Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes

Herman J. Bierens and Li Wang
Department of Economics and CAPCP*
Pennsylvania State University
University Park, PA 16802

April 29, 2008

Abstract

In this paper we propose a weighted integrated conditional moment (ICM) test of the validity of parametric specifications of conditional distribution models for stationary time series, by extending the weighted ICM test of Bierens (1984) for time series regression models to complete parametric conditional distribution specifications.

*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

1 Introduction

Time series models aim to represent conditional means, moments, and/or conditional distributions relative to the entire past of the time series involved, even if the model employs only a finite number of lagged conditioning variables. The past time series involved refers to all lagged dependent variables, as for example is the case for ARMA models, and possibly all present and past exogenous variables, as for example is the case for ARMAX models. The consistency and asymptotic normality of parameter estimators of time series models require various conditions on the model variables conditional on their infinite past. For instance, the asymptotic normality and asymptotic efficiency of maximum likelihood estimators hinge on the condition that the score vectors are martingale differences relative to their entire past. If the model is only correctly specified conditional on a finite number of past variables rather than on the whole past these results may not hold.

It is possible that the conditional mean or conditional distribution of a time series is correctly specified conditional on a finite number of lagged variables, but is incorrect when the infinite past is conditioned on. We will give an example in section 2. Therefore, to test the validity of a time series model specification consistently, we need to condition on the entire past of the time series involved.

The tricky issue of how to condition on the whole past will be dealt with along the approach in Bierens (1984), by conducting a sequence of ICM tests $\widehat{B}_{n,m}$, say, where m is the number of lagged conditioning variables involved and n is the number of observations of the (vector) time series Y_t involved. Each ICM test $\widehat{B}_{n,m}$ is conducted similar to Bierens and Wang (2008) for the i.i.d. case, with $(Y_{t-1}', Y_{t-2}', ..., Y_{t-m}')'$ the vector of conditioning variables. Thus, $\widehat{B}_{n,m}$ is based on the integrated squared difference between the empirical characteristic function of $(Y_t', Y_{t-1}', Y_{t-2}', ..., Y_{t-m}')'$ and the corresponding empirical characteristic function implied by the estimated conditional distribution model for Y_t. Given an arbitrary $\alpha \in (0,1)$ and a subsequence ℓ_n of the sample size n, the actual weighted ICM (WICM) test statistic is

$$\widehat{W}_n = \sum_{m=1}^{\ell_n} \alpha^m \widehat{B}_{n,m}.$$

The asymptotic null distribution of this test is case dependent. Therefore, critical values and/or p-values have to be derived via a bootstrap method.

Since this test is based on characteristic functions, it has the unique advantage that it is applicable to any type of conditional distribution; continuous, discrete or mixed continuous-discrete (for example Tobit type models), as long as the time series involved are strictly stationary. With some modifications this test can even handle singular conditional distributions, for example stochastic dynamic general equilibrium macro-economic models. This test is consistent against all stationary alternatives and has nontrivial power against $\sqrt{n}$-local alternatives. To the best of our knowledge no other consistent test for parametric conditional time series distributions has been proposed yet in the literature, despite consistency claims made by some authors.

Conditional characteristic functions often do not have a closed form expression and then have to be computed numerically. To avoid this computational burden, we propose a Weighted Simulated ICM (WSIMC) test where the conditional characteristic function of the estimated model is replaced with an simulated counterpart based on a single random drawing from this conditional distribution. The WSICM test has an easy-to-compute closed-form expression, and all theoretical properties of the exact WICM test carry over.

This paper is organized as follows. Section 2 reviews the literature on time series specification testing. In Section 3 we state the maintained hypothesis on the data generating process and the parametric model. In Section 4 we discuss the identification of the alternative hypothesis via characteristic functions. In Section 5 we derive the asymptotic properties of our test under the null hypothesis. A simulated version of our test is proposed in Section 5. A limited Monte Carlo study will be presented in Section 6. Finally, in Section 7 we will make some concluding remarks and propose directions for further research.

As to notations, the indicator function will be denoted by $I(.)$, the vector norm $||x||$ is the Euclidean norm if $x \in \mathbb{R}^d$ and $||x|| = \sqrt{x'\overline{x}}$ if $x \in \mathbb{C}^d$, where the bar denotes the complex conjugate. In the case $x = a + i.b \in \mathbb{C}$ this norm becomes the absolute value: $|x| = \sqrt{x.\overline{x}} = \sqrt{a^2 + b^2}$. The matrix norm $||A||$ is the maximum absolute value of the elements involved, regardless whether the elements of A are real or complex valued. Finally, we adopt the convention that the derivative of a function to a row vector is a column vector of partial derivatives, e.g., $\partial\left(x'Ax\right)/\partial x' = 2Ax$, $\partial\left(x'Ax\right)/\partial x = x'A$.

2 Literature Review

Recall that a test is called consistent if its power against any deviation of the null hypothesis approaches one as the sample size goes to infinity. However, most test require some maintained hypotheses on the data, so that the consistency concept is relative to these maintained hypotheses. For example, a time series specification test may be consistent against all stationary alternatives but not against nonstationary alternatives.

The first consistent test for the specification of functional form of cross-section regression models was proposed by Bierens (1982), and later named by Bierens and Ploberger (1997) the Integrated Conditional Moment (ICM) test. The key idea of the ICM test is that the null hypothesis is transformed to a testable sufficient and necessary equivalent hypothesis consisting of an infinite number of orthogonality conditions formed by products of model errors and special weight functions of the explanatory variables. The features of these weight functions are characterized by Stinchcombe and White (1998). The ICM test was generalized to time series regression models by Bierens (1984), De Jong (1996) and Bierens and Ploberger (1997).

A necessary condition for the consistency of tests of time series hypotheses is that the information set conditioned on contains the entire past of the time series involved. In particular, for testing the functional form of time series regression models this condition implies that the null hypothesis involved is that the model errors are martingale differences with respect to the σ-algebra generated by this information set. For example, consider the AR(1) model

$$Y_t = \alpha + \beta Y_{t-1} + U_t, \;\; |\beta| < 1. \tag{1}$$

The condition for the validity of this model as the best one-step-ahead forecasting scheme for Y_t is that U_t is a martingale difference process with respect to the σ-algebra $\mathcal{F}_{-\infty}^{t-1} = \sigma\left(\{Y_{t-j}\}_{j=1}^{\infty}\right)$ generated by the information set $\mathcal{I}_{t-t} = \{Y_{t-j}, \; j \geq 1\}$, i.e.,

$$E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0 \text{ a.s.}, \tag{2}$$

and thus $E\left[Y_t|\mathcal{F}_{-\infty}^{t-1}\right] = \alpha + \beta Y_{t-1}$. Of course, the latter implies that also $E\left[Y_t|Y_{t-1}\right] = \alpha + \beta Y_{t-1}$ a.s., but not the other way around. We will discuss the literature using this AR(1) model as an example of the null hypothesis.

Most specification tests for regression-type time series models proposed in the statistical and econometric literature, including Bierens and Ploberger

(1997), only test implications of the martingale difference hypothesis rather than this hypothesis itself. To the best of our knowledge the only two exceptions are the ICM tests of Bierens (1984) and De Jong (1996).

Bierens (1984) proposed to compute a sequence of ICM test statistics $\widehat{B}_m$ of the null hypotheses

$$E\left[U_t|\mathcal{F}_{t-m}^{t-1}\right]=0 \text{ a.s.}, \tag{3}$$

where $\mathcal{F}_{t-m}^{t-1}=\sigma\left(\{Y_{t-j}\}_{j=1}^{m}\right)$, and then use $\sum_{m=1}^{\ell_n}\omega_m\widehat{B}_m$ as the actual test statistic, where ω_m is a sequence of positive weights satisfying $\sum_{m=1}^{\infty}\omega_m<\infty$ and ℓ_n is a subsequence of n.

De Jong (1996) has extended the approaches in Bierens' (1982, 1990) to an ICM test of the martingale difference hypothesis (2), as follows. He identifies the null hypothesis (2) versus the alternative

$$\Pr\left(E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right]=0\right)<1 \tag{4}$$

via the contents of a set $S\subset\mathbb{R}^{\infty}$ of the type

$$\mathbf{S}=\left\{\boldsymbol{\xi}=(\xi_1',\xi_2',\xi_3',....)'\in\boldsymbol{\Xi}:E\left[U_t\exp\left(\sum_{j=1}^{\infty}\xi_j'\Psi(Y_{t-j})\right)\right]=0\right\},$$

where $\boldsymbol{\Xi}$ is a compact metric space in $\mathbb{R}^{\infty}$, and Ψ is a bounded one-to-one mapping. In particular, de Jong specifies $\boldsymbol{\Xi}=\times_{j=1}^{\infty}\left[-c.j^{-2},c.j^{-2}\right]^k$ for some constant $c>0$, where k is the dimension of $\Psi(Y_{t-j})$. Under the null hypothesis (2), $\mathbf{S}=\boldsymbol{\Xi}$, whereas under the alternative, $\mathbf{S}$ is "almost empty". Therefore, a consistent ICM test of the null hypothesis (2) can be based on the integral

$$\int_{\boldsymbol{\Xi}}\left(\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\widehat{U}_t\exp\left(\sum_{j=1}^{t-1}\xi_j'\Phi(Y_{t-j})\right)\right)^2 d\boldsymbol{\xi},$$

where the $\widehat{U}_t$'s are the regression residuals and n is the sample size.

Hong (1999) proposed a test for time series independence using a generalized spectral density, where the autocorrelation function in the standard spectral density is replaced by the difference between the joint characteristic function and the product of two marginal characteristic functions. If there is pairwise independence, then these differences are zero. Su and White (2007) also use characteristic functions in testing serial independence. Hong and Lee (2005) test pair-wise independence of the regression errors, using the approach in Hong (1999). However, independence of regression errors is too strong a condition for model validity because the only requirement for correctness of conditional mean time series models is that the model errors are martingale differences. Moreover, pairwise independence does not imply the martingale difference hypothesis.

Escanciano and Velasco (2006) propose to test the martingale difference hypothesis using the same pairwise implications as those in Hong (1999). The generalized spectral density they use is based on the covariance between the regression errors U_t and particular functions of each of the lagged conditioning variables Y_{t-m}. Thus, these authors test the null hypothesis $\sup_{m\geq 1}|E[U_t|Y_{t-m}]| = 0$ a.s., rather than the martingale difference hypothesis itself.

Dominguez and Lobato (2003) and Stute et al. (2006) propose tests of the hypothesis (3) for fixed m based on moment conditions of the form $E\left[U_t\prod_{j=1}^m I(Y_{t-j} \leq y_j)\right] = 0$ for all conformable nonrandom vectors y_j.

Before discussing the literature on testing the validity of parametric conditional distribution specifications for time series data, let us explain first what we mean by "validity", on the basis of the AR(1) model (1) augmented with the assumption $U_t|Z_{t-1} \sim N\left[0, \sigma^2\right]$. The conditional distribution of this model given Y_{t-1} takes the form

$$\begin{aligned} G_{t-1}(y|\theta) &= \sigma^{-1}\Phi\left((y-\alpha-\beta Y_{t-1})/\sigma\right) \\ \theta &= (\alpha, \beta, \sigma)', \end{aligned} \tag{5}$$

where Φ is the c.d.f. of the standard normal distribution. This functional specification is correct for any stationary Gaussian process Y_t because then $(Y_t, Y_{t-1})'$ has a bivariate normal distribution. As is well-known, in this case $E\left[Y_t|Y_{t-1}\right]$ is linear in Y_{t-1}, say $E\left[Y_t|Y_{t-1}\right] = \alpha + \beta Y_{t-1}$, $U_t = Y_t - E\left[Y_t|Y_{t-1}\right] \sim N\left[0, \sigma^2\right]$ for some σ, and U_t and Y_{t-1} are independent. However, in general $\Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right] \neq \Phi\left((y-\alpha-\beta Y_{t-1}-)/\sigma\right)$. For example, let Y_t be the MA(1) process $Y_t = V_t - \gamma V_{t-1}$ with $|\gamma| < 1$, and V_t Gaussian white noise with variance σ_V^2. Then

$$F_{t-1}(y) = \Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right] = \sigma_V^{-1}\Phi\left(\frac{y-\sum_{j=1}^{\infty}\gamma^j Y_{t-j}}{\sigma_V}\right) \tag{6}$$

and $G_{t-1}(y|\theta) = \Pr\left[Y_t \leq y|Y_{t-1}\right]$.

In testing dynamic distribution specifications, White (1987) used the fact that if the distribution is correctly specified, then the negative of the Fisher information matrix is equal to the variance of score function. Note that this equality is just an implication of the null hypothesis. Hence, accepting this equality does not necessarily mean that the null hypothesis is true, rendering this test inconsistent.

Bai's (2003) test of the validity of conditional distribution models for time series is based on the well-known fact that for a univariate time series process Y_t with absolutely continuous conditional distribution of the type (6), $U_t = F_{t-1}(Y_t)$ is independent uniformly $[0,1]$ distributed. Therefore, given the specification $G_{t-1}(y|\theta)$ of $F_{t-1}(y)$, Bai proposes a Kolmogorov-type test based on an empirical process of the form

$$\widehat{V}_n(u) = (1/\sqrt{n}) \sum_{t=1}^{n} \left[I\left(G_{t-1}(Y_t|\widehat{\theta}) \leq u \right) - u \right], \ u \in [0,1],$$

where $\widehat{\theta}$ is a (quasi-) maximum likelihood estimator. To get an asymptotically distribution free test, Bai uses the Khmaladze (1981) martingale transformation, which yields a correction term $\widehat{K}_n$, say, such that under the null hypothesis,

$$V_n(u) = \widehat{V}_n(u) - \widehat{K}_n = (1/\sqrt{n}) \sum_{t=1}^{n} \left[I\left(U_t \leq u\right) - u \right] + o_p(1),$$

where $\theta = p\lim_{n\to\infty} \widehat{\theta}$. Under the null hypothesis, V_n converges weakly to a standard Brownian bridge. However, in the case of the incorrect null model (5) $U_t = G_{t-1}(Y_t|\theta)$ is also uniformly [0,1] distributed, but no longer independent. Then under some regularity conditions, V_n still converges weakly to a limit process, although due to the dependence of U_t this limit process is no longer a standard Brownian bridge. Thus, Bai's test is not consistent. One of the reasons for this inconsistency is that the independence condition for the U_t's is not part of the test, as only the uniformity condition is tested. Another reason is given in Bierens and Wang (2008).

Bai and Chen (2007) have extended Bai's (2003) test to vector time series processes. Corradi and Swanson (2006) use the same uniform transformation as in Bai (2003) to extend the conditional Kolmogorov test to time series. But instead of using the Khmaladze (1981) martingale transformation to get an asymptotically distribution free test, they used bootstrap critical values.

Li and Tkacz (2006) propose a specification test based on a comparison of a parametric conditional density with a nonparametrically estimated conditional density function, weighted with a nonparametric kernel estimator of the density of the (finite-dimensional) vector of conditioning variables. They claim consistency of their test, even in the title of their paper.[1]

To the best of our knowledge there does not yet exist a test for the validity of parametric distributions for time series data that is consistent against all stationary alternatives. In this paper we will propose such a test.

3 Data Generating Process and Model

Throughout we will assume that

Assumption 1. *The data generating process Y_t is a strictly stationary p-variate vector time series process defined on a common probability space $\{\Omega, \mathcal{F}, P\}$, with a vanishing memory.*

The latter concept is defined in Bierens (2004, Ch. 7) as follows.

Definition 1. *Denote by $\mathcal{F}_{t-m}^{t-1}$ the σ-algebra generated by $Y_{t-1}, Y_{t-2}, ..., Y_{t-m}$: $\mathcal{F}_{t-m}^{t-1} = \sigma\left(Y_{t-1}, Y_{t-2}, ..., Y_{t-m}\right)$, and let $\mathcal{F}_{-\infty}^{t-1} = \sigma\left(\cup_{m=1}^{\infty}\mathcal{F}_{t-m}^{t-1}\right)$, which is the σ-algebra generated by $\{Y_{t-j}\}_{j=1}^{\infty}$. Then $\mathcal{F}_{-\infty} = \cap_t \mathcal{F}_{-\infty}^{t-1}$ is the remote σ-algebra involved. The time series process Y_t has a vanishing memory if for all sets $A \in \mathcal{F}_{-\infty}$, either $P(A) = 1$ or $P(A) = 0$.*

As is well known from Kolmogorov's zero-one law, independent processes have a vanishing memory in this sense, but this property carries over to quite general stationary processes. See for example Bierens (2004, Ch. 7). Moreover, under Assumption 1,

$$p\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} Y_t = E\left[Y_1\right], \text{ provided that } E\left[\|Y_1\|\right] < \infty.$$

See Bierens (2004, Ch.7). Furthermore, under Assumption 1 the stochastic properties of Y_t are completely determined by the conditional distribution

[1] Other authors who makes unjustified consistency claims are Li (1999) and Chen and Fan (1999).

function

$$F_{t-1}(y) = E\left[I(Y_t \le y) \left| \mathcal{F}_{-\infty}^{t-1}\right.\right], \; y \in \mathbb{R}^p$$

Let $G_{t-1}(y|\theta)$, $\theta \in \Theta$, be a family of parametric distributions of Y_t conditional on $\mathcal{F}_{-\infty}^{t-1}$, where $\Theta \subset \mathbb{R}^k$ is a compact parameter space. Note that the conditional distribution functions $G_{t-1}(y|\theta)$ may depend on the entire sequence $\{Y_{t-j}\}_{j=1}^{\infty}$, as is the case for MA models, but for the time being we will ignore this problem. Moreover, it is reasonable to assume that Θ is chosen such that

Assumption 2. *For all $\theta \in \Theta$ the support of $G_{t-1}(y|\theta)$ is the same as the support of $F_{t-1}(y)$.*

The null and alternative hypotheses involved are

$$H_0\text{:} \quad \begin{array}{l}\text{There exists an interior point } \theta_0 \in \Theta \text{ such that} \\ G_{t-1}(y|\theta_0) = F_{t-1}(y) \text{ a.s. for all } y \in \mathbb{R}^p,\end{array} \tag{7}$$

$$H_1\text{:} \quad \begin{array}{l}\text{For all } \theta \in \Theta \text{ there exists an } y \in \mathbb{R}^p \text{ such that} \\ \Pr\left[G_{t-1}(y|\theta) = F_{t-1}(y)\right] < 1.\end{array} \tag{8}$$

respectively. It will be assumed that θ_0 has been estimated by maximum likelihood (ML), with ML estimator $\widehat{\theta}$, and that under H_0 all the conditions for consistency and asymptotic normality of $\widehat{\theta}$ are satisfied. In particular

Assumption 3. *Under the null hypothesis* (7),

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = -\Sigma^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\right) + o_p\left(1\right)$$

where $U_t \in \mathbb{R}^k$ is a martingale difference process with respect to. the filtration $\mathcal{F}_{-\infty}^{t-1}$, satisfying the conditions of the martingale difference central limit theorem:[2]

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t \xrightarrow{d} N_k\left[0, \Sigma\right], \; \det\left(\Sigma\right) > 0.$$

[2]See for example McLeish (1974).

The U_t's are of course the vectors of scores of the log-likelihood $\ln L_n(\theta)$, with

$$\Sigma = -\lim_{n\to\infty} n^{-1} E\left[\partial^2 \ln L_n(\theta) / (\partial\theta\partial\theta')\big|_{\theta=\theta_0}\right] = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E\left[U_t U_t'\right].$$

Under H_1 the estimator $\widehat{\theta}$ is a Quasi ML (QML) estimator. It is pretty standard to set forth mild condition such that $\widehat{\theta}$ converges in probability to a point in Θ, namely the point

$$\theta_* = \arg\max_{\theta\in\Theta} \lim_{n\to\infty} E\left[\ln L_n(\theta)/n\right].$$

Therefore, we assume that

Assumption 4. *Under the alternative hypothesis* (8), $p\lim_{n\to\infty}\widehat{\theta} = \theta_*$.

The assumption that $\widehat{\theta}$ is a (quasi-) ML estimator is not essential. Any estimator satisfying Assumption 3 will do, for example, GMM estimators.

4 Identifying the Alternative Hypothesis Via Empirical Characteristic Functions

The null and alternative hypotheses can, in theory, be identified via the conditional characteristic functions of $G_{t-1}(y|\theta)$ and $F_{t-1}(y)$:

$$\varphi_{t-1}(\tau|\theta) = \int_{\mathbb{R}^p} \exp(i.\tau' y) dG_{t-1}(y|\theta), \tag{9}$$

$$\psi_{t-1}(\tau) = \int_{\mathbb{R}^p} \exp(i.\tau' y) dF_{t-1}(y) = E\left[\exp(i.\tau' Y_t)\left|\mathcal{F}_{-\infty}^{t-1}\right.\right] \tag{10}$$

respectively. As is well known, H_0 is true if and only if $\varphi_{t-1}(\tau|\theta_0) \equiv \psi_{t-1}(\tau)$ a.s. for all $\tau \in \mathbb{R}^p$, whereas under H_1, $\inf_{\theta\in\Theta}\sup_{\tau\in\mathbb{R}^p}|\varphi_{t-1}(\tau|\theta) - \psi_{t-1}(\tau)| > 0$ a.s. Moreover, if Y_t is bounded then the latter is true if and only if in an arbitrary open neighborhood N_0 of the origin of $\mathbb{R}^p$,

$$\inf_{\theta\in\Theta}\sup_{\tau\in N_0}|\varphi_{t-1}(\tau|\theta) - \psi_{t-1}(\tau)| > 0 \text{ a.s.},$$

due to the well-known fact that characteristic functions of bounded random variables [or vectors] coincide everywhere if they coincide in an arbitrary neighborhood of zero [or the zero vector]. Therefore, for the time being we will assume that Y_t is a bounded time series process, because then we know where to look for possible discrepancies between $\varphi_{t-1}(\tau|\theta)$ and $\psi_{t-1}(\tau)$.

However, although $\varphi_{t-1}(\tau|\theta)$ can be determined from the model distribution $G_{t-1}(y|\theta)$, it is difficult if not impossible to estimate $\psi_{t-1}(\tau)$ consistently. The following lemma provides a solution to this problem.

Lemma 1. *Let Assumption 1 hold, with Y_t a bounded process: there exists an $M \in (0,\infty)$ such that* $\Pr[\|Y_t\| \leq M] = 1$. *Denote*

$$\begin{aligned}
\varphi^{m+1}(\tau|\theta) &= E\left[\int_{\mathbb{R}^p} \exp(i.\tau_0' y) dG_{t-1}(y|\theta) \exp\left(i\sum_{j=1}^{m} \tau_j' Y_{t-j}\right)\right] \\
\psi^{m+1}(\tau) &= E\left[\exp\left(i\sum_{j=0}^{m} \tau_j' Y_{t-j}\right)\right] \\
\tau &= (\tau_0', \tau_1', ..., \tau_m')' \in \times_{j=0}^{m} \Upsilon \\
S_{m+1} &= \left\{\tau \in \times_{j=0}^{m}\Upsilon : \left|\varphi^{m+1}(\tau|\theta_*) - \psi^{m+1}(\tau)\right| > 0\right\}
\end{aligned}$$

where $\Upsilon \subset \mathbb{R}^p$ is a hypercube centered around the origin of $\mathbb{R}^p$ and θ_* is defined by Assumption 4. Under H_1, for all but a finite number of m's, S_{m+1} has positive Lebesgue measure.

Proof: Appendix.

Of course, under H_0 the Lebesgue measure of S_{m+1} is zero.

This result suggests that a test for H_0 can be based on the empirical counterparts of $\varphi^{m+1}(\tau|\theta)$ and $\psi^{m+1}(\tau)$:

$$\begin{aligned}
\widehat{\varphi}^{m+1}(\tau|\theta) &= \frac{1}{n}\sum_{t=1}^{n} \int_{\mathbb{R}^p} \exp(i.\tau_0' y) dG_{t-1}(y|\theta) \exp\left(i\sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \\
\widehat{\psi}^{m+1}(\tau) &= \frac{1}{n}\sum_{t=1}^{n} \exp\left(i\sum_{j=0}^{m} \tau_j' Y_{t-j}\right)
\end{aligned}$$

In particular, if

Assumption 5. *For each* $\tau_0 \in \Upsilon$, $\int_{\mathbb{R}^p} \exp(i.\tau_0' y) dG_{t-1}(y|\theta)$ *is a.s. continuous in* $\theta \in \Theta$, *where* Υ *and* Θ *are compact,*

then by the uniform weak law of large numbers for strictly stationary time series with vanishing memory (see Bierens 2004, Theorem 7.8(b), p. 189),

$$
\begin{aligned}
p \lim_{n\to\infty} \sup_{\tau \in \times_{j=0}^m \Upsilon} \left| \widehat{\psi}^{m+1}(\tau) - \psi^{m+1}(\tau) \right| &= 0 \\
p \lim_{n\to\infty} \sup_{\tau \in \times_{j=0}^m \Upsilon, \theta \in \Theta} \left| \widehat{\varphi}^{m+1}(\tau|\theta) - \varphi^{m+1}(\tau|\theta) \right| &= 0
\end{aligned}
$$

hence by Assumption 4,[3]

$$
p \lim_{n\to\infty} \sup_{\tau \in \times_{j=0}^m \Upsilon} \left| \widehat{\varphi}^{m+1}(\tau|\widehat{\theta}) - \varphi^{m+1}(\tau|\theta_*) \right| = 0
$$

and thus

Lemma 2. *Under Assumptions 1-5, the boundedness condition in Lemma 1 and* H_1,

$$
p \lim_{n\to\infty} \int_{\times_{j=0}^m \Upsilon} \left| \widehat{\varphi}^{m+1}(\tau|\widehat{\theta}) - \widehat{\psi}^{m+1}(\tau) \right|^2 d\tau > 0
$$

for all but a finite number of m*'s.*

5 The Weighted ICM Test and its Asymptotic Null Distribution

Consider the empirical process

$$
\begin{aligned}
\widehat{h}_{n,m}(\tau) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\exp(i.\tau_0' Y_t) - \varphi_{t-1}(\tau_0|\widehat{\theta}) \right) \exp\left(i \sum_{j=1}^m \tau_j' Y_{t-j} \right), \quad (11) \\
\tau &= (\tau_0', \tau_1', ..., \tau_m')' \in \times_{j=0}^m \Upsilon,
\end{aligned}
$$

[3] Together with the measurability conditions in Bierens (2004, Theorem 7.8(b), condition (a)), which we will not make explicit.

where $\varphi_{t-1}(\tau_0|\widehat{\theta})$ is defined by (9) and Υ is a compact set in $\mathbb{R}^p$.

For given m let

$$\widehat{B}_{n,m} = \int_{\times_{j=0}^{m}\Upsilon} \left|\widehat{h}_{n,m}(\tau)\right|^2 d\mu_m(\tau)$$

where $\mu_m(\tau)$ is the uniform probability measure on $\times_{j=0}^{m}\Upsilon$, i.e.

$$d\mu_m(\tau) = \frac{d\tau}{\int_{\times_{j=0}^{m}\Upsilon} d\tau}$$

5.1 Weak convergence

In this subsection it will be shown that under H_0, and for fixed m, $\widehat{h}_{n,m} \Rightarrow h_m$, where $h_m(\tau)$ is a zero-mean Gaussian process on $\times_{j=0}^{m}\Upsilon$, so that by the continuous mapping theorem,

$$\widehat{B}_{n,m} \xrightarrow{d} B_m = \int_{\times_{j=0}^{m}\Upsilon} |h_m(\tau)|^2 d\mu_m(\tau). \tag{12}$$

As to the general notion of weak convergence, consider a sequence of random elements $h_n(\beta)$ of a metric space $\mathcal{C}(\mathbf{B})$ of functions on a compact subset $\mathbf{B}$ of an Euclidean space. In our case $\mathcal{C}(\mathbf{B})$ is the metric space of complex-valued continuous functions on $\mathbf{B} = \times_{j=0}^{m}\Upsilon$, endowed with the "sup" metric. Weak convergence can be defined in various equivalent ways[4], but the one that delivers the result (12) directly is the following: h_n converges weakly to h, $h_n \Rightarrow h$, if for all bounded continuous real functions f on $\mathcal{C}(\mathbf{B})$,

$$\lim_{n\to\infty} E[f(h_n)] = E[f(h)]. \tag{13}$$

For example, let for $h \in \mathcal{C}(\mathbf{B})$, $f(h) = \gamma\left(\int_{\mathbf{B}} |h(\beta)|^2 d\mu(\beta)\right)$, where μ is a probability measure on $\mathbf{B}$ and γ is an arbitrary bounded continuous real function on $\mathbb{R}$. Then (13) implies

$$\int_{\mathbf{B}} |h_n(\beta)|^2 d\mu(\beta) \xrightarrow{d} \int_{\mathbf{B}} |h(\beta)|^2 d\mu(\beta).$$

See, for example, Theorem 6.18 in Bierens (2004).

[4] See for example Billingsley (1968) or Van der Vaart and Wellner (1996).

The necessary and sufficient conditions for weak convergence are that h_n is tight and the finite distributions of h_n converge. The latter means that for arbitrary $\beta_1, \beta_2, ..., \beta_k$ in $\mathbf{B}$,

$$(h_n(\beta_1), h_n(\beta_2), ..., h_n(\beta_k)) \xrightarrow{d} (h(\beta_1), h(\beta_2), ..., h(\beta_k)) . \tag{14}$$

The tightness concept is a generalization of the stochastic boundedness concept for sequences of random variables: For each $\varepsilon \in (0,1)$ there exists compact set $K \subset \mathcal{C}(\mathbf{B})$ such that $\inf_{n\geq 1} \Pr[h_n \in K] > 1 - \varepsilon$.

According to Billingsley (1968, Theorem 8.2), the following two conditions are sufficient for the tightness of h_n:
(a) For each $\eta > 0$ and each $\beta \in \mathbf{B}$ there exists a $\delta > 0$ such that

$$\sup_{n\geq 1} \Pr[|h_n(\beta)| > \delta] \leq \eta \tag{15}$$

(b) For each $\eta > 0$ and $\delta > 0$ there exists an $\varepsilon > 0$ such that

$$\sup_{n\geq 1} \Pr\left[\sup_{||\beta_1-\beta_2||<\varepsilon} |h_n(\beta_1) - h_n(\beta_1)| \geq \delta\right] \leq \eta. \tag{16}$$

Condition (a) is a pointwise stochastic boundedness condition, which holds if for each $\beta \in \mathbf{B}$, $h_n(\beta)$ converges in distribution, hence this condition follows from the condition (14). Condition (b) is also known as the stochastic equicontinuity condition, which is the difficult part of the tightness proof.

5.2 Eliminating the ML Estimator

To prove $\widehat{h}_{n,m} \Rightarrow h_m$ we first need to get rid of the ML estimator $\widehat{\theta}$ in the expression (11), using Assumption 3, as follows. Write (11) as

$$\widehat{h}_{n,m}(\tau) = \widehat{h}_{1,n,m}(\tau) - \widehat{h}_{2,n,m}(\tau|\widehat{\theta}),$$

where

$$\begin{aligned}\widehat{h}_{1,n,m}(\tau) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\exp\left(i.\tau_0'Y_t\right) - \varphi_{t-1}\left(\tau_0|\theta_0\right)\right), \\ &\quad \times \exp\left(i\sum_{j=1}^{m}\tau_j'Y_{t-j}\right)\end{aligned} \tag{17}$$

$$
\begin{aligned}
\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\varphi_{t-1} \left(\tau_0 | \widehat{\theta} \right) - \varphi_{t-1} \left(\tau_0 | \theta_0 \right) \right) \\
&\quad \times \exp \left(i \sum_{j=1}^{m} \tau_j' Y_{t-j} \right)
\end{aligned} \tag{18}
$$

Next, assume that

Assumption 6. *Under H_0 the conditional characteristic function $\varphi_{t-1}(\tau|\theta)$ defined by (9) is a.s. twice continuously differentiable on an open neighborhood $\Theta_0 \subset \Theta$ of θ_0 with vector of first derivatives satisfying*

$$
\begin{aligned}
E \left[\sup_{\tau_0 \in \Upsilon, \theta \in \Theta_0} \left| \partial \varphi_{t-1} \left(\tau_0 | \theta \right) / \partial \theta_j \right| \right] &< \infty, \\
E \left[\sup_{\tau_0 \in \Upsilon, \theta \in \Theta_0} \left| \partial^2 \varphi_{t-1} \left(\tau_0 | \theta \right) / \left(\partial \theta_{j_1} \partial \theta_{j_2} \right) \right| \right] &< \infty,
\end{aligned}
$$

for $j, j_1, j_2 = 1, 2, ..., k$.

Denote

$$
\Delta \varphi_{t-1} \left(\tau | \theta \right) = \frac{\partial \varphi_{t-1} \left(\tau | \theta \right)}{\partial \theta'}, \ \Delta^2 \varphi_{t-1} \left(\tau | \theta \right) = \frac{\partial^2 \varphi_{t-1} \left(\tau | \theta \right)}{\partial \theta \partial \theta'}.
$$

It follows from Assumptions 3 and 6 and Taylor's theorem that

$$
\begin{aligned}
&\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) = \sqrt{n} \left(\widehat{\theta} - \theta_0 \right)' \frac{1}{n} \sum_{t=1}^{n} \Delta \varphi_{t-1} \left(\tau_0 | \theta_0 \right) \exp \left(i \sum_{j=1}^{m} \tau_j' Y_{t-j} \right) \\
&+ \frac{1}{2} \sqrt{n} \left(\widehat{\theta} - \theta_0 \right)' \frac{1}{n} \sum_{t=1}^{n} \left(\mathrm{Re} \left[\Delta^2 \varphi_{t-1} \left(\tau_0 | \widetilde{\theta}_1 \right) \exp \left(i \sum_{j=1}^{m} \tau_j' Y_{t-j} \right) \right] \right. \\
&\times \left(\widehat{\theta} - \theta_0 \right) \\
&+ i. \frac{1}{2} \sqrt{n} \left(\widehat{\theta} - \theta_0 \right)' \frac{1}{n} \sum_{t=1}^{n} \mathrm{Im} \left[\Delta^2 \varphi_{t-1} \left(\tau_0 | \widetilde{\theta}_2 \right) \exp \left(i \sum_{j=1}^{m} \tau_j' Y_{t-j} \right) \right] \\
&\times \left(\widehat{\theta} - \theta_0 \right)
\end{aligned} \tag{19}
$$

where $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ are mean values satisfying $\left\|\widetilde{\theta}_j - \theta_0\right\| \leq \left\|\widehat{\theta} - \theta_0\right\|$, $j = 1, 2$. Hence

$$\left|\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) - \sqrt{n}\left(\widehat{\theta} - \theta_0\right)' \frac{1}{n}\sum_{t=1}^{n} \Delta\varphi_{t-1}\left(\tau_0|\theta_0\right) \exp\left(i\sum_{j=1}^{m} \tau_j' Y_{t-j}\right)\right|$$
$$\leq \frac{1}{2\sqrt{n}} \left\|\sqrt{n}\left(\widehat{\theta} - \theta_0\right)\right\|^2 \frac{1}{n}\sum_{t=1}^{n} \sup_{\tau_0\in\Upsilon,\theta\in\Theta_0} \left\|\Delta^2\varphi_{t-1}\left(\tau_0|\theta\right)\right\|$$
$$+O_p\left(1/\sqrt{n}\right) = O_p\left(1/\sqrt{n}\right)$$

Note that the first O_p term is due to

$$\lim_{n\to\infty} \Pr\left[\widetilde{\theta}_1 \in \Theta_0\right] = 1, \lim_{n\to\infty} \Pr\left[\widetilde{\theta}_2 \in \Theta_0\right] = 1$$

and the second O_p term is due to the fact that by Assumptions 1 and 6 the weak law of large numbers applies:

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\tau_0\in\Upsilon,\theta\in\Theta_0} \left\|\Delta^2\varphi_{t-1}\left(\tau_0|\theta\right)\right\| \xrightarrow{p} E\left[\sup_{\tau_0\in\Upsilon,\theta\in\Theta_0} \left\|\Delta^2\varphi_{t-1}\left(\tau_0|\theta\right)\right\|\right].$$

Moreover, using Theorem 7.8(b) in Bierens (2004) it can be shown that

$$\frac{1}{n}\sum_{t=1}^{n} \Delta\varphi_{t-1}\left(\tau_0|\theta_0\right) \exp\left(i\sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \xrightarrow{p} b_m(\tau|\theta) \tag{20}$$
$$= E\left[\Delta\varphi_{t-1}\left(\tau_0|\theta\right) \exp\left(i\sum_{j=1}^{m} \tau_j' Y_{t-j}\right)\right]$$

uniformly in $\tau = (\tau_0', \tau_1', ..., \tau_m')' \in \times_{j=0}^{m}\Upsilon$. Thus it follows from Assumption 3 that

$$\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) = -b_m(\tau|\theta_0)'\Sigma^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\right) + o_p\left(1\right) \tag{21}$$

where the $o_p\left(1\right)$ term is uniform in $\tau \in \times_{j=0}^{m}\Upsilon$.

Combining the results (17) and (21), $\widehat{h}_{n,m}(\tau)$ can be written as

$$\widehat{h}_{n,m}(\tau) = \widetilde{h}_{n,m}(\tau) + o_p\left(1\right)$$

where

$$\widetilde{h}_{n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \phi_{m,t}(\tau) \tag{22}$$

with

$$\begin{aligned} \phi_{m,t}(\tau) &= \left(\exp\left(i.\tau_0' Y_t\right) - \varphi_{t-1}\left(\tau_0|\theta_0\right)\right) \exp\left(i \sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \\ &\quad + b_m(\tau|\theta_0)' \Sigma^{-1} U_t. \end{aligned} \tag{23}$$

5.3 Tightness and Convergence Results

Note that pointwise in $\tau \in \times_{j=0}^{m} \Upsilon$, $\phi_{m,t}(\tau)$ is a (complex-valued) martingale difference process, i.e., $\phi_{m,t}(\tau)$ is measurable $\mathcal{F}_{-\infty}^{t}$ and $E\left[\phi_{m,t}(\tau)|\mathcal{F}_{-\infty}^{t-1}\right] = 0$ a.s., hence by the martingale difference central limit theorem (see McLeish 1974),

$$\begin{pmatrix} \operatorname{Re}\left[\widetilde{h}_{n,m}(\tau)\right] \\ \operatorname{Im}\left[\widetilde{h}_{n,m}(\tau)\right] \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \operatorname{Re}\left[h_m(\tau)\right] \\ \operatorname{Im}\left[h_m(\tau)\right] \end{pmatrix}$$

for fixed m and $n \to \infty$, where the latter is a bivariate zero-mean random vector. The same result holds for $\widehat{h}_{n,m}(\tau)$. Similarly, it follows that

Lemma 3. *Let m be fixed. Under H_0 and Assumptions 1-6 the finite distributions of $\widehat{h}_{n,m}(\tau)$ converge.*

Because $b_m(\tau|\theta_0)$ is uniformly continuous on $\times_{j=0}^{m} \Upsilon$, it follows straightforwardly from (21) that $\widehat{h}_{2,n,m}(\tau|\widehat{\theta})$ is tight. Therefore, the tightness of $\widehat{h}_{n,m}(\tau)$ follows from the following lemma.

Lemma 4. *Let Y_t be bounded and m be fixed. Under H_0 and Assumptions 1-6 the process $\widehat{h}_{1,n,m}(\tau)$ is tight.*

Proof: Appendix.
Consequently:

Theorem 1. *Let Y_t be bounded and m be fixed. Under H_0 and Assumptions 1-6, $\widehat{h}_{n,m} \Rightarrow h_m$ on $\times_{j=0}^{m} \Upsilon$, where h_m is a complex-valued zero-mean*

Gaussian process with covariance function

$$\Gamma_m(\tau,\varsigma) = E\left[\phi_{m,t}(\tau)\overline{\phi_{m,t}(\varsigma)}\right], \tag{24}$$

where $\phi_{m,t}$ *is defined by* (23).[5] *Thus by the continuous mapping theorem,*

$$\widehat{B}_{n,m} = \int_{\times_{j=0}^{m}\Upsilon} \left|\widehat{h}_{n,m}(\tau)\right|^2 d\mu_m(\tau) \xrightarrow{d} B_m = \int_{\times_{j=0}^{m}\Upsilon} |h_m(\tau)|^2 d\mu_m(\tau)$$

for each non-negative integer m*, whereas under* H_1, $p\lim_{n\to\infty}\widehat{B}_{n,m}/n > 0$ *for all but a finite number of* m*'s.*

5.4 Weighted ICM Test

The weighted ICM test statistic takes the form

$$\widehat{W}_n = \sum_{m=1}^{\ell_n} \alpha^m \widehat{B}_{n,m}$$

where $\alpha \in (0,1)$ is arbitrary, and so is the subsequence ℓ_n of n as long as $\lim_{n\to\infty}\ell_n = \infty$. To prove that under H_0 and the conditions of Theorem 1, $\widehat{W}_n \xrightarrow{d} \sum_{m=1}^{\infty}\alpha^m B_m$, we need the following result.

Lemma 5. *Under* H_0 *and Assumptions 1-6,* $\sup_{m\geq 1} E[B_m] < \infty$ *and* $\sup_{m\geq 1}\widehat{B}_{n,m} = O_p(1)$.

Proof: Appendix

Combining this result with the results in Theorem 1 it follows that

Theorem 2. *Choose a constant* $\alpha \in (0,1)$ *and a subsequence* ℓ_n *of* n*. Under the conditions of Theorem 1,* $\widehat{W}_n = \sum_{m=1}^{\ell_n}\alpha^m\widehat{B}_{n,m} \xrightarrow{d} \sum_{m=1}^{\infty}\alpha^m B_m = W$ *if* H_0 *is true, whereas* $p\lim_{n\to\infty}\widehat{W}_n/n > 0$ *if* H_1 *is true.*

Proof: Appendix

[5] The bar in (24) denotes the complex conjugate of $\phi_{m,t}$.

6 The Weighted Simulated ICM Test

The theoretical conditional characteristic function poses a computational challenge, because often conditional distributions have no closed-form expression for their characteristic functions. To cope with this problem, we propose a Weighted Simulated Integrated Conditional Moment (WSICM) test, similar to the i.i.d. case considered in Bierens and Wang (2008), as follows. The idea is to replace the estimated conditional characteristic function $\varphi_{t-1}(\tau|\widehat{\theta})$ in the empirical process $\widehat{h}_m(\tau)$ defined by (11) with $\exp\left(i.\tau'\widetilde{Y}_t\right)$, where $\tilde{Y}_t$ is a random drawing from the estimated conditional null distribution $G_{t-1}(y|\widehat{\theta})$. Note that $\tilde{Y}_t$ has to be drawn from $G_{t-1}(y|\widehat{\theta})$ conditional on the actual past data.

The process (11) now becomes

$$\widehat{h}_{S,n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp\left(i.\tau_0'Y_t\right)) - \exp\left(i.\tau_0'\widetilde{Y}_t\right)\right) \exp\left(i \sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \quad (25)$$

Note that

$$\widehat{h}_{S,n,m}(\tau) = \widehat{h}_{n,m}(\tau) - \widetilde{h}_{S,n,m}(\tau),$$

where $\widehat{h}_{n,m}(\tau)$ is defined by (11) and

$$\widetilde{h}_{S,n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp\left(i.\tau_0'\widetilde{Y}_t\right) - \varphi_{t-1}\left(\tau_0|\widehat{\theta}\right)\right) \exp\left(i \sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \quad (26)$$

Similar to the proof of Lemma 4 it can be shown that conditional on all past and future data, i.e., conditional on the σ-algebra $\mathcal{F}_D = \sigma\left(\cup_t \mathcal{F}_{-\infty}^t\right)$, the process $\widetilde{h}_{S,n,m}$ is tight and is therefore tight unconditionally as well. Consequently, $\widetilde{h}_{S,n,m}$ converges weakly to a zero mean Gaussian process $h_{S,m}^*$, say. Moreover, denoting

$$\phi_{S,m,t}(\tau) = \left(\exp\left(i.\tau'Y_t\right) - \varphi_{t-1}\left(\tau|\theta_0\right)\right) \exp\left(i \sum_{j=1}^{m} \tau_j' Y_{t-j}\right) \quad (27)$$

which is similar to (23) but without the term $b_m(\tau|\theta_0)'\Sigma^{-1}U_t$, it is obvious that $(1/\sqrt{n})\sum_{t=1}^{n} \phi_{S,m,t}(\tau) \Rightarrow h_{S,m}^*(\tau)$ as well, hence the covariance function of $h_{S,m}^*$ is

$$\Gamma_{S,m}\left(\tau,\varsigma\right) = E\left[\phi_{S,m,t}(\tau)\overline{\phi_{S,m,t}(\varsigma)}\right] \quad (28)$$

Furthermore, it is not hard to verify that $h^*_{S,m}$ is independent of the Gaussian process h_m in Theorem 1. Therefore, the following results hold.

Theorem 3. *Under H_0 and the conditions of Theorem 1, $\widehat{h}_{S,n,m} \Rightarrow h_{S,m}$, where $\widehat{h}_{S,n,m}$ is the empirical process* (25) *and $h_{S,m}$ is a complex-valued zero-mean Gaussian process on $\times_{j=0}^{m}\Upsilon$ with covariance function $\Gamma_m(\tau,\varsigma) + \Gamma_{S,m}(\tau,\varsigma)$, with* Γ_m and $\Gamma_{S,m}$ *defined by* (24) *and* (28), *respectively. Thus by the continuous mapping theorem,*

$$\widehat{B}_{S,n,m} = \int_{\times_{j=0}^{m}\Upsilon} \left|\widehat{h}_{S,n,m}(\tau)\right|^2 d\mu_m(\tau) \stackrel{d}{\to} B_{S,m} = \int_{\times_{j=0}^{m}\Upsilon} \left|h_{S,m}(\tau)\right|^2 d\mu_m(\tau) \tag{29}$$

for fixed non-negative integers m, whereas under H_1, $p\lim_{n\to\infty}\widehat{B}_{S,n,m}/n > 0$ for all but a finite number of m's.

It is also easy to verify that Lemma 5 carries over. Consequently, Theorem 2 carries over to the SWICM test.

Theorem 4. *Choose a constant $\alpha \in (0,1)$ and a subsequence ℓ_n of n. Under the conditions of Theorem 1,*

$$\widehat{W}_{S,n} = \sum_{m=1}^{\ell_n} \alpha^m \widehat{B}_{S,n,m} \stackrel{d}{\to} \sum_{m=1}^{\infty} \alpha^m B_{S,m} = W_S \tag{30}$$

if H_0 is true, whereas $p\lim_{n\to\infty}\widehat{W}_{S,n}/n > 0$ if H_1 is true.

7 Standardization and Bounded Tranformation

The assumption that the process Y_t is bounded is not restrctive because without loss of genality we may replace Y_t and $\widetilde{Y}_t$ by bounded one-to-one transformations $\Phi(Y_t)$ and $\Phi(\widetilde{Y}_t)$, respectively. However, as argued in Bierens and Wang (2008) for the cross-section case, it is important for the preservation of the finite sample power of the WSICM test to standardize the variables involved before transforming them by a bounded one-to-one mapping Φ, as otherwise some or all the components of $\Phi(Y_t)$ and/or $\Phi(\widetilde{Y}_t)$ may become approximately constants. In particular, Bierens and Wang (2008) propose to standardize each component $Y_{j,t}$ of Y_t by $\overline{Y}_{j,t} = \sigma_{j,n}^{-1}(Y_{j,t} - \mu_{n,j})$, where for example $\mu_{j,n}$ is the sample mean of the $Y_{j,t}$'s and $\sigma_{j,n}$ is the corresponding sample standard error, and then taking the the arctan (.) transformation.

An alternative way to choose the location and scale parameters $\mu_{j,n}$ and $\sigma_{j,n}$, respectively, proposed by Bierens and Wang (2008) is to base them on empirical quantiles of the $Y_{j,t}$'s such that, for example, $(1/n)\sum_{t=1}^{n} I(|\overline{Y}_{j,t}| \leq 1) \approx 0.9$. The reason for the latter is that the arctan(.) function has still substantial variation on the interval $[-1,1]$: $\min_{-1\leq x\leq 1} d\arctan(x)/dx = 1/2$.

However, adopting the same standardization procedures in the time series case would create additional dependence between $\Phi(Y_t)$ and $\Phi(Y_{t-m})$ due to the common location and scale parameters. To avoid this problem, we propose to standardize each component $Y_{j,t}$ by

$$\overline{Y}_{j,t} = \frac{Y_{j,t} - \mu_{j,t-1}}{\sigma_{j,t-1}},\ \sigma_{j,t-1} > 0 \text{ for } t \geq 1$$

for example, where $\mu_{j,t-1}$ and $\sigma_{j,t-1}$ are functions of $Y_{j,1}, ..., Y_{j,t-1}$ only, and then taking the arctan transformation:

$$\Phi(Y_t) = \Psi_p\left(\Sigma_{Y,t-1}^{-1}(Y_t - \mu_{Y,t-1})\right) \tag{31}$$

$$\Phi(\widetilde{Y}_t) = \Psi_p\left(\Sigma_{Y,t-1}^{-1}\left(\widetilde{Y}_t - \mu_{Y,t-1}\right)\right) \tag{32}$$

where

$$\begin{aligned} \Psi_p\left((x_1,..,x_p)'\right) &= (\arctan(x_1), ..., \arctan(x_p))' \\ \mu_{Y,t-1} &= (\mu_{1,t-1}, ..., \mu_{p,t-1})', \\ \Sigma_{Y,t-1} &= \operatorname{diag}(\sigma_{1,t-1}), ..., \sigma_{p,t-1})) \end{aligned}$$

For example, choose

$$\mu_{j,t-1} = \frac{1}{t-1}\sum_{m=1}^{t-1} Y_{j,m},\ \sigma_{j,t-1} = 1 + \sqrt{\frac{1}{t-1}\sum_{m=1}^{t-1} Y_{j,m}^2 - \mu_{j,t-1}^2}$$

for $t \geq 2$ and $\mu_{j,t-1} = 0,\ \sigma_{j,t-1} = 1$ for $t \leq 1$. Alternatively, as motivated by Bierens and Wang (2008), choose

$$\begin{aligned}\mu_{j,t-1} &= \frac{1}{2}\left(Q_{j,t-1}(0.95) + Q_{j,t-1}(0.05)\right),\\ \sigma_{j,t-1} &= \frac{1}{2}\left(Q_{j,t-1}(0.95) - Q_{j,t-1}(0.05)\right),\end{aligned}$$

for $t \geq 2$ and $\mu_{j,t-1} = 0$, $\sigma_{j,t-1} = 1$ for $t \leq 1$, where

$$Q_{j,t-1}(\alpha) = \arg \max_{\frac{1}{t-1}\sum_{m=1}^{t-1} I(Y_{j,m} \leq x) \leq \alpha} x$$

is the $\alpha \times 100\%$ sample quantile of $Y_{j,1}, ..., Y_{j,t-1}$.

Denoting $\overline{Y}_t = \Sigma_{Y,t-1}^{-1}(Y_t - \mu_{Y,t-1})$ it follows trivially that

$$\begin{aligned}\Pr\left[\overline{Y}_t \leq y | \mathcal{F}_{-\infty}^{t-1}\right] &= \Pr\left[Y_t \leq \Sigma_{Y,t-1} y + \mu_{Y,t-1} | \mathcal{F}_{-\infty}^{t-1}\right]\\ &= F_{t-1}(\Sigma_{Y,t-1} y + \mu_{Y,t-1}) = \overline{F}_{t-1}(y),\end{aligned}$$

say, with corresponding specification

$$\overline{G}_{t-1}(y|\theta) = G_{t-1}(\Sigma_{Y,t-1} y + \mu_{Y,t-1}|\theta)$$

Therefore, all our asymptotic results carry over if we replace Y_t and $\widetilde{Y}_t$ by (31) and (32), respectively.

8 Monte Carlo Simulations

To check the small sample performance of the WSICM test, we have generated Gaussian MA(1) processes $Y_t = U_t - \theta U_{t-1}$ for $\theta = 0, 0.3, 0.6, 0.9$, respectively. For each of these processes we test the null hypothesis that Y_t is a Gaussian AR(p) process, for $p = 0, 1, 2, 3$, and sample sizes $n = 200, 600$.[6]

The parameter α in (30) has been chosen $\alpha = 0.9$, and the subsequence ℓ_n in (30) has been chosen $\ell_n = \left[n^{1/3}\right]$,[7] so that $\ell_n = 6$ for $n = 200$ and $\ell_n = 8$ for $n = 600$. The integration range Υ of the SICM statistic $\widehat{B}_{S,n,m}$ in (29) has been chosen $\Upsilon = [-5, 5]$. Finally, the bootstrap sample size is 500, the significance level is 10% and the number of replications is 100.

[6] Corresponding to 50 years of quarterly and monthly data, respectively.
[7] The notation $[x]$ indicates the largest integer $\leq x$.

Admittedly, the number of replications is rather small, but that is due to computational constraints. In particular, even with only 100 replications it took about one hour and 15 minutes on a PC to compute a single entry in Table 1 for $n = 200$, and eight hours for $n = 600$.

Table 1: Simulation results

	$n = 200$					$n = 600$			
$\theta \setminus p$	0	1	2	3	$\theta \setminus p$	0	1	2	3
0.0	13	8	15	11	0.0	16	8	12	11
0.3	37	12	12	10	0.3	87	12	10	10
0.6	90	16	10	7	0.6	100	49	19	9
0.9	97	31	11	11	0.9	100	78	14	12

As expected, the power results improve with the length of the time series, and the power decreases with θ. However, it is puzzling that the test has no power for $p > 1$ in the case $\theta = 0.9$. To partly explain this, note that for each lag length m the SICM test $\widehat{B}_{S,n,m}$ in (29) compares the AR(p) null distribution with the linear projection of Y_t on $Y_{t-1}, ..., Y_{t-m}$, which is an AR(m) model,

$$Y_t = \sum_{j=1}^{m} \beta_{j,m} Y_{t-j} + V_{m,t}, \quad V_{m,t} \sim N\left[0, \sigma_m^2\right], \tag{33}$$

where due the Gaussianity of Y_t the residual $V_{m,t}$ is independent of $Y_{t-1}, ..., Y_{t-m}$. The residual process $V_{m,t}$ is not independent, though. The AR(p) null model is of course also the linear projection of Y_t on $Y_{t-1}, ..., Y_{t-p}$,

$$Y_t = \sum_{j=1}^{p} \beta_{j,p} Y_{t-j} + V_{p,t}, \quad V_{p,t} \sim N\left[0, \sigma_p^2\right]. \tag{34}$$

For $m \leq p$ the test $\widehat{B}_{S,n,m}$ effectively compares an AR(m) model with itself because then by the law of iterated expectations,

$$\begin{aligned} & E\left[E\left[\exp\left(i.\tau Y_t\right) \middle| Y_{t-1}, ..., Y_{t-p}\right] \middle| Y_{t-1}, ..., Y_{t-m}\right] \\ & \qquad = E\left[\exp\left(i.\tau Y_t\right) \middle| Y_{t-1}, ..., Y_{t-m}\right] \end{aligned}$$

Thus, only the terms $\alpha^m \widehat{B}_{S,n,m}$ for $m > p$ in the WSICM statistic $\widehat{W}_{S,n} = \sum_{m=1}^{\ell_n} \alpha^m \widehat{B}_{S,n,m}$ contribute to the power of the test. Still, why they don't for $p > 1$ in the cases under review is an open question.

On the other hand, Bai's (2003) test will have no power in either of these cases, because (34) implies that the actual conditional distribution of Y_t given $Y_{t-1}, ..., Y_{t-p}$ is

$$\begin{aligned} F\left(y|Y_{t-1}, ..., Y_{t-p}\right) &= \Pr\left[Y_t \le y|Y_{t-1}, ..., Y_{t-p}\right] \\ &= \frac{\exp\left(-\frac{1}{2\sigma_p^2}\left(y - \sum_{j=1}^{p}\beta_{j,p}Y_{t-j}\right)^2\right)}{\sigma_p\sqrt{2\pi}} \end{aligned}$$

so that $\widetilde{U}_t = F\left(Y_t|Y_{t-1}, ..., Y_{t-p}\right)$ is uniformly $[0,1]$ distributed, although serially dependent. Bai's test only tests the uniformity hypothesis. If the AR(p) null model were correct, the $\widetilde{U}_t$'s would also be independent, but Bai's test does not check for that.

9 Conclusions

This paper extends Bierens (1984) weighted ICM test for functional forms to the test for the validity of parametric specifications of conditional distributions. The test is done by conducting a sequence of simulated ICM tests with an increasing number of lagged conditioning variables. The test statistic is a weighted sum of these simulated ICM test statistics. Preliminary simulations for Gaussian MA(1) data generating processes and Gaussian AR(p) null models show that in principle this test works, but that the small sample power deteriorates for $p \ge 2$. Why this is the case is yet unknown. Our conjecture is that this problem is typical for Gaussian processes. These issues will be addressed in future research.

10 Appendix

10.1 Proof of Lemma 1

It is well-known [see for example Theorem 3.12 in Bierens (2004)] that pointwise in τ_0,

$$\begin{aligned} \lim_{m\to\infty} E\left[\varphi_{t-1}\left(\tau_0|\theta_*\right) - \psi_{t-1}\left(\tau_0\right)|\mathcal{F}_{t-m}^{t-1}\right] &= E\left[\varphi_{t-1}\left(\tau_0|\theta_*\right) - \psi_{t-1}\left(\tau_0\right)|\mathcal{F}_{-\infty}^{t-1}\right] \\ &= \varphi_{t-1}\left(\tau_0|\theta_*\right) - \psi_{t-1}\left(\tau_0\right) \text{ a.s.} \end{aligned} \quad (35)$$

Let Y_t be a univariate time series. Without loss of generality we may interpret $\varphi_{t-1}(\tau_0|\theta_*)$ as

$$\varphi_{t-1}(\tau_0|\theta_*) = E\left[\exp\left(i.\tau_0\widetilde{Y}_t\right)\middle|\mathcal{F}_{-\infty}^{t-1}\right]$$

where $\widetilde{Y}_t$ is a bounded time series proces satisfying

$$G_{t-1}(y|\theta_*) = E\left[I(\widetilde{Y}_t \leq y)|\mathcal{F}_{-\infty}^{t-1}\right].$$

Then

$$\begin{aligned} E\left[\varphi_{t-1}(\tau_0|\theta_*) - \psi_{t-1}(\tau_0)|\mathcal{F}_{t-m}^{t-1}\right] &= \sum_{j=0}^{\infty}\frac{(i.\tau_0)^j}{j!}E\left[\widetilde{Y}_t^j - Y_t^j\middle|\mathcal{F}_{t-m}^{t-1}\right], \\ \varphi_{t-1}(\tau_0|\theta_*) - \psi_{t-1}(\tau_0) &= \sum_{j=0}^{\infty}\frac{(i.\tau_0)^j}{j!}E\left[\widetilde{Y}_t^j - Y_t^j\middle|\mathcal{F}_{-\infty}^{t-1}\right] \end{aligned}$$

Clearly, under H_1,

$$\Pr\left(E\left[\widetilde{Y}_t^{j_0} - Y_t^{j_0}\middle|\mathcal{F}_{-\infty}^{t-1}\right] = 0\right) < 1$$

for at least one $j_0 > 0$. For such a j,

$$\lim_{m\to\infty} E\left[\widetilde{Y}_t^j - Y_t^j\middle|\mathcal{F}_{t-m}^{t-1}\right] = E\left[\widetilde{Y}_t^j - Y_t^j\middle|\mathcal{F}_{-\infty}^{t-1}\right]$$

which implies that for all but a finite number of m's,

$$\Pr\left(E\left[\widetilde{Y}_t^{j_0} - Y_t^{j_0}\middle|\mathcal{F}_{t-m}^{t-1}\right] = 0\right) < 1$$

a.s. It follows now from Theorem 1 in Bierens and Ploberger (1997) that the set

$$\left\{(\tau_1,...,\tau_m)' \in [-c,c]^m : E\left[\left(\widetilde{Y}_t^{j_0} - Y_t^{j_0}\right)\exp\left(i.\sum_{j=1}^{m}\tau_j'Y_{t-j}\right)\right] = 0\right\}$$

has zero Lebesgue measure and is nowhere dense in $[-c,c]^m$. Consequently, for each of these m's there exists a τ_0 such that the set

$$\left\{(\tau_1,...,\tau_m)' \in [-c,c]^m : \right.$$
$$\left. E\left[\left(\exp\left(i.\tau_0\widetilde{Y}_t\right) - \exp(i.\tau_0 Y_t)\right)\exp\left(i.\sum_{j=1}^{m}\tau_j'Y_{t-j}\right)\right] = 0\right\}$$

has zero Lebesgue measure and is nowhere dense in $[-c,c]^m$. The result of Lemma 1 follows now straightforwardly, using the continuity of characteristic functions.

10.2 Proof of Lemma 4

We will prove Lemma 4 for the case $Y_t \in [-M, M]$ a.s. and $\Upsilon = [-c, c]$ where $c > 1$, as follows. Write (17) as

$$\widehat{h}_{1,n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp\left(i.\tau_0 Y_t\right) - E_{t-1}\left[\exp\left(i.\tau_0 Y_t\right)\right]\right) \exp\left(i \sum_{j=1}^{m} \tau_j Y_{t-j}\right)$$

where $E_{t-1}[\cdot]$ denotes $E\left[\cdot|\mathcal{F}_{-\infty}^{t-1}\right]$. Using the series expansion of the complex exp function, we can write

$$\begin{aligned}
\widehat{h}_{1,n,m}(\tau) &= \sum_{k=0}^{\infty} \frac{i^k}{k!} \tau_0^k \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(Y_t^k - E_{t-1}\left[Y_t^k\right]\right) \prod_{j=1}^{m} \left(\sum_{s=0}^{\infty} \frac{i^s}{s!} \tau_j^s Y_{t-j}^s\right) \\
&= \sum_{k_0=0}^{\infty} \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \frac{i^{\sum_{j=0}^{m} k_j}}{\prod_{j=0}^{m} k_j!} \prod_{j=0}^{m} \tau_j^{k_j} \\
&\qquad \times \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(Y_t^{k_0} - E_{t-1}\left[Y_t^{k_0}\right]\right) \prod_{j=1}^{m} Y_{t-j}^{k_j} \qquad (36)
\end{aligned}$$

Moreover, for $\tau, \varsigma \in [-c, c]^{m+1}$ and $||\tau - \varsigma|| \le \varepsilon < 1$ we have the inequality

$$\begin{aligned}
\left|\prod_{j=0}^{m} \tau_j^{k_j} - \prod_{j=0}^{m} \varsigma_j^{k_j}\right| &\le \left|\tau_0^{k_0} - \varsigma_0^{k_0}\right| c^{\sum_{j=1}^{m} k_k} + c^{k_0} \left|\prod_{j=1}^{m} \tau_j^{k_j} - \prod_{j=1}^{m} \varsigma_j^{k_j}\right| \\
&\le \sum_{j=1}^{k_0} \binom{k_0}{j} \left|\tau_0 - \varsigma_0\right|^j c^{\sum_{j=0}^{m} k_k} + c^{k_0} \left|\prod_{j=1}^{m} \tau_j^{k_j} - \prod_{j=1}^{m} \varsigma_j^{k_j}\right| \\
&\le \varepsilon \sum_{j=1}^{k_0} \binom{k_0}{j} c^{\sum_{j=0}^{m} k_k} + c^{k_0} \left|\prod_{j=1}^{m} \tau_j^{k_j} - \prod_{j=1}^{m} \varsigma_j^{k_j}\right| \\
&\le \varepsilon . 2^{k_0} c^{\sum_{j=0}^{m} k_k} + c^{k_0} \left|\prod_{j=1}^{m} \tau_j^{k_j} - \prod_{j=1}^{m} \varsigma_j^{k_j}\right|
\end{aligned}$$

hence by induction

$$\left|\prod_{j=0}^{m}\tau_j^{k_j}-\prod_{j=0}^{m}\varsigma_j^{k_j}\right|\leq\varepsilon.c^{\sum_{j=0}^{m}k_k}\sum_{j=0}^{m}2^{k_j}<\varepsilon.m\left(2c\right)^{\sum_{j=0}^{m}k_k}$$

Consequently, for $\varepsilon<1$,

$$\begin{aligned}
&E\left[\sup_{\|\tau-\varsigma\|\leq\varepsilon}\left|\widehat{h}_{1,n,m}(\tau)-\widehat{h}_{1,n,m}(\varsigma)\right|\right]\\
&\quad\leq\varepsilon.m\sum_{k_0=0}^{\infty}\sum_{k_1=0}^{\infty}\cdots\sum_{k_m=0}^{\infty}\frac{1}{\prod_{j=0}^{m}k_j!}\left(2c\right)^{\sum_{j=0}^{m}k_k}\\
&\qquad\times\sqrt{E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(Y_t^{k_0}-E_{t-1}\left[Y_t^{k_0}\right]\right)\prod_{j=1}^{m}Y_{t-j}^{k_j}\right)^2\right]}\\
&\quad\leq\varepsilon.m\sqrt{2}\sum_{k_0=0}^{\infty}\sum_{k_1=0}^{\infty}\cdots\sum_{k_m=0}^{\infty}\frac{1}{\prod_{j=0}^{m}k_j!}\left(2c\right)^{\sum_{j=0}^{m}k_k}\prod_{j=0}^{m}M^{k_j}\\
&\quad=\varepsilon.m\sqrt{2}\exp\left(2(m+1)cM\right)
\end{aligned}$$

Thus for fixed m the stochastic equicontinuity condition (16) holds, so that $\widehat{h}_{1,n,m}$ is tight. The generalization of this results to higher dimensions and more general spaces is straightforward.

10.3 Proof of Lemma 5

To prove of $\sup_{m\geq 1}E\left[B_m\right]<\infty$, note that

$$E\left[B_m\right]=\int_{\times_{j=0}^{m}\Upsilon}\Gamma_m\left(\tau,\tau\right)d\mu_m\left(\tau\right)=\int_{\times_{j=0}^{m}\Upsilon}E\left[|\phi_{m,t}(\tau)|^2\right]d\mu_m\left(\tau\right)$$

where Γ_m is the covariance function (24). Moreover, observe from (23) that $\phi_{m,t}(\tau)$ can be written as

$$\begin{aligned}
&\phi_{m,t}(\tau)\\
&=\exp\left(i\sum_{j=0}^{m}\tau_j'Y_{t-j}\right)-E_{t-1}\left[\exp\left(i\sum_{j=0}^{m}\tau_j'Y_{t-j}\right)\right]+b_m(\tau|\theta_0)'\Sigma^{-1}U_t.
\end{aligned}$$

$$= \cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right] + \mathrm{Re}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t$$
$$+i\left(\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right] + \mathrm{Im}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t\right)$$

where again $E_{t-1}[\cdot]$ denotes $E\left[\cdot|\mathcal{F}_{-\infty}^{t-1}\right]$. Moreover, observe from (20) and Assumption 6 that

$$\begin{aligned}\sup_{\tau\in\times_{j=0}^{m}\Upsilon}\|b_m(\tau|\theta_0)\|' &\leq \sup_{\tau_0\in\Upsilon} E\left[\|\Delta\varphi_{t-1}(\tau_0|\theta_0)\|\right] \\ &\leq E\left[\sup_{\tau_0\in\Upsilon}\|\Delta\varphi_{t-1}(\tau_0|\theta_0)\|\right] < \infty\end{aligned}$$

Hence

$$\begin{aligned}&|\phi_{m,t}(\tau)|^2 \\ &\leq \left(\cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right] + \mathrm{Re}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t\right)^2 \\ &+ \left(\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right] + \mathrm{Im}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t\right)^2 \\ &\leq 2\left(\cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\cos\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right]\right)^2 \\ &+2\left(\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right) - E_{t-1}\left[\sin\left(\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right]\right)^2 \\ &+2\left(\mathrm{Re}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t\right)^2 + 2\left(\mathrm{Im}\left[b_m(\tau|\theta_0)'\right]\Sigma^{-1}U_t\right)^2 \\ &\leq 8 + 2\left(E\left[\sup_{\tau_0\in\Upsilon}\|\Delta\varphi_{t-1}(\tau_0|\theta_0)\|\right]\right)^2 U_t'\Sigma^{-2}U_t\end{aligned}$$

It follows now easily from Assumption 3 that

$$E[B_m] \leq 8 + 2\left(E\left[\sup_{\tau_0\in\Upsilon}\|\Delta\varphi_{t-1}(\tau_0|\theta_0)\|\right]\right)^2 \mathrm{trace}\left(\Sigma^{-1}\right) < \infty.$$

Next, observe from (21) that

$$
\begin{aligned}
&E\left[\left|\widehat{h}_{1,n,m}(\tau)\right|^2\right] \\
&= 2E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\cos\left(i\sum_{j=0}^{m}\tau_j Y_{t-j}\right) - E_{t-1}\left[\cos\left(i\sum_{j=0}^{m}\tau_j Y_{t-j}\right)\right]\right)\right)^2\right] \\
&+2E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\sin\left(i\sum_{j=0}^{m}\tau_j Y_{t-j}\right) - E_{t-1}\left[\sin\left(i\sum_{j=0}^{m}\tau_j Y_{t-j}\right)\right]\right)\right)^2\right] \\
&\leq 2
\end{aligned}
$$

and from (19) that

$$
\begin{aligned}
&\left|\widehat{h}_{2,n,m}(\tau|\widehat{\theta})\right| \\
&\leq \left\|\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\right\|.\frac{1}{n}\sum_{t=1}^{n}\sup_{\tau_0\in\Upsilon}\left\|\Delta\varphi_{t-1}\left(\tau_0|\theta_0\right)\right\| \\
&+\frac{1}{2\sqrt{n}}\left\|\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\right\|^2\frac{1}{n}\sum_{t=1}^{n}\sup_{\tau_0\in\Upsilon}\left\|\Delta^2\varphi_{t-1}\left(\tau_0|\widetilde{\theta}_1\right)\right\| \\
&+\frac{1}{2\sqrt{n}}\left\|\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\right\|^2\frac{1}{n}\sum_{t=1}^{n}\sup_{\tau_0\in\Upsilon}\left\|\Delta^2\varphi_{t-1}\left(\tau_0|\widetilde{\theta}_2\right)\right\| \\
&= O_p(1)
\end{aligned}
$$

uniformly in τ and m. Hence

$$
\begin{aligned}
\widehat{B}_{n,m} &= \int_{\times_{j=0}^{m}\Upsilon}\left|\widehat{h}_{n,m}(\tau)\right|^2 d\mu_m(\tau) \\
&\leq 2\int_{\times_{j=0}^{m}\Upsilon}\left|\widehat{h}_{1,n,m}(\tau)\right|^2 d\mu_m(\tau) + 2\int_{\times_{j=0}^{m}\Upsilon}\left|\widehat{h}_{2,n,m}(\tau|\widehat{\theta})\right|^2 d\mu_m(\tau) \\
&= O_p(1)
\end{aligned}
$$

uniformly in m.

10.4 Proof of Theorem 2

It is easy to verify that under the conditions of Theorem 1, for fixed m,

$$
\left(\widehat{h}_{n,1},\widehat{h}_{n,2},...,\widehat{h}_{n,m}\right) \Rightarrow (h_1,h_2,...,h_m)
$$

hence for any positive integer K,

$$\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \xrightarrow{d} \sum_{m=1}^{K} \alpha^m B_m$$

Moreover, it is obvious from Lemma 5 that for $K \to \infty$,

$$\sum_{m=1}^{K} \alpha^m B_m \xrightarrow{d} \sum_{m=1}^{\infty} \alpha^m B_m = W,$$

say. Due to Lemma 5 we can choose K so large that for abitrary small $\varepsilon > 0$,

$$\begin{aligned} \limsup_{n\to\infty} \Pr\left[\sup_{m\geq 1} \widehat{B}_{n,m} > \varepsilon\left(1-\alpha\right)/\alpha^{K+1}\right] &< \varepsilon \\ \Pr\left[\sum_{m=K+1}^{\infty} \alpha^m B_m > \varepsilon\right] &< \varepsilon \end{aligned}$$

Next, let x be a continuity point of the distribution of W and observe that for $\ell_n > K$

$$\begin{aligned} I\left(\widehat{W}_n \leq x\right) &\leq I\left(\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \leq x\right) \\ I\left(\widehat{W}_n \leq x\right) &= I\left(\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \leq x - \sum_{m=K+1}^{\ell_n} \alpha^m \widehat{B}_{n,m}\right) \\ &\geq I\left(\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \leq x - \sup_{m\geq 1} \widehat{B}_{n,m}\alpha^{K+1}/\left(1-\alpha\right)\right) \\ &\geq I\left(\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \leq x - \varepsilon\right) I\left(\sup_{m\geq 1} \widehat{B}_{n,m}\alpha^{K+1}/\left(1-\alpha\right) \leq \varepsilon\right) \\ &\geq I\left(\sum_{m=1}^{K} \alpha^m \widehat{B}_{n,m} \leq x - \varepsilon\right) - I\left(\sup_{m\geq 1} \widehat{B}_{n,m} > \varepsilon\left(1-\alpha\right)/\alpha^{K+1}\right) \end{aligned}$$

Taking expectation and then letting $n \to \infty$ yield

$$\begin{aligned} \limsup_{n\to\infty} \Pr\left[\widehat{W}_n \leq x\right] &\leq \Pr\left[\sum_{m=1}^{K} \alpha^m B_m \leq x\right] \\ \liminf_{n\to\infty} \Pr\left[\widehat{W}_n \leq x\right] &\geq \Pr\left[\sum_{m=1}^{K} \alpha^m B_m \leq x - \varepsilon\right] - \varepsilon \end{aligned}$$

Next, letting $K \to \infty$ yields

$$\begin{aligned}
\limsup_{n\to\infty} \Pr\left[\widehat{W}_n \leq x\right] &\leq \Pr\left[W \leq x\right] \\
\liminf_{n\to\infty} \Pr\left[\widehat{W}_n \leq x\right] &\geq \Pr\left[W \leq x - \varepsilon\right] - \varepsilon
\end{aligned}$$

Finally, letting $\varepsilon \downarrow 0$ yields

$$\lim_{n\to\infty} \Pr\left[\widehat{W}_n \leq x\right] = \Pr\left[W \leq x\right].$$

References

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

Bai, J. and Chen, Z., 2007, Testing Multivariate Distributions in GARCH Models, *Journal of Econometrics (*forthcoming).

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., 2004, *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge, UK: Cambridge University Press.

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

Bierens, H. J. and Wang, L., 2008, 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.

Chen, X. and Fan, Y., 1999, Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series, *Journal of Econometrics*, 91, 373-401.

Corradi, V. and Swanson, N., 2006, Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification, *Journal of Econometrics*, 779-806.

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

Domínguez, M. and Lobato, I., 2003, Testing the Martingale Difference Hypothesis, *Econometric Reviews*, 22, 4, 351 - 377.

Escanciano, J. C. and Velasco, C., 2006, Generalized Spectral Density for the Martingale Difference Hypothesis Test, *Journal of Econometrics*, 134, 151-185.

Hong, Y., 1999, Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach, *Journal of the American Statistical Association*, 94, 1201-1220.

Hong Y. and Lee, Y., 2005, Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form, *Review of Economic Studies*, 72, 499-541.

Khmaladze, E., 1981, Martingale Approach in the Goodness-of-fit Tests, *Theory of Probability and its Applications,* XXVI, 240-257.

Li, F. and Tkacz, G. A., 2006, Consistent Bootstrap Test for Conditional Density Functions with Time-series Data, *Journal of Econometrics*, 133, 863-886.

Li, Q., 1999, Consistent Model Specification Tests for Time Series Econometric Models, *Journal of Econometrics*, 92, 101-147.

McLeish, D. L., 1974, Dependent Central Limit Theorems and Invariance Principles, *Annals of Probability*, 2, 620-628.

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

Stute, W., Quindimil, M., Manteiga, W. and Koul, H., 2006, Model Checks of Higher Order Time Series, *Statistics and Probability Letters*, 76, 1385-1396.

Su, L. and White, H., 2007, A Consistent Characteristic Function-Based Test for Conditional Dependence, *Journal of Econometrics,* 141, 807-834.

Van der Vaart, A.W. and Wellner, J.A., 1996, *Weak Convergence and Empirical Processes*, New York: Springer.

White, H., 1987, Specification Testing in Dynamic Models, in: Truman F.Bewley (ed.), *Advances in Econometrics, Fifth World Congress, Vol.I,* Cambridge, UK: Cambridge University Press, 1-58.