (\partial/\partial\xi^T)b(\theta,\xi)$ *uniformly on* $\Theta \times \Xi$; $\sup_{\theta\in\Theta,\xi\in\Xi}\|(\partial/\partial\xi^T)b(\theta,\xi)\| < \infty$; $(1/n)\sum_{t=1}^{n} E\{u_t^2[(\partial/\partial\theta^T)f_t(\theta_0)] \times [(\partial/\partial\theta)f_t(\theta_0)]\} \to A_2$, *where* A_2 *is finite. There exists a continuous function* $\Gamma(\xi_1,\xi_2)$ *on* $\Xi\times\Xi$ *such that* $(1/n)\sum_{t=1}^{n} E(u_t^2|\mathscr{F}_{t-1})\phi_t(\xi_1)\phi_t(\xi_2) \to \Gamma(\xi_1,\xi_2)$ *uniformly on* $\Xi\times\Xi$, *while pointwise on* $\Xi\times\Xi$, $(1/n)\sum_{t=1}^{n} u_t^2\phi_t(\xi_1)\phi_t(\xi_2) \to \Gamma(\xi_1,\xi_2)$, $(1/n)\sum_{t=1}^{n} E[u_t^2\phi_t(\xi_1)\phi_t(\xi_2)] \to \Gamma(\xi_1,\xi_2)$. *Moreover, for some* $\delta > 0$, $\limsup_{n\to\infty}\sup_{\xi\in\Xi}(1/n)\sum_{t=1}^{n} E|u_t\phi_t(\xi)|^{2+\delta} < \infty$. *There exists a continuous function* $\eta(\xi)$ *on* Ξ *such that* $(1/n)\sum_{t=1}^{n} g_t\phi_t(\xi) \to \lim_{n\to\infty}(1/n)\sum_{t=1}^{n} E[g_t\phi_t(\xi)] = \eta(\xi)$ *uniformly on* Ξ.

ASSUMPTION B: $f_t(\theta)$, $(\partial/\partial\theta^T)f_t(\theta)$, g_t, *and* $w_t(\xi)$ *are stationary. Moreover,* $\operatorname{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} g_t w_t(\xi) = E[g_1 w_1(\xi)]$ *and* $\operatorname{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} w_t(\xi)(\partial/\partial\theta^T)f_t(\theta_0) = E[w_1(\xi)(\partial/\partial\theta^T)f_1(\theta_0)]$ *uniformly on* Ξ. *Furthermore,* $\operatorname{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} g_t(\partial/\partial\theta^T)f_t(\theta_0) = E[g_1(\partial/\partial\theta^T)f_1(\theta_0)]$.

ASSUMPTION C: *The errors* u_t*'s in the models* (1) *and* (5) *are normally distributed:* $u_t|\mathscr{F}_{t-1} \sim N(0,\sigma^2)$. *Moreover, the exogenous variables* x_t*'s (c.f. Assumption A.*1) *are weakly exogenous in the sense of Engle, Hendry, and Richard* (1983). *Furthermore,* $g_t = 0$ *for* $t < 1$, *and under the null hypothesis,* $\operatorname{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} g_t^2$ *exists, is constant, and finite.*

Proofs

PROOF OF THEOREM 1: Similar to Theorem 1 in Bierens (1982, 1990) and Theorem 3.3.4 in Bierens (1994b).

PROOF OF THEOREM 2: We need to show that the finite distributions of the process z_n converge to normal distributions, and that z_n is tight. Cf. Billingsley (1968). The asymptotic normality of the finite distributions of z_n follows easily from the Liapunov-type version in Bierens (1994b, Theorem 6.1.7) of McLeish's (1974) martingale difference central limit theorem. The tightness of z_n follows from Lemma A.1 below, choosing $K_t = \sup_{\xi\in\Xi}\|(\partial/\partial\xi^T)\phi_t(\xi)\|$.

LEMMA A.1: *Let* u_t *be a martingale difference sequence, i.e.,* u_t *is measurable with respect to* $\mathscr{F}_t$, *where* $\mathscr{F}_t$ *is an increasing sequence of* σ*-algebras and* $E(u_t|\mathscr{F}_{t-1}) = 0$ *a.s. Moreover, let* $\phi_t(\xi)$ *be a sequence of random function on a compact subset* Ξ *of a Euclidean space such that* $|\phi_t(\xi_1) - \phi_t(\xi_2)| \le K_t\|\xi_1 - \xi_2\|$ *for each* ξ_1, ξ_2 *in* Ξ, *where* $\phi_t(\xi)$ *and* K_t *are measurable with respect to* $\mathscr{F}_{t-1}$ *and* $\limsup_{n\to\infty}(1/n)\sum_{t=1}^{n} E[u_t^2K_t^2] < \infty$. *Finally, let for one arbitrary* ξ_0 *in* Ξ, $\limsup_{n\to\infty}(1/n)\sum_{t=1}^{n} E[u_t^2\phi_t(\xi_0)^2] < \infty$. *Then the sequence of random functions* $z_n(\xi) = (1/\sqrt{n})\sum_{t=1}^{n} u_t\phi_t(\xi)$ *is tight on* Ξ.

PROOF: Choose an arbitrary $\varepsilon > 0$. We prove the lemma by showing the existence of a sequence of tight random functions $v_n(\xi)$ on Ξ such that $P[z_n = v_n] \ge 1 - \varepsilon$. Denote $A_t(\xi) = \sum_{j=1}^{t} u_j^2\phi_j(\xi)^2$, $B_t = \sum_{j=1}^{t} u_j^2K_j^2$. Now choose an $M > 0$ and define the stopping time $\tau(M) = \sup\{t \le n \mid A_t(\xi_0) \le nM, B_t \le nM\}$ for an arbitrary ξ_0 in Ξ. Since $A_t(\xi_0)$ and B_t are monotonic nondecreasing, and

$\limsup_{n\to\infty}(1/n)E[A_n(\xi_0)] < \infty$, $\limsup_{n\to\infty}(1/n)E[B_n] < \infty$ by the conditions of the lemma under review, it follows from Chebyshev's inequality applied to $A_n(\zeta_0)$ and B_n that there exists an M_ε such that $P[\tau(M_\varepsilon) = n] \geq 1 - \varepsilon$. Next, define $v_n(\xi) = z_{\tau(M_\varepsilon)}(\xi)$. Then $P[z_n = v_n] \geq P[\tau(M_\varepsilon) = n] \geq 1 - \varepsilon$. We show that v_n is tight by applying the Kolmogorov-Cencov criterion (c.f. Kunita (1990, Theorem 1.4.7, p. 38)), i.e., if for some γ, $\delta > 0$ there exists a constant C such that for every ξ_0, ξ_1, ξ_2 in Ξ, $E|v_n(\xi_0)|^\gamma \leq C$, and $E|v_n(\xi_1) - v_n(\xi_2)|^\gamma \leq C\|\xi_1 - \xi_2\|^{k+\delta}$, where k is the dimension of Ξ, then v_n is tight. Now utilize Burkholder's inequality (c.f. Chow and Teicher (1988, p. 396)), i.e., if f_n is a martingale and $S_n = \Sigma_{i=1}^n (f_i - f_{i-1})^2$, then for m, $n \geq 1$, $|E[f_n^m]| \leq C_m E[S_n^{m/2}]$, where $C_m < \infty$ is a constant which does not depend on n. Moreover, n can be an arbitrary adapted and bounded stopping time. Applying this inequality to v_n yields $E|v_n(\xi_0)|^{2k+2} \leq C_{2k+2}(1/n^{k+1}) E(\Sigma_{t=1}^{\tau(M)} u_t^2 \phi_t(\xi_0)^2)^{k+1} \leq C_{2k+2} M^{k+1}$, where the second inequality follows from the definition of the stopping time $\tau(M)$. This proves the first part of the Kolmogorov-Cencov criterion, for $\gamma = 2k+2$.

Finally, again using Burkholder's inequality, the Lipschitz condition on ϕ_t and the definition of $\tau(M)$, it follows that

$$E|v_n(\xi_1) - v_n(\xi_2)|^{2k+2} = (1/n^{k+1})E\left(\left|\sum_{t=1}^{\tau(M)} u_t(\phi_t(\xi_1) - \phi_t(\xi_2))\right|^{2k+2}\right)$$

$$\leq C_k(1/n^{k+1})E\left(\sum_{i=1}^{\tau(M)} u_t^2(\phi_t(\xi_1) - \phi_t(\xi_2))^2\right)^{k+1}$$

$$\leq C_k(1/n^{k+1}E\left(\sum_{i=i}^{\tau(M)} u_t^2 K_t^2\right)^{k+1} \|\xi_1 - \xi_2\|^{2k+2}$$

$$\leq C_k\|\xi_1 - \xi_2|^{2k+2} M^{k+1}.$$

This result proves the second part of the Kolmogorov-Cencov criterion and hence the tightness of v_n. *Q.E.D.*

PROOF OF COROLLARY 1: Denoting $T_i = \Sigma_{j\neq i}(\eta_j + \varepsilon_j\sqrt{\lambda_j})^2$, the corollary follows from repeated application of the easy inequality $P(T \leq K) \leq P(\lambda_i \varepsilon_i^2 + T_i \leq K)$, where the inequality is strict if $\eta_i \neq 0$. Thus if at least one $\eta_i \neq 0$, then the conclusion of the corollary holds. It is easy to verify that condition (14) guarantees this.

PROOF OF LEMMA 1: This lemma is a straightforward further elaboration of Mercer's theorem (cf. Dunford and Schwartz (1963, p. 1088)), mimicking the properties of eigenvalues and eigenvectors of positive definite symmetric matrices.

PROOF OF LEMMA 2: First, observe from Theorem 1 that the limiting distribution of the ICM test statistic under this "large" local alternative is $T(c) = \Sigma_{i=1}^\infty (c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2$, ε_i is i.i.d. $N(0,1)$. Next, let Φ and φ be the c.d.f. and the density, respectively, of the standard normal distribution. It is easy to verify that for $M > 0$ and $x - M > 0$ or $x + M < 0$, $\varphi(|x| + M) \leq \Phi(x+M) - \Phi(x-M) \leq \varphi(|x| - M)$. Taking logs, dividing by x^2, and letting $|x| \to \infty$, it follows that for every $M > 0$, $\lim_{|x|\to\infty} x^{-2}\ln[\Phi(x+M) - \Phi(x-M)] = -1/2$. In its turn this result implies that for $M_i > 0$,

$$\text{(A.1)} \qquad \lim_{c\to\infty} c^{-2}\ln\left\{P\left[\left(c\eta_i + \varepsilon_i\sqrt{\lambda_i}\right)^2 \leq M_i\right]\right\} = -(1/2)\eta_i^2/\lambda_i.$$

Note that this result also holds if $\eta_i^2/\lambda_i = 0$.

The result (A.1) now enables us to prove the equality in (17) in two steps. First, we establish the upperbound of the limit (17), and then the lowerbound.

Step 1: For every $K>0$ and every natural number $N>1$ we have $P[T(c)\leq K]\leq P[\bigcap_{i=1}^{N}\{(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i})^2\leq K\}]=\prod_{i=1}^{N}P[(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i})^2\leq K]$; hence it follows from (A.1) that for arbitrary $N>1$, $\limsup_{c\to\infty}c^{-2}\ln[P(T(c)\leq K)]\leq -(1/2)\sigma^2\sum_{i=1}^{N}\eta_i^2/\lambda_i$. Letting $N\to\infty$, this result implies

$$\text{(A.2)}\qquad \limsup_{c\to\infty}c^{-1}\ln[P(T(c)\leq K)]\leq -\frac{1}{2}\sigma^2\sum_{i=1}^{\infty}\eta_i^2/\lambda_i.$$

Step 2: For arbitrary $K>0$ and natural numbers $N\geq 1$ we have

$$\text{(A.3)}\qquad P(T(c)\leq K)\geq P\left(\bigcap_{i=1}^{N}\left\{\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2N}\right\}\times\bigcap\left\{\sum_{i=N+1}^{\infty}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}\right\}\right)$$
$$=\left(\prod_{i=1}^{N}P\left[\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2N}\right]\right)P\left(\sum_{i=N+1}^{\infty}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}\right).$$

Moreover, for arbitrary $\delta\in(0,K/2)$ we have

$$\text{(A.4)}\qquad P\left(\sum_{i=N+1}^{\infty}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}\right)$$
$$\geq\liminf_{L\to\infty}P\left(\sum_{i=N+1}^{L}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}-\delta\right)$$
$$\times\liminf_{L\to\infty}P\left(\sum_{i=L+1}^{\infty}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\delta\right)$$
$$=\liminf_{L\to\infty}P\left(\sum_{i=N+1}^{\infty}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}-\delta\right),$$

where the equality follows from Chebyshev inequality and the fact that by Lemma 1, $\sum_{i=L+1}^{\infty}\eta_i^2$ and $\sum_{i=L+1}^{\infty}\lambda_i$ converge to zero for $L\to\infty$. Furthermore,

$$\liminf_{L\to\infty}P\left(\sum_{i=N+1}^{L}\left(c\sigma\eta_i+\varepsilon_i\sqrt{\lambda_i}\right)^2\leq\frac{K}{2}-\delta\right)$$
$$=\liminf_{L\to\infty}\int I\left(\sum_{i=N+1}^{L}\lambda_i x_i^2\leq\frac{K}{2}-\delta\right)\times\prod_{i=N+1}^{L}\left(\frac{\exp\left(-\frac{1}{2}\left(x_i-c\sigma\eta_i/\sqrt{\lambda_i}\right)^2\right)}{\sqrt{2\pi}}dx_i\right)$$
$$=\liminf_{L\to\infty}\int I\left(\sum_{i=N+1}^{L}\lambda_i x_i^2\leq\frac{K}{2}-\delta\right)\frac{\exp\left(-\frac{1}{2}\sum_{i=N+1}^{L}x_i^2\right)}{\sqrt{2\pi}^{L-N}}$$
$$\times\exp\left(c\sigma\sum_{i=N+1}^{L}x_i\eta_i/\sqrt{\lambda_i}\right)\exp\left(-\frac{1}{2}c^2\sigma^2\sum_{i=N+1}^{L}\eta_i^2/\lambda_i\right)dx_{N+1}\cdots dx_L$$

$$\geq \exp\left(-\frac{1}{2}c^2\sigma^2 \sum_{i=N+1}^{\infty} \eta_i^2/\sqrt{\lambda_i}\right)$$

$$\times \operatorname{liminf}_{L\to\infty} P\left(\sum_{i=N+1}^{L} \lambda_i \varepsilon_i^2 \leq \frac{K}{2} - \delta \wedge \sum_{i=N+1}^{L} \varepsilon_i \sigma \eta_i/\sqrt{\lambda_i} \geq 0\right);$$

hence,

$$\text{(A.5)} \qquad \operatorname{liminf}_{c\to\infty} c^{-2} \ln\left[\operatorname{liminf}_{L\to\infty} P\left(\sum_{i=N+1}^{L} \left(c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i}\right)^2 \leq \frac{K}{2} - \delta\right)\right]$$

$$\geq -\frac{1}{2}\sigma^2 \sum_{i=N+1}^{\infty} \eta_i^2/\lambda_i + \operatorname{liminf}_{c\to\infty} c^{-2} \ln\left[\operatorname{liminf}_{L\to\infty} P\left(\sum_{i=N+1}^{L} \lambda_i \varepsilon_i^2 \leq \frac{K}{2} - \delta \wedge \sum_{i=N+1}^{L} \varepsilon_i \sigma\eta_i/\sqrt{\lambda_i} \geq 0\right)\right]$$

$$= -\frac{1}{2}\sigma^2 \sum_{i=N+1}^{\infty} \eta_i^2/\lambda_i,$$

where the last conclusion follows from the fact that the liminf of the probability at the right-hand side of (A.5) is positive. Combining (A.1), (A.3), (A.4), and (A.5) now yields

$$\text{(A.6)} \qquad \operatorname{liminf}_{c\to\infty} c^{-2} \ln[P(T(c) \leq K)] \geq -\frac{1}{2}\sigma^2 \sum_{i=1}^{\infty} \eta_i^2/\lambda_i.$$

The equality in (17) now follows from (A.2) and (A.6).

Step 3: In order to prove the inequality in (17), observe that by Assumption A and the conditions in (16), $\lim_{n\to\infty}(1/n)\sum_{t=1}^{n} E[\phi_t(\xi_1)\phi_t(\xi_2)] = \Gamma(\xi_1, \xi_2)/\sigma^2$. Moreover, recall that by Assumption A.5 and Lemma 1, $\lim_{n\to\infty}(1/n)\sum_{t=1}^{n} E[g_t \int \phi_t(\xi)\psi_i(\xi)\,d\mu(\xi)] = \eta_i$. Now let β_i be a sequence of coefficients. Then it follows from Lemma 1 that

$$\text{(A.7)} \qquad \lim_{n\to\infty}(1/n)\sum_{t=1}^{n} E\left(g_t - \sum_{i=1}^{\infty} \beta_i \int_{\Xi} \phi_t(\xi)\psi_i(\xi)\,d\mu(\xi)\right)^2$$

$$= 1 - 2\sum_{i=1}^{\infty} \beta_i \int_{\Xi} \eta(\xi)\psi_i(\xi)\,d\mu(\xi)$$

$$+ \sigma^{-2} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \beta_i\beta_j \int_{\Xi}\int_{\Xi} \Gamma(\xi_1, \xi_2)\psi_i(\xi_1)\psi_j(\xi_2)\,d\mu(\xi_1)\,d\mu(\xi_2)$$

$$= 1 - 2\sum_{i=1}^{\infty} \beta_i\eta_i + \sum_{i=1}^{\infty} \beta_i^2\lambda_i/\sigma^2 = 1 - \sigma^2 \sum_{i=1}^{\infty} \eta_i^2/\lambda_i \geq 0,$$

where the last equality holds for $\beta_i = \sigma^2\eta_i/\lambda_i$. Note that $\lambda_i = 0$ implies $\eta_i = 0$, as otherwise we can choose β_i such that the left-hand side of (A.7) becomes negative. Therefore we may assume that $\eta_i^2/\lambda_i = 0$ if $\lambda_i = 0$.

PROOF OF THEOREM 4: It follows from (A.7) that Theorem 4 is true if

$$v_t = g_t - \sigma^2 \sum_{i=1}^{\infty} (\eta_i/\lambda_i) \int \phi_t(\xi)\psi_i(\xi)\, d\mu(\xi) = 0 \quad \text{a.s.} \tag{A.8}$$

Note that

$$\int \phi_t(\xi)\psi_i(\xi)\, d\mu(\xi)$$
$$= \int w_t(\xi)\psi_i(\xi)\, d\mu(\xi) - \int b(\theta_0,\xi)^T \psi_i(\xi)\, d\mu(\xi) A(\theta_0)^{-1}(\partial/\partial\theta^T) f_t(\theta_0)$$

(cf. (8)); hence if g_t and $(\partial/\partial\theta^T)f_t(\theta_0)$ are measurable with respect to $\mathfrak{I}_{t-1}$, then so is v_t, and therefore $v_t = 0$ a.s. if and only if $E[v_t w_t(\xi)] \equiv 0$. Of course, the latter also holds if the ICM test is consistent. Since by stationarity (Assumption B) and the conditions in (16), $\Gamma(\xi_1,\xi_2) = \sigma^2 E[\phi_t(\xi_1)\phi_t(\xi_2)] = \sigma^2 E[\phi_t(\xi_1)w_t(\xi_2)]$, and by the orthogonality condition (18), $\eta(\xi) = E[g_t w_t(\xi)]$, it follows easily from Lemma 1 that $E[v_t w_t(\xi)] = \eta(\xi) - \sum_{i=1}^{\infty}\eta_i\psi_i(\xi)$. But $\eta(\xi) = \sum_{i=1}^{\infty}\eta_i\psi_i(\xi)$ a.s. μ because $\{\psi_i(\ )\}$ is a complete orthonormal basis of the space $L^2(\mu)$, so that any function η in $L^2(\mu)$ can be written (a.s. μ) as a linear combination of the $\psi_i(\xi)$'s, with (Fourier) coefficients η_i. Hence (A.8) holds.

PROOF OF LEMMA 3: Let $L_n = dP_{1,n}/dP_{0,n}$. Observe that

$$\begin{aligned} P_{g,n}(\tau_n \neq \gamma_n) &\le P_{g,n}(\{L_n > m\} \cup \{L_n \le m \wedge \tau_n \neq \gamma_n\}) \\ &\le P_{g,n}(\{L_n > m\}) + P_{g,n}(\{L_n \le m \wedge \tau_n \neq \gamma_n\}) \\ &\le \int_{\{L_n > m\}} L_n\, dP_{0,n} + \int_{\{L_n \le m \wedge \tau_n \neq \gamma_n\}} L_n\, dP_{0,n} \\ &\le \int_{\{L_n > m\}} L_n\, dP_{0,n} + mP_{0,n}(\tau_n \neq \gamma_n). \end{aligned}$$

Thus for the arbitrary m we have

$$\operatorname{limsup}_{n\to\infty} P_{g,n}(\tau_n \neq \gamma_n) \le \operatorname{limsup}_{n\to\infty} \int_{\{L_n > m\}} L_n\, dP_{0,n}. \tag{A9}$$

Now if under $P_{0,n}$, $L_n \to L$ in distribution, where L is a continuously distributed random variable satisfying $E(L) = 1$, then it follows from Lemma 6.12 in Strasser (1985, p. 36) that L_n is uniformly $(P_{0,n})$-integrable and that therefore, by increasing m, we can make the right-hand side of (A9) arbitrarily small. But due to Assumption C we have under $P_{0,n}$,

$$\ln(L_n) = \frac{1}{\sigma^2}\left((1/\sqrt{n})\sum_{t=1}^{n} u_t g_t - \frac{1}{2}(1/n)\sum_{t=1}^{n} g_t^2\right) \to N(-\omega^2/2, \omega^2)$$

in distribution, where $\omega^2 = \operatorname{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} g_t^2/\sigma^2$; hence $L = \exp(-1/2\omega^2)\exp(\omega\varepsilon)$ with $\varepsilon \sim N(0,1)$, and obviously $E(L) = 1$.

PROOF OF LEMMA 4: Suppose first that the competing test γ_n is an exact α-level test, and that τ_n is an exact α-level LR test: $\tau_n = \rho_n$. Let $K_{\alpha,n}$ be the corresponding α-fractile of the likelihood ratio L_n. It is not hard to verify that

$$\gamma_n - \rho_n = I(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) - I(\{\gamma_n \neq \rho_n\} \cap \{L_n \ge K_{\alpha,n}\}).$$

Since under the null, $E(\gamma_n) = E(\rho_n) = \alpha$, these two equalities imply that

$$P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) = \frac{1}{2}P_{0,n}(\gamma_n \neq \rho_n). \tag{A10}$$

Since we have assumed that the tests γ_n and ρ_n are essentially different, the "liminf" of the right-hand side probability is bounded away from zero; hence there exists a $\delta_0 > 0$ such that

(A11) $$\liminf_{n\to\infty} P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n} - \delta\}) > 0 \quad \text{if} \quad 0 \leq \delta < \delta_0.$$

Next, observe that

(A12) $$E_1(\gamma_n) = \int \gamma_n \, dP_{1,n} = \int \gamma_n L_n \, dP_{0,n} = E_0[(L_n - K_{\alpha,n})\gamma_n] + \alpha K_{\alpha,n},$$

and similarly for ρ_n. Thus, using the fact that $\rho_n = I(L_n > K_{\alpha,n})$, we have

(A13) $$\begin{aligned} E_1(\rho_n) - E_1(\gamma_n) &= E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)] \\ &= E_0[(L_n - K_{\alpha,n})(-\gamma_n)I(L_n < K_{\alpha,n} - \delta)] \\ &\quad + E_0[(L_n - K_{\alpha,n})(-\gamma_n)I(K_{\alpha,n} - \delta \leq L_n \leq K_{\alpha,n})] \\ &\quad + E_0[(L_n - K_{\alpha,n})(1-\gamma_n)I(L_n > K_{\alpha,n})] \\ &\geq \delta E_0(\gamma_n I(L_n < K_{\alpha,n} - \delta)) \\ &= \delta P_{0,n}(\{\rho_n \neq \gamma_n\} \cap \{L_n < K_{\alpha,n} - \delta\}). \end{aligned}$$

Choosing $\delta \in (0, \delta_0)$, and taking "liminf"'s in (A13), it follows now from (A11) and (A13) that $\liminf_{n\to\infty}(E_1(\rho_n) - E_1(\gamma_n)) > 0$. This result carries over to the general case where ρ_n is replaced by the asymptotically equivalent test τ_n and the exact α-level test γ_n is replaced by an asymptotic α-level test. Denoting $\alpha_n = E_0(\gamma_n)$, where $\alpha_n \to \alpha$, equality (A10) then only holds in the limit, α in (A12) needs to be replaced by α_n, and consequently inequality (A13) now holds in "liminf," which is just fine.

Proof of Lemma 5: The conditions of Lemma 5 imply those of Lemmas 3 and 4.

Proof of Lemma 6: Define $m(N) = \inf\{n\colon P[|A_{n,N} - B_N| > 1/N] < 1/N\}$, and let $N_n^* = \max\{N\colon m(N) \leq n\}$. N_n^* is monotonic and cannot be bounded as otherwise there would exist a constant c such that $N_n^* \leq c$ for all n, therefore $m(c+1) > n$ for all n. Thus the existence of a sequence N_n converging to infinity and bounded by N_n^* is guaranteed. Then for all $\varepsilon > 0$, $P(|A_{n,N_n}| > \varepsilon) \leq P(|A_{n,N_n} - B_{N_n}| > \varepsilon) + P(|B_{N_n}| > \varepsilon)$. Clearly, the right-hand side converges to zero.

Proof of Lemma 7: Let $x_i = \sum_{j=i}^N y_j^2$. Then $\sum_{j=1}^N x_j = 1$ implies $\sum_{j=1}^N j y_j^2 = 1$. The linear programming problem involved can now be put in a Lagrange framework, with Lagrange function $L(y_1, \ldots, y_N, \mu) = \sum_{j=1}^N c_j \sum_{i=j}^N y_i^2 + \mu(1 - \sum_{j=1}^N j y_j^2)$. The solution involved follows now easily from the first-order conditions, in particular the conditions $(\partial/\partial y_k)L(y_1, \ldots, y_N, \mu) = 2k y_k[(1/k)\sum_{j=1}^k c_j - \mu] = 0$.

REFERENCES

Andrews, D. W. K., and W. Ploberger (1993): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative," Cowles Foundation Discussion Paper No. 1058, Yale University.
——— (1994): "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.
Bierens, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.
——— (1984): "Model Specification Testing of Time Series Regressions," *Journal of Econometrics*, 26, 323–353.
——— (1987): "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands," *Journal of Econometrics*, 35, 161–190.
——— (1988): "ARMA Memory Index Modeling of Economic Time Series (with Discussion)," *Econometric Theory*, 4, 35–59.
——— (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.
——— (1994a): "Comment on: 'Artificial Neural Networks: An Econometric Perspective'," *Econometric Reviews*, 13, 93–97.

——— (1994b): *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models*. Cambridge: Cambridge University Press.

BIERENS, H. J., AND J. HARTOG (1988): "Nonlinear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function," *Journal of Econometrics*, 38, 269–299.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: John Wiley.

CHOW, Y. S., AND H. TEICHER (1988): *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer.

DE JONG, R. M. (1996): "On the Bierens Test Under Data Dependence," *Journal of Econometrics*, 72, 1–32.

DE JONG, R. M., AND H. J. BIERENS (1994): "On the Limiting Behavior of a Chi-Square Type Test If the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 10, 70–90.

DUNFORD, N., AND J. T. SCHWARTZ (1963): *Linear Operators, Part II: Spectral Theory*. New York: Wiley Interscience.

ENGLE, R. F., D. F. HENDRY, AND J-F. RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865–890.

GOZALO, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9, 451–477.

HANSEN, B. E. (1996): "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

HARDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric Versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947.

HONG, Y., AND H. WHITE (1996): "Consistent Specification Testing via Nonparametric Series Regressions," *Econometrica*, 63, 1133–1160.

HOROWITZ, J., AND W. HARDLE (1994): "Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory*, 10, 821–848.

KUNITA, H. (1990): *Stochastic Flows and Stochastic Differential Equations*. Cambridge: Cambridge University Press.

MCLEISH, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability*, 2, 620–628.

LEE, T. H., H. WHITE, AND C. W. J. GRANGER (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269–290.

NEWEY, W. K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

PLOBERGER, W., AND H. J. BIERENS (1996): "Asymptotic Power of the Integrated Conditional Moment Test Against Global and Large Local Alternatives," Discussion Paper, CentER, Tilburg University, The Netherlands.

STINCHCOMBE, M. B., AND H. WHITE (1993): "Consistent Specification Tests Using Duality and Banach Space Limit Theory," Discussion Paper 93-14, Department of Economics, University of California, San Diego.

STRASSER, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. New York: De Gruyter.

TAUCHEN, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

WHITE, H. (1989): "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks," *Proceedings of the International Joint Conference on Neural Networks, Washington, D.C.*. San Diego: SOS Printing, II: 451–455.

WOOLDRIDGE, J. M. (1992): "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452–475.

YATCHEW, A. J. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.