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

Herman J. Bierens*
Pennsylvania State University
University Park, PA 16802

Li Wang
Department of Public Health Sciences
Penn State College of Medicine
Hershey, PA 17033

June 18, 2014

Abstract

In this paper we propose a weighted simulated integrated conditional moment (WSICM) test of the validity of parametric specifications of conditional distribution models for stationary time series data, by combining the weighted ICM test of Bierens (1984) for time series regression models with the Simulated ICM test of Bierens and Wang (2012) of conditional distribution models for cross-section data. 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.

*Professor Emeritus of Economics. Please address correspondence by e-mail only to hbierens@psu.edu, as I no longer have an office at Penn State.

Keywords Stationary time series; Parametric conditional distributions; Simulated integrated method of moments; Consistent tests; Characteristic functions.

JEL Classification C12; C15; C22; C32; C52

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 of the 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.

In order to motivate our approach, consider the null hypothesis that a univariate time series Y_t is generated by the following ARMAX(1,1) process:

$$\begin{aligned} &Y_t = \alpha + \beta Y_{t-1} + \gamma' X_t + U_t - \delta U_{t-1}, \\ &\text{where } |\beta| < 1,\ |\delta| < 1,\ \beta \neq \delta,\ U_t \sim \text{i.i.d. } N(0, \sigma^2), \end{aligned} \tag{1.1}$$

with $X_t \in \mathbb{R}^k$ an observable stationary vector time series process of exogenous variables. Then

$$Y_t = \frac{\alpha}{1-\delta} + \sum_{j=0}^{\infty} \delta^j (\beta - \delta, \gamma') Z_{t-1-j} + U_t, \text{ where } Z_{t-1-j} = (Y_{t-1-j}, X'_{t-j})',$$

hence, conditional on the σ-algebra $\mathcal{F}_{-\infty}^{t-1} = \sigma(\{Z_{t-1-j}\}_{j=0}^{\infty})$ generated by the one-sided infinite sequence $\{Z_{t-1-j}\}_{j=0}^{\infty}$, and with $\theta = (\alpha, \beta, \gamma', \sigma)'$, the distribution $G_{t-1}(y|\theta)$ of Y_t is normal with conditional mean $\mu_{t-1}(\theta) = \alpha/(1-\delta) + \sum_{j=0}^{\infty} \delta^j (\beta - \delta, \gamma') Z_{t-1-j}$ and variance σ^2:

$$G_{t-1}(y|\theta) = \int_{-\infty}^{y} \exp\left(-(v - \mu_{t-1}(\theta))^2/(2\sigma^2)\right) / \left(\sigma\sqrt{2\pi}\right) \mathrm{d}v.$$

Now let $F_{t-1}(y) = E[I(Y_t \leq y)|\mathcal{F}_{-\infty}^{t-1}]$ be the actual conditional distribution of Y_t given $\mathcal{F}_{-\infty}^{t-1}$, and let Θ be the parameter space for θ. Then the null hypothesis to be tested is that

$$H_0 : \text{For some } \theta_0 \in \Theta,\ \Pr\left[\sup_{y\in\mathbb{R}}|G_{t-1}(y|\theta_0) - F_{t-1}(y)| = 0\right] = 1$$

against the alternative hypothesis that the null hypothesis is false:

$$H_1 : \text{For all } \theta \in \Theta,\ \Pr\left[\sup_{y\in\mathbb{R}}|G_{t-1}(y|\theta) - F_{t-1}(y)| = 0\right] < 1.$$

As is well-known, these hypotheses can also be expressed equivalently in terms of conditional characteristic functions, as follows. Let[1]

$$\begin{aligned} \varphi_{t-1}(\tau|\theta) &= \int_{-\infty}^{\infty} \exp(\mathbf{i}.\tau.y)\mathrm{d}G_{t-1}(y|\theta) = \exp(\mathbf{i}.\tau.\mu_{t-1}(\theta))\exp(-\sigma^2\tau^2/2) \\ \psi_{t-1}(\tau) &= \int_{-\infty}^{\infty} \exp(\mathbf{i}.\tau.y)\mathrm{d}F_{t-1}(y) = E\left[\exp(\mathbf{i}.\tau.Y_t)\left|\mathcal{F}_{-\infty}^{t-1}\right.\right]. \end{aligned}$$

Then H_0 and H_1 are equivalent to

$$\begin{aligned} H_0 &: \text{For some } \theta_0 \in \Theta,\ \Pr\left[\sup_{\tau\in\mathbb{R}}|\varphi_{t-1}(\tau|\theta_0) - \psi_{t-1}(\tau)| = 0\right] = 1, \\ H_1 &: \text{For all } \theta \in \Theta,\ \Pr\left[\sup_{\tau\in\mathbb{R}}|\varphi_{t-1}(\tau|\theta) - \psi_{t-1}(\tau)| = 0\right] < 1, \end{aligned}$$

respectively.

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 (2012)[2] for the i.i.d. case, with

[1] Here and in the sequel $\mathbf{i}$ denotes the complex number $\mathbf{i} = \sqrt{-1}$.

[2] See also Bierens (2014b) for extensions and corrections of Bierens and Wang (2012).

$(Z'_{t-1}, Z'_{t-2}, ..., Z'_{t-m})'$ the vector of conditioning variables. Thus, $\widehat{B}_{n,m}$ is based on the integrated squared difference between the empirical characteristic function of $(Y'_t, Z'_{t-1}, Z'_{t-2}, ..., Z'_{t-m})'$ and the corresponding empirical characteristic function implied by the estimated conditional distribution model for Y_t. Given an arbitrary subsequence L_n of the sample size n, the actual weighted ICM (WICM) test statistic is $\widehat{W}_n = \sum_{m=1}^{L_n} \omega_m \widehat{B}_{n,m}$, where ω_m is a given sequence of positive weights satisfying $\sum_{m=1}^{\infty} \omega_m < \infty$.

It will be shown that under H_0, $\widehat{W}_n \xrightarrow{\text{d}} W$, whereas under H_1, $p\lim_{n\to\infty} \widehat{B}_{n,m}/n > 0$ for all but a finite number of m's, so that under H_1, $p\lim_{n\to\infty} \widehat{W}_n /n > 0$. The distribution of W is case dependent. However, p-values can be derived via a bootstrap method.

Since this test is based on characteristic functions, as in Bierens and Wang (2012), 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. 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 whose tests are not truly based on the entire past of the time series involved.

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 (WSICM) 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 6. Our test depends crucially on the fact that conditional characteristic functions of bounded random variables are equal if and only if they are equal in an arbitrary open neighborhood of zero, or the origin of the Euclidean space involved in the multivariate case. Section 7 discuss various ways to transform the random variables in our test statistic to bounded random variables while preserving enough variation. Section 8 shows how the p-value of our test can be computed via a parametric bootstrap method. In section 9 we conduct a limited Monte Carlo simulation, and in section 10 we apply our test to a few GARCH models for a financial time series. Finally, in Section 11 we make some concluding remarks.

As to notations, the indicator function will be denoted by $I(.)$, the vector norm $||x||$ is the Euclidean norm $||x|| = \sqrt{x'x}$ 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 $x = a + \mathbf{i}.b \in \mathbb{C}$ with $\mathbf{i} = \sqrt{-1}$ 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. The double-arrow $\Rightarrow$ indicates weak convergence.[3] Finally, we adopt the convention that the derivative of a function to a row [column] vector is a column [row] vector of partial derivatives.[4]

2. LITERATURE REVIEW ON TESTING TIME SERIES MODELS

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

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

[4] For example, $\partial\left(x'b\right)/\partial x' = b$, $\partial\left(x'b\right)/\partial x = \partial\left(b'x\right)/\partial x = b'$.

regression models by Bierens (1984), de Jong (1996) and Bierens and Ploberger (1997).

The literature on consistent model specification testing has focused almost entirely on conditional expectation models. Up to the early nineties the only papers on this topic were Bierens (1982, 1984, 1990). After 1990 two strands of literature emerged: (1) de Jong (1996), Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Boning and Sowell (1999), Fan and Li (2000) and Escanciano (2006) for Integrated Conditional Moment (ICM) and related tests, and (2) Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Gozalo (1993), Horowitz and Härdle (1994), Hong and White (1995), Li and Wang (1998), Zheng (1996) and Lavergne and Vuong (2000), among others, for tests based on comparisons of parametric functional forms with corresponding nonparametric or semi-nonparametric estimates.

However, the literature on *consistent* specification testing of conditional distribution models is very limited;[5] to the best of our knowledge it consists only of three papers, Andrews (1997), Zheng (2000) and Bierens and Wang (2012), for i.i.d. data. As motivated below, we have excluded Bai (2003) and Bai and Chen (2008) from this list as their tests are not consistent.

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

$$Y_t = \alpha + \beta Y_{t-1} + U_t, \ \ |\beta| < 1. \tag{2.1}$$

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 $\mathcal{F}_{-\infty}^{t-1} = \sigma(\{Y_{t-j}\}_{j=1}^{\infty})$, i.e.,

$$E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0 \text{ a.s.}, \tag{2.2}$$

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.

[5] Of course, there are many more tests for the validity of conditional distributions, but none of these tests are consistent.

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

$$E\left[U_t|\mathcal{F}_{t-m}^{t-1}\right] = 0 \text{ a.s.}, \tag{2.3}$$

where $\mathcal{F}_{t-m}^{t-1} = \sigma(\{Y_{t-j}\}_{j=1}^m)$, and then use $\sum_{m=1}^{L_n} \omega_m \widehat{B}_m$ 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 (2.2), as follows. He identifies the null hypothesis (2.2) versus the alternative

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

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

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

where $\boldsymbol{\Xi}$ is a compact metric space in $\mathbb{R}^\infty$, and Φ 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 $\Phi(Y_{t-j})$. Under the null hypothesis (2.2), $\mathbf{S} = \boldsymbol{\Xi}$, whereas under the alternative, $\mathbf{S}$ is "almost empty". Therefore, a consistent ICM test of the null hypothesis (2.2) can be based on the integral

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

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 (2.3) 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 (2.1) augmented with the assumption $U_t|Y_{t-1} \sim N\left[0,\sigma^2\right]$. The conditional distribution of this model given Y_{t-1} takes the form

$$\begin{aligned} G_{t-1}(y|\theta) &= \sigma^{-1}\Gamma\left((y-\alpha-\beta Y_{t-1})/\sigma\right) \\ \theta &= (\alpha,\beta,\sigma)', \end{aligned} \tag{2.4}$$

where $\Gamma(.)$ is the cumulative distribution function 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 $\Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right] \neq \Gamma\left((y-\alpha-\beta Y_{t-1})/\sigma\right)$. 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

$$F_{t-1}(y) = \Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right] = \sigma_V^{-1}\Gamma\left(\frac{y-\sum_{j=1}^{\infty}\gamma^j Y_{t-j}}{\sigma_V}\right) \tag{2.5}$$

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

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 (2.5), for example, $U_t = F_{t-1}(Y_t)$ 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 a pivotal test, Bai uses the Khmaladze (1981) martingale transformation, which yields a correction term $\widehat{K}_n$, say, such that under the null hypothesis,

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

Under the null hypothesis, V_n converges weakly to a standard Brownian bridge. However, in the case of the incorrect null model (2.4) $U_t = G_{t-1}(Y_t|\theta)$ is also uniformly [0,1] distributed for $\theta = p\lim_{n\to\infty}\widehat{\theta}$, 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.

Bai and Chen (2008) 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 a pivotal 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. Due to the latter, their test is not consistent, and for the same reason neither are the tests developed by Li (1999) and Chen and Fan (1999).

A somewhat related paper to ours is Corradi and Swanson (2007). They compare the joint distributions of simulated data from competing DSGE models, including a finite number of lags, with the corresponding joint empirical distribution via an ICM type criterion to select the best fitting model. The critical values involved are computed via a block-bootstrap approach.

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 $Y_{t-1}, Y_{t-2}, ..., Y_{t-m}$: $\mathcal{F}_{t-m}^{t-1} = \sigma(Y_{t-1}, Y_{t-2}, ..., Y_{t-m})$, 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 $\{Y_{t-j}\}_{j=1}^{\infty}$: $\mathcal{F}_{-\infty}^{t-1} = \sigma(\{Y_{t-j}\}_{j=1}^{\infty})$. 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,

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

See Bierens (2004, Theorem 7.4, p. 184). Furthermore, under Assumption 1 the stochastic properties of Y_t are completely determined by the conditional distribution function

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

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 and convex parameter space. Thus, we do not consider models that involve exogenous variables, as in the ARMAX(1,1) case (1.1). The reason is notational convenience. The extension to models with exogenous variables is almost trivial and will therefore be left to the reader.

Furthermore, it is reasonable to assume that $G_{t-1}(y|\theta)$ is specified 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

$$\begin{aligned} &H_0\textbf{:} \;\; \text{There exists an interior point } \theta_0 \in \Theta \text{ such that} \\ &\qquad \Pr\left[\sup\nolimits_{y\in\mathbb{R}^p}|G_{t-1}(y|\theta_0) - F_{t-1}(y)| = 0\right] = 1, \qquad (3.1) \\ &H_1\textbf{:} \;\; \text{For all } \theta \in \Theta, \\ &\qquad \Pr\left[\sup\nolimits_{y\in\mathbb{R}^p}|G_{t-1}(y|\theta) - F_{t-1}(y)| = 0\right] < 1, \qquad (3.2) \end{aligned}$$

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* (3.1),

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -\Sigma^{-1}\left(n^{-1/2}\sum_{t=1}^{n} U_t\right) + o_p(1),$$

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:[6] $n^{-1/2}\sum_{t=1}^{n} U_t \xrightarrow{\text{d}} N_k[0,\Sigma]$, $\det(\Sigma) > 0$.

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

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

[6] See for example McLeish (1974).

Under H_1 the estimator $\widehat{\theta}$ is a Quasi ML (QML) estimator. It is a standard exercise to set forth mild condition such that $\widehat{\theta}$ converges in probability to a point in Θ, namely the point $\theta_* = \arg\max_{\theta\in\Theta} \lim_{n\to\infty} E\left[\ln L_n(\theta)/n\right].$ See for example White (1994). Therefore, rather than stating these conditions we assume that

Assumption 4. *Under the alternative hypothesis* (3.2), $p\lim_{n\to\infty} \widehat{\theta} = \theta_* \in \Theta.$

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 $F_{t-1}(y)$:

$$\varphi_{t-1}(\tau|\theta) = \int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau' y) \mathrm{d}G_{t-1}(y|\theta), \tag{4.1}$$

$$\psi_{t-1}(\tau) = \int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau' y) \mathrm{d}F_{t-1}(y) = E\left[\exp(\mathbf{i}.\tau' Y_t) \left| \mathcal{F}_{-\infty}^{t-1}\right.\right], \tag{4.2}$$

respectively. As is well known, H_0 is true if and only if $\Pr[\sup_{\tau\in\mathbb{R}^p} |\varphi_{t-1}(\tau|\theta_0) - \psi_{t-1}(\tau)| = 0] = 1$, whereas under H_1, $\Pr[\sup_{\tau\in\mathbb{R}^p} |\varphi_{t-1}(\tau|\theta_*) - \psi_{t-1}(\tau)| = 0] < 1$. 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 $\mathbb{R}^p$, $\Pr[\sup_{\tau\in N_0} |\varphi_{t-1}(\tau|\theta_*) - \psi_{t-1}(\tau)| = 0] < 1$, 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 in principle $\varphi_{t-1}(\tau|\theta)$ 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, as the latter may depend on the entire past $\{Y_{t-j}\}_{j=1}^{\infty}$ of the time series involved. The following lemma provides a solution to this problem.

Lemma 1. *Let Assumptions 1, 2 and 4 hold, with Y_t a bounded process. Denote*

$$\varphi^{m+1}(\tau|\theta) = E\left[\int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau_0' y) \mathrm{d}G_{t-1}(y|\theta) \exp\left(\mathbf{i}\sum_{j=1}^{m} \tau_j' Y_{t-j}\right)\right]$$

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

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.

The result of Lemma 1 yields the following corollary. Denote

$$
\begin{aligned}
G_{t-1}^m(y|\theta) &= E[G_{t-1}(y|\theta)|Y_{t-1}, Y_{t-2}, ..., Y_{t-m}], \\
F_{t-1}^m(y) &= E[F_{t-1}(y)|Y_{t-1}, Y_{t-2}, ..., Y_{t-m}],
\end{aligned}
$$

Then:

Theorem 1. *Under H_1 and Assumptions 1, 2 and 4,* $\Pr[\sup_{y\in\mathbb{R}^p}|G_{t-1}^m(y|\theta_*) - F_{t-1}^m(y)| = 0] < 1$ *for all but a finite number of m's, regardless whether Y_t is bounded or not.*

Proof. Appendix.

Note that by the law of iterated expectations,

$$
\begin{aligned}
\varphi^{m+1}(\tau|\theta) &= E\left[\int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau_0' y) \mathrm{d}G_{t-1}(y|\theta) \exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right)\right] \\
&= E\left[\int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau_0' y) \mathrm{d}G_{t-1}^m(y|\theta) \exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right)\right], \\
\psi^{m+1}(\tau) &= E\left[\exp\left(\mathbf{i}\sum_{j=0}^{m}\tau_j' Y_{t-j}\right)\right] \\
&= E\left[\int_{\mathbb{R}^p} \exp(\mathbf{i}.\tau_0' y) \mathrm{d}F_{t-1}^m(y) \exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right)\right],
\end{aligned}
$$

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

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

In particular, if

Assumption 5. *For each $\tau_0 \in \Upsilon$, $\int_{\mathbb{R}^p}\exp(\mathbf{i}.\tau_0'y)\mathrm{d}G_{t-1}(y|\theta)$ 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),

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

Hence by Assumption 4,[7]

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

and thus

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

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

for all but a finite number of m's.

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

5. THE WEIGHTED ICM TEST AND ITS ASYMPTOTIC NULL DISTRIBUTION

Consider the empirical process

$$\begin{aligned} \widehat{h}_{n,m}(\tau) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\exp\left(\mathbf{i}.\tau_0'Y_t\right)-\varphi_{t-1}(\tau_0|\widehat{\theta})\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right), \quad (5.1)\\ \tau &= (\tau_0',\tau_1',...,\tau_m')'\in\Upsilon^{m+1}, \end{aligned}$$

where $\varphi_{t-1}(\tau_0|\widehat{\theta})$ is defined by (4.1) and Υ is a compact set in $\mathbb{R}^p$ centered around the origin of $\mathbb{R}^p$.

For given m, let

$$\widehat{B}_{n,m} = \int_{\Upsilon^{m+1}} \left|\widehat{h}_{n,m}(\tau)\right|^2 \mathrm{d}\mu_m(\tau)$$

where $\mu_m(\tau)$ is the uniform probability measure on Υ^{m+1}, i.e.

$$\mathrm{d}\mu_m(\tau) = \frac{\mathrm{d}\tau}{\int_{\Upsilon^{m+1}} 1.\mathrm{d}\tau}$$

5.1. Weak Convergence

In this subsection we will set forth conditions such that under H_0, and for fixed m, $\widehat{h}_{n,m} \Rightarrow h_m$, where $h_m(\tau)$ is a zero-mean complex valued Gaussian process on Υ^{m+1}, so that by the continuous mapping theorem,

$$\widehat{B}_{n,m} \xrightarrow{\mathrm{d}} B_m = \int_{\Upsilon^{m+1}} |h_m(\tau)|^2 \mathrm{d}\mu_m(\tau).$$

As is well known[8], the necessary and sufficient conditions for weak convergence are that $\widehat{h}_{n,m}$ is tight and the finite distributions of $\widehat{h}_{n,m}$ converge. The latter means that for arbitrary $\tau_1, \tau_2, ..., \tau_k$ in Υ^{m+1},

$$\left(\widehat{h}_{n,m}(\tau_1), \widehat{h}_{n,m}(\tau_2), ..., \widehat{h}_{n,m}(\tau_k)\right)' \xrightarrow{\mathrm{d}} (h_m(\tau_1), h_m(\tau_2), ..., h_m(\tau_k))'. \quad (5.2)$$

According to Billingsley (1968, Theorem 8.2), the following two conditions are sufficient for the tightness of $\widehat{h}_{n,m}$:

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

(a) For each $\eta > 0$ and each $\tau \in \Upsilon^{m+1}$ there exists a $\delta > 0$ such that

$$\sup_{n\geq 1} \Pr[|\widehat{h}_{n,m}(\tau)| > \delta] \leq \eta$$

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

$$\sup_{n\geq 1} \Pr\left[\sup_{||\tau_1-\tau_2||<\varepsilon} |\widehat{h}_{n,m}(\tau_1) - \widehat{h}_{n,m}(\tau_1)| \geq \delta\right] \leq \eta. \tag{5.3}$$

Condition (a) is a pointwise stochastic boundedness condition, which holds if for each $\tau \in \Upsilon^{m+1}$, $\widehat{h}_{n,m}(\tau)$ converges in distribution, hence this condition follows from condition (5.2). 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 (5.1), using Assumption 3, as follows. Write (5.1) as

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

where

$$\begin{aligned}\widehat{h}_{1,n,m}(\tau) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\exp\left(\mathbf{i}.\tau_0'Y_t\right) - \varphi_{t-1}\left(\tau_0|\theta_0\right)\right), \\ &\quad \times \exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right) \qquad (5.4) \\ \widehat{h}_{2,n,m}(\tau|\widehat{\theta}) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varphi_{t-1}\left(\tau_0|\widehat{\theta}\right) - \varphi_{t-1}\left(\tau_0|\theta_0\right)\right) \\ &\quad \times \exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right) \qquad (5.5)\end{aligned}$$

Next, assume that

Assumption 6. *Under H_0 the conditional characteristic function $\varphi_{t-1}(\tau|\theta)$ defined by (4.1) is a.s. twice continuously differentiable on an open neighborhood $\Theta_0 \subset \Theta$ of θ_0 with first and second partial derivatives satisfying*

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

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

Denote

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

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

$$
\begin{aligned}
&\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) = \sqrt{n}\left(\widehat{\theta}-\theta_0\right)'\frac{1}{n}\sum_{t=1}^{n}\Delta\varphi_{t-1}(\tau_0|\theta_0)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right) \\
&+\frac{1}{2}\sqrt{n}\left(\widehat{\theta}-\theta_0\right)'\frac{1}{n}\sum_{t=1}^{n}\left(\mathrm{Re}\left[\Delta^2\varphi_{t-1}\left(\tau_0|\widetilde{\theta}_1\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right)\right]\right. \\
&\times\left(\widehat{\theta}-\theta_0\right) \\
&+\mathbf{i}.\frac{1}{2}\sqrt{n}\left(\widehat{\theta}-\theta_0\right)'\frac{1}{n}\sum_{t=1}^{n}\mathrm{Im}\left[\Delta^2\varphi_{t-1}\left(\tau_0|\widetilde{\theta}_2\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j'Y_{t-j}\right)\right] \\
&\times\left(\widehat{\theta}-\theta_0\right) \qquad (5.6)
\end{aligned}
$$

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

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

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:

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

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

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} \Delta\varphi_{t-1}\left(\tau_0|\theta_0\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right) &\xrightarrow{\text{p}} b_m(\tau|\theta_0) \\ &= E\left[\Delta\varphi_{t-1}\left(\tau_0|\theta_0\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right)\right]\end{aligned} \tag{5.7}$$

uniformly in $\tau = (\tau_0', \tau_1', ..., \tau_m')' \in \Upsilon^{m+1}$. Thus it follows from Assumption 3 and (5.5) that

$$\widehat{h}_{2,n,m}(\tau|\widehat{\theta}) = -b_m(\tau|\theta_0)'\Sigma^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\right) + o_p\left(1\right), \tag{5.8}$$

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

Combining the results (5.4) and (5.8), $\widehat{h}_{n,m}(\tau)$ can be written as

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

where

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

with

$$\begin{aligned}\phi_{m,t}(\tau) \quad = \quad & \left(\exp\left(\mathbf{i}.\tau_0' Y_t\right) - \varphi_{t-1}\left(\tau_0|\theta_0\right)\right)\exp\left(\mathbf{i}\sum_{j=1}^{m}\tau_j' Y_{t-j}\right) \\ & + b_m(\tau|\theta_0)'\Sigma^{-1}U_t.\end{aligned} \tag{5.9}$$

5.3. Tightness and Convergence Results

Note that pointwise in $\tau \in \Upsilon^{m+1}$, $\phi_{m,t}(\tau)$ 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),

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

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. *Under H_0 and Assumptions 1-6 the finite distributions of $\widehat{h}_{n,m}(\tau)$ converge for each $m \geq 1$.*

Because $b_m(\tau|\theta_0)$ is uniformly continuous on Υ^{m+1}, it follows straightforwardly from (5.8) 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 2. *Let Y_t be bounded and m be fixed. Under H_0 and Assumptions 1-6, $\widehat{h}_{n,m} \Rightarrow h_m$ on Υ^{m+1}, where h_m is a complex-valued zero-mean Gaussian process with covariance function*

$$\Gamma_m(\tau, \varsigma) = E\left[\phi_{m,t}(\tau)\overline{\phi_{m,t}(\varsigma)}\right], \tag{5.10}$$

with $\phi_{m,t}$ defined by (5.9). Thus by the continuous mapping theorem,

$$\widehat{B}_{n,m} = \int_{\Upsilon^{m+1}} \left|\widehat{h}_{n,m}(\tau)\right|^2 \mathrm{d}\mu_m(\tau) \xrightarrow{\mathrm{d}} B_m = \int_{\Upsilon^{m+1}} \left|h_m(\tau)\right|^2 \mathrm{d}\mu_m(\tau)$$

for each integer $m \geq 1$, 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

Recall that the weighted ICM test statistic takes the form $\widehat{W}_n = \sum_{m=1}^{L_n} \omega_m \widehat{B}_{n,m}$, where ω_m is an a priori chosen positive sequence of weights satisfying $\sum_{m=1}^{\infty} \omega_m < \infty$, and L_n is any subsequence of n satisfying $\lim_{n\to\infty} L_n = \infty$.

To prove that under H_0 and the conditions of Theorem 2, $\widehat{W}_n \xrightarrow{d} W$, we need the following result.

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

Proof. Appendix.

Combining this result with the results in Theorem 2 it follows that

Theorem 3. *Under the conditions of Theorem 2,* $\widehat{W}_n = \sum_{m=1}^{L_n} \omega_m \widehat{B}_{n,m} \xrightarrow{d} \sum_{m=1}^{\infty} \omega_m B_m = W$ *if* H_0 *is true, whereas* $p\lim_{n\to\infty} \widehat{W}_n/n > 0$ *if* H_1 *is true.*

Proof. Appendix.

Remark: Although the results of Theorem 3 hold for any positive sequence ω_m satisfying $\sum_{m=1}^{\infty} \omega_m < \infty$ and any subsequence L_n of n, it is obvious that the choice of the weights ω_m matters for the small sample power, and so does the subsequence L_n. However, there is no way to devise an optimal choice for ω_m and L_n unless the true conditional distribution is known.

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 (2012), as follows. The idea is to replace the estimated conditional characteristic function $\varphi_{t-1}(\tau|\widehat{\theta})$ in the empirical process $\widehat{h}_{n,m}(\tau)$ defined by (5.1) with $\exp\left(\mathbf{i}.\tau'\widetilde{Y}_t\right)$, 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 (5.1) now becomes

$$\widehat{h}_{S,n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp\left(\mathbf{i}.\tau_0' Y_t\right)\right) - \exp\left(\mathbf{i}.\tau_0' \widetilde{Y}_t\right)\right) \exp\left(\mathbf{i} \sum_{j=1}^{m} \tau_j' Y_{t-j}\right). \quad (6.1)$$

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 (5.1) and

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

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

$$\mathcal{D} = \sigma\left(\{Y_t\}_{t=-\infty}^{\infty}\right)$$

generated by the entire time series Y_t 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

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

which is similar to (5.9) 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, where the latter is a zero-mean Gaussian process with covariance function

$$\Gamma_{S,m}(\tau, \varsigma) = E\left[\phi_{S,m,t}(\tau)\overline{\phi_{S,m,t}(\varsigma)}\right]. \quad (6.2)$$

Furthermore, it is not hard to verify that $h^*_{S,m}$ is independent of the Gaussian process h_m in Theorem 2. Therefore, the following results hold.

Theorem 4. *Under H_0 and the conditions of Theorem 2, $\widehat{h}_{S,n,m} \Rightarrow h_{S,m}$, where $\widehat{h}_{S,n,m}$ is the empirical process* (6.1) *and $h_{S,m}$ is a complex-valued zero-mean Gaussian process on Υ^{m+1} with covariance function $\Gamma_m(\tau,\varsigma) + \Gamma_{S,m}(\tau,\varsigma)$, with* Γ_m and $\Gamma_{S,m}$ *defined by* (5.10) *and* (6.2), *respectively. Thus by the continuous mapping theorem,*

$$\widehat{B}_{S,n,m} = \int_{\Upsilon^{m+1}} \left|\widehat{h}_{S,n,m}(\tau)\right|^2 \mathrm{d}\mu_m(\tau) \quad (6.3)$$
$$\xrightarrow{\mathrm{d}} B_{S,m} = \int_{\Upsilon^{m+1}} \left|h_{S,m}(\tau)\right|^2 \mathrm{d}\mu_m(\tau)$$

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 3 carries over to the WSICM test.

Theorem 5. *Choose the sequences ω_m and L_n as before. Under the conditions of Theorem 3,*

$$\widehat{W}_{S,n} = \sum_{m=1}^{L_n} \omega_m \widehat{B}_{S,n,m} \xrightarrow{\mathrm{d}} \sum_{m=1}^{\infty} \omega_m B_{S,m} = W_S \tag{6.4}$$

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 TRANSFORMATION

The assumption that the process Y_t is bounded is not restrictive because without loss of generality 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 (2012) 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 (2012) 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 $\arctan(.)$ transformation.

An alternative way to choose the location and scale parameters $\mu_{j,n}$ and $\sigma_{j,n}$, respectively, proposed by Bierens and Wang (2012) 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 $[-1, 1]$:

$$\min_{-1\leq x\leq 1} \frac{\mathrm{d}\arctan(x)}{\mathrm{d}x} = 1/2.$$

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

$$\overline{Y}_{j,t}=\frac{Y_{j,t}-\mu_{j,t-1}}{\sigma_{j,t-1}},\ \sigma_{j,t-1}>0 \text{ for } t\geq 1 \tag{7.1}$$

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:

$$\Phi\left(Y_t\right) = \Psi_p\left(\Sigma_{Y,t-1}^{-1}\left(Y_t-\mu_{Y,t-1}\right)\right) \tag{7.2}$$

$$\Phi(\widetilde{Y}_t) = \Psi_p\left(\Sigma_{Y,t-1}^{-1}\left(\widetilde{Y}_t-\mu_{Y,t-1}\right)\right) \tag{7.3}$$

where

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

For example, choose

$$\mu_{j,t-1}=\frac{1}{t-1}\sum_{s=1}^{t-1}Y_{j,s},\ \sigma_{j,t-1}^2=\frac{1}{t-1}\sum_{s=1}^{t-1}\left(Y_{j,s}-\mu_{j,s}\right)^2 \tag{7.4}$$

for $t\geq 3$ and $\mu_{j,t-1}=0,\ \sigma_{j,t-1}=1$ for $t\leq 2$. Alternatively, as motivated by Bierens and Wang (2012), choose

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

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

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

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 (7.2) and (7.3), respectively.

8. PARAMETRIC BOOTSTRAP

In this section we set forth mild additional conditions for the asymptotic validity of the following parametric bootstrap approach, which is an adaptation of the bootstrap method proposed by Li and Tkacz (2006) and has been used in Bierens and Wang (2012) for the i.i.d. case.

For notational convenience, suppose that the time series process Y_t is univariate and bounded, and that the conditional null distribution model $G_{t-1}(y|\theta)$ depends only on a finite number m of lagged Y_t's, i.e.,

$$G_{t-1}(y|\theta) = F(y|X_{t,m}, \theta) \text{ with } X_{t,m} = (Y_{t-1}, Y_{t-2}, ..., Y_{t-m})',$$

where $F(y|X_{t,m}, \theta)$ is absolutely continuous with conditional density $f(y|X_{t,m}, \theta)$. Moreover, assume that the null hypothesis is true.

Generate M bootstrap samples

$$\{(\widetilde{Y}_{b,t}, X_{t,m})\}_{t=1}^{n}, \ b = 1, ..., M,$$

where $\widetilde{Y}_{b,t}$ is a random drawing from $F(y|X_{t,m}, \widehat{\theta})$ in bootstrap sample b given $X_{t,m}$ and the ML estimator $\widehat{\theta}$. Let $\widetilde{\theta}_b$ be the ML estimator on the basis of this bootstrap sample, i.e., $\widetilde{\theta}_b = \arg\max_{\theta\in\Theta} \ln \widehat{L}_{b,n}(\theta)$, where $\ln \widehat{L}_{b,n}(\theta) = \sum_{t=1}^{n} \ell(\widetilde{Y}_{b,t}, X_{t,m}; \theta)$ with

$$\ell(y, X_{t,m}; \theta) = \ln(f(y|X_{t,m}, \theta)).$$

Then the bootstrap ICM test statistic (in the exact ICM case) is

$$\widehat{B}_{b,n,m} = \int_{\Upsilon^{m+1}} \left|\widehat{h}_{b,n,m}(\tau)\right|^2 \mathrm{d}\mu_m(\tau),$$

where

$$\widehat{h}_{b,n,m}(\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp\left(\mathbf{i}.\tau_0 \widetilde{Y}_{b,t}\right) - \varphi(\tau_0 | X_{t,m}; \widetilde{\theta}_b) \right) \exp\left(\mathbf{i}.\tau_1' X_{t,m}\right)$$

with $\tau = (\tau_0, \tau_1')'$ and

$$\varphi(\tau_0 | X_j; \widetilde{\theta}_b) = \int \exp(\mathbf{i}.\tau_0 y) f(y | X_{t,m}; \widetilde{\theta}_b) \mathrm{d}y.$$

In Bierens and Wang (2012) we have set forth conditions such that in the i.i.d. case, $(\widetilde{\theta}_b - \widehat{\theta}) \xrightarrow{\text{p}} 0$. Along the same lines, using the uniform law of large numbers for dependent processes,[9] it can be straightforwardly shown that this result carries over in the present case. Therefore, we simply assume that

Assumption 7. $p\lim_{n\to\infty}(\widetilde{\theta}_b - \widehat{\theta}) = 0$.

The next step is to show that $\sqrt{n}(\widetilde{\theta}_b - \widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$. For this we need the following standard regularity conditions on the vector $\Delta\ell(y, X_{t,m}; \theta) = (\partial/\partial\theta')\ell(y, X_{t,m}; \theta)$ and the matrix $\Delta^2\ell(y, X; \theta) = (\partial^2/(\partial\theta\partial\theta'))\ell(y, X_{t,m}; \theta)$, similar to Assumption 5 in Bierens and Wang (2012).

Assumption 8. *The elements and components, respectively, of the integrals* $\int \Delta(y, X_{t,m}; \theta).f(y|X_{t,m}; \theta)\mathrm{d}y$ *and* $\int \Delta^2\ell(y, X_{t,m}; \theta) f(y|X_{t,m}; \theta)\mathrm{d}y$ *are a.s. continuous on an arbitrary small open neighborhood* Θ_0 *of* θ_0, *with* $\Theta_0 \subset \Theta$, *and*

$$\int \Delta\ell(y, X_{t,m}; \theta) f(y|X_{t,m}; \theta)\mathrm{d}y = \frac{\partial}{\partial\theta'} \int \ell(y, X_{t,m}; \theta) f(y|X_{t,m}; \theta)\mathrm{d}y = 0$$

on Θ_0. *Moreover, for an arbitrarily small* $\delta > 0$,[10]

$$\begin{aligned} E\left[\sup_{\theta\in\Theta_0} \left\| \int \Delta\ell(y, X_{t,m}; \theta) f(y|X_{t,m}; \theta) \right\|^{2+\delta} \mathrm{d}y \right] &< \infty, \\ E\left[\sup_{\theta\in\Theta_0} \left\| \int \Delta^2\ell(y, X_{t,m}; \theta) f(y|X_{t,m}; \theta) \right\| \mathrm{d}y \right] &< \infty. \end{aligned} \quad (8.1)$$

[9] See for example Bierens (2004, Theorem 7.8(a), p. 187).

[10] Recall that the matrix norm $||.||$ in (8.1) is the maximum absolute value of the elements of the matrix involved.

Let $\mathcal{D} = \sigma(\{Y_t\}_{t=-\infty}^{\infty})$ be the σ-algebra generated by the entire time series Y_t. Then *conditional on* $\mathcal{D}$,

$$U_{t,m,n} = \Delta\ell(\widetilde{Y}_{b,j}, X_{t,m}; \widehat{\theta}) - \int \Delta\ell(y, X_{t,m}; \widehat{\theta}) f(y|X_{t,m}; \widehat{\theta}) \mathrm{d}y$$

is a double array of independent random vectors, for which Liapounov's central limit theorem applies. See for example Chung (1974, p. 200). It follows now similar to Bierens and Wang (2012) that conditional on $\mathcal{D}$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Delta\ell\left(\widetilde{Y}_{b,j}, X_j; \widehat{\theta}\right) \xrightarrow{\mathrm{d}} N_k[0, \Sigma], \tag{8.2}$$

where Σ is defined in Assumption 3. This result can also be proved using the martingale difference central limit theorem. See McLeish (1974) for the latter. It is a standard exercise to verify from Assumptions 7-8 and (8.2) that conditional on $\mathcal{D}$,

$$\sqrt{n}(\widetilde{\theta}_b - \widehat{\theta}) \xrightarrow{\mathrm{d}} N_k\left[0, \Sigma^{-1}\right].$$

It follows now similar to Bierens and Wang (2012) that conditional on $\mathcal{D}$, $\widehat{h}_{b,n,m} \Rightarrow h_{b,m}$ as $n \to \infty$, where $h_{b,m}$ is a zero-mean complex valued Gaussian process on Υ^{m+1} with the same covariance function as the limiting process h_m in Theorem 2. Moreover, conditional on $\mathcal{D}$, the processes $h_{b,m}$ are independent across the bootstrap samples b and therefore unconditionally as well. Hence,

$$\widehat{B}_{b,n,m} \xrightarrow{\mathrm{d}} B_{b,m}$$

where the $B_{b,m}$'s are independent replications of B_m in Theorem 2. Consequently, for each bootstrap sample b,

$$\widehat{W}_{b,n} = \sum_{m=1}^{L_n} \omega_m \widehat{B}_{b,n,m} \xrightarrow{\mathrm{d}} \sum_{m=1}^{\infty} \omega_m B_{b,m} = W_b,$$

where the latter are independent replications of W in Theorem 3, so that the p-value of the test can be approximated by

$$\widehat{p}_{n,M} = \frac{1}{M} \sum_{b=1}^{M} I\left(\widehat{W}_{b,n} > \widehat{W}_n\right).$$

Strictly speaking, we should let M go to infinity with the sample size n at some rate, like in Corradi and Swanson (2007), in order to get a consistent estimate of the actual p-value. However, this bootstrap procedure is computationally intensive, so that too large an M may not be practical in empirical applications. Moreover, the limited Monte Carlo simulation in the next section demonstrates that the bootstrap rejection rates under the null hypothesis are close to the nominal sizes for fixed $M = 500$.

Finally, it is not hard to verify that this bootstrap approach remains valid for the WSICM test.

9. A LIMITED MONTE CARLO STUDY

In order to check the performance of the WSICM test and its bootstrap procedure, we consider four data-generating processes:

$$
\begin{array}{lll}
\text{AR(1)} & Y_t = \beta Y_{t-1} + \varepsilon_t, & |\beta| < 1, \\
\text{MA(1)} & Y_t = (\varepsilon_t - \theta\varepsilon_{t-1})/\sqrt{1+\theta^2}, & |\theta| < 1, \\
\text{ARCH(1)} & Y_t = \varepsilon_t.\sigma_t, & \\
 & \sigma_t^2 = (1-\theta) + \theta.Y_{t-1}^2, & 0 < \theta < 1, \\
\text{GARCH(1,1)} & Y_t = \varepsilon_t.\sigma_t, & \\
 & \sigma_t^2 = (1-\theta)(1-\lambda) + \lambda\sigma_{t-1}^2 + \theta(1-\lambda)Y_{t-1}^2, & 0 < \theta, \lambda < 1,
\end{array}
$$

where $\varepsilon_t \sim$ i.i.d. $N(0,1)$. The reason for the particular parametrization in the last three cases is to enforce the restriction $E[Y_t^2] = 1$ in order to keep the comparison fair.

The null hypothesis to be tested is that Y_t is a stationary Gaussian AR(1) process:

$$H_0\text{:} \quad Y_t = \alpha + \beta Y_{t-1} + \sigma.\varepsilon_t, \ |\beta| < 1, \ \text{where } \varepsilon_t \sim \ \text{i.i.d. } N(0,1).$$

We consider two sample sizes, $n = 200$ and $n = 600$, corresponding to 50 years of quarterly and monthly data, respectively. The maximum number of lags in the WSICM test is $L_n = \sqrt{n}$, rounded off to the nearest integer. Thus, $L_n = 14$ for $n = 200$ and $L_n = 24$ for $n = 600$. The WSICM weights have been chosen $\omega_m = \alpha^m$ with $\alpha = 0.9$, and the integration range has been chosen $\Upsilon = [-5, 5]$. We will use 500 bootstrap samples to compute bootstrap p-values, and 1000 replications to approximate the 5% and 10% rejection rates.

In the AR(1) case we conduct the WSICM test for $\beta = 0, 0.25, 0.5, 0.75$, in the other cases for $\theta = 0.2, 0.4., 0.6, 0.8$ and in the GARCH(1,1) case also for $\lambda = 0.25, 0.5, 0.75$. The results are as follows.

9.1. AR(1)

In this case the null hypothesis is true, so this is a check on the size properties of the bootstrap procedure.

TABLE 1. AR(1) rejection rates (%)

$n = 200$			$n = 600$		
β	5%	10%	β	5%	10%
0.00	5.1	9.9	0.00	5.6	11.6
0.25	5.9	10.3	0.25	5.5	10.4
0.50	5.6	11.6	0.50	3.7	9.5
0.75	5.5	9.2	0.75	4.6	9.4

As expected, the value of β does not matter. The results indicate that the bootstrap procedure works quite well. However, the variations of the rejection rates indicate that 1000 replications or 500 bootstrap samples may not be enough, but going beyond these numbers would take too much computing time.[11]

9.2. MA(1)

As is well-known, the MA(1) process involved can be written as a Gaussian AR(∞) process:

$$Y_t = -\sum_{m=1}^{\infty} \theta^m Y_{t-m} + U_t, \text{ where } U_t = \sigma_U.\varepsilon_t, \ \sigma_U^2 = (1+\theta^2)^{-1}, \tag{9.1}$$

whereas under the null hypothesis one would expect that for some $\beta \in (-1, 1)$,

$$Y_t = \beta Y_{t-1} + V_t, \text{ where } V_t = \sigma_V.\varepsilon_t. \tag{9.2}$$

At first sight there seems to be a huge difference between (9.1) and (9.2). However, the WSICM test has difficulties to detect this difference:

[11] This Monte Carlo analysis took several weeks to conduct on a PC.

TABLE 2. MA(1) rejection rates (%)

$n = 200$			$n = 600$		
θ	5%	10%	θ	5%	10%
0.2	5.3	10.3	0.2	5.6	10.4
0.4	5.2	11.4	0.4	6.1	11.2
0.6	6.4	12.9	0.6	12.5	20.5
0.8	8.1	16.0	0.8	22.3	35.4

To explain these results, note that βY_{t-1} in (9.2) can be interpreted as the projection of Y_t on Y_{t-1}, i.e.,

$$\beta = \frac{E[Y_t Y_{t-1}]}{E[Y_{t-1}^2]} = \frac{-\theta}{1+\theta^2}$$

with projection residual

$$\begin{aligned} V_t &= Y_t + \frac{\theta}{1+\theta^2} Y_{t-1} \\ &= U_t - \frac{\theta^3}{1+\theta^2} U_{t-1} - \frac{\theta^2}{1+\theta^2} U_{t-2} \end{aligned} \tag{9.3}$$

and corresponding variance

$$\sigma_V^2 = \sigma_U^2 \left(1 + \frac{\theta^4}{1+\theta^2}\right) = \frac{1+\theta^2+\theta^4}{(1+\theta^2)^2}$$

Since V_t is a projection residual, we have $E[V_t Y_{t-1}] = 0$, so that by the joint normality of V_t and Y_{t-1}, the latter pair is independent. Moreover, it follows straightforwardly from (9.3) that V_t is independent of all but one lagged Y_t's, with exception of Y_{t-2}:

$$\begin{aligned} E[V_t Y_{t-2}] &= -\frac{\theta^2}{(1+\theta^2)^2} \\ &= \begin{cases} -0.036982 & \text{for} \quad \theta = 0.2 \\ -0.118906 & \text{for} \quad \theta = 0.4 \\ -0.194637 & \text{for} \quad \theta = 0.6 \\ -0.237954 & \text{for} \quad \theta = 0.8 \end{cases}, \\ E[V_t Y_{t-m}] &= 0 \text{ for } m = 1 \text{ and } m \geq 3. \end{aligned}$$

This is the main difference between (9.1) and (9.2).

Another way to look at the difference between (9.1) and (9.2) is to compare the mean square error of the implied conditional expectations:

$$\begin{aligned} E\left[\left(\beta Y_{t-1} + \sum_{m=1}^{\infty} \theta^m Y_{t-m}\right)^2\right] &= \frac{\theta^4}{(1+\theta^2)^2} \\ &= \begin{cases} 0.001479 & \text{for} \quad \theta = 0.2 \\ 0.019025 & \text{for} \quad \theta = 0.4 \\ 0.070069 & \text{for} \quad \theta = 0.6 \\ 0.152290 & \text{for} \quad \theta = 0.8 \end{cases} \end{aligned}$$

Thus, the mean square error involved is very small, in particular for $\theta < 0.8$, which explains why it is difficult to detect the difference between (9.1) and (9.2).

Finally, let us compare the conditional characteristic functions of Y_t and $\widetilde{Y}_t = \beta Y_{t-1} + \widetilde{V}_t$, respectively, where $\widetilde{V}_t = \sigma_V \widetilde{\varepsilon}_t$ with $\widetilde{\varepsilon}_t$ a random drawing from the $N(0,1)$ distribution. It follow from (9.1) and $Y_t = U_t - \theta U_{t-1}$ that the sequences $\{Y_{t-m}\}_{m=1}^{\infty}$ and $\{U_{t-m}\}_{m=1}^{\infty}$ generate the same σ-algebra: $\mathcal{F}_{-\infty}^{t-1} = \sigma(\{Y_{t-m}\}_{m=1}^{\infty}) = \sigma(\{U_{t-m}\}_{m=1}^{\infty})$, hence, the conditional characteristic function of Y_t relative to $\mathcal{F}_{-\infty}^{t-1}$ takes the form

$$\begin{aligned} \varphi_{t-1}(\tau) &= E[\exp(\mathbf{i}.\tau.Y_t)|\mathcal{F}_{-\infty}^{t-1}] \\ &= E[\exp(\mathbf{i}.\tau.U_t)\exp(-\mathbf{i}.\tau.\theta U_{t-1})|\mathcal{F}_{-\infty}^{t-1}] \\ &= \exp\left(-\frac{1}{2}\tau^2.\sigma_U^2\right).\exp(-\mathbf{i}.\tau.\theta U_{t-1}). \end{aligned}$$

whereas the conditional characteristic function of $\widetilde{Y}_t$ relative to $\mathcal{F}_{-\infty}^{t-1}$ takes the form

$$\begin{aligned} \widetilde{\varphi}_{t-1}(\tau) &= E[\exp(\mathbf{i}.\tau.\widetilde{Y}_t)|\mathcal{F}_{-\infty}^{t-1}] \\ &= E[\exp(\mathbf{i}.\tau.\widetilde{V}_t)\exp(\mathbf{i}.\tau.\beta Y_{t-1})|\mathcal{F}_{-\infty}^{t-1}] \\ &= \exp(-\tau^2.\sigma_V^2/2).\exp(\mathbf{i}.\tau.\beta Y_{t-1}) \\ &= \exp\left(-\frac{1}{2}\tau^2\sigma_U^2\left(1 + \frac{\theta^4}{1+\theta^2}\right)\right) \\ &\quad \times \exp\left(\mathbf{i}.\tau.\left(\frac{-\theta}{1+\theta^2}U_{t-1} + \frac{\theta^2}{1+\theta^2}U_{t-2}\right)\right) \end{aligned}$$

After some tedious but straightforward calculations it can be shown that

$$E[|\widetilde{\varphi}_{t-1}(\tau) - \varphi_{t-1}(\tau)|^2] = \exp(-\tau^2.\sigma_U^2) - \exp\left(-\tau^2\sigma_U^2\left(1 + \frac{\theta^4}{1+\theta^2}\right)\right)$$

and maximizing the latter expression to τ yields

$$\begin{aligned}\sup_{\tau\in\mathbb{R}} E[|\widetilde{\varphi}_{t-1}(\tau) - \varphi_{t-1}(\tau)|^2] &= \left(\frac{1+\theta^2}{1+\theta^2+\theta^4}\right)^{(1+\theta^2)\theta^{-4}} \frac{\theta^4}{1+\theta^2+\theta^4} \\ &= \begin{cases} 0.000566 & \text{for} \quad \theta = 0.2 \\ 0.008030 & \text{for} \quad \theta = 0.4 \\ 0.033474 & \text{for} \quad \theta = 0.6 \\ 0.081849 & \text{for} \quad \theta = 0.8 \end{cases}\end{aligned}$$

Thus, the difference between the two conditional characteristic functions is very small, which explains once more why our WSICM test has low power against the MA(1) case.

9.3. ARCH(1)

The results for the ARCH(1) case are as follows.

TABLE 3. ARCH(1) rejection rates (%)

$n = 200$			$n = 600$		
θ	5%	10%	θ	5%	10%
0.2	8.3	13.4	0.2	10.8	16.4
0.4	14.8	21.7	0.4	40.9	53.2
0.6	36.5	45.0	0.6	87.5	91.5
0.8	66.0	74.0	0.8	99.7	99.8

The dramatic increase in power with θ has two reasons. Obviously, the dependence of σ_t^2 on Y_{t-1}^2 increases with θ. But the constant $1 - \theta$ decreases with θ, which enhances the effect of Y_{t-1}^2 on σ_t^2.

9.4. GARCH(1,1)

The results for the GARCH(1,1) case are displayed in Table 4.

TABLE 4. GARCH(1,1) rejection rates (%)

$n = 200$

$\lambda = 0.25$			$\lambda = 0.50$			$\lambda = 0.75$		
θ	5%	10%	θ	5%	10%	θ	5%	10%
0.2	5.3	11.5	0.2	7.0	11.8	0.2	5.3	11.6
0.4	10.3	18.6	0.4	7.7	12.1	0.4	6.3	12.3
0.6	26.3	37.5	0.6	18.5	29.2	0.6	8.3	15.2
0.8	55.6	65.3	0.8	40.5	50.6	0.8	20.7	27.9

$n = 600$

$\lambda = 0.25$			$\lambda = 0.50$			$\lambda = 0.75$		
θ	5%	10%	θ	5%	10%	θ	5%	10%
0.2	7.8	13.1	0.2	6.8	11.2	0.2	5.6	10.8
0.4	26.2	37.9	0.4	15.6	23.6	0.4	7.9	14.2
0.6	70.3	78.2	0.6	51.8	62.1	0.6	20.7	28.8
0.8	96.5	98.2	0.8	89.5	92.7	0.8	59.2	67.2

It is a standard exercise to verify that

$$\sigma_t^2 = (1-\theta) + \theta(1-\lambda)\sum_{m=0}^{\infty}\lambda^m Y_{t-1-m}^2.$$

The latter explains why the power increases with θ and decreases with λ.

10. AN EMPIRICAL APPLICATION TO STOCK RETURNS

10.1. Data and Model

To verify that the WSICM test has power in practice, we apply the test to the daily (log) stock returns of Agilent Technologies[12] over the period 1/5/2009-9/20/2013 (Source: Wessa 2013). Thus, the length of the time series involved is $n = 1187$.

The standard model for stock returns Y_t is the GARCH(p, q) model proposed by Bollerslev (1986):

$$\begin{aligned} Y_t &= \mu + \varepsilon_t.\sigma_t(\theta), \text{ where} \\ \varepsilon_t &\sim \text{i.i.d. } N(0,1), \end{aligned}$$

[12] Agilent Technologies produces hi-tech measurement equipments. See http://www.home.agilent.com/

$$\begin{aligned}\sigma_t^2(\theta) &= \delta + \sum_{m=1}^{p} \beta_m (Y_{t-m} - \mu)^2 + \sum_{k=1}^{q} \gamma_k \sigma_{t-k}^2(\theta), \\ \theta &= (\mu, \delta, \beta_1, ..., \beta_p, \gamma_1, ..., \gamma_q)', \\ \delta &> 0,\ \beta_m \geq 0,\ \gamma_k \geq 0,\ p \geq 1,\ q \geq 0.\end{aligned}$$

In the case $q = 0$ this model becomes an ARCH(p) model proposed by Engle (1982). As is well-known, p has to be at least 1 because otherwise $\sigma_t^2(\theta)$ becomes constant.

The null hypothesis to be tested is that for given p and q this model (including the normality of the ε_t's) is the data-generating process for the stock returns involved.

10.2. Test Options and Bootstrap Setup

The construction of the WSICM test statistic requires to choose the integration range Υ in (6.3) and the weights ω_m and maximum lag L_n in the WSICM test (6.4). We have chosen $\Upsilon = [-c, c]$ with $c = 5$,[13] $\omega_m = \alpha^m$ with $\alpha = 0.9$, and $L_n = \sqrt{n}$, rounded off to the nearest integer. Thus, because $n = 1187$, the maximum lag is $L_n = 34$. All the random variables in the WSICM test statistic have been rescaled similar to (7.1) and (7.4), and then made bounded by using the arctan(.) function.

Moreover, since Y_t is not observed for $t \leq 0$, the conditional variance $\sigma_t^2(\theta)$ for $t \leq 0$ has been approximated by replacing the Y_{t-m}'s by the sample mean $\overline{Y}$, so that for $t \leq 0$, $\sigma_t^2(\theta) \approx \left(1 - \sum_{k=1}^{q} \gamma_k\right)^{-1} \left(\delta + \sum_{m=1}^{p} \beta_m (\overline{Y} - \mu)^2\right)$.

Given the ML estimates $\widehat{\theta} = (\widehat{\mu}, \widehat{\delta}, \widehat{\beta}_1, ..., \widehat{\beta}_p, \widehat{\gamma}_1, ..., \widehat{\gamma}_q)'$ and the actual data Y_t, $t = 1, ..., n$, each bootstrap sample is generated as $Y_{b,t} = \widehat{\mu} + \widetilde{\varepsilon}_t . \sigma_t(\widehat{\theta})$, $t = 1, ..., n$, where the $\widetilde{\varepsilon}_t$'s are random drawings from the standard normal distribution. Note that $\sigma_t(\widehat{\theta})$ depends on the actual time series Y_t only because we condition on the data. The next step is to obtain new ML estimates on the basis of the conditional log-likelihood

$$\mathcal{L}_n(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left((Y_{b,t} - \mu)^2 / \sigma_t^2(\theta) - \ln(\sigma_t^2(\theta)) \right) - \frac{1}{2} n . \ln(2\pi).$$

However, that is too much of a computational burden. Therefore, instead of a full ML estimation round, we have used the well-known fact that a single Newton

[13] The value $c = 5$ has been chosen on the basis of the simulation results in Bierens and Wang (2012).

step starting from a $\sqrt{n}$-consistent estimator $\widehat{\theta}$, i.e., $\sqrt{n}(\widehat{\theta}-\theta_0)=O_p(1)$ with θ_0 the true parameter, yields an estimate $\widetilde{\theta}$ that is asymptotically equivalent to the full ML estimator. Moreover, it is a standard ML exercise to verify that under the null hypothesis the same applies to the quasi-Newton step estimator

$$\widetilde{\theta}=\widehat{\theta}+\left(\frac{1}{n}\sum_{t=1}^{n}s_t(\widehat{\theta})s_t(\widehat{\theta})'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}s_t(\widehat{\theta})\right),$$

where $s_t(\theta)=-0.5\frac{\partial}{\partial\theta'}\left((Y_{b,t}-\mu)^2/\sigma_t^2(\theta)-\ln(\sigma_t^2(\theta))\right)$ is a score vector. The simulated bootstrap sample is now generated as $\widetilde{Y}_{b,t}=\widetilde{\mu}+\widetilde{\varepsilon}_t.\sigma_t(\widetilde{\theta})$, where again the $\widetilde{\varepsilon}_t$'s are random drawings from the standard normal distribution.

Finally, the bootstrap p-values will be calculated using 500 bootstrap samples.

10.3. GARCH(1,1) Model

In first instance we have estimated and tested a GARCH(1,1) model for the Agilent stock returns. The ML estimation and WSICM test results are:[14]

TABLE 5. GARCH(1,1)

θ	$\widehat{\theta}$	t-value	LL	WSICM	p-value
μ	0.001638	3.132	2892.89	16.4782	0.00000
δ	0.000037	7.588	Inf. crit.		
β_1	0.208171	16.674	−7.69899	(HQ)	
γ_1	0.738860	70.246	−7.68831	(SC)	

Here LL is the log-likelihood, and HQ and SC are the Hannan-Quinn (1979) and Schwarz (1978) information criteria, respectively.

The (bootstrap) p-value is the number of times that a bootstrap WSICM statistic $\widetilde{W}_{n,b}$ exceeded the value $\widehat{W}_{S,n}=16.4782$, divided by 500. Obviously, $\widetilde{W}_{n,b}>\widehat{W}_{S,n}$ never happened. Thus, the GARCH(1,1) model is strongly rejected by the WSICM test.

In contrast, applying the WSICM test to an artificially generated GARCH(1,1) time series of length $n=1187$ with the same parameters as the ML estimates in Table 5 yields bootstrap p-value 0.94800.[15]

[14] The GARCH estimation and WSICM test have been conducted via the latest version of the free econometric software package *EasyReg International* (See Bierens 2014c).

[15] In principle we could repeat this simulation procedure 1000 times, for example, to determine the size properties of the WSICM test. However, the computation of the WSICM test and its

There are various possible reasons for this rejections. First, it may be possible that the specifications of p and/or q are incorrect. Another reason may be that the normality assumption $\varepsilon_t \sim$ i.i.d. $N(0,1)$ does not hold. A third reason may be that the GARCH functional form of $\sigma_t^2(\theta)$ is incorrect.

10.4. GARCH(4,5) Model

To check whether the rejection is due to incorrect values of p and/or q, we have started off with a GARCH(6,6) model. The ML estimates of β_6 and γ_6 were both insignificant, but some other β's and γ's were significant at the 1% level, which is evidence against a common root problem. Therefore, we have re-estimated the model as a GARCH(5,5) model and conducted a series of Wald tests to clean up the model. In particular, the Wald test of the null hypothesis $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \beta_5 = 0$ could not be rejected at the 10% significance level. Consequently, we have re-estimated the model as a GARCH(4,5) model with $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$. The ML estimation and WSICM test results involved are presented in Table 6.

Somewhat surprisingly, this GARCH(4,5) model is also strongly rejected by the WSICM test, although the HQ and SC information criteria favor this model over the GARCH(1,1) specification.

TABLE 6. GARCH(4,5)

θ	$\widehat{\theta}$	t-value	LL.	WSICM	p-value
μ	0.001841	3.472	2922.36	14.3459	0.02800
δ	0.000059	4.349	Inf. crit.		
β_1	0.102315	5.047	-7.73874	(HQ)	
β_2	0.135668	4.651	-7.72007	(SC)	
β_3	0.063527	2.616			
β_4	0.122836	4.966			
γ_5	0.476575	13.215			

Again, we have applied the WSICM test to an artificially generated GARCH(4,5) time series with the same parameters and structure as in Table 6, resulting in bootstrap p-value 0.58600.

To check for possible non-normality, we have estimated the density of the ε_t's by nonparametric kernel density estimation on the basis of the estimated ε_t's, $\widehat{\varepsilon}_t = (Y_t - \widehat{\mu})/\sigma_t(\widehat{\theta})$, with standard normal kernel and optimal bandwidth $n^{-1/5}$. In

bootstrap p-value takes about half an hour on a 32-bit Windows XP PC, so that in this case a full Monte Carlo simulation would take about 20 days!

Figure 1 this kernel density estimator (solid curve) is compared with the standard normal density (dashed curve). Clearly, they do not match, which may be the reason for the rejection of the GARCH(4,5) model.

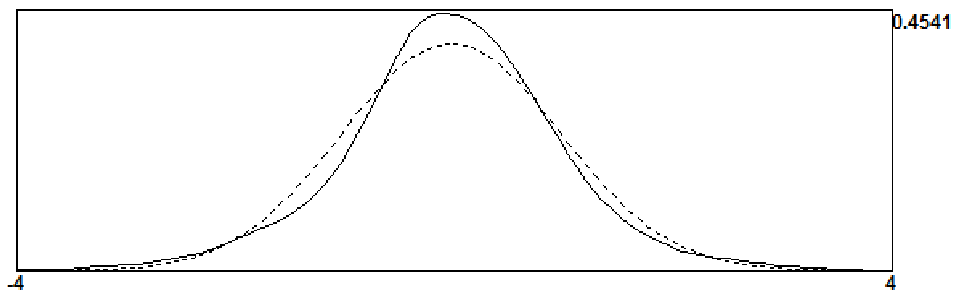

FIGURE 1. Kernel density estimator of ε_t (solid curve)

Of course, it may also be possible that the functional form of $\sigma_t^2(\theta)$ is not of the GARCH type. Various alternative specifications of conditional variances have been proposed in the literature. See for example Nelson (1991) and Rabemananjara and Zakoian (1993). However, exploring these alternative specifications is beyond the scope of this paper.

The rejection of this Gaussian GARCH model raises the question how to distinguish between rejection due to failure of distributional assumptions, i.e., the standard normality of the ε_t's in this case, or failure due to incorrect specification of the parametric part of the model, i.e., the specification of the conditional variance in the GARCH case. A possible solution is to specify a semi-nonparametric GARCH model, with the distribution of the ε_t's specified semi-nonparametrically as suggested in Bierens (2014a, Section 8), and then adapt the current WSICM test to semi-nonparametric conditional distributions. However, this is far from trivial and clearly beyond the scope of the present paper.

11. CONCLUDING REMARKS

This paper extends Bierens (1984) weighted ICM test for functional forms of conditional expectation models to a test for the validity of parametric specifications of conditional distributions for stationary time series data, along the approach in Bierens and Wang (2012). The test is done by conducting a sequence of simulated ICM tests as proposed by Bierens and Wang (2012) for cross-section data, with an increasing number of lagged conditioning variables. The actual test statistic is a weighted sum of these simulated ICM test statistics. This test is consistent against all stationary 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. Finally, similar to Bierens and Wang (2012) it can be shown that this test has nontrivial power against $\sqrt{n}$-local alternatives.

12. APPENDIX

12.1. Proof of Lemma 1

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

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

Let Y_t be a bounded univariate time series and let $\Upsilon = [-c, c]$, where $c > 0$ is arbitrary. Without loss of generality we may interpret $\varphi_{t-1}\left(\tau_0|\theta_*\right)$ as

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

where $\widetilde{Y}_t$ is a random drawing from $G_{t-1}(y|\theta_*)$. Then

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

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 $j_0 > 0$ as otherwise $\varphi_{t-1}\left(\tau_0|\theta_*\right) = \psi_{t-1}\left(\tau_0\right)$ a.s. for all τ_0. For such a j_0,

$$\lim_{m\to\infty} E\left[\widetilde{Y}_t^{j_0} - Y_t^{j_0}\middle| \mathcal{F}_{t-m}^{t-1}\right] = E\left[\widetilde{Y}_t^{j_0} - Y_t^{j_0}\middle| \mathcal{F}_{-\infty}^{t-1}\right]$$

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

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

In its turn this result implies that with positive probability the conditional distribution of $\widetilde{Y}_t$ given $\mathcal{F}_{t-m}^{t-1}$, is unequal to the conditional distribution of Y_t given $\mathcal{F}_{t-m}^{t-1}$, so that the joint distribution of $(\widetilde{Y}_t, Y_{t-1}, ..., Y_{t-m})'$ is unequal to the joint distribution of $(Y_t, Y_{t-1}, ..., Y_{t-m})'$, and so are the corresponding characteristic functions:

$$\begin{aligned}
&\varphi^{m+1}(\tau|\theta_*) - \psi^{m+1}(\tau) \\
&= E\left[\left(\exp\left(\mathbf{i}.\tau_0\widetilde{Y}_t\right) - \exp\left(\mathbf{i}.\tau_0 Y_t\right)\right)\exp\left(\mathbf{i}.\sum_{j=1}^{m}\tau_j Y_{t-j}\right)\right] \\
&\neq 0 \text{ for some } \tau = (\tau_0, \tau_1, ..., \tau_m)' \in \mathbb{R}^{m+1}.
\end{aligned} \tag{12.1}$$

Now suppose that the set $S_{m+1} = \{\tau \in [-c, c]^{m+1} : |\varphi^{m+1}(\tau|\theta_*) - \psi^{m+1}(\tau)| > 0\}$ has zero Lebesgue measure. Then by the boundedness of $(\widetilde{Y}_t, Y_{t-1}, ..., Y_{t-m})'$ and $(Y_t, Y_{t-1}, ..., Y_{t-m})'$ it follows that $\varphi^{m+1}(\tau|\theta_*) = \psi^{m+1}(\tau)$ for all $\tau \in \mathbb{R}^{m+1}$, which contradicts (12.1). This proves Lemma 1 for the univariate case. The multivariate case follows similarly.

12.2. Proof of Theorem 1

Again, let Y_t be a univariate time series and let $\widetilde{Y}_t$ be a random drawing from $G_{t-1}(y|\theta_*)$. If Y_t is bounded then Theorem 1 follows straightforwardly from Lemma 1 and its proof. Therefore, suppose that Y_t is not bounded. Let $\Phi(y)$ be the logistic distribution function, $\Phi(y) = (1 + \exp(-y))^{-1}$, for example, which has inverse $\Phi^{-1}(u) = \ln(u/(1-u))$, $u \in (0,1)$. Let $\overline{G}_{t-1}(u|\theta_*)$ be the conditional distribution of $\Phi(\widetilde{Y}_t)$ given $\mathcal{F}_{-\infty}^{t-1}$, i.e.

$$\overline{G}_{t-1}(u|\theta_*) = \int_{-\infty}^{\infty} I(\Phi(y) \leq u)\mathrm{d}G_{t-1}(y|\theta_*) = G_{t-1}(\Phi^{-1}(u)|\theta_*) \tag{12.2}$$

and similarly,

$$\overline{F}_{t-1}(u) = F_{t-1}(\Phi^{-1}(u)) \tag{12.3}$$

is the conditional distribution of $\Phi(Y_t)$ given $\mathcal{F}_{-\infty}^{t-1}$. Moreover, note that

$$\overline{G}_{t-1}^m(u|\theta_*) \stackrel{\text{def.}}{=} E[\overline{G}_{t-1}(u|\theta_*)|Y_{t-1}, ..., Y_{t-m}]$$

$$
\begin{aligned}
&= E[\overline{G}_{t-1}(u|\theta_*)|\Phi(Y_{t-1}), ..., \Phi(Y_{t-m})] \\
\overline{F}_{t-1}^m(u) &\stackrel{\text{def.}}{=} E[\overline{F}_{t-1}(u)|Y_{t-1}, ..., Y_{t-m}] \\
&= E[\overline{F}_{t-1}(u)|\Phi(Y_{t-1}), ..., \Phi(Y_{t-m})]
\end{aligned}
$$

where the second equalities follow from the fact that

$$(Y_{t-1}, ..., Y_{t-m})' \text{ and } (\Phi(Y_{t-1}), ..., \Phi(Y_{t-m}))'$$

generate the same σ-algebra $\mathcal{F}_{t-m}^{t-1}$. Hence it follows from Lemma 1 that

$$\Pr\left[\sup_{u\in[0,1]} |\overline{G}_{t-1}^m(u|\theta_*) - \overline{F}_{t-1}^m(u)| = 0\right] < 1$$

for all but a finite number of m's, which by (12.2) and (12.3) implies that

$$\Pr\left[\sup_{y\in\mathbb{R}} |G_{t-1}^m(y|\theta_*) - F_{t-1}^m(y)| = 0\right] < 1$$

for all but a finite number of m's.

The proof of the multivariate case is similar.

12.3. Proof of Lemma 4

We will prove Lemma 4 for the case $Y_t \in [-M, M]$ a.s. and $\Upsilon = [-c, c]$ where $c > 0$, as follows. Write (5.4) as

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

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

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

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

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

hence by induction

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

Consequently, for $\varepsilon < 1$,

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

Thus for fixed m the stochastic equicontinuity condition (5.3) holds, so that $\widehat{h}_{1,n,m}$ is tight. The generalization of this results to higher dimensions and more general spaces is straightforward.

12.4. Proof of Lemma 5

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

$$E\left[B_m\right] = \int_{\Upsilon^{m+1}} \Gamma_m\left(\tau,\tau\right) \mathrm{d}\mu_m\left(\tau\right) = \int_{\Upsilon^{m+1}} E\left[\left|\phi_{m,t}(\tau)\right|^2\right] \mathrm{d}\mu_m\left(\tau\right)$$

where Γ_m is the covariance function (5.10). Moreover, observe from (5.9) that $\phi_{m,t}(\tau)$ can be written as

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

where again $E_{t-1}\left[\cdot\right]$ denotes $E\left[\cdot|\mathcal{F}_{-\infty}^{t-1}\right]$. Furthermore, observe from (5.7) and Assumption 6 that

$$\begin{aligned}
\sup_{\tau\in\Upsilon^{m+1}} \|b_m(\tau|\theta_0)\| &\leq \sup_{\tau_0\in\Upsilon} E\left[\|\Delta\varphi_{t-1}\left(\tau_0|\theta_0\right)\|\right] \\
&\leq E\left[\sup_{\tau_0\in\Upsilon}\|\Delta\varphi_{t-1}\left(\tau_0|\theta_0\right)\|\right] < \infty.
\end{aligned}$$

Hence

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

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

It follows now easily from Assumption 3 that

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

Next, observe from (5.8) that

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

and from (5.6) that

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

uniformly in τ and m. Hence

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

uniformly in m.

12.5. Proof of Theorem 3

Assume that the entire past of the time series Y_t involved is observable, so that $\widehat{B}_{n,m}$ can be computed for all m. Then it follows from Lemma 5 that under H_0,

$$\sum_{m=1}^{\infty} \omega_m \widehat{B}_{n,m} - \sum_{m=1}^{L_n} \omega_m \widehat{B}_{n,m} = O_p\left(\sum_{m=L_n+1}^{\infty} \omega_m\right) = o_p(1), \tag{12.4}$$

$$\sum_{m=1}^{\infty} \omega_m B_m - \sum_{m=1}^{L_n} \omega_m B_m = O_p\left(\sum_{m=L_n+1}^{\infty} \omega_m\right) = o_p(1), \tag{12.5}$$

Moreover, for any fixed integer $\ell > 1$ and $m = 1, .., \ell-1$, each $\widehat{B}_{n,m}$ can be written as

$$\widehat{B}_{n,m} = \int_{\Upsilon^{m+1}} \left|\widehat{h}_{n,\ell}((\tau', 0'_{\ell-m})')\right|^2 \mathrm{d}\mu_m(\tau),$$

where $0_{\ell-m}$ is the zero vector in $\mathbb{R}^{\ell-m}$, hence by Theorem 2 and the continuous mapping theorem,

$$\begin{aligned}
\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} &= \sum_{m=1}^{\ell-1} \omega_m \int_{\Upsilon^{m+1}} \left|\widehat{h}_{n,\ell}((\tau', 0'_{\ell-m})')\right|^2 \mathrm{d}\mu_m(\tau) \\
&\quad + \omega_\ell \int_{\Upsilon^{\ell+1}} \left|\widehat{h}_{n,\ell}(\tau)\right|^2 \mathrm{d}\mu_\ell(\tau) \\
&\stackrel{\mathrm{d}}{\to} \sum_{m=1}^{\ell-1} \omega_m \int_{\Upsilon^{m+1}} \left|h_\ell((\tau', 0'_{\ell-m})')\right|^2 \mathrm{d}\mu_m(\tau) \\
&\quad + \omega_\ell \int_{\Upsilon^{\ell+1}} \left|h_\ell(\tau)\right|^2 \mathrm{d}\mu_\ell(\tau) \\
&= \sum_{m=1}^{\ell} \omega_m B_m.
\end{aligned}$$

To prove that the latter convergence result carries over for $\ell = \infty$, observe that for $x > 0$,

$$\begin{aligned}
\Pr\left[\sum_{m=1}^{\infty} \omega_m \widehat{B}_{n,m} \le x\right] &= \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \sum_{m=\ell+1}^{\infty} \omega_m \widehat{B}_{n,m}\right] \\
&\le \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x\right] \\
&\ge \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \left(\sum_{m=\ell+1}^{\infty} \omega_m\right) \sup_{k\ge 1} \widehat{B}_{n,k}\right].
\end{aligned}$$

The latter probability can be bounded from below further as follows. For arbitrary $\varepsilon > 0$,

$$\begin{aligned}
&\Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \left(\sum_{m=\ell+1}^{\infty} \omega_m\right) \sup_{k\ge 1} \widehat{B}_{n,k}\right] \\
&= \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \left(\sum_{m=\ell+1}^{\infty} \omega_m\right) \sup_{k\ge 1} \widehat{B}_{n,k} \wedge \sup_{k\ge 1} \widehat{B}_{n,k} \le \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty} \omega_m}\right] \\
&+ \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \left(\sum_{m=\ell+1}^{\infty} \omega_m\right) \sup_{k\ge 1} \widehat{B}_{n,k} \wedge \sup_{k\ge 1} \widehat{B}_{n,k} > \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty} \omega_m}\right] \\
&\ge \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \varepsilon \wedge \sup_{k\ge 1} \widehat{B}_{n,k} \le \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty} \omega_m}\right] \\
&= \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \varepsilon\right] \\
&- \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \varepsilon \wedge \sup_{k\ge 1} \widehat{B}_{n,k} > \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty} \omega_m}\right] \\
&\ge \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x - \varepsilon\right] - \Pr\left[\sup_{k\ge 1} \widehat{B}_{n,k} > \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty} \omega_m}\right]
\end{aligned}$$

Hence,

$$\limsup_{n\to\infty} \Pr\left[\sum_{m=1}^{\infty} \omega_m \widehat{B}_{n,m} \le x\right] \le \lim_{n\to\infty} \Pr\left[\sum_{m=1}^{\ell} \omega_m \widehat{B}_{n,m} \le x\right]$$

$$= \Pr\left[\sum_{m=1}^{\ell}\omega_m B_m \le x\right] \tag{12.6}$$

$$\begin{aligned}\liminf_{n\to\infty}\Pr\left[\sum_{m=1}^{\infty}\omega_m \widehat{B}_{n,m} \le x\right] &\ge \lim_{n\to\infty}\Pr\left[\sum_{m=1}^{\ell}\omega_m \widehat{B}_{n,m} \le x-\varepsilon\right] \\ &\quad - \limsup_{n\to\infty}\Pr\left[\left(\sum_{m=\ell+1}^{\infty}\omega_m\right)\sup_{k\ge 1}\widehat{B}_{n,k} > \varepsilon\right] \\ &= \Pr\left[\sum_{m=1}^{\ell}\omega_m B_m \le x-\varepsilon\right] \\ &\qquad - \limsup_{n\to\infty}\Pr\left[\sup_{k\ge 1}\widehat{B}_{n,k} > \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty}\omega_m}\right] \end{aligned} \tag{12.7}$$

Furthermore, it follows straightforwardly from the trivial inequality

$$E\left[\left|\sum_{m=1}^{\ell}\omega_m B_m - \sum_{m=1}^{\infty}\omega_m B_m\right|\right] \le \sum_{m=\ell+1}^{\infty}\omega_m \sup_{k\ge 1} E[B_k]$$

and Chebyshev inequality for first moments that

$$p\lim_{\ell\to\infty}\sum_{m=1}^{\ell}\omega_m B_m = \sum_{m=1}^{\infty}\omega_m B_m,$$

hence

$$\lim_{\ell\to\infty}\Pr\left[\sum_{m=1}^{\ell}\omega_m B_m \le x\right] = \Pr\left[\sum_{m=1}^{\infty}\omega_m B_m \le x\right]$$

where now x is an arbitrary continuity point of the distribution of $\sum_{m=1}^{\infty}\omega_m B_m$. Also, note that

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\Pr\left[\sup_{k\ge 1}\widehat{B}_{n,k} > \frac{\varepsilon}{\sum_{m=\ell+1}^{\infty}\omega_m}\right] = 0.$$

It follows now from (12.6) and (12.7) by letting $\ell\to\infty$ that

$$\begin{aligned}\limsup_{n\to\infty}\Pr\left[\sum_{m=1}^{\infty}\omega_m \widehat{B}_{n,m} \le x\right] &\le \Pr\left[\sum_{m=1}^{\infty}\omega_m B_m \le x\right], \\ \liminf_{n\to\infty}\Pr\left[\sum_{m=1}^{\infty}\omega_m \widehat{B}_{n,m} \le x\right] &\ge \Pr\left[\sum_{m=1}^{\infty}\omega_m B_m \le x-\varepsilon\right],\end{aligned}$$

hence, letting $\varepsilon \downarrow 0$ yields

$$\lim_{n\to\infty} \Pr\left[\sum_{m=1}^{\infty} \omega_m \widehat{B}_{n,m} \leq x\right] = \Pr\left[\sum_{m=1}^{\infty} \omega_m B_m \leq x\right]$$

for all continuity points x of the distribution of $\sum_{m=1}^{\infty} \omega_m B_m$. In other words,

$$\sum_{m=1}^{\infty} \omega_m \widehat{B}_{n,m} \xrightarrow{\mathrm{d}} \sum_{m=1}^{\infty} \omega_m B_m \tag{12.8}$$

Combining (12.4) and (12.8) yields the result of Theorem 3 under H_0. The result under H_1 follows trivially from Theorem 2.

ACKNOWLEDGMENT

The helpful comments of the Editor, an Associate Editor, and two referees, leading to substantial improvement of the paper over previous versions, are gratefully acknowledged.

REFERENCES

Andrews, D. W. (1997). A conditional Kolmogorov test. *Econometrica* 65:1097-1128.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85:531-549.

Bai, J., Chen, Z. (2008). Testing multivariate mistributions in GARCH models. *Journal of Econometrics* 143:19-36.

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. (2004). *Introduction to the Mathematical and Statistical Foundations of Econometrics.* Cambridge, U.K.: Cambridge University Press.

Bierens, H. J. (2014a). Consistency and asymptotic normality of sieve ML estimators under low-level conditions. *Econometric Theory 30:* Forthcoming.

Bierens, H. J. (2014b). Addendum to Integrated conditional moment tests for parametric conditional distributions.[16] Unpublished manuscript, downloadable

[16] See the reference Bierens and Wang (2012).

from URL http://econ.la.psu.edu/~hbierens/ADDENDUM_BW2012.PDF.

Bierens, H. J. (2014c). *EasyReg International, Version June 12, 2014.* Downloadable from URL http://econ.la.psu.edu/~hbierens/EASYREG.HTM.

Bierens, H. J., Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* 65:1129-1151.

Bierens, H. J., Wang, L. (2012). Integrated conditional moment tests for parametric conditional distributions. *Econometric Theory* 28:328-362.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307-327.

Boning, B., Sowell, F. (1999). Optimality of the integrated conditional moment test. *Econometric Theory* 15:710-719.

Chen, X., Fan, Y. (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91:373-401.

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

Corradi, V., Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133:779-806.

Corradi, V., Swanson, N. R. (2007). Evaluation of dynamic stochastic general equilibrium models based on distributional comparison of historical and simulated data. *Journal of Econometrics* 136:699-723.

de Jong, R. (1996). The Bierens test under data dependence. *Journal of Econometrics* 72:1-32.

Domínguez, M., Lobato, I. (2003). Testing the martingale difference hypothesis. *Econometric Reviews* 22:351-377.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987-1007.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* 22:1030-1051.

Escanciano, J. C., Velasco, C. (2006). Generalized spectral density for the martingale difference hypothesis test. *Journal of Econometrics* 134:151-185.

Hannan, E. J., Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society Series B* 41:190-195.

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., 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.

Fan, Y. Q., Li, Q. (2000). Consistent model specification tests: Kernel based tests versus Bierens' ICM tests. *Econometric Theory* 16:1016-1041.

Gozalo, P. L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9:451- 477.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21:1926-1947.

Hong, Y., White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica* 63:1133-1159.

Horowitz, J. T., Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10:821-848.

Lavergne, P., Vuong, Q. H. (2000). A nonparametric significance test. *Econometric Theory* 16:576-601.

Khmaladze, E. (1981). Martingale approach in the goodness-of-fit tests. *Theory of Probability and its Applications* 26:240-257.

Li, F., 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 mpecification tests for time series econometric models. *Journal of Econometrics* 92:101-147.

Li, Q., Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87:145-165.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Annals of Probability* 2:620-628.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* 59:347-370.

Rabemananjara, R., Zakoian, J. M. (1993). Threshold Arch models and asymmetries in volatility. *Journal of Applied Econometrics* 8:31-49.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461-464.

Stinchcombe, M., White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14:295-325.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics* 25:613-641.

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

Su, L., White, H. (2007). A Consistent characteristic function-based test for conditional dependence. *Journal of Econometrics* 141:807-834.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Wessa, P. (2013). Free statistics software. Office for Research Development and Education. http://www.wessa.net/.

White, H. (1987). Specification testing in dynamic models. In: Bewley, T. F., ed. *Advances in Econometrics, Fifth World Congress, Vol.I.* Cambridge, U.K.: Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge, U.K.: Cambridge University Press.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory* 8:452-475.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory* 8:435-451.

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75:263-289.

Zheng, J.X. (2000). A consistent test of conditional parametric distributions. *Econometric Theory* 16:667-691.