Consistent Model Specification Tests*

Herman J. Bierens
Pennsylvania State University
Department of Economics and CAPCP†
University Park, PA 16802, USA

April 27, 2009

Abstract

This paper reviews the literature on tests for the correct specification of the functional form of parametric conditional expectation and conditional distribution models. In particular I will discuss various versions of the Integrated Conditional Moment (ICM) test and the ideas behind them.

1 Introduction

Quoting Hausman (1978), "Specification tests form one of the most important areas for research in econometrics", because the correct specification of a model constitutes a fundamental assumption for its estimation and inference.

The Hausman (1978) specification test is based on the difference between an efficient estimator under the null and a non-efficient estimator. White (1981) utilized a specification-robust estimator of nonlinear regression models to test conditional mean specifications. White (1982) compared two different

---

[*] This paper has been prepared for a keynote address given at the *Primer Congreso Científico Internacional en Economía y Finanzas,* May 18-22, 2009, in Quito, Ecuador.

[†] Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation

expressions of the Fisher information matrix, which should be equal if the conditional distribution is correctly specified.

Newey (1985) proposed the conditional moment (CM) test and suggested that both the Hausman (1978) and the White (1981) methods can be viewed as special cases of a CM test. The idea behind the CM test is that a correct model specification implies that certain conditional moments are zero, which can be converted to unconditional moment restrictions by multiplying these conditional moments by instrumental variables. The sample counterparts of these unconditional moments are the basis for a CM test.

These tests are not consistent because they only employ a finite number of moment restrictions implied by the model. Bierens (1982) and Holly (1983) observed this inconsistency for the Hausman test. In general, for tests based on a finite number of moment restrictions one can always construct a data-generating process for which the null hypothesis is false but the moment restrictions involved hold.

Given a parametric specification $f\left(x,\theta\right),\ \theta \in \Theta$, of a conditional expectation $E[Y|X=x]$, where Y is the dependent variable, $X \in \mathbb{R}^k$ is a vector of explanatory variables and $\Theta \subset \mathbb{R}^m$ is a compact parameter space, the corresponding nonlinear regression model $Y = f\left(X,\theta_0\right) + U$ is correctly specified if

$$H_0 : \exists\theta_0 \in \Theta : \Pr\left(f\left(X,\theta_0\right) = E[Y|X]\right) = 1 \tag{1}$$

A test for this null hypothesis is consistent if it has asymptotic power one against the general alternative that the null hypothesis is false:

$$H_1 : \forall\theta \in \Theta : \Pr\left(f\left(X,\theta\right) = E[Y|X]\right) < 1 \tag{2}$$

The first consistent specification test for nonlinear regression models was proposed by me, in Bierens (1982). This test is also a conditional moment tests, but based on uncountable many moments conditions of the form

$$E[(Y - f\left(X,\theta_0\right)).w(\xi'\Phi(X))] = 0, \tag{3}$$

where Φ is a bounded one-to-one mapping and $w(u) = \exp\left(i.u\right)$ is the complex exponential function. Given a random sample $\{(Y_j, X_j)\}_{j=1}^{n}$ from (Y, X) and the NLLS estimator $\widehat{\theta}_n$ for θ_0, the empirical counter-part of the expectation in (3), times $\sqrt{n}$, is

$$\widehat{Z}_n(\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widehat{U}_j w(\xi'\Phi(X_j)), \tag{4}$$

where $\widehat{U}_j = Y_j - f(X_j, \widehat{\theta}_n)$. Therefore, I proposed to test the validity of the nonlinear regression model via the ICM test statistic

$$\widehat{T}_n = \int_{\Xi} \left| \widehat{Z}_n(\xi) \right|^2 d\mu(\xi), \tag{5}$$

where Ξ is a closed and bounded hypercube around the origin of $\mathbb{R}^k$ and $\mu(.)$ is the uniform probability measure on Ξ. I proved that under the alternative H_1, $\widehat{T}_n/n$ converges in probability to a positive constant, whereas under H_0, $\widehat{T}_n \xrightarrow{d} T$ for some random variable T. However, at that time I was not able to establish the type of the limiting distribution T, but only that its first moment is of the form $E\left[T\right] = E\left[g\left(X, \theta_0\right)\right]$ with $g\left(x, \theta\right)$ a known function on $\mathbb{R}^k \times \Theta$. Therefore, I proposed to derive upper bounds of the critical values based on Chebyshev's inequality for first moments. Two years later I generalized these results to nonlinear time series regressions. See Bierens (1984).

Eight years after my paper Bierens (1982) I had finally figured out what the limiting null distribution of the ICM test looks like. In Bierens (1990) I proved for the case $w(u) = \exp(u)$ that under the null hypothesis (1), $\widehat{Z}_n(\xi) \Rightarrow Z(\xi)$, where $Z(\xi)$ is a zero-mean Gaussian process on Ξ, so that the limiting null distribution T is of the form

$$T = \int_{\Xi} \left| Z(\xi) \right|^2 d\mu(\xi). \tag{6}$$

Building further on these results, Bierens and Ploberger (1997) (henceforth BP) showed for a general class of real valued weight functions $w(z)$ that the asymptotic null distribution takes the form $T = \sum_{j=1}^{\infty} \lambda_j \varepsilon_j^2$, where the ε_j's are independent $N(0,1)$ distributed and the λ_j's are the decreasingly ordered eigenvalues of the covariance function $\Gamma(\xi_1, \xi_2) = E[Z(\xi_1)Z(\xi_2)]$. Moreover, BP also proved that the ICM test has non-trivial $\sqrt{n}$ local power and is admissible. Very recently the local power properties of the ICM test has been analyzed further by Escanciano (2009).

As to the choice of the weight function $w(z)$, BP showed that the only requirement is that $w(u)$ is a power function in an open neighborhood of $u = 0$ such that $(d/du)^s w(u) \mid_{u=0} \neq 0$ for all but a finite number of natural numbers s, which is the case for $w\left(u\right) = \exp\left(i.u\right)$ used in Bierens (1982, 1984), $w\left(u\right) = \exp\left(u\right)$ used in Bierens (1990) and, for example, $w\left(u\right) = \cos(u) + \sin(u)$. Stinchcombe and White (1998) have shown that the ICM test is consistent for a wide range of non-polynomial analytical weight functions $w(u)$, for which they cast the name "totally revealing".

The choice of the probability measure μ in (5) not critical either. Any probability measure μ that is absolutely continuous with respect to the Lebesgue measure will do. However, Boning and Sowell (1999) showed that the ICM test is the best ICM test according to the weighted average power criterion considered by Andrews and Ploberger (1994) if μ is the uniform probability measure on Ξ.

Up to the early nineties, my papers Bierens (1982, 1984, 1990) were the only literature on consistent model specification testing. In the mid-nineties a related strand of statistics literature emerged, starting with Stute (1997). Stute's approach is based on the fact that $\Pr\left(E[U|X]=0\right)<1$ is equivalent to $E\left[U.I(X\leq x)\right]\neq 0$ for some conformable vector x, where here and in the sequel $I(.)$ denotes the indicator function. Therefore, various consistent test can be based on the empirical process $I_n(x)=n^{-1/2}\sum_{j=1}^{n}\widehat{U}_jI(X_j\leq x)$, including an ICM type test.

Escanciano (2006) observes that $\Pr\left(E[U|X]=0\right)=1$ is equivalent to $E\left[U.I\left(\beta'X\leq u\right)\right]=0$ for all $\beta\in\mathbb{R}^k$, normalized to unit length $\beta'\beta=1$, and all $u\in\mathbb{R}$, which in its turn is equivalent to $\int\int_{-\infty}^{\infty}\left(E\left[U.I\left(\beta'X\leq u\right)\right]\right)^2$ $dF(u|\beta)d\mu(\beta)=0$, where $F(u|\beta)$ is the distribution function of $\beta'X$ and μ is the uniform probability measure on the hyperball $\beta'\beta=1$. Therefore, Escanciano (2006) proposes a combined Bierens-Stute ICM test based on the empirical version of the latter integral, $\frac{1}{n}\sum_{j=1}^{n}\int\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{U}_i.I\left(\beta'X_i\leq\beta'X_j\right)\right)^2$ $d\mu(\beta)$.

Another strand of literature on model specification testing emerged in the nineties in the econometrics literature. These tests are based on comparisons in various ways of parametric functional forms and corresponding nonparametric or semi-nonparametric estimates. See, for example, Härdle and Mammen (1993), Gozalo (1993), Horowitz and Härdle (1994), Hong and White (1995), Li and Wang (1998), Zheng (1996), and Lavergne and Vuong (2000), among others. In particular, Härdle and Mammen's (1993) test is the integrated squared difference between the parametric fit and the nonparametric fit of a regression model. Moreover, the similar but independently developed tests of Li-Wang (1998) and Zheng (1996) are based on the trivial equality $E\left[(E[U|X])^2h(X)\right]=E\left[E\left(U.E[U|X]\right)h(X)\right]$, where $h\left(x\right)$ is the density of X. Replacing $E[U|X=x]$ by the corresponding kernel regression estimator $\widehat{\mu}\left(x\right)$ on the basis of the estimated residuals $\widehat{U}_j$ and $h\left(x\right)$ by a kernel density estimator $\widehat{h}\,(x)$, the Li-Wang (1998) and Zheng (1996) test statistics take the form $V_n = \frac{1}{n}\sum_{j=1}^{n}\widehat{U}_j.\widehat{\mu}(X_j)\widehat{h}(X_j)$. However, these tests have only non-trivial power against local alternatives that approach the null at a slower rate than $1/\sqrt{n}$, due to the slower rate of convergence than $\sqrt{n}$ of nonparametric kernel estimators.

Although at first sight the ICM test and the nonparametric kernel regression based tests for nonlinear regression models seem fundamentally different, Fan and Li (2000) have shown that the ICM test can be viewed as a special case of a the kernel-based test but with a fixed bandwidth.

The literature on consistent specification testing of conditional distribution models is rather limited. Zheng (2000) proposed a test for the validity of conditional densities by comparing a parametric conditional density with a corresponding nonparametric kernel estimator via the Kullback-Leibler (1951) information criterion. Thus, this test is only applicable to absolutely continuous conditional distribution models. Zheng (2000) test has non-trivial local power, but only against local alternatives that approach the null at a slower rate than $1/\sqrt{n}$. Andrews (1988) extended the Pearson's Chi-square test to a test for parametric conditional distributions. This test is based on partitioning the dependent and explanatory variables in cells, and then comparing the frequencies involved with the frequencies implied by the model. However, is unknown what the best way is to choose these cells. See Justel et. al. (1997). Andrews (1997) generalized the Kolmogorov test for testing unconditional distribution to a Conditional Kolmogorov (CK) test for testing general conditional distributions. In particular, the CK test statistic takes the form

$$\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(I(Y_j\leq Y_i)-F(Y_i|X_j,\widehat{\theta})\right)I(X_j\leq X_i)\right| \tag{7}$$

where $F(y|X_j,\widehat{\theta})$ is the estimated conditional distribution model. Since the asymptotic null distribution is case-dependent, a bootstrap method is used to derive critical values. This test is consistent, and has non-trivial power against $\sqrt{n}$ local alternatives. However, a practical problem with the CK test is that if the dimension of X_j is large the inequality $X_j < X_i$ for $i \neq j$ may never happen, even for quite a large sample size n.

Bai (2003) proposes a test for the validity of absolutely continuous conditional distribution models based on the well-known fact that plugging in an absolutely continuous distributed random variable in its conditional distribution function yields an uniformly $[0,1]$ distributed random variable. Although Bai's test aims to test conditional time series distribution models, it applies to cross-section models as well. In particular, with $F(y|X_j,\widehat{\theta})$ the estimated conditional distribution of Y_j given X_j, Bai's test is based on an empirical process of the form

$$\widehat{V}(u) = n^{-1/2}\sum_{j=1}^{n}\left(I\left(F(Y_j|X_j,\widehat{\theta}) \leq u\right) - u\right),\ u \in [0,1].$$

However, Bai's test is not consistent. Admittedly, Bai (2003) did not claim consistency, but only that his test has non-trivial power against $\sqrt{n}$ local alternatives. This demonstrates that non-trivial local power does not imply consistency against fixed alternatives.

Recently, Bierens and Wang (2009) have extended the ICM test to a consistent tests for parametric conditional distribution specifications.. This test is formed on the basis of the integrated squared difference between the empirical characteristic function of the actual data and the characteristic function implied by the model. This test is consistent, and has nontrivial power against $\sqrt{n}$ -local alternatives. To avoid numerical evaluation of the conditional characteristic function of the model distribution, a simulated integrated conditional moment (SICM) test is proposed, where each theoretical conditional characteristic function is replaced by a simulated counterpart, based on a single random drawing from the corresponding conditional distribution.

2 The asymptotic null distribution of the ICM statistic

Under the null hypothesis and standard regularity conditions, the NLLS estimator $\widehat{\theta}_n$ satisfies

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = A^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left.\frac{\partial f(X_j,\theta)}{\partial\theta'}\right|_{\theta=\theta_0} U_j + o_p(1)$$

where $U_j = Y_j - f(X_j,\theta_0) = Y_j - E[Y_j|X_j]$ and

$$A = p\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial f(X_j,\theta)}{\partial\theta'}\right)\left.\left(\frac{\partial f(X_j,\theta)}{\partial\theta'}\right)'\right|_{\theta=\theta_0}.$$

Using this standard result, it is straightforward to verify (see Bierens 1982) that,

$$\widehat{Z}_n(\xi) = \widetilde{Z}_n(\xi) + o_p(1),$$

uniformly on Ξ, where

$$\widetilde{Z}_n(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \phi_j(\xi)$$

with

$$\begin{aligned} \phi_j(\xi) &= w\left(\xi' \Phi(x_j)\right) - b(\xi)' A^{-1} \left. \frac{\partial f(X_j, \theta)}{\partial \theta'} \right|_{\theta = \theta_0}, \\ b(\xi) &= p \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \left. \frac{\partial f(X_j, \theta)}{\partial \theta'} \right|_{\theta = \theta_0} w\left(\xi' \Phi(X_j)\right). \end{aligned}$$

It has been shown by Bierens (1990) that under some mild regularity conditions (among which the assumption that the function $w(.)$ is real-valued), and the null hypothesis (2), $\widetilde{Z}_n(\xi) \Rightarrow Z(\xi)$ on Ξ,[1] where $Z(\xi)$ is a zero-mean Gaussian process on Ξ, with covariance function

$$\Gamma(\xi_1, \xi_2) = E\left[Z(\xi_1) Z(\xi_2)\right] = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} E\left[U_j^2 \phi_j(\xi_1) \phi_j(\xi_2)\right] \tag{8}$$

Consequently, under the null hypothesis (2),

$$\widehat{T}_n = \int_{\Xi} \left| \widehat{Z}_n(\xi) \right|^2 d\mu(\xi) \xrightarrow{d} T = \int_{\Xi} \left| Z(\xi) \right|^2 d\mu(\xi).$$

Moreover, as said before, BP have shown that the asymptotic null distribution of the ICM statistic is of the type

$$T = \int_{\Xi} \left| Z(\xi) \right|^2 d\mu(\xi) = \sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2, \tag{9}$$

where the ε_i's are i.i.d. $N(0,1)$ and the λ_i's are the eigenvalues of the covariance function Γ. In particular, the λ_i's are the solutions of the eigenvalue problem

$$\lambda . \psi(\xi_1) = \int_{\Xi} \Gamma(\xi_1, \xi_2) \psi(\xi_2) d\mu(\xi_2)$$

---

[1] See for example Billingsley (1968) for the meaning of weak convergence, denoted by "$\Rightarrow$".

with corresponding eigenfunctions $\psi_i(\xi)$ satisfying $\int_\Xi \psi_i(\xi)\psi_j(\xi)d\mu(\xi) = I(i = j)$. A related result is Mercer's theorem, which states that

$$\Gamma(\xi_1, \xi_2) = \sum_{i=1}^{\infty} \lambda_i \psi_i(\xi_1)\psi_i(\xi_2)$$

Consequently,

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

As to the source of the ε_i's, it is possible to write the random function $Z(\xi)$ as

$$Z(\xi) = \sum_{i=1}^{\infty} \gamma_i \psi_i(\xi),$$

where $\gamma_i = \int_\Xi Z(\xi)\psi_i(\xi)d\mu(\xi)$, hence

$$\int_\Xi |Z(\xi)|^2 \, d\mu(\xi) = \sum_{i=1}^{\infty} \gamma_i^2.$$

Since $Z(\xi)$ is zero-mean Gaussian, the γ_i's are zero-mean Gaussian as well, with covariance

$$\begin{aligned} E\left[\gamma_i \gamma_j\right] &= \int_\Xi \int_\Xi E\left[Z(\xi_1)Z(\xi_2)\right] \psi_i(\xi_1)\psi_j(\xi_2) d\mu(\xi_1) d\mu(\xi_2) \\ &= \int_\Xi \left(\int_\Xi \Gamma(\xi_1, \xi_2)\psi_j(\xi_2)d\mu(\xi_2) \right) \psi_i(\xi_1)d\mu(\xi_1) \\ &= \lambda_j \int_\Xi \psi_i(\xi_j)\psi_i(\xi_1)d\mu(\xi_1) \\ &= \lambda_j I(i = j). \end{aligned}$$

Hence, $\varepsilon_i = \gamma_i/\sqrt{\lambda_i} \sim N(0, 1)$, and $E[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$. Since the ε_i's are zero-mean Gaussian and uncorrelated, they are independent.

The eigenvalues λ_i depend on the covariance function $\Gamma(\xi_1, \xi_2)$, which in its turn depends on the distribution of (Y, X) and the conditional expectation model $f(X, \theta_0)$. Therefore, the limiting null distribution T is case-dependent. However, with the eigenvalues λ_j sorted in decreasing order, BP proved that

$$\frac{\sum_{j=1}^{\infty} \lambda_j \varepsilon_j^2}{\sum_{j=1}^{\infty} \lambda_j} \leq \sup_{m \geq 1} \frac{1}{m} \sum_{j=1}^{m} \varepsilon_j^2 = \overline{T}. \tag{10}$$

Therefore, BP proposed to use

$$\widetilde{T}_n = \frac{\int_\Xi \left|\widehat{Z}_n(\xi)\right|^2 d\mu(\xi)}{\int_\Xi \widehat{\Gamma}_n(\xi,\xi)d\mu(\xi)} \tag{11}$$

as the actual ICM test statistic, where $\widehat{\Gamma}_n(\xi,\xi)$ is a consistent estimator of the variance function $\Gamma(\xi,\xi)$, and to use upper bounds of the critical values derived from $\overline{T}$. As to the latter, $\Pr(\overline{T} \geq 4.26) = 0.05$ and $\Pr(\overline{T} \geq 3.23) = 0.1$, so that the values 4.26 and 3.23 can be used as upper bounds of the 5% and 10% critical values, respectively.

3 Consistency of the ICM test

The consistency of the ICM test hinges on the following fundamental theorem:

Theorem 1: *Let U be a random variable satisfying $E[|U|] < \infty$, and $P[E(U|X) = 0] < 1$, where $X \in \mathbb{R}^k$ is a bounded random vector.*
(a) *Let $w(u)$ be a complex or real valued function that is infinitely many times differentiable in $u = 0$ and satisfies the condition that $(d/du)^s w(u) \mid_{u=0} \neq 0$ for all but a finite number of natural numbers s. Then for every $\varepsilon > 0$ there exists a $\xi \in \mathbb{R}^k$ such that $E[u.w(\xi' X)] \neq 0$ and $\|\xi\| < \varepsilon$.*
(b) *If in addition $w(u)$ is a power series in an open neighborhood of $u = 0$, i.e., for some $\delta > 0$, $w(u) = \sum_{s=0}^{\infty} (\gamma_s/s!)\, u^s$ for $|u| < \delta$, where $\gamma_s = (d/du)^s w(u) \mid_{u=0}$, then the set*

$$\left\{\xi \in \mathbb{R}^k : E[U.w(\xi' X)] = 0\right\}$$

has Lebesgue measure zero and is nowhere dense.

Proof: See Bierens (1982) for part (a) with $w(u) = \exp(i.u)$, Bierens (1990) for the case $w(u) = \exp(u)$, and Theorem 1 in BP for the general case (b). See also Stinchcombe and White (1998) for further elaborations on this theorem, and Bierens (1994, Ch. 3) for the details of the proof of Theorem 1 for the cases $w(u) = \exp(i.u)$ and $w(u) = \exp(u)$.

The condition that the random vector X is bounded can be get rid of by replacing X by $\Phi(X)$, where Φ is a Borel measurable bounded one-to-one mapping, because the σ-algebra generated by X is then the same as the

σ-algebra generated by $\Phi(X)$, hence conditioning on $\Phi(X)$ is equivalent to conditioning on X. Therefore, denoting

$$\begin{aligned}\phi(\theta,\xi) &= E\left[(Y-f(X,\theta))w(\xi'\Phi(X))\right]\\ \Upsilon\left(\theta\right) &= \int_{\Xi}\left|\phi(\theta,\xi)\right|^2 d\mu(\xi)\end{aligned}$$

where the probability measure μ is absolutely continuous with respect to the Lebesgue measure, it follows from Theorem 1 that the null hypothesis (1) is equivalent to

$$H_0 : \exists\theta_0 \in \Theta : \Upsilon\left(\theta_0\right) = 0$$

whereas the alternative hypothesis (2) is equivalent to

$$H_1 : \forall\theta \in \Theta : \Upsilon\left(\theta\right) > 0$$

The empirical counter-parts of $\phi(\theta,\xi)$ and $\Upsilon\left(\theta\right)$ are

$$\begin{aligned}\widehat{\phi}_n(\theta,\xi) &= \frac{1}{n}\sum_{j=1}^{n}(Y_j-f(X_j,\theta))w(\xi'\Phi(X_j)\\ \widehat{\Upsilon}_n\left(\theta\right) &= \int_{\Xi}\left|\widehat{\phi}_n(\theta,\xi)\right|^2 d\mu(\xi).\end{aligned}$$

Under standard regularity condition it follows from Jennrich's (1969) uniform strong law of large numbers that $p\lim_{n\to\infty}\sup_{(\theta,\xi)\in\Theta\times\Xi}\left|\widehat{\phi}_n(\theta,\xi)-\phi(\theta,\xi)\right| = 0$, hence

$$p\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\widehat{\Upsilon}_n\left(\theta\right)-\Upsilon\left(\theta\right)\right| = 0. \tag{12}$$

Next, let

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

which is the NLLS estimator of θ_0. Under standard conditions,

$$p\lim_{n\to\infty}\widehat{\theta}_n = \theta_* \in \Theta \tag{13}$$

regardless wether the model is true or not. Of course, under the null hypothesis (2), $\theta_* = \theta_0$. Moreover, note that $\widehat{T}_n/n = \widehat{\Upsilon}_n(\widehat{\theta}_n)$. It follows therefore straightforwardly from (12) and (13) that $p\lim_{n\to\infty}\widehat{T}_n/n = \Upsilon\left(\theta_*\right)$, hence

$$p\lim_{n\to\infty}\widehat{T}_n/n = \Upsilon\left(\theta_0\right) = 0$$

under H_0 and

$$p\lim_{n\to\infty}\widehat{T}_n/n = \Upsilon\left(\theta_*\right) > 0$$

under H_1. The latter establishes the consistency of the ICM test.

4 The ICM test of the martingale difference hypothesis

The asymptotic theory of the BP version of the ICM test has been developed for time series data. However, if applied to time series regression models this version of the ICM test is no longer consistent. The reason is that time series regression model aim to represent conditional expectations relative to all lagged dependent variables and possibly other current and all lagged exogenous variables. For example, consider the ARX(1) model

$$Y_t = \alpha + \beta Y_{t-1} + \gamma' X_t + U_t, \tag{14}$$

where $X_t \in \mathbb{R}^{k-1}$ is a vector of exogenous variables. Denote

$$V_t = (Y_t, X_{t+1}')' \in \mathbb{R} \times \mathbb{R}^{k-1}. \tag{15}$$

The ARX(1) model is correctly specified if and only if

$$E\left[Y_t|V_{t-1}, V_{t-2}, V_{t-3}, ...\right] = \alpha + \beta Y_{t-1} + \gamma' X_t.$$

This condition is equivalent to the condition that the error process U_t is a martingale difference process with respect to the σ-algebra

$$\mathcal{F}_{-\infty}^{t} = \sigma\left(V_t, V_{t-1}, V_{t-2}, ...\right)$$

generated by the sequence V_{t-j}, $j \geq 0$, i.e., $E[|U_t|] < \infty$, U_t is measurable $\mathcal{F}_{-\infty}^{t}$, and $E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0$ a.s.

The BP version of the ICM test only allows to test the null hypothesis $E\left[U_t|\mathcal{F}_{t-m}^{t-1}\right] = 0$ for some finite $m \geq 1$, where $\mathcal{F}_{t-m}^{t-1} = \sigma\left(\{V_{t-j}\}_{j=1}^{m}\right)$ is the σ-algebra generated by $V_{t-1}, ..., V_{t-m}$. In particular, the empirical process (4) now becomes

$$\begin{aligned} \widehat{Z}_{m,n}(\xi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{U}_t.w\left(\sum_{j=1}^{m}\xi_j'\Phi(V_{t-j})\right), \\ \xi &= (\xi_1', ..., \xi_m')' \in \mathbb{R}^{k.m} \end{aligned}$$

where the $\widehat{U}_t$'s are the model residuals, with corresponding the ICM statistic

$$\widehat{T}_{m,n} = \int_{\Xi_{k.m}} \left|\widehat{Z}_{m,n}(\xi)\right|^2 d\mu_m(\xi),$$

where $\Xi_{k.m}$ is a closed and bounded hypercube around the origin of $\mathbb{R}^{k.m}$ and $\mu_m(.)$ is the uniform probability measure on $\Xi_{k.m}$.

Bierens (1984) and De Jong (1996) have, in different ways, extended the ICM test to a consistent test of the martingale difference hypothesis. Bierens (1984) proposed to use a test statistic of the form $\sum_{m=1}^{L_n} p_m \widehat{T}_{m,n}$, where L_n is a subsequence of n and p_m is a positive-valued sequence of weights satisfying $\sum_{m=1}^{\infty} p_m = 1$. The consistency of this version of the ICM test follows from the fact that if $\Pr\left(E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0\right) < 1$ then there exists a finite $m \geq 1$ such that $\Pr\left(E\left[U_t|\mathcal{F}_{t-m}^{t-1}\right] = 0\right) < 1$.

De Jong's (1996) approach is the following. For two infinite sequences of points in $\mathbb{R}^k \times \mathbb{R}^\infty$, ξ and ζ, given by $\xi = (\xi_1', \xi_2', ...)'$ and $\zeta = (\zeta_1', \zeta_2', ...)'$, where $\xi_j, \zeta_j \in \mathbb{R}^k$, define the norm

$$\|\xi - \zeta\| = \sqrt{\sum_{j=1}^{\infty} j^2 |\xi_j - \zeta_j|^2}, \tag{16}$$

where $|\xi_j - \zeta_j|$ is the Euclidean norm on $\mathbb{R}^k$. Next, define the space Ξ as

$$\Xi = \{\xi \in \mathbb{R}^{k+1} \times \mathbb{R}^\infty : |\xi_j| \leq c.j^{-2}, \ \forall j \geq 1\}, \tag{17}$$

for some constant $c > 0$. Note that Ξ has finite Lebesgue measure. With this definition $(\Xi, \|\cdot\|)$ is a compact metric space.

Following Bierens (1990), De Jong now proposes to use the weight function

$$w_t(\xi) = \exp\left(\sum_{j=1}^{t-1} \xi_j' \Phi(V_{t-j})\right),$$

where again Φ is a bounded one-to-one mapping, and μ is the uniform probability measure on Ξ. However, in view of Theorem 1, De Jong's results carry over to the more general case

$$w_t(\xi) = w\left(\sum_{j=1}^{t-1} \xi_j' \Phi(V_{t-j})\right), \tag{18}$$

where $w(\cdot)$ is a real-valued function satisfying the conditions of Theorem 1. The consistency of De Jong's ICM test now follows from the following generalization of Theorem 1.

Theorem 2: *Let* U_t *be a stationary process satisfying* $E[|U_t|] < \infty$, *and let* V_t *be a* k*-variate time series process such that* (U_t, V_t) *is stationary. Let* $(\Xi, \|\cdot\|)$ *be defined by* (16) *and* (17), *and let*

$$\overline{w}_t(\xi) = w\left(\sum_{j=1}^{\infty} \xi_j' \Phi(V_{t-j})\right), \tag{19}$$

where $w(\cdot)$ *satisfies the conditions of Theorem 1. Then*

$$P\left[E(U_t|V_{t-1}, V_{t-2}, V_{t-3}, ...) = 0\right] < 1$$

if and only if the set

$$\{\xi \in \Xi : \ E\left(u_t \overline{w}_t(\xi)\right) = 0\}$$

has Lebesgue measure zero and is nowhere dense in Ξ.

Proof: De Jong (1997).

The actual test statistic is now similar to (11), and the upper bounds of the critical values still apply.

5 Testing an ARCH specification

Consider a standard ARCH(1) process

$$Y_t = \mu + \varepsilon_t \sqrt{\alpha + \beta_1 \left(Y_{t-1} - \mu\right)^2}$$

for financial time series, in particular stock returns, where ε_t is an independent process with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = 1$, hence $E[Y_t] = \mu$, and $\alpha > 0$, $\beta_1 \in (0, 1)$. This model can be rewritten as

$$Y_t^2 = \beta_0 + \beta_1 Y_{t-1}^2 - \beta_2 Y_{t-1} + U_t, \tag{20}$$

where $\beta_0 = (1 - \beta_1)\mu^2 + \alpha > 0$, $\beta_1 \in (0, 1)$, $\beta_2 = 2\mu\beta_1 > 0$, and

$$U_t = (\varepsilon_t - 1)\left(\alpha + \beta\left(Y_{t-1} - \mu\right)^2\right) + 2\varepsilon_t \mu \sqrt{\alpha + \beta\left(Y_{t-1} - \mu\right)^2}.$$

We may interpret $\beta_0 + \beta_1 Y_{t-1}^2 - \beta_2 Y_{t-1}$ as the linear projection of Y_t^2 on $(1, Y_{t-1}^2, Y_{t-1})$. Clearly, under the null hypothesis that Y_t is an ARCH(1) process, U_t is a martingale difference process: $E\left[U_t | \mathcal{F}_{-\infty}^{t-1}\right] = 0$, where $\mathcal{F}_{-\infty}^{t-1} = \sigma\left(\{Y_{t-j}\}_{j=1}^{\infty}\right)$.

Now suppose that in reality, $Y_t = \mu + \varepsilon_t h_{t-1}$, where $h_{t-1} > 0$ is measurable w.r.t. $\mathcal{F}_{-\infty}^{t-1}$, and that for all μ, $\alpha > 0$ and $\beta_1 \in (0,1)$,

$$\Pr\left[h_{t-1}^2 = \alpha + \beta_1 (Y_{t-1} - \mu)^2\right] < 1.$$

Then it is easy to verify from $E\left[Y_t^2 | \mathcal{F}_{-\infty}^{t-1}\right] = \mu^2 + h_{t-1}^2$ that then

$$\Pr\left(E\left[U_t | \mathcal{F}_{-\infty}^{t-1}\right] = 0\right) < 1.$$

Therefore, the ARCH(1) hypothesis can be tested by testing the correctness of the regression model (20), using the ICM test, and similarly for the ARCH(p) hypothesis.

6 The ICM test for conditional distribution specifications

Often econometric models take the form of a conditional density, distribution or probability. For example, if the dependent variable Y is a count starting from zero, with no upper bound, a popular model for the conditional distribution of Y given a vector X of covariates is the conditional Poisson model

$$\Pr[Y = m | X] = \exp\left(-\exp(\alpha + \beta' X)\right) \frac{\exp\left(m(\alpha + \beta' X)\right)}{m!}.$$

More generally, consider a parametric specification $F(y|X, \theta)$, $\theta \in \Theta$, where again Θ is the parameter space, of a conditional distribution $\Pr[Y \leq y | X]$, $Y \in \mathbb{R}^m$, $X \in \mathbb{R}^k$.

To test the correctness of such a model, Bierens and Wang (2009) propose a Simulated ICM (SICM) test, based on the well-known fact that distributions are equal if and only if their characteristic functions are identical, and if the random variables involved are bounded then their joint distribution is completely determined by the shape of its characteristic function in an arbitrary open neighborhood of the zero vector. Thus, let

$$\begin{aligned} \varphi(\tau, \xi) &= E\left[\exp\left(i.(\tau' \Phi_1(Y) + \xi' \Phi_2(X))\right)\right], \\ i &= \sqrt{-1},\ \tau \in \mathbb{R}^m,\ \xi \in \mathbb{R}^k, \end{aligned}$$

be the characteristic function of $(\Phi_1(Y), \Phi_2(X))$, where Φ_1 and Φ_2 are bounded one-to-one mappings on $\mathbb{R}^m$ and $\mathbb{R}^k$, respectively. Similarly, let

$$\begin{aligned} \psi(\tau, \xi|\theta) &= E\left[\exp\left(i.(\tau'\Phi_1(\widetilde{Y}(\theta)) + \xi'\Phi_2(X))\right)\right] \\ &= E\left[\exp\left(i.\xi'\Phi_2(X)\right)\int \exp\left(i.\tau'\Phi_1(y)\right) dF(y|X,\theta)\right] \end{aligned}$$

where $\widetilde{Y}(\theta)$ is a random drawing from the conditional distribution $F(y|X, \theta)$. If for some $\theta_0 \in \Theta$, $\psi(\tau, \xi|\theta_0) = \varphi(\tau, \xi)$ in an open neighborhood of the origin of $\mathbb{R}^m \times \mathbb{R}^k$ then the joint distribution of $\left(\Phi_1(\widetilde{Y}(\theta_0)), \Phi_2(X)\right)$ is equal to the joint distribution of $(\Phi_1(Y), \Phi_2(X))$, which in its turn implies that the joint distribution of $\left(\widetilde{Y}(\theta_0), X\right)$ is equal to the joint distribution of (Y, X). Therefore, the null hypothesis

$H_0 : \Pr\left(\sup_{y\in\mathbb{R}^m} \left|\Pr\left[Y \le y|X\right] - F(y|X, \theta_0)\right| = 0\right) = 1$ for some $\theta_0 \in \Theta$,
is equivalent to $\psi(\tau, \xi|\theta_0) = \varphi(\tau, \xi)$ on $(-c, c)^{m+k}$, where $c > 0$ is arbitrary, and the general alternative that H_0 is false:

$H_1 : \Pr\left(\sup_{y\in\mathbb{R}^m} \left|\Pr\left[Y \le y|X\right] - F(y|X, \theta)\right| = 0\right) < 1$ for all $\theta \in \Theta$,
is equivalent to the statement that for all $\theta \in \Theta$, the set

$$\left\{(\tau, \xi) \in (-c, c)^{m+k} : \psi(\tau, \xi|\theta_0) \ne \varphi(\tau, \xi)\right\}$$

has positive Lebesgue measure.

Given a random sample $\{(Y_j, X_j)\}_{j=1}^n$ from (Y, X) and random drawings $\widetilde{Y}_j$ from $F(y|X_j, \widehat{\theta})$, $j = 1, ..., n$, where $\widehat{\theta}$ is the ML estimator of θ_0, Bierens and Wang (2009) propose the SICM test

$$\widetilde{T}_n\left(c\right) = \frac{1}{\left(2c\right)^{m+k}} \int_{[-c,c]^k} \int_{[-c,c]^m} \left|\widetilde{Z}_n^{(s)}(\tau, \xi)\right|^2 d\tau d\xi$$

where

$$\begin{aligned} \widetilde{Z}_n^{(s)}(\tau, \xi) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\exp\left(i.\tau'\Phi_1(Y_j)\right) - \exp\left(i.\tau'\Phi_1(\widetilde{Y}_j)\right)\right) \\ &\quad \times \exp\left(i.\xi'\Phi_2(X_j)\right) \end{aligned}$$

The null distribution of the SICM test is similar to the ICM test in BP, i.e.,

$$\widetilde{Z}_n^{(s)}(\tau, \xi) \Rightarrow Z(\tau, \xi) \quad \text{on} \quad [-c, c]^{k+m}$$

where Z is a complex-valued zero mean Gaussian process, so that

$$\widetilde{T}_n(c) \xrightarrow{d} T(c) = \frac{1}{(2c)^{m+k}} \int_{[-c,c]^k} \int_{[-c,c]^m} |Z(\tau,\xi)|^2 \, d\tau d\xi.$$

The critical values involved have to be derived via a bootstrap method, as upper bounds of the critical values are no longer available. The SICM test is consistent, and has nontrivial $\sqrt{n}$ local power.

As to the choice of the constant c, it is shown that under H_0, on any bounded interval $[\underline{c}, \overline{c}]$ with $\underline{c} > 0$, $\widetilde{T}_n(c) \Rightarrow T(c)$, so that $\max_{\underline{c} \leq c \leq \overline{c}} \widetilde{T}_n(c) \xrightarrow{d} \max_{\underline{c} \leq c \leq \overline{c}} T(c)$. In practice this maximum can be computed via grid search. The resulting test is called the MAXSICM test.

References

Andrews, D.W., 1988, Chi-Square Diagnostic Tests for Econometric Models: Theory, *Econometrica*, 56, 1419-1453.

Andrews, D.W., 1997, A Conditional Kolmogorov Test, *Econometrica*, 65, 1097-1128.

Andrews, D.W., & W. Ploberger, 1994, Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative, *Econometrica*, 62, 1383-1414.

Bai, J., 2003, Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics*, 85, 531-549.

Bierens, H.J., 1982, Consistent Model Specification Tests, *Journal of Econometrics*, 20, 105-134.

Bierens, H.J., 1984, Model Specification Testing of Time Series Regressions, *Journal of Econometrics,* 26, 323-353.

Bierens, H.J.,1990, A Consistent Conditional Moment Test of Functional Form, *Econometrica*, 58, 1443-1458.

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. & W. Ploberger, 1997, Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica*, 65, 1129-1151.

Bierens, H.J. & L. Wang, 2009, Integrated Conditional Moment Tests for Parametric Conditional Distributions, Working paper (http://econ.la.psu.edu/~hbierens/ICM_IID.PDF)

Billingsley, P., 1968, *Convergence of Probability Measures.* New York: John Wiley.

Boning, B. & F. Sowell, 1999, Optimality for the Integrated Conditional Moment Test, *Econometric Theory*, 15, 710-719.

De Jong, R., 1996, The Bierens Test Under Data Dependence, *Journal of Econometrics*, 72, 1-32.

Escanciano, J.C. ,2006, A Consistent Diagnostic Test for Regression Models Using Projections, *Econometric Theory*, 22, 1030-1051.

Escanciano, J.C., 2009, On the Lack of Power of Omnibus Specification Tests, *Econometric Theory*, 25, 162-194.

Fan, Y.Q., & Q. Li, 2000, Consistent Model Specification Tests: Kernel Based Tests Versus Bierens' ICM tests, *Econometric Theory*, 16, 1016-1041.

Gozalo, P.L., 1993, A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models, *Econometric Theory*, 9, 451- 477.

Härdle, W. & E. Mammen, 1993, Comparing Nonparametric Versus Parametric Regression Fits, *Annals of Statistics*, 21, 1926-1947.

Hausman, J., 1978, Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

Holly, A., 1982, A Remark on Hausman's Specification Test, *Econometrica*, 50, 749-760.

Hong, Y. & H. White, 1995, Consistent Specification Testing Via Nonparametric Series Regression, *Econometrica*, 63, 1133-1159.

Horowitz, J.T. & W, Härdle, 1994, Testing a Parametric Model Against a Semiparametric Alternative, *Econometric Theory*, 10, 821848.

Jennrich, R. I. ,1969, Asymptotic Properties of Non-Linear Least Squares Estimators, *Annals of Mathematical Statistics*, 40, 633-643.

Justel, A., D. Pena & R. Zamar, 1997, A Multivariate Kolmogorov-Smirnov Test of Goodness of Fit, *Statistics and Probability Letters*, 35, 251-259.

Li, F. & G. Tkacz, 1996, A Consistent Bootstrap Test for Conditional Density Functions with Time-Series Data, *Journal of Econometrics*, 133, 863-886.

Li, Q. & S. Wang, 1998, A Simple Consistent Bootstrap Test for a Parametric Regression Function, *Journal of Econometrics*, 87, 145-165.

Kullback, L., & R.A. Leibler, 1951, On Information and Sufficiency, *Annals of Mathematical Statistics,* 22, 79-86.

Newey, W.K., 1985, Maximum Likelihood Specification Testing and Conditional Moment Tests, *Econometrica*, 53, 1047-1070.

Stinchcombe, M. & H. White, 1998, Consistent Specification Testing With

Nuisance Parameters Present Only Under the Alternative, *Econometric Theory*, 14, 295-325.

Stute, W., 1997, Nonparametric Model Checks for Regression, *Annals of Statistics*, 25, 613-641.

White, H., 1981, Consequences and Detection of Misspecified Nonlinear Regression Models, *Journal of the American Statistical Association*, 76, 419-433.

Zheng, J.X., 1996, A Consistent Test of Functional Form via Nonparametric Estimation Techniques, *Journal of Econometrics*, 75, 263-289.

Zheng, J.X., 2000, A Consistent Test of Conditional Parametric Distributions, *Econometric Theory*, 16, 667-691.