A Consistent Test for the Martingale Di¤erence Hypothesis

Manuel A. Domínguez and Ignacio N. Lobato

Instituto Tecnológico Autónomo de México

January 14, 2000

Abstract

This paper considers testing that an economic time series follows a martingale difference process. The martingale di¤erence hypothesis has been typically tested using information contained in the second moments of a process, that is, using test statistics based on the sample autocovariances or in the periodograms. Tests based on these statistics are inconsistent since they just test necessary conditions of the null hypothesis. In this paper we consider tests that are consistent against all …xed alternatives and against Pitman´s local alternatives. Since the asymptotic distributions of the tests statistics depend on the data generating process, the tests are implemented using a modi…cation of the wild bootstrap procedure. The paper justi…es theoretically the proposed tests and examines their …nite sample behavior by means of Monte Carlo experiments. In addition we include an application to exchange rate data.

JEL nos: C12, C14, C15, C22 and C52.

Keywords: nonlinear dependence, nonparametric, correlation, bootstrap.

Please send further communications to:

Ignacio N. Lobato

Centro de Investigación Económica

Instituto Tecnológico Autónomo de México

Av. Camino Sta. Teresa #930

10700 México D.F.

MEXICO

1 Introduction

In Economics it is frequently assumed that an economic time series follows a martingale diﬀerence sequence (MDS) given some information set. For instance, it is a common implication in Rational Expectations models (see, for instance, Hall (1978)). A MDS process is deﬁned as a process that has constant mean (usually zero) given some information set (that typically includes just its past values). Testing the MDS assumption is rather challenging. In Econometrics the common way of testing this property has been testing that the process is uncorrelated. Hence, in the time domain the test statistics typically employed have been based on the sample autocorrelations while in the frequency domain they have been based on the periodograms. The test statistic most commonly employed has been the Box and Pierce (1970) Q statistic for testing that a process is uncorrelated of a given order p (that is, the ﬁrst p autocorrelations are equal to zero). For the general case of testing that a process is uncorrelated of any order, alternative statistics have been proposed by Durlauf (1991), Anderson (1993) or Hong (1996).

Notice that all these procedures do not test that the considered process is a MDS but that it is uncorrelated. This distinction is crucial when nonlinear dependence is present, as it commonly happens with ﬁnancial data. For processes with bounded second moment, a MDS is an uncorrelated sequence, but an uncorrelated sequence is not necessarily a MDS. An uncorrelated process cannot be forecasted using linear functions of lagged values, while a MDS cannot be forecasted using either linear or nonlinear functions of past values. Hence, for uncorrelated non-MDS processes the previous tests have no asymptotic power (see, for instance, examples in Section 4). The fact that these tests are inconsistent can be understood since they only employ information contained in the second moments of the process. Contrary to these commonly employed tests, this paper provides consistent tests for the null hypothesis that the process has zero conditional expectation given the information set composed by the current value of some exogenous variables and some ﬁnite numbers of lags of both the own process and some exogenous variables.

Consistent tests for the MDS assumption can be established using recently developed statistical theory on speciﬁcation testing. There are basically two approaches to constructing consistent tests. First, to employ tests based on checking an inﬁnite number of orthogonality conditions (see, for instance, Bierens (1984, 1990), Stute (1997), Andrews (1997), Bierens and Ploberger (1997) and Koul and Stute (1999)). Second, to employ tests based on smoothed nonparametric estimates of the conditional expectation function (see, for instance, Härdle and Mammen (1993), Hong and White (1995), Zheng (1996) and Li (1999)).

Test statistics based on the first approach do not demand the selection of user-chosen tuning parameters. However, they have the disadvantage of (typically) having non-standard asymptotic null distributions. This is not a serious drawback, though, since critical values can be estimated by the bootstrap. In addition, these tests are consistent against Pitman´s local alternatives but inconsistent against certain local alternatives to the null, see Rosenblatt (1975) or Horowitz and Spokoiny (1999).

Tests based on the second approach have the advantages of having standard asymptotic null distributions and being consistent against Rosenblatt's (1975) local alternatives, but three inconveniences. First, they require stronger smoothness assumptions on the data generating process (DGP). Second, they have asymptotically no power against Pitman´s alternative hypotheses tending to the null at the parametric rate. Third, their main disadvantage is that they require a user-chosen smoothing parameter and, in practice, statistical inference is quite sensitive to the selection of this number.

For these reasons, in this paper we employ a test based on the first approach. Since the asymptotic distribution of the considered test statistic depends on the specific DGP, standard asymptotic inference procedures are not feasible. In this paper we propose and justify rigorously to estimate the distribution of the test statistic by using a modification of the wild bootstrap.

The organization of the paper is the following. In Section 2 we review the different approaches to construct consistent tests, and motivate the selection of our test statistics which are introduced in Section 3. Section 4 analyzes the bootstrap tests and in Section 5 we present a Monte Carlo study of their finite sample performance. Section 6 reports an empirical application to exchange rates and Section 7 concludes and establishes some directions of further research. Proofs are in the Appendix.

2 Consistent Hypothesis Testing

Let y_t be an ergodic and strictly stationary process of which a sample of $n + p$ observations, $(y_{-p+1}, \ldots, y_0, y_1, \ldots, y_n)$; is available. Let $\tilde{x}_t = (x_{1,t},\ldots, x_{K,t})'$ be a $K \times 1$ ergodic and stationary stochastic vector process of conditioning variables. We employ the superscript to denote vectors. Notice that no assumptions are made about the moment structure of $\tilde{x}_t$ and, in particular, all their moments could be unbounded. Of each variable $\{x_{i,t}\}$, $i = 1, \ldots K$, a sample of size $n + p_i$ is observed. Denote the whole observed sample by

$X_n = (y_{-p+1}, ..., y_0, y_1, ..., y_n, x_{1,-p_1+1}, ..., x_{1,n}, ..., x_{K,-p_K+1}, ..., x_{K,n})$. The null hypothesis of interest that we consider in this paper is testing mean independence with respect to the information set $I_{t,\mathbf{p}} = \{y_{t-1}, ..., y_{t-p}, x_{1,t}, x_{1,t-1}, ..., x_{1,t-p_1}, ..., x_{K,t}, x_{K,t-1}, ..., x_{K,t-p_K}\}$ where $\mathbf{p} = (p, p_1, ..., p_K)'$ and $p, p_1, ..., p_K$ are any natural numbers. Notice that the conditioning information set includes the past p values of the considered process and current and past values of the other conditioning processes.

Thus, the considered null hypothesis is

$$H_0 : E(y_t / I_{t,\mathbf{p}}) = \mu \quad a.s. \tag{1}$$

for some unknown $\mu \in \mathbb{R}$, and the alternative

$$H_A : E(y_t / I_{t,\mathbf{p}}) = \mu\left(I_{t,\mathbf{p}}\right) \quad a.s.; \tag{2}$$

where $\mu(\cdot)$ is some unknown measurable function of $I_{t,\mathbf{p}}$ from $\mathbb{R}^P$ into $\mathbb{R}$, where $P = p + K + \sum_{j=1}^{K} p_j$, such that $\Pr\left(\mu\left(I_{t,\mathbf{p}}\right) = \mu\right) < 1$. A process that verifies (1) is said to be a martingale difference sequence of order p with respect to its past and of orders p_i with respect to x_i, for $i = 1, ..., K$, (more briefly, we say that y_t is a MDS of orders $\mathbf{p}$). In Sections 5 and 6 we consider the special case of testing that a process is just a martingale difference sequence of order p with respect to its own past. We establish the theoretical results for the general case since economic theory establishes the orthogonality with respect to the agent information set that typically includes a set of additional explanatory variables.

In order to obtain consistent tests of the null hypothesis (1) there are two approaches: the use of tests based on empirical processes indexed by classes of functions and the use of tests based on nonparametric estimates. Both approaches are nonparametric in spirit. For simplicity, we call the integrated approach to the first (since the corresponding tests require the selection of an integrating measure) and the smoothing approach to the second (since the corresponding tests require the selection of a smoothing number). Both are based on the following equivalence which is based on the definition of conditional expectation (see, for instance, p.63 in Brockwell and Davies (1993))

$$H_0 \Leftrightarrow E((y_t - \mu)W(I_{t,\mathbf{p}})) = 0 \quad a.s.$$

for any bounded measurable weighting function $W(\cdot)$ with respect to $I_{t,\mathbf{p}}$. The tests are based on evaluating the discrepancy of the sample analog of $E((y_t - \mu)W(I_{t,\mathbf{p}}))$ to zero.

Notice that any such test involves testing an infinite number of orthogonality conditions. The smoothing approach reduces the problem of testing H_0 to testing a unique, appropriately chosen, orthogonality restriction. Namely, it employs $W^S(\mathbf{Y}_{t,P}) = E((y_t - \mu)|\mathbf{Y}_{t,P})$. Hence, this methodology is based on the following equivalence

$$H_0 \Leftrightarrow E((y_t - \mu)W^S(\mathbf{Y}_{t,P})) = 0.$$

Notice that this approach, although implicitly, involves testing an infinite number of orthogonality restrictions as well, as we show now. First, express $W^S(\mathbf{Y}_{t,P})$ as a linear combination of a basis $\{w_i\} = \{w_i(\mathbf{Y}_{t,P})\}$ of the space of functions with finite second moment

$$W^S(\mathbf{Y}_{t,P}) = \sum_i \alpha_i w_i,$$

where $\alpha_i = E[W^S(\mathbf{Y}_{t,P})w_i]$. Second, apply the law of iterated expectations to obtain $\alpha_i = E[E((y_t - \mu)|\mathbf{Y}_{t,P})w_i] = E[(y_t - \mu)w_i]$. Finally, testing $E((y_t - \mu)W^S(\mathbf{Y}_{t,P})) = 0$ is equivalent to testing $E((y_t - \mu)\sum_i \alpha_i w_i) = \sum_i \alpha_i E((y_t - \mu)w_i) = \sum_i \alpha_i^2 = 0$, that is, testing that $\alpha_i = 0$ for all i.

Since the function $W^S(\mathbf{Y}_{t,P})$ is unknown, evaluating its sample analog will require the use of nonparametric estimation techniques, that is, the introduction of a user-chosen smoothing number. This approach presents three drawbacks. First, its main problem is that statistical inference is sensitive to the selection of the smoothing parameter. There has been considerable research on how to select this parameter automatically from the sample for estimation problems (see Marron (1988) for a survey). Unfortunately, there is not any completely satisfactory answer yet and furthermore, most of this research has focused on estimation rather that hypothesis testing. Second, tests based on the nonparametric approach have no power against alternative hypotheses tending to the null hypothesis at the $n^{-1/2}$ rate. Third, this literature needs to impose strong smoothness conditions on the function $\mu(\mathbf{Y}_{t,P})$, see, for instance, Zheng (1996) or Li (1999).

The integrated approach tests H_0 by selecting a family of functions $\mathcal{W}$ so that H_0 holds if and only if $(y_t - \mu)$ is orthogonal to every member of $\mathcal{W}$ almost surely. Depending on the choice of $\mathcal{W}$ the corresponding test resembles certain classical goodness of fit tests employed in the statistical literature as we see now. There are two types of integrated tests. The first type has been employed by Bierens (1984, 1990), De Jong (1996), Bierens and Ploberger (1997) and Stinchcombe and White (1999). They proposed testing procedures based on families of analytic functions $\mathcal{W} = \{W(y;\beta);\ \beta \in \Upsilon,\ \Upsilon \in \mathbb{R}^P\}$ and prove that

$$H_0 \Leftrightarrow R^W(\beta) = E((y_t - \mu)W(\mathbf{Y}_{t,P};\beta)) = 0; \quad \text{for any } \beta \in \Upsilon,$$

where, in general, the set Υ is an infinite dense set, for instance, any neighborhood of $0 \in \mathbb{R}^p$. In particular, Bierens (1990) considers $W = \{\exp(i\beta'\mathbf{x}); \beta \in \Upsilon; \mathbf{x} \in \mathbb{R}^p\}$. Hence, for this choice these tests resemble goodness of fit tests based on the characteristic function. Note that under both the null and the alternative hypotheses, $R^W(\beta) \in C[\Upsilon]$, the space of continuous functions on Υ. We have called $[A]$ the closure of A and, in case the function is not defined for every β in the frontier of A, we extend the process by considering that $R^W(\beta) = \lim_{\beta_n \to \beta} R^W(\beta_n)$. The dependence of the nuisance parameter vector β is avoided by applying a norm of the space $C[\Upsilon]$ onto the function $R^W(\beta)$. Recall that a norm φ, that is a positive continuous functional, verifies that for any $f \in C[\Upsilon]$; $\varphi(f) = 0 \iff f = 0$. A main problem with this approach is that, the application of the norm requires the selection of an arbitrary measure on Υ.

The second type of integrated tests has been employed by Brunk (1970), Su and Wei (1991), An and Bing (1991), Delgado (1993), Andrews (1997), Stute (1997) and Koul and Stute (1999). They have considered the family $W_1 = \{I(\mathbf{x}_{t,p} \leq \mathbf{x}); \mathbf{x} \in \mathbb{R}^p\}$ where $I(A)$ is the indicator function of the event A and $\mathbf{x} \leq \mathbf{y}$ denotes that each element of $\mathbf{x}$ is less or equal that the corresponding of $\mathbf{y}$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$. For this family, the nuisance parameters are evaluated in the support of the conditioning vector, $\mathbf{x}_{t,p}$, and hence, the natural integrating measure is the joint empirical distribution function of the vector $\mathbf{x}_{t,p}$. Therefore, the corresponding tests resemble goodness of fit tests based on the distribution function, such as, the Cramer-von Mises test and the Kolmogorov-Smirnov test. Hence, the advantage of this family is that the arbitrariness involved in selecting an integrating measure disappears. In fact, Koul and Stute (1999) have shown that this family can be used to build consistent tests for H_0 for the case $p = 1$ and $K = 0$.

Since both the smoothing approach and the first type of integrated tests present the problem of arbitrarily selecting the smoothing number and the integrating measure, respectively, the test procedure proposed in this paper belongs to the second type of integrated tests.

3 A Consistent Test for the Martingale Difference Hypothesis of order P

We assume that $Ey_t^{4+\delta} < \infty$, for some $\delta > 0$, and that y_t given $Y_{t,P}$ has a continuous bounded density function. The proposed test is based on the following equivalence

$$H_0 \iff R(x) = 0 \text{ for almost all } x \in \mathbb{R}^P;$$

where

$$R(x) = E((y_t - \mu)I(Y_{t,P} \le x)) = \int (s-\mu)I(y \le x)dF(s;y) \in C[\mathbb{R}]^P; \qquad (3)$$

where $F(s;y)$ is the joint distribution function of the vector $(y_t; Y_{t,P})$. In order to evaluate the distance of $R(x)$ to zero, a norm has to be chosen. Denote the general norm by

$$T_P = \psi(R(x)). \qquad (4)$$

The two norms considered in this paper are the Cramer-von Mises norm, that is,

$$\psi_2(R(x)) = \left(\int [R(x)]^2 dF(\infty;x)\right)^{1/2} \qquad (5)$$

where $F(\infty;x) = \lim_{s\to\infty} F(s;x)$, and the Kolmogorov-Smirnov norm, that is,

$$\psi_\infty(R(x)) = \sup_x |R(x)|. \qquad (6)$$

A general consistent test would be based on the sample analog of (4). In particular, the tests considered in this paper are based on the empirical versions of (5) and (6). Next, we provide explicit formulae for these two test statistics.

Denote by F_n the empirical distribution function of $\left(y_t; Y_{t,P}\right)'$ and by $\bar{y}$ the sample mean $\bar{y} = \int s\, dF_n(s;y) = n^{-1}\sum_{t=1}^n y_t$. Notice that $\bar{y} = \bar{y}^C + o_p(n^{-1/2})$ where $\bar{y}^C$ is the usual definition for the sample mean that takes into account all the available observations ($\bar{y}^C = (n+p)^{-1}\sum_{t=-p+1}^n y_t$). We estimate the function $R(x)$ given in (3) by its sample analog

$$R_n(x) = \int (s-\bar{y})I(y \le x)dF_n(s;y) = \frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})I(Y_{t,P} \le x). \qquad (7)$$

Notice that $R_n(x) \in D[\mathbb{R}]^P$, where $D[\mathbb{R}]^P$ is the natural extension of the cadlag space $D[\mathbb{R}]$ considered by Koul and Stute (1999). Also, for a fixed x; under the null hypothesis, $R_n(x) = O_p(n^{-1/2})$; but under the alternative $\sqrt{n}R_n(x)$ diverges to infinity for some x

(as we will show in Theorems 1 and 2). Hence, test procedures are based on $\sqrt{n}R_n(\mathbf{x})$. The general test statistic considered is the empirical analog of (4) and we denote it by $T_{P,n} = \varphi(\sqrt{n}R_n(\mathbf{x}))$. The two particular test statistics considered here are the Cramer-von Mises test statistic

$$C_{P,n} = \int \left[\sqrt{n}R_n(\mathbf{x})\right]^2 dF_n(\infty;\mathbf{x}) = n \int \left[\int (s - \bar{y}) I(\mathbf{y} \le \mathbf{x}) dF_n(s;\mathbf{y})\right]^2 dF_n(\infty;\mathbf{x})$$

where $F_n(\infty;\mathbf{x}) = \lim_{s\to\infty} F_n(s;\mathbf{x})$; and the Kolmogorov-Smirnov statistic

$$K_{P,n} = \max_{i=1,\ldots,n} \left|\sqrt{n}R_n(\mathbf{Y}_{i,P})\right| = \max_{i=1,\ldots,n} \left|\frac{1}{\sqrt{n}} \sum_{j=1} (y_j - \bar{y}) I(\mathbf{Y}_{j,P} \le \mathbf{Y}_{i,P})\right|.$$

In order to consider the asymptotic distribution of the test statistic $T_{P,n}$, we need first to consider the asymptotic probability law of the process $\sqrt{n}R_n(\mathbf{x})$. It turns out to be Gaussian with the asymptotic covariance matrix depending on the DGP as the following Theorem shows.

Theorem 1. Under the null hypothesis (1)

$$\sqrt{n}R_n(\mathbf{x}) \Rightarrow B(\mathbf{x}); \tag{8}$$

where $\Rightarrow$ denotes weak convergence and $B(\mathbf{x})$ denotes a centered continuous Gaussian process in $D[\mathbb{R}]^P$ with covariance given by

$$\begin{aligned} \Sigma(\mathbf{x};\mathbf{x}') &= E[\sigma^2(\mathbf{Y}_{t,P}) I_t(\mathbf{x} \wedge \mathbf{x}')] - F(\infty;\mathbf{x}) E[\sigma^2(\mathbf{Y}_{t,P}) I_t(\mathbf{x}')] \\ &\quad - F(\infty;\mathbf{x}') E[\sigma^2(\mathbf{Y}_{t,P}) I_t(\mathbf{x})] + \sigma^2 F(\infty;\mathbf{x}') F(\infty;\mathbf{x}); \end{aligned} \tag{9}$$

where $\sigma^2(\mathbf{Y}_{t,P})$ is the unknown variance of y_t conditional on $\mathbf{Y}_{t,P}$, which is known to be finite, $I_t(\mathbf{x}) = I(\mathbf{Y}_{t,P} \le \mathbf{x})$ and $\mathbf{x} \wedge \mathbf{x}'$ denotes the vector whose $i-$th component is the minimum of the i-th components of the vectors $\mathbf{x}$ and $\mathbf{x}'$.

Since $R_n(-\infty) = R_n(\infty) = 0$; the asymptotic distribution is a tied-down Gaussian process. Notice, however, that the covariance structure depends on the DGP. Theorem 1 is a natural extension of Theorem 2.2 in Koul and Stute (1999), see their Remark 2.4.

When the interest resides in testing that a process is a MDS with respect to the past p values of the process, Theorem 1 particularizes to the following corollary.

Corollary 1. Let $\Omega_{t-1,p} = \{y_{t-1}, \ldots, y_{t-p}\}$ and the null hypothesis of interest be

$$H_0 : E(y_t / \Omega_{t-1,p}) = \mu \quad a.s. \tag{10}$$

for some unknown $\mu \in \mathbb{R}$ and the alternative

$$H_A : E(y_t / \mathcal{F}_{t-1,p}) = \mu_y(\mathbf{Y}_{t-1}) \quad a.s.; \tag{11}$$

where $\mu_y(\cdot)$ is some unknown measurable function of $\mathbf{Y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ from $\mathbb{R}^p$ into $\mathbb{R}$ such that $P(\mu_y(\mathbf{Y}_{t-1}) = \mu) < 1$. Let $\sqrt{n}R_{y,n}(\mathbf{x}) = n^{-1/2} \sum_{t=1}^{n} (y_t - \bar{y}) I(\mathbf{Y}_{t-1} \le \mathbf{x})$. Then, under the null hypothesis

$$\sqrt{n}R_{y,n}(\mathbf{x}) \Rightarrow B_y(\mathbf{x});$$

where $B_y(\mathbf{x})$ denotes a centered continuous Gaussian process in $D[\mathbb{R}]^p$ with covariance given by

$$\begin{aligned} \Sigma_y(\mathbf{x}, \mathbf{y}) = {} & E[\sigma^2(\mathbf{Y}_{t-1}) I^y_{t-1}(\mathbf{x} \wedge \mathbf{y})] - F(\infty; \mathbf{x}) E[\sigma^2(\mathbf{Y}_{t-1}) I^y_{t-1}(\mathbf{y})] \\ & - F(\infty; \mathbf{y}) E[\sigma^2(\mathbf{Y}_{t-1}) I^y_{t-1}(\mathbf{x})] + \sigma^2 F(\infty; \mathbf{y}) F(\infty; \mathbf{x}); \end{aligned}$$

where $\sigma^2(\mathbf{Y}_{t-1})$ is the unknown variance of y_t conditional on $\mathbf{Y}_{t-1}$ and $I^y_{t-1}(\mathbf{x}) = I(\mathbf{Y}_{t-1} \le \mathbf{x})$.

Next we provide three remarks about Theorem 1.

Remark 1. Notice that under the assumption of conditional homoskedasticity, that is $\sigma^2(\mathbf{I}_{t,p}) = \sigma^2$, the asymptotic covariance matrix (9) reduces to

$$\Sigma(\mathbf{x}, \mathbf{y}) = \sigma^2 [F(\infty; \mathbf{x} \wedge \mathbf{y}) - F(\infty; \mathbf{y}) F(\infty; \mathbf{x})]. \tag{12}$$

If, in addition, the conditioning set $\mathbf{I}_{t,p}$ includes only one variable (either one own lagged value or one contemporaneous or lagged value of a conditioning variable), using the classical quantile transformation, equation (12) simplifies to $\Sigma(\tau, \rho) = \sigma^2[(\tau \wedge \rho) - (\rho\tau)]$, and so, the process $B(\mathbf{x})$ follows a standard Brownian Bridge. Hence, for this restrictive case, inference is straightforward since the critical values are already tabulated, see Shorack and Wellner (1986, pp. 143-147).

Remark 2. A Gaussian process similar to (8) has been previously discussed in Koul and Stute (1999), see their equation at the end of p.211. The null hypothesis considered by Koul and Stute (1999) is different to (1). They considered the case where $\mathbf{I}_{t,p}$ only includes one past value of the process (hence, no additional conditioning variables or lagged values of the process are allowed), but allow for a general functional form for the conditional expectation function. For the even more specific case of conditional homoskedasticity (in which $\sigma^2(\mathbf{Y}_{t-1}) = \sigma^2$), they proposed a transformation of $\sqrt{n}R_n(\mathbf{x})$ to obtain a pivotal test statistic. However, as we have seen in the previous remark, for this special context, we do not need to transform the statistic since we already have a pivotal distribution. Koul and

Stute (1999) only justified the transformation for this restrictive case and the extension to the general case does not appear to be feasible.

Remark 3. The functional central limit theorem (8) can be obtained under alternative sets of assumptions. Here we have followed Koul and Stute (1999) and assumed ergodicity and strict stationarity with finite fourth moment. Some of these assumptions could be weakened at the cost of strenghtening others. For instance, the strict stationarity assumption could have been removed at the cost of strengthening to strong mixing the condition on the dependence of the process (see, for instance, Andrews and Pollard (1994) and references therein). Continuity of the density of y_t given $\mathbf{Y}_{t,P}$ could also be weakened by assuming that the conditional second moment $E(y_t^2/\mathbf{Y}_{t,P})$ is a continuous function, see Koul and Stute (1999, p.219).

Next, we provide two Theorems about the behavior of the process $\sqrt{n}R_n(\mathbf{x})$ under the alternative hypothesis.

Theorem 2. Under the alternative hypothesis (2), there exists a $T \subset \mathbb{R}^P$ such that $\Pr\left(\mathbf{Y}_{t,P} \in T\right) > 0$ and for all $\mathbf{x} \in T$; $R_n(\mathbf{x}) \Rightarrow R(\mathbf{x}) \neq 0$. Hence, under the alternative hypothesis (2), $\sqrt{n}R_n(\mathbf{x})$ diverges.

The next Theorem shows the behavior of the process $\sqrt{n}R_n(\mathbf{x})$ under a sequence of alternative hypotheses tending to the null at the rate $n^{-1/2}$. Consider the following sequence of alternative hypotheses

$$H_{A,n} : E(y_t/\mathbf{Y}_{t,P}) = \mu + \frac{g(\mathbf{Y}_{t,P})}{\sqrt{n}} \quad a.s. \tag{13}$$

for any function $g(\cdot)$ such that $\Pr(g(\mathbf{Y}_{t,P}) = \text{constant}) < 1$.

Theorem 3. Under the sequence of alternative hypotheses (13)

$$\sqrt{n}R_n(\mathbf{x}) \Rightarrow B(\mathbf{x}) + G(\mathbf{x});$$

where $G(\mathbf{x}) = E(g(\mathbf{Y}_{t,P})w_t(\mathbf{x})) \neq 0$, where $w_t(\mathbf{x}) = I_t(\mathbf{x}) - F(\mathbf{x})$ and $K(\mathbf{x};\mathbf{y})$ is given in (9).

Using the previous three theorems and the Continuous Mapping Theorem the following corollary establishes the asymptotic behavior of the general test statistic $T_{P,n}$.

Corollary 2. Under the null hypothesis (1), $T_{P,n} \Rightarrow \varphi(B(\mathbf{x}))$; under the alternative hypothesis (2), $T_{P,n}$ diverges; and under the sequence of alternative hypotheses (13), $T_{P,n} \Rightarrow \varphi(B(\mathbf{x}) + G(\mathbf{x}))$.

Notice that the asymptotic null distribution of $T_{P,n}$ is given by $\varphi(B(\mathbf{x}))$ that depends on the specific DGP. Hence, standard asymptotic inference procedures cannot be applied.

4 The Bootstrap Test

Since the asymptotic distribution of $\sqrt{n}R_n(\theta)$ depends, in general, on the DGP, the one corresponding to $T_{\theta,n}$ also depends on the DGP. Hence, the theory in the previous section cannot be automatically applied for statistical inference: In this section we propose to estimate this unknown distribution using a modification of the wild bootstrap. Notice that our solution is valid for the general case, that is, when additional conditioning variables or more than one lagged value of the process are included in the conditioning set, and conditional heteroskedasticity is present. These generalizations are important because economic theory typically includes additional explanatory variables and conditional heteroskedasticity is a well-known feature of financial data.

Next, we explain and justify theoretically the proposed bootstrap-based test procedure. Let $\varepsilon_t = (y_t - \mu)$ and notice that uniformly in θ

$$
\begin{aligned}
\sqrt{n}R_n(\theta) &= \frac{1}{\sqrt{n}}\sum_{t=1} \varepsilon_t I_t(\theta) - \frac{1}{\sqrt{n}}\sum_{s=1}\frac{1}{n}\sum_{t=1} \varepsilon_t I_s(\theta) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1} \varepsilon_t [I_t(\theta) - F_n(\infty,\theta)] \\
&= \frac{1}{\sqrt{n}}\sum_{t=1} \varepsilon_t [I_t(\theta) - F(\infty,\theta)] + o(1) \quad \text{a.s.} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1} \varepsilon_t w_t(\theta) + o(1) \quad \text{a.s.};
\end{aligned}
$$

where in the third equality we have used a Glivenko-Cantelli type Strong Law of Large Numbers for stationary ergodic processes. The main idea is to estimate the distribution of $\sqrt{n}R_n(\theta)$ by the distribution of

$$
\sqrt{n}R_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1} \hat{\varepsilon}_t \hat{w}_t(\theta) W_t;
$$

where $\hat{\varepsilon}_t = (y_t - \bar{y})$; $\hat{w}_t(\theta) = I_t(\theta) - F_n(\infty,\theta)$ and W_t is a sequence of independent random variables with zero mean, unit variance and bounded support. This procedure has been called a wild bootstrap (see Wu (1986) or Mammen (1993)).

Next, we justify the bootstrap test procedure by providing a Theorem that establishes the consistency of the bootstrapped process $\sqrt{n}R_n^*(\theta)$. This means that asymptotically the probability law of $\sqrt{n}R_n^*(\theta)$ given the data X_n is the null asymptotic distribution of $\sqrt{n}R_n(\theta)$.

Theorem 4. Under either the null hypothesis (1) or under the alternative hypothesis (2) or under the sequence of alternative hypotheses (13),

$$\sqrt{n}R_n^*(\cdot) \Rightarrow_* B(\cdot) \quad a.s.,$$

where $\Rightarrow_*$ a.s. denotes weak convergence almost surely under the bootstrap law, that is,

$$P(\sqrt{n}R_n^*(\cdot) \leq s \mid X_n) \to_{a.s.} P(\sqrt{n}R_n(\cdot) \leq s) \text{ as } n \to \infty.$$

Therefore, the asymptotic distribution of $\sqrt{n}R_n(\cdot)$ can be estimated with that of $\sqrt{n}R_n^*(\cdot)$. Hence, defining $T_{\phi,n}^* = \varphi(\sqrt{n}R_n^*(\cdot))$, the asymptotic distribution of $T_{\phi,n}$ can be estimated with that of $T_{\phi,n}^*$ that is given by $\varphi(B(\cdot))$ as the following corollary (that is a straightforward application of the Continuous Mapping Theorem) shows.

Corollary 3. Under (1) or (2) or (13), $T_{\phi,n}^* \Rightarrow_* \varphi(B(\cdot))$ a.s..

Corollaries 2 and 3 justify that the asymptotic critical values of $T_{\phi,n}$ can be estimated with those of $T_{\phi,n}^*$. In practice, the critical values of $T_{\phi,n}^*$ are approximated by simulations. Hence, the proposed general bootstrap test consists in the following steps:

a) Calculate the test statistic $T_{\phi,n}$.

b) Generate $\{W_t\}$ a sequence of n bounded independent random variables with zero mean and unit variance. This sequence is serially independent and is also independent of the original sample X_n.

c) Compute $\sqrt{n}R_n^*(\cdot) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} b_t(\cdot)W_t$. Then compute $T_{\phi,n}^* = \varphi(\sqrt{n}R_n^*(\cdot))$.

d) Repeat steps b) and c) B times where in step b) each sequence $\{W_t\}$ is independent of each other. This produces a set of B independent (conditionally in the sample) values of $T_{\phi,n}^*$ that share the asymptotic distribution of $T_{\phi,n}$.

e) Let $T_{\phi[\alpha]}^*$ be the $\alpha-$quantile of the empirical distribution of the B values of $T_{\phi,n}^*$. The proposed test rejects the null hypothesis if $T_{\phi,n} > T_{\phi[\alpha]}^*$.

Corollaries 2 and 3 establish that under the null hypothesis (1) $T_{\phi,n}$ and $T_{\phi,n}^*$ share the same asymptotic distribution for almost all samples. Hence, the rejection probability (RP) of the bootstrap test converges to α (the theoretical level). Besides, since under the alternative hypothesis (2) $T_{\phi,n}$ diverges while $T_{\phi,n}^*$ remains bounded, the RP under (2) converges to 1. Formally,

$$P(T_{\phi,n} > T_{\phi[\alpha]}^*) \to \begin{cases} \alpha & \text{under (1);} \\ 1 & \text{under (2);} \\ C & \text{under (13)} \end{cases}$$

where $\alpha < C < 1$. Hence, the proposed bootstrap test has an α asymptotic level, it is consistent and it is able to detect alternatives tending to the null al the $n^{-1/2}$ rate.

5 Finite Sample Performance

In this section we examine the finite sample performance of the Cramer-von Mises and the Kolmogorov-Smirnov test (for simplicity, C_p and K_p, respectively) for the case in which no other explanatory variables are considered. Hence, the considered null hypothesis is (10) and the alternative is (11). We employ two Data Generating Processes under the null hypothesis and several uncorrelated and correlated processes under the alternative.

The two MDS uncorrelated processes are a sequence of independent and identically distributed (i.i.d.) N(0,1) variates and a GARCH(1,1) process, that is, $y_t = \varepsilon_t \sigma_t$ where $\sigma_t^2 = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, and $\{\varepsilon_t\}$ is a sequence of i.i.d. N(0, 1) random variables. We have chosen three specifications. We employ $w = 0.001$ and the following combinations for $(\alpha; \beta) : (0.01; 0.97)$; $(0.09; 0.89)$; and $(0.09; 0.90)$. These cases were employed in Lobato, Nankervis and Savin (1999) to compare the finite sample properties of the Box-Pierce Q statistic and Q^*, a modified Q statistic. Notice that the second and third GARCH models have unbounded eight and sixth moment, respectively.

The non-MDS processes are a nonlinear moving average (NLMA) process, a chaotic process and a bilinear process. The NLMA process is given by $y_t = \varepsilon_{t-1}\varepsilon_{t-2}(\varepsilon_{t-2} + \varepsilon_t + 1)$ where $\{\varepsilon_t\}$ is as above. The chaotic process is given by $y_t = 4z_{t-1}(1 - z_{t-1})$ with z_0 distributed as a uniform in [0,1]. The bilinear process is given by $y_t = \varepsilon_t + b_1\varepsilon_{t-1}y_{t-1} + b_2\varepsilon_{t-1}y_{t-2}$ where $\{\varepsilon_t\}$ is as above. Two combinations for $(b_1; b_2)$ were chosen, $(b_1; b_2) = (0.15; 0.05)$ and $(0.25; 0.15)$. Notice that usual test procedures for uncorrelatedness, such as Box and Pierce's Q, or Q^* or the statistics proposed by Robinson (1991), Durlauf (1991), Anderson (1993) or Hong (1996) have asymptotic no power against the NLMA model or the chaotic process.

We consider three values for $p = 1; 2$ and 3; one sample size $n = 100$ for the experiments under the null hypothesis and three sample sizes $n = 100; 500$ and 1000 for the experiments under the alternative. Notice that different values for p correspond to different null hypotheses, and hence, the number p cannot be seen as a smoothing number. In all replications 200 pre-sample data values were generated and discarded. The number of Monte Carlo experiments is 3000 and the number of bootstrap replications is $B = 500$. Random numbers were generated using the IMSL ggnml subroutine. Computations have been carried out in Fortran 90. The code is available from the authors. In these finite sample exercises, as well as in the empirical application in the next section, we follow Mammen (1993) and Stute, Manteiga and Presedo (1998) and the employed sequence $\{W_t\}$ is an i.i.d sequence of Bernoulli variates W where $P(W = 0.5(1 - \sqrt{5})) = (1 + \sqrt{5})/2\sqrt{5}$ and $P(W = 0.5(1 + \sqrt{5})) = 1 - (1 + \sqrt{5})/2\sqrt{5}$. Notice that the third moment of W is equal to 1, and hence, this selection of $\{W_t\}$ guarantees that the first three moments of the bootstrap series coincide with the first three moments of the original series. In the previous references it was shown that this particular choice of W leads to very accurate finite sample behavior.

In Tables 1 and 2 we report the empirical rejection probabilities (RP's) associated with three nominal levels 10%, 5% and 1%, for experiments under the null and the alternative, respectively. Table 1 shows that for a sample size as small as 100 the empirical RP's under the null are very close to the nominal levels for all DGP's considered. Notice that the finite sample behavior is very similar in all GARCH cases suggesting that the proposed test procedures are quite insensitive to thick tails. Notice also that, in most of the cases (28 out of 36), the K_p test rejects more often than the C_p test.

Table 2 shows that typically we need at least sample sizes of about 500 in order to have reasonable power against a wide range of alternatives. Notice, however, than in some cases, such as the first bilinear process for the $p = 3$ case, $n=500$ is not big enough. Also note that the empirical power always increases with n but decreases with p. In general, it can be expected that no test will dominate others in the sense of having more empirical power for all cases. In our experiments the K_p test has typically more empirical power than the C_p test for the NLMA and the bilinear cases, while the C_p test has more empirical power than the K_p test for the chaotic process. Notice also that in the bilinear examples both tests are comparable for the $p = 1$ case, but as p increases the K_p test has more empirical power than the C_p test.

6 Empirical Application

In this Section we examine whether the daily log price changes of the British pound in terms of the U.S. dollar (BP/USD) follows a martingale difference sequence up to order p with respect to its own past. We consider three values for $p = 1, 2,$ and $3.$ This series has been studied before in Hsieh (1989), Gallant, Hsieh and Tauchen (1991) and Bera and Higgins (1997) among others. For the sample period 1974-1983, Hsieh (1989) and Gallant, Hsieh and Tauchen (1991) found that GARCH models were not satisfactory. On the contrary, Bera and Higgins preferred a GARCH model rather than a bilinear model for the period 1985-1991. Recently, Brooks and Hinich (1999) have reported evidence (based on bicorrelations) against the MDS property of exchange rate returns.

Results for the Cramer-von Mises and the Kolmogorov-Smirnov tests are reported in Table 3 for both periods. Notice that, in order to facilitate interpretation, p-values are reported. The number of bootstrap replications, B, is 500.

In the first column of Table 3 we report the results for the period January 2nd, 1974 to December 31st, 1993. For $p = 1$; there is strong evidence against the MDS hypothesis what agrees with the results found by Hsieh (1989) and Gallant, Hsieh and Tauchen (1991). Notice, however, that for $p = 2$ the Cramer-von Mises test does not reject the MDS hypothesis (although the Kolmogorov-Smirnov rejects), while both test do not reject for $p = 3$, indicating that a sensible model for this data should be a MDS process of order 3 with respect to its past and not necessarily a MDS process of order 1 with respect to its past.

In the second column of Table 3 we consider the data from December 12th, 1985 to February 28th, 1991. For this sample, Bera and Higgins considered two alternative models: a GARCH(1,1) model (an example of a MDS) and a bilinear model (an example of a non-MDS). Bera and Higgins computed formal tests to discriminate between both models such as Cox (1961) and Vuong (1989) tests and also compared the two models using some measures of out-of-sample predictive ability. They found that the general evidence favored the GARCH (1,1) model in detriment of the bilinear model. The results in the second column of Table 3 agree with this result: for the sample period 1985-1991, the null hypotheses that the process is a MDS of order p cannot be rejected for any of the considered values of p.

7 Conclusions and Further Research

In this paper we have analyzed consistent tests for the MDS assumption. Contrary to the commonly employed tests, the proposed tests are able to detect failures of the MDS assumption for uncorrelated processes. In fact, the proposed tests are consistent, that is, whenever a process does not follow a martingale difference of orders P, the tests will have asymptotic unit power. Since the asymptotic distribution of the test statistics are not standard and, in fact, they depend on the specific data generating process, we could either transform the test statistics to find ones whose asymptotic distributions were pivotal or use the bootstrap to estimate the asymptotic distributions. The transformation proposed by Koul and Stute (1999) is not valid for our case; alternative transformations, such as the one proposed by Ming (1999) present several problems such as requiring conditional homoskedasticity or demanding the selection of a user-chosen smoothing number. Hence, we have proposed (and justified theoretically) to implement the test using a modification of the wild bootstrap procedure. We have also shown that the proposed test is very simple to use in practice and performs remarkably well in finite samples. Finally, we have applied the proposed tests to the British pound vs. the U.S. dollar exchange rate for two different periods and found, in general, evidence in favor of the MDS hypotheses. Of course, more exhaustive studies for this and for other currencies are needed.

We finish this section with some suggestions on further research. First, in this paper we have considered the case of testing that a process is a martingale difference sequence of orders p. However, the martingale hypothesis is typically stated involving an infinite number of lags. Analyzing this case is a challenging problem. De Jong (1996) presents a consistent test (that belongs to the first type of integrated tests described in Section 2) for this hypothesis. His Monte Carlo results indicate that his test has very low power except for extreme cases. In fact, we have applied his test to the examples in Section 5 and we have found that his test has no power for the considered cases. Our tests could be extended to cover the $p = \infty$ case but evidence in Table 2 suggests that the test may also present a finite sample power problem.

Second, in this paper we have employed the wild bootstrap, but alternative bootstrap procedures, such as the naive bootstrap or some blocking bootstrap, could have been employed. For instance, in the simplest case where the information set only contains lagged values of the relevant process, the naive bootstrap is based on resampling with replacement from $\{r_t\} = \{(y_t; y_{t-1}; \ldots; y_{t-1;p})'\}$ for $t = 1; \ldots; n$, to obtain $\{r_t^* = (r_{t;1}^*; \ldots; r_{t;p+1}^*)'$; $t = 1; \ldots; n\}$, so that the test statistics are based on

$$\sqrt{n}R_n^*(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1} \xi_t^*(\theta) - \sqrt{n}R_n(\theta)$$

where $\xi_t^*(\theta) = (r_{t;1}^* - \overline{r^*})I(r_{t;2}^* \leq \tau_1; \ldots; r_{t;p+1}^* \leq \tau_p)$ and $\overline{r^*} = n^{-1}\sum_{t=1}^{n} r_{t;1}^*$. Another alternative bootstrap procedure is to generate the bootstrap series using some blocking bootstrap scheme, such as moving blocks bootstrap.

Third, in this paper we have considered testing that the conditional mean is constant, but the more general null hypothesis

$$E[\psi(y_t; \theta_{t;p}|\theta_{t;p})] = 0$$

where ψ is a given function, could be tested using similar procedures to the ones considered here (for instance, testing for conditional homoskedasticity).

Tables

Table 1

Percentage of rejections at nominal 10%, 5% and 1% levels. The …rst DGP is an i.i.d N(0,1) sequence. The others are GARCH(1,1) processes. The sample size is 100. The number of replications is 3000. The number of bootstrap replications is 500:

		IID		GARCH1		GARCH2		GARCH3	
p		C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p
	10%	9.58	9.98	10.4	10.2	10.4	10.4	10.6	10.4
1	5%	4.74	5.08	4.80	5.70	5.17	5.83	5.08	5.50
	1%	1.00	1.26	1.13	1.03	1.23	1.27	1.22	1.36
	10%	10.1	10.3	11.2	12.0	11.2	11.3	11.1	11.4
2	5%	4.68	5.18	6.17	5.93	5.97	5.90	5.76	6.16
	1%	1.34	1.12	1.37	1.33	1.37	1.60	1.20	1.32
	10%	9.30	10.3	11.2	11.3	10.7	10.9	10.4	10.6
3	5%	4.62	5.22	5.60	6.10	5.33	5.83	5.04	5.26
	1%	0.88	1.02	1.00	1.27	0.90	1.27	0.76	1.02

Table 2

Percentage of rejections at nominal 10%, 5% and 1% levels. The first DGP is a non-linear moving average model, $y_t = \varepsilon_{t-1}\varepsilon_{t-2}(\varepsilon_{t-2} + \varepsilon_t + 1)$ where $\{\varepsilon_t\}$ is an i.i.d N(0,1) sequence. The second DGP is a chaotic process given by $y_t = 4z_{t-1}(1 - z_{t-1})$ with z_0 distributed as a uniform in [0,1]. The sample sizes are 100, 500 and 1000. The number of replications is 3000. The number of bootstrap replications is 500:

		NLMA						Chaotic					
p/n		100		500		1000		100		500		1000	
		C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p
	10%	27.8	27.9	68.3	73.5	91.5	93.1	100	100	100	100	100	100
1	5%	16.8	16.7	53.5	60.0	83.9	86.7	100	100	100	100	100	100
	1%	4.60	4.50	28.8	35.9	53.1	64.7	100	100	100	100	100	100
	10%	20.0	19.7	53.0	61.9	76.1	87.3	96.5	90.2	100	100	100	100
2	5%	10.6	10.6	36.5	48.0	62.8	77.9	84.6	75.3	100	100	100	100
	1%	2.70	2.90	15.7	25.1	30.0	52.5	46.2	42.2	100	100	100	100
	10%	13.9	17.8	39.6	55.1	64.6	83.9	54.3	40.0	100	100	100	100
3	5%	6.30	9.67	24.4	40.7	46.1	74.3	35.8	24.8	100	99.9	100	100
	1%	1.23	2.10	7.43	19.6	15.5	46.8	13.1	8.67	96.1	89.4	100	100

Table 2 (continued)

Percentage of rejections at nominal 10%, 5% and 1% levels. The first DGP is a bilinear model, $y_t = \varepsilon_t + 0{:}15\varepsilon_{t-1}y_{t-1} + 0{:}05\varepsilon_{t-1}y_{t-2}$, where $\{\varepsilon_t\}$ is as above. The second DGP is a bilinear model, $y_t = \varepsilon_t + 0{:}25\varepsilon_{t-1}y_{t-1} + 0{:}15\varepsilon_{t-1}y_{t-2}$, where $\{\varepsilon_t\}$ is as above. The sample sizes are 100, 500 and 1000. The number of replications is 3000. The number of bootstrap replications is 500:

		Bilinear 1						Bilinear 2					
p/n		100		500		1000		100		500		1000	
		C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p	C_p	K_p
	10%	18.0	18.7	72.2	66.2	97.4	94.7	40.9	42.4	99.5	99.1	100	100
1	5%	8.73	10.8	50.6	47.8	92.9	86.3	23.8	27.7	98.1	96.4	100	100
	1%	1.97	2.62	18.0	20.8	56.0	52.9	7.11	10.5	82.4	82.0	99.5	99.6
	10%	13.2	13.3	38.9	41.5	76.0	77.1	20.6	25.3	88.6	90.8	100	100
2	5%	6.93	6.87	20.8	26.4	57.3	62.5	10.7	15.3	70.7	81.9	99.3	99.9
	1%	1.60	1.80	4.63	9.97	14.9	29.2	2.77	4.92	26.5	56.9	78.1	95.5
	10%	11.0	13.0	26.9	32.1	55.6	65.1	17.3	21.1	73.0	82.1	98.7	99.3
3	5%	5.23	7.00	13.8	21.1	35.0	50.1	8.72	12.5	50.2	70.8	93.5	98.5
	1%	1.23	1.83	3.33	8.20	7.27	21.0	2.03	3.76	16.0	47.6	54.5	90.2

Table 3

p-values for the Cramer-von Mises (C_p) and the Kolmogorov-Smirnov (K_p) tests for daily returns of the exchange rate of the British pound vs. the U.S. dollar. The …rst sample period is from January 2nd, 1974 to December 31st, 1983. The second sample period covers from December 12th, 1985 to February 28th, 1991. The number of bootstrap replications is 500:

	74-83	85-91
n	2557	1311
C_1	0.014	0.322
K_1	0.006	0.202
C_2	0.218	0.788
K_2	0.024	0.508
C_3	0.594	0.842
K_3	0.282	0.436

Appendix

Proof of Theorem 1

We need to show that the finite dimensional distributions of the process $\sqrt{n}R_n(\gamma)$ are asymptotically normal with the appropriate covariance matrix and that the process $\sqrt{n}R_n(\gamma)$ is tight. Both conditions hold in this multidimensional context using procedures similar to those in Koul and Stute (1999). In this appendix K denotes some generic positive finite constant.

Convergence of finite-dimensional distributions refers to the weak convergence of vectors of the form $(\sqrt{n}R_n(\gamma_1), \sqrt{n}R_n(\gamma_2), \ldots, \sqrt{n}R_n(\gamma_k))$; for arbitrary $k \in \mathbb{N}$ and $\tau_i \in \mathbb{R}^P$; $i = 1, 2, \ldots, k$. This result can be obtained using the Corollary 3.1 in Hall and Heyde (1980).

In order to prove tightness, some definitions are required. Let $\left\{ W_n(\gamma) : \gamma \in \mathbb{R}^P;\ n = 1, 2, \ldots \right\}$ be a sequence of stochastic processes on some set D. Then, $\left\{ W_n(\gamma) \right\}$ is tight if and only if for any $\delta > 0$ there exists a compact set $K \subset D$ such that

$$\sup_n \Pr\left(W_n(\gamma) \in K \right) > 1 - \delta. \tag{14}$$

Let $D_1 = (\mathbf{s}^1, \mathbf{t}^1] = \times_{k=1}^P (s_k^1, t_k^1]$; and $D_2 = (\mathbf{s}^2, \mathbf{t}^2] = \times_{k=1}^P (s_k^2, t_k^2]$ be two intervals of $\mathbb{R}^P$. Then, D_1 and D_2 are neighbor intervals if and only if for some $j \in \{1, 2, \ldots, P\}$, $(s_j^1, t_j^1] \neq (s_j^2, t_j^2]$ and $\times_{k \neq j}(s_k^1, t_k^1] = \times_{k \neq j}(s_k^2, t_k^2]$; that is, if and only if they are next to each other and share the j-th face. The stochastic process indexed by the intervals is defined as

$$W_n(D_j) = \sum_{e_1=0} \cdots \sum_{e_P=0} (-1)^{P - \sum_j e_j} W_n\left(s_1^j + e_1(t_1^j - s_1^j), \cdots, s_P^j + e_P(t_P^j - s_P^j) \right).$$

In this proof we verify Chentsov´s criterion that is a sufficient condition for (14), see Billingsley (1968) and Koul and Stute (1999).

In our case,

$$\sqrt{n}R_n(\gamma) = \frac{1}{\sqrt{n}} \sum_{t=1} \varepsilon_t I(\mathbf{e}_{t,P} \le \gamma) - F(\infty, \gamma) \frac{1}{\sqrt{n}} \sum_{t=1} \varepsilon_t + o(1); \quad a.s..$$

The second term is tight since $F(\infty, \gamma)$ is continuous. The first term can be written as the following process indexed by the intervals

$$\sqrt{n}R_n(D_j) = \frac{1}{\sqrt{n}} \sum_{t=1} \left[\varepsilon_t I_t(D_j) \right];$$

where $I_t(D_j) = I\left(\tilde{e}_{t,p} \in D_j\right)$. For instance, in the $p = 2$ case, $I_t(D_j) = I_t(t_1^j, t_2^j) - I_t(s_1^j, t_2^j) - I_t(t_1^j, s_2^j) + I_t(s_1^j, s_2^j)$. Then

$$\begin{aligned} & E\left(\left[\sqrt{n}R_n(D_1)\right]^2\left[\sqrt{n}R_n(D_2)\right]^2\right) \\ = {} & \frac{1}{n^2}E\left\{\sum_{t=1}\sum_{s=1}\sum_{u=1}\sum_{v=1}[\varepsilon_t I_t(D_1)][\varepsilon_s I_s(D_1)][\varepsilon_u I_u(D_2)][\varepsilon_v I_v(D_2)]\right\}. \end{aligned}$$

Using that ε_t is a centered MDS, the non-zero terms are those such that the greater subindex appears at least twice. Moreover, notice that when a subindex appears three times, the corresponding term is zero using that D_1 and D_2 are disjoint sets. Therefore,

$$\begin{aligned} E\left(\left[\sqrt{n}R_n(D_1)\right]^2\left[\sqrt{n}R_n(D_2)\right]^2\right) &= \frac{1}{n^2}E\left\{\sum_{t=1}\left[\varepsilon_t^2 I_t(D_1)\right]\left(\sum_{s=1}^{t-1}[\varepsilon_s I_s(D_2)]\right)^2\right\} \\ &+ \frac{1}{n^2}E\left\{\sum_{t=1}\left[\varepsilon_t^2 I_t(D_2)\right]\left(\sum_{s=1}^{t-1}[\varepsilon_s I_s(D_1)]\right)^2\right\}. \end{aligned}$$

Under the assumptions of the Theorem, these expectations exist. Note that both terms are analyzed similarly since the only difference is the index set D_j. Using that $\left(\sum_{i=1}^{l} a_i\right)^2 \le 2l\sum_{i=1}^{l} a_i^2$; the first term is bounded by

$$\frac{4P}{n^2}\sum_{s=1} E\left\{\sum_{t=1}\left[\varepsilon_t^2 I_t(D_1)\right]\left[\varepsilon_{t-s}^2 I_{t-s}(D_2)\right]\right\} \tag{15}$$

$$+\frac{2}{n^2}E\left\{\sum_{t=1}\left[\varepsilon_t^2 I_t(D_1)\right]\left(\sum_{s=1}^{t-P-1}[\varepsilon_s I_s(D_2)]\right)^2\right\}. \tag{16}$$

First, consider any term in (15). For any $s = 1, \dots, P$, using the law of iterated expectations, and defining $\Omega_{t-1} = \lim_{p\to\infty}\tilde{\Omega}_{t-1,p}$

$$\begin{aligned} E\left\{\left[\varepsilon_t^2 I_t(D_1)\right]\left[\varepsilon_{t-s}^2 I_{t-s}(D_2)\right]\right\} &= E\left\{\sigma^2(\Omega_{t-1}) I_t(D_1)\left[\varepsilon_{t-s}^2 I_{t-s}(D_2)\right]\right\} \\ = {} & E\left\{E\left[\sigma^2(\Omega_{t-1}) I_t(D_1)\varepsilon_{t-s}^2 I_{t-s}(D_2) \mid \Omega_{t-s}\right]\right\}. \end{aligned} \tag{17}$$

Note that $I_t(D_1)$ depends on two types of variables, namely $\tilde{e}_{t,p}^{(1)} = \{y_{t-1}, \dots, y_{t-s}, x_{1,t}, x_{1,t-1}, \dots, x_{1,t-s+1}, \dots, x_{K,t}, x_{K,t-1}, \dots, x_{K,t-s+1}\}$ and $\tilde{e}_{t,p}^{(2)} = \{y_{t-s-1}, \dots, y_{t-p}, x_{1,t-s}, \dots, x_{1,t-p_1}, \dots, x_{K,t-s}, \dots, x_{K,t-p_K}\}$. Notice that $\tilde{e}_{t,p}^{(2)}$ is Ω_{t-s} measurable while $\tilde{e}_{t,p}^{(1)}$ is affected by the integration of the inside conditional expectation. Let $f_s(\mathbf{e} \mid \Omega_{t-s})$ be the density of $\tilde{e}_{t,p}^{(1)}$ conditional on Ω_{t-s}. Now, arrange the interval D_1 in some way according with the decomposition of

$\mathbf{e}_{t,p}$ into $\mathbf{e}^{(1)}_{t,p}$ and $\mathbf{e}^{(2)}_{t,p}$, and call $D_1^{(1)}$ and $D_1^{(2)}$ to those subsets. Then, equation (17) can be rewritten as

$$= E\left\{\int \sigma^2(\mathbf{e};\theta_{t-s})\, I(\mathbf{e}\in D_1^{(1)};\ \mathbf{e}^{(2)}_{t,p}\in D_1^{(2)})(e_s-\mu)^2 f(\mathbf{e}\mid\theta_{t-s})\, I_{t-s}(D_2)\, d\mathbf{e}\right\}$$

where e_s $(=y_{t-s})$ is the s-th coordinate of $\mathbf{e}$. Using Fubini´s theorem and Hölder´s inequality, the last expression is bounded by

$$\int_{D_1^{(1)}} E\left(\sigma^2(\mathbf{e};\theta_{t-s})\, I(\mathbf{e}^{(2)}_{t,p}\in D_1^{(2)})(e_s-\mu)^2 f(\mathbf{e}\mid\theta_{t-s})\, I_{t-s}(D_2)\right) d\mathbf{e}$$

$$\le \int_{D_1^{(1)}}\left(E\left[\sigma^2(\mathbf{e};\theta_{t-s})\, I(\mathbf{e}^{(2)}_{t,p}\in D_1^{(2)})(e_s-\mu)^2 f(\mathbf{e}\mid\theta_{t-s})\right]^{1+\delta}\right)^{1/(1+\delta)} d\mathbf{e}\cdot(E\, I_{t-s}(D_2))^{\delta/(1+\delta)}$$

$$\le \mu_{1,s}(D_1\cup D_2)\,[\mu_2(D_1\cup D_2)]^{\delta/(1+\delta)}$$

with $0<\delta<1$,

$$\mu_{1,s}(D_1\cup D_2)=\int_{D_1^{(1)}\cup D_2^{(1)}}\left(E\left[\sigma^2(\mathbf{e};\theta_{t-s})\, I(\mathbf{e}^{(2)}_{t,p}\in D_1^{(2)})(e_s-\mu)^2 f(\mathbf{e}\mid\theta_{t-s})\right]^{1+\delta}\right)^{1/(1+\delta)} d\mathbf{e}$$

and

$$\mu_2(D_1\cup D_2)=E\, I_{t-s}(D_1\cup D_2)\,.$$

Second, consider (16). Applying the Law of Iterated expectation, for any t;

$$E\left\{\left[\varepsilon_t^2 I_t(D_1)\right]\left(\sum_{s=1}^{t-P-1}[\varepsilon_s I_s(D_2)]\right)^2\right\} = E\left\{\sigma^2(\theta_{t-1})\, I_t(D_1)\left(\sum_{s=1}^{t-P-1}[\varepsilon_s I_s(D_2)]\right)^2\right\}$$

$$= E\left\{\int_{D_1}\sigma^2(\mathbf{e};\theta_{t-P-1})\ f(\mathbf{e}\mid\theta_{t-P-1})\, d\mathbf{e}\left(\sum_{s=1}^{t-P-1}[\varepsilon_s I_s(D_2)]\right)^2\right\}.$$

Using Fubini's theorem and Hölder's inequality, the last expression is bounded by

$$\int_{D_1}\left\{E\left[\sigma^2(\theta_{t-1})\right]^2 f^2\left(\mathbf{e}\mid\theta_{t-P-1,p}\right)\right\}^{1/2} d\mathbf{e}\left[E\left(\sum_{s=1}^{t-P-1}[\varepsilon_s I_s(D_2)]\right)^4\right]^{1/2}. \qquad (18)$$

Now, notice that the integral is bounded by $\mu_3(D_1\cup D_2)$; where

$$\mu_3(D_1\cup D_2)=\int_{D_1\cup D_2}\left\{E\left[\sigma^2(\mathbf{e};\theta_{t-P-1})\right]^2 f^2\left(\mathbf{e}\mid\theta_{t-P-1,p}\right)\right\}^{1/2} d\mathbf{e}.$$

In addition, using that $\varepsilon_s I_s(D_2)$ is a MDS, and applying Burkholder's inequality, the expression in brackets in (18) is bounded by

$$K E\left(\sum_{s=1}^{t-P-1}\left[\varepsilon_s^2 I_s(D_2)\right]\right)^2 \le K(t-P-1)^2 E\left[\varepsilon_2^4 I_1(D_2)\right]\cdot$$

$$K\,(t - P - 1)^2\,\mu_4(D_1 \cup D_2)\,;$$

where $\mu_4(D_1 \cup D_2) = E\left[\varepsilon_2^4 I_1(D_1 \cup D_2)\right]$: Hence,

$$E\left(\left(\sqrt{n}R_n(D_1)\right)^2\left(\sqrt{n}R_n(D_2)\right)^2\right)$$

$$\frac{2}{n^2}\left(n\sum_{s=1}\mu_{1;s}(D_1 \cup D_2)\left[\mu_2(D_1 \cup D_2)\right]^{\delta=1+\delta} + \sum_{t=1}(t - P - 1)\left[\mu_3(D_1 \cup D_2)\right]\left(\mu_4(D_1 \cup D_2)\right)^{1=2}\right)$$

$$\frac{2nP}{n^2}\mu_1(D_1 \cup D_2)\left[\mu_2(D_1 \cup D_2)\right]^{\delta=1+\delta} + \frac{2}{n^2}\sum_{t=1} n\left[\mu_3(D_1 \cup D_2)\right]\left(\mu_4(D_1 \cup D_2)\right)^{1=2}$$

$$K\left[\mu_1(D_1 \cup D_2)\left[\mu_2(D_1 \cup D_2)\right]^{\delta=1+\delta} + \left[\mu_3(D_1 \cup D_2)\right]\left(\mu_4(D_1 \cup D_2)\right)^{1=2}\right]; \qquad (19)$$

where $\mu_1(D_1 \cup D_2) = \sum_{s=1}^{P}\mu_{1;s}(D_1 \cup D_2)$: Equation (19) is a Chentsov´s inequality in the multidimensional case (see Gaenssler and Stute, 1979, p 215) and the proof of tightness is finished.

Proof of Theorem 2

Using a Uniform Strong Law of Large numbers for stationary ergodic sequences as in Koul and Stute (1999)

$$R_n(\theta) = E\left(\varepsilon_t\left[I_t(\theta)\right]\right) + O\left(n^{-1=2}\right); \quad a{:}s{:}{:}$$

Under H_1; there exists a $T \subset R^P$ such that $E\left(\varepsilon_t\left[I_t(\theta)\right]\right) \neq 0$ for $\theta \in T$ with $\Pr\left(\theta_{t;\theta} \in T\right) > 0$: Therefore, for $\theta \in T$; $R_n(\theta) \Rightarrow R(\theta)$, and, hence, $\sqrt{n}R_n(\theta)$ diverges to infinity almost surely.

Proof of Theorem 3

$$\sqrt{n}R_n(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}\left(y_t - \mu - \frac{g\left(\theta_{t;\theta}\right)}{\sqrt{n}}\right)I_t(\theta) - \frac{1}{\sqrt{n}}\sum_{t=1}\left(\bar{y} - \mu - \frac{g\left(\theta_{t;\theta}\right)}{\sqrt{n}}\right)I_t(\theta)$$

and defining $\xi_t = y_t - \mu - n^{-1=2}g\left(\theta_{t;\theta}\right)$;

$$\begin{aligned}
\sqrt{n}R_n(\theta) &= \frac{1}{\sqrt{n}}\sum_{t=1}\xi_t I_t(\theta) - \frac{1}{\sqrt{n}}\sum_{t=1}\left(\left(\frac{1}{n}\sum_{s=1}y_s\right) - \mu - \frac{g\left(\theta_{t;\theta}\right)}{\sqrt{n}}\right)I_t(\theta) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}\xi_t I_t(\theta) + \frac{1}{n}\sum_{t=1}\left(g\left(\theta_{t;\theta}\right)I_t(\theta)\right. \\
&\quad - \frac{1}{n}\sum_{t=1}I_t(\theta)\frac{1}{\sqrt{n}}\sum_{s=1}\xi_s - \frac{1}{n}\sum_{t=1}I_t(\theta)\frac{1}{n}\sum_{s=1}\left.g\left(\theta_{s;\theta}\right)\right) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}\xi_t \hat{w}_t(\theta) + \frac{1}{n}\sum_{t=1}g\left(\theta_{t;\theta}\right)\hat{w}_t(\theta)
\end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1} \psi_t w_t(\theta) + \frac{1}{n} \sum_{t=1} g\left(\varepsilon_{t,\beta}\right) w_t(\theta) + o(1) \text{ a.s..}$$

Apply the functional Central Limit Theorem and the Strong Law of Large Numbers for ergodic and stationary processes as in Koul and Stute (1999) to the first and second term of the last expression respectively and the result follows.

Proof of Theorem 4

We need to show that the finite dimensional distributions of the process $\sqrt{n}R_n^*(\theta)$ are asymptotically normal and that the process $\sqrt{n}R_n^*(\theta)$ is tight, conditionally on the sample.

First, define for any $k \in N$ and any $\lambda_j \in R^k$ such that $\|\lambda\| = 1$, $L_n^* = \sum_{j=1}^k \lambda_j \sqrt{n}R_n^*(\theta_j) = n^{-1/2} \sum_{t=1}^n W_t \hat{e}_t \hat{w}_t$ where $\hat{w}_t = \hat{w}_t(\theta_1, \ldots, \theta_k; \lambda) = \sum_{j=1}^k \lambda_j \hat{w}_t(\theta_j)$. Also call $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n \hat{e}_t^2 \hat{w}_t^2$ and $\sigma^2 = E(\hat{\sigma}^2)$. Then rewrite L_n^* as $L_n^* = \sum_{t=1}^n W_t (\sqrt{n}\hat{\sigma})^{-1} \hat{e}_t \hat{w}_t \hat{\sigma} = l^* \hat{\sigma}$ where $l^* = \sum_{t=1}^n \xi_{nt}^*$ with $\xi_{nt}^* = W_t (\sqrt{n}\hat{\sigma})^{-1} \hat{e}_t \hat{w}_t$. Now, using standard bootstrap notation, call E^* and V^* to the expectation and the variance taken given the sample. Then, $E^*(l^*) = \sum_{t=1}^n (\sqrt{n}\hat{\sigma})^{-1} \hat{e}_t \hat{w}_t E(W_t) = 0$, while $V^*(l^*) = \sum_{t=1}^n (\sqrt{n}\hat{\sigma})^{-2} (\hat{e}_t \hat{w}_t)^2 V(W_t) = 1$. In addition, ξ_{nt}^* and ξ_{ns}^* are independent conditionally on the sample X_n, since W_t is independent of W_s for $t \neq s$. Finally, using that W_t, $\hat{w}_t$ and $\hat{\sigma}$ are bounded and $\hat{\sigma}$ is bounded away from zero almost surely,

$$\sum_{t=1} E^*(\xi_{nt}^{*2} I(|\xi_{nt}^*| > \delta) \leq \frac{K}{n} \sum_{t=1} \hat{e}_t^2 I(|\hat{e}_t| > \delta' \sqrt{n}) \text{ a.s.}$$

for some positive constants δ and δ'. This last expression converges almost surely to zero as in Stute, Manteiga and Presedo (1998). Hence, the triangular array $\{\xi_{nt}^*\}$ satisfies the conditions of the central limit theorem of Lindeberg-Feller, conditionally on almost all samples, so that $l^* \Rightarrow_* N(0;1)$ a.s., and consequently, using a Strong Law of Large numbers for $\hat{\sigma}^2$, $L_n^* \Rightarrow_* N(0; \sigma^2)$ a.s..

Second, we prove that under either the null hypothesis (1) or under the alternative hypothesis (2) or under the sequence of alternative hypotheses (13), $\sqrt{n}R_n^*(\theta)$ is tight in $D[R]^P$. In this case, we can express the process indexed by the intervals as

$$\sqrt{n}R_n^*(D_j) = \frac{1}{\sqrt{n}} \sum_{t=1} [\hat{e}_t w_t(D_j)] W_t$$

where, for $t = 1, \ldots, n$, and for $j = 1, 2$, we define $w_t(D_j) = I\left(\varepsilon_{t,\beta} \in D_j\right) - \Pr_n\left(\varepsilon_{t,\beta} \in D_j\right)$, and $\Pr_n\left(\varepsilon_{t,\beta} \in D_j\right) = n^{-1} \# \{\varepsilon_{t,\beta} \in D_j\}$. For instance, for the $p = 2$ case, $w_t(D_j) = \left(\hat{w}_t(t_1^j, t_2^j) - \hat{w}_t(s_1^j, t_2^j) - \hat{w}_t(t_1^j, s_2^j) + \hat{w}_t(s_1^j, s_2^j)\right)$. Then

$$E^* \left(\left(\sqrt{n}R_n^*(D_1)\right)^2 \left(\sqrt{n}R_n^*(D_2)\right)^2 \right) \tag{20}$$

$$= \frac{1}{n^2}\sum_{t=1}\sum_{s=1}\sum_{u=1}\sum_{v=1}[\hat{b}_t w_t(D_1)][\hat{b}_s w_s(D_1)][\hat{b}_u w_u(D_2)][\hat{b}_v w_v(D_2)]\,E^*(W_tW_sW_uW_v)$$

$$= \frac{1}{n^2}\sum_{t=1}\sum_{s=1}\hat{b}_t^2\hat{b}_s^2E^*\left(W_t^2W_s^2\right)\left(w_t^2(D_1)w_s^2(D_2)+2w_t(D_1)w_t(D_2)w_s(D_1)w_s(D_2)\right)$$

since the expected value of the rest of the terms is zero (notice that $E^*\{W_tW_sW_uW_v\}=0$ for all values of $t;s;u;v$ except when two pairs with the same subindex appear). Furthermore, since $0 \leq w_t^2(D_j) \leq |w_t(D_j)|$, for $j=1;2$, expression (20) is bounded above by

$$\left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2|w_t(D_1)|\right)\left(\frac{1}{n}\sum_{s=1}\hat{b}_s^2|w_s(D_2)|\right)$$

$$+2\left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2|w_t(D_1)||w_t(D_2)|\right)\left(\frac{1}{n}\sum_{s=1}\hat{b}_s^2|w_s(D_1)||w_s(D_2)|\right)$$

$$\leq \left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2\left(I_t(D_1)+\Pr{}_n\left(\varepsilon^*_{t;n}\in D_1\right)\right)\right)\left(\frac{1}{n}\sum_{s=1}\hat{b}_s^2\left(I_s(D_2)+\Pr{}_n\left(\varepsilon^*_{s;n}\in D_2\right)\right)\right)$$

$$+\frac{1}{2}\left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2[|w_t(D_1)|+|w_t(D_2)|]\right)\left(\frac{1}{n}\sum_{s=1}\hat{b}_s^2[|w_s(D_1)|+|w_s(D_2)|]\right)$$

$$\leq \left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2\left(I_t(D_1\cup D_2)+\Pr{}_n\left(\varepsilon^*_{t;n}\in D_1\cup D_2\right)\right)\right)^2$$

$$+2\left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2\left[I_t(D_1\cup D_2)+\Pr{}_n\left(\varepsilon^*_{t;n}\in D_1\cup D_2\right)\right]\right)^2$$

$$= 3\left(\frac{1}{n}\sum_{t=1}\hat{b}_t^2\left[I_t(D_1\cup D_2)+\Pr{}_n\left(\varepsilon^*_{t;n}\in D_1\cup D_2\right)\right]\right)^2$$

$$\xrightarrow{a.s.} 3\left[E\left(\varepsilon_1^2 I_1(D_1\cup D_2)\right)+\Pr\left(\varepsilon^*_{1;n}\in D_1\cup D_2\right)\right]^2 :$$

This is a Chenstov´s inequality in the multidimensional case and the proof of tightness is finished.

REFERENCES

AN, H.-Z. and BING, C. (1991), "A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series", International Statistical Review, 59, 287-307.

ANDERSON, T.W. (1993), "Goodness of …t tests for spectral distributions", Annals of Statistics, 21, 830-847.

ANDREWS, D.W.K.(1997), "A conditional Kolmogorov test", Econometrica, 65, 1097-1128.

ANDREWS, D.W.K. and POLLARD, D. (1994), "An introduction to functional central limit theorems for dependent stochastic processes", International Statistical Review, 62, 119-132.

BIERENS, H. (1984), "Model speci…cation testing of time series regressions", Journal of Econometrics, 26, 323-353.

BIERENS, H. (1990), "A consistent conditional moment test of functional form", Econometrica, 58, 1443-1458.

BIERENS, H. and PLOBERGER, W. (1997), "Asymptotic theory of integrated conditional moment test", Econometrica, 65, 1129-1151.

BILLINGSLEY, P. (1968), Convergence of Probability Measures, Wiley, New York.

BOX, G.E.P. and PIERCE, D. A (1970), "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", Journal of the American Statistical Association, 65, 1509-1526.

BROCKWELL, P. and DAVIES, R.A. (1992), Time series: theory and practice, Springer Verlag, New York.

BROOKS, C. and HINICH, M.J. (1999), "Cross-correlations and cross-bicorrelations in Sterling exchange rates", Journal of Empirical Finance, 6, 385-404.

BRUNK, H. D. (1970), "Estimation for isotonic regression," in Nonparametric Techniques in Statistical Inference, Ed. M.L. Puri, pp. 177-197, Cambridge: Cambridge University Press.

COX, D.R. (1961), "Tests of separate families of hypotheses", in Proceedings of the 4th Berkeley Symposium, Berkeley: University of California Press, 105-123.

DE JONG, R. M. (1996), "The Bierens test under data dependence", Journal of Econometrics, 72, 1-32.

DE JONG, R. M. and BIERENS, H.J. (1994), "On limit behavior of a Chi-Square type test if the number of conditional moments tested approaches in…nity", Econometric Theory, 9, 70-90.

DELGADO, M. (1993), "Testing the equality of nonparametric regression curves", Statistics & Probability Letters, 17, 199-204.

DURLAUF, S. N. (1991), "Spectral based testing of the martingale hypothesis", Journal of Econometrics, 50, 355-376.

GALLANT, A.R., HSIEH, D.A., and TAUCHEN, G. (1991), "On …tting a recalcitrant series: the pound/dollar exchange rate, 1974-1983", in Nonparametric and semiparametric methods in econometrics and statistics, eds W.A. Barnett, J. Powell, and G. Tauchen, U.K: Cambridge University Press, 199-240.

HALL, P. and HEYDE, C.C. (1980), Martingale Limit Theory and its Application, Academic Press, New York.

HALL, R.E. (1978), "Stochastic umplications of the life cycle - permanent income hypothesis: theory and evidence", Journal of Political Economy, 86, 971-987.

HÄRDLE, W. and MAMMEN, E. (1993), "Comparing nonparametric versus parametric regression …ts", Annals of Statistics, 21, 1926-1947.

HONG, Y. and WHITE, H. (1995), "Consistent speci…cation testing via nonparametric series regressions", Econometrica, 63, 1133-1160.

HONG, Y.(1996), "Consistent testing for serial correlation of unknown form", Econometrica, 64, 837-864.

HOROWITZ, J. and SPOKOINY, V. G.(1999), "An adaptive, rate-optimal test of a parametric model against a nonparametric alternative", manuscript, University of Iowa.

HSIEH, D.A. (1989), "Testing for nonlinear dependence in daily foreign exchange rates", Journal of Business, 62, 339-368.

KOUL, H. L. and STUTE, W. (1999), "Nonparametric model checks for time series", Annals of Statistics, 27, 204-236.

LI, Q. (1999), "Consistent model speci…cation tests for time series econometric models", Journal of Econometrics, 92, 101-147

LOBATO, I., NANKERVIS, J., and SAVIN, N.E. (1999), "Testing for autocorrelation using a modi…ed Box-Pierce Q test", forthcoming, International Economic Review.

MAMMEN, E. (1993), "Bootstrap and Wild Bootstrap for High Dimensional Linear Models", Annals of Statistics, 21, 255-285.

MARRON, S.J. (1988), "Automatic Smoothing Parameter Selection: A Survey", Empirical Economics, 13, 187-208.

MING, X. (1999), "One-One Transformation of Multidimensional General Distribution to Multidimensional Uniform Distribution", Econometric Theory, 15, 429-430.

ROBINSON, P. M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression", Journal of Econometrics, 47, 67-84.

ROSENBLATT, M. (1975), "A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and A Test of Independence", Annals of Statistics, 3, 1-14.

SHORACK, G.R. and WELLNER, J.A. (1986), Empirical processes with applications to statistics, Wiley, New York.

STINCHCOMBE, M. and WHITE, H. (1998), "Consistent speci…cation 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., MANTEIGA, W.G. and PRESEDO, M. (1998), "Bootstrap approximations in model checks for regression", Journal of the American Statistical Association, 83, 141-149.

SU, J.Q. and WEI, L.J. (1991), "A lack of …t test for the mean function in a generalized linear model", Journal of the American Statistical Association, 86, 420-426.

VUONG, Q. H. (1989), "Likelihood ratio tests for model selection and non-nested hypothesis", Econometrica, 57, 307-333.

WU, C. F. J. (1986): "Jacknife, bootstrap and other resampling methods in regression analysis (with discussion)", Annals of Statistics, Vol. 14 pp. 1261-1350.

ZHENG, X. (1996), "A consistent test of functional form via nonparametric estimation technique", Journal of Econometrics, 75, 263-289.