Integrated Conditional Moment Testing of Conditional Heteroskedasticity Models*

Herman J. Bierens
Pennsylvania State University

Santiago Mira
University of Las Palmas, Gran Canaria

April 1999

Abstract

In this paper we propose a consistent Integrated Conditional Moment (ICM) test of the functional form of a conditional heteroskedasticity model, for example a GARCH specification, which is asymptotically independent of the ICM test of the specification of the underlying conditional expectation model, under the null hypothesis that both models are correctly specified.

1 Introduction

The Integrated Conditional Moment (ICM) estimator of a parameter vector is derived by minimizing the test statistic of the ICM test of Bierens(1982) and Bierens and Ploberger (1997) to the parameters. The advantage of this estimator is that its objective function automatically yields a consistent test of the null hypothesis that the model involved is correctly specified. We apply this approach to estimation and testing of ARCH, GARCH and other conditional heteroskedasticity models. Our main aim is to construct asymptotically independent ICM tests of the correctness of both the specification of the conditional expectation model and the conditional variance model.

*Incomplete working paper.

2 The ICM test

2.1 The Bierens-Ploberger ICM test of a conditional expectation model

The ICM test is based on the following 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 the set*

$$\{s \in \mathbb{N} : (d/du)^s w(u) \mid_{u=0} = 0\}$$

is finite. 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 Bierens and Ploberger (1997) for the general case. Examples of suitable functions $w(u)$ in the general case are $w(u) = \cos(u) + \sin(u)$, and $w(u) = 1/[1 + \exp(c - u)]$ for $c \neq 0$. See also Stinchcombe and White (1998) for further elaborations on this theorem.

The condition that the random vector x is bounded can be get rid off by replacing x with $\Phi(x)$, 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. See Bierens (1982, 1990).

Theorem 1 suggests that, given a random sample (y_t, x_t), $t = 1, .., n$, $x_t \in \mathbb{R}^k$, and a conditional expectation model $E(y_t|x_t) = g(x_t, \theta_0)$, the null hypothesis $P[E(y_t|x_t) = g(x_t, \theta_0)] = 1$ for some θ_0, can be consistently tested on the basis of the Integrated Conditional Moment (ICM) statistic

$$\int |\widehat{z}(\xi)|^2 \, d\mu(\xi),$$

where

$$\widehat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{u}_t w\left(\xi' \Phi(x_t)\right),$$

with

$$\widehat{u}_t = y_t - g(x_t, \widehat{\theta}),$$

where $\widehat{\theta}$ is the nonlinear least squares estimator of θ_0, Φ is a bounded one-to-one mapping, $w(.)$ is a weighting function satisfying the conditions of Theorem 1, and μ a probability measure on a compact set $\Xi \subset \mathbb{R}^k$ with positive Lebesgue measure, which is absolute continuous with respect to Lebesgue measure. The ICM test was proposed by Bierens (1982), for the case $w(u) = \exp(i.u)$, Ξ a hypercube in $\mathbb{R}^k$, μ the Lebesgue measure on Ξ, and i.i.d. observations (y_t, x_t).

It has been shown by Bierens (1990) and Bierens and Ploberger (1997) that under some mild regularity conditions (among which the assumption that the function $w(.)$ is real-valued), and the null hypothesis involved, $\widehat{z} \Rightarrow z$ on Ξ, where z is a zero-mean Gaussian process with covariance function $\Gamma(\xi_1, \xi_2) = E\left[z(\xi_1) z(\xi_2)\right]$, hence $\int |\widehat{z}(\xi)|^2 d\mu(\xi) \to \int |z(\xi)|^2 d\mu(\xi)$ in distribution, whereas under the general alternative that the null is false, $\widehat{z}(\xi)/\sqrt{n} \to \eta(\xi) = p\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \widehat{u}_t w\left(\xi' \Phi(x_t)\right)$ in probability, uniformly on Ξ, where $\eta(\xi) \neq 0$ except on a set with zero Lebesgue measure. Consequently, under the alternative $(1/n) \int |\widehat{z}(\xi)|^2 d\mu(\xi) \to \int |\eta(\xi)|^2 d\mu(\xi) > 0$, a.s.

The asymptotic null distribution of the ICM statistic is of the type

$$\int |z(\xi)|^2 d\mu(\xi) = \sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2,$$

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. Moreover,

$$\frac{\int |z(\xi)|^2 d\mu(\xi)}{\int \Gamma(\xi, \xi) d\mu(\xi)} = \frac{\sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2}{\sum_{i=1}^{\infty} \lambda_i} \leq \sup_{m \geq 1} \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i^2 = \overline{T},$$

say, so that asymptotic critical values can be derived from the latter distribution. The actual test statistic of the ICM test is therefore

$$\widehat{T}_{ICM} = \frac{\int |\widehat{z}(\xi)|^2 d\mu(\xi)}{\int \widehat{\Gamma}(\xi, \xi) d\mu(\xi)}, \tag{1}$$

where $\widehat{\Gamma}(\xi_1, \xi_2)$ is a consistent estimator of $\Gamma(\xi_2, \xi_2)$, uniformly on $\Xi \times \Xi$.

The asymptotic null distribution of $\widehat{T}_{ICM}$ is case-dependent, because the eigenvalues λ_i depend on the distribution of (y_t, x_t) and the conditional expectation model $g(x_t, \theta_0)$, but is dominated by the distribution of $\overline{T}$. Thus, denoting the $1-\alpha$ quantile of $\overline{T}$ by T_α, i.e., $P(\overline{T} \geq T_\alpha) = \alpha$, the null hypothesis is rejected at the $\alpha \times 100\%$ significance level if $\widehat{T}_{ICM} \geq T_\alpha$. The values of T_α for $\alpha = 0.10, 0.05, 0.01$ can be found in Bierens and Ploberger (1997).

2.2 De Jong's ICM test

The Bierens-Ploberger version of the ICM test allows for consistently testing of linear and nonlinear ARX models, but not for ARMAX models, because, given a k-variate vector time series process

$$x_t = (y_t, x_t^{*\prime})' \in \mathbb{R} \times \mathbb{R}^{k-1}, \tag{2}$$

an ARMAX model represents the conditional expectation of the dependent variable y_t relative to *all* lagged x_t's. Bierens (1984) and De Jong (1996) have, in different ways, extended the ICM test to the case where x is infinite dimensional, i.e., $x = (x'_{t-1}, x'_{t-2}, ...)'$, in order to accommodate conditioning on the infinite past of a k-variate time series process x_t. In this paper we shall adopt the approach of De Jong (1996).

The space $(\Xi, \|\cdot\|)$ defined in De Jong (1996) is given as follows. For two infinite sequences of points in $\mathbb{R}^\infty$, 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

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

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

$$\Xi = \{\xi \in \mathbb{R}^\infty : a_j \leq \xi_j \leq b_j, \forall j \geq 1\},$$

where $a_j < b_j$ and $|a_j|, |b_j| \leq cj^{-2}$ for some constant $c > 0$. With this definition $(\Xi, \|\cdot\|)$ is a compact metric space, and therefore it is totally bounded. Following Bierens (1990), De Jong now proposes to use the weight function

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

and the Lebesgue measure on Ξ as the measure μ. However, in view of Theorem 1, De Jong's results carry over to the more general case

$$w_t(\xi) = w\left(\sum_{j=1}^{t} \xi_j' \Phi(x_{t-j})\right), \tag{3}$$

where $w(\cdot)$ is a real-valued function satisfying the conditions of Theorem 1(b), and Φ is a bounded one-to-one mapping. If $w(\cdot)$ is real valued but only satisfies the conditions of Theorem 1(a), we have to choose $a_j < 0 < b_j, \forall j$. In this case the results of Theorem 1(b) read:

THEOREM 2: *Let u_t be a random variable satisfying $E|u_t| < \infty$, and let x_t be a k-variate time series process, such that (u_t, x_t) is stationary. Let $(\Xi, \|\cdot\|)$ be defined as in De Jong* (1996), *and let*

$$\overline{w}_t(\xi) = w\left(\sum_{j=1}^{\infty} \xi_j' \Phi(x_{t-j})\right), \tag{4}$$

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

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

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 the same as (1).

Note that if we choose

$$w(\cdot) = \cos(\cdot) + \sin(\cdot), \tag{5}$$
$$\Xi = \times_{j=1}^{\infty} \left\{ \times_{\ell=1}^{k} \left[-cj^{-2}, cj^{-2}\right] \right\}$$

for some constant $c > 0$, and $\mu(\xi)$ a symmetric measure on Ξ, then

$$\int_{\Xi} w_t(\xi)^2 d\mu(\xi) = 1,$$

and consequently the denominator of (1) becomes,

$$\int \widehat{\Gamma}(\xi, \xi) d\mu(\xi) = (1/n) \sum_{t=1}^{n} \widehat{u}_t^2.$$

The latter result is one of the reasons why we favor the weight function (5).

2.3 An ICM test for ARCH

The results of De Jong can be straightforwardly applied to construct an ICM test of ARCH and GARCH, given the correctness of the underlying conditional expectation model. For example, consider the ARCH(1) model

$$y_t = \beta_0' \widetilde{x}_t + u_t,$$

$$u_t = v_t \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}, \quad E_{t-1}(v_t) = 0, \quad E_{t-1}(v_t^2) = 1,$$

where $\widetilde{x}_t = (1, x_t')'$, with x_t defined by (2), and

$$E_{t-1}(\bullet) = E(\bullet | x_{t-1}, x_{t-2}, ...).$$

Let $\widehat{\alpha}_0, \widehat{\alpha}_1$, and $\widehat{\beta}$ be the maximum likelihood estimators of the parameters involved (assuming that the v_t's are i.i.d. $N(0,1)$), and let $\widehat{e}_t = \widehat{u}_t^2 - \widehat{\alpha}_0 - \widehat{\alpha}_1 \widehat{u}_{t-1}^2$ be the ARCH(1) residuals, with $\widehat{u}_t = y_t - \widehat{\beta}_0' \widetilde{x}_t$. Then under the null hypothesis that the conditional variance of u_t is ARCH(1), the integral $\int \left| (1/\sqrt{n}) \sum_{t=1}^n \widehat{e}_t w_t(\xi) \right|^2 d\mu(\xi)$ converges in distribution, whereas under the appropriate choice of $w_t(\xi)$ and $\mu(\xi)$ this integral converges to infinity if the ARCH(1) model is incorrect.

However, the problem is that this ICM test is not independent of the ICM test of the null hypothesis that the conditional expectation of u_t is zero with probability 1, because both limiting null distributions depend on the parameter estimators of the conditional expectation model. This can be verified from

$$\begin{aligned} \widehat{e}_t &= \left(u_t - \left(\widehat{\beta} - \beta_0 \right)' \widetilde{x}_t \right)^2 - \widehat{\alpha}_0 - \widehat{\alpha}_1 \left(u_{t-1} - \left(\widehat{\beta} - \beta_0 \right)' \widetilde{x}_{t-1} \right)^2 \\ &= e_t - (\widehat{\alpha}_0 - \alpha_0) - (\widehat{\alpha}_1 - \alpha_1) u_{t-1}^2 \\ &\quad -2 u_t \widetilde{x}_t' \left(\widehat{\beta} - \beta_0 \right) + \left(\widehat{\beta} - \beta_0 \right)' \left(\widetilde{x}_t \widetilde{x}_t' - \widehat{\alpha}_1 \widetilde{x}_{t-1} \widetilde{x}_{t-1}' \right) \left(\widehat{\beta} - \beta_0 \right) \\ &\quad +2 \widehat{\alpha}_1 u_{t-1} \widetilde{x}_{t-1}' \left(\widehat{\beta} - \beta_0 \right), \end{aligned}$$

where

$$e_t = u_t^2 - \alpha_0 - \alpha_1 u_{t-1}^2 = \left(v_t^2 - 1 \right) \left(\alpha_0 + \alpha_1 u_{t-1}^2 \right),$$

so that under the null hypothesis,

$$(1/\sqrt{n}) \sum_{t=1}^n \widehat{e}_t w_t(\xi) = (1/\sqrt{n}) \sum_{t=1}^n \left(v_t^2 - 1 \right) \left(\alpha_0 + \alpha_1 u_{t-1}^2 \right) w_t(\xi)$$

$$
\begin{aligned}
&-\sqrt{n}\left(\begin{array}{c}\widehat{\alpha}_0-\alpha_0\\ \widehat{\alpha}_1-\alpha_1\end{array}\right)'\frac{1}{n}\sum_{t=1}^{n}\left(\begin{array}{c}1\\ u_{t-1}^2\end{array}\right)w_t(\xi)\\
&+2\sqrt{n}\left(\widehat{\beta}-\beta_0\right)'\frac{1}{n}\sum_{t=1}^{n}\left(\alpha_1 u_{t-1}\widetilde{x}_{t-1}\right)w_t(\xi)\\
&+O_p(1/\sqrt{n}),
\end{aligned}
$$

where the O_p term is uniform on Ξ. The null distribution of the ICM test of the null hypothesis $E_{t-1}(u_t)=0$ is based on

$$
\begin{aligned}
(1/\sqrt{n})\sum_{t=1}^{n}\widehat{u}_t w_t(\xi) &= (1/\sqrt{n})\sum_{t=1}^{n}v_t\left(\sqrt{\alpha_0+\alpha_1 u_{t-1}^2}\right)w_t(\xi)\\
&\quad -\sqrt{n}\left(\widehat{\beta}-\beta_0\right)'\frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_t w_t(\xi).
\end{aligned}
$$

Although

$$
(1/\sqrt{n})\sum_{t=1}^{n}v_t\left(\sqrt{\alpha_0+\alpha_1 u_{t-1}^2}\right)w_t(\xi)
$$

and

$$
(1/\sqrt{n})\sum_{t=1}^{n}\left(v_t^2-1\right)\left(\alpha_0+\alpha_1 u_{t-1}^2\right)w_t(\xi)
$$

converge to independent Gaussian processes if the v_t's are i.i.d. $N(0,1)$, conditional on $x_{t-1},x_{t-2,....,}$due to the fact that then $E_{t-1}\left[\left(v_t^2-1\right)v_t\right]=0$, the two ICM tests themselves will be dependent because they have $\sqrt{n}\left(\widehat{\beta}-\beta_0\right)$ in common.

A possible solution to this problem is to use *independent* estimators of β_0 in constructing the two ICM tests, for example, by using the following ICM estimators:

$$
\widehat{\beta}_0=\arg\min\beta\int\left|(1/\sqrt{n})\sum_{t=1}^{n}\left(y_t-\beta'\widetilde{x}_t\right)w_t(\xi)\right|^2 d\mu(\xi),
$$

and

$$
\begin{aligned}
\left(\begin{array}{c}\widehat{\alpha}_0\\ \widehat{\alpha}_1\\ \widehat{\beta}_1\end{array}\right) &= \arg\min_{\alpha_0,\alpha_1,\beta}\int\left|(1/\sqrt{n})\sum_{t=1}^{n}\left(\left(y_t-\beta'\widetilde{x}_t\right)^2\right.\right.\\
&\quad \left.\left.-\alpha_0-\alpha_1\left(y_{t-1}-\beta'\widetilde{x}_{t-1}\right)^2\right)w_t(\xi)\right|^2 d\mu(\xi).
\end{aligned}
$$

As will be shown, the estimators $\widehat{\beta}_0$ and $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1')'$ are asymptotically independent if $E_{t-1}\left[\left(v_t^2-1\right)v_t\right]=0$, because

$$\sqrt{n}\left(\widehat{\beta}_0-\beta_0\right)=(1/\sqrt{n})\sum_{t=1}^{n}v_t g_t+o_p(1)$$

$$\sqrt{n}\begin{pmatrix}\widehat{\alpha}_0-\alpha_0\\ \widehat{\alpha}_1-\alpha_1\\ \widehat{\beta}_1-\beta_0\end{pmatrix}=(1/\sqrt{n})\sum_{t=1}^{n}\left(v_t^2-1\right)h_t+o_p(1),$$

where the weights g_t and h_t depend on $\widetilde{x}_t, \widetilde{x}_{t-1}, v_{t-1}, w_t(\xi)$, and $\mu(\xi)$.

3 The ICM test based on the ICM estimator

3.1 The model: Examples

Let $x_t=(y_t, x_t^{*\prime})'$ be a strictly stationary k-variate time series process [cf. (2)], and consider a possibly implicit single-equation time series model

$$f(y_t, y_{t-1}, ..., y_{t-p}, x_{t-1}^*, .., x_{t-q}^*, \theta_0)=u_t, \tag{6}$$

where f is a known function, θ_0 an unknown parameter vector, and u_t an error term which satisfies the condition

$$H_0: E(u_t|x_{t-1}, x_{t-2}, ...)=0 \ a.s. \tag{7}$$

Most single-equation time series models can be cast in this framework, especially if we allow the lag lengths p and/or q to be infinite in order to include ARMAX and GARCH models as well. However, for the time being we shall assume that both p and q are finite, and that x_t is observable for $t=-\max(p,q), ..., n$.

The ARX(1) regression model with ARCH(1) errors provides various examples of model (6). Let

$$y_t=\beta_0'\widehat{x}_{t-1}+v_t\sqrt{\alpha_{0,0}+\alpha_{0,1}\varepsilon_{t-1}^2(\beta_0)}, \tag{8}$$

where $\alpha_{0,0}>0,\ 0<\alpha_{0,1}<1,\ \theta_0=(\alpha_{0,0}, \alpha_{0,1}, \beta_0')'$, and

$$\varepsilon_t(\beta)=y_t-\beta'\widehat{x}_{t-1}, \quad \varepsilon_t^2(\beta)=\left(\varepsilon_t(\beta)\right)^2,$$

with $\widehat{x}_{t-1} = (1, y_{t-1}, x^*_{t-1})$. The correctness of this ARX-ARCH model hinges on two crucial hypotheses:

$$E_{t-1}(v_t) = 0 \; a.s., \tag{9}$$

and

$$E_{t-1}\left(v_t^2\right) = 1 \; a.s. \tag{10}$$

There are various models f that correspond to this ARX-ARCH model. First, the ARX part fits in this framework:

$$\begin{aligned} f(y_t, x^*_{t-1}, \theta_0) &= \varepsilon_t(\beta_0) = u_t, \\ u_t &= v_t\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)}, \end{aligned} \tag{11}$$

where now $\theta_0 = \beta_0$. However, various version of the ARCH part can be formulated. For example, let:

$$\begin{aligned} f(y_t, y_{t-1}, x^*_{t-1}, x^*_{t-2}, \theta_0) &= \varepsilon^2_t(\beta_0) - \alpha_{0,0} - \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0) = u_t, \\ u_t &= \left(v_t^2 - 1\right)\left(\alpha_{0,0} + \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)\right), \end{aligned} \tag{12}$$

where again $\theta_0 = (\alpha_{0,0}, \alpha_{0,1}, \beta_0')'$. Then under the hypotheses (9) and (10),

$$\begin{aligned} & E_{t-1}\left[f(y_t, y_{t-1}, x^*_{t-1}, x^*_{t-2}, \theta)\right] \\ = \; & -(\alpha_0 - \alpha_{0,0}) - (\alpha_1 - \alpha_{0,1})\,\varepsilon^2_{t-1}(\beta_0) + 2\alpha_1\varepsilon_{t-1}(\beta_0)\widehat{x}'_{t-2}(\beta - \beta_0) \\ & +(\beta - \beta_0)'\left(\widehat{x}_{t-1}\widehat{x}'_{t-1} - \alpha_1\widehat{x}_{t-2}\widehat{x}'_{t-2}\right)(\beta - \beta_0), \end{aligned}$$

where $\theta = (\alpha_0, \alpha_1, \beta')'$, hence

$$E_{t-1}\left[f(y_t, y_{t-1}, x^*_{t-1}, x^*_{t-2}, \theta)\right] = 0 \text{ a.s. if and only if } \theta = \theta_0. \tag{13}$$

An alternative ARCH(1) specification of f is

$$f(y_t, y_{t-1}, x^*_{t-1}, x^*_{t-2}, \theta_0) = \frac{\varepsilon^2_t(\beta_0)}{\alpha_{0,0} + \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)} - 1 = u_t, \tag{14}$$

$$u_t = v_t^2 - 1, \tag{15}$$

and again (13) holds if both (9) and (10) hold. The same applies to the alternative ARCH(1) model

$$f(y_t, y_{t-1}, x^*_{t-1}, x^*_{t-2}, \theta_0) = \frac{\varepsilon^2_t(\beta_0) - \alpha_{0,0} - \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)}{\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)}} = u_t, \tag{16}$$

$$u_t = \left(v_t^2 - 1\right)\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon^2_{t-1}(\beta_0)} \tag{17}$$

The latter model has the advantage (as will turn out later) that the factor $\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon_{t-1}^2(\beta_0)}$ is the same as in the ARX model (11).

Note that if (9) does not hold for some β_0, it is still possible that for some β_0, $E_{t-1}\left(\varepsilon_t^2(\beta_0)\right) = \alpha_{0,0} + \alpha_{0,1}\varepsilon_{t-1}^2(\beta_0)$. Therefore, the hypothesis (9) has to be tested separately.

3.2 The generic model

For notational convenience we shall suppress the random arguments of f, and write the model in generic form as

$$f_t(\theta_0) = u_t \in \mathbb{R}, \ \theta_0 \in \Theta \subset \mathbb{R}^m,$$

where θ_0 is such that

$$H_0 : E_{t-1}(u_t) = 0 \text{ a.s.}$$

We may interpret $f_t(\theta_0)$ as a short-hand notation for the left-hand side of (6), and $E_{t-1}(u_t)$ as a short-hand notation for the left-hand side of (7).

Throughout we assume that

ASSUMPTION 1: *The parameter space Θ is compact and convex, and $f_t(\theta)$ is a.s. twice continuously differentiable on Θ.*

Also, we need the identification condition that

ASSUMPTION 2: *For any pair $(\theta_1, \theta_2) \in \Theta \times \Theta$,*

$$P\left(E_{t-1}\left[f_t(\theta_1)\right] = E_{t-1}\left[f_t(\theta_2)\right]\right) = 1$$

implies $\theta_1 = \theta_2$.

Moreover, we may replace any reference to the time series process x_t by references to the σ-algebra $\mathcal{F}_{t-1}$ generated by $\{x_{t-j},\ j \geq 1\}$ and/or the σ-algebra $\mathcal{F}_{0,t-1}$ generated by $\{x_{t-j},\ j = 1,..,t\}$:

ASSUMPTION 3: $\{f_t(\theta),\ \theta \in \Theta\}$ *is strictly stationary, and measurable $\mathcal{F}_t$, where $\{\mathcal{F}_t\}$ is a monotonic increasing sequence of σ-algebras: $\forall t : \mathcal{F}_{t-1} \subset \mathcal{F}_t$.*

Now $E_{t-1}(u_t)$ is a short-hand notation for $E\left(u_t|\mathcal{F}_{t-1}\right)$.

Note that for the weight functions $\overline{w}_t(\xi)$, $w_t(\xi)$, and the metric space Ξ in Theorem 2, the σ-algebra generated by $\{x_{t-j},\ j \geq 1\}$ is the same as the σ-algebra generated by $\{\overline{w}_t(\xi),\ \xi \in \Xi\}$, and the σ-algebra generated by $\{x_{t-j},\ j = 1,..,t\}$ is the same as the σ-algebra generated by $\{w_t(\xi),\ \xi \in \Xi\}$. Moreover, it is not hard to verify that we can choose w and Φ such that $\lim_{t\to\infty} E \int_\Xi (\overline{w}_t(\xi) - w_t(\xi))^2 d\xi = 0$. In view of Theorem 2 we may therefore assume, without loss of generality, that:

ASSUMPTION 4:
(a) *The weight functions* $\overline{w}_t(\xi)$ *and* $w_t(\xi)$ *are random functions on a compact metric space* $(\Xi, \|\cdot\|)$, *such that*

$$\lim_{t\to\infty} E \int_\Xi (\overline{w}_t(\xi) - w_t(\xi))^2 d\xi = 0.$$

(b) *The process* $\{\overline{w}_t(\xi), \xi \in \Xi\}$ *is measurable* $\mathcal{F}_{t-1}$, *and the process* $\{w_t(\xi), \xi \in \Xi\}$ *is measurable* $\mathcal{F}_{0,t-1}$, *where* $\mathcal{F}_{0,t}$ *is a monotonic increasing sequence of sub-*σ*-algebras of* $\mathcal{F}_{t-1}$: $\forall t : \mathcal{F}_{0,t-1} \subset \mathcal{F}_{0,t} \subset \mathcal{F}_t$.
(c) *For any* $\mathcal{F}_t$*-measurable random variable* u_t *for which* $E[|u_t|] < \infty$,

$$P[E(u_t|\mathcal{F}_{t-1}) = 0] < 1$$

implies that the set $\{\xi \in \Xi : E(u_t\overline{w}_t(\xi)) = 0\}$ *has Lebesgue measure zero and is nowhere dense in* Ξ.

3.3 The ICM estimator of θ_0

The ICM estimator of θ_0 is:

$$\widehat{\theta}_{ICM} = \arg\min_{\theta\in\Theta} \widehat{Q}_{ICM}(\theta),$$

where

$$\widehat{Q}_{ICM}(\theta) = \int_\Xi \left| \frac{1}{n} \sum_{t=1}^{n} f_t(\theta) w_t(\xi) \right|^2 d\mu(\xi),$$

with μ as in Theorem 2.

A more general estimator would be the conditional moment estimator proposed by Carrasco and Florens (1997), which is based on an objective function of the form

$$\begin{aligned} \widetilde{Q}_\Lambda(\theta) \quad &= \quad \int_\Xi \int_\Xi \left(\frac{1}{n} \sum_{t=1}^{n} f(y_t, x_{t-1}, \theta) w_t(\xi_1) \right) \Lambda(\xi_1, \xi_2) \\ &\quad \times \left(\frac{1}{n} \sum_{t=1}^{n} f(y_t, x_{t-1}, \theta) w_t(\xi_2) \right) d\mu(\xi_1) d\mu(\xi_2), \end{aligned}$$

where $\Lambda(\xi_1, \xi_2)$ is a positive definite function on $\Xi \times \Xi$. Carrasco and Florens (1997) derive the optimal $\Lambda(\xi_1, \xi_2)$ as the inverse of a particular positive definite function, similarly to the GMM case. However, this inverse may not exist, and if it exists it is quite difficult to compute. Therefore we prefer our approach. Moreover, in the context of testing the ARCH specification, our objective is not to efficiently estimating θ_0, but to construct independent ICM tests of the correctness of the specification of conditional expectation and conditional variance models.

It can be shown, similarly to Bierens (1990), De Jong (1997), and Bierens and Ploberger (1997), that under H_0, Assumptions 1-4, and the additional assumptions in the Appendix, that $\widehat{\theta}_{ICM}$ is consistent, and that

$$\widehat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t w_t(\xi) \Rightarrow z(\xi), \tag{18}$$

where z is a zero mean Gaussian process with covariance function

$$\Gamma(\xi_1, \xi_2) = E\left[z(\xi_1) z(\xi_2)\right] = E\left[u_t^2 \overline{w}_t(\xi_1) \overline{w}_t(\xi_2)\right].$$

This function can be consistently estimated by

$$\widehat{\Gamma}(\xi_1, \xi_2) = \frac{1}{n} \sum_{t=1}^{n} \left[f_t\left(\widehat{\theta}_{ICM}\right)\right]^2 w_t(\xi_1) w_t(\xi_2),$$

uniformly on $\Xi \times \Xi$. Denoting,

$$b(\xi) = E\left[\frac{\partial f_t(\theta_0)}{\partial \theta'} \overline{w}_t(\xi)\right],$$

and

$$A = \int_{\Xi} b(\xi) b(\xi)' d\mu(\xi)$$

it can now easily be shown, using (18) and standard asymptotic theory, that

$$\sqrt{n}\left(\widehat{\theta}_{ICM} - \theta_0\right) \to -A^{-1} \int_{\Xi} b(\xi) z(\xi) d\mu(\xi) \text{ in distr.}$$

The function $b(\xi)$ can be consistently estimated by

$$\widehat{b}(\xi) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial f_t\left(\widehat{\theta}_{ICM}\right)}{\partial \widehat{\theta}'_{ICM}}\right) w_t(\xi),$$

uniformly on Ξ, hence

$$\widehat{A} = \int_\Xi \widehat{b}(\xi)\widehat{b}(\xi)' d\mu(\xi)$$

is a consistent estimator of the matrix A.

Finally, observe that

$$\int_\Xi b(\xi)z(\xi)d\mu(\xi) \sim N_m\left[0, \Omega\right],$$

where

$$\begin{aligned}\Omega &= \int_\Xi\int_\Xi \Gamma(\xi_1,\xi_2)b(\xi_1)b(\xi_2)'d\mu(\xi_1)d\mu(\xi_2) \\ &= E\left[u_t^2\left(\int_\Xi \overline{w}_t(\xi_1)b(\xi_1)d\mu(\xi_1)\right)\left(\int_\Xi \overline{w}_t(\xi_2)b(\xi_2)d\mu(\xi_2)\right)'\right].\end{aligned}$$

This matrix can be consistently estimated by

$$\widehat{\Omega} = \frac{1}{n}\sum_{t=1}^n \left(f_t(\widehat{\theta}_{ICM})\right)^2\left(\int_\Xi w_t(\xi)\widehat{b}(\xi)d\mu(\xi)\right)\left(\int_\Xi w_t(\xi)\widehat{b}(\xi)d\mu(\xi)\right)'$$

Summarizing, we have:

THEOREM 3: *Under Assumptions 1-4, the null hypothesis H_0, and the additional assumptions in the Appendix, the ICM estimator is asymptotically normally distributed:*

$$\sqrt{n}\left(\widehat{\theta}_{ICM} - \theta_0\right) \to -A^{-1}\int_\Xi b(\xi)z(\xi)d\mu(\xi) \ \sim N_m\left[0, A^{-1}\Omega A^{-1}\ \right],$$

and $p\lim_{n\to\infty} \widehat{A}^{-1}\widehat{\Omega}\widehat{A}^{-1} = A^{-1}\Omega A^{-1}$.

3.4 The ICM test on the basis of the ICM estimator

Next, let us look at the asymptotic distribution of the ICM statistic

$$n\widehat{Q}_{ICM}(\widehat{\theta}_{ICM}) = \int_\Xi \left|\frac{1}{\sqrt{n}}\sum_{t=1}^n f_t(\widehat{\theta}_{ICM})w_t(\xi)\right|^2 d\mu(\xi),$$

on which the ICM test is based. We have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_t(\widehat{\theta}_{ICM})w_t(\xi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t w_t(\xi) \\ &\quad + b(\xi)'\sqrt{n}\left(\widehat{\theta}_{ICM}-\theta_0\right)+o_p(1),\end{aligned}$$

uniformly on Ξ, hence

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_t(\widehat{\theta}_{ICM})w_t(\xi) &\Rightarrow z(\xi)-b(\xi)'A^{-1}\int_{\Xi} b(\xi_*)z(\xi_*)d\mu(\xi_*) \\ &= z^*(\xi),\ \ say,\end{aligned}$$

where $z^*(\xi)$ is a zero-mean Gaussian process on Ξ with covariance function

$$\begin{aligned}\Gamma_*(\xi_1,\xi_2) &= \Gamma(\xi_1,\xi_2)-\int_{\Xi}\Gamma(\xi_1,\xi_*)b(\xi_*)'d\mu(\xi_*)A^{-1}b(\xi_2) \\ &\quad -b(\xi_1)'A^{-1}\int_{\Xi}\Gamma(\xi_*,\xi_2)b(\xi_*)d\mu(\xi_*)+b(\xi_1)'A^{-1}\Omega A^{-1}b(\xi_2).\end{aligned}$$

Thus we have

$$\begin{aligned}n\widehat{Q}_{ICM}(\widehat{\theta}_{ICM}) &\to \int_{\Xi} z(\xi)^2 d\mu(\xi)-\left(\int_{\Xi} z(\xi)b(\xi)d\mu(\xi)\right)'A^{-1}\int_{\Xi} b(\xi_*)z(\xi_*)d\mu(\xi_*) \\ &= \int_{\Xi} z^*(\xi)^2 d\mu(\xi)\end{aligned}$$

in distr. Note that

$$\begin{aligned}\int_{\Xi}\Gamma_*(\xi,\xi)d\mu(\xi) &= \int_{\Xi}\Gamma(\xi,\xi)d\mu(\xi)-\int_{\Xi} b(\xi)'A^{-1}\Omega A^{-1}b(\xi)d\mu(\xi) \\ &= \int_{\Xi}\Gamma(\xi,\xi)d\mu(\xi)-\text{trace}\left(A^{-1}\Omega\right),\end{aligned}$$

which can be estimated consistently by

$$\int_{\Xi}\widehat{\Gamma}_*(\xi,\xi)d\mu(\xi)=\int_{\Xi}\widehat{\Gamma}(\xi,\xi)d\mu(\xi)-\text{trace}\left(\widehat{A}^{-1}\widehat{\Omega}\right).$$

Therefore, the ICM test statistic for testing the null hypothesis (10) is

$$\widehat{T}_{ICM}=\frac{n\widehat{Q}_{ICM}(\widehat{\theta}_{ICM})}{\int_{\Xi}\widehat{\Gamma}(\xi,\xi)d\mu(\xi)-\text{trace}\left(\widehat{A}^{-1}\widehat{\Omega}\right)}.$$

Finally, we recall that the weight functions $w_t(\xi)$ and the measure $\mu(\xi)$ can be chosen such that $\int_\Xi w_t(\xi)^2 d\mu(\xi) = 1$, so that then

$$\int_\Xi \widehat{\Gamma}(\xi,\xi)d\mu(\xi) = \frac{1}{n}\sum_{t=1}^{n}\left(f_t(\widehat{\theta}_{ICM})\right)^2.$$

3.5 The one Newton step approximation of the ICM estimator

Since the objective function $\widehat{Q}_{ICM}(\theta)$ of the ICM estimator is a double sum:

$$\begin{aligned}\widehat{Q}_{ICM}(\theta) &= \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} f\left(y_s, x_{s-1}, \theta\right) f\left(y_t, x_{t-1}, \theta\right)\int_\Xi w_s(\xi)w_t(\xi)d\mu(\xi) \\ &= \frac{1}{n^2}\sum_{t=1}^{n} f\left(y_t, x_{t-1}, \theta\right)^2 \int_\Xi w_t(\xi)^2 d\mu(\xi) \\ &\quad + \frac{2}{n^2}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n} f\left(y_s, x_{s-1}, \theta\right) f\left(y_t, x_{t-1}, \theta\right)\int_\Xi w_s(\xi)w_t(\xi)d\mu(\xi),\end{aligned}$$

the actual computation of the ICM estimator $\widehat{\theta}_{ICM} = \arg\min_{\theta\in\Theta}\widehat{Q}_{ICM}(\theta)$ is quite laborious (even if we choose the weights $w_t(\xi)$ and $\mu(\xi)$ such that $\int_\Xi w_t(\xi)^2 d\mu(\xi) = 1$). However, there is no need to go all the way with minimizing $\widehat{Q}_{ICM}(\theta)$. As is well known from maximum likelihood theory, if one has an initial estimator $\widetilde{\theta}$ such that

$$\sqrt{n}\left(\widetilde{\theta} - \theta_0\right) = O_p(1),$$

then a single Newton step starting from $\widetilde{\theta}$ yields an estimator with the same asymptotic normal distribution as the maximum likelihood estimator. For example, in the case of ARCH(1) model (12), we may choose

$$\widetilde{\theta} = (\widehat{\alpha}'_{OLS}, \widehat{\beta}'_{OLS})',$$

where $\widehat{\alpha}_{OLS}$ is the vector of OLS estimators of α_0 and α_1 in the linear regression of $\widehat{u}_t^2$ on $\widehat{u}_{t-1}^2$, with $\widehat{u}_t = y_t - \widehat{\beta}'_{OLS}\widehat{x}_{t-1}$, and $\widehat{\beta}_{OLS}$ the OLS estimator of β_0. It is a standard exercise to prove that this result carries over to the ICM estimator, as follows.

Let

$$\widetilde{\eta}(\xi) = \frac{1}{n}\sum_{t=1}^{n} f_t(\widetilde{\theta})w_t(\xi),$$

$$\widetilde{b}(\xi) = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial f_t\left(\widetilde{\theta}\right)}{\partial\widetilde{\theta}'}\right)w_t(\xi),$$

and

$$\widetilde{\theta}_{NEWT} = \widetilde{\theta} - \left(\int_{\Xi}\widetilde{b}(\xi)\widetilde{b}(\xi)'d\mu(\xi)\right)^{-1}\int_{\Xi}\widetilde{b}(\xi)\widetilde{\eta}(\xi)d\mu(\xi).$$

Then

$$\sqrt{n}\left(\widetilde{\theta}_{NEWT} - \widehat{\theta}_{ICM}\right) = o_p(1).$$

Moreover, using the second-order Taylor expansion of $n\widehat{Q}_{ICM}(\widetilde{\theta}_{NEWT})$ around $\widehat{\theta}_{ICM}$, it follows that

$$n\widehat{Q}_{ICM}(\widetilde{\theta}_{NEWT}) = n\widehat{Q}_{ICM}(\widehat{\theta}_{ICM}) + o_p(1).$$

Therefore, rather than using $\widehat{\theta}_{ICM}$, we may use the one Newton step estimator $\widetilde{\theta}_{NEWT}$.

4 Estimation and testing of ARX-ARCH models

In the ARCH(1) cases (12) and (14), the ICM estimator $\widehat{\theta}_{ICM}$ also yields an estimator $\widehat{\beta}$, say, of β_0: $\widehat{\theta}_{ICM} = (\widehat{\alpha}', \widehat{\beta}')'$, where $\widehat{\alpha}' = (\widehat{\alpha}_0, \widehat{\alpha}_1)$. Moreover, the limiting normal distribution of $\widehat{\theta}_{ICM}$, and the null distribution of the ICM test statistic $\widehat{T}_{ICM}$, are determined by the Gaussian process $z(\xi) = z_2(\xi)$, say, which is the limit process of

$$\begin{aligned}\widehat{z}_2(\xi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t w_t(\xi) \\ &= \begin{cases} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(v_t^2-1\right)\left(\alpha_{0,0}+\alpha_{0,1}\varepsilon_{t-1}^2\right)w_t(\xi) & \text{in the case (12),} \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(v_t^2-1\right)w_t(\xi) & \text{in the case (14),} \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(v_t^2-1\right)\sqrt{\alpha_{0,0}+\alpha_{0,1}\varepsilon_{t-1}^2}w_t(\xi) & \text{in the case (16).} \end{cases}\end{aligned}$$

Similarly, the limiting normal distribution of the ICM estimator $\widetilde{\beta}_{ICM}$ of β_0,and the null distribution of the corresponding ICM test statistic $\widetilde{T}_{ICM}$, in the case of the conditional expectation model for y_t,

$$\begin{aligned} f(y_t, x_{t-1}, \beta) &= \varepsilon_t(\beta_0), \\ f(y_t, x_{t-1}, \beta_0) &= u_t = v_t\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon_{t-1}^2(\beta_0)}, \\ E_{t-1}(v_t) &= 0 \ a.s., \end{aligned} \tag{19}$$

are determined by the limiting Gaussian process $z_1(\xi)$ of

$$\widehat{z}_1(\xi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_t\left(\sqrt{\alpha_{0,0} + \alpha_{0,1}\varepsilon_{t-1}^2(\beta_0)}\right)w_t(\xi).$$

Moreover

$$\begin{pmatrix} \widehat{z}_1(\xi) \\ \widehat{z}_2(\xi) \end{pmatrix} \Rightarrow \begin{pmatrix} z_1(\xi) \\ z_2(\xi) \end{pmatrix}.$$

Now if under both null hypotheses $E_{t-1}(v_t) = 0$ a.s., $E_{t-1}(v_t^2) = 1$ a.s., we also have that

$$E_{t-1}(v_t^3) = 0 \ a.s., \tag{20}$$

then

$$E\left[z_1(\xi_1)z_2(\xi_2)\right] = 0,$$

hence the Gaussian processes $z_1(\xi)$ and $z_2(\xi)$ are independent, and asymptotically so are $\left(\widehat{\theta}_{ICM}, \widehat{T}_{ICM}\right)$ and $\left(\widetilde{\beta}_{ICM}, \widetilde{T}_{ICM}\right)$.

By a similar argument it follows that asymptotically the OLS estimator $\widetilde{\beta}_{OLS}$ of β_0, and the corresponding test statistic $\widetilde{T}_{BP}$ of the ICM test of Bierens and Ploberger (1997), are independent of $\left(\widehat{\theta}_{ICM}, \widehat{T}_{ICM}\right)$. These results carry straightforwardly over to ARX(p) models with ARCH(q) errors.

An alternative approach to consistently testing the ARCH(1) specification would be to conduct the ICM test of Bierens and Ploberger (1997) on the basis of say the maximum likelihood estimator $\widehat{\theta}_{ML}$ of θ_0. The null distribution of this ICM test then depends on the Gaussian limit process $z_{ML}(\xi)$ of

$$\widehat{z}(\xi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f(y_t, x_{t-1}, \widehat{\theta}_{ML})w_t(\xi),$$

where f is defined by (12). However, it is not too hard to verify that this test will not be asymptotically independent of the ICM test $\widetilde{T}_{BP}$ of the conditional expectation model (19). This is the main reason for considering ICM estimation of the parameters of ARCH models.

5 Estimation and testing of general conditional heteroskedasticity models

The problem of estimation and consistently testing of an ARCH model using the ICM approach can be cast be cast in the following general framework:

Let for $t = 1, .., n$,

$$\begin{array}{llll} H_{0,1}: & \exists \beta_0 \in \Theta_\beta: & f_{1,t}(\beta_0) = u_{1,t}, & E_{t-1}(u_{1,t}) = 0 \ a.s., \\ H_{0,2}: & \exists(\alpha_0', \beta_0')' \in \Theta_\alpha \times \Theta_\beta: & f_{2,t}(\alpha_0, \beta_0) = u_{2,t}, & E_{t-1}(u_{2,t}) = 0 \ a.s., \end{array} \tag{21}$$

$$\Theta_\alpha \times \Theta_\beta \subset \mathbb{R}^p \times \mathbb{R}^q,$$

where $(f_{1,t}(\beta), f_{2,t}(\alpha, \beta))$ is a stationary process on $\Theta_\alpha \times \Theta_\beta$, and $f_{1,t}(\beta)$, $f_{2,t}(\alpha, \beta)$ are (a.s.) twice continuously differentiable on $\Theta_\alpha \times \Theta_\beta$. Again, the conditional expectation $E_{t-1}(\cdot)$ is taken relative to the σ-algebra $\mathcal{F}_{t-1}$ generated by $\{\overline{w}_t(\xi),\ \xi \in \Xi\}$.

We recall that in the ARCH case,

$$u_{1,t} = v_t \sqrt{E_{t-1}\left(u_{1,t}^2\right)},\ E_{t-1}\left(v_t\right) = 0 \ a.s.,$$

$$u_{2,t} = \left(v_t^2 - 1\right) g_{t-1},\ \ E_{t-1}\left(v_t^2\right) = 1 \ a.s.,$$

where g_{t-1} is measurable $\mathcal{F}_{t-1}$. For example, $g_{t-1} = E_{t-1}\left(u_{1,t}^2\right)$ in the case (12), and $g_{t-1} = 1$ in the case (14).

Denoting

$$\widehat{z}_j(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{j,t} w_t(\xi),\ \ j = 1, 2,$$

it follows that

$$\widehat{z}(\xi) = \begin{pmatrix} \widehat{z}_1(\xi) \\ \widehat{z}_2(\xi) \end{pmatrix} \Rightarrow z(\xi) = \begin{pmatrix} z_1(\xi) \\ z_2(\xi) \end{pmatrix},$$

If we are willing to assume that v_t is independent of the events in $\mathcal{F}_{t-1}$ (which is the usual assumption in ARCH and GARCH models), then it is possible to make the ICM tests of $H_{0,1}$ and $H_{0,2}$ independent, as follows.

Let μ be a probability measure on Ξ which is absolutely continuous with respect to Lebesgue measure. Let

$$w_t(\xi) = w\left(\sum_{j=1}^{t} \xi_j' \Phi(x_{t-j})\right).$$

Denote

$$\begin{aligned}\widehat{Q}_1(\beta) &= \int_{\Xi} \left| \frac{1}{n} \sum_{t=1}^{n} f_{1,t}(\beta) w_t(\xi) \right|^2 d\mu(\xi), \\ \overline{Q}_1(\beta) &= \int_{\Xi} \left| E\left(f_{1,t}(\beta)\overline{w}_t(\xi)\right)\right|^2 d\mu(\xi),\end{aligned}$$

$$\begin{aligned}\widehat{\beta}_1 &= \arg\min_{\beta\in\Theta_\beta} \widehat{Q}_1(\beta), \\ \overline{\beta}_1 &= \arg\min_{\beta\in\Theta_\beta} \overline{Q}_1(\beta),\end{aligned}$$

$$\begin{aligned}\widehat{Q}_2(\alpha,\beta) &= \int_{\Xi} \left| \frac{1}{n} \sum_{t=1}^{n} f_{2,t}(\alpha,\beta) w_t(\xi) \right|^2 d\mu(\xi), \\ \overline{Q}_2(\alpha,\beta) &= \int_{\Xi} \left| E\left(f_{2,t}(\alpha,\beta)\overline{w}_t(\xi)\right)\right|^2 d\mu(\xi),\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta}_2 \end{pmatrix} &= \arg \min_{\alpha\in\Theta_\alpha,\beta\in\Theta_\beta} \widehat{Q}_2(\alpha,\beta), \\ \begin{pmatrix} \overline{\alpha} \\ \overline{\beta}_2 \end{pmatrix} &= \arg \min_{\alpha\in\Theta_\alpha,\beta\in\Theta_\beta} \overline{Q}_2(\alpha,\beta).\end{aligned}$$

Then

LEMMA 1: *Under standard regularity conditions,* $p\lim_{n\to\infty}\widehat{\alpha} = \overline{\alpha}$, $p\lim_{n\to\infty}\widehat{\beta}_1 = \overline{\beta}_1$, *and* $p\lim_{n\to\infty}\widehat{\beta}_2 = \overline{\beta}_2$. *If* $H_{0,1}$ *and* $H_{0,2}$ *are true then* $\overline{\alpha} = \alpha_0$ *and* $\overline{\beta}_1 = \overline{\beta}_2 = \beta_0$.

Next, denote

$$
\begin{aligned}
b_1(\xi) &= E\left[\left(\partial f_{1,t}(\overline{\beta}_1)/\partial\overline{\beta}_1'\right)\overline{w}_t(\xi)\right], \\
b_2(\xi) &= E\left[\left(\begin{array}{c}\partial f_{2,t}\left(\overline{\alpha},\overline{\beta}_2\right)/\partial\overline{\alpha}' \\ \partial f_{2,t}\left(\overline{\alpha},\overline{\beta}_2\right)/\partial\overline{\beta}_2'\end{array}\right)\overline{w}_t(\xi)\right],
\end{aligned}
$$

$$
\begin{aligned}
\widehat{b}_1(\xi) &= \frac{1}{n}\sum_{t=1}^{n}\left(\partial f_{1,t}(\widehat{\beta}_1)/\partial\widehat{\beta}_1'\right)w_t(\xi), \\
\widehat{b}_2(\xi) &= \frac{1}{n}\sum_{t=1}^{n}\left(\begin{array}{c}\partial f_{2,t}\left(\widehat{\alpha},\widehat{\beta}_2\right)/\partial\widehat{\alpha}' \\ \partial f_{2,t}\left(\widehat{\alpha},\widehat{\beta}_2\right)/\partial\widehat{\beta}_2'\end{array}\right)w_t(\xi),
\end{aligned}
$$

$$
A_j = \int_{\Xi} b_j(\xi)b_j(\xi)'d\mu(\xi),\ \widehat{A}_j = \int_{\Xi}\widehat{b}_j(\xi)\widehat{b}_j(\xi)'d\mu(\xi),\ j = 1,2,
$$

$$
\begin{aligned}
\Gamma_j(\xi_1,\xi_2) &= E\left[u_{j,t}^2\overline{w}_t(\xi_1)\overline{w}_t(\xi_2)\right],\ j = 1,2, \\
\widehat{\Gamma}_j(\xi_1,\xi_2) &= \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_{j,t}^2 w_t(\xi_1)w_t(\xi_2),\ j = 1,2,
\end{aligned}
$$

where $\widehat{u}_{1,t} = f_{1,t}(\widehat{\beta}_1)$, $\widehat{u}_{2,t} = f_{2,t}\left(\widehat{\alpha},\widehat{\beta}_2\right)$, and

$$
\begin{aligned}
\Omega_j &= \int_{\Xi}\int_{\Xi}\Gamma_j(\xi_1,\xi_2)b_j(\xi_1)b_j(\xi_2)'d\mu(\xi_1)d\mu(\xi_2),\ j = 1,2, \\
\widehat{\Omega}_j &= \int_{\Xi}\int_{\Xi}\widehat{\Gamma}_j(\xi_1,\xi_2)\widehat{b}_j(\xi_1)\widehat{b}_j(\xi_2)'d\mu(\xi_1)d\mu(\xi_2),\ j = 1,2.
\end{aligned}
$$

Then

THEOREM 4: *Under standard regularity conditions and the null hypotheses $H_{0,1}$ and $H_{0,2}$,*

$$
\begin{aligned}
\sqrt{n}(\widehat{\beta}_1 - \beta_0) &\to -A_1^{-1}\int_{\Xi} b_1(\xi)z_1(\xi)d\mu(\xi) \sim N_q(0, A_1^{-1}\Omega_1 A_1^{-1}), \\
\sqrt{n}\left(\begin{array}{c}\widehat{\alpha} - \alpha_0 \\ \widehat{\beta}_2 - \beta_0\end{array}\right) &\to -A_2^{-1}\int_{\Xi} b_2(\xi)z_2(\xi)d\mu(\xi) \sim N_q(0, A_2^{-1}\Omega_2 A_2^{-1}),
\end{aligned}
$$

where $z_j(\xi)$ *is a zero mean Gaussian process on* Ξ *with covariance function* $\Gamma_j(\xi_1, \xi_2)$, $j = 1, 2$. *Moreover,*

$$p \lim_{n\to\infty} \widehat{A}_j^{-1}\widehat{\Omega}_j\widehat{A}_j^{-1} = A_j^{-1}\Omega_j A_j^{-1},\ j = 1, 2.$$

Furthermore, the test statistics of the ICM tests converge in distribution, i.e.,

$$\begin{aligned}\widehat{T}_1 &= \frac{n\widehat{Q}_1(\widehat{\beta}_1)}{\int_\Xi \widehat{\Gamma}_1(\xi,\xi)d\mu(\xi) - \text{trace}\left(\widehat{A}_1^{-1}\widehat{\Omega}_1\right)} \to \frac{\int_\Xi z_1^*(\xi)^2 d\mu(\xi)}{\int_\Xi E\left[z_1^*(\xi)^2\right] d\mu(\xi)} = T_1,\ \textit{say},\\ \widehat{T}_2 &= \frac{n\widehat{Q}_2(\widehat{\alpha},\widehat{\beta}_2)}{\int_\Xi \widehat{\Gamma}_2(\xi,\xi)d\mu(\xi) - \text{trace}\left(\widehat{A}_2^{-1}\widehat{\Omega}_2\right)} \to \frac{\int_\Xi z_2^*(\xi)^2 d\mu(\xi)}{\int_\Xi E\left[z_2^*(\xi)^2\right] d\mu(\xi)} = T_2,\ \textit{say},\end{aligned}$$

where

$$z_j^*(\xi) = z_j(\xi) - b_j(\xi)' A_j^{-1} \int_\Xi b_j(\xi_*) z_j(\xi_*) d\mu(\xi_*),\ j = 1, 2.$$

The null distributions T_j, $j = 1, 2$, *are of the type*

$$T_j \sim \frac{\sum_{i=1}^\infty \lambda_{i,j}\varepsilon_{i,j}^2}{\sum_{i=1}^\infty \lambda_{i,j}},$$

where for each j, *the* $\varepsilon_{i,j}$*'s are i.i.d.* $N(0, 1)$, *and the* $\lambda_{i,j}$*'s are the eigenvalues of the covariance function of* $z_j^*(\xi)$. *Moreover,* $T_j \leq \overline{T}_j$ *a.s., where* $\overline{T}_j = \sup_{N\geq 1} \frac{1}{N}\sum_{i=1}^N \varepsilon_{i,j}^2$.

If under $H_{0,1}$ *and* $H_{0,2}$ *also*

$$E_{t-1}\left[u_{1,t}u_{2,t}\right] = 0\ a.s., \tag{22}$$

then z_1 *and* z_2 *are independent, hence* $\widehat{\beta}_1$ *and* $\left(\widehat{\alpha}, \widehat{\beta}_2\right)$ *are asymptotically independent, and the asymptotic null distributions* T_1 *and* T_2 *of the ICM tests are independent.*

Finally, if the null hypothesis $H_{0,j}$ *is false, then* $p\lim_{n\to\infty} \widehat{T}_j/n > 0$.

The hypothesis (22) can be tested consistently by the ICM test, on the basis of the model $f_{1,t}(\beta_0)f_{2,t}(\alpha_0, \beta_0) = u_{1,t}u_{2,t}$, similarly to model $f_{2,t}(\alpha_0, \beta_0) = u_{2,t}$, but the test involved will not be asymptotically independent of the ICM tests $\widehat{T}_1$ and $\widehat{T}_2$, except if $E_{t-1}\left[u_{1,t}^2 u_{2,t}\right] = 0$ a.s. or $E_{t-1}\left[u_{1,t}u_{2,t}^2\right] = 0$ a.s. However, if one is willing to assume that under $H_{0,1}$ and $H_{0,2}$,

$$(u_{1,t}, u_{2,t}) \text{ is independent of the instruments } x_{t-j},\ j \geq 1, \tag{23}$$

then there is no need for condition (22) in order to achieve independence of the ICM tests, because then we can may rearrange model (21) as follows:

$$\begin{aligned} f_{1,t}(\beta_0) &= u_{1,t}, \\ f^*_{2,t}(\alpha_0, \beta_0|\widehat{\gamma}) &= f_{2,t}(\alpha_0, \beta_0) - \widehat{\gamma} f_{1,t}(\beta_0) = u_{2,t} - \widehat{\gamma} u_{1,t}, \end{aligned}$$

where

$$\widehat{\gamma} = \frac{\sum_{t=1}^n \widehat{u}_{1,t}\widehat{u}_{2,t}}{\sum_{t=1}^n \widehat{u}^2_{1,t}},$$

with $\widehat{u}_{1,t}$ and $\widehat{u}_{2,t}$ determined on the basis of initial consistent estimates of α_0 and β_0. Then under $H_{0,1}$ and $H_{0,2}$,

$$p \lim_{n\to\infty} \widehat{\gamma} = \gamma_0 = \frac{E\left(u_{1,t}u_{2,t}\right)}{E\left(u^2_{1,t}\right)},$$

and consequently $E\left[u_{1,t}\left(u_{2,t} - \gamma_0 u_{1,t}\right)\right] = 0$. The ICM test $\widehat{T}^*_2$ of the null hypothesis

$$H_{0,2}{:}\exists(\alpha_0', \beta_0')' \in \Theta_\alpha \times \Theta_\beta{:}\ f^*_{2,t}(\alpha_0, \beta_0|\gamma_0) = u^*_{2,t},\ E_{t-1}(u^*_{2,t}) = 0\ a.s.,$$

has now a null distribution which is determined by the Gaussian process $z^*_2 = z_2 - \gamma_0 z_1$. Under condition (23), z^*_2 is independent of z_1, because

$$\begin{aligned} E\left[z^*_2(\xi_1)z_1(\xi_2)\right] &= E\left[u_{1,t}u^*_{2,t}\overline{w}_t(\xi_1)\overline{w}_t(\xi_2)\right] \\ &= E\left[u_{1,t}\left(u_{2,t} - \gamma_0 u_{1,t}\right)\right] E\left[\overline{w}_t(\xi_1)\overline{w}_t(\xi_2)\right] = 0, \end{aligned}$$

where the second equality follows from condition (23). Hence the ICM tests $\widehat{T}_1$ and $\widehat{T}^*_2$ are asymptotically independent.

REFERENCES

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: Cambridge University Press

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

Carrasco, M., and J-P. Florens (1997): "Generalization of GMM to a Continuum of Moment Conditions", Working paper, Department of Economics, Ohio State University.

De Jong, R. M. (1996): "On the Bierens Test Under Data Dependence", *Journal of Econometrics* 72, 1-32.

Stinchcombe, M. B., and H. White (1998), "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative", *Econometric Theory* 14, 295-325.