Addendum to:
Consistent Model Specification Tests

1. Introduction

In this addendum to Bierens (1982) [B82 hereafter], I will further elaborate on the proposed Integrated Conditional Moment (ICM) test. In particular, I will derive the asymptotic null distribution of this ICM test, propose much sharper upper bounds of the critical values than the ones based on Chebyshev's inequality, and show how to compute bootstrap critical values.

The consistency of the ICM test requires that the variables on which the model error are conditioned are bounded, or are made bounded by a bounded one-to-one mapping. In the latter case I proposed to standardize these variables first by subtracting their sample means and then dividing the centered variables by their sample standard errors. I will now provide the proof that this procedure does not affect the asymptotic results.

Moreover, the finite sample performance of the ICM test will be demonstrated by a numerical example, and in the last section I will suggest avenues to optimize the finite sample power of the ICM test.

The paper B82 is to the best of my knowledge the first paper ever to propose a consistent test of the null hypothesis that the functional form of a (non)linear regression model is correctly specified as a conditional expectation, against all deviations of the null hypothesis. I wrote it, and its companion paper Bierens (1984), in the fall of 1981 while enjoying the hospitality of the University of Minnesota, Minneapolis, as a postdoc. See Pinkse (2013) for more on how these papers came about.

Knowing that two distributions are equal if and only if their characteristic functions are identical, I came to realize that the same applies to Fourier transforms of Borel measurable functions, in the following way.

Theorem 1.1. *Given a random vector* $X \in \mathbb{R}^k$ *and a pair* $g_1(x)$, $g_2(x)$ *of Borel measurable real functions on* $\mathbb{R}^k$ *satisfying* $E[|g_1(X)|] < \infty$, $E[|g_2(X)|] < \infty$, *with corresponding Fourier transforms* $f_1(t) = E[g_1(X)\exp(\mathbf{i}.t'X)]$, $f_2(t) = E[g_2(X)\exp(\mathbf{i}.t'X)]$,[1] $t \in \mathbb{R}^k$, *we have that if* $\Pr[g_1(X) = g_2(X)] < 1$ *then there*

[1] Here and in the sequel the bold $\mathbf{i}$ represents the complex numer $\mathbf{i} = \sqrt{-1}$.

exists a $t_0 \in \mathbb{R}^k$ *for which* $f_1(t_0) \neq f_2(t_0)$, *whereas of course* $\Pr[g_1(X) = g_2(X)] = 1$ *implies* $f_1(t) \equiv f_2(t)$.

Proof. See the proof of part I of Theorem 1 in B82, with $r(x) = g_1(x) - g_2(x)$, and Theorem 3.1.1 in Bierens (1994, p.50). ■

In this addendum I will focus on model specification test 1 in B82, which later by Bierens and Ploberger (1997) was named the Integrated Conditional Moment (ICM) test. Since the purpose of this addendum is to derive the limiting null distribution of the ICM test, I will ignore model specification test 2 in B82, except for a few words at the end of the last section.

This addendum employs elements of measure-theoretical probability theory and complex calculus, at the level of Bierens (2004, Ch. 1-3, 6 & Appendix III). The reader is assumed to be familiar with this material.

2. Nonlinear regression

2.1. Model and assumptions

Following B82, I will consider the nonlinear regression model

$$Y = f(X, \theta_0) + U, \ \theta_0 \in \Theta, \tag{2.1}$$

where

Assumption 2.1.
(1) Y *is the dependent variable satisfying* $E[Y^2] < \infty$;
(2) $X \in \mathbb{R}^k$ *is a vector of explanatory variables;*
(3) $\Theta \subset \mathbb{R}^m$ *is a compact and convex parameter space;*
(4) $f(x, \theta)$ *is a given real function on* $\mathbb{R}^k \times \Theta$, *which for each* $\theta \in \Theta$ *is Borel measurable in* x, *and for each* $x \in \mathbb{R}^k$ *is continuous in* θ;
(5) $E[\sup_{\theta \in \Theta} f(X, \theta)^2] < \infty$;
(6) $\theta_0 = \arg\min_{\theta \in \Theta} E[(Y - f(X, \theta))^2]$ *is a unique interior point of* Θ;
(7) $f(X, \theta)$ *is a.s. twice continuously differentiable on* Θ;
(8) *Denoting* $\nabla f(x, \theta) = (\partial/\partial\theta')f(x, \theta)$ *and* $\nabla^2 f(x, \theta) = (\partial/\partial\theta)(\partial/\partial\theta')f(x, \theta)$ *we have*

$$E\left[\sup_{\theta \in \Theta} ||\nabla f(X, \theta)||^2\right] < \infty,$$

$$E\left[\sup_{\theta\in\Theta}|Y-f(X,\theta)|.||\nabla^2 f(X,\theta)||\right] \quad < \quad \infty,$$

where in the latter case and in the sequel $||.||$ *is the matrix norm*

$$||A|| = \sqrt{\mathrm{trace}(A'A)}.$$

(9) *For all* $\theta \in \Theta$ *the matrix*

$$A(\theta) = E[(\nabla f(X,\theta))(\nabla f(X,\theta))'] \tag{2.2}$$

is positive definite.

Since the components of $\nabla f(x,\theta)$ are limits of $f(x,\theta)$, and limits of Borel measurable functions are Borel measurable itself, part (4) of Assumption 2.1 implies that for each $\theta \in \Theta$ the components of $\nabla f(x,\theta)$ are Borel measurable. The same applies to the elements of $\nabla^2 f(x,\theta)$. C.f. Assumption 3 in B82. Part (8) of Assumption 2.1 is a slight modification of the corresponding parts of Assumption 4 in B82. The reasons for this modification is to avoid the homoskedasticity condition in Assumption 2 in B82, and that I need these modified conditions for the bootstrap procedure in section 6 below.

Note that by Assumption 2.1 and Schwartz inequality,

$$\begin{aligned} &E\left[\sup_{\theta\in\Theta}\left\|(\partial/\partial\theta')(Y-f(X,\theta))^2\right\|\right] \\ &\leq 2\left(\sqrt{E[Y^2]}+\sqrt{E\left[\sup_{\theta\in\Theta}f(X,\theta)^2\right]}\right).\sqrt{E\left[\sup_{\theta\in\Theta}||\nabla f(X,\theta)||^2\right]} < \infty, \end{aligned}$$

so that by the dominated convergence theorem,

$$(\partial/\partial\theta')E[(Y-f(X,\theta))^2] = E[(\partial/\partial\theta')(Y-f(X,\theta))^2]$$

and thus by the first-order condition for a minimum of $E[(Y-f(X,\theta))^2]$,

$$E[U.\nabla f(X,\theta_0)] = (\partial/\partial\theta')E[(Y-f(X,\theta))^2]\big|_{\theta=\theta_0} = 0. \tag{2.3}$$

2.2. Misspecification

The (nonlinear) regression model (2.1) is correctly specified as a conditional expectation model if for some $\theta_0 \in \Theta$,

$$H_0 : \Pr\left(E[Y|X] = f(X, \theta_0)\right) = 1.$$

If so, this θ_0 is just the one defined in Assumption 2.1(6). Therefore, the alternative hypothesis that the model is misspecified can be formulated as

$$H_1 : \Pr\left(E[Y|X] = f(X, \theta_0)\right) < 1.$$

In other words, the model (2.1) is correctly specified if the error term U satisfies $\Pr(E[U|X] = 0) = 1$, and is misspecified if $\Pr(E[U|X] = 0) < 1$.

2.3. Nonlinear least squares

Next, consider the nonlinear least squares (NLLS) estimator

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^{n}(Y_j - f(X_j, \theta))^2,$$

of θ_0, where

Assumption 2.2. $(Y_1, X_1')', (Y_2, X_2')', ..., (Y_n, X_n')'$ *is a random sample from the distribution of* $(Y, X')'$;

It is a standard M-estimation exercise[2] to verify that

Theorem 2.1. *Under Assumptions 2.1 and 2.2,* $\widehat{\theta}_n \stackrel{\text{a.s.}}{\to} \theta_0$, *and*

$$\begin{aligned} \sqrt{n}(\widehat{\theta}_n - \theta_0) &= A(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} U_j \nabla f(X_j, \theta_0) + o_p(1) \\ &\stackrel{\text{d}}{\to} N_m\left(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}\right), \end{aligned}$$

where $U_j = Y_j - f(X_j, \theta_0)$, $A(\theta)$ *is defined by* (2.2) *and*

$$B(\theta) = E[(Y - f(X, \theta))^2(\nabla f(X, \theta))(\nabla f(X, \theta))']. \quad (2.4)$$

[2] See for example Bierens (2004, Ch.6).

Note that by (2.3), $E[U_j \nabla f(X_j, \theta_0)] = 0$, and that $\mathrm{Var}[U_j \nabla f(X_j, \theta_0)] = B(\theta_0)$, so that by the standard multivariate central limit theorem, $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \nabla f(X_j, \theta_0) \xrightarrow{\mathrm{d}} N_m(0, B(\theta_0))$.

Theorem 2.1 is related to Theorems 3 and 4 in B82, except that now I don't assume homoskedasticity under H_0, and that due to the i.i.d. condition (Assumption 2.2), Theorem 2.1 does not require that H_0 is true.

3. The ICM test

Since in general, $E[Y|X] = g(X)$ a.s., where $g(x)$ is a Borel measurable real function on $\mathbb{R}^k$,[3] it follows from Theorem 1.1 above, with $g_1(x) = g(x)$ and $g_2(x) = f(x, \theta_0)$, that H_0 is false if and only if there exists a $t_0 \in \mathbb{R}^k$ such that

$$E[(Y - f(X, \theta_0)) \exp(\mathbf{i}.t_0' X)] \neq 0. \tag{3.1}$$

Observe from the proof of part I of Theorem 1 in B82 that this result is obtained by transforming the Fourier transform $E[r(X) \exp(\mathbf{i}.t' X)]$ of $r(X) = E[Y - f(X, \theta)|X]$ into a difference of two characteristic functions. However, it follows from Theorem 6.5.5 in Chung (1974, p. 185) that two distinct characteristic functions on $\mathbb{R}$ may coincide in an arbitrarily large interval $[-c, c]$. This implies that, in the case $k = 1$, for each constant $c > 0$ there exists a function $r_c(x)$ and distribution function $F_c(x)$ such that, with X a random drawing from F_c, $E[r_c(X) \exp(\mathbf{i}.t.X)] = 0$ on $[-c, c]$ while $\Pr[r_c(X) = 0] < 1$. Consequently, there are cases where (3.1) only holds for t_0 far away from the origin of $\mathbb{R}^k$.

Part II of Theorem 1 in B82 remedies this problem: If X is bounded then under H_1 one can find such a t_0 in an arbitrary small neighborhood of the origin of $\mathbb{R}^k$.

The boundedness assumption is not too restrictive because for any bounded Borel measurable one-to-one mapping $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ with Borel measurable inverse, conditioning on X is equivalent to conditioning on $Z = \Phi(X)$. Thus, under H_1, and with θ_0 defined in Assumption 2.1(6), there exists a t_0 in an arbitrarily small open neighborhood of the origin of $\mathbb{R}^k$ such that $E[(Y - f(X, \theta_0)) \exp(\mathbf{i}.t_0' Z)] \neq 0$, and by continuity of $E[(Y - f(X, \theta_0)) \exp(\mathbf{i}.t' Z)]$ the latter is nonzero in an open neighborhood of t_0 as well.

[3] See Chung (1974, Theorem 9.1.2) or Bierens (2004, Theorem 3.10).

Consequently, for any compact subset Υ of $\mathbb{R}^k$ containing the zero vector in its interior, for example the hypercube

$$\Upsilon = \times_{i=1}^{k}[-\tau_i, \tau_i],\ \ \tau_i > 0 \text{ for } i = 1, 2, ..., k, \tag{3.2}$$

and any absolutely continuous probability measure μ on Υ, for example the uniform probability measure for which

$$\mathrm{d}\mu(t) = \frac{I\left(t \in \Upsilon\right)}{\int_{\Upsilon} 1.\mathrm{d}t}\mathrm{d}t, \tag{3.3}$$

where here and in the sequel $I(.)$ is the indicator function, we have that under H_1,

$$\eta = \int_{\Upsilon} \left|E[(Y - f(X, \theta_0)) \exp(\mathbf{i}.t'Z)]\right|^2 \mathrm{d}\mu(t) > 0, \tag{3.4}$$

whereas of course under H_0, $\eta = 0$.

In view of Theorem 3.3.4 in Bierens (1994) and Theorem 1 in Bierens and Ploberger (1997), part II of Theorem 1 in B82 can be generalized as follows.

Theorem 3.1. *Let V be a random variable satisfying $E[|V|] < \infty$ and let $Z \in \mathbb{R}^k$ be a bounded random vector. If $\Pr(E[V|Z] = 0) < 1$ then the set*

$$S = \{t \in \mathbb{R}^k : E[V \exp(\mathbf{i}.t'Z) = 0\}$$

has Lebesgue measure zero and is nowhere dense.

Therefore, the compact set Υ may be chosen anywhere in $\mathbb{R}^k$ as long as it has positive Lebesgue measure. In particular, there is no need to require that Υ contains the zero vector in its interior. On the other hand, the choice (3.2) for Υ together with the uniform probability measure μ in (3.3) simplify the computation of the ICM test statistic. See Lemma 5.2 below. Moreover, Boning and Sowell (1999) have shown that, given the compact set Υ, the uniform probability measure (3.3) is optimal in the sense that then the ICM test has the greatest weighted average local power as defined in Andrews and Ploberger (1994).

Next, denote

$$\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widehat{U}_j \exp(\mathbf{i}.t'Z_j),\ \text{ where } Z_j = \Phi(X_j), \tag{3.5}$$

with Φ the bounded one-to-one mapping mentioned before and $\widehat{U}_j = Y_j - f(X_j, \widehat{\theta}_n)$ the NLLS residuals. Under H_1 and Assumptions 2.1 and 2.2,,

$$\widehat{W}_n(t)/\sqrt{n} \xrightarrow{\text{p}} E[(Y - f(X, \theta_0)) \exp(\mathbf{i}.t'Z)] = \varsigma(t),$$

say, pointwise in $t \in \mathbb{R}^k$, where $\varsigma(t) \neq 0$ for some t in an open subset of Υ. Hence by the dominated convergence theorem and (3.4)

$$\widehat{\eta}_n = \frac{1}{n} \int_{\Upsilon} |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t) \xrightarrow{\text{p}} \eta = \int_{\Upsilon} |\varsigma(t)|^2 \mathrm{d}\mu(t) > 0 \tag{3.6}$$

under H_1.

On the other hand, under H_0 it follows from Theorem 2.1 and the mean value theorem that

$$\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - E[Y_j|X_j])\phi_j(t) + o_p(1) = W_n(t) + o_p(1), \tag{3.7}$$

where

$$\phi_j(t) = \exp(\mathbf{i}.t'Z_j) - b(t, \theta_0)' A(\theta_0)^{-1} \nabla f(X_j, \theta_0), \tag{3.8}$$

$$b(t, \theta) = E[\nabla f(X, \theta) \exp(\mathbf{i}.t'Z)], \tag{3.9}$$

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - E[Y_j|X_j])\phi_j(t), \tag{3.10}$$

and the $o_p(1)$ term in (3.7) is uniform in $t \in \Upsilon$. Hence, under H_0 the test statistic $n.\widehat{\eta}_n$ of the ICM test involved satisfies

$$n.\widehat{\eta}_n = \int_{\Upsilon} |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t) = \int_{\Upsilon} |W_n(t)|^2 \mathrm{d}\mu(t) + o_p(1). \tag{3.11}$$

In the fall of 1981 when I wrote the paper B82 I did not know how to derive the limiting distribution of the integral $\int_{\Upsilon} |W_n(t)|^2 \mathrm{d}\mu(t)$ because I was at that time unaware of the concept of weak convergence of random functions (c.f. Billingsley 1968). Therefore, I proposed to use upper bounds of the critical values based on Chebyshev's inequality for first moments, taking into account that $\int_{\Upsilon} E[|W_n(t)|^2]\mathrm{d}\mu(t)$ does not depend on n.

In the next section I will briefly explain what weak convergence means and how it applies to $W_n(t)$, and show that $\int_{\Upsilon} |W_n(t)|^2 \mathrm{d}\mu(t)$ converges in distribution.

4. The null distribution of the ICM test

The complex-valued empirical process $W_n(t)$ in (3.10) can be split-up in a real and imaginary part as

$$\begin{aligned} W_n(t) &= \mathrm{Re}[W_n(t)] + \mathbf{i}.\,\mathrm{Im}[W_n(t)] \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} U_j\,\mathrm{Re}[\phi_j(t)] + \mathbf{i}.\frac{1}{\sqrt{n}}\sum_{j=1}^{n} U_j\,\mathrm{Im}[\phi_j(t)] \end{aligned}$$

where $U_j = Y_j - E[Y_j|X_j]$, and

$$\begin{aligned} \mathrm{Re}[\phi_j(t)] &= \cos(t'Z_j) - \mathrm{Re}[b(t,\theta_0)]'A(\theta_0)^{-1}\nabla f(X_j,\theta_0) \\ &\quad \text{with } \mathrm{Re}[b(t,\theta_0)] = E[\cos(t'Z)\nabla f(X,\theta_0)], \\ \mathrm{Im}[\phi_j(t)] &= \sin(t'Z_j) - \mathrm{Im}[b(t,\theta_0)]'A(\theta_0)^{-1}\nabla f(X_j,\theta_0) \\ &\quad \text{with } \mathrm{Im}[b(t,\theta_0)] = E[\sin(t'Z)\nabla f(X,\theta_0)]. \end{aligned}$$

where $Z = \Phi(X)$, $Z_j = \Phi(X_j)$. Since obviously $\mathrm{Re}[\phi_j(t)]$ and $\mathrm{Im}[\phi_j(t)]$ are continuous random functions, and t is confined to a compact set Υ, the sequence W_n is a sequence of random elements of the space $\mathbb{C}(\Upsilon)$ of complex-valued continuous functions on Υ. So the questions arise how to define distributions, and the notion of convergence in distribution, for random functions.

4.1. Weak convergence of random functions

For a sequence X_n of random variables[4] defined on a common probability space $\{\Omega, \mathcal{F}, P\}$, with corresponding distribution functions F_n, it is well-known that convergence in distribution of X_n to X, denoted by $X_n \stackrel{\mathrm{d}}{\to} X$, means that there exists a distribution function $F(x)$ on $\mathbb{R}$ represented by X such that $\lim_{n\to\infty} F_n(x) = F(x)$ pointwise in the continuity points of F. Moreover, in terms of induced probability measures on the Borel sets $B \subset \mathbb{R}$, i.e.,

$$\mu_n(B) = P\left(\{\omega \in \Omega : X_n(\omega) \in B\}\right),\ \mu(B) = P\left(\{\omega \in \Omega : X(\omega) \in B\}\right),$$

$X_n \stackrel{\mathrm{d}}{\to} X$ is equivalent to the statement that for all Borel sets $B \subset \mathbb{R}$ with border ∂B,

$$\lim_{n\to\infty} \mu_n(B) = \mu(B) \text{ if } \mu(\partial B) = 0. \tag{4.1}$$

[4]Not to be confused with the vectors of regressors X_j in $Y_j = f(X_j,\theta_0) + U_j$.

For example, the border of $B_0 = (-\infty, x_0]$ is the singleton $\partial B_0 = \{x_0\}$, hence $\mu(\partial B_0) = \lim_{\delta \downarrow 0}(F(x_0) - F(x_0 - \delta)) = 0$ if and only if x_0 is a continuity point of F. If so, $\mu_n(B_0) = F_n(x_0) \to F(x_0) = \mu(B_0)$.

Convergence in distribution in the form (4.1) carries over to random functions W_n in $\mathbb{C}(\Upsilon)$ provided that we can define Borel sets in $\mathbb{C}(\Upsilon)$. Recall that the collection $\mathcal{B}$ of Borel sets in $\mathbb{R}$ is defined as the smallest σ-algebra containing the collection of all the open sets in $\mathbb{R}$, and similarly for the space $\mathbb{C}$ of complex numbers. Therefore, if we define a metric on $\mathbb{C}(\Upsilon)$ then we can define open sets in $\mathbb{C}(\Upsilon)$ and thus define Borel sets in $\mathbb{C}(\Upsilon)$ in the same way as for $\mathbb{C}$. Since $\mathbb{C}(\Upsilon)$ contains all the continuous functions on Υ, and Υ is compact, an appropriate metric is the "sup" metric

$$||f - g||_{\sup} = \sup_{t \in \Upsilon} |f(t) - g(t)| = \sup_{t \in \Upsilon} \sqrt{(\mathrm{Re}[f(t) - g(t)])^2 + (\mathrm{Im}[f(t) - g(t)])^2}$$

with corresponding norm $||f||_{\sup} = \sup_{t \in \Upsilon} |f(t)|$. Endowed with this metric, $\mathbb{C}(\Upsilon)$ becomes a metric space, for which we can define open sets and Borel sets. Thus, let $\mathcal{B}(\Upsilon)$ now be the smallest σ-algebra containing the collection of all the open sets in $\mathbb{C}(\Upsilon)$.

Since $W_n(t)$ is a complex-valued random function on Υ defined on the probability space $\{\Omega, \mathcal{F}, P\}$, it is actually a mapping $W_n(t, \omega) : \Upsilon \times \Omega \to \mathbb{C}$ such that for all Borel sets B in $\mathbb{C}$ and all $t \in \Upsilon$, $\{\omega \in \Omega : W_n(t, \omega) \in B\} \in \mathcal{F}$. More simply stated, for each $t \in \Upsilon$, $W_n(t)$ is a well-defined complex-valued random variable.

As a (well-defined) random *function*, W_n is a mapping $W_n(., \omega) : \Omega \to \mathbb{C}(\Upsilon)$ such that

$$\text{for all } B \in \mathcal{B}(\Upsilon), \ \{\omega \in \Omega : W_n(., \omega) \in B\} \in \mathcal{F}.$$

Therefore, the induced probability measure of W_n can now be defined as

$$\upsilon_n(B) = P\left(\{\omega \in \Omega : W_n(., \omega) \in B\}\right), \ B \in \mathcal{B}(\Upsilon).$$

Let W be another random element of $\mathbb{C}(\Upsilon)$, with induced probability measure

$$\upsilon(B) = P\left(\{\omega \in \Omega : W(., \omega) \in B\}\right), \ B \in \mathcal{B}(\Upsilon).$$

Then similar to (4.1) we can define "convergence in distribution" of W_n to W, which is now called *weak convergence* and denoted by $W_n \Rightarrow W$, by

$$\lim_{n \to \infty} \upsilon_n(B) = \upsilon(B) \text{ for all } B \in \mathcal{B}(\Upsilon) \text{ for which } \upsilon(\partial B) = 0.$$

For real-valued random variables X_n and X, $X_n \xrightarrow{\text{d}} X$ is equivalent to the statement that for all continuous and bounded real functions φ on $\mathbb{R}$, $\lim_{n\to\infty} E[\varphi(X_n)] = E[\varphi(X)]$. See for example Bierens (2004, Theorem 6.18, p.150). This property carries over to weak convergence: $W_n \Rightarrow W$ if and only if for all bounded and uniformly continuous real functions Ψ on $\mathbb{C}(\Upsilon)$,

$$\lim_{n\to\infty} E[\Psi(W_n)] = E[\Psi(W)]. \tag{4.2}$$

The latter is often used as the definition of $W_n \Rightarrow W$. See Billingsley (1968, Theorem 2.1, p. 11).

Moreover, $X_n \xrightarrow{\text{d}} X$ implies that X_n is stochastically bounded: For each $\varepsilon \in (0,1)$ there exists an $M \in (0,\infty)$ such that $\inf_{n\geq 1} \Pr[|X_n| \leq M] > 1-\varepsilon$, or equivalently in terms of the induced probability measures μ_n, $\inf_{n\geq 1} \mu_n(K) > 1-\varepsilon$, where $K = [-M, M]$. See for example Bierens (2004, Theorem 6.26, p. 158). Thus, stochastic boundedness is a necessary condition for convergence in distribution. A similar condition is necessary for weak convergence of random functions. This condition is called tightness, and is in the present case defined as follows.

Definition 4.1. *A sequence of random functions $W_n \in \mathbb{C}(\Upsilon)$ with corresponding induced probability measures υ_n is tight on Υ if for each $\varepsilon \in (0,1)$ there exists a compact set $K \subset \mathbb{C}(\Upsilon)$ such that $\inf_{n\geq 1} \upsilon_n(K) > 1-\varepsilon$.*

Obviously, for sequences of random variables and vectors the notion of tightness is the same as the notion of stochastic boundedness.

To prove that the empirical process W_n in (3.10) is tight, it suffices to prove that $\mathrm{Re}[W_n]$ and $\mathrm{Im}[W_n]$ are tight. The latter can easily be verified from the tightness result in Bierens (1990, Lemma 3) and the more general tightness result in Bierens and Ploberger (1996, Lemma A.1). Thus,

Lemma 4.1. *The empirical process W_n in* (3.10) *is tight on Υ.*

Another necessary condition for weak convergence is that the finite-dimensional distributions converge.

Definition 4.2. *The finite-dimensional distributions of a sequence of random functions $W_n \in \mathbb{C}(\Upsilon)$ converge if for arbitrary $m \in \mathbb{N}$ and arbitrary distinct points $t_1, t_2, ..., t_m \in \Upsilon$,*

$(\mathrm{Re}[W_n(t_1)], \mathrm{Im}[W_n(t_1)], \mathrm{Re}[W_n(t_2)], \mathrm{Im}[W_n(t_2)], ..., \mathrm{Re}[W_n(t_m)], \mathrm{Im}[W_n(t_m)])'$
$\xrightarrow{\mathrm{d}} (\mathrm{Re}[W(t_1)], \mathrm{Im}[W(t_1)], \mathrm{Re}[W(t_2)], \mathrm{Im}[W(t_2)], ..., \mathrm{Re}[W(t_m)], \mathrm{Im}[W(t_m)])'$,
where $W \in \mathbb{C}(\Upsilon)$.

It follows straightforwardly from (3.10) and the standard central limit theorem that:

Lemma 4.2. *Under* H_0 *the finite-dimensional distributions of the empirical process* W_n *in* (3.10) *converge to zero-mean multivariate normal distributions corresponding to a zero-mean Gaussian random element* W *of* $\mathbb{C}(\Upsilon)$.

Similar to zero-mean multivariate normal distributions, the induced probability measure of W is completely determined by its covariance function

$$\Gamma(t_1, t_2) = E[W(t_1)\overline{W(t_2)}], \ (t_1, t_2) \in \Upsilon \times \Upsilon,$$

where $\overline{W(t_2)}$ is the complex-conjugate of $W(t_2)$.

The tightness condition together with the condition that the finite-dimensional distributions converge imply weak convergence. See Billingsley (1968, Theorem 8.1). Thus, the conditions in Definitions 4.1 and 4.2 are necessary *and* sufficient for weak convergence. Consequently, it follows from (3.10), Lemmas 4.1 and 4.2 and the easy equality $E[W(t_1)\overline{W(t_2)}] = E[W_n(t_1)\overline{W_n(t_2)}]$ that the following result holds.

Theorem 4.1. *Under* H_0 *and Assumptions 2.1 and 2.2 the empirical process* W_n *in* (3.10) *satisfies* $W_n \Rightarrow W$ *on* Υ, *where* W *is a zero-mean Gaussian random element of* $\mathbb{C}(\Upsilon)$ *with covariance function*

$$\begin{aligned} \Gamma(t_1, t_2) &= E[(Y_1 - f(X_1, \theta_0))^2 \phi_1(t_1)\overline{\phi_1(t_2)}], \\ &= E[(Y_1 - E[Y_1|X_1])^2 \phi_1(t_1)\overline{\phi_1(t_2)}], \ (t_1, t_2) \in \Upsilon \times \Upsilon. \end{aligned} \quad (4.3)$$

So what can now be said about the integral $\int_\Upsilon |W_n(t)|^2 \mathrm{d}\mu(t)$? The answer will be given in the next subsection.

4.2. Convergence of the ICM test statistic

Recall from (4.2) that $W_n \Rightarrow W$ is equivalent to $\lim_{n\to\infty} E[\Psi(W_n)] = E[\Psi(W)]$ for all bounded and uniformly continuous real functions Ψ on $\mathbb{C}(\Upsilon)$. Examples of such functions Ψ are

$$\Psi_1(f|\xi) = \cos\left(\xi \int_\Upsilon |f(t)|^2 \mathrm{d}\mu(t)\right), \ \Psi_2(f|\xi) = \sin\left(\xi \int_\Upsilon |f(t)|^2 \mathrm{d}\mu(t)\right), \ f \in \mathbb{C}(\Upsilon),$$

for arbitrary fixed $\xi \in \mathbb{R}$. Thus, $W_n \Rightarrow W$ implies

$$\lim_{n\to\infty} E[\Psi_1(W_n|\xi)] = E[\Psi_1(W|\xi)], \ \lim_{n\to\infty} E[\Psi_2(W_n|\xi)] = E[\Psi_2(W|\xi)]$$

and thus

$$\begin{aligned}
\lim_{n\to\infty} E\left[\exp\left(\mathbf{i}.\xi \int_\Upsilon |W_n(t)|^2 \mathrm{d}\mu(t)\right)\right] &= \lim_{n\to\infty} E[\Psi_1(W_n|\xi)] + \mathbf{i}. \lim_{n\to\infty} E[\Psi_2(W_n|\xi)] \\
&= E[\Psi_1(W|\xi)] + \mathbf{i}.E[\Psi_2(W|\xi)] \\
&= E\left[\exp\left(\mathbf{i}.\xi \int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t)\right)\right],
\end{aligned}$$

pointwise in $\xi \in \mathbb{R}$. In other words, the characteristic function of $\int_\Upsilon |W_n(t)|^2 \mathrm{d}\mu(t)$ converges pointwise to the characteristic function of $\int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t)$, which implies that

$$\int_\Upsilon |W_n(t)|^2 \mathrm{d}\mu(t) \xrightarrow{\mathrm{d}} \int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t). \tag{4.4}$$

It follows now from (3.5), (3.6), (3.11) and (4.4) that:

Theorem 4.2. *Under H_0 and Assumptions 2.1 and 2.2 the ICM test statistic*

$$\widehat{V}_n = n.\widehat{\eta}_n = \int_\Upsilon |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t) \tag{4.5}$$

converges in distribution to

$$V = \int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t), \tag{4.6}$$

where W is given in Theorem 4.1, whereas under H_1, $p\lim_{n\to\infty} \widehat{\eta}_n = \eta > 0$, hence $p\lim_{n\to\infty} \widehat{V}_n = \infty$.

Actually, the result under H_0 in Theorem 4.2 is a special case of the continuous mapping theorem:

Theorem 4.3. *Let W_n be a sequence of random elements of a metric space Ξ such that $W_n \Rightarrow W$. Then for any continuous mapping $\Psi : \Xi \to \mathbb{R}^m$, $\Psi(W_n) \stackrel{d}{\to} \Psi(W)$.*

Therefore, in the present case, with $\Xi = \mathbb{C}(\Upsilon)$, endowed with the "sup" metric, we also have that under H_0, $\sup_{t\in\Upsilon} |W_n(t)| \stackrel{d}{\to} \sup_{t\in\Upsilon} |W(t)|$, for example, and thus by (3.10), $\sup_{t\in\Upsilon} |\widehat{W}_n(t)| \stackrel{d}{\to} \sup_{t\in\Upsilon} |W(t)|$ as well.

The distribution of V in (4.6) is still unknown though. Moreover, since the induced probability measure of W is completely determined by its covariance function, which by (3.8), (3.9) and (4.3) depends on the joint distribution of U and X, the bounded one-to-one mapping Φ in $Z = \Phi(X)$, and the functional form of $f(x, \theta)$, so does the distribution of V. Therefore, the ICM test is non-pivotal, that is, its null distribution is case dependent. How to solve this problem will be shown in the next two sections.

5. Sharper upper bounds of the critical values

5.1. Series representation on the null distribution

In the addendum to Bierens and Wang (2012) [see Chapter 6] I have shown that the following general result holds.

Lemma 5.1. *Let W be a zero mean complex-valued continuous Gaussian random function on a compact subset Υ of a Euclidean space, and let μ be a probability measure on Υ. There exists a non-negative sequence ω_j satisfying $\sum_{j=1}^{\infty} \omega_j < \infty$ such that $\int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t) = \sum_{j=1}^{\infty} \omega_j \varepsilon_j^2$, where the ε_j's are independent standard normally distributed random variables.*

Proof. See Theorem 5.1 in the addendum to Bierens and Wang (2012) in Chapter 6. This theorem is a revised version of Lemma 4 in Bierens and Wang (2012), which appears to be incorrect. ■

Lemma 5.1 yields two corollaries. First, similar to Bierens and Ploberger (1997, Corollary 1) it follows from Lemma 5.1 that

Theorem 5.1. *The ICM test under review has nontrivial $\sqrt{n}$ local power.*

Second, similar to Bierens and Ploberger (1997, Theorem 7) it follows from Lemma 5.1 that the following result holds.

Theorem 5.2. *Let the conditions of Lemma 5.1 hold, and let*

$$\overline{\chi}_1^2 = \sup_{n\geq 1} \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j^2.$$

Then

$$\Pr\left[\left(\int_\Upsilon E[|W(t)|^2]\mathrm{d}\mu(t)\right)^{-1} \int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t) > y\right] \leq \Pr[\overline{\chi}_2^2 > y]$$

for all $y > 0$. *Therefore, for* $\alpha \in (0,1)$ *and* $c(\alpha)$ *such that* $\Pr[\overline{\chi}_1^2 > c(\alpha)] = \alpha$,

$$\Pr\left[\int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t) > c(\alpha). \int_\Upsilon E[|W(t)|^2]\mathrm{d}\mu(t)\right] \leq \alpha.$$

The values of $c(\alpha)$ for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$ have been calculated in Bierens and Ploberger (1997), i.e.,

$$c(0.01) = 6.81,\ c(0.05) = 4.26,\ c(0.10) = 3.23 \tag{5.1}$$

In contrast, the corresponding upper bounds of the critical values derived from Chebyshev's inequality for first moments, i.e.,

$$\Pr\left[\int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t) > b(\alpha) \int_\Upsilon E[|W(t)|^2]\mathrm{d}\mu(t)\right] \leq \frac{1}{b(\alpha)},$$

are $b(\alpha) = 1/\alpha$, which are obviously much larger than $c(\alpha)$.

5.2. A consistent estimator of the covariance function

In order to be able to use the upper bounds (5.1) of the critical values in practice we need a consistent estimator of the covariance function $\Gamma(t_1, t_2) = E[W(t_1)\overline{W(t_2)}]$ in (4.2). To construct such a consistent estimator, denote for $\theta \in \Theta$ and $(t_1, t_2) \in \Upsilon \times \Upsilon$,

$$\begin{aligned}\Gamma_*(t_1, t_2, \theta) \quad = \quad & \gamma(t_1, t_2, \theta) - b(t_1, \theta)' A(\theta)^{-1} \overline{d(t_2, \theta)} \\ & - d(t_1, \theta)' A(\theta)^{-1} \overline{b(t_2, \theta)} \\ & + b(t_1, \theta)' A(\theta)^{-1} B(\theta) A(\theta)^{-1} \overline{b(t_2, \theta)},\end{aligned} \tag{5.2}$$

where $b(t,\theta)$ is defined in (3.9), $A(\theta)$ and $B(\theta)$ are defined in (2.2) and (2.4), respectively, and

$$\begin{aligned}\gamma(t_1,t_2,\theta) &= E[(Y-f(X,\theta))^2\exp(\mathbf{i}(t_1-t_2)'Z)],\\ d(t,\theta) &= E\left[(Y-f(X,\theta))^2\nabla f(X,\theta)\exp(\mathbf{i}.t'Z)\right].\end{aligned}$$

Then it is not too hard to verify that under H_0,

$$\Gamma(t_1,t_2) = E\left[U_j^2\phi_j(t_1)\overline{\phi_j(t_2)}\right] = \Gamma_*(t_1,t_2,\theta_0).$$

The empirical counter-part of (5.2) is

$$\begin{aligned}\Gamma_n(t_1,t_2,\theta) &= \gamma_n(t_1,t_2,\theta) - b_n(t_1,\theta)'A_n(\theta)^{-1}\overline{d_n(t_2,\theta)}\\ &\quad -d_n(t_1,\theta)'A_n(\theta)^{-1}\overline{b_n(t_2,\theta)}\\ &\quad +b_n(t_1,\theta)'A_n(\theta)^{-1}B_n(\theta)A_n(\theta)^{-1}\overline{b_n(t_2,\theta)},\end{aligned}$$

where

$$\left.\begin{array}{l}\gamma_n(t_1,t_2,\theta) = (1/n)\sum_{j=1}^n(Y_j-f(X_j,\theta))^2\exp(\mathbf{i}(t_1-t_2)'Z_j),\\ b_n(t,\theta) = (1/n)\sum_{j=1}^n\nabla f(X_j,\theta)\exp(\mathbf{i}.t'Z_j),\\ d_n(t,\theta) = (1/n)\sum_{j=1}^n(Y_j-f(X_j,\theta))^2\nabla f(X_j,\theta)\exp(\mathbf{i}.t'Z_j),\\ A_n(\theta) = (1/n)\sum_{j=1}^n(\nabla f(X_j,\theta))(\nabla f(X_j,\theta))',\\ B_n(\theta) = (1/n)\sum_{j=1}^n(Y_j-f(X_j,\theta))^2(\nabla f(X_j,\theta))(\nabla f(X_j,\theta))'.\end{array}\right\} \quad (5.3)$$

Now under Assumptions 2.1 and 2.2,

$$p\lim_{n\to\infty}\sup_{(t_1,t_2,\theta)\in\Upsilon\times\Upsilon\times\Theta}|\Gamma_n(t_1,t_2,\theta)-\Gamma_*(t_1,t_2,\theta)| = 0.$$

This implies the following result.

Theorem 5.3. *Denote $\widehat{\Gamma}_n(t_1,t_2) = \Gamma_n(t_1,t_2,\widehat{\theta}_n)$. Under Assumptions 2.1 and 2.2, $\widehat{\Gamma}_n(t_1,t_2) \xrightarrow{p} \Gamma_*(t_1,t_2,\theta_0)$ uniformly on $\Upsilon\times\Upsilon$. Under H_0, $\Gamma_*(t_1,t_2,\theta_0)$ is equal to the covariance function $\Gamma(t_1,t_2)$ in Theorem 4.1.*

5.3. The standardized ICM test and its implementation

It follows now from Lemma 5.1 and Theorem 5.3 that, with

$$\widehat{T}_n = \frac{\int_\Upsilon|\widehat{W}_n(t)|^2\mathrm{d}\mu(t)}{\int_\Upsilon\widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)}, \quad (5.4)$$

the test statistic of the standardized ICM test, and $c(\alpha)$ the upper bound of the $\alpha \times 100\%$ critical value,

$$\limsup_{n\to\infty} \Pr[\widehat{T}_n > c(\alpha)] \leq \alpha \text{ under } H_0,$$

whereas by (3.6),

$$p\lim_{n\to\infty} \frac{\widehat{T}_n}{n} = \frac{\int_\Upsilon |\varsigma(t)|^2 \mathrm{d}\mu(t)}{\int_\Upsilon \Gamma_*(t,t,\theta_0)\mathrm{d}\mu(t)} > 0 \text{ under } H_1.$$

To make this standardized ICM test operational we need closed form expressions for the integrals $\int_\Upsilon |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t)$ and $\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)$. They are provided in the following lemma, for the case (3.2) with μ the uniform probability measure on Υ.

Lemma 5.2. *With Υ defined by* (3.2) *and μ the uniform probability measure on Υ the integral $\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)$ takes the form*

$$\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t) = \widehat{\sigma}_n^2 - 2.\mathrm{trace}\left[\widehat{A}_{1,n}^{-1}\widehat{C}_{2,n}\right] + \mathrm{trace}\left[\widehat{A}_{1,n}^{-1}\widehat{A}_{2,n}\widehat{A}_{1,n}^{-1}\widehat{C}_{1,n}\right],$$

where

$$\begin{aligned}
\widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{j=1}^n \widehat{U}_j^2, \\
\widehat{A}_{1,n} &= \frac{1}{n}\sum_{j=1}^n \nabla f(X_j,\widehat{\theta}_n)\nabla f(X_j,\widehat{\theta}_n)', \\
\widehat{A}_{2,n} &= \frac{1}{n}\sum_{j=1}^n \widehat{U}_j^2 \nabla f(X_j,\widehat{\theta}_n)\nabla f(X_j,\widehat{\theta}_n)', \\
\widehat{C}_{1,n} &= \frac{1}{n^2}\sum_{j_1=1}^n\sum_{j_2=1}^n \nabla f(X_{j_1},\widehat{\theta}_n)\nabla f(X_{j_2},\widehat{\theta}_n)' P_{j_1,j_2}, \\
\widehat{C}_{2,n} &= \frac{1}{n^2}\sum_{j_1=1}^n\sum_{j_2=1}^n \widehat{U}_{j_1}^2 \nabla f(X_{j_1},\widehat{\theta}_n)\nabla f(X_{j_2},\widehat{\theta}_n)' P_{j_1,j_2},
\end{aligned}$$

with

$$P_{j_1,j_2} = \prod_{i=1}^k \frac{\sin(\tau_i(Z_{i,j_1} - Z_{i,j_2}))}{\tau_i(Z_{i,j_1} - Z_{i,j_2})},$$

where $Z_{i.j}$ is components i of $Z_j = \Phi(X_j)$. Moreover,

$$\int_\Upsilon |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t) = \widehat{\sigma}_n^2 + 2\frac{1}{n}\sum_{j_1=1}^{n-1}\sum_{j_2=j_1+1}^{n} \widehat{U}_{j_1}\widehat{U}_{j_2}P_{j_1,j_2}.$$

The latter expression is in essence equation (14) in B82. The expression for $\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)$ can be verified after some tedious but straightforward complex calculus exercises.

6. Bootstrap

In this section I will adapt the bootstrap procedure in Hansen (1996) to the ICM test. The following theorem illustrates the idea behind this bootstrap procedure.

Theorem 6.1. *Denote for some fixed $M \in \mathbb{N}$ and $i = 1, 2, ..., M$,*

$$\begin{aligned} W_{i,n}(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_{i,j}(Y_j - f(X_j,\theta_0))\exp(\mathbf{i}.t'Z_j) \\ &\quad - b(t,\theta_0)'A(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_{i,j}(Y_j - f(X_j,\theta_0))\nabla f(X_j,\theta_0) \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_{i,j}(Y_j - f(X_j,\theta_0))\phi_j(t), \end{aligned}$$

where the $\varepsilon_{i,j}$'s are independent random drawings from the standard normal distribution, and $\phi_j(t)$ is defined in (3.8). Under Assumptions 2.1 and 2.2 the empirical processes $W_{i,n}$ satisfy $W_{i,n} \Rightarrow W_i^*$ on Υ, where the W_i^*'s are zero-mean Gaussian random elements of $\mathbb{C}(\Upsilon)$ with covariance function

$$\begin{aligned} \Gamma_*(t_1,t_2) &= E[(Y_1 - f(X_1,\theta_0))^2\phi_1(t_1)\overline{\phi_1(t_2)}] \\ &= \Gamma(t_1,t_2) + E[(E[Y_1|X_1] - f(X_1,\theta_0))^2\phi_1(t_1)\overline{\phi_1(t_2)}], \end{aligned}$$

$(t_1,t_2) \in \Upsilon \times \Upsilon$, with $\Gamma(t_1,t_2)$ the covariance function (4.3) of W in Theorem 4.1. Hence,

$$V_{i,n} = \int_\Upsilon |W_{i,n}(t)|^2\mathrm{d}\mu(t) \xrightarrow{\mathrm{d}} V_i^* = \int_\Upsilon |W_i^*(t)|^2\mathrm{d}\mu(t). \tag{6.1}$$

Moreover, under H_0,

$$V_{i,n} = \int_{\Upsilon} |W_{i,n}(t)|^2 \mathrm{d}\mu(t) \xrightarrow{\mathrm{d}} V_i = \int_{\Upsilon} |W_i(t)|^2 \mathrm{d}\mu(t) \tag{6.2}$$

where the W_i's are zero-mean Gaussian random elements of $\mathbb{C}(\Upsilon)$ with covariance function $\Gamma(t_1, t_2)$ of W in Theorem 4.1. Consequently, the random variables $V_1^*, V_2^*, ..., V_M^*$ are i.i.d., and under H_0, $V, V_1, V_2, ..., V_M$ are i.i.d., where V is defined in Theorem 4.2.

Proof. The results up to (6.2) follow similar to Theorems 4.1 and 4.2.

Denote

$$\begin{aligned} W_n^M(\xi) &= (W_{1,n}(t_1), W_{2,n}(t_2), ..., W_{M,n}(t_M))', \\ W^M(\xi) &= (W_1(t_1), W_2(t_2), ..., W_M(t_M))', \text{ with} \\ \xi &= (t_1', t_2', ..., t_M')' \in \Upsilon^M, \end{aligned}$$

which are random elements in the space $\mathbb{C}^M(\Upsilon^M)$ of M-dimensional complex-valued continuous functions on Υ^M.

Since similar to Lemma 4.1, each $W_{i,n}$ is tight, it follows that W_n^M is tight. Moreover, it is easy to verify that the finite-dimensional distributions of W_n^M converge to the corresponding finite-dimensional distributions of W^M. Furthermore, since for $\xi_1 = (t_{1,1}', t_{1,2}', ..., t_{1,M}')' \in \Upsilon^M$ and $\xi_2 = (t_{2,1}', t_{2,2}', ..., t_{2,M}')' \in \Upsilon^M$,

$$E\left[W_n^M(\xi_1)\overline{W_n^M(\xi_2)}'\right] = \mathrm{diag}\left(\Gamma_*(t_{1,1}, t_{2,1}), \Gamma_*(t_{1,2}, t_{2,2}), ..., \Gamma_*(t_{1,M}, t_{2,M})\right)$$

is a complex valued diagonal $M \times M$ matrix and does not depend on n, we must have that

$$E\left[W^M(\xi_1)\overline{W^M(\xi_2)}'\right] = \mathrm{diag}\left(\Gamma_*(t_{1,1}, t_{2,1}), \Gamma_*(t_{1,2}, t_{2,2}), ..., \Gamma_*(t_{1,M}, t_{2,M})\right)$$

as well. The latter implies that the W_i's are i.i.d., and therefore so are the V_i's.

Note that this result does not depend on whether H_0 is true or not, except that under H_0, $\Gamma_*(t_1, t_2) \equiv \Gamma(t_1, t_2)$, and therefore, with W and V defined in Theorem 4.1 and 4.2, respectively, the W_i's are i.i.d. W and thus the V_i's are i.i.d. V. Moreover, $E[W_n(t_1)\overline{W_{i,n}(t_2)}] = 0$ and therefore $E[W(t_1)\overline{W_i(t_2)}] = 0$ as well. Thus, W is independent of the W_i's, which implies that V is independent of $V_1, V_2, ..., V_M$. ■

If it were possible to compute the $V_{i,n}$'s then we could use

$$G_{n,M}(v) = \frac{1}{M}\sum_{i=1}^{M} I(V_{i,n} \leq v)$$

as an estimate the distribution function $G(v) = \Pr[V \leq v]$, and approximate the asymptotic $\alpha \times 100\%$ critical value of the ICM test by the $1-\alpha$ quantile of $G_{n,M}(v)$. However, the $V_{i,n}$'s can not be computed directly.

On the other hand, one may consider to approximate each $V_{i,n}$ by its empirical counter-part

$$\widetilde{V}_{i,n} = \int_{\Upsilon} |\widetilde{W}_{i,n}(t)|^2 \mathrm{d}\mu(t), \tag{6.3}$$

where

$$\begin{aligned}\widetilde{W}_{i,n}(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_{i,j}(Y_j - f(X_j,\widehat{\theta}_n))\exp(\mathbf{i}.t'Z_j) \\ &- b_n(t,\widehat{\theta}_n)' A_n(\widehat{\theta}_n)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_{i,j}(Y_j - f(X_j,\widehat{\theta}_n))\nabla f(X_j,\widehat{\theta}_n),\end{aligned} \tag{6.4}$$

with the $\varepsilon_{i,j}$'s the same as in Theorem 6.1, and $b_n(t,\theta)$ and $A_n(\theta)$ are defined in (5.3).

Indeed, if

Assumption 6.1. $E\left[\sup_{\theta\in\Theta}(Y - f(X,\theta))^2||\nabla f(X,\theta)||^2\right] < \infty$,

then

Theorem 6.2. *Under Assumptions 2.1, 2.2 and 6.1,*

$$\sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)| = o_p(1) \; for \; i = 1, 2, ..., M. \tag{6.5}$$

Consequently, with $\widetilde{V}_{i,n}$ *defined by* (6.3) *and* V_i *by* (6.1) *we have*

$$\left(\widetilde{V}_{1,n}, \widetilde{V}_{2,n}, ..., \widetilde{V}_{M,n}\right)' \xrightarrow{\mathrm{d}} (V_1^*, V_2^*, ..., V_M^*)', \tag{6.6}$$

where $V_1^*, V_2^*, ..., V_M^*$ *are i.i.d. Moreover, under* H_0 *and with* $\widehat{V}_n$ *the ICM test statistic* (4.5) *we have*

$$\left(\widehat{V}_n, \widetilde{V}_{1,n}, \widetilde{V}_{2,n}, ..., \widetilde{V}_{M,n}\right)' \xrightarrow{\text{d}} (V, V_1, V_2, ..., V_M)', \tag{6.7}$$

where $V, V_1, V_2, ..., V_M$ *are i.i.d.*

Proof. Observe that

$$\begin{aligned}
&\widetilde{W}_{i,n}(t) - W_{i,n}(t) \\
&= -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_{i,j}(f(X_j, \widehat{\theta}_n) - f(X_j, \theta_0)) \exp(\mathbf{i}.t'Z_j) \\
&- \left(b_n(t, \widehat{\theta}_n)' A_n(\widehat{\theta}_n)^{-1} - b(t, \theta_0)' A(\theta_0)^{-1}\right) \\
&\quad \times \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_{i,j}(Y_j - f(X_j, \widehat{\theta}_n)) \nabla f(X_j, \widehat{\theta}_n) \\
&+ b(t, \theta_0)' A(\theta_0)^{-1} \\
&\quad \times \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_{i,j} \left(f(X_j, \widehat{\theta}_n) \nabla f(X_j, \widehat{\theta}_n) - f(X_j, \theta_0)) \nabla f(X_j, \theta_0)\right).
\end{aligned}$$

Using the mean value theorem it follows that

$$\begin{aligned}
&\sup_{t \in \Upsilon} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_{i,j} \left(f(X_j, \widehat{\theta}_n) - f(X_j, \theta_0)\right) \exp(\mathbf{i}.t'Z_j) \right\| \\
&\quad \leq ||\sqrt{n}(\widehat{\theta}_n - \theta_0)|| \sup_{(\theta,t) \in \Theta \times \Upsilon} \left\| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{i,j} \nabla f(X_j, \theta) \exp(\mathbf{i}.t'Z_j) \right\| \\
&\quad = o_p(1),
\end{aligned} \tag{6.8}$$

where the latter follows from the fact that by Jennrich's (1969) uniform strong law of large numbers (USLLN),[5] applied to the real and imaginary parts separately,

$$\sup_{(\theta,t) \in \Theta \times \Upsilon} \left\| (1/n) \sum_{j=1}^{n} \varepsilon_{i,j} \nabla f(X_j, \theta) \exp(\mathbf{i}.t'Z_j) \right\| \xrightarrow{\text{a.s.}} 0,$$

[5] See also Bierens (2004, Theorem 6.13, p.149) for a more detailed proof of Jennrich's USLLN.

whereas by Theorem 2.1, $||\sqrt{n}(\widehat{\theta}_n - \theta_0)|| = O_p(1)$.

Similarly, it follows from the USLLN that $\sup_{(t,\theta)\in\Upsilon\times\Theta} ||b_n(t,\theta) - b(t,\theta)|| \overset{\text{a.s.}}{\to} 0$ and $\sup_{\theta\in\Theta} ||A_n(\theta) - A(\theta)|| \overset{\text{a.s.}}{\to} 0$, hence

$$\sup_{t\in\Upsilon} \left\| A_n(\widehat{\theta}_n)^{-1} b_n(t,\widehat{\theta}_n) - A(\theta_0)^{-1} b(t,\theta_0) \right\| \overset{\text{a.s.}}{\to} 0. \tag{6.9}$$

Next, let $\mathcal{D}_n = \sigma(\{(Y_j, X_j')'\}_{j=1}^n)$ be the σ-algebra generated by the data. Then

$$\begin{aligned}
&E\left[\left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_{i,j}(Y_j - f(X_j,\widehat{\theta}_n))\nabla f(X_j,\widehat{\theta}_n) \right\|^2\right] \\
&= E\left(E\left[\left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_{i,j}(Y_j - f(X_j,\widehat{\theta}_n))\nabla f(X_j,\widehat{\theta}_n) \right\|^2 \middle| \mathcal{D}_n\right]\right) \\
&\leq E\left[\frac{1}{n}\sum_{j=1}^n (Y_j - f(X_j,\widehat{\theta}_n))^2 ||\nabla f(X_j,\widehat{\theta}_n)||^2\right] \\
&\leq E\left[\sup_{\theta\in\Theta}(Y - f(X,\theta))^2 ||\nabla f(X,\theta)||^2\right] < \infty
\end{aligned}$$

where the latter is due to Assumption 6.1. This implies that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_{i,j}(Y_j - f(X_j,\widehat{\theta}_n))\nabla f(X_j,\widehat{\theta}_n) = O_p(1). \tag{6.10}$$

At this point it follows now from (6.8), (6.9) and (6.10) that

$$\sup_{t\in\Upsilon} \left| \widetilde{W}_{i,n}(t) - W_{i,n}(t) \right| \leq \sup_{t\in\Upsilon} ||A(\theta_0)^{-1} b(t,\theta_0)|| \times c_n + o_p(1).$$

where $\sup_{t\in\Upsilon} ||A(\theta_0)^{-1} b(t,\theta_0)|| < \infty$ and

$$c_n = \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_{i,j} \left(f(X_j,\widehat{\theta}_n)\nabla f(X_j,\widehat{\theta}_n) - f(X_j,\theta_0)\nabla f(X_j,\theta_0) \right) \right\|^2.$$

Thus, to prove (6.5) it remains to show that $c_n = o_p(1)$, as follows.

Denote $d_n(\theta) = (1/n)\sum_{j=1}^n \|f(X_j,\theta)\nabla f(X_j,\theta) - f(X_j,\theta_0)\nabla f(X_j,\theta_0)\|^2$ and observe that

$$E[c_n|\mathcal{D}_n] = d_n(\widehat{\theta}_n). \tag{6.11}$$

Since by Assumption 6.1, $E\left[\sup_{\theta\in\Theta} d_n(\theta)\right] < \infty$, it follows that the USLLN is applicable, i.e., $\sup_{\theta\in\Theta} |d_n(\theta) - E[d_n(\theta)]| \stackrel{\text{a.s.}}{\to} 0$, which in its turn implies that

$$d_n(\widehat{\theta}_n) \stackrel{\text{a.s.}}{\to} E[d_n(\theta_0)] = 0. \tag{6.12}$$

Now by Chebyshev's inequality for conditional probabilities and conditional first moments it follows from (6.11) that for arbitrary $\delta \in (0,1)$,

$$\Pr[c_n > \delta^{-1} d_n(\widehat{\theta}_n) | \mathcal{D}_n] \leq \delta E[c_n|\mathcal{D}_n]/d_n(\widehat{\theta}_n) = \delta$$

hence for an arbitrary $\varepsilon > 0$,

$$\begin{aligned}
\delta &\geq E\left(\Pr[c_n > \delta^{-1} d_n(\widehat{\theta}_n)|\mathcal{D}_n]\right) \\
&= \Pr[c_n > \delta^{-1} d_n(\widehat{\theta}_n)] \\
&\geq \Pr[c_n > \delta^{-1} d_n(\widehat{\theta}_n) \text{ and } d_n(\widehat{\theta}_n) \leq \delta.\varepsilon] \\
&\geq \Pr[c_n > \varepsilon \text{ and } d_n(\widehat{\theta}_n) \leq \delta.\varepsilon] \\
&= \Pr[c_n > \varepsilon] - \Pr[c_n > \varepsilon \text{ and } d_n(\widehat{\theta}_n) > \delta.\varepsilon] \\
&\geq \Pr[c_n > \varepsilon] - \Pr[d_n(\widehat{\theta}_n) > \delta.\varepsilon],
\end{aligned}$$

and thus by (6.12), $\limsup_{n\to\infty} \Pr[c_n > \varepsilon] \leq \delta$. Letting $\delta \downarrow 0$ it follows that $\lim_{n\to\infty} \Pr[c_n > \varepsilon] = 0$, which proves that $c_n = o_p(1)$.

The result (6.6) follows from Theorem 6.1 and the easy (in)equalities

$$\begin{aligned}
\left|\widetilde{V}_{i,n} - V_{i,n}\right| &= \left|\int_\Upsilon |\widetilde{W}_{i,n}(t)|^2 \mathrm{d}\mu(t) - \int_\Upsilon |W_{i,n}(t)|^2 \mathrm{d}\mu(t)\right| \\
&= \left|\int_\Upsilon \left(\left(\mathrm{Re}[\widetilde{W}_{i,n}(t)]\right)^2 - (\mathrm{Re}[W_{i,n}(t)])^2\right) \mathrm{d}\mu(t)\right. \\
&\quad \left. + \int_\Upsilon \left(\left(\mathrm{Im}[\widetilde{W}_{i,n}(t)]\right)^2 - (\mathrm{Im}[W_{i,n}(t)])^2\right) \mathrm{d}\mu(t)\right| \\
&\leq \int_\Upsilon \left(\mathrm{Re}[\widetilde{W}_{i,n}(t)] - \mathrm{Re}[W_{i,n}(t)]\right)^2 \mathrm{d}\mu(t) \\
&\quad + 2\int_\Upsilon \left|\mathrm{Re}[\widetilde{W}_{i,n}(t)] - \mathrm{Re}[W_{i,n}(t)]\right| . \left|\mathrm{Re}[W_{i,n}(t)]\right| \mathrm{d}\mu(t) \\
&\quad + \int_\Upsilon \left(\mathrm{Im}[\widetilde{W}_{i,n}(t)] - \mathrm{Im}[W_{i,n}(t)]\right)^2 \mathrm{d}\mu(t) \\
&\quad + 2\int_\Upsilon \left|\mathrm{Im}[\widetilde{W}_{i,n}(t)] - \mathrm{Im}[W_{i,n}(t)]\right| . \left|\mathrm{Im}[W_{i,n}(t)]\right| \mathrm{d}\mu(t)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)|^2 + 2.\sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)| \\
&\quad \times \left(\int_\Upsilon |\mathrm{Re}[W_{i,n}(t)]|\, \mathrm{d}\mu(t) + \int_\Upsilon |\mathrm{Im}[W_{i,n}(t)]|\, \mathrm{d}\mu(t) \right) \\
&\leq \sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)|^2 + 2.\sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)| \\
&\quad \times \left(\sqrt{\int_\Upsilon (\mathrm{Re}[W_{i,n}(t)])^2 \mathrm{d}\mu(t)} + \sqrt{\int_\Upsilon (\mathrm{Im}[W_{i,n}(t)])^2 \mathrm{d}\mu(t)} \right) \\
&\leq \sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)|^2 + 4.\sup_{t\in\Upsilon} |\widetilde{W}_{i,n}(t) - W_{i,n}(t)|.\sqrt{V_{i,n}} \\
&= o_p(1)
\end{aligned}
$$

where the latter follows from (6.5) and the well-known fact that $V_{i,n} \xrightarrow{\mathrm{d}} V_i$ implies $V_{i,n} = O_p(1)$.

Finally (6.7) follows from (6.6) and Theorem 6.1. ■

These results suggest to use

$$
\widetilde{G}_{n,M}(v) = \frac{1}{M} \sum_{i=1}^{M} I(\widetilde{V}_{i,n} \leq v)
$$

as an estimate the distribution function

$$
G(v) = \Pr[V_1^* \leq v],
$$

which under H_0 is equal to $\Pr[V \leq v]$.

Indeed,

Theorem 6.3. *Under the conditions of Theorem 6.2, and for arbitrary* $\delta > 0$,

$$
\limsup_{n\to\infty} \Pr\left[|\widetilde{G}_{n,M}(v) - G(v)| > \delta \right] \leq \frac{G(v)(1-G(v))}{\delta^2 M} \leq \frac{1}{4\delta^2 M} \tag{6.13}
$$

hence

$$
\lim_{M\to\infty} \limsup_{n\to\infty} \Pr\left[|\widetilde{G}_{n,M}(v) - G(v)| > \delta \right] = 0, \tag{6.14}
$$

pointwise in $v > 0$. *Moreover, given a* $\beta \in (0,1)$, *let* $\widetilde{q}_{n,M}(\beta)$ *be the* β*-quantile of* $\widetilde{G}_{n,M}(v)$, *i.e.,* $\widetilde{q}_{n,M}(\beta) = \arg\min_{\widetilde{G}_{n,M}(v) > \beta} v$, *and let* $q(\beta)$ *be the* β*-quantile of*

$G(v)$, *i.e.*, $q(\beta) = \arg\min_{G(v)>\beta} v$. *Then* (6.14) *implies*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \Pr\left[|\widetilde{q}_{n,M}(\beta) - q(\beta)| > \delta\right] = 0. \tag{6.15}$$

Proof. Note that

$$\begin{aligned}
&E\left[\left(\widetilde{G}_{n,M}(v) - G(v)\right)^2\right] \\
&= \frac{1}{M^2}\sum_{i_1=1}^{M}\sum_{i_2=1}^{M} E\left[\left(I(\widetilde{V}_{i_1,n} \le v) - G(v)\right)\left(I(\widetilde{V}_{i_2,n} \le v) - G(v)\right)\right] \\
&= \frac{1}{M^2}\sum_{i_1=1}^{M}\sum_{i_2=1}^{M} E\left[I(\widetilde{V}_{i_1,n} \le v) I(\widetilde{V}_{i_2,n} \le v)\right] + G(v)^2 \\
&\qquad - 2G(v)\frac{1}{M}\sum_{i=1}^{M} E\left[I(\widetilde{V}_{i,n} \le v)\right] \\
&= \frac{1}{M^2}\sum_{i_1=1}^{M}\sum_{i_2=1}^{M} \Pr\left[\widetilde{V}_{i_1,n} \le v \text{ and } \widetilde{V}_{i_2,n} \le v\right] + G(v)^2 \\
&-2G(v)\frac{1}{M}\sum_{i=1}^{M} \Pr\left[\widetilde{V}_{i,n} \le v\right].
\end{aligned}$$

It follows from Theorem 6.2 that for $i_1 \ne i_2$,

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left[\widetilde{V}_{i_1,n} \le v \text{ and } \widetilde{V}_{i_2,n} \le v\right] &= \Pr\left[V_{i_1}^* \le v \text{ and } V_{i_2}^* \le v\right] \\
&= \Pr\left[V_{i_1}^* \le v\right] \Pr\left[V_{i_2}^* \le v\right] \\
&= G(v)^2
\end{aligned}$$

whereas for $i_1 = i_2 = i$, $\Pr[\widetilde{V}_{i_1,n} \le v \text{ and } \widetilde{V}_{i_2,n} \le v] = \Pr[\widetilde{V}_{i,n} \le v]$, and

$$\lim_{n\to\infty} \Pr\left[\widetilde{V}_{i,n} \le v\right] = G(v).$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} E\left[\left(\widetilde{G}_{n,M}(v) - G(v)\right)^2\right] &= \frac{1}{M}G(v) + \frac{M^2 - M}{M^2}G(v)^2 - G(v)^2 \\
&= \frac{1}{M}\left(G(v) - G(v)^2\right) \le \frac{1}{4M},
\end{aligned}$$

which by Chebyshev's inequality implies (6.13).

To prove (6.15) observe that

$$\begin{aligned}
&\Pr\left[\widetilde{q}_{n,M}(\beta) > q(\beta)+\varepsilon\right] \\
&\quad \leq \Pr\left[\widetilde{G}_{n,M}\left(\widetilde{q}_{n,M}(\beta)\right) \geq \widetilde{G}_{n,M}\left(q(\beta)+\varepsilon\right)\right] \\
&\quad = \Pr\left[\widetilde{G}_{n,M}\left(q(\beta)+\varepsilon\right) \leq \beta\right] \\
&\quad = \Pr\left[\widetilde{G}_{n,M}\left(q(\beta)+\varepsilon\right) - G(q(\beta)+\varepsilon) \leq \beta - G(q(\beta)+\varepsilon)\right] \\
&\quad \leq \Pr\left[\left|\widetilde{G}_{n,M}\left(q(\beta)+\varepsilon\right) - G(q(\beta)+\varepsilon)\right| \geq G(q(\beta)+\varepsilon) - \beta\right]
\end{aligned}$$

and similarly

$$\begin{aligned}
&\Pr\left[\widetilde{q}_{n,M}(\beta) < q(\beta)-\varepsilon\right] \\
&\quad \leq \Pr\left[\left|\widetilde{G}_{n,M}\left(q(\beta)-\varepsilon\right) - G(q(\beta)-\varepsilon)\right| \geq \beta - G(q(\beta)-\varepsilon)\right].
\end{aligned}$$

It follows now easily from (6.14) that (6.15) holds. ■

Now the asymptotic $\alpha \times 100\%$ critical value of the ICM test is $q(1-\alpha)$ with corresponding bootstrap critical value $\widetilde{q}_{n,M}(1-\alpha)$.

Note that the integral $\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)$ in (5.4) is applicable to each of the integrals $\int_\Upsilon |\widetilde{W}_{i,n}(t)|^2\mathrm{d}\mu(t)$ in (6.4) as well. Therefore, denoting

$$\widetilde{T}_{i,n} = \frac{\int_\Upsilon |\widetilde{W}_{i,n}(t)|^2\mathrm{d}\mu(t)}{\int_\Upsilon \widehat{\Gamma}_n(t,t)\mathrm{d}\mu(t)}, \tag{6.16}$$

Theorem 6.2 yields the following straightforward corollary.

Theorem 6.4. *Under Assumptions 2.1, 2.2 and 6.1, and with $\widetilde{T}_{i,n}$ defined by* (6.16), *we have*

$$\left(\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\mathrm{d}} (T_1^*, T_2^*, ..., T_M^*)',$$

where $T_1^*, T_2^*, ..., T_M^*$ are i.i.d. Moreover, under H_0 and with $\widehat{T}_n$ the ICM test statistic (5.4) *we have*

$$\left(\widehat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\mathrm{d}} (T, T_1, T_2, ..., T_M)',$$

where $T, T_1, T_2, ..., T_M$ are i.i.d.

In the proof of Theorem 6.3 I have implicitly assumed that the distribution function G is continuous on $(0, \infty)$. This is indeed true, as will be shown more generally by the following theorem.

Theorem 6.5. *Let $W(t)$ be a complex-valued zero mean Gaussian random element of the space $\mathbb{C}(\Upsilon)$ of complex-valued continuous functions on a compact subset Υ of a Euclidean space, and let $V = \int_\Upsilon |W(t)|^2 \mathrm{d}\mu(t)$, where μ is a probability measure on Υ. Then the distribution function $G(x) = \Pr[V \leq x]$ of V is continuous on $(0, \infty)$.*

Proof. Recall from Lemma 5.1 that $V \sim \sum_{j=1}^{\infty} \omega_j \chi_{1,j}^2$, where the $\chi_{1,j}^2$'s are independent χ_1^2 distributed random variables, and the ω_j's are positive numbers satisfying $\sum_{j=1}^{\infty} \omega_j < \infty$. Thus,

$$G(x) = \Pr\left[\sum_{j=1}^{\infty} \omega_j \chi_{1,j}^2 \leq x\right].$$

Because distribution functions are always right-continuous it suffices to show that $G(x)$ is left-continuous, i.e., $\lim_{\varepsilon \downarrow 0} G(x - \varepsilon) = G(x)$, as follows. Denote

$$G_K(x) = \Pr\left[\sum_{j=1}^{K} \omega_j \chi_{1,j}^2 \leq x\right]$$

It is not too hard to verify that $G_K(x)$ is continuous for each $K \in \mathbb{N}$. Since by Chebyshev's inequality for first moments,

$$\sum_{j=1}^{\infty} \omega_j \chi_{1,j}^2 - \sum_{j=1}^{K} \omega_j \chi_{1,j}^2 = \sum_{j=K+1}^{\infty} \omega_j \chi_{1,j}^2 \xrightarrow{\text{p}} 0 \text{ as } K \to \infty,$$

it follows that $\lim_{K\to\infty} G_K(x) = G(x)$ pointwise in the continuity points of $G(x)$.

Now let x_0 be a discontinuity point of $G(x)$,[6] so that

$$\lim_{\delta \downarrow 0} (G(x_0) - G(x_0 - \delta)) = \varepsilon > 0.$$

[6] Recall that the set of discontinuity points of a distribution function on $\mathbb{R}$ is countable.

Hence, for all $\delta > 0$ such that $x_0 - \delta$ is a continuity point of G,

$$\varepsilon \leq G(x_0) - G(x_0 - \delta) = G(x_0) - \lim_{K\to\infty} G_K(x_0 - \delta) \tag{6.17}$$

Since G_K is continuous, for each K there exists a $\delta_K > 0$ such that $G_K(x_0) - G_K(x_0 - \delta_K) < \varepsilon/2$, and since for $\delta \in (0, \delta_K]$, $G_K(x_0) - G_K(x_0 - \delta) < \varepsilon/2$ as well, we may without loss of generality assume that $\lim_{K\to\infty} \delta_K = 0$. Then for arbitrary $\delta > 0$ and K so large that $\delta > \delta_K$, we have $G_K(x_0 - \delta_K) \geq G_K(x_0 - \delta)$, hence

$$\begin{aligned} G(x_0) - G_K(x_0 - \delta) &\leq G(x_0) - G_K(x_0 - \delta_K) \\ &= G(x_0) - G_K(x_0) + G_K(x_0) - G_K(x_0 - \delta_K) \\ &\leq G(x_0) - G_K(x_0) + \varepsilon/2 \\ &\leq \varepsilon/2 \end{aligned}$$

where the last inequality follows from $G(x_0) \leq G_K(x_0)$. Again, assuming that $x_0 - \delta$ is a continuity point of G, we have

$$G(x_0) - G(x_0 - \delta) = G(x_0) - \lim_{K\to\infty} G_K(x_0 - \delta) \leq \varepsilon/2. \tag{6.18}$$

Since (6.17) and (6.18) contradict, it follows that G has no discontinuity points. ■

7. Standardization

In equation (126) in B82 I proposed to standardize each component $X_{i,j}$ of X_j by $(X_{i,j} - \overline{X}_{i,n})/S_{i,n}$, where $\overline{X}_{i,n}$ and $S_{i,n}$ are the sample mean and the sample standard error, respectively, of the $X_{i,j}$'s, and then use the arctan(.) transformation to get the corresponding component of $Z_{i,j}$, i.e.,

$$\widetilde{Z}_{n,i,j} = \arctan\left((X_{i,j} - \overline{X}_{i,n})/S_{i,n}\right). \tag{7.1}$$

The reason for the standardization of $X_{i,j}$ in (7.1) is to preserve sufficient variation in $Z_{i,j}$.

As I admitted at the end of section 9 in B82, the asymptotic theory in the paper does not account for $Z_{i,j}$'s of this type, but I also stated that it is not too hard to verify that after some minor modifications the results in the paper carry over.

In Bierens and Wang (2012) it has been shown that in their case this type of data dependent standardization does not affect the asymptotic results involved, provided that $\overline{X}_{i,n} = \xi_i + O_p(1/\sqrt{n})$ and $S_{i,n} = \sigma_i + O_p(1/\sqrt{n})$, where $\xi_i = E[X_i]$, $\sigma_i = \sqrt{\mathrm{var}(X_i)}$, which requires that the components X_i of X have finite fourth moments. Under these conditions it is indeed not too hard to verify that in the present case the standardization (7.1) does not affect the results in B82. On the other hand, if only $E[X_i^2] < \infty$ then $\overline{X}_{i,n} = \xi_i + o_p(1)$ and $S_{i,n} = \sigma_i + o_p(1)$. Proving that under the latter condition the standardization (7.1) does not affect the asymptotic results is much harder than I thought, but my conjecture turns out to be true.

Theorem 7.1. *Suppose that $E[||X||^2] < \infty$. Let $\widetilde{Z}_{n,j} = (\widetilde{Z}_{n,1,j}, \widetilde{Z}_{n,2,j}, ..., \widetilde{Z}_{n,k,j})'$, with $\widetilde{Z}_{n,i,j} = \psi((X_{i,j} - \overline{X}_{i,n})/S_{i,n})$, where $\overline{X}_{i,n}$ and $S_{i,n}$ are the sample mean and sample standard error, respectively, of component i of the $X_{i,j}$'s, and $\psi(x)$ is a bounded strictly monotonic continuously differentiable real function on $\mathbb{R}$ with bounded first derivative $\psi'(x)$. Let $Z_j = (Z_{1,j}, Z_{2,j}, ..., Z_{k,j})'$, where for $i = 1, 2, .., k$, $Z_{i,j} = \psi((X_{i,j} - \xi_i)/\sigma_i)$, with $\xi_i = E[X_{i,1}]$ and $\sigma_i = \sqrt{E[X_{i,1}^2] - \xi_i^2}$. Denote for these Z_j's,*

$$\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - f(X_j, \widehat{\theta}_n)) \exp(\mathbf{i}.t' Z_j)$$

and let

$$\widetilde{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_j - f(X_j, \widehat{\theta}_n)) \exp(\mathbf{i}.t' \widetilde{Z}_{n,j}).$$

Then under Assumptions 2.1 and 2.2,

$$\sup_{t \in \Upsilon} |\widehat{W}_n(t) - \widetilde{W}_n(t)| \xrightarrow{\text{p}} 0, \tag{7.2}$$

if H_0 is true, whereas under H_1,

$$\sup_{t \in \Upsilon} |\widehat{W}_n(t)/\sqrt{n} - \widetilde{W}_n(t)/\sqrt{n}| \xrightarrow{\text{p}} 0. \tag{7.3}$$

Consequently, all the results in the previous sections carry over if $\widehat{W}_n(t)$ and the integral $\int_\Upsilon |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t)$ are replaced by $\widetilde{W}_n(t)$ and $\int_\Upsilon |\widetilde{W}_n(t)|^2 \mathrm{d}\mu(t)$, respectively.

Proof. For notational convenience I will assume that the X_j's are univariate and satisfy $E[X_j^2] < \infty$. Once I have proven Theorem 7.1 for this case, the generalization to the case $X_j \in \mathbb{R}^k$ is easy.

Note that the conditions on ψ imply that its inverse ψ^{-1} is continuous and strictly monotonic. However, in this proof I will use $\psi(x) = \arctan(x)$, for which $\sup_x|\psi(x)| = \pi/2$, $\psi'(x) = (1+x^2)^{-1}$ and $\psi^{-1}(y) = \tan(y)$.

Due to the i.i.d. condition in Assumption 2.2, it follows from Kolmogorov's strong law of large numbers[7] that $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j \overset{\text{a.s.}}{\to} E[X] = \xi_0$, say, and $S_n = \sqrt{\frac{1}{n}\sum_{j=1}^n X_j^2 - \overline{X}_n^2} \overset{\text{a.s.}}{\to} \sqrt{E[X^2] - \xi_0^2} = \sigma_0$, say, so that

$$\overline{X}_n \overset{\text{p}}{\to} \xi_0 \text{ and } S_n \overset{\text{p}}{\to} \sigma_0.$$

I will only use the latter results. Note that $\sigma_0 > 0$ as otherwise X is a.s. constant.

Let $\Xi = [\underline{\xi}, \overline{\xi}]$ with $-\infty < \underline{\xi} < \xi_0 < \overline{\xi} < \infty$, and let $\Sigma = [\underline{\sigma}, \overline{\sigma}]$ with $0 < \underline{\sigma} < \sigma_0 < \overline{\sigma} < \infty$. Moreover, modify the random functions $\widehat{W}_n$ in (3.5) and W_n in (3.10) to

$$\begin{aligned}
\widehat{W}_n(t,\xi,\sigma) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n \widehat{U}_j \exp(\mathbf{i}.t.Z_j(\xi,\sigma)), \\
\widetilde{W}_n(t,\xi,\sigma) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n U_j \phi_j(t,\xi,\sigma),
\end{aligned}$$

respectively, where

$$Z_j(\xi,\sigma) = \arctan((X_j - \xi))/\sigma),$$

$\widehat{U}_j = Y_j - f(X_j, \widehat{\theta}_n)$ is the NLLS residual, $U_j = Y_j - f(X_j, \theta_0)$ is the actual residual, and

$$\begin{aligned}
\phi_j(t,\xi,\sigma) &= \exp(\mathbf{i}.t.Z_j(\xi,\sigma)) - b(t,\xi,\sigma,\theta_0)'A(\theta_0)^{-1}\nabla f(X_j,\theta_0), \text{ with} \\
b(t,\xi,\sigma,\theta) &= E[\nabla f(X_1,\theta)\exp(\mathbf{i}.t.Z_1(\xi,\sigma))].
\end{aligned}$$

Note that similar to (3.7) it follows from Theorem 2.1 and the mean value theorem that under H_0,

$$\sup_{(t,\xi,\sigma)' \in \Upsilon\times\Xi\times\Sigma} |\widehat{W}_n(t,\xi,\sigma) - \widetilde{W}_n(t,\xi,\sigma)| = o_p(1), \tag{7.4}$$

[7] See for example Bierens (2004, Theorem 6.6, p. 144).

uniformly on $\Upsilon \times \Xi \times \Sigma$, where now $\Upsilon = [\underline{\tau}, \overline{\tau}]$ with $-\infty < \underline{\tau} < \overline{\tau} < \infty$.

Consider the following two conjectures.

Conjecture A. *Under* H_0,

$$\widetilde{W}_n(t, \xi, \sigma) - \widetilde{W}_n(t, \xi_0, \sigma_0) \Rightarrow W(t, \xi, \sigma) - W(t, \xi_0, \sigma_0) \text{ on } \Upsilon \times \Xi \times \Sigma,$$

where $W(t, \xi, \sigma)$ *is a complex valued zero-mean continuous Gaussian process on* $\Upsilon \times \Xi \times \Sigma$.

Conjecture B. *Conjecture A implies that under* H_0,

$$\sup_{t \in \Upsilon} \left| \widetilde{W}_n(t, \overline{X}_n, S_n) - \widetilde{W}_n(t, \xi_0, \sigma_0) \right| = o_p(1). \tag{7.5}$$

If these two conjecture are true then it follows from (7.4) and (7.5) that part (7.2) of Theorem 7.1 holds.

I will prove Conjecture B first, as follows.

Proof of Conjecture B. Denote for $\varepsilon > 0$,

$$\Xi_0(\varepsilon) = [\xi_0 - \varepsilon, \xi_0 + \varepsilon], \ \Sigma_0(\varepsilon) = [\sigma_0 - \varepsilon, \sigma_0 + \varepsilon],$$

where ε is so small that $\Xi_0(\varepsilon) \subset \Xi$ and $\Sigma_0(\varepsilon) \subset \Sigma$. Moreover, denote

$$\begin{aligned} \widehat{D}_n &= \sup_{t \in \Upsilon} \left| \widetilde{W}_n(t, \overline{X}_n, S_n) - \widetilde{W}_n(t, \xi_0, \sigma_0) \right|, \\ \widetilde{D}_n(\varepsilon) &= \sup_{(t,\xi,\sigma)' \in \Upsilon \times \Xi_0(\varepsilon) \times \Sigma_0(\varepsilon)} \left| \widetilde{W}_n(t, \xi, \sigma) - \widetilde{W}_n(t, \xi_0, \sigma_0) \right|, \\ D(\varepsilon) &= \sup_{(t,\xi,\sigma)' \in \Upsilon \times \Xi_0(\varepsilon) \times \Sigma_0(\varepsilon)} \left| W(t, \xi, \sigma) - W(t, \xi_0, \sigma_0) \right|, \end{aligned}$$

and note that by the tightness of $W(t, \xi, \sigma)$,

$$\lim_{\varepsilon \downarrow 0} D(\varepsilon) = 0 \text{ a.s.} \tag{7.6}$$

It follows from Conjecture A and the continuous mapping theorem that

$$\widetilde{D}_n(\varepsilon) \xrightarrow{\text{d}} D(\varepsilon), \tag{7.7}$$

and since $\Pr[(\overline{X}_n, S_n) \in \Xi_0(\varepsilon) \times \Sigma_0(\varepsilon)] \to 1$, we have

$$\lim_{n\to\infty} \Pr[\widehat{D}_n \leq \widetilde{D}_n(\varepsilon)] = 1. \tag{7.8}$$

Let $\delta > 0$ be arbitrary, and observe that

$$\begin{aligned}
\Pr\left[\widehat{D}_n \leq \delta\right] &\geq \Pr\left[\widehat{D}_n \leq \delta \text{ and } \widehat{D}_n \leq \widetilde{D}_n(\varepsilon)\right] \\
&\geq \Pr\left[\widetilde{D}_n(\varepsilon) \leq \delta \text{ and } \widehat{D}_n \leq \widetilde{D}_n(\varepsilon)\right] \\
&= \Pr\left[\widetilde{D}_n(\varepsilon) \leq \delta\right] \\
&\quad - \Pr\left[\widetilde{D}_n(\varepsilon) \leq \delta \text{ and } \widehat{D}_n > \widetilde{D}_n(\varepsilon)\right] \\
&\geq \Pr\left[\widetilde{D}_n(\varepsilon) \leq \delta\right] - \Pr\left[\widehat{D}_n > \widetilde{D}_n(\varepsilon)\right].
\end{aligned}$$

Hence by (7.7) and (7.8),

$$\begin{aligned}
\liminf_{n\to\infty} \Pr\left[\widehat{D}_n \leq \delta\right] &\geq \liminf_{n\to\infty} \Pr\left[\widetilde{D}_n(\varepsilon) \leq \delta\right] \\
&\quad - \limsup_{n\to\infty} \Pr\left[\widehat{D}_n > \widetilde{D}_n(\varepsilon)\right] \\
&= \Pr[D(\varepsilon) \leq \delta] \to 1 \text{ as } \varepsilon \downarrow 0,
\end{aligned}$$

where the latter follows from (7.6). Thus, $\widehat{D}_n = o_p(1)$, which is the result in Conjecture B.

Proof of Conjecture A. We need to established first that $\mathrm{Re}[\widetilde{W}_n(t,\xi,\sigma)]$ and $\mathrm{Im}[\widetilde{W}_n(t,\xi,\sigma)]$ are tight on $\Upsilon \times \Xi \times \Sigma$, so that then $\widetilde{W}_n(t,\xi,\sigma) - \widetilde{W}_n(t,\xi_0,\sigma_0)$ is tight.

According to Lemma A.1 in Bierens and Ploberger (1997),

$$\mathrm{Re}[\widetilde{W}_n(t,\xi,\sigma)] = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \,\mathrm{Re}[\phi_j(t,\xi,\sigma)]$$

is tight under the following conditions:
(a) For each observation index j there exists a random variable K_j such that

$$|\mathrm{Re}[\phi_j(t_1,\xi_1,\sigma_1)] - \mathrm{Re}[\phi_j(t_2,\xi_2,\sigma_2)]| \leq K_j ||(t_1,\xi_1,\sigma_1)' - (t_2,\xi_2,\sigma_2)'||$$

for all pairs $(t_1, \xi_1, \sigma_1)'$, $(t_2, \xi_2, \sigma_2)'$ in $\Upsilon \times \Xi \times \Sigma$.
(b) $\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} E[U_j^2 K_j] < \infty$.
(c) $\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} E[U_j^2(\mathrm{Re}[\phi_j(t,\xi,\sigma)])^2] < \infty$ for some $(t,\xi,\sigma)' \in \Upsilon\times\Xi\times\Sigma$.

To prove condition (a), recall that

$$\begin{aligned}
\mathrm{Re}[\phi_j(t,\xi,\sigma)] &= \cos(t.Z_j(\xi,\sigma)) - \mathrm{Re}[b(t,\xi,\sigma,\theta_0)']A(\theta_0)^{-1}\nabla f(X_j,\theta_0),\\
\mathrm{Re}[b(t,\xi,\sigma,\theta_0)] &= E[\nabla f(X_1,\theta_0)\cos(t.Z_1(\xi,\sigma))].
\end{aligned}$$

With $\psi(x) = \arctan(x)$, it follows from the mean value theorem that

$$\begin{aligned}
&|Z_j(\xi_1,\sigma_1) - Z_j(\xi_2,\sigma_2)| \le |(X_j-\xi_1)/\sigma_1 - (X_j-\xi_2)/\sigma_2|\\
&\quad\le \frac{1}{\inf_{\sigma\in\Sigma}\sigma}|\xi_1-\xi_2| + \left(\frac{1}{\inf_{\sigma\in\Sigma}\sigma^2}|X_j| + \sup_{\xi\in\Xi}|\xi|\right)|\sigma_1-\sigma_2|,
\end{aligned}$$

hence

$$\begin{aligned}
&|\cos(t_1.Z_j(\xi_1,\sigma_1)) - \cos(t_2.Z_j(\xi_2,\sigma_2))|\\
&\quad\le |\cos(t_1.Z_j(\xi_1,\sigma_1)) - \cos(t_2.Z_j(\xi_1,\sigma_1))|\\
&\qquad\quad + |\cos(t_2.Z_j(\xi_1,\sigma_1)) - \cos(t_2.Z_j(\xi_2,\sigma_2))|\\
&\quad\le \frac{1}{2}\pi|t_1-t_2| + |Z_j(\xi_1,\sigma_1) - Z_j(\xi_2,\sigma_2)|.\sup_{t\in\Upsilon}|t|\\
&\quad\le \frac{1}{2}\pi.|t_1-t_2| + \frac{\sup_{t\in\Upsilon}|t|}{\inf_{\sigma\in\Sigma}\sigma}|\xi_1-\xi_2|\\
&\qquad\quad + \left(\frac{\sup_{t\in\Upsilon}|t|}{\inf_{\sigma\in\Sigma}\sigma^2}|X_j| + \sup_{t\in\Upsilon}|t|.\sup_{\xi\in\Xi}|\xi|\right)|\sigma_1-\sigma_2|\\
&\quad\le (\alpha+\beta|X_j|)\sqrt{(t_1-t_2)^2 + (\xi_1-\xi_2)^2 + (\sigma_1-\sigma_2)^2},
\end{aligned}$$

where

$$\alpha = \sqrt{3}\left(\frac{1}{2}\pi + \frac{\sup_{t\in\Upsilon}|t|}{\inf_{\sigma\in\Sigma}\sigma} + \sup_{t\in\Upsilon}|t|.\sup_{\xi\in\Xi}|\xi|\right),\ \beta = \sqrt{3}\frac{\sup_{t\in\Upsilon}|t|}{\inf_{\sigma\in\Sigma}\sigma^2}.$$

Similarly, we have

$$\begin{aligned}
&||A(\theta_0)^{-1}\,\mathrm{Re}[b(t_1,\xi_1,\sigma_1,\theta_0)] - A(\theta_0)^{-1}\,\mathrm{Re}[b(t_2,\xi_2,\sigma_2,\theta_0)]||\\
&\le E\left[||A(\theta_0)^{-1}\nabla f(X_1,\theta)||.|\cos(t_1.Z_1(\xi_1,\sigma_1)) - \cos(t_2.Z_1(\xi_2,\sigma_2))|\right]\\
&\le E\left[||A(\theta_0)^{-1}\nabla f(X,\theta)||(\alpha+\beta|X|)\right]\sqrt{(t_1-t_2)^2 + (\xi_1-\xi_2)^2 + (\sigma_1-\sigma_2)^2}
\end{aligned}$$

Hence, condition (a) holds, with

$$K_j = \alpha + \beta|X_j| + E\left[||A(\theta_0)^{-1}\nabla f(X,\theta_0)||(\alpha + \beta|X|)\right].||\nabla f(X_j,\theta_0)||.$$

It is easy to verify from Assumptions 2.1 and 2.2 that

$$\begin{aligned} E[U_j^2 K_j] &= \alpha E[U^2] + \beta E[U^2|X|] \\ &\quad + E\left[||A(\theta_0)^{-1}\nabla f(X,\theta_0)||(\alpha + \beta|X|)\right].E\left[U^2||\nabla f(X,\theta_0)||\right] \\ &< \infty, \end{aligned}$$

so that condition (b) holds. Finally condition (c) holds almost trivially.

Thus, $\mathrm{Re}[\widetilde{W}_n(t,\xi,\sigma)]$ is tight on $\Upsilon \times \Xi \times \Sigma$, and similarly the same applies to $\mathrm{Im}[\widetilde{W}_n(t,\xi,\sigma)]$.

Now Conjecture A is true if the finite-dimensional distributions of $\widetilde{W}_n(t,\xi,\sigma) - \widetilde{W}_n(t,\xi_0,\sigma_0)$ converge to the corresponding finite-dimensional distributions of $W(t,\xi,\sigma) - W(t,\xi_0,\sigma_0)$. This is easy to prove using the standard multivariate central limit theorem.

The proof of (7.2) is now complete. The proof of (7.3) is easy and is therefore left to the reader. ■

8. A numerical example

To demonstrate the small-sample performance of the ICM test, I have drawn for $j = 1,2,3,...,500$ the random variables $X_{1,j}$, $X_{2,j}$ and U_j independently from the standard normal distribution, and set

$$Y_j = 1 + X_{1,j} + X_{2,j} + 0.5X_{1,j}.X_{2,j} + U_j.$$

The null hypothesis to be tested is that

$$H_0 : E\left[Y_j|X_{1,j},X_{2,j}\right] = \theta_1 X_{1,j} + \theta_2 X_{2,j} + \theta_3 \text{ a.s.} \tag{8.1}$$

for some parameter vector $\theta = (\theta_1,\theta_2,\theta_3)'$, which is obviously false. On the other hand, the error term $U_j^* = 0.5X_{1,j}.X_{2,j} + U_j$ of the linear regression involved satisfies $E[U_j^*] = 0$, $E[U_j^* X_{1,j}] = 0$ and $E[U_j^* X_{2,j}] = 0$, hence

$$\begin{aligned} \theta_0 &= (1,1,1)' \\ &= \arg\min_{(\theta_1,\theta_2,\theta_3)' \in \mathbb{R}^3} E\left[(Y_j - \theta_1 X_{1,j} - \theta_2 X_{2,j} - \theta_3)^2\right]. \end{aligned}$$

As to the ICM test, the integration range Υ as been chosen as in (3.2) with $\tau_1 = \tau_2 = 5$, and μ is the uniform probability measure on Υ. The explanatory variables $X_{1,j}$ and $X_{2,j}$ have been standardized and transformed according to (7.1).

The ICM test statistic (5.4) involved takes the value $\widehat{T}_n = 3.85$. Comparing this result with the upper bounds of the critical values in (5.1), i.e., $c(0.01) = 6.81$, $c(0.05) = 4.26$, $c(0.10) = 3.23$, the tentative conclusion is that H_0 is rejected at a significance level between 5% and 10%. However, these upper bounds may be conservative. Therefore I have computed the bootstrap p-value and critical values on the basis of the results in Theorem 6.4, with bootstrap sample size $M = 500$. The bootstrap critical values, comparable to (5.1), are

$$b(0.01) = 1.87,\ b(0.05) = 1.59,\ b(0.10) = 1.41, \tag{8.2}$$

for significance levels 1%, 5% and 10%, respectively, and the bootstrap p-value is virtually zero. Thus, the false null hypothesis is rejected at any conventional significance level.

Comparing (8.2) with (5.1) we see that the latter upper bounds are indeed way too conservative, although less conservative than the ones in B82. Nevertheless, if for example the ICM test static (5.4) exceeds the 1% or 5% upper bounds of the critical values one can be sure that the test statistic will also exceed the 1% or 5% bootstrap critical values, respectively, without conducting the actual bootstrap.

Moreover, note that the bootstrap critical values will differ from case to case because the ICM test is non-pivotal.

Finally, consider the case $Y_j = 1 + X_{1,j} + X_{2,j} + U_j$, for which the null hypothesis (8.1) is true. The ICM test statistic (5.4) involved now takes the value $\widehat{T}_n = 0.67$ with bootstrap p-value 0.844 and bootstrap critical values $b(0.01) = 1.96$, $b(0.05) = 1.67$ and $b(0.10) = 1.43$.

9. Concluding remarks

In this addendum to B82 I have brought the proposed ICM test (test 1) up to date by deriving its limiting null distribution, new sharper upper bound of the asymptotic critical values, and bootstrap critical values, using the knowledge I have now but lacked in 1981 when this paper was conceived. Also, I have now been able to substantiate the claim in B82 that the data-dependent standardization of the X variables in the bounded one-to-one mapping Φ does not affect the asymptotic theory in the paper. This issue has bothered me for a long time.

There are still a few loose ends, though, in particular regarding the choice of the compact set Υ. As to the latter, let me focus on the case (3.2), in the form

$$\Upsilon(\xi) = \times_{i=1}^{k}[-\xi_i, \xi_i],\ \xi = (\xi_1, \xi_2, ..., \xi_k)' \in \mathbb{R}_+^k = \times_{i=1}^{k}(0, \infty),$$

and let $\mu(.|\xi)$ be the uniform probability measure on $\Upsilon(\xi)$, so that

$$\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t|\xi) = \frac{\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t}{2^k \prod_{i=1}^{k} \xi_i}. \tag{9.1}$$

The question now is: Can we choose ξ such that $p\lim_{n\to\infty} n^{-1} \int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}\mu(t|\xi)$ is maximal under H_1 while preserving convergence in distribution of (9.1) under H_0?

The answer is yes, provided that ξ is confined to a compact or finite subset Ξ of $\mathbb{R}_+^k$. Then under H_0, $\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t \Rightarrow \int_{\Upsilon(\xi)} |W_n(t)|^2 \mathrm{d}t$ on Ξ, as can be shown similar to Bierens and Wang (2012). Hence by the continuous mapping theorem (Theorem 4.3),

$$\sup_{\xi\in\Xi} \frac{\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t}{2^k \prod_{i=1}^{k} \xi_i} \xrightarrow{\mathrm{d}} \sup_{\xi\in\Xi} \frac{\int_{\Upsilon(\xi)} |W(t)|^2 \mathrm{d}t}{2^k \prod_{i=1}^{k} \xi_i} \text{ under } H_0,$$

whereas

$$\frac{1}{n} \sup_{\xi\in\Xi} \frac{\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t}{2^k \prod_{i=1}^{k} \xi_i} \xrightarrow{\mathrm{p}} \sup_{\xi\in\Xi} \frac{\int_{\Upsilon(\xi)} \left| p\lim_{n\to\infty} \widehat{W}_n(t)/\sqrt{n} \right|^2 \mathrm{d}t}{2^k \prod_{i=1}^{k} \xi_i} > 0 \text{ under } H_1.$$

The critical values of this test can still be computed via the bootstrap procedure in section 6. Moreover, the upper bounds of the asymptotic critical values in Theorem 5.2 are still applicable to

$$\sup_{\xi\in\Xi} \frac{\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t}{\int_{\Upsilon(\xi)} \widehat{\Gamma}_n(t,t) \mathrm{d}t},$$

where $\widehat{\Gamma}_n(t_1, t_1)$ is a uniformly consistent estimator of the covariance function $\Gamma(t_1, t_1) = E[W(t_1)\overline{W(t_2)}]$ on $\overline{\Upsilon} \times \overline{\Upsilon}$, with $\overline{\Upsilon}$ a compact subset of $\mathbb{R}^k$ containing $\cup_{\xi\in\Xi}\Upsilon(\xi)$.

Except if we choose for Ξ a finite set, a practical problem with these versions of the ICM test is the computation of the suprema involved, because the functions $(2^k \prod_{i=1}^k \xi_i)^{-1} \int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t$ and $\int_{\Upsilon(\xi)} |\widehat{W}_n(t)|^2 \mathrm{d}t \ / \int_{\Upsilon(\xi)} \widehat{\Gamma}_n(t,t) \mathrm{d}t$ are highly nonlinear in ξ.

Finally, a few comments about test 2 in B82. This test grew out of frustration with my inability to derive the null distribution of the ICM test (test 1). However, in hindsight test 2 is a pretty bad idea because it is obvious from the asymptotic normality result (24) in B82 that the rate of convergence of the proposed test statistic under H_0 is lower than $\sqrt{n}$, and via the approach in Bierens and Ploberger (1996) it can be shown that this test has no power against $\sqrt{n}$ local alternatives.

References

Andrews, D. W. K., and W. Ploberger (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, reprinted in Chapter 2.

Bierens, H. J. (1984): "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353, reprinted in Chapter 3.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form", *Econometrica* 58, 1443-1458, reprinted in Chapter 4.

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

Bierens, H. J. (2004): *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests", *Econometrica* 65, 1129-1151, reprinted in Chapter 5.

Bierens, H. J., and L. Wang (2012): "Integrated Conditional Moment Tests for Parametric Conditional Distributions", *Econometric Theory* 28, 328-362, reprinted in Chapter 6.

Billingsley, P. (1968): *Convergence of Probability Measures*, Wiley & Sons.

Boning, W. B., and F. Sowell (1999): "Optimality for the Integrated Conditional Moment Test", *Econometric Theory* 15, 710-718.

Chung, K. L. (1974): *A Course in Probability Theory*, Academic Press.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter is not Identified

Under the Null Hypothesis", *Econometrica* 64, 413-430.

Jennrich, R. I. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643.

Pinkse, J. (2013): "The ET Interview: Herman Bierens", *Econometric Theor*y 29, 590-608.