Generalized spectral tests for the martingale difference hypothesis

J. Carlos Escanciano[a,*], Carlos Velasco[b]

[a]*Facultad de Económicas, Universidad de Navarra, Edificio Biblioteca (Entrada Este), Pamplona, 31080 Navarra, Spain*
[b]*Departamento de Economía, Universidad Carlos III de Madrid, Spain*

Abstract

This article proposes a test for the martingale difference hypothesis (MDH) using dependence measures related to the characteristic function. The MDH typically has been tested using the sample autocorrelations or in the spectral domain using the periodogram. Tests based on these statistics are inconsistent against uncorrelated non-martingales processes. Here, we generalize the spectral test of Durlauf (1991) for testing the MDH taking into account linear and nonlinear dependence. Our test considers dependence at all lags and is consistent against general pairwise nonparametric Pitman's local alternatives converging at the parametric rate $n^{-1/2}$, with n the sample size. Furthermore, with our methodology there is no need to choose a lag order, to smooth the data or to formulate a parametric alternative. Our approach could be extended to specification testing of the conditional mean of possibly nonlinear models. The asymptotic null distribution of our test depends on the data generating process, so a bootstrap procedure is proposed and theoretically justified. Our bootstrap test is robust to higher order dependence, in particular to conditional heteroskedasticity. A Monte Carlo study examines the finite sample performance of our test and shows that it is more powerful than some competing tests. Finally, an application to the S&P 500 stock index and exchange rates highlights the merits of our approach.
© 2005 Elsevier B.V. All rights reserved.

JEL classification: C12

Keywords: Martingale difference hypothesis; Hilbert spaces; Generalized spectral distribution; Characteristic function; S&P 500 stock index; Exchange rates

---

*Corresponding author.
E-mail address:* jescanci@unav.es (J.C. Escanciano).

1. Introduction

The concept of martingale or martingale difference sequence (MDS) is central in many areas of economics and finance. Many economic theories in a dynamic context, such as the market efficiency hypothesis, rational expectations or optimal consumption smoothing deliver such dependence restrictions on the underlying economic variables, see e.g. Hall (1978) or Lo (1997). The martingale difference hypothesis (MDH) states that the best predictor, in the sense of least mean square error, of the future values of a time series given the current information set is just the unconditional expectation. Hence, past information does not help to improve the forecast of future values of a MDS. More formally, given a real valued stationary time series $\{Y_t\}_{-\infty}^{\infty}$, we tackle the problem of testing that almost surely (a.s)

$$E[Y_t | Y_{t-1}, Y_{t-2}, \ldots] \quad \mu, \quad \mu \in \mathbb{R}. \tag{1}$$

In the literature there have been several proposals for testing the MDH, some of them using a spectral approach which has been shown useful in studying serial dependence, see e.g. Hong (1999), Paparoditis (2000) or Delgado et al. (2004) for recent references. Durlauf (1991) proposed a spectral distribution based test using the fact that, under the MDH, the standardized spectral distribution function is a straight line. Deo (2000) has extended Durlauf's (1991) test to allow for some types of conditional heteroskedasticity. However, these tests are suitable for testing lack of correlation but not for the MDH. In fact, they are not consistent against non martingale difference sequences with zero autocorrelations, that is, when only nonlinear dependence is present (for instance, see in Section 5 the NLMA process). The inconsistency of these tests arises because they only employ information contained in the second moments of the process, which in the case of nonlinear dependence can not completely characterize the conditional mean. The objective of this paper is to develop a test that can overcome this difficulty, that is, a test consistent against a large class of uncorrelated non martingale processes.

One way to circumvent this problem is to take into account higher moments, for instance the third order cumulants used in Hinich and Patterson (1992), but again this test is not consistent against non martingale difference sequences with zero third order cumulants.

We use instead an alternative methodology based on the following equivalence principle. Let I_t $\{Y_t, Y_{t-1}, \ldots\}$ and $\mathscr{F}_t$ be the information set at time t and the σ field generated by I_t, respectively. Then,

$$E[Y_t | I_{t-1}] \quad \mu \text{ a.s., } \mu \in \mathbb{R} \iff E[(Y_t \quad \mu)w(I_{t-1})] \quad 0, \tag{2}$$

for any bounded $\mathscr{F}_{t-1}$ measurable weighting function $w(\cdot)$. Consistent tests can be based on the discrepancy of the sample analog of $E[(Y_t \quad \mu)w(I_{t-1})]$ to zero.

The problem of testing over all possible weighting functions can be reduced to testing the orthogonality condition over a parametric family of functions, see Stinchcombe and White (1998). However, how to choose this family is not an easy problem. In the specification testing literature several tests have been proposed using this approach. The exponential function has been considered in Bierens (1984, 1990),

de Jong (1996), Bierens and Ploberger (1997) or Hong (1999a, b), among others. On the other hand, indicator functions have been used in Stute (1997), Koul and Stute (1999), Park and Whang (1999), Whang (2000) or Domínguez and Lobato (2003). Although much effort has been devoted to testing with these families, none of the above papers has compared both approaches because the different properties of the weighting function usually demand alternative technical frameworks.

Among all the tests based on the equivalence (2), only de Jong (1996) considers the case in which the conditioning variable I_{t-1} is infinite dimensional. de Jong (1996) generalized Bierens' tests to time series, and although this test has the appealing property of considering an increasing number of lags as the sample size increases, it requires numerical integration with dimension equal to the sample size, which makes this test unfeasible in applications where the sample size is usually large, e.g. financial applications.

Our approach has its foundations in a series of works due to Hong (1999a, b). Hong (1999a) introduces a generalized spectral density as a new tool for testing interesting hypotheses in a nonlinear time series framework. In particular, among other tests, Hong (1999a) proposed a MDH test using a smoothed kernel estimator of a generalized spectral density with standard normal asymptotic distribution. Hong's (1999a) MDH test has the appealing property of being asymptotic distribution free, but also presents the serious drawback that the asymptotic null distribution is only obtained under independence. Recently, Hong and Lee (2003, 2005) have extended Hong's (1999a) MDH test to a class of processes which may display conditional dependence in second and higher moments. The main differences between our test and the MDH test of Hong and Lee (2003, 2005) are threefold. First, we do not assume any asymptotic independence under the null of MDH and we only assume finite variance. The fourth moment necessary in Hong and Lee (2003, 2005) may look restrictive. It rules out many empirically relevant GARCH processes whose fourth moments are often found to be infinite. Second, our test does not depend on a kernel and a bandwidth parameter but is based on a generalized spectral distribution instead of a generalized spectral density function. And third, under dependence the asymptotic null distribution of our test depends on the data generating process (DGP) and is no longer standard. Hence a bootstrap approach will be considered and justified theoretically.

Our test is more in the spirit of Hong's (1999b) test, which is based on a generalized spectral distribution function. In fact, our test can be derived by appropriately differentiating this generalized spectral distribution. Note that Hong's (1999b) test is a test for serial independence, and not a proper test for the MDH. In fact, his test is not valid in the presence of conditional heteroskedasticity, e.g. for ARCH models which are MDS but dependent processes. Our approach synthesizes the approaches of Hong (1999b) and Deo (2000) allowing for serial dependence under the null of MDH, in particular conditional heteroskedasticity and higher order dependence.

A related work to our problem is Kuan and Lee (2003). These authors propose a MDH test based on the exponential function with asymptotic χ_2^2 distribution. Unlike other related papers, Kuan and Lee do not consider a "functional space" approach,

because they first integrate the auxiliary parameter and afterwards consider a norm, consequently the omnibus property is lost. As a matter of fact, it is not difficult to show that their test is equivalent to check for correlation (lack thereof) between Y_t and a unique function $w(\cdot)$ in (2), say $w_0(\cdot)$, so eventually their test has power only against some particular alternatives, more specifically those correlated with the function $w_0(\cdot)$. These limitations are confirmed in our simulations.

The layout of the article is as follows, in Section 2 we propose the test and in Section 3 we study its asymptotic distribution under the null and under fixed and local alternatives. In Section 4 we propose and justify a bootstrap approach. In Section 5 we provide a simulation exercise comparing among different MDH tests proposed in the literature and we apply our methodology to the daily S&P 500 stock index and some exchange rates returns. We finish in Section 6 with further research and some conclusions. All proofs are gathered in an appendix. Throughout, $|A|$, A^{T} and A^{c} denote the usual Euclidean norm, the matrix transpose and the complex conjugate, respectively, i $\sqrt{}$ 1 and unless indicated all limits are taken as the sample size $n \rightarrow \infty$. Here and in the sequel C is a generic constant that may change from one expression to another.

2. The generalized spectral martingale test

Given raw data $\{Y_t\}_{t\ \ 1}^n$ we are interested in testing the hypothesis (1),

$\mathrm{H}_0^* :\ \{Y_t\}$ is a MDS.

Alternatively, one may test the hypotheses $\mathrm{E}[Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_{t-P}]\quad \mu$ with P tending to infinite with the sample size, but this approach brings some problems up, especially from a practical point of view as can be seen in de Jong's (1996) test which involves P dimensional integration. We could also consider P fixed, as in Koul and Stute (1999), Park and Whang (1999), Kuan and Lee (2003) or Domínguez and Lobato (2003), but this is unsatisfactory from a theoretical point of view because there could be structure in the conditional mean at omitted lags. Often, the maximum power is achieved by using the correct lag order of the alternative. However, prior information on the conditional mean structure is usually not available. We consider a pairwise approach that takes into account all lags available in the sample and at the same time avoids high dimensional integration, delivering a test easy to compute. There is a price to pay for the computational feasibility, the omnibus property of the test procedure is lost. Our pairwise approach represents a good compromise between generality and simplicity, leading to a test that is consistent against a broad class of alternatives. More specifically, it is consistent against all pairwise alternatives, including most empirically relevant models. Extensions to higher order dependence are possible and are considered in the final section.

Our proposal for testing the MDH consists in checking all the pairwise implications of (1), namely, our null hypothesis is now

$$\mathrm{H}_0 : m_j(y)\quad 0\quad \forall j \geqslant 1,\ \text{a.s.}, \tag{3}$$

where $m_j(y) \equiv \mathrm{E}[Y_t - \mu \mid Y_{t-j} = y]$ a.s. are the pairwise regression functions. The alternative is the negation of the null (3), i.e.,

H_A : there exists a $j \geqslant 1$ such that $P(m_j(Y_{t-j}) \neq 0) > 0.$

To this end, the conditional mean dependence measures

$$\gamma_j(x) \equiv \mathrm{E}[(Y_t - \mu)\mathrm{e}^{\mathrm{i}x Y_{t-j}}],$$

can be viewed as a generalization of the usual autocovariances to measure the conditional mean dependence in a nonlinear time series framework. We use the following characterization of the null hypothesis (3), see Theorem 1 in Bierens (1982),

$$\mathrm{H}_0 \Longleftrightarrow \gamma_j(x) = 0 \quad \forall j \geqslant 1, \quad \text{almost everywhere (a.e.).} \tag{4}$$

The sample counterpart of $\gamma_j(x)$ based on a sample $\{Y_t\}_{t=1}^n$ is

$$\widehat{\gamma}_j(x) = \frac{1}{n-j}\sum_{1+j}^{n}(Y_t - \overline{Y}_{n-j})\mathrm{e}^{\mathrm{i}x Y_{t-j}},$$

with

$$\overline{Y}_{n-j} = \frac{1}{n-j}\sum_{1+j}^{n} Y_t.$$

Define $\gamma_{-j}(\cdot) = \gamma_j(\cdot)$ for $j \geqslant 1$, and consider the Fourier transform of the functions $\gamma_j(x)$,

$$f(\varpi, x) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j(x)\mathrm{e}^{-\mathrm{i}j\varpi} \quad \forall \varpi \in [-\pi, \pi], \ x \in \mathbb{R}. \tag{5}$$

Notice that $f(\varpi, x)$ exists if

$$\sup_{x \in \mathbb{R}}\sum_{j=-\infty}^{\infty}|\gamma_j(x)| < \infty,$$

which holds under a proper mixing condition. To test the serial independence between Y_t and Y_{t-j} Hong (1999a) proposed the measures

$$\sigma_j(y, x) = cov(\mathrm{e}^{\mathrm{i}y Y_t}, \mathrm{e}^{\mathrm{i}x Y_{t-|j|}}) = \varphi_{|j|}(y, x) - \varphi(y)\varphi(x),$$

where $\varphi_j(y, x)$ is the joint characteristic function of (Y_t, Y_{t-j}) and $\varphi(y)$ is the marginal characteristic function. He considered the Fourier transform of $\sigma_j(y, x)$

$$g(\varpi, y, x) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(y, x)\mathrm{e}^{-\mathrm{i}j\varpi} \quad \forall \varpi \in [-\pi, \pi], (y, x) \in \mathbb{R}^2,$$

as a generalized spectral density function. Hong (1999a) used kernel estimators of $g(\varpi, y, x)$ and its derivatives to test relevant hypotheses in time series. In particular, he showed that the standard spectral density can be derived by properly

differentiating the generalized spectral density. Note that by similar arguments

$$\frac{\partial \sigma_j(y,x)}{\partial y}\bigg|_{y=0} = i\gamma_j(x), \tag{6}$$

and then

$$\frac{\partial g(\varpi,y,x)}{\partial y}\bigg|_{y=0} = if(\varpi,x), \tag{7}$$

which relates the generalized spectral density of Hong (1999a) with the generalized spectral density $f(\varpi,x)$. The introduction of the auxiliary parameters y and x, renders to $g(\varpi,y,x)$ be able to capture all pairwise dependencies. The same argument shows that $f(\varpi,x)$ is able to capture pairwise non martingale difference alternatives with zero autocorrelations. Furthermore, the conventional spectral density function can be obtained under finite variance by differentiating $f(\varpi,x)$ with respect to x at the origin, or equivalently by differentiating $g(\varpi,y,x)$ with respect y and x at $(0,0)$, see Hong (1999a) for details. The generalized spectral density $f(\varpi,x)$ contains the same information about the null hypothesis H$_0$ as the whole sequence $\{\gamma_j(x)\}_{j=0}^{\infty}$. In particular, the null hypothesis (3) is equivalent to $f_0(\varpi,x) = (2\pi)^{-1}\gamma_0(x)$. Hong (1999a) proposed the estimators

$$\widehat{f}(\varpi,x) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1}\left(1-\frac{|j|}{n}\right)^{1/2}k\left(\frac{j}{p}\right)\widehat{\gamma}_j(x)e^{-ij\varpi}$$

and

$$\widehat{f}_0(\varpi,x) = \frac{1}{2\pi}\widehat{\gamma}_0(x)$$

to test the MDH, where $k(\cdot)$ is a symmetric kernel and p a bandwidth parameter. He considered a standardization of a L_2 distance using a weighting function $W(\cdot)$

$$L_2^2(p) = \frac{\pi}{2}\int_{\mathbb{R}}\int_{-\pi}^{\pi}n|\widehat{f}(\varpi,x)-\widehat{f}_0(\varpi,x)|^2W(dx)\,d\varpi$$
$$= \sum_{j=1}^{n-1}(n-j)k^2\left(\frac{j}{p}\right)\int_{\mathbb{R}}|\widehat{\gamma}_j(x)|^2W(dx). \tag{8}$$

The centering and scaling factors in the standardization of $L_2^2(p)$ to obtain an asymptotic standard normal null distribution depend on the higher dependence structure between the errors and the regressors and are given in Hong and Lee (2003, 2005). Instead of considering kernel estimation as in Hong (1999a), we propose to use a generalized spectral distribution function

$$H(\lambda,x) = 2\int_0^{\lambda\pi}f(\varpi,x)\,d\varpi \quad \forall\lambda\in[0,1], x\in\mathbb{R},$$

that is

$$H(\lambda, x) \quad \gamma_0(x)\lambda + 2 \sum_{j \ 1}^{\infty} \gamma_j(x) \frac{\sin j\pi\lambda}{j\pi}, \tag{9}$$

which is related via (7) with the generalized spectral distribution considered in Hong (1999b) to test for serial independence. The new test is based on the sample analogue of (9),

$$\widehat{H}(\lambda, x) \quad \widehat{\gamma}_0(x)\lambda + 2 \sum_{j \ 1}^{n-1} \left(1 \quad \frac{j}{n}\right)^{1/2} \widehat{\gamma}_j(x) \frac{\sin j\pi\lambda}{j\pi},$$

with $(1 \quad j/n)^{1/2}$ a finite sample correction factor as in Hong (1999a). The effect of this correction factor is to put less weight on very large lags, for which we have less sample information. The test with this factor shows a better finite sample performance, specially in terms of the empirical power of the test. It has no effect on the asymptotic distribution and can be deleted without changing the asymptotic theory. Because (3) is equivalent to $H(\lambda, x) \quad \gamma_0(x)\lambda$, consistent tests can be based on the discrepancy between $\widehat{H}(\lambda, x)$ and $\widehat{H}_0(\lambda, x)$: $\widehat{\gamma}_0(x)\lambda$. That is, we consider the process

$$S_n(\lambda, x) \quad \left(\frac{n}{2}\right)^{1/2} \{\widehat{H}(\lambda, x) \quad \widehat{H}_0(\lambda, x)\} \quad \sum_{j \ 1}^{n-1} (n \quad j)^{1/2} \widehat{\gamma}_j(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}, \tag{10}$$

to test H_0.

Durlauf (1991) proposed a MDH test based on the usual standardized spectral distribution function, that is, a standardization of

$$\mathrm{DUR}(\lambda) \quad \sum_{j \ 1}^{n-1} n^{1/2}\widehat{\gamma}_{j1} \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}, \tag{11}$$

where $\widehat{\gamma}_{j1}$ is the usual sample autocovariance at lag j. The process $S_n(\lambda, x)$ can be viewed as a generalization of the process used in Durlauf (1991). To see this, fix the sample and consider the formal Taylor expansion of $\widehat{\gamma}_j(x)$ around $x \quad 0$,

$$\widehat{\gamma}_j(x) \quad \sum_{k \ 1}^{\infty} \frac{\mathrm{i}^k x^k}{k!} \widetilde{\gamma}_{jk}, \tag{12}$$

where

$$\widetilde{\gamma}_{jk} \quad \frac{1}{n \quad j} \sum_{1+j}^{n} (Y_t \quad \overline{Y}_{n-j}) Y_{t-j}^k.$$

Hence, substituting (12) into (10) yields

$$S_n(\lambda, x) \quad \sum_{k \ 1}^{\infty} \sum_{j \ 1}^{n-1} (n \quad j)^{1/2} \frac{\mathrm{i}^k x^k}{k!} \widetilde{\gamma}_{jk} \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}. \tag{13}$$

7

It can be seen from (13) that $S_n(\lambda, x)$ can be expressed as an infinite order polynomial in x, where the linear term, i.e. the term corresponding to $k = 1$, is

$$\mathrm{i}x \sum_{j=1}^{n-1} (n-j)^{1/2} \widetilde{\gamma}_{j1} \frac{\sqrt{2}\sin j\pi\lambda}{j\pi},$$

which is equal to $\mathrm{i}x DUR(\lambda)$ except for the use of $(n-j)^{1/2}$ and $\overline{Y}_{n-j}$ instead of $n^{1/2}$ and the usual sample mean $\overline{Y}_n = n^{-1}\sum_{t=1}^{n} Y_t$, respectively. Note that $S_n(\lambda, x)$ takes into account all the infinite terms in k in (13). For instance, the coefficient of the second term in (13) is a measure of the linear dependence between Y_t and Y_{t-j}^2, that is, the so called ARCH in mean effects. Therefore, $S_n(\lambda, x)$ uses the dependence measures $\widetilde{\gamma}_{jk}$ for all values of j and k whereas $DUR(\lambda)$ only considers the case $k = 1$.

In order to evaluate the distance of $S_n(\lambda, x)$ to zero, a norm has to be chosen. One norm considered in practice is the Cramér–von Mises norm

$$D_n^2 = \int_{\mathbb{R}} \int_0^1 |S_n(\lambda, x)|^2 W(\mathrm{d}x)\,\mathrm{d}\lambda = \sum_{j=1}^{n-1} (n-j) \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\widehat{\gamma}_j(x)|^2 W(\mathrm{d}x), \qquad (14)$$

where $W(\cdot)$ is a weighting function satisfying some mild conditions (seee Assumption A2 below). Our generalized spectral martingale test rejects the null hypothesis (3) for large values of D_n^2. Notice that D_n^2 uses all the $n-1$ lags contained in the sample, so we have not to choose any lag order parameter. Note the similarity between (8) and (14). $L_2^2(p)$ and D_n^2 differ only in the weighting scheme. The optimal weighting scheme depends on the alternative at hand. $L_2^2(p)$ involves the choice of a kernel and smoothing parameter and although the inferences could be sensitive to these choices, smoothing may give more flexibility in directing the power towards some desired directions. On the other hand, D_n^2 has the attractive convenience of being free of choosing any smoothing parameter or kernel. Therefore, D_n^2 should be viewed as not competing but as a complement to $L_2^2(p)$.

3. Asymptotic theory

In this section, we first establish the null limit distribution of the process $S_n(\lambda, x)$. Let $\eta = (\lambda, x) \in \Pi = [0,1] \times (-\infty, \infty)$ and v the product measure of $W(\cdot)$ and the Lebesgue measure on $[0,1]$, i.e. $\mathrm{d}v(\eta) \equiv \mathrm{d}v(\lambda, x) = W(\mathrm{d}x)\,\mathrm{d}\lambda$. Then, we can consider the process $S_n(\eta) \equiv S_n(\lambda, x)$ as a random element in the Hilbert space $L_2(\Pi, v)$ of all square integrable functions (with respect to the measure v) with the inner product

$$\langle f, g \rangle = \int_\Pi f(\eta) g^c(\eta)\,\mathrm{d}v(\eta) = \int_\Pi f(\lambda, x) g^c(\lambda, x) W(\mathrm{d}x)\,\mathrm{d}\lambda.$$

$L_2(\Pi, v)$ is endowed with the natural Borel σ–field induced by the norm $\|f\| = \langle f, f \rangle^{1/2}$, see Chapter VI in Parthasarathy (1967) for convergence results on Hilbert spaces. For recent applications in the econometrics literature see Politis and Romano (1994), Chen and White (1996, 1998) or Chen and Fan (1999). If Z is a $L_2(\Pi, v)$ valued random element and has probability distribution μ_Z, we say that Z has mean m if $\mathrm{E}[\langle Z, h \rangle] = \langle m, h \rangle$ $\forall h \in L_2(\Pi, v)$. If $\mathrm{E}\|Z\|^2 < \infty$ and Z has zero mean,

then, the covariance operator of Z (or μ_Z), $C_Z(\cdot)$ say, is a continuous, linear, symmetric, positive definite operator from $L_2(\Pi, v)$ to $L_2(\Pi, v)$ defined by

$$C_Z(h) \quad \mathrm{E}[\langle Z, h \rangle Z].$$

An operator s on a Hilbert space is called nuclear if it can be represented as $s(h) \quad \sum_{j=1}^{\infty} l_i \langle h, f_j \rangle f_j$, where $\{f_j\}$ is an orthonormal basis of the Hilbert space and $\{l_i\}$ is a real sequence such that $\sum_{j=1}^{\infty} |l_i| < \infty$. It is easy to show, see e.g. Bosq (2000), that the covariance operator $C_Z(\cdot)$ is a nuclear operator provided that $\mathrm{E}\|Z\|^2 < \infty$. Let $\Rightarrow$ denote weak convergence in the Hilbert space $L_2(\Pi, v)$ endowed with the norm metric. To derive the asymptotic theory we consider the following assumptions.

Assumption A1. A1(a): $\{Y_t\}$ is a strictly stationary and ergodic process.
A1(b): $\mathrm{E}|Y_1|^2 < C$.

Assumption A2. W is a probability measure on $\mathbb{R}$, absolutely continuous with respect to Lebesgue measure.

Note that Assumption A1(a) is mild, in particular it allows us to consider conditional heteroskedastic processes. A1(b) only assumes finite variance in contrast with the eighth moment necessary in Durlauf (1991) and Deo (2000) or the fourth moment in Hong and Lee (2003, 2005).

The choice of the weighting function $W(\cdot)$ has implications on the power performance of the test. The optimal choice of $W(\cdot)$ depends on the true alternative at hand. Also, it can be shown (cf. Neuhaus, 1976) that the directions of maximum power of the test depend on the inner product considered in $L_2(\Pi, v)$, which at the same time depends basically on the function $W(\cdot)$. But on the other hand, because an analytic characteristic function is determined in a neighborhood of the origin, it appears that the most relevant point is the behaviour of $W(\cdot)$ at the origin. Here, following Epps and Pulley (1983) three considerations influence the choice of $W(\cdot)$. The first is that it should assign high weight where $|\gamma_j(x)|$ is large under the alternative. The second is that $W(\cdot)$ should give high weight where the statistics $\widehat{\gamma}_j(x)$ are precise estimators. It can be shown that this precision is greatest near the origin. A final practical consideration is that, the choice of $W(\cdot)$ should be such that D_n^2 has a closed form, examples of these weighting functions are the normal or exponential cumulative distribution functions (cdf.). For instance, in the exponential case with parameter β, as in Kuan and Lee (2003), we have that

$$D_n^2 \quad \sum_{j=1}^{n-1} (n \quad j) \frac{1}{(j\pi)^2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} (Y_t \quad \overline{Y}_{n-j})(Y_s \quad \overline{Y}_{n-j}) \frac{1 + \beta(Y_{t-j} \quad Y_{s-j})}{1 + \beta^2 (Y_{t-j} \quad Y_{s-j})^2}.$$

For some discussion about the choice of β see Kuan and Lee (2003). On the other hand, if we take $W(\cdot)$ as the cdf. of a standard normal random variable, then

$$D_n^2 \quad \sum_{j=1}^{n-1} (n \quad j) \frac{1}{(j\pi)^2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} (Y_t \quad \overline{Y}_{n-j})(Y_s \quad \overline{Y}_{n-j}) \exp(\quad 0.5(Y_{t-j} \quad Y_{s-j})^2).$$

(15)

These choices of $W(\cdot)$ avoid numerical integration in the computation of D_n^2. In any case, any $W(\cdot)$ satisfying A2 will deliver a consistent test against all pairwise alternatives. Some simulations in Hong (1999a) confirm that the power results for similar tests to D_n^2 are not too sensitive to the choice of $W(\cdot)$. It can be shown that the choice of $W(\cdot)$ is similar to the choice of the kernel function in non parametric smoothing, see Fan and Li (2000). Finally, note that unbounded support of $W(\cdot)$ is generally necessary because the characteristic functions of two different distributions may coincide on a finite interval. However, when Y_t is bounded, any continuous $W(\cdot)$ with a bounded support containing the origin will ensure consistency against all pairwise non MDS alternatives.

3.1. Asymptotic null distribution

Using a martingale central limit theorem it can be shown, see the proof of Theorem 1, that the finite dimensional projections $\langle S_n, h \rangle$, with $h \in L_2(\Pi, v)$, converge in distribution under (1) and Assumptions A1 A2 to a Gaussian random variable with zero mean and asymptotic variance

$$\sigma_h^2 \quad \sum_{j\ 1}^{\infty} \sum_{k\ 1}^{\infty} \mathrm{E}[(Y_1 \quad \mu)^2 \int_{\Pi \times \Pi} h(\eta) h^c(\eta') \phi_{t-j}^c(x) \phi_{t-k}(x') \Psi_j(\lambda) \Psi_k(\lambda') \, dv(\eta) \, dv(\eta')],$$

(16)

where $\phi_t(x): \exp(\mathrm{i} x Y_t) \quad \varphi(x)$, $\Psi_j(\lambda): (\sqrt{2} \sin j\pi\lambda)/j\pi$, $\eta \quad (\lambda, x)$ and $\eta' \quad (\lambda', x')$.

The next result extends the convergence of the finite dimensional projections to weak convergence on the Hilbert space $L_2(\Pi, v)$.

Theorem 1. *Under Assumptions A1, A2 and (1), the process $S_n(\eta)$ converges weakly to $S(\eta)$ on the Hilbert space $L_2(\Pi, v)$, where $S(\eta)$ is a Gaussian process with zero mean and covariance operator satisfying $\sigma_h^2 \quad \langle C_S(h), h \rangle$, where σ_h^2 is given in (16).*

The next corollary is a simple consequence of the Continuous Mapping Theorem, see e.g. Billingsley's (1968) Theorem 5.1, and our Theorem 1.

Corollary 1. *Under the Assumptions of Theorem 1,*

$$D_n^2 \xrightarrow{\mathrm{d}} D_\infty^2: \int_{\mathbb{R}} \int_0^1 |S(\lambda, x)|^2 W(\mathrm{d}x) \, d\lambda.$$

To study the properties of D_∞^2, let l_i and ψ_i be the solutions of the eigenvalue problem $C_S(\psi_i) \quad l_i \psi_i$. Because $C_S(\cdot)$ is a nuclear operator, we have that $\{l_i, \psi_i\}_i^{\infty}{}_1$ is a complete sequence of eigenelements of it, i.e. $\{l_i\}_i^{\infty}{}_1$ are real valued and positive, and the corresponding eigenfunctions $\{\psi_i\}_i^{\infty}{}_1$ form a complete orthonormal basis for $L_2(\Pi, v)$. Hence any $L_2(\Pi, v)$ valued random element has a Fourier expansion in terms of $\{\psi_i\}_i^{\infty}{}_1$. In particular, we have the expansions (in distribution)

$$S_n(\eta) \quad \sum_{i\ 1}^{\infty} \sqrt{l_i} \varepsilon_{ni} \psi_i(\eta),$$

10

$$S(\eta) \quad \sum_{i\ 1}^{\infty} \sqrt{l_i} \varepsilon_i \psi_i(\eta),$$

where

$$\varepsilon_i \quad l_i^{-1/2} \int_{\Pi} S(\eta) \psi_i(\eta) \, dv(\eta),$$

and

$$\varepsilon_{ni} \quad l_i^{-1/2} \int_{\Pi} S_n(\eta) \psi_i(\eta) \, dv(\eta).$$

Note that by Theorem 1 $\{\varepsilon_i\}_{i\ 1}^{\infty}$ are i.i.d. N(0, 1) random variables and $\{\varepsilon_{ni}\}_{i\ 1}^{\infty}$ are at least uncorrelated with unit variance. Then, by Parseval's identity

$$D_{\infty}^2 \quad \int_{\mathbb{R}} \int_0^1 |S(\eta)|^2 \, dv(\eta) \quad \sum_{i\ 1}^{\infty} (\varepsilon_i \sqrt{l_i})^2. \tag{17}$$

Therefore, the asymptotic distribution of D_n^2 can be expressed as a weighted sum of independent χ_1^2 random variables with weights (l_i) depending on the DGP. The principal components $\{\varepsilon_i\}_{i\ 1}^{\infty}$ are involved in the directions of maximum power of D_{∞}^2, see Neuhaus (1976).

3.2. Consistency

The consistency properties of the generalized spectral martingale test based on rejecting H_0 for large values of D_n^2 are stated in the following theorems.

Theorem 2. *Under Assumptions A1 and A2,*

$$\frac{1}{n} D_n^2 \quad \overset{P}{\to} \quad \sum_{j\ 1}^{\infty} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\gamma_j(x)|^2 \, W(dx).$$

Under the alternative H_A there exists at least one $j \geqslant 1$ such that $\gamma_j(x) \neq 0$ for some subset of $\mathbb{R}$ with positive Lebesgue measure, cf. Theorem 1 in Bierens (1982). Then, since $W(\cdot)$ is absolutely continuous with respect to Lebesgue measure, the test will be consistent because

$$\sum_{j\ 1}^{n-1} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\gamma_j(x)|^2 \, W(dx) > 0.$$

That is, the test is consistent against all pairwise alternatives of the null (3), a property not attainable by the tests of Durlauf (1991), Deo (2000), Domínguez and Lobato (2003) and Kuan and Lee (2003). Note however, that there could exist non MDS such that $\gamma_j(x) \quad 0, \forall j \geqslant 1$, and our test will be not able to detect such alternatives that are not included by definition in H_A. An example of such alternatives is $Y_t \quad \varepsilon_{t-1}\varepsilon_{t-3} + \varepsilon_t$, where $\varepsilon_t \sim$ i.i.d N(0, 1). It is interesting to formally characterize the class of non MDS that do not satisfy H_A, but this is beyond the scope of this paper.

11

To gain some insights in the consistency properties of the test, next theorem shows the behaviour of $S_n(\eta)$ under a sequence of alternative hypotheses tending to the null at the parametric rate $n^{-1/2}$. Consider the non parametric local alternatives,

$$H_{A,n} : E[Y_t \quad \mu|\mathscr{F}_{t-1}] \quad \frac{g_t}{\sqrt{n}}, \text{ a.s.,} \tag{18}$$

where the sequence $\{g_t\}$ satisfies the following assumption.

Assumption A3. $\{g_t\}$ *is measurable with respect to* I_{t-1} ($\mathscr{F}_{t-1}$ *measurable*), *zero mean, strictly stationary, ergodic and square integrable sequence such that there exists a* $j \geqslant 1$ *with* $E[g_t|Y_{t-j}] \neq 0$ *with positive probability.*

Theorem 3. *Under the sequence of alternative hypothesis* (18) *satisfying A3 and Assumptions A1 and A2*

$$S_n(\eta) \Rightarrow S(\eta) + G(\eta),$$

where $S(\eta)$ *is the process defined in Theorem 1 and* $G(\eta)$ *is the deterministic function*

$$G(\lambda, x) \quad \sum_{j \ 1}^{\infty} \mu_j(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi},$$

where $\mu_j(x)$: $E[g_t \exp(ix Y_{t-j})]$.

Under the local alternatives $H_{A,n}$ and using A3, there exists at least one $j \geqslant 1$ such that $\mu_j(x) \neq 0$ in a set of positive measure, and therefore

$$\sum_{j \ 1}^{\infty} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\mu_j(x)|^2 W(dx) > 0.$$

Then, D_n^2 can detect all pairwise alternatives (18) satisfying A3 that converge to the null (3) at the parametric rate $n^{-1/2}$, a property which does not hold for those MDH tests based on smoothing approaches, e.g. Hong and Lee (2003, 2005).

Again, the next corollary is an immediate consequence of the Continuous Mapping Theorem and Theorem 3.

Corollary 2. *Under the local alternatives* (18) *and Assumptions A1, A2 and A3,*

$$D_n^2 \xrightarrow{d} \int_{\Pi} |S(\eta) + G(\eta)|^2 \, dv(\eta).$$

It is easy to show that

$$\int_{\Pi} |S(\eta) + G(\eta)|^2 \, dv(\eta) \quad \sum_{i \ 1}^{\infty} (\tau_i + \varepsilon_i \sqrt{l_i})^2,$$

where ε_i, l_i and ψ_i are as before, and $\{\tau_i\}_{i \ 1}^{\infty}$ are the Fourier coefficients of $G(\eta)$, that is

$$\tau_i \quad l_i^{-1/2} \int_{\mathbb{R}} \int_0^1 G(\eta)\psi_i(\eta) \, dv(\eta).$$

Now, we consider the "large" local alternatives

$$\mathrm{H}_{A,n}^{\mathrm{L}}(c): \mathrm{E}[Y_t \quad \mu|\mathscr{F}_{t-1}] \quad \frac{c\sigma g_t}{\sqrt{n}}, \text{ a.s.}, \tag{19}$$

with g_t verifying A3 and $\mathrm{E}[g_t^2] \quad 1$. Define the true error u_t: $Y_t \quad \mathrm{E}[Y_t|\mathscr{F}_{t-1}]$ and assume that $\mathrm{E}[u_t^2|\mathscr{F}_{t-1}] \quad \sigma^2$ a.s. Then, from Theorem 4 of Bierens and Ploberger (1997) we obtain the following result related with the rate of the asymptotic power function of the generalized spectral martingale test,

$$\Pi_{\mathrm{GSM}}(c): \quad \lim_{n\longrightarrow\infty} \mathrm{Pr}(D_n^2 \text{ rejects } \mathrm{H}_0|\mathrm{H}_{A,n}^{\mathrm{L}}(c)).$$

Corollary 3. *Under the sequence of alternative hypothesis* (19) *and Assumptions A1, A2 and A3, we have for any positive constant K,*

$$\lim_{c\longrightarrow\infty} c^{-2} \ln \mathrm{Pr}(D_{\infty}^2(c) \leqslant K) \quad \frac{1}{2}.$$

If the test has non trivial local power, then $\Pi_{\mathrm{GSM}}(c)$ approaches 1 at an exponential rate as $c \to \infty$. This result is even stronger if it is compared with the t statistic for $\delta \quad 0$ in the following regression:

$$Y_t \quad \mu + \delta g_t^* + v_t,$$

where g_t^* is some "guess" of g_t which is assumed to satisfy A3 and $\mathrm{E}[g_t^{*2}] \quad 1$. If $\rho \quad Corr(g_t, g_t^*)$ and $\Pi_t(c): \lim_{n\to\infty} \mathrm{Pr}(t \text{ test rejects } \mathrm{H}_0|\mathrm{H}_{A,n}^{\mathrm{L}}(c))$, then it is proved in Theorem 5 of Bierens and Ploberger (1997) that

$$\lim_{c\longrightarrow\infty} c^{-2} \ln(1 \quad \Pi_t(c)) \quad \frac{\rho^2}{2}.$$

This result implies that if the correlation coefficient ρ involved is not equal to 1 or 1, then there exists a c_0 such that $\Pi_{\mathrm{GSM}}(c) > \Pi_t(c)$ for $c > c_0$, that is, as long as the correlation between g_t^* and g_t is not perfect, our test is more powerful than the t test uniformly for large c's.

We end this section by proving an optimality power property of our test that can be derived under the following assumption.

Assumption A4. *The true error is conditionally Gaussian, i.e.*, $u_t|\mathscr{F}_{t-1} \sim \mathrm{N}(0, \sigma^2)$.

Note that Assumption A4 does not imply that Y_t is Gaussian. For alternatives for which Y_t is non Gaussian, it is expected that our generalized spectral martingale tests will be more powerful than correlation based tests as confirmed in our simulations. The next result, which follows from Theorem 6 of Bierens and Ploberger (1997), ensures that there does not exist a test uniformly more powerful than our test against all local alternatives satisfying A3.

Corollary 4. *Under Assumptions* A1 A4 *the generalized spectral martingale test is asymptotically admissible, that is, there does not exist a test that is uniformly more powerful than the generalized spectral martingale test against all local alternatives* (18) *satisfying* A3.

13

4. The bootstrap test

The asymptotic distribution of the generalized spectral martingale test statistic can be expressed as that of a weighted sum of independent χ_1^2 random variables and, although the weights depend on the DGP, they can be approximated by various methods. A simple multiple χ^2 approximation is due to Satterthwaite (1941, 1946). Under independence this approximation might be accurate, see Kock and Yang (1986) for an application. However, in the present context and under the null of MDS, the asymptotic distribution of $S_n(\eta)$ depends, in general, on the DGP in a very complicated way, and the same problem arises with the asymptotic distribution of the corresponding norm D_n^2.

One possibility to solve this problem is to estimate the distribution of $S_n(\eta)$ by that of

$$S_n^*(\lambda, x) \quad \sum_{j\,1}^{n-1}(n \quad j)^{1/2}\widehat{\gamma}_j^*(x)\frac{\sqrt{2}\sin j\pi\lambda}{j\pi},$$

with

$$\widehat{\gamma}_j^*(x) \quad \frac{1}{n \quad j}\sum_{t\,1+j}^{n}(Y_t \quad \overline{Y}_{n-j})\widehat{\phi}_{t-j}(x)w_t,$$

where $\widehat{\phi}_{t-j}(x) \quad e^{ixY_{t\,j}} \quad (n \quad j)^{-1}\sum_{t\,1+j}^{n}e^{ixY_{t\,j}}$ and $\{w_t\}$ is a sequence of independent random variables with zero mean, unit variance, bounded support and also independent of the sequence $\{Y_t\}_{t\,1}^{n}$. This procedure is similar to the wild bootstrap used in Wu (1986), Liu (1988) or Mammen (1993). Also, our bootstrap approximation is related to the fixed design wild bootstrap for a constant model ($E[Y_t|I_{t-1}] \quad \mu$), see e.g. Goncalves and Kilian (2004). Examples of $\{w_t\}$ sequences are i.i.d. Bernoulli variates with

$$P(w_t \quad 0.5(1 \quad \sqrt{5})) \quad (1+\sqrt{5})/2\sqrt{5}, \tag{20}$$

and

$$P(w_t \quad 0.5(1+\sqrt{5})) \quad 1 \quad (1+\sqrt{5})/2\sqrt{5}, \tag{21}$$

used in e.g. Mammen (1993), Stute et al. (1998) or more recently in Li et al. (2003), or $P(w_t \quad 1) \quad 0.5$ and $P(w_t \quad 1) \quad 0.5$ as in Liu (1988) or de Jong (1996), for other sequences see Mammen (1993). Note that the third moment of w_t in those cases is equal to 1, and hence the first three moments of the bootstrap series coincide with the three moments of the original series. These properties have implications on the second order asymptotic properties of the bootstrap approximation, see Liu (1988). Next theorem shows the validity of the bootstrap and allows us to approximate the critical values of the test.

Theorem 4. *Assume A1 and A2, then under the null hypothesis* (1), *under any fixed alternative hypothesis or under the local alternatives* (18)

$$S_n^*(\eta) \underset{*}{\Rightarrow} S(\eta), a.s. \text{ in } L_2(\Pi, v),$$

where $S(\eta)$ is the process defined in Theorem 1 and $\Rightarrow$ a.s. denote weak convergence almost surely under the bootstrap law, that is, if the sample is χ_n,

$$\rho_w(\mathscr{L}(S_n^*(\eta)|\chi_n), \mathscr{L}(S(\eta))) \to 0 \text{ a.s. as } n \to \infty$$

where $\mathscr{L}(S_n^*(\eta)|\chi_n)$ stands for the law of $S_n^*(\eta)$ given the sample, and ρ_w is any metric which metricizes weak convergence in $L_2(\Pi, v)$, see Politis and Romano (1994).

Therefore, we can approximate the asymptotic distribution of the process $S_n(\eta)$ by that of $S_n^*(\eta)$. In particular, we can simulate the critical values for the test statistic D_n^2 by the following algorithm:

1. Calculate the test statistic D_n^2 with the original sample $\{Y_t\}_{t=1}^n$.
2. Generate $\{w_t\}_{t=1}^n$, a sequence of independent random variables with zero mean, unit variance and bounded support and independent of the sample $\{Y_t\}_{t=1}^n$.
3. Compute $\widehat{\gamma}_j^*(x)$ for $j = 1,\ldots,n-1$, with $\{w_t\}_{t=1}^n$. Then, compute $S_n^*(\eta)$ and D_n^{*2}.
4. Repeat steps 2 and 3 B times and compute the empirical $(1-\alpha)$th sample quantile of D_n^{*2} with the B values, $D_{n,\alpha}^{*2}$ say. The proposed test rejects the null hypothesis at the significance level α if $D_n^2 > D_{n,\alpha}^{*2}$.

Note that given the result obtained in Theorem 4, the proposed bootstrap test has a correct asymptotic level, is consistent and is able to detect alternatives tending to the null at the parametric rate $n^{-1/2}$.

5. Empirical evidence

In order to examine the finite sample performance of the proposed test we carry out a simulation experiment with some DGP under the null and under the alternative. We compare our test with those of Deo (2000), Domínguez and Lobato (2003), Hong and Lee (2003, 2005) and Kuan and Lee (2003). We briefly describe them to clarify our simulation setup.

To compute D_n^2, we choose the standard normal cdf. as the weighting function $W(\cdot)$, then D_n^2 simplifies to (15). In the simulations we standardize D_n^2 by

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \overline{Y}_n)^2.$$

Hong and Lee (2003, 2005) have proposed recently a diagnostic test for the adequacy of a parametric conditional mean under possible conditional heteroskedasticity. This test can be adapted as a MDH test by taking as the residuals $\{\widehat{\varepsilon}_t \equiv Y_t - \overline{Y}_n\}_{t=1}^n$. They consider the test statistic

$$\mathrm{HL}_n(p) = [L_2^2(p) - \widehat{C}_1(p)]/[\sqrt{\widehat{D}_1(p)}], \tag{22}$$

where now $L_2^2(p)$ is given by

$$L_2^2(p) = \sum_{j=1}^{n-1} (n-j) k^2 \left(\frac{j}{p}\right) \int_{\mathbb{R}} |\widehat{\sigma}_j(x)|^2 W(\mathrm{d}x).$$

with

$$\widehat{\sigma}_j(x) = \frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \overline{Y}_n) \widehat{\psi}_{t-j}(x)$$

and $\widehat{\psi}_t(x) = \exp(\mathrm{i}x(Y_t - \overline{Y}_n)) - \widehat{\varphi}(x)$, and $\widehat{\varphi}(x) = n^{-1} \sum_{t=1}^{n} \exp(\mathrm{i}x(Y_t - \overline{Y}_n))$. The centering and scaling factors are, respectively,

$$\widehat{C}_1(p) = \sum_{j=1}^{n-1} (n-j)^{-1} k^2 \left(\frac{j}{p}\right) \sum_{t=j+1}^{n} (Y_t - \overline{Y}_n)^2 \int_{\mathbb{R}} |\widehat{\psi}_{t-j}(x)|^2 W(\mathrm{d}x)$$

and

$$\widehat{D}_1(p) = 2 \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} k^2 \left(\frac{j}{p}\right) k^2 \left(\frac{l}{p}\right)$$

$$\times \int_{\mathbb{R}^2} \left| \frac{1}{n - \max(j,l)} \sum_{t=\max(j,l)+1}^{n} (Y_t - \overline{Y}_n)^2 \widehat{\psi}_{t-j}(x) \widehat{\psi}_{t-l}(y) \right|^2 W(\mathrm{d}x) W(\mathrm{d}y).$$

Under some assumptions and the null of MDH they showed that $\mathrm{HL}_n(p)$ converges to a standard normal random variable. For the simulations we use again the cdf. of standard normal random variable as the integrating function W and the Daniell kernel $k(z) = \sin(\pi z)/(\pi z)$ as in the simulations of Hong and Lee (2003, 2005). For the values of the bandwidth we consider $p = 2, 6$ and 10. For the empirical level of the $\mathrm{HL}_n(p)$ test we consider in the simulations the asymptotic critical values and for the power we consider the empirical critical values obtained under 5000 simulations of model 1 (IID) below.

Domínguez and Lobato (2003) have proposed a generalization of Koul and Stute (1999) for a fixed number of lags P, $P \geqslant 1$. We denote by CvM_P and KS_P their Cramér von Mises and Kolmogorov Smirnov statistics, respectively. Their tests statistics are

$$\mathrm{CvM}_P: \quad \frac{1}{\widehat{\sigma}^2 n^2} \sum_{j=1}^{n} \left[\sum_{t=1}^{n} (Y_t - \overline{Y}_n) I(\widetilde{Y}_{t,P} \leqslant \widetilde{Y}_{j,P}) \right]^2,$$

$$\mathrm{KS}_P: \quad \max_{1 \leqslant i \leqslant n} \left| \frac{1}{\widehat{\sigma} \sqrt{n}} \sum_{t=1}^{n} (Y_t - \overline{Y}_n) I(\widetilde{Y}_{t,P} \leqslant \widetilde{Y}_{i,P}) \right|,$$

where $\widetilde{Y}_{t,P} = (Y_{t-1}, \ldots, Y_{t-P})$ are the P lagged values of the series. They also considered a wild bootstrap approach. In the simulations we consider the values $P = 1, 2$ and 3, although only $P = 1$ and 3 are reported.

[Deo (2000)](#) considers a corrected version of [Durlauf's (1991)](#) test to take into account some kinds of conditional heteroskedasticity. The corrected version of the test statistic is

$$\text{DURC:} \quad \sum_{j=1}^{n-1} n\widehat{a}_j^2 \left(\frac{1}{j\pi}\right)^2,$$

where

$$\widehat{a}_j : \ \widehat{\rho}_j \left[(n-j)^{-1} \sum_{t=1+j}^{n} (Y_t - \overline{Y}_n)^2 (Y_{t-j} - \overline{Y}_n)^2 \right]^{-1/2},$$

and

$$\widehat{\rho}_j : \ (n-j)^{-1} \sum_{t=1+j}^{n} (Y_t - \overline{Y}_n)(Y_{t-j} - \overline{Y}_n).$$

We have considered the factor $(n-j)^{-1}$ in $\widehat{\rho}_j$ instead of n^{-1} used in [Deo (2000)](#) as in the definition of D_n^2. Note that we do not consider a kernel or weighting function in DURC. Under the null hypothesis of MDS and some additional assumptions (see [Deo, 2000](#)),

$$\text{DURC} \xrightarrow{d} \int_0^1 B^2(t)\,dt \text{ as } n \to \infty,$$

where $B(t)$ is the standard Brownian bridge on [0,1]. The 10%, 5% and 1% asymptotic critical values are obtained from [Shorack and Wellner (1986, p. 147)](#) and are 0.347, 0.461 and 0.743, respectively. We have also used in the simulations bootstrap critical values. Since the results do not differ substantially we only report the results based on the asymptotic critical values.

Recently, [Kuan and Lee (2003)](#) have proposed a MDH test related to the characteristic function which is robust to conditional heteroskedasticity and with asymptotic χ_2^2 distribution. Their test statistic is

$$\text{KL}_P : \ \frac{n-P}{\widehat{\sigma}_{c,g}^2 \widehat{\sigma}_{s,g}^2 - \widehat{\sigma}_{cs,g}^2} [\widehat{\sigma}_{s,g}^2 \overline{\Psi}_{c,g}^2 + \widehat{\sigma}_{c,g}^2 \overline{\Psi}_{s,g}^2 - 2\widehat{\sigma}_{cs,g}^2 \overline{\Psi}_{c,g}^2 \overline{\Psi}_{s,g}^2],$$

where for $j = c, s$

$$\overline{\Psi}_{j,g} : \ \frac{1}{n-P} \sum_{t=P+1}^{n} \Psi_{j,g}(Y_t, \widetilde{Y}_{t,P}),$$

$$\widehat{\sigma}_{j,g}^2 : \ \frac{1}{n-P} \sum_{t=P+1}^{n} \Psi_{j,g}^2(Y_t, \widetilde{Y}_{t,P}),$$

$$\widehat{\sigma}_{cs,g}^2 : \ \frac{1}{n-P} \sum_{t=P+1}^{n} \Psi_{c,g}(Y_t, \widetilde{Y}_{t,P})\Psi_{s,g}(Y_t, \widetilde{Y}_{t,P}),$$

and

$$\Psi_{c,g}(Y_t, \widetilde{Y}_{t,P}): \quad Y_t \int_{\mathbb{R}^{P+}} \cos(x^T \widetilde{Y}_{t,P}) g(x)\, \mathrm{d}x,$$

$$\Psi_{s,g}(Y_t, \widetilde{Y}_{t,P}): \quad Y_t \int_{\mathbb{R}^{P+}} \sin(x^T \widetilde{Y}_{t,P}) g(x)\, \mathrm{d}x.$$

Kuan and Lee (2003) assume that μ 0 and a fixed number of lags in the conditioning set, P say. Instead of consider a norm as in D_n^2, they first integrate with respect to x with the weight function $g(x)$, which plays the same role as our $W(\mathrm{d}x)$, and consider a norm on the sample means $\overline{\Psi}_{j,g}$. They supply formulae for $\Psi_{j,g}(Y_t, \widetilde{Y}_{t,P})$ using a multivariate exponential density function $g(x)$, see Kuan and Lee (2003, p. 9). We choose in the simulations as $g(\cdot)$ the density of the multivariate exponential distribution with parameter $\widetilde{\beta}$ $(n^{-1}\sum_t^n {}_1 Y_t^2)^{-1/2}$, see Kuan and Lee (2003) for details. We consider P 1 and 3.

In the sequel ε_t and u_t are independent sequences of i.i.d. N(0, 1). The models used in the simulations include first three MDS:

1. Independent and identically distributed N(0, 1) variates (IID).
2. GARCH(1,1) processes, Y_t $\varepsilon_t \sigma_t$, with σ_t^2 $0.001 + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$, and the following combinations for (α, β): (0.01, 0.97), (0.09, 0.89) and (0.09, 0.90). We denote these processes by GARCH1, GARCH2 and GARCH3, respectively.
3. Stochastic volatility model (SV), Y_t $\varepsilon_t \exp(\sigma_t)$, with σ_t $0.936\sigma_{t-1} + 0.32 u_t$.

And the following non martingale difference sequences:

4. Non linear moving average model (NLMA): Y_t $\varepsilon_{t-1}\varepsilon_{t-2}(\varepsilon_{t-2} + \varepsilon_t + 1)$.
5. Bilinear Processes: Y_t $\varepsilon_t + b_1\varepsilon_{t-1}Y_{t-1} + b_2\varepsilon_{t-1}Y_{t-2}$, with parameter values (b_1, b_2): (0.15, 0.05) and (0.25, 0.15). We call these processes BIL I and BIL II, respectively.
6. Fractional integrated model ARFIMA(0, d, 0), that is, $(1 \quad L)^d Y_t$ ε_t, with d 0.3 and where L is the usual back shift operator, i.e. LY_t Y_{t-1}.
7. The sum of a white noise and the first difference of a stationary autoregressive process of order one (NDAR):

$$Y_t \quad \varepsilon_t + X_t \quad X_{t-1},$$

$$X_t \quad 0.85 X_{t-1} + u_t.$$

8. Threshold autoregressive model of order one (TAR(1)):

$$Y_t \quad 0.5 Y_{t-1} + \varepsilon_t \text{ if } Y_{t-1} \geqslant 1 \quad \text{and} \quad Y_t \quad 0.4 Y_{t-1} + \varepsilon_t \text{ if } Y_{t-1} < 1.$$

9. First order exponential autoregressive model (EXP(1)): Y_t $0.6 Y_{t-1}$ $\exp(\quad 0.5 Y_{t-1}^2) + \varepsilon_t$.

Models 1, 2, 4 and 5 are used in Domínguez and Lobato (2003). Models 3, 6 and 7 are used in Deo (2000), while models 8 and 9 are used in Hong (1999b). Models 1, 2, 5 and 8 are also considered in Kuan and Lee (2003). As was noted by Domínguez and Lobato, the second and third GARCH models have unbounded eighth and sixth moment, respectively. Note that Deo (2000) assumed a bounded eighth moment. However, we conjecture that this assumption can be simplified by using a Hilbert space approach similar to that considered in our test statistics.

We consider for the experiments under the null a sample size of $n = 100$ and under the alternative $n = 100$ and 300, the number of Monte Carlo experiments is 1000 and the number of bootstrap replications is $B = 300$. We consider a nominal level of 5%, the results with other nominal levels are similar. In all the replications 200 pre sample data values were generated and discarded. Random numbers were generated using IMSL ggnml subroutine. For the bootstrap, we employ a sequence $\{w_t\}$ of $i.i.d.$ Bernoulli variates given in (20) (21). The simulation program was written in Fortran 90 and run on a PC using a P 4 processor at 1,7 GHz. The computational time for computing D_n^2 and the bootstrap p value with $B = 300$ at 10%, 5% and 1% levels, for the NLMA model is 9″ for $n = 100$ and 4′10″ for $n = 300$. Computational time with other models were very similar.

In Table 1 we report the empirical rejections probabilities (RP) associated with the models 1 3. The results for D_n^2, $HL_n(p)$, CvM_P, KS_P, KL_P and DURC show that the empirical size properties of the tests are appropriate, although KL_3 shows some size distortions for the SV model. From these results with the GARCH models we conclude that D_n^2 is robust to thick tails.

In Tables 2 and 3 we report the empirical power against the models 4 9 for the sample size $n = 100$ and 300, respectively. The results show that for almost all nonlinear alternatives D_n^2 has more empirical power than the other tests and maintains good properties for linear models (ARFIMA). Hong and Lee (2003, 2005) test $HL_n(p)$ achieves its maximum empirical power at $p = 2$ and is highly affected by the choice of the bandwidth parameter p. The results for $HL_n(2)$ are comparable to

Table 1
Size of tests at 5%

$n = 100, B = 300$	IID	GARCH1	GARCH2	GARCH3	SV
D_n^2	4.8	5.2	5.4	5.1	5.9
$HL_n(2)$	7.1	7.0	6.6	6.7	7.1
$HL_n(6)$	5.0	5.4	5.2	5.6	5.0
$HL_n(10)$	3.8	5.2	5.2	5.7	3.8
CvM_1	4.6	4.6	5.5	5.6	5.3
KS_1	5.3	5.1	4.8	6.3	4.6
CvM_3	5.8	5.8	5.2	5.5	4.2
KS_3	6.6	6.2	5.1	6.1	6.1
KL_1	5.4	5.7	4.9	3.6	6.7
KL_3	6.5	4.9	2.9	4.9	9.3
DURC	5.4	5.4	5.9	5.8	4.3

Table 2
Power of tests at 5%

$n = 100$	NLMA	BIL-I	BIL-II	ARFIMA	NDAR	TAR(1)	EXP(1)
D_n^2	19.0	25.4	59.5	80.0	6.5	72.4	66.6
$HL_n(2)$	10.2	24.9	51.5	75.1	5.6	71.8	64.3
$HL_n(6)$	6.7	16.5	43.1	74.3	4.9	59.6	53.1
$HL_n(10)$	4.8	14.4	34.7	70.1	4.5	50.2	44.3
CvM_1	18.3	11.5	25.8	84.3	7.7	50.7	65.6
KS_1	18.4	12.6	30.1	79.4	8.5	47.6	65.8
CvM_3	6.4	5.4	11.0	73.5	6.5	15.0	26.5
KS_3	7.4	8.7	14.1	69.5	7.6	20.8	30.7
KL_1	12.5	5.9	13.0	66.5	2.7	42.6	49.4
KL_3	10.2	5.8	7.3	35.3	7.4	16.0	14.0
DURC	5.3	6.3	11.3	87.6	7.2	19.5	49.5

Table 3
Power of tests at 5%

$n = 300$	NLMA	BIL-I	BIL-II	ARFIMA	NDAR	TAR(1)	EXP(1)
D_n^2	41.9	66.6	98.5	100.0	9.2	99.9	98.7
$HL_n(2)$	33.3	58.4	96.3	96.5	6.0	99.6	99.0
$HL_n(6)$	21.3	49.3	95.2	95.3	4.7	99.0	98.0
$HL_n(10)$	18.3	45.0	91.3	91.5	7.0	98.7	96.6
CvM_1	35.5	30.5	80.5	100.0	9.6	97.9	98.7
KS_1	38.3	30.8	80.5	100.0	9.7	95.3	98.9
CvM_3	14.2	9.4	27.2	99.8	4.6	47.6	72.6
KS_3	24.2	11.0	45.4	99.8	4.6	63.6	84.8
KL_1	21.8	8.4	41.6	99.0	2.6	92.2	98.0
KL_3	13.4	8.2	12.6	91.8	7.4	46.6	40.6
DURC	6.4	7.2	21.2	100.0	9.0	51.5	93.8

those of D_n^2 except for the NLMA alternative. In the NLMA case DURC has no power, as expected, because this NLMA process is uncorrelated. For the bilinear cases, D_n^2 and $HL_n(2)$ have excellent empirical power whereas KL_P and DURC have low power. For the Deo's (2000) model NDAR, all the test statistics have low empirical power. For the TAR(1) and EXP(1) models D_n^2 has the highest empirical power, followed by $HL_n(2)$. Note that CvM_P and KS_P are largely affected by the curse of dimensionality for moderate values of P. In all cases D_n^2 outperforms KL_P for all alternatives considered and all values of P. This confirms that for all these alternatives it can be much better to consider a pairwise than a joint approach as taken in Domínguez and Lobato (2003) or Kuan and Lee (2003). In Table 3, we show the empirical power with the sample size n 300. As was expected, all the empirical powers increase and D_n^2 maintains its overall superiority. For complete ness, we show in Fig. 1 the plots of simulated $|\hat{\gamma}_j(x)|^2$ for j 1 and for the models
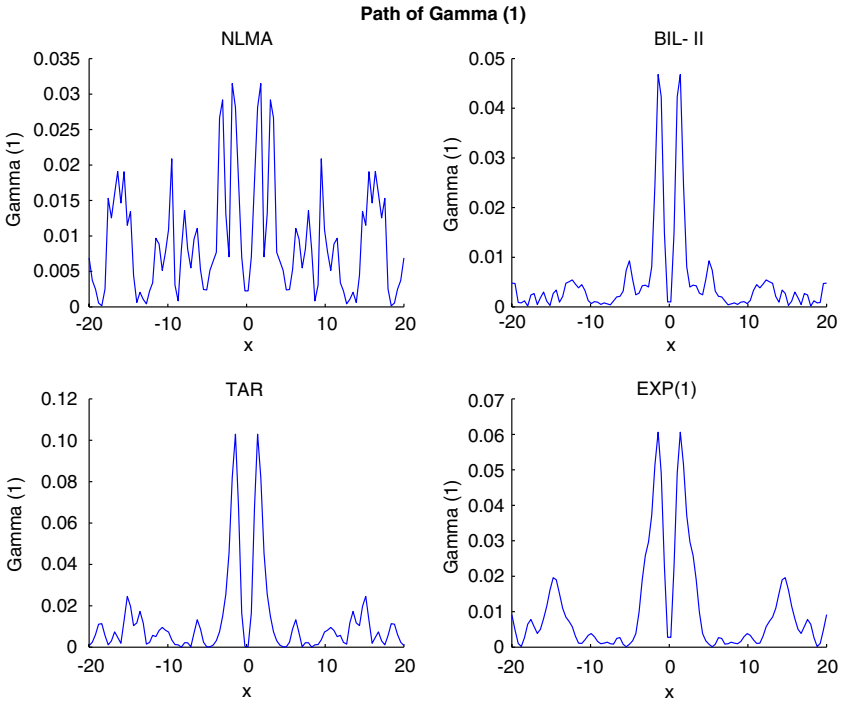
Path of Gamma (1)

Fig. 1. Path of simulated $|\hat{\gamma}_1(x)|^2$ for models NLMA, BIL-II, TAR(1) and EXP(1), respectively. The number of observations is $n = 300$.

NLMA, BIL II, TAR(1) and EXP(1), respectively. The number of observations used is n 300. These graphs can help to indicate how the W function "finds" the dependence structure, putting high weigh in intervals close to the origin.

Now, we present applications of our generalized spectral martingale test to the daily closed S&P 500 stock index and some exchange rates returns. We consider three sample periods for the S&P 500 stock index, first from 2 January 1990 until 31 December 1993, the second period from 3 January 1994 until 31 December 1997 and third from 2 January 1998 until 28 August 2002. With a total of 1013, 1011 and 1170 observations, respectively. We consider the returns of the index obtained as the log differences of the data. The implementation is as in the Monte Carlo simulations. To facilitate interpretations we show the p values for the three periods in Table 4.

We observe that for the first period all the tests fail to reject the MDH, whereas in the second period, although DURC finds no correlation in the data and CvM_3, KS_3, KL_1 and KL_3 fail to reject the MDH, our generalized spectral test statistic D_n^2 and $HL_n(p)$ strongly reject the null hypothesis of MDS. Again, this shows the ability of

21

Table 4
P-values for the S&P 500 stock index

	1990 1993	1994 1997	1998 2002
D_n^2	0.280	0.000	0.080
$HL_n(2)$	0.097	0.000	0.327
$HL_n(6)$	0.301	0.000	0.246
$HL_n(10)$	0.347	0.000	0.415
CvM_1	0.300	0.004	0.156
KS_1	0.390	0.006	0.194
CvM_3	0.882	0.758	0.076
KS_3	0.634	0.342	0.030
KL_1	0.670	0.108	0.331
KL_3	0.737	0.392	0.252
DURC	0.130	0.730	0.650

our test to detect possibly uncorrelated non martingale difference sequences. Note that for the second period, CvM_P, KS_P deliver contradictory results for different values of P, possibly due to the curse of dimensionality. For the third period, D_n^2 also questions the MDH whereas $HL_n(p)$, KL_1, KL_3, and DURC strongly support the MDH.

We next examine the martingale properties of some exchange rates returns studied previously by Fong and Ouliaris (1995) or Kuan and Lee (2003) among others. Also recently, Hong and Lee (2003, 2005) have studied the MDH properties of a related data set. Hong and Lee (2003, 2005) consider a wild bootstrap approximation for the Hong's (1999) MDH test. Although, they do not prove the theoretical validity of this bootstrap approximation, our Theorem 4 can serve as a theoretical justification of this resampling procedure. The data set consists in five 760 weekly exchange rate returns on the Canadian Dollar (Can), the German Deutschmark (Dm), the French Franc (Fr), the sterling Pound (£) and the Japanese Yen (¥), from August 14, 1974 to March 29, 1989. The empirical results are reported in Table 5. Again, we use the same implementation as in the Monte Carlo experiments and we show the empirical p values. The results for the Canadian dollar are inconclusive with our test statistic D_n^2 and contradictory with the statistics CvM_P, KS_P and KL_P. $HL_n(p)$ clearly rejects the MDH for all values of p. Also, DURC finds no correlation in Can at 5% level. Our test D_n^2 rejects the MDH for all the remaining exchange rates returns at 5% level and also at the 1% in Dm and ¥. Note that Fr and £ seem to be serially uncorrelated, although they are not MDS. The test statistic $HL_n(p)$ also rejects the MDH for these exchange rates returns. Therefore, our new test statistic D_n^2 finds nonlinear dependence in the conditional mean of these exchange rates, contrasting with earlier studies which conclude that exchange rate changes are very nearly to be unpredictable given past prices and in agreement with the results of Kuan and Lee (2003) and Hong and Lee (2003, 2005). The nonlinearity in the conditional mean suggest that additional effort has to be dedicated to investigate the form of such nonlinearity in the conditional mean before modelling the conditional variance.

Table 5
P-values for the exchange rates returns

n = 760	Can	Dm	Fr	£	¥
D_n^2	0.050	0.000	0.010	0.032	0.000
$HL_n(2)$	0.001	0.000	0.000	0.008	0.000
$HL_n(6)$	0.000	0.000	0.000	0.003	0.000
$HL_n(10)$	0.000	0.000	0.000	0.016	0.000
CvM_1	0.024	0.000	0.020	0.044	0.000
KS_1	0.044	0.000	0.030	0.082	0.000
CvM_3	0.122	0.006	0.086	0.056	0.000
KS_3	0.378	0.002	0.080	0.252	0.004
KL_1	0.000	0.005	0.033	0.114	0.002
KL_3	0.1959	0.000	0.000	0.000	0.000
DURC	0.071	0.010	0.095	0.580	0.010

6. Conclusions and further extensions

In this article we have synthesized the approaches of Hong (1999a) and Durlauf (1991) introducing a new test for the MDH which is based on a functional of the difference between a generalized spectral distribution function under the null and under the alternative. We extend Durlauf's (1991) test which uses the second moment information, in the sense that we our test uses all the information contained in the pairwise regression functions. In fact, as was shown by Hong (1999a), Durlauf's (1991) test statistic can be obtained by differentiating our test statistic with respect to the auxiliary parameter at the origin. The auxiliary parameter x, renders a test statistic able to pick any pairwise alternative of the MDH. In particular, our test is able to detect failures of the MDS assumption for processes that are uncorrelated, as has been shown in the Monte Carlo simulations and the applications. The asymptotic null distribution depends on the DGP and hence, we have proposed to implement the test via a bootstrap procedure.

We have carried out an empirical comparison with the MDH tests of Deo (2000), Domínguez and Lobato (2003), Hong and Lee (2003, 2005) and Kuan and Lee (2003) and the results have shown that our test has for almost all nonlinear alternatives more empirical power. Note also that our test avoids the choice of any lag order or smoothing parameter and overcomes the problem of the curse of dimensionality which may affect other tests, for instance Domínguez and Lobato's (2003) test when P is high or even moderate. These facts confirm that in practice it could be better to consider a pairwise approach than a joint approach as in Domínguez and Lobato (2003) or Kuan and Lee (2003). One obvious limitation of our approach is that our test is not consistent against pairwise MDS which are non MDS. To solve this problem one can apply the same methodology but considering as the target the hypothesis

$$E[Y_t \quad \mu | Y_{t-j}, Y_{t-j-1}, \ldots, Y_{t-j-P+1}] \quad 0 \text{ a.s. for all } 1 \leqslant j \leqslant n \quad P$$

23

for P fixed. Let $\widetilde{x}_P = (x_1, x_2, \ldots, x_P)$ and $\widetilde{Y}_{tjP} = (Y_{t-j}, Y_{t-j-1}, \ldots, Y_{t-j-P+1})$. Hence, defining the measures

$$\gamma_j(\widetilde{x}_P) = \mathrm{E}[(Y_t - \mu)\exp(\mathrm{i}\widetilde{x}_P^T \widetilde{Y}_{tjP})],$$

we test

$$\mathrm{H}_0 : \gamma_j(\widetilde{x}_P) = 0 \qquad \forall j \geqslant 1, \widetilde{x}_P \in \mathbb{R}^P.$$

Let $n_P = n - P$ and $n_{jP} = n - j - P + 1$. Proceeding in exactly the same way as above, an extended test can be based on the statistic

$$\widehat{H}(\lambda, \widetilde{x}_P) = \widehat{\gamma}_0(\widetilde{x}_P)\lambda + 2\sum_{j=1}^{n_P} \left(\frac{n_{jP}}{n}\right)^{1/2} \widehat{\gamma}_j(\widetilde{x}_P)\frac{\sin j\pi\lambda}{j\pi},$$

where

$$\widehat{\gamma}_j(\widetilde{x}_P) = \frac{1}{n_{jP}}\sum_{t=P+j}^{n} (Y_t - \overline{Y}_{n_{jP}})\exp(\mathrm{i}\widetilde{x}_P^T \widetilde{Y}_{tjP}).$$

We consider the statistic $\widehat{H}(\lambda, \widetilde{x}_P)$ as a $L_2(\Pi, \nu)$ valued random element with $\Pi = [0, 1] \times \mathbb{R}^P$. The Cramér von Mises test in this case is

$$D_n^2 = \sum_{j=1}^{n_P} \frac{n_{jP}}{(j\pi)^2} \int_{\mathbb{R}^P} |\widehat{\gamma}_j(\widetilde{x}_P)|^2 \, \mathrm{d}W(\widetilde{x}_P),$$

which involves P dimensional integration. As in the univariate case, $W(\cdot)$ can be chosen such that D_n^2 has a closed expression, examples of this are the normal or exponential multivariate cdf. All the results shown in this paper hold trivially for this multivariate case. The lack of theoretical complication in the multivariate situation is one of the advantages of our Hilbert space approach over a ''sup'' norm approach as in Hong (1999b).

Finally, we make some comments on how to extent our approach to test the correct specification of a conditional mean of linear and nonlinear time series models. It is well known that the correct specification for the conditional mean is equivalent to a MDS property on the unobservable errors. Then, we could apply our methodology to the residuals to test this MDS property. The main problem is to take into account the parameter uncertainty. In particular, the bootstrap approximation as is proposed here is no longer valid and needs some modification, although other resampling or asymptotic approximations could be considered. These topics are currently being investigated.

Acknowledgements

Appendix. Proofs

Let $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ be the real and imaginary parts, respectively, of the complex number A. First, consider two useful lemmas.

Lemma 1. *Suppose we have a random element in $L_2(\Pi, v)$ of the form $h_n(\eta) = \sum_{j=1}^{n-1} h_{j,n}(x)\sqrt{2}\sin j\pi\lambda/j\pi$. Assume A2 and that*

(i) $\int_{\mathbb{R}} \mathrm{E}|h_{j,n}(x)|^2 W(\mathrm{d}x) < C$ *uniformly in $j \geqslant 1$.*
(ii) $\sup_{x\in[-a,a]} |h_{j,n}(x)| = o_p(1)$ $\forall 1 \leqslant j < n$, $\forall a > 0$.

Then, $h_n(\eta)$ converges in probability to zero in $L_2(\Pi, v)$, i.e. $\|h_n(\eta)\|^2 = o_p(1)$.

Proof. Decompose for each K fixed with $1 \leqslant K \leqslant n-1$

$$h_n(\eta) = \sum_{j=1}^{K} h_{j,n}(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi} + \sum_{j=K+1}^{n-1} h_{j,n}(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}$$
$$=: h_{1K}(\eta) + h_{2K}(\eta).$$

Now, for all $a > 0$,

$$\|h_{1K}(\eta)\|^2 = \sum_{j=1}^{K} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |h_{j,n}(x)|^2 \, \mathrm{d}W(x)$$
$$= \sum_{j=1}^{K} \frac{1}{(j\pi)^2} \int_{[-a,a]} |h_{j,n}(x)|^2 \, \mathrm{d}W(x)$$
$$+ \sum_{j=1}^{K} \frac{1}{(j\pi)^2} \int_{\mathbb{R}\setminus[-a,a]} |h_{j,n}(x)|^2 \, \mathrm{d}W(x). \tag{23}$$

Thus, by (ii) for the first summand on the right hand side of (23) and by (i) and letting $a \to +\infty$ for the second summand, we have that $\|h_{1K}(\eta)\|^2 = o_p(1)$. Next, use (i) to show that

$$\mathrm{E}\|h_{2K}(\eta)\|^2 = \sum_{j=K+1}^{n-1} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} \mathrm{E}|h_{j,n}(x)|^2 \, \mathrm{d}W(x)$$
$$\leqslant C \sum_{j=K+1}^{\infty} \frac{1}{(j\pi)^2} \to 0 \text{ as } K \to \infty.$$

Finally, Theorem 4.2 of Billingsley (1968) yields that $\|h_n(\eta)\|^2 = o_p(1)$ by letting first $n \to \infty$ and then $K \to \infty$. This finishes the proof of Lemma 1. $\square$

Let define

$$Z_n(\lambda, x) := \sum_{j=1}^{n-1} (n-j)^{1/2} \widehat{r}_j(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi},$$

where

$$\widehat{r}_j(x) = \frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu)\phi_{t-j}(x).$$

The next lemma states that the norm of the process $S_n(\lambda, x)$ can be approximated by the norm of $Z_n(\lambda, x)$.

Lemma 2. *Under* (1) *and the assumptions A1 and A2*,

$$\left| \|S_n(\eta)\|^2 - \|Z_n(\eta)\|^2 \right| \xrightarrow{P} 0.$$

Proof. It is easy to show that

$$\widehat{\gamma}_j(x) - \widehat{r}_j(x) = \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu) \right]\left[\frac{1}{n-j} \sum_{t=1+j}^{n} \phi_{t-j}(x) \right], \tag{24}$$

and therefore, we have that $S_n(\eta) = Z_n(\eta) - R_n(\eta)$ where

$$R_n(\eta) = \sum_{j=1}^{n-1} (n-j)^{1/2} \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu) \right]\left[\frac{1}{n-j} \sum_{t=1+j}^{n} \phi_{t-j}(x) \right] \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}. \tag{25}$$

By simple algebra

$$\int_{\mathbb{R}} \int_0^1 |S_n(\eta)|^2 \, dW \, d\lambda = \int_{\mathbb{R}} \int_0^1 |Z_n(\eta)|^2 \, dW \, d\lambda$$
$$+ \int_{\mathbb{R}} \int_0^1 |R_n(\eta)|^2 \, dW \, d\lambda$$
$$- 2\operatorname{Re}\left\{ \int_{\mathbb{R}} \int_0^1 Z_n(\eta) R_n^c(\eta) \, dW \, d\lambda \right\}. \tag{26}$$

Let define $\overline{\phi}_{n-j}(x) := (n-j)^{-1}\sum_{t=1+j}^{n}\phi_{t-j}(x)$, then

$$\|R_n(\eta)\|^2 = \sum_{j=1}^{n-1} \frac{(n-j)}{(j\pi)^2} \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu) \right]^2 \int_{\mathbb{R}} |\overline{\phi}_{n-j}(x)|^2 W(dx).$$

From A1 and (1) we obtain that

$$\frac{1}{(n-j)^{1/2}} \sum_{t=1+j}^{n} (Y_t - \mu) = O_p(1) \quad \forall 1 \leqslant j < n.$$

Whereas Theorem 2 in Jennrich (1969) and the Glivenko-Cantelli's Theorem for stationary and ergodic sequences, see e.g. Dehling and Philipp (2002, p. 4), yield the

following uniform ergodic theorem (UET):

$$\sup_{x\in[-a,a]} |\overline{\phi}_{n-j}(x)| \quad o_p(1) \quad \forall 1 \leqslant j < n, \ \forall a > 0.$$

Thus, applying Lemma 1 we have that $\|R_n(\eta)\|^2 \quad o_p(1)$. On the other hand, it is easy to show that under (1)

$$E \int_{\mathbb{R}} \int_0^1 |Z_n(\eta)|^2 \, dW \, d\lambda \quad \sum_{j\ 1}^{n-1} \frac{(n\ j)}{(j\pi)^2} \, E \int_{\mathbb{R}} |\widehat{r}_j(x)|^2 \, dW(x) \quad O(1). \qquad (27)$$

Combining (26), (27) and Cauchy Schwartz's inequality we conclude the proof. $\quad\square$

Proof of Theorem 1. By Lemma 2 we only need to show that the finite dimensional projections $\langle Z_n(\eta), h \rangle$ are asymptotically normal $\forall h \in L_2(\Pi, v)$ with the appropriate asymptotic variance, and that the sequence $\{Z_n(\eta)\}$ is tight, see e.g. Parthasarathy (1967). The idea of the proof is first, to prove these facts for a partition version of $Z_n(\eta)$, and then, to show that the remainder is negligible. These steps are considered in Theorems A.1 A.3 below hold. From these theorems and Theorem 4.2 of Billingsley (1968) Theorem 1 follows. We write for some integer K,

$$Z_n(\eta) \quad \sum_{j\ 1}^{K} (n\ j)^{1/2} \widehat{r}_j(x) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi} + \sum_{j\ K+1}^{n-1} (n\ j)^{1/2} \widehat{r}_j(x) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi}$$
$$: \ Z_n^K(\eta) + R_n^K(\eta), \qquad (28)$$

say.

Theorem A.1. *Under the conditions of Theorem* 1, *for an arbitrary but fixed integer* K *the finite dimensional distributions of* $Z_n^K(\eta)$, $\langle Z_n^K(\eta), h \rangle$, *converge to those of* $Z^K(\eta)$, $\langle Z^K(\eta), h \rangle$, $\forall h \in L_2(\Pi, v)$, *where* $Z^K(\eta)$ *is a Gaussian process with zero mean and asymptotic projected variances*

$$\sigma_{h,K}^2 : \ Var[\langle Z^K, h \rangle]$$
$$\sum_{j\ 1}^{K} \sum_{k\ 1}^{K} E\left[(Y_1 \quad \mu)^2 \int_{\Pi \times \Pi} h(\eta) h^c(\eta') \phi_{t-j}^c(x) \phi_{t-k}(x') \Psi_j(\lambda) \Psi_k(\lambda') \, dv(\eta) \, dv(\eta') \right]$$

Theorem A.2. *Under the conditions of Theorem* 1, *for an arbitrary but fixed integer* K *the sequence* $\{Z_n^K(\eta)\}$ *is tight.*

Theorem A.3. *Under conditions of Theorem* 1, *the process* $R_n^K(\eta)$ *verifies for all* $\varepsilon > 0$

$$\underset{K\longrightarrow\infty}{\text{Lim}} \ \underset{n\longrightarrow\infty}{\text{Lim}} \ P[\|R_n(\eta)\| > \varepsilon] \quad 0.$$

Proof of Theorem A.1. Note that

$$\langle Z_n^K, h \rangle \quad \sum_{t=2}^{n} (Y_t \quad \mu) \left\{ \sum_{j=1}^{(t-1)\wedge K} (n \quad j)^{-1/2} \int_{\Pi} \phi_{t-j}(x) h^c(\eta) \Psi_j(\lambda) \, \mathrm{d}\nu(\eta) \right\}$$

$$: \quad \sum_{t=2}^{n} S_{h,t}^K \quad \sum_{t=2}^{K+1} S_{h,t}^K + \sum_{t=K+2}^{n} S_{h,t}^K, \tag{29}$$

where $S_{h,t}^K(\eta): (Y_t \quad \mu) Q_{h,t}^K$ and $Q_{h,t}^K$ is implicitly defined. Under (1), $\{S_{h,t}^K, \mathscr{F}_t\}$ is an adapted MDS with stationary and ergodic differences for $t \geqslant K + 2$. Applying Markov's inequality it is easy to show that the first summand in the right hand side of (29) goes to zero in probability. For the second, the central limit theorem (CLT) for martingales with stationary and ergodic differences (Billingsley (1961)) states that the process converges to a normal distribution. Now we check that the limit variance is the appropriate, under (1) and stationarity

$$\hat{\sigma}_{h,K}: \quad Var[\langle Z_n^K, h \rangle] \quad \sum_{t=2}^{n} \mathrm{E}[S_{h,t}^K S_{h,t}^K] \quad \sum_{t=2}^{n} \mathrm{E}[(Y_t \quad \mu)^2 |Q_{h,t}^K|^2]$$

$$: \quad \sum_{j=1}^{K} \sum_{k=1}^{K} (n \quad j)^{-1/2} (n \quad k)^{-1/2} \sum_{t=1+k\vee j}^{n} \mathrm{E}[(Y_t \quad \mu)^2 a_{tjk}]$$

$$\sum_{j=1}^{K} \sum_{k=1}^{K} (n \quad j)^{-1/2} (n \quad k)^{-1/2} (n \quad k \vee j) \mathrm{E}[(Y_1 \quad \mu)^2 a_{1jk}]$$

$$\to \sum_{j=1}^{K} \sum_{k=1}^{K} \mathrm{E}[(Y_1 \quad \mu)^2 a_{1jk}] \text{ as } n \to \infty, \tag{30}$$

where

$$a_{tjk}: \quad \int_{\Pi \times \Pi} h(\eta) h^c(\eta') \phi_{t-j}^c(x) \phi_{t-k}(x') \Psi_j(\lambda) \Psi_k(\lambda') \, \mathrm{d}\nu(\eta) \, \mathrm{d}\nu(\eta').$$

Then, Theorem A.1 follows. $\square$

Proof of Theorem A.2. We apply Theorem 2.1 of Politis and Romano (1994). Again, we write

$$Z_n^K \quad n^{-1/2} \sum_{t=2}^{n} (Y_t \quad \mu) \left\{ \sum_{j=1}^{(t-1)\wedge K} \left(\frac{n}{n \quad j} \right)^{1/2} \phi_{t-j}(x) \Psi_j(\lambda) \right\}$$

$$: \quad n^{-1/2} \sum_{t=2}^{n} S_{n,t}^K \quad n^{-1/2} \sum_{t=2}^{K+1} S_{n,t}^K + n^{-1/2} \sum_{t=K+2}^{n} S_{n,t}^K, \tag{31}$$

where $S_{n,t}^K$ is implicitly defined. For each fixed $K \geqslant 1$ we have that $n^{-1/2} \sum_{t=2}^{K+1} S_{n,t}^K$ is tight because it is a finite sum and each summand is tight, see Theorem 1.4 in

[Billingsley (1968)]. Then, we concentrate on the second summand in the right hand side of (31). To verify Theorem 2.1 in [Politis and Romano (1994)] we have to show that the following conditions hold:

(i) $E\|S_{n,t}^K\|^2 < \infty$

(ii) For each integer $m \geqslant 2$, $(S_{n,K+2}^K, S_{n,K+3}^K, \ldots, S_{n,K+m}^K)$ regarded as a random element of $L_2(\Pi, v)$ converges in distribution to $(S_{K+2}^K, S_{K+3}^K, \ldots, S_{K+m}^K)$, say.

(iii) For each integer $m \geqslant 2$, $E[\langle S_{n,K+2}^K, S_{n,K+m}^K \rangle] \to E[\langle S_{K+2}^K, S_{K+m}^K \rangle]$ as $n \to \infty$.

(iv) $\mathrm{Lim}_{n \to \infty} \sum_{t\,K+2}^n E[\langle S_{n,K+2}^K, S_{n,t}^K \rangle] \quad \sum_{t\,K+2}^\infty E[\langle S_{K+2}^K, S_t^K \rangle] < \infty$, and the last series converges absolutely.

(v) $Var[\langle Z_n^K, h \rangle] \to \sigma_{h,K}^2$: $Var[\langle Z^K, h \rangle]$ as $n \to \infty$.
  Conditions (i) and (ii) are trivially satisfied. Conditions (iii) and (iv) follow easily from (1), because in that case $\{S_t^K, \mathscr{F}_t\}$ is an adapted MDS. Finally, condition (v) has been proved in (30). $\square$

Proof of Theorem A.3. By Tchebychev's inequality it is sufficient to prove that

$$\mathrm{Lim}_{K \to \infty} \mathrm{Lim}_{n \to \infty} E[\|R_n^K(\eta)\|^2] \quad 0.$$

Then under the null and A1(b),

$$E[\|R_n^K(\eta)\|^2] \quad n^{-1} \sum_{t\,2+K}^n E\left[(Y_t \quad \mu)^2 \sum_{j\,1+K}^{t-1} \left(\frac{n}{n\ \ j}\right) \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\phi_{t-j}(x)|^2 W(\mathrm{d}x)\right]$$

$$\leqslant C \sum_{j\,1+K}^\infty \left(\frac{n}{n\ \ j}\right) \frac{1}{(j\pi)^2} \to 0.$$

by first letting $n \to \infty$ and then $K \to \infty$. $\square$

Proof of Theorem 2. Write

$$\frac{1}{n} D_n^2 \quad \sum_{j\,1}^{n-1} \frac{n\ \ j}{n} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\widehat{\gamma}_j(x)|^2 W(\mathrm{d}x) \quad A_1 + A_2 + A_3,$$

where

$$A_1 \quad \sum_{j\,1}^{n-1} \frac{n\ \ j}{n} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\widehat{\gamma}_j(x) \quad \gamma_j(x)|^2 W(\mathrm{d}x),$$

$$A_2 \quad \sum_{j\,1}^{n-1} \frac{n\ \ j}{n} \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\gamma_j(x)|^2 W(\mathrm{d}x)$$

and

$$A_3 \quad \sum_{j=1}^{n-1} \frac{n}{n} \frac{j}{(j\pi)^2} \frac{2}{(j\pi)^2} \operatorname{Re}\left[\int_{\mathbb{R}} (\widehat{\gamma}_j(x) \quad \gamma_j(x))\gamma_j^c(x) W(\mathrm{d}x)\right].$$

We shall prove that $A_1 \quad o_P(1)$. Using a UET for stationary and ergodic sequence is easy to show that

$$\sup_{x\in[-a,a]} |\widehat{\gamma}_j(x) \quad \gamma_j(x)| \quad o_p(1) \quad \forall 1\leqslant j<n, \ \forall a>0.$$

Then, from Lemma 1 we have that $A_1 \quad o_P(1)$. Using this and that $A_2 \quad O(1)$, from the Cauchy Schwartz's inequality we also conclude that $A_3 \quad o_P(1)$. $\quad\square$

Proof of Theorem 3. Let $v_{t,n}: Y_t \quad \mu \quad n^{-1/2}g_t$ and $\widehat{\phi}_{t-j}(x): \mathrm{e}^{ixY_{t\,j}} \quad (n \quad j)^{-1} \sum_{t=1+j}^{n} \mathrm{e}^{ixY_{t\,j}}$. Then, by simple algebra

$$\widehat{\gamma}_j(x) \quad (n \quad j)^{-1} \sum_{t=1+j}^{n} \left(Y_t \quad \mu \quad \frac{g_t}{\sqrt{n}}\right)\mathrm{e}^{ixY_{t\,j}}$$

$$(n \quad j)^{-1} \sum_{t=1+j}^{n} \left(\overline{Y}_{n-j} \quad \mu \quad \frac{g_t}{\sqrt{n}}\right)\mathrm{e}^{ixY_{t\,j}}$$

$$(n \quad j)^{-1} \sum_{t=1+j}^{n} v_{t,n}\widehat{\phi}_{t-j}(x) + n^{-1/2}(n \quad j)^{-1} \sum_{t=1+j}^{n} g_t\widehat{\phi}_{t-j}(x)$$

$$: \widehat{\gamma}_{1j}(x) + n^{-1/2}\widehat{\gamma}_{2j}(x).$$

Then, substituting in $S_n(\eta)$ yields

$$S_n(\eta) \quad \sum_{j=1}^{n-1} (n \quad j)^{1/2}\widehat{\gamma}_j(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}: \ S_{1n}(\eta) + G_n(\eta),$$

where

$$S_{1n}(\eta): \sum_{j=1}^{n-1} (n \quad j)^{1/2}\widehat{\gamma}_{1j}(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}$$

and

$$G_n(\eta): \sum_{j=1}^{n-1} n^{-1/2}(n \quad j)^{1/2}\widehat{\gamma}_{2j}(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}.$$

Notice that under $\mathrm{H}_{A,n}$, $v_{t,n}$ is a MDS with respect to the σ field $\mathscr{F}_t$, so that by Theorem 1 it is straightforward to show that $S_{1n}(\eta)$ converges to $S(\eta)$. On the other hand, by a UET for ergodic stationary time series

$$\sup_{x\in[-a,a]} |\widehat{\gamma}_{2j}(x) \quad \mu_j(x)| \quad o_p(1) \quad \forall 1\leqslant j<n, \ \forall a>0.$$

Hence, using Lemma 1 is easy to show that $G_n(\eta)$ converges in probability in $L_2(\Pi, \nu)$ to $G(\eta)$. Thus, from Slutsky's Theorem

$$S_n(\eta) \Rightarrow S(\eta) + G(\eta).$$

This proves the Theorem. $\quad \square$

Proof of Theorem 4. We need to show that the process $S_n^*(\eta)$ (conditionally on the sample) has the same asymptotic finite projections as the process $S_n(\eta)$ and that $S_n^*(\eta)$ is tight. First, define similarly to Lemma 2

$$Z_n^*(\eta) : \sum_{j=1}^{n-1} (n-j)^{1/2} \widehat{r}_j^*(x) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi},$$

where

$$\widehat{r}_j^*(x) \quad \frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu)\phi_{t-j}(x)w_t.$$

Then, some algebra shows that

$$\widehat{\gamma}_j^*(x) \quad \widehat{r}_j^*(x) \quad (\widehat{\varphi}_{n-j}(x) \quad \varphi(x)) \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu)w_t \right]$$

$$(\overline{Y}_n - \mu) \left[\frac{1}{n-j} \sum_{t=1+j}^{n} \phi_{t-j}(x)w_t \right],$$

where $\widehat{\varphi}_{n-j}(x) : (n-j)^{-1} \sum_{t=1+j}^{n} e^{i x Y_{t-j}}$. Now, we follow similar arguments as in the proof of Lemma 2 and write $S_n^*(\eta) \quad Z_n^*(\eta) \quad R_n^*(\eta)$ where $R_n^*(\eta)$ is implicitly defined. By a simple inequality

$$\int_{\mathbb{R}} \int_0^1 |R_n^*(\eta)|^2 \, dW \, d\lambda \leqslant 2 \int_{\mathbb{R}} \int_0^1 |R_{1n}^*(\eta)|^2 \, dW \, d\lambda + 2 \int_{\mathbb{R}} \int_0^1 |R_{2n}^*(\eta)|^2 \, dW \, d\lambda,$$

where

$$\| R_{1n}^*(\eta) \|^2 \quad \sum_{j=1}^{n-1} \frac{(n-j)}{(j\pi)^2} \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu)w_t \right]^2 \int_{\mathbb{R}} |\widehat{\varphi}_{n-j}(x) \quad \varphi(x)|^2 W(dx),$$

and

$$\| R_{2n}^*(\eta) \|^2 \quad \sum_{j=1}^{n-1} \frac{(n-j)}{(j\pi)^2} (\overline{Y}_n - \mu)^2 \int_{\mathbb{R}} \left| \frac{1}{n-j} \sum_{t=1+j}^{n} \phi_{t-j}(x)w_t \right|^2 W(dx).$$

Now, using standard bootstrap notation, denote by E^* and V^* the expectation and the variance, respectively, given the sample. Then, it is not difficult to show that

$$E^* \left[\frac{1}{n-j} \sum_{t=1+j}^{n} (Y_t - \mu)w_t \right]^2 \quad O_P((n-j)^{-1})$$

and

$$\text{E}^* \left| \frac{1}{n \ j} \sum_{t \ 1+j}^{n} \phi_{t-j}(x) w_t \right|^2 \quad \text{O}_\text{P}((n \ j)^{-1}),$$

where the last inequality holds uniformly in x. So, from these inequalities, an Ergodic Theorem for $(\overline{Y}_n \quad \mu)$, a UET for $|\widehat{\varphi}_{n-j}(x) \quad \varphi(x)|$ and a partition argument as in Lemma 1 we conclude that $\|R_{1n}^*(\eta)\|^2 \quad o_\text{P}(1)$ a.s. and $\|R_{2n}^*(\eta)\|^2 \quad o_\text{P}(1)$ a.s. Therefore,

$$\left| \int_\mathbb{R} \int_0^1 |S_n^*(\lambda, x)|^2 W(\mathrm{d}x) \, \mathrm{d}\lambda \quad \int_\mathbb{R} \int_0^1 |Z_n^*(\lambda, x)|^2 W(\mathrm{d}x) \, \mathrm{d}\lambda \right| \xrightarrow{\text{P}} 0 \text{ a.s..}$$

To show that the finite dimensional projections of $Z_n^*(\eta)$ converge (conditional on the original sample) to those of $S(\eta)$ a.s. for all samples, we consider a partition argument similar to that used in the proof of Theorem 1. Decompose analogously to (28) $Z_n^*(\eta) \quad Z_n^{K*}(\eta) + R_n^{K*}(\eta)$ for fixed K. We shall show that first, $\langle Z_n^{K*}(\eta), h \rangle$ converges (conditionally on the original data) in distribution to $\langle Z^K(\eta), h \rangle$ a.s. for all samples and $\forall h \in L_2(\Pi, v)$, and second that remainder $R_n^{K*}(\eta)$ is negligible a.s. for all samples. To this end, write

$$\langle Z_n^{K*}(\eta), h \rangle \quad n^{-1/2} \sum_{t\ 2}^{n} (Y_t \quad \mu) \left\{ \sum_{j\ 1}^{(t-1)\wedge K} \left(\frac{n}{n \ j} \right) \int_\Pi^{1/2} \phi_{t-j}(x) h^c(\eta) \Psi_j(\lambda) \, \mathrm{d}v(\eta) \right\} w_t$$

$$n^{-1/2} \sum_{t\ 2}^{n} (Y_t \quad \mu) Q_{h,t}^K w_t \quad \sum_{t\ 2}^{n} \zeta_{nt}^{K*},$$

where $\zeta_{nt}^{K*} \quad n^{-1/2}(Y_t \quad \mu) Q_{h,t}^K w_t$ and $Q_{h,t}^K$ is implicitly defined. Then

$$\text{E}^* \left(\sum_{t\ 2}^{n} \zeta_{nt}^{K*} \right) \quad \sum_{t\ 2}^{n} n^{-1/2} (Y_t \quad \mu) Q_{h,t}^K \text{E}(w_t) \quad 0,$$

while

$$V^* \left(\sum_{t\ 2}^{n} \zeta_{nt}^{K*} \right) \quad \sum_{t\ 2}^{n} n^{-1} (Y_t \quad \mu)^2 |Q_{h,t}^K|^2 V(w_t)$$

$$\sum_{t\ 2}^{n} n^{-1} (Y_t \quad \mu)^2 |Q_{h,t}^K|^2 \colon \ \widetilde{\sigma}_{h,K}^2.$$

Note that conditional on the original data, ζ_{nt}^{K*} is an independent (not identically distributed) array of random variables. Note that using a SLLN for stationary and ergodic sequences is easy to show that $\widetilde{\sigma}_{h,K}^2 \to \sigma_{h,K}^2$ a.s.. Then, we will verify a Lindeberg Feller's condition. On the other hand using that $Q_{h,t}^K$ and w_t are bounded

$$\sum_{t\ 2}^{n} \text{E}^*[|\zeta_{nt}^{K*}|^2 1(|\zeta_{nt}^{K*}| > \delta)] \leqslant \frac{C}{n} \sum_{t\ 2}^{n} (Y_t \quad \mu)^2 1(|(Y_t \quad \mu)| > \delta' \sqrt{n})] \text{ a.s.}$$

for some positive constants δ and δ'. By A1 the last expression converges almost surely to zero, see Stute et al. (1998, p. 149). Then the triangular array $\{\zeta_{nt}^{K*}\}$ satisfies the conditions of the Lindeberg Feller's CLT, conditionally on almost all samples, so that $\sum_{t\;2}^{n} \zeta_{nt}^{K*} \Rightarrow_* N(0, \sigma_{h,K}^2)$ a.s.. Next, arguing as in the proof of Theorem A.3,

$$
E^*[\|R_n^{K*}(\eta)\|^2] \quad n^{-1} \sum_{t\;2+K}^{n} (Y_t \quad \mu)^2 \sum_{j\;1+K}^{t-1} \left(\frac{n}{n\;j}\right) \frac{1}{(j\pi)^2} \int_{\mathbb{R}} |\phi_{t-j}(x)|^2 W(dx)
$$

$$
\leqslant C\left(n^{-1} \sum_{t\;2+K}^{n} (Y_t \quad \mu)^2\right) \sum_{j\;1+K}^{\infty} \left(\frac{n}{n\;j}\right) \frac{1}{(j\pi)^2} \to 0 \text{ a.s.},
$$

by first letting $n \to \infty$ and then $K \to \infty$. Finally, we have to prove the tightness of the sequence $\{S_n^*(\eta)\}$ a.s.. Write

$$
S_n^*(\eta) \quad n^{-1/2} \sum_{t\;2}^{n} (Y_t \quad \overline{Y}_n)\left\{\sum_{j\;1}^{(t-1)} \left(\frac{n}{n\;j}\right)^{1/2} \widehat{\phi}_{t-j}(x))\Psi_j(\lambda)\right\} w_t
$$

$$
: \quad n^{-1/2} \sum_{t\;2}^{n} S_{n,t}^*.
$$

Note that $S_{n,t}^*$ and $S_{n,s}^*$ are independent given the sample for $s \neq t$, thus it is sufficient for the tightness that $E^*\|S_{n,t}^*\|^2 < \infty$ a.s. for all samples, which is trivially satisfied, see example 1.8.5 in van der Vaart and Wellner (1996). The proof is finished. $\square$

References

Bierens, H.J., 1982. Consistent model specification tests. Journal of Econometrics 20, 105 134.

Bierens, H.J., 1984. Model specification testing of time series regressions. Journal of Econometrics 26, 323 353.

Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443 1458.

Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129 1151.

Billingsley, P., 1961. The Lindeberg Levy theorem for martingales. Proceedings of the American Mathematical Society 12, 788 792.

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

Bosq, D., 2000. Linear Processes in Function Spaces. Theory and Applications, Lecture Notes in Statistics, vol. 149, Springer, Berlin.

Chen, X., White, H., 1996. Laws of large numbers for Hilbert space-valued mixingales with applications. Econometric Theory 12, 284 304.

Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. Econometric Theory 14, 260 284.

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

de Jong, R.M., 1996. The Bierens' tests under data dependence. Journal of Econometrics 72, 1 32.

Dehling, H., Philipp, W., 2002. Empirical Process Techniques for Dependent Data. In: Dehling, H., Mikosch, T., Sørensen, M. (Eds.), Empirical Process Techniques for Dependent Data. Birkhäuser, Basel, pp. 3 113.

Delgado, M., Hidalgo, J., Velasco, C., 2004. Distribution free goodness-of-fit tests for linear models. The Annals of Statistics, to appear.

Deo, R.S., 2000. Spectral tests of the Martingale Hypothesis under conditional heteroskedasticity. Journal of Econometrics 99, 291 315.

Domínguez, M., Lobato, I.N., 2003. A consistent test for the martingale difference hypothesis. Econometric Reviews 22, 351 377.

Durlauf, S., 1991. Spectral-based test for the martingale hypothesis. Journal of Econometrics 50, 1 19.

Epps, T.W., Pulley, L.B., 1983. A test for normality based on the empirical characteristic function. Biometrika 70, 723 726.

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

Fong, W.M., Ouliaris, S., 1995. Spectral tests of the martingale hypothesis for exchange rates. Journal of Applied Econometrics 10, 255 271.

Goncalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89 120.

Hall, R.E., 1978. Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. Journal of Political Economy 86, 971 987.

Hinich, M., Patterson, D., 1992. A new diagnostic test of model inadequacy which uses the martingale difference criterion. Journal of Time Series Analysis 13, 233 252.

Hong, Y., 1999a. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. Journal of the American Statistical Association 84, 1201 1220.

Hong, Y., 1999b. Testing serial independence via the empirical characteristic function, Preprint.

Hong, Y., Lee, T.H., 2003. Inference on predictability of foreign exchange rate changes via generalized spectrum and nonlinear time series models. Review of Economics and Statistics 85, 1048 1062.

Hong, Y., Lee, Y.J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Review of Economic Studies 72, 499 541.

Jennrich, R.I., 1969. Asymptotic properties of nonlinear least squares estimators. Annals of Mathematical Statistics 40, 633 643.

Kock, B., Yang, S., 1986. A method for testing the independence of two series that accounts potential pattern in the cross correlations. Journal of the American Statistical Association 81, 533 544.

Koul, H.L., Stute, W., 1999. Nonparametric model checks for time series. The Annals of Statistics 27, 204 236.

Kuan, C.-M., Lee, W., 2003. A new test of the martingale difference hypothesis. Preprint.

Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. Journal of Econometrics 112, 295 325.

Liu, R.Y., 1988. Bootstrap procedures under some non-i.i.d models. The Annals of Statistics 16, 1696 1708.

Lo, A.W., 1997. Market Efficiency: Stock Market Behaviour in Theory and Practice, vols. I and II. Edward Elgar.

Mammen, E., 1993. Bootstrap and wild bootstrap for high-dimensional linear models. The Annals of Statistics 21, 255 285.

Neuhaus, G., 1976. Asymptotic power properties of the Cramér von Mises test under contiguous alternatives. Journal of Multivariate Analysis 6, 95 110.

Paparoditis, E., 2000. Spectral density based goodness-of-fit tests for time series models. Scandinavian Journal of Statistics 27, 143 176.

Park, J.Y., Whang, Y.J., 1999. Testing for the martingale hypothesis. Preprint.

Parthasarathy, K.R., 1967. Probability Measures on Metric Spaces. Academic Press, New York.

Politis, D., Romano, J., 1994. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. Statistica Sinica 4, 461 476.

Satterthwaite, F.E., 1941. Synthesis of variance. Psychometrica 6, 309 316.

Satterthwaite, F.E., 1946. An approximate distribution of estimates of variances components. Biometrics Bulletin 2, 110 114.

Shorack, G., Wellner, J., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

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. The Annals of Statistics 25, 613–641.

Stute, W., Gonzalez-Manteiga, W., Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141–149.

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

Whang, Y.-J., 2000. Consistent bootstrap tests of parametric regression functions. Journal of Econometrics 98, 27–46.

Wu, C.F.J., 1986. Jacknife, bootstrap and other resampling methods in regression analysis with discussion. The Annals of Statistics 14, 1261–1350.