Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 02/05

GOODNESS-OF-FIT TESTS FOR LINEAR AND NON-LINEAR TIME SERIES MODELS

J. Carlos Escanciano

Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

GOODNESS-OF-FIT TESTS FOR LINEAR AND NON-LINEAR TIME SERIES MODELS

J. Carlos Escanciano

Working Paper No.02/05
February 2005
JEL No. C12.

ABSTRACT

In this article we study a general class of goodness-of-fit tests for the conditional mean of a linear or nonlinear time series model. Among the properties of the proposed tests are that they are suitable when the conditioning set is infinite-dimensional; are consistent against a broad class of alternatives including Pitman's local alternatives converging at the parametric rate ; and do not need to choose a lag order depending on the sample size or to smooth the data. It turns out that the asymptotic null distributions of the tests depend on the data generating process, so a new bootstrap procedure is proposed and theoretically justified. The proposed bootstrap tests are robust to higher order dependence, in particular to conditional heteroskedasticity of unknown form. A simulation study compares the finite sample performance of the proposed and competing tests and shows that our tests can play a valuable role in time series modeling. Finally, an application to an economic price series highlights the merits of our approach.

Juan Carlos Escanciano Reyero
Universidad de Navarra
Departamento de Métodos Cuantitativos
Campus Universitario
31080 Pamplona
jescanci@unav.es

Acknowledgments

Juan Carlos Escanciano is Assistant Professor at Universidad de Navarra, Facultad de Económicas, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain, e-mail: jescanci@unav.es. The author thanks Carlos Velasco for many helpful comments and suggestions. Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON.

GOODNESS-OF-FIT TESTS FOR LINEAR AND NON-LINEAR TIME SERIES MODELS

J. Carlos Escanciano*

January 18, 2005

Abstract

In this article we study a general class of goodness-of-fit tests for the conditional mean of a linear or nonlinear time series model. Among the properties of the proposed tests are that they are suitable when the conditioning set is infinite-dimensional; are consistent against a broad class of alternatives including Pitman's local alternatives converging at the parametric rate $n^{-1/2}$, with n the sample size; and do not need to choose a lag order depending on the sample size or to smooth the data. It turns out that the asymptotic null distributions of the tests depend on the data generating process, so a new bootstrap procedure is proposed and theoretically justified. The proposed bootstrap tests are robust to higher order dependence, in particular to conditional heteroskedasticity of unknown form. A simulation study compares the finite sample performance of the proposed and competing tests and shows that our tests can play a valuable role in time series modeling. Finally, an application to an economic price series highlights the merits of our approach.

Keywords and Phrases: Diagnostic test; Model adequacy; Nonlinear spectral analysis; Wild bootstrap; Conditional mean.

1. INTRODUCTION

In this paper we develop some methodology for testing the goodness-of-fit of a parametric conditional mean of a linear or nonlinear time series model. The proposed tests are suitable when the conditioning set is infinite-dimensional. More precisely, let $\{(Y_t, \mathbf{Z}'_{t-1})\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$, where $Y_t \in \mathbb{R}$ is the dependent (predicted) variable and $\mathbf{Z}_{t-1} = (Y_{t-1}, \mathbf{X}'_{t-1})' \in \mathbb{R}^m$, $m \in \mathbb{N}$, is the explanatory random vector containing the lagged value of the dependent variable and other explanatory variables $\mathbf{X}_{t-1}$, say. In this paper we are mainly concerned with the case in which the conditioning set at time $t-1$

*Juan Carlos Escanciano is Assistant Professor at Universidad de Navarra, Facultad de Económicas, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain, e-mail: jescanci@unav.es. The author thanks Carlos Velasco for many helpful comments and suggestions. The paper has also benefited from the comments of two referees, the associate editor and the joint editor. Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON.

is given by $\mathbf{I}_{t-1} = (\mathbf{Z}'_{t-1}, \mathbf{Z}'_{t-2}, ...)'$. From Probability Theory we know that under integrability of Y_t we can write the tautological expression

$$Y_t = m(\mathbf{I}_{t-1}) + \varepsilon_t,$$

where $m(\mathbf{I}_{t-1}) = E[Y_t \mid \mathbf{I}_{t-1}]$ is the conditional mean almost surely (a.s.) given the conditioning set $\mathbf{I}_{t-1}$, and $\varepsilon_t = Y_t - E[Y_t \mid \mathbf{I}_{t-1}]$ is, by construction, a martingale difference sequence (mds) with respect to $\mathcal{F}_{t-1}$, the $\sigma-$field generated by $\mathbf{I}_{t-1}$, i.e., $\mathcal{F}_{t-1} = \sigma(\mathbf{I}_{t-1}) \equiv \sigma(\mathbf{Z}_s : s \leq t-1,\ s \in \mathbb{Z})$.

Then, in parametric time series modeling one assumes the existence of a parametric family of functions $\mathcal{M} = \{f(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ and considers the following regression model

$$Y_t = f(\mathbf{I}_{t-1}, \boldsymbol{\theta}) + e_t(\boldsymbol{\theta}), \tag{1}$$

where $f(\mathbf{I}_{t-1}, \boldsymbol{\theta})$ is a parametric specification for the conditional mean $m(\mathbf{I}_{t-1})$, and $\{e_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is a sequence of disturbances of the model $\mathcal{M}$. Examples of specification (1) include ARMA, ARMAX, bilinear, nonlinear moving average, Markov-switching, smooth transition, exponential and threshold autoregressive models among many others, see, e.g., Tong (1990) or Fan and Yao (2003). Our main goal in this paper is to test the null hypothesis that $m(\cdot) \in \mathcal{M}$, i.e.,

$$H_0 : E[Y_t \mid \mathbf{I}_{t-1}] = f(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0) \text{ a.s., for some } \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p,$$

against the alternative that $m(\cdot) \notin \mathcal{M}$, or equivalently

$$H_1 : P(E[Y_t \mid \mathbf{I}_{t-1}] = f(\mathbf{I}_{t-1}, \boldsymbol{\theta})) < 1\ ,\ \text{for all } \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p.$$

From the regression model (1), the correct specification is tantamount to

$$E[e_t(\boldsymbol{\theta}_0) \mid \mathbf{I}_{t-1}] = 0 \text{ a.s., for some } \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p. \tag{2}$$

Parametric time series modeling continues to be attractive among practitioners because the parameter $\boldsymbol{\theta}_0$ together with the functional form $f(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0)$ describes, in a concise way, the relationship between the response Y_t and the conditioning set $\mathbf{I}_{t-1}$. A lack of fit in the postulated conditional mean can lead to misleading conclusions and statistical inferences, and to suboptimal point forecasts. To give an example, misspecifications in the conditional mean may deliver inconsistent estimations of the parameter $\boldsymbol{\theta}_0$. Therefore, in order to prevent wrong conclusions, every statistical inference which is based on the model $\mathcal{M}$ should be accompanied by a proper model check, i.e., a test for H_0.

There is a huge literature on testing the correct specification of a time series model. A large body of this literature uses the fact that under our assumptions $\sigma(\mathbf{I}^e_{t-1}) \subset \sigma(\mathbf{I}_{t-1})$, where $\mathbf{I}^e_{t-1} = (e_{t-1}(\boldsymbol{\theta}_0), e_{t-2}(\boldsymbol{\theta}_0), ...)'$, and thus, condition (2) yields that the error sequence $\{e_t(\boldsymbol{\theta}_0)\}$ satisfies

$$E[e_t(\boldsymbol{\theta}_0) \mid \mathbf{I}^e_{t-1}] = 0 \text{ a.s., for some } \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p. \tag{3}$$

The latter condition motivates that many specification tests for the conditional mean are based on checking for serial dependence (lack thereof) of the unobserved errors $\{e_t(\boldsymbol{\theta}_0)\}$. In particular, classical Portmanteau tests are based on checking the serial uncorrelatedness of the errors $\{e_t(\boldsymbol{\theta}_0)\}$, see e.g. Box and Pierce (1970), Ljung and Box (1978), or more recently, Paparoditis (2000), Peña and Rodriguez (2001) or Delgado, Hidalgo and Velasco (2003). However, it is well-known that, in general, the serial uncorrelatedness of the errors $\{e_t(\boldsymbol{\theta}_0)\}$ neither imply (3) nor (2), and therefore, these tests may not be able to detect some misspecifications in the conditional mean. As a matter of fact, correlation-based tests are inconsistent for testing H_0 in any direction where the error sequence $\{e_t(\boldsymbol{\theta}_0)\}$ is uncorrelated, see the simulations below for some examples. Uncorrelatedness is only one of the implications of the correct specification but does not characterize it in the presence of non-Gaussanity and/or nonlinearity. On the other hand, other tests consider the stronger hypothesis that the errors $\{e_t(\boldsymbol{\theta}_0)\}$ are serially independent, see the *BDS* test of Brock, Dechert and Scheinkman (1987) based on the correlation integral or the generalized spectral test of Hong and Lee (2003). However, testing for serial independence of the errors is a more restrictive condition than (3) and, in particular, it is possible that those tests reject a correct null model because of higher order dependence.

It is important to emphasize that most tests in the time series modeling literature have considered a finite number of lags in the conditioning set, i.e., they test for

$$E[e_t(\boldsymbol{\theta}_0) \mid \mathbf{I}_{t-1}^d] = 0 \text{ a.s., for some } \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p, \tag{4}$$

where $\mathbf{I}_{t-1}^d = (\mathbf{Z}_{t-1}', \mathbf{Z}_{t-2}', ..., \mathbf{Z}_{t-d}')'$, and d is finite, $d \in \mathbb{N}$. See Escanciano (2004) for a review of the literature on testing (4). Those tests are inconsistent against alternatives in H_1 satisfying (4). In addition, if the d used in (4) is large, most of these tests are highly affected by the so-called "curse of dimensionality" problem, see, e.g., the smoothing-based tests of Härdle and Mammen (1993) or the indicator-based test of Stute (1997). To make a test based on (4) consistent against H_1, it seems natural to consider d in (4) going to infinity with the sample size. In fact, de Jong (1996) has generalized Bierens' (1982) test to the case of $d \to \infty$ as $n \to \infty$. However, de Jong's test requires numerical integration with dimension equals to the sample size, which makes this test infeasible in applications where the sample size is usually large, e.g. financial applications. Therefore, we observe that when a large number of lags is used in the conditioning set most existing tests have poor power performance in finite samples, due to the loss of a large number of degrees of freedom, to the problem of the curse of dimensionality or to the integration in spaces of large dimensions.

To alleviate some of these problems and, at the same time, consider information at all lags, Hong (1999) has introduced a generalized spectral density as a new tool for testing interesting hypotheses in a nonlinear time series framework. Rather recently, Hong and Lee (2004) have extended Hong's (1999) ideas to a test for H_0 under processes which may display conditional dependence at second and higher conditional moments. Hong and Lee's (2004) test is based on the fact that, under (3), the pairwise error regressions $E[e_t(\boldsymbol{\theta}_0) \mid e_{t-j}(\boldsymbol{\theta}_0)]$ vanish (a.s.) $\forall j \geq 1$. Under integrability of $e_t(\boldsymbol{\theta}_0)$, this is in turn equivalent to the fact that the pairwise measures of dependence $\gamma_j^e(x, \boldsymbol{\theta}_0) = E[e_t(\boldsymbol{\theta}_0) \exp(ixe_{t-j}(\boldsymbol{\theta}_0))]$, $j \geq 1, x \in \mathbb{R}$, are identically equal to the zero function almost everywhere (a.e.), where $i = \sqrt{-1}$ is the imaginary unit. With this in mind, Hong and Lee's (2004) test is based on a Fourier transform of the measures $\{\gamma_j^e(\cdot, \boldsymbol{\theta}_0)\}_{j=1}^{\infty}$. From similar arguments, under H_0

$$\gamma_j(\boldsymbol{\theta}_0) = E[e_t(\boldsymbol{\theta}_0) \mid \mathbf{Z}_{t-j}] = 0 \text{ a.s. } \forall j, j \geq 1, \text{ for some } \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p. \tag{5}$$

Then, by choosing appropriately a parametric family of functions $\{w(\mathbf{Z}_{t-j}, \mathbf{x}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$, cf. Lemma 1 below, condition (5) can be equivalently expressed as

$$\gamma_{j,w}(\mathbf{x}, \boldsymbol{\theta}_0) = E[e_t(\boldsymbol{\theta}_0) w(\mathbf{Z}_{t-j}, \mathbf{x})] = 0 \text{ a.e. in } \Upsilon \subset \mathbb{R}^s, s \in \mathbb{N}, j \geq 1. \tag{6}$$

As we shall show below, usual examples of weighting functions w satisfying previous equivalence are $w(\mathbf{Z}_{t-j}, \mathbf{x}) = 1(\mathbf{Z}_{t-j} \leq \mathbf{x})$ with $\mathbf{x} \in [-\infty, \infty]^m$, where $1(A)$ denotes the indicator of the event A, or $w(\mathbf{Z}_{t-j}, \mathbf{x}) = \exp(i\mathbf{x}'\mathbf{Z}_{t-j})$ with $\mathbf{x} \in \mathbb{R}^m$, see Lemma 1. The tests proposed in this paper are then based on an integrated Fourier transform of the measures $\{\gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)\}_{j=1}^{\infty}$.

Although our tests and the test of Hong and Lee (2004) are founded on a generalized spectral approach they are different at least in four aspects. First, here we are mainly concerned with the problem of testing (2), whereas Hong and Lee (2004) test for (3). Second, our methodology is based on a generalized spectral distribution function of $\{\gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)\}_{j=1}^{\infty}$ contrary to the generalized spectral density function approach used in Hong and Lee (2004) and based on $\{\gamma_j^e(\cdot, \boldsymbol{\theta}_0)\}_{j=1}^{\infty}$. Therefore, unlike their test, our tests do not depend on any kernel and bandwidth choices, which are necessary for the consistent estimation of the density function. Note that our methodology is more general because is not restricted to exponential-based weighting families. Third, we overcome the technical problem of considering different weighting families w in (6) through a Hilbert space approach for the asymptotic theory that not only allows us to consider smooth and non-smooth $w's$ and multivariate $\mathbf{x}'s$, but also requires weaker conditions than other existing approaches. And fourth, in general, the asymptotic null distributions of our tests depend on the data generating process (DGP) and are no longer standard. Hence, a bootstrap approach to approximate the asymptotic critical values of our tests will be considered and justified theoretically.

We summarize the main characteristics of our tests as follows; (i) they are consistent against a broad class of linear and nonlinear alternatives to H_0, as we shall show in an extensive simulation experiment below; (ii) are consistent against pairwise Pitman's local alternatives converging at the parametric rate $n^{-1/2}$; (iii) incorporate information on the serial dependence from all lags and, at the same time, avoid the problem of the curse of dimensionality or high-dimensional integration; (iv) do not depend on any smoothing parameter or kernel; (v) are valid under fairly general regularity conditions on the underlying DGP, in particular, no mixing conditions are imposed; and (vi) are simple to compute.

A related work has been considered in Escanciano and Velasco (2003) for testing the martingale difference hypothesis using the exponential weighting function. In fact, Escanciano and Velasco's (2003) test can be viewed as the simple hypothesis here, i.e., the case in which $\boldsymbol{\theta}_0$ is known. In the present paper, we are more interested in the use of a general weighting function w and, more important, in the composite case in which $\boldsymbol{\theta}_0$ is unknown and has to be estimated from the sample. Note that this is not a trivial extension. In particular, the bootstrap approach of Escanciano and Velasco (2003) is not valid here, and a more involved resampling procedure is needed in the composite case.

The layout of the article is as follows, in Section 2 we present the generalized spectral distribution based-tests for testing H_0. In Section 3, we study the asymptotic distribution of our tests under the null, fixed and local alternatives. In Section 4, we propose and justify theoretically a bootstrap method to implement the tests. We make an extensive simulation exercise and an empirical application in Section 5, comparing with competing tests. Finally, we conclude in Section 6 with some conclusions and further research. All proofs are gathered in an appendix. Throughout, A^c, A' and $|A|$ denote the complex conjugate, the matrix transpose and the Euclidean norm of A, respectively. Unless indicated, all convergences are taken as the sample size $n \to \infty$. In the sequel C is a generic constant that may change from one expression to another.

2. THE INTEGRATED GENERALIZED SPECTRAL TESTS

The main purpose of this paper is to test for H_0 when the conditioning set at time $t-1$ is infinite-dimensional and is given by $\mathbf{I}_{t-1} = (\mathbf{Z}'_{t-1}, \mathbf{Z}'_{t-2}, ...)'$, with $\mathbf{Z}_{t-j} \in \mathbb{R}^m$, $j \geq 1$, $m \in \mathbb{N}$. One possible approach for testing H_0 is to consider (4) with d tending to infinite with the sample size. However, this approach delivers some important problems such as the curse of dimensionality or integration in high-dimensional spaces, hindering its application in cases where the sample size is large and the d demanded is also large, e.g. high frequency data. To avoid those problems, we propose a pairwise approach based on (6) that, although checking for a necessary but not sufficient condition of (2), it is general enough to pick a broad class of alternatives and delivers simple tests statistics.

The equivalence between (5) and (6) plays a crucial role in our subsequent work. The following Lemma gives sufficient conditions on the parametric family $\mathcal{F} = \{w(\mathbf{Z}, \mathbf{x}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ to satisfy this equivalence, where hereafter $\mathbf{Z}$ is a random vector with the same distribution as $\mathbf{Z}_t$, $t \in \mathbb{Z}$. We need some definitions. Let denote by $C_b(\mathbb{R}^m)$ the space of all bounded, continuous complex-valued functions on $\mathbb{R}^m$. We say that a class of functions in $C_b(\mathbb{R}^m)$, $\mathcal{F}$ say, is a vector lattice if it is a vector space that is closed under taking positive parts: if $f \in \mathcal{F}$, then $f^+ = \max\{f, 0\} \in \mathcal{F}$. We say that $\mathcal{F} \subset C_b(\mathbb{R}^m)$ is an algebra if is a vector space that is closed under taking products, i.e., if $f, g \in \mathcal{F}$, then $f \cdot g \in \mathcal{F}$. Also, $\mathcal{F}$ separates points of $\mathbb{R}^m$ if, for every pair $x \neq y \in \mathbb{R}^m$, there exists a function $f \in \mathcal{F}$, with $f(x) \neq f(y)$. A function is analytic if is locally equal to its Taylor expansion at each point of its domain. Finally, a class of Borel sets of $\mathbb{R}^m$, $\mathcal{B}$ say, is a separating class if two Borel probability measures that agree in $\mathcal{B}$ necessarily agree also on the whole Borel σ-field of $\mathbb{R}^m$, see Billingsley (1999, p. 9).

Lemma 1 *The following conditions are sufficient for the class of functions $\mathcal{F} = \{w(\mathbf{Z}, \mathbf{x}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ to satisfy the equivalence between (5) and (6):*

(a) $\mathcal{F} \subset C_b(\mathbb{R}^m)$ is a vector lattice that contains the constant functions and separates points of $\mathbb{R}^m$.

(b) $\mathcal{F} \subset C_b(\mathbb{R}^m)$ is an algebra that contains the constant functions and separates points of $\mathbb{R}^m$.

(c) $\mathcal{F} = \{w(\mathbf{x}'\mathbf{Z}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ and w is an analytic function which is non-polynomial.

(d) $\mathcal{F} = \{1(\mathbf{Z} \in B_{\mathbf{x}}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ and $\{B_{\mathbf{x}}\}_{\mathbf{x} \in \Upsilon}$ is a separating class of Borel sets of $\mathbb{R}^m$.

Examples of families satisfying (c) in the previous Lemma are $w(\mathbf{Z}, \mathbf{x}) = \exp(i\mathbf{x}'\mathbf{Z})$, $w(\mathbf{Z}, \mathbf{x}) = \sin(\mathbf{x}'\mathbf{Z})$ or $w(\mathbf{Z}, \mathbf{x}) = 1/(1+\exp(\mathbf{x}'\mathbf{Z}))$, all of them with Υ a compact set of $\mathbb{R}^m$ containing the origin, see Bierens and Ploberger (1997) or Stinchcombe and White (1998). Whereas $w(\mathbf{Z}, \mathbf{x}) = 1(\mathbf{Z} \leq \mathbf{x})$ satisfies (d). Throughout the paper we shall assume that the family of functions $\mathcal{F} = \{w(\mathbf{Z}, \mathbf{x}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ satisfies at least one of the sufficient conditions of Lemma 1.

To consider simultaneously all the dependence measures $\{\gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)\}$, we define $\gamma_{-j,w}(\cdot, \boldsymbol{\theta}_0) = \gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)$ for $j \geq 1$, and write the Fourier transform of the functions $\{\gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)\}_{j=-\infty}^{\infty}$, i.e.,

$$f_w(u, \mathbf{x}, \boldsymbol{\theta}_0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{j,w}(\mathbf{x}, \boldsymbol{\theta}_0) e^{-iju} \qquad \forall u \in [-\pi, \pi], \mathbf{x} \in \Upsilon, \tag{7}$$

which contains the same information about H_0 as the whole sequence $\{\gamma_{j,w}(\mathbf{x}, \boldsymbol{\theta}_0)\}_{j=0}^{\infty}$. Note that under H_0, $f_w(u, \mathbf{x}, \boldsymbol{\theta}_0) \equiv f_{0,w}(\mathbf{x}, \boldsymbol{\theta}_0) = (2\pi)^{-1}\gamma_{0,w}(\mathbf{x}, \boldsymbol{\theta}_0)$. Using a similar idea, Hong and Lee (2004) have proposed a test for H_0 based on the Fourier transform $f_{HL}(u, x, \boldsymbol{\theta}_0)$, where $f_{HL}(u, x, \boldsymbol{\theta}_0)$ is the same as $f_w(u, \mathbf{x}, \boldsymbol{\theta}_0)$ but with the measures $\{\gamma_j^e(x, \boldsymbol{\theta}_0)\}$ replacing $\{\gamma_{j,w}(\mathbf{x}, \boldsymbol{\theta}_0)\}$. Hong and Lee's (2004) test statistic is an standardization of an L_2-distance between kernel estimators of f_{HL} under H_1 and under H_0, see Section 5 below.

The novel approach here is to avoid kernel estimation by considering a generalized spectral distribution function based on the dependence measures $\{\gamma_{j,w}(\cdot, \boldsymbol{\theta}_0)\}_{j=-\infty}^{\infty}$, i.e., based on the integral of $f_w(u, \mathbf{x}, \boldsymbol{\theta}_0)$

$$H_w(\lambda, \mathbf{x}, \boldsymbol{\theta}_0) = 2\int_0^{\lambda\pi} f_w(u, \mathbf{x}, \boldsymbol{\theta}_0) du \qquad \forall \lambda \in [0, 1], \mathbf{x} \in \Upsilon,$$

that is,

$$H_w(\lambda,\mathbf{x},\boldsymbol{\theta}_0)=\gamma_{0,w}(\mathbf{x},\boldsymbol{\theta}_0)\lambda+2\sum_{j=1}^{\infty}\gamma_{j,w}(\mathbf{x},\boldsymbol{\theta}_0)\frac{\sin j\pi\lambda}{j\pi}. \tag{8}$$

Now, suppose we have a random sample $\{Y_t,\widehat{\mathbf{I}}_{t-1}\}_{t=1}^{n}$ of size n which is used to estimate the model $f(\mathbf{I}_{t-1},\boldsymbol{\theta})$. Here $\widehat{\mathbf{I}}_{t-1}$ is the information set observed at time $t-1$ that contains $(\mathbf{Z}_{t-1}',\mathbf{Z}_{t-2}',...,\mathbf{Z}_0')'$ and that may contain some initial values. We obtain residuals $\widehat{e}_t\equiv\widehat{e}_t(\boldsymbol{\theta}_n)=Y_t-f(\widehat{\mathbf{I}}_{t-1},\boldsymbol{\theta}_n)$ where $\boldsymbol{\theta}_n$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\theta}_0$, e.g. the conditional nonlinear least squares estimator (NLSE). The sample version of $\gamma_{j,w}(\mathbf{x},\boldsymbol{\theta}_0)$ is then given by

$$\widehat{\gamma}_{j,w}(\mathbf{x},\boldsymbol{\theta}_n)=\frac{1}{n_j}\sum_{t=j}^{n}\widehat{e}_t w(\mathbf{Z}_{t-j},\mathbf{x}),\ j\geq 1,\ n_j=n-j+1.$$

Hence, the sample analogue of (8) is

$$\widehat{H}_w(\lambda,\mathbf{x},\boldsymbol{\theta}_n)=\widehat{\gamma}_{0,w}(\mathbf{x},\boldsymbol{\theta}_n)\lambda+2\sum_{j=1}^{n}\widehat{\gamma}_{j,w}(\mathbf{x},\boldsymbol{\theta}_n)(n_j/n)^{1/2}\frac{\sin j\pi\lambda}{j\pi},$$

with $(n_j/n)^{1/2}$ a finite sample correction factor which does not affect the asymptotic theory and delivers a better finite sample performance of the test procedure. The effect of this correction factor is to put less weight on very large lags, for which we have less sample information. Under the null hypothesis, $H_w(\lambda,\mathbf{x},\boldsymbol{\theta}_0)=\gamma_{0,w}(\mathbf{x},\boldsymbol{\theta}_0)\lambda$, and therefore, tests can be based on the discrepancy between $\widehat{H}_w(\lambda,\mathbf{x},\boldsymbol{\theta}_n)$ and $\widehat{H}_{0,w}(\lambda,\mathbf{x},\boldsymbol{\theta}_n)=\widehat{\gamma}_{0,w}(\mathbf{x},\boldsymbol{\theta}_n)\lambda$, i.e.

$$S_{n,w}(\lambda,\mathbf{x},\boldsymbol{\theta}_n)=\left(\frac{n}{2}\right)^{1/2}\{\widehat{H}_w(\lambda,\mathbf{x},\boldsymbol{\theta}_n)-\widehat{H}_{0,w}(\lambda,\mathbf{x},\boldsymbol{\theta}_n)\}=\sum_{j=1}^{n}n_j^{1/2}\widehat{\gamma}_{j,w}(\mathbf{x},\boldsymbol{\theta}_n)\frac{\sqrt{2}\sin j\pi\lambda}{j\pi}.$$

In order to evaluate the distance from $S_{n,w}(\lambda,\mathbf{x},\boldsymbol{\theta}_n)$ to zero, a norm has to be chosen. We consider the usual Cramér-von Mises (CvM) norm

$$D_{n,w}^2(\boldsymbol{\theta}_n)=\int_{\Pi}\left|S_{n,w}(\lambda,\mathbf{x},\boldsymbol{\theta}_n)\right|^2W(d\mathbf{x})d\lambda=\sum_{j=1}^{n}\frac{n_j}{(j\pi)^2}\int_{\Upsilon}\left|\widehat{\gamma}_{j,w}(\mathbf{x},\boldsymbol{\theta}_n)\right|^2W(d\mathbf{x}), \tag{9}$$

where $W(\cdot)$ is an integrating function depending on the weighting family w and satisfying some mild conditions (see Assumption A5 below). Therefore, our tests consist in rejecting H_0 for "large" values of $D_{n,w}^2(\boldsymbol{\theta}_n)$. Note that $D_{n,w}^2(\boldsymbol{\theta}_n)$ uses all lags contained in the sample, does not depend on any lag order or kernel function and is very simple to compute, see Section 5. On the other hand, the range of possibilities in the choice of w and W gives flexibility for $D_{n,w}^2(\boldsymbol{\theta}_n)$ in directing the power against some desired directions. Next section justifies inferences based on the asymptotic theory.

3. ASYMPTOTIC THEORY

3.1 Asymptotic Null Distribution

To elaborate the asymptotic theory we consider the following assumptions. Let denote $\mathbf{g}(\mathbf{I}_{t-1}, \boldsymbol{\theta}) \equiv \mathbf{g}_t(\boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta}')f(\mathbf{I}_{t-1}, \boldsymbol{\theta})$ and $w(\mathbf{Z}_{t-j}, \mathbf{x}) \equiv w_{t-j}(\mathbf{x})$. Recall that $\varepsilon_t = Y_t - E[Y_t \mid \mathbf{I}_{t-1}]$, $\mathbf{I}_{t-1} = (\mathbf{Z}'_{t-1}, \mathbf{Z}'_{t-2}, ...)'$, and that under H_0, $e_t(\boldsymbol{\theta}_0) = \varepsilon_t$ a.s.

Assumption A1:

A1(a): $\{Y_t, \mathbf{Z}_{t-1}\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process.

A1(b): $E[\varepsilon_1^2] < C$.

Assumption A2: The response function $f_t(\cdot) \equiv f(\mathbf{I}_{t-1}, \cdot)$ is specified stationary, ergodic and is twice continuously differentiable on Θ. The function $\mathbf{g}_t(\boldsymbol{\theta}_0)$ is stationary and ergodic, $\mathcal{F}_{t-1}$-measurable and there exists an integrable function $\mathbf{M}(\mathbf{I}_{t-1})$ with $|\mathbf{g}(\mathbf{I}_{t-1}, \boldsymbol{\theta})| \leq \mathbf{M}(\mathbf{I}_{t-1})$.

Assumption A3:

A3(a): The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter $\boldsymbol{\theta}_0$ belongs to the interior of Θ. There exists a unique $\boldsymbol{\theta}_* \in \Theta$ such that $|\boldsymbol{\theta}_n - \boldsymbol{\theta}_*| = o_P(1)$. Obviously, under H_0, $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$.

A3(b): The estimator θ_n satisfies the following asymptotic expansion under H_0

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{h}(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}_0) + o_P(1),$$

where $\mathbf{h}(\cdot)$ is such that $E[\mathbf{h}(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}_0)] = 0$ and $\mathbf{L}(\boldsymbol{\theta}_0) = E[\mathbf{h}(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}_0)\mathbf{h}'(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}_0)]$ exists and is positive definite.

Assumption A4: The integrating function $W(\cdot)$ is a probability distribution function absolutely continuous with respect to Lebesgue measure. The weighting function $w(\cdot)$ is such that the equivalence between (5) and (6) holds, and is uniformly bounded on compacta. Also, $w(\cdot)$ satisfies the following uniform law of large numbers (ULLN)

$$\sup_{\mathbf{x}\in\Upsilon_c}\left|n^{-1}\sum_{t=1}^{n}\zeta_t w(\boldsymbol{\xi}_t, \mathbf{x}) - E[\zeta_t w(\boldsymbol{\xi}_t, \mathbf{x})]\right| \longrightarrow 0 \text{ a.s.},$$

whenever $\{(\zeta_t, \boldsymbol{\xi}'_t),\ t = 0, \pm1, ...\}$ is a strictly stationary and ergodic process with $\zeta_t \in \mathbb{R}$, $\boldsymbol{\xi}_t \in \mathbb{R}^m$, $E|\zeta_1| < \infty$, and Υ_c is any compact subset of $\Upsilon \subset \mathbb{R}^s$.

Assumption A5: The observed information set available at period t, $\widehat{\mathbf{I}}_t$, may contain some assumed initial values and satisfies $\lim_{n\to\infty} \sum_{t=1}^{n} \left(E\sup_{\boldsymbol{\theta}\in\Theta}(f(\mathbf{I}_{t-1}, \boldsymbol{\theta}) - f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}))^2\right)^{1/2} \leq C$.

Assumption A1 is a mild condition on the DGP, that permits but does not require $var(Y_t) < \infty$. Here, we only assume finite variance for the errors ε_t, whereas most works in the literature assume fourth bounded moments. This fourth moment assumption may look restrictive, it rules out many empirically relevant GARCH processes whose fourth moments are often found to be infinite. Note that unlike most existing test under time series, we do not need of any mixing or asymptotic independence assumption to derived the asymptotic theory, see e.g. A.2 in Hong and Lee (2004). These asymptotic independence concepts are difficult to check in practice, whereas the martingale difference errors assumption used in our asymptotic theory is implied from H_0. A1 can be extended to non-stationary sequences using the results of Jakubowski (1980) at the cost of complicating further the notation. Assumption A2 is on the model and is standard in the conditional mean specification literature, see e.g. Koul and Stute (1999). Assumption A3 is satisfied under mild conditions, for instance, for the NLSE or for its robust modifications (under further regularity assumptions), see Chapter 5 in Koul (2002) or Chapter 6 in Hall and Heyde (1980). Examples of $W(\cdot)$ include the cumulative distributions functions (cdf) of a N(0,1), Double Exponential or the Student's t_ν distribution. The continuity assumption of W is essential to gain consistency in the test procedure, see Escanciano and Velasco (2003) for further discussions on the choice of W. All previous examples of functions w satisfy A4. A5 is a condition on the truncation of the information set $\widehat{\mathbf{I}}_{t-1}$ and is similar to Assumption A4 in Hong and Lee (2004). Those authors show that A5 is satisfied for some standard examples, e.g. ARMA(1,1) models, under mild conditions on the conditional mean parameters.

To elaborate the asymptotic theory we need some further notation. Let $\Pi = [0,1] \times \Upsilon$ and $\boldsymbol{\eta}=(\lambda, \mathbf{x}')' \in \Pi$. We first establish the null limit distribution of the process $S_n(\lambda, \mathbf{x}, \boldsymbol{\theta}_n) \equiv S_n(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ under H_0. We consider $S_n(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ as a random element on the Hilbert space $L_2(\Pi, \nu)$ of all complex-valued and square ν-integrable functions on Π, where ν is the product measure of the W-measure and the Lebesgue measure on [0,1]. In $L_2(\Pi, \nu)$ we define the inner product

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

$L_2(\Pi, \nu)$ is endowed with the natural Borel σ-field induced by the norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$, see Chapter VI in Parthasarathy (1967) for convergence results on Hilbert spaces. If Z is an $L_2(\Pi, \nu)$-valued random variable, we say that Z has mean m if $E[\langle Z, h\rangle] = \langle m, h\rangle \; \forall h \in L_2(\Pi, \nu)$. If $E\|Z\|^2 < \infty$ and Z has zero mean, then the covariance operator of Z, C_Z say, is defined by $C_Z(h) = E[\langle Z, h\rangle Z]$. Let $\Longrightarrow$ denote weak convergence in the Hilbert space $L_2(\Pi, \nu)$ endowed with the norm metric. Also, denote by $\xrightarrow{L_2}$ convergence in probability in $L_2(\Pi, \nu)$, i.e., $Z_n \xrightarrow{L_2} Z \Longleftrightarrow \|Z_n - Z\| \xrightarrow{P} 0$. Let define $\Psi_j(\lambda) = \sqrt{2}(\sin j\pi\lambda)/j\pi$, $\mathbf{b}_j(\mathbf{x}, \boldsymbol{\theta}_0) = E[w_{t-j}(\mathbf{x})\mathbf{g}_t(\boldsymbol{\theta}_0)]$, $\mathbf{G}_w(\boldsymbol{\eta}) \equiv \mathbf{G}_w(\boldsymbol{\eta}, \boldsymbol{\theta}_0) = \sum_{j=1}^{\infty} \mathbf{b}_j(\mathbf{x}, \boldsymbol{\theta}_0)\Psi_j(\lambda)$ and

$$\sigma_h^2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} E[\varepsilon_1^2 \int_{\Pi\times\Pi} h(\boldsymbol{\eta}_1) h^c(\boldsymbol{\eta}_2) w_{1-j}^c(\mathbf{x}) w_{1-k}(\mathbf{y}) \Psi_j(\lambda)\Psi_k(\varpi) d\nu(\boldsymbol{\eta}_1) d\nu(\boldsymbol{\eta}_2)], \; h \in L_2(\Pi, \nu), \quad (10)$$

with $\boldsymbol{\eta}_1 = (\lambda, \mathbf{x}')'$ and $\boldsymbol{\eta}_2 = (\varpi, \mathbf{y}')'$. Let $\mathbf{V}$ be a normal random vector with zero mean and variance-covariance matrix given by $\mathbf{L}(\boldsymbol{\theta}_0)$, and let $S_w^0(\cdot)$ be a Gaussian process in $L_2(\Pi,\nu)$ with zero mean and covariance operator $C_{S_w^0}$ satisfying $\sigma_h^2 = \langle C_{S_w^0}(h), h\rangle$, $\forall h \in L_2(\Pi,\nu)$, where σ_h^2 is defined in (10). Then, under Assumptions A1-A5 we establish the asymptotic null distribution of $S_{n,w}$ in the following Theorem.

Theorem 1 *Under Assumptions A1-A5 and H_0, the process $S_{n,w}$ converges weakly to S_w on $L_2(\Pi,\nu)$, where $S_w(\cdot)$ has the same distribution as $S_w^0(\cdot) - \mathbf{G}_w'(\cdot)\mathbf{V}$, with*

$$Cov(S_w^0(\boldsymbol{\eta}), \mathbf{V}) = \sum_{j=1}^{\infty} E[\varepsilon_t w_{t-j}(\mathbf{x})\mathbf{h}(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}_0)]\Psi_j(\lambda).$$

The next corollary follows from the Continuous Mapping Theorem (Billingsley 1999, Theorem 2.7) and Theorem 1.

Corollary 1 *Under the Assumptions of Theorem 1,*

$$D_{n,w}^2(\boldsymbol{\theta}_n) \xrightarrow{d} D_{\infty,w}^2(\boldsymbol{\theta}_0) = \int |S_w(\lambda, \mathbf{x}, \boldsymbol{\theta}_0)|^2 W(d\mathbf{x})d\lambda.$$

For $w_{t-j}(\mathbf{x}) = 1(\mathbf{Z}_{t-j} \leq \mathbf{x})$, it is natural to choose $W(\cdot)$ as $F_n(\cdot)$, the empirical distribution function based on $\{\mathbf{Z}_{t-1}\}_{t=1}^n$. In the next corollary we shall show that the use of the empirical distribution function $F_n(\mathbf{x})$ instead of the true continuous cdf, $F(\mathbf{x})$ say, does not affect the asymptotic null distribution of the CvM test for the indicator case.

Corollary 2 *Under the assumptions of Theorem 1 and the continuity of $F(\mathbf{x})$,*

$$D_{n,I}^2 \xrightarrow{d} \int (S_I(\lambda, \mathbf{x}, \boldsymbol{\theta}_0))^2 F(d\mathbf{x})d\lambda.$$

To end this section, it is important to remark that the asymptotic null distribution of $D_{n,w}^2$ depends in a complex way on the DGP as well as the hypothesized model under the null, so critical values have to be tabulated for each model and each DGP, making the application of these asymptotic results difficult in practice. To overcome this problem we shall propose to implement the tests with the assistance of a bootstrap procedure in Section 4.

3.2 Consistency and Local Alternatives

The consistency properties of the tests are stated in the following theorems. Let consider the global alternative

$$H_a : Y_t = f_t(\boldsymbol{\theta}_0) + a_t + \varepsilon_t,$$

where $\{a_t\}$ is strictly stationary, ergodic, with a_t $\mathcal{F}_{t-1}$-measurable and with $E|a_1| < \infty$. The next theorem shows the asymptotic behaviour of $S_{n,w}$ under the global alternative H_a.

Theorem 2 *Under Assumptions A1-A5 and H_a,*

$$n^{-1/2}S_{n,w}(\boldsymbol{\eta},\boldsymbol{\theta}_n)\xrightarrow{L_2}L_w(\boldsymbol{\eta})=\sum_{j=1}^{\infty}\varsigma_j(\mathbf{x})\Psi_j(\lambda),$$

where $\varsigma_j(\mathbf{x})=E[a_tw_{t-j}(\mathbf{x})]$.

Let denote by Ξ the class of alternatives $\{a_t\}$ for which it holds that under H_a there exists at least one $j\geq 1$, such that $\varsigma_j(\mathbf{x})\neq 0$ for some subset of Υ with positive Lebesgue measure. From Lemma 1 we observe that Ξ is the class of alternatives $\{a_t\}$ for which it holds that under H_a there exists at least one $j\geq 1$, such that $E[a_t\mid\mathbf{Z}_{t-j}]\neq 0$ with positive P-measure. Then, if $\{a_t\}\in\Xi$, because $W(\cdot)$ is absolutely continuous with respect the Lebesgue measure, $n^{-1}D^2_{n,w}$ will converge to a positive constant under H_a, and consequently, our test statistic $D^2_{n,w}$ will be consistent against H_a. It is important to mention that although the set Ξ forms a large class of alternatives, it does not cover all possible alternatives. The test statistic $D^2_{\infty,w}$ will not be able to detect those alternatives in the complement of Ξ. We illustrate this with the following

Example 1 *Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) zero mean random variables. Let define the process $Y_t=\varepsilon_{t-1}\varepsilon_{t-3}+\varepsilon_t$. In this case, $\mathbf{Z}_{t-1}=(Y_{t-1},\varepsilon_{t-1})'$, $\mathbf{I}_{t-1}=(\mathbf{Z}'_{t-1},\mathbf{Z}'_{t-2},...)'$, $\boldsymbol{\theta}_0$ is known and equal to zero, $f_t(\boldsymbol{\theta}_0)\equiv 0$ and $a_t=\varepsilon_{t-1}\varepsilon_{t-3}$. Then, it is easy to show that $\varsigma_j(\mathbf{x})=E[\varepsilon_{t-1}\varepsilon_{t-3}w_{t-j}(\mathbf{x})]=0$ a.e., $\forall j\geq 1$.*

To gain insight of the consistency properties of the tests, the next theorem shows the behavior of our tests statistics under a sequence of alternative hypotheses tending to the null at the parametric rate $n^{-1/2}$. Let consider the local alternatives

$$H_{a,n}:Y_{t,n}=f_t(\boldsymbol{\theta}_0)+\frac{a_t}{\sqrt{n}}+\varepsilon_t, \tag{11}$$

where $\{a_t\}$ is as in H_a. To proceed further, we need an assumption related to the behavior of the estimator under these local alternatives.

*Assumption A6***:** The estimator $\boldsymbol{\theta}_n$ satisfies the following asymptotic expansion under $H_{a,n}$

$$\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}_0)=\boldsymbol{\xi}_a+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{h}(Y_t,\mathbf{I}_{t-1},\boldsymbol{\theta}_0)+o_P(1),$$

where the function $\mathbf{h}(\cdot)$ is as in A3 and $\boldsymbol{\xi}_a\in\mathbb{R}^p$.

Theorem 3 *Under the sequence of alternatives hypotheses (11) and Assumptions A1-A6,*

$$S_{n,w}\Longrightarrow S_w+L_w(\cdot)-\mathbf{G}'_w(\cdot)\boldsymbol{\xi}_a,$$

where S_w and L_w are the processes defined in Theorem 1 and Theorem 2, respectively. Then

$$D^2_{n,w}(\boldsymbol{\theta}_n)\xrightarrow{d}\int\left|S_w(\boldsymbol{\eta},\boldsymbol{\theta}_0)+L_w(\boldsymbol{\eta})-\mathbf{G}'_w(\boldsymbol{\eta})\boldsymbol{\xi}_a\right|^2d\nu(\boldsymbol{\eta}).$$

Furthermore, it can be shown that our tests will have nontrivial power against the local nonparametric alternatives (11) in Ξ not collinear to the score $\mathbf{g}_t(\boldsymbol{\theta}_0)$, see Escanciano (2004) for further discussions on the local power properties of residual-marked tests. This property is not attainable for those tests using lag-bandwidth parameters or a fixed number of lags, e.g. Koul and Stute (1999), Fan and Huang (2001) or Hong and Lee (2004).

4. BOOTSTRAP APPROXIMATIONS

Resampling methods have been used extensively in the model checks literature of regression and time series models, see e.g. Härdle and Mammen (1993), Stute, Gonzalez-Manteiga and Presedo-Quindimil (1998) or more recently Li, Hsiao and Zinn (2003) in an i.i.d context, or Franke, Kreiss and Mammen (2002) for time series sequences. It is shown, in these papers, that the most relevant bootstrap method for regression problems is the wild bootstrap (WB) introduced in Wu (1986) and Liu (1988). Here we extend the WB to our present context. In time series, the block bootstrap is the oldest and best known bootstrap method. Particularly well-suited in our set-up is the stationary bootstrap of Politis and Romano (1994), because it is justified for general Hilbert-valued random variables. Unfortunately, the block bootstrap involves the choice of the block size and, more important, estimation errors that converge to zero relatively slow, see Härdle, Horowitz and Kreiss (2003) for a discussion on these issues. Here we approximate the asymptotic null distribution of $S_{n,w}$ by that of

$$S^*_{n,w}(\lambda, \mathbf{x}, \boldsymbol{\theta}^*_n) = \sum_{j=1}^{n} n_j^{1/2} \widehat{\gamma}^*_j(\mathbf{x}, \boldsymbol{\theta}^*_n) \Psi_j(\lambda),$$

with

$$\widehat{\gamma}^*_j(\mathbf{x}, \boldsymbol{\theta}^*_n) = \frac{1}{n_j} \sum_{t=j}^{n} \widehat{e}^*_t(\boldsymbol{\theta}^*_n) w_{t-j}(\mathbf{x}),$$

and where the fixed-design wild bootstrap (FDWB) residuals $\widehat{e}^*_t(\boldsymbol{\theta}^*_n)$, $1 \leq t \leq n$, are obtained from the following algorithm:

Step 1 Estimate the original model and obtain the residuals $\widehat{e}_t(\boldsymbol{\theta}_n)$.

Step 2 Generate WB residuals according to $\widehat{e}^*_t(\boldsymbol{\theta}_n) = \widehat{e}_t(\boldsymbol{\theta}_n) V_t$ for $1 \leq t \leq n$, with $\{V_t\}$ a sequence of i.i.d random variables with zero mean, unit variance, bounded support and also independent of the sequence $\{(Y_t, \widehat{\mathbf{I}}'_{t-1})'\}_{t=1}^{n}$.

Step 3 Given $\boldsymbol{\theta}_n$ and $\widehat{e}^*_t(\boldsymbol{\theta}_n)$, generate bootstrap data for the dependent variable Y^*_taccording to

$$Y^*_t = f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n) + \widehat{e}^*_t(\boldsymbol{\theta}_n) \text{ for } 1 \leq t \leq n.$$

Step 4 Compute $\boldsymbol{\theta}^*_n$ from the data $\{Y^*_t, \widehat{\mathbf{I}}_{t-1}\}_{t=1}^{\infty}$ and compute the FDWB residuals $\widehat{e}^*_t(\boldsymbol{\theta}^*_n) = Y^*_t - f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}^*_n)$ for $t = 1, ..., n$.

Examples of $\{V_t\}$ sequences are i.i.d. Bernoulli variates with $P(V_t = 0.5(1-\sqrt{5})) = b$ and $P(V_t = 0.5(1+\sqrt{5})) = 1-b$, with $b = (1+\sqrt{5})/2\sqrt{5}$, used in e.g. Mammen (1993), Stute, Gonzalez-Manteiga and Presedo-Quindimil (1998), or $P(V_t = 1) = 0.5$ and $P(V_t = -1) = 0.5$, as in Liu (1988) or de Jong (1996), for other sequences see Mammen (1993). The next theorem justifies theoretically the bootstrap approximation. We use the concept of convergence in distribution in probability one, a less restrictive concept is convergence in distribution in probability, see Giné and Zinn (1990) for more detailed discussions on these concepts. We need an additional assumption on the bootstrap estimator. In the remaining of this section and using standard bootstrap notation, denote by E^* the expectation operator given the sample $\{(Y_t, \widehat{\mathbf{I}}'_{t-1})'\}_{t=1}^n$. Let define $\mathbf{L}^*(\boldsymbol{\theta}_n) = E^*[\mathbf{h}(Y_t^*, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n)\mathbf{h}'(Y_t^*, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n)]$.

Assumption A7:

A7(a): The estimator $\boldsymbol{\theta}_n^*$ satisfies the following asymptotic expansion

$$\sqrt{n}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{h}(Y_t^*, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n) + o_P(1), \text{ a.s.}$$

where the function $\mathbf{h}(\cdot)$ is as in A3 with

A7(b): $E^*[\mathbf{h}(Y_t^*, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n)] = 0$, a.s..

A7(c): $\mathbf{L}^*(\boldsymbol{\theta}_n)$ exists and is positive definite (a.s.) with $\mathbf{L}(\boldsymbol{\theta}_n) \to \mathbf{L}(\boldsymbol{\theta}_*)$ a.s..

A7(d): $n^{-1}\sum_{t=1}^{n} E^*[e_t(\boldsymbol{\theta}_n)w_{t-j}(\mathbf{x})\mathbf{V}_t\mathbf{h}(Y_t^*, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n)] \to E[e_t(\boldsymbol{\theta}_*)w_{t-j}(\mathbf{x})\mathbf{h}(Y_t, \widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_*)]$ a.s..

In many cases, the function $\mathbf{h}(\cdot)$ required in A3(b) and A7 can be expressed as $\mathbf{h}(Y_t, \mathbf{I}_{t-1}, \boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\theta})\mathbf{k}(\mathbf{I}_{t-1}, \boldsymbol{\theta})$ for some function $\mathbf{k}(\cdot)$, see e.g. the NLSE or, more generally, estimators resulting from a martingale estimating equation, see Heyde (1997). Then, in those cases A7 is satisfied under some mild conditions on the function $\mathbf{k}(\cdot)$.

Theorem 4 *Assume A1-A7, then, under the null hypothesis H_0, under any fixed alternative hypothesis or under the local alternatives (11),*

$$S_{n,w}^* \underset{*}{\Longrightarrow} \widetilde{S}_w, \text{ a.s.,}$$

where $\widetilde{S}_w$ is the same Gaussian process of Theorem 1 but with $\boldsymbol{\theta}_*$ replacing $\boldsymbol{\theta}_0$ and $\underset{*}{\Longrightarrow}$ denote weak convergence almost surely under the bootstrap law, see Giné and Zinn (1990).

5. FINITE SAMPLE PERFORMANCE AND EMPIRICAL APPLICATION

In order to examine the finite sample performance of the proposed tests we carry out a simulation experiment with some DGP under the null and under the alternative. In the simulations we consider the univariate case, i.e., $m = 1$ with $Z_t = Y_t$, $t \in \mathbb{Z}$. We compare our tests with the test of Bierens (1982), the generalized spectral test of Hong and Lee (2004), the multivariate bootstrap version of Koul and Stute (1999) proposed in Escanciano (2004) and the usual Portmanteau test of Ljung and Box (1978). We briefly describe our simulation setup. We denote by $D^2_{n,I}$ and $D^2_{n,C}$ our new CvM tests corresponding to $w(Y_t, x) = 1(Y_t \leq x)$ and $w(Y_t, x) = \exp(iY_t x)$, respectively. They are given by

$$D^2_{n,I} = \sum_{j=1}^{n} \frac{n_j}{n(j\pi)^2} \sum_{t=1}^{n} \widehat{\gamma}^2_{j,I}(Y_{t-1}, \boldsymbol{\theta}_n),$$

with $\widehat{\gamma}_{j,I}(x, \boldsymbol{\theta}_n) = (\widehat{\sigma}_e n_j)^{-1} \sum_{t=j}^{n} \widehat{e}_t(\boldsymbol{\theta}_n) 1(Y_{t-j} \leq x)$, $\widehat{\sigma}^2_e = n^{-1} \sum_{t=1}^{n} \widehat{e}^2_t(\boldsymbol{\theta}_n)$, and

$$D^2_{n,C} = \sum_{j=1}^{n} \frac{1}{\widehat{\sigma}^2_e n_j (j\pi)^2} \sum_{t=j}^{n} \sum_{s=j}^{n} \widehat{e}_t(\boldsymbol{\theta}_n) \widehat{e}_s(\boldsymbol{\theta}_n) \exp(-0.5 \cdot (Y_{t-j} - Y_{s-j})^2),$$

where we have considered as W the empirical cdf $F_n(\cdot)$ based on $\{Y_{t-1}\}_{t=1}^{n}$ and the cdf of a standard normal random variable, Φ say, respectively. The results with other weighting functions W in $D^2_{n,C}$ are similar, see Hong (1999) and Hong and Lee (2004) who documented a similar situation. These are representatives of the CvM tests based on a smooth and non-smooth weighting function, which are the most used in the literature.

In an $ARMA(p, q)$ framework Ljung and Box (1978) proposed a diagnostic test based on the classical Portmantau's statistic

$$LB_m = n(n+2) \sum_{j=1}^{m} (n-j)^{-1} \widehat{\rho}^2_{e,j},$$

where $\widehat{\rho}_{e,j}$ is the residual autocorrelation coefficient at lag j from the $ARMA(p, q)$ fitted model. Under i.i.d errors and H_0 the asymptotic distribution of LB_m can be approximated by a χ^2_{m-p-q} distribution ($m > p + q$).

Hong and Lee (2004) have proposed a diagnostic test for the adequacy of a parametric conditional mean under possible conditional heteroskedasticity. Their test statistic is

$$HL_n(p) = \left[L^2_2(p) - \widehat{C}_1(p) \right] / \left[\sqrt{\widehat{D}_1(p)} \right], \tag{12}$$

where

$$L^2_2(p) = \sum_{j=1}^{n} n_j k^2 \left(\frac{j}{p} \right) \int_{\mathbb{R}} \left| \widehat{\gamma}^e_j(x, \boldsymbol{\theta}_n) \right|^2 W(dx), \tag{13}$$

$\widehat{\gamma}^e_j(x, \boldsymbol{\theta}_n) = n_j^{-1} \sum_{t=j}^{n} e_t(\boldsymbol{\theta}_n) \widehat{\psi}_{t-j}(x)$, $\widehat{\psi}_t(x) = \exp(ixe_t(\boldsymbol{\theta}_n)) - \widehat{\varphi}(x)$, $\widehat{\varphi}(x) = n^{-1} \sum_{t=1}^{n} \exp(ixe_t(\boldsymbol{\theta}_n))$, $k(\cdot)$ is a symmetric kernel, p is a bandwidth parameter and $W(\cdot)$ is an integrating function. 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.

In the more general case considered by Hong and Lee (2004) the centering and scaling factors are, respectively,

$$\widehat{C}_1(p) = \sum_{j=1}^{n} n_j^{-1} k^2\left(\frac{j}{p}\right) \sum_{t=j}^{n} \widehat{e}_t^2(\boldsymbol{\theta}_n) \int_{\mathbb{R}} \left|\widehat{\psi}_{t-j}(x)\right|^2 W(dx),$$

and

$$\widehat{D}_1(p) = 2\sum_{j=1}^{n-1}\sum_{l=1}^{n-1} k^2\left(\frac{j}{p}\right) k^2\left(\frac{l}{p}\right) \int_{\mathbb{R}^2} \left| \frac{1}{n-\max(j,l)+1} \sum_{t=\max(j,l)}^{n} \widehat{e}_t^2(\boldsymbol{\theta}_n) \widehat{\psi}_{t-j}(x) \widehat{\psi}_{t-l}(y) \right|^2 W(dx)W(dy).$$

Under some assumptions and H_0, they showed that $HL_n(p)$ converges to a standard normal random variable. For the simulations for $HL_n(p)$ 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 (2004).

In order to compare the pairwise approach with the case of considering a fixed number of lags in the conditioning set, we examine the finite sample properties of the CvM test of Bierens (1982) and the CvM and Kolmogorov-Smirnov (KS) tests of a multivariate version of Koul and Stute (1999) proposed in Escanciano (2004). The Bierens' test is based on the weighting function $w(\mathbf{I}_{t-1}^d, \mathbf{x}) = \exp(i\mathbf{x}'\mathbf{I}_{t-1}^d)$ and the cdf of a multivariate standard normal random vector as the integrating measure, where $\mathbf{I}_{t-1}^d = (Y_{t-1}, ..., Y_{t-d})'$ is the d-lagged values of the series. Under this setup Bierens' test is

$$CvM_{\exp,d} = n^{-1}\widehat{\sigma}_e^{-2} \sum_{t=1}^{n}\sum_{s=1}^{n} e_t(\boldsymbol{\theta}_n) e_s(\boldsymbol{\theta}_n) \exp(-\frac{1}{2}\left|\mathbf{I}_{t-1}^d - \mathbf{I}_{s-1}^d\right|^2).$$

Escanciano (2004) has recently proposed some diagnostic tests based on residual marked empirical processes. He justifies theoretically the FDWB approximation for a large class of residual marked tests, including the Bierens' test and the multivariate extension of Koul and Stute (1999) as special cases. We denote by CvM_d and KS_d the CvM and KS statistics proposed in Escanciano (2004), which are, respectively,

$$CvM_d = \frac{1}{\widehat{\sigma}_e^2 n^2} \sum_{j=1}^{n} \left[\sum_{t=1}^{n} \widehat{e}_t(\boldsymbol{\theta}_n) 1(\mathbf{I}_{t-1}^d \leq \mathbf{I}_{j-1}^d) \right]^2$$

and

$$KS_d = \max_{1 \leq i \leq n} \left| \frac{1}{\widehat{\sigma}_e \sqrt{n}} \sum_{t=1}^{n} \widehat{e}_t(\boldsymbol{\theta}_n) 1(\mathbf{I}_{t-1}^d \leq \mathbf{I}_{i-1}^d) \right|.$$

Throughout ε_t and v_t are independent sequences of $i.i.d.$ $N(0,1)$. We have considered the empirical level at the 5% and samples sizes $n = 100$ and $n = 300$. The results with other significance levels are similar. For the sake of space, we only present in simulations the case $n = 100$. The number of Monte Carlo experiments is 1000 and the number of bootstrap replications is $B = 500$. In all the replications 200 pre-sample data values of the processes were generated and discarded. Random numbers were generated using IMSL ggnml subroutine. We employ a sequence $\{V_t\}$ of i.i.d Bernoulli variates satisfying $P(V_t = 0.5(1-\sqrt{5})) = b$ and $P(V_t = 0.5(1+\sqrt{5})) = 1-b$, with $b = (1+\sqrt{5})/2\sqrt{5}$. The power in the non-bootstrap cases is level-adjusted by using the empirical values obtained under 5000 simulations of Model 1 below, although the difference is not substantial.

We use the FDWB approximation described before in our tests $D^2_{n,I}$ and $D^2_{n,C}$, and in $CvM_{\exp,d}$, CvM_d and KS_d, see Escanciano (2004). To examine the impact of the different parameters on the tests we consider p from 2 to 11 in $HL_n(p)$, m from 2 to 11 in LB_m, and $d = 1, 3, 5, 7, 9$ and 11 in $CvM_{\exp,d}$, CvM_d and KS_d. To simplify notation, we use in Figures the names D_i, $D_{\exp}$, $CvM_{\exp}$, CvM, KS, HL and LB to denote $D^2_{n,I}$, $D^2_{n,C}$, $CvM_{\exp,d}$, CvM_d, KS_d, $HL_n(p)$ and LB_m, respectively. The X-axes in all figures corresponds with the values of m, p and d.

Our null model is an AR(1) model: $Y_t = a + bY_{t-1} + \varepsilon_t$. We examine the adequacy of this model under the following DGP:

1. AR(1) model: $Y_t = 0.6Y_{t-1} + \varepsilon_t$.
2. AR(1) model with exponential centered noise $\epsilon_t \sim Exp(1) : Y_t = 0.6Y_{t-1} + \epsilon_t$.
3. AR(1) model with heteroskedasticity (ARHET): $Y_t = 0.6Y_{t-1} + h_t\varepsilon_t$; $h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.6Y_{t-2}^2$.
4. AR(1) model plus a bilinear term (AR-BIL): $Y_t = 0.6Y_{t-1} + 0.4Y_{t-1}\varepsilon_t + \varepsilon_t$.
5. AR(2) model: $Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$.
6. ARMA(1,1) model: $Y_t = 0.6Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$.
7. Bilinear model (BIL): $Y_t = 0.6Y_{t-1} + 0.7\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$.
8. Nonlinear moving average model (NLMA): $Y_t = 0.6Y_{t-1} + 0.7\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t$.
9. Threshold autoregressive model (TAR): $Y_t = 0.6Y_{t-1} + \varepsilon_t$ if $Y_{t-1} < 1$ and $Y_t = -0.5Y_{t-1} + \varepsilon_t$ if $Y_{t-1} \geq 1$.
10. Sign autoregressive model (SIGN): $Y_t = sign(Y_{t-1}) + 0.43\varepsilon_t$, where $sign(x) = 1(x > 0) - 1(x < 0)$.
11. Temp Map model (TEM MAP): $Y_t = \alpha^{-1}Y_{t-1}$ if $0 \leq Y_{t-1} < \alpha$ and $Y_t = (1-\alpha)^{-1}(1 - Y_{t-1})$ if $\alpha \leq Y_{t-1} \leq 1$, where $\alpha = 0.49999$ and Y_0 is generated from the uniform distribution on [0,1].
12. Nonlinear autoregressive model (NAR): $Y_t = 0.6Y_{t-1} + 0.7sin(0.3\pi Y_{t-2}) + \varepsilon_t$.

Models 1 and 5 to 11 have been considered in Hong and Lee (2003) and are well described there. Models 2 to 4 are introduced to examine the empirical size of tests. Model 12 is introduced to compare indicator and exponential functions, see e.g. Eubank and Hart (1992).

To compute all the tests statistics we consider the usual least squares residuals from model 1, i.e., $\widehat{e}_t(\boldsymbol{\theta}_n) = Y_t - \widehat{a} - \widehat{b}Y_{t-1}$ for $t = 1, ..., n$. In Figures 1 and 2 we report the empirical rejections probabilities (RP) associated with the models 1 and 5 to 11 to examine the empirical level and power of the tests. The tests statistics D_i, $D_{\exp}$, CvM, KS show good empirical level properties. The empirical size of $CvM_{\exp}$ decreases as d increases, this is a general property of this test in all simulations and might be due to numerical problems with the characteristic weighting function, that is, because the distance $\left|\mathbf{I}_{t-1}^d - \mathbf{I}_{s-1}^d\right|^2$ increases very fast with d, and hence, the weights are very near to zero when d is relatively large. Hong and Lee's (2004) test HL presents some underrejection for all null models. A similar situation is reported in Hong and Lee (2004). The empirical level of LB is more or less satisfactory except in the heteroskedastic model ARHET. This is a well-known result and to solve this problem one can robustify the Ljung and Box (1978) test, see e.g. Deo (2000).

Our tests D_i and $D_{\exp}$ have excellent empirical power against the AR(2), TAR, SIGN, TEM MAP and NAR models, and moderate empirical power against ARMA(1,1), BIL and NLMA models. Neither D_i nor $D_{\exp}$ outperforms the other. The Cramér-von Mises and Kolmogorov-Smirnov statistics CvM and KS present similar behavior with decreasing empirical power for large values of d, as expected, given the sparseness of the data in high dimensional spaces. This is the case also for $CvM_{\exp}$ but with more variation with respect to d. In general, $CvM_{\exp}$ presents better empirical power properties than CvM and KS. For AR(2), ARMA(1,1) and NAR models their maximum power is achieved at $d = 3$, whereas for the other models is at $d = 1$. Their power is very sensitive to d, see e.g. AR(2) model. Hong and Lee's (2004) test HL has good empirical power against AR(2) and ARMA(1,1), and moderate power against the rest of models. It is very sensitive on p for AR(2) and NAR models, and roughly stable for the other models. Ljung and Box's (1978) test LB has excellent empirical power against the linear models AR(2) and ARMA(1,1), and cannot detect some nonlinear alternatives such as the BIL, NLMA, TAR, and TEM MAP alternatives.

Generally speaking, D_i and $D_{\exp}$ have omnibus power against all linear and nonlinear alternatives, for TAR, SIGN, TEM MAP and NAR models achieve the best empirical power overall. Moreover, their size behavior is robust to different models with the same specified conditional mean. From the present simulations and other considered in a previous version of this paper we conclude that, in general, the pairwise approach is better than the approach based on a fixed number of lags, in some cases uniformly in d. The latter approach, gives reasonable empirical power properties if the lag order d is chosen appropriately, but is very sensitive to this choice. In addition, the pairwise approach avoids the choice of lag order parameters and overcomes the problem of the curse of dimensionality, which affects the tests when d is large or even moderate.

Now, we consider an application to model a weekly egg prices series of a German agricultural market between April 1967 and May 1990. This series is of length 300 and is the first quarter of a longer series extensively analyzed by Finkenstädt (1995), and also studied in Fan and Yao (2003, p. 113-117), where different linear models have been proposed. Using different criteria these authors choose a linear model $ARMA(1,2)$ for the differenced series, X_t say, where $X_t = Y_t - Y_{t-1}$, and estimate the model

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}. \tag{14}$$

As in Fan and Yao (2003), we subtract the sample mean from the data before the fitting. Our estimations using the NLSE for $(\varphi_1, \theta_1, \theta_2)$ are respectively, $(0.888, 0.594, 0.381)$. The Ljung and Box's (1978) test LB_m shows that there is no linear dependence structure in the residuals with a sequence of p-values that attain the minimum value of 0.080 at lag $m = 7$. The empirical p-values obtained for $D^2_{n,I}$, $D^2_{n,C}$, $CvM_{\exp,2}$, CvM_2 and KS_2 are 0.046, 0.004, 0.942, 0.872 and 0.956, respectively. For values of d larger than 2 the behaviour of $CvM_{\exp,d}$, CvM_d and KS_d is similar. Therefore, we observe that although the tests based on a fixed number of lags fail to reject the $ARMA(1,2)$ model, our new pairwise tests are able to reject it. These findings may be due to the present of high order subtle dependence structure that the pairwise tests are able to detect because they consider information from all lags contained in the sample.

6. SUMMARY AND CONCLUSIONS

In this paper we have presented goodness-of-fit tests for linear and nonlinear time series models using a generalized spectral distribution approach very convenient when the conditioning set is infinite-dimensional. We think that our tests provide a good compromise between generality, simplicity and feasibility. The present paper, jointly with the works by Hong and his coauthors, shows that the generalized spectral approach can be a useful tool for studying serial dependence in a time series framework. Here, we provide an alternative proposal to Hong and we extend the generalized spectral approach to other weighting functions including but not restricting to exponential functions. We use Hilbert space methods which allow us to elaborate a unified asymptotic theory for the tests statistics. The smoothing approach used by Hong may give more flexibility in the weighting scheme, compare (13) and (9). However, the choice of a kernel and a smoothing parameter affects the inferences in finite samples, as we have shown in the extensive simulation experiment. On the other hand, the range of possibilities in the choice of w and W gives flexibility for $D^2_{n,w}$ in directing the power against some desired directions. We would like to stress here that although $D^2_{n,w}$ has the attractive convenience of being free of choosing any smoothing parameter or kernel, should be viewed as not competing but as a complement to the smoothing kernel density approach.

Among the appealing properties of our tests are that; first they are free of choosing any lag order or smoothing parameters; second, they present excellent empirical size and power properties, as is shown in the extensive simulation experiment; third, they avoid the curse of dimensionality problem and high dimensional integration that affects other tests proposed in the literature; and four, they are robust to higher order dependence, which is of crucial importance in conditional moments modeling. The price to pay for such good properties is that our tests are not consistent against all alternatives because we check only pairwise implications of the correct specification, that the weights in our tests are $1/(j\pi)^2$, which heavily downweigh the contribution of $\widehat{\gamma}_{j,w}(\mathbf{x}, \boldsymbol{\theta}_n)$ to detect a lack of fit in the conditional mean, and that under H_0 the asymptotic null distribution depends on the DGP, so bootstrap procedures have to be used. To solve the first problem, one possibility is to apply the spectral methodology proposed here to the measures

$$\gamma_{jh}(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_0) = E[e_t(\boldsymbol{\theta}_0) w(\mathbf{Z}_{t-j}, \mathbf{x}, \mathbf{Z}_{t-h}, \mathbf{y})], \; j \geq 1, h \geq 1,$$

where, for example, $w(\mathbf{Z}_{t-j}, \mathbf{x}, \mathbf{Z}_{t-h}, \mathbf{y}) = 1(\mathbf{Z}_{t-j} \leq \mathbf{x}, \mathbf{Z}_{t-h} \leq \mathbf{y})$, and consider a generalized bispectrum approach, i.e., the Fourier transform of the double sequence $\{\gamma_{jh}\}$. For the second problem, our weights represent, in some sense, the price to pay for considering all lags contained in the sample and avoiding the choice of lag-bandwidth parameters. Once these two properties are sacrificed there are many possibilities to avoid such weights. For instance, one can consider Portmanteau-type tests

$$\sum_{j=1}^{m} a_{n,j} \int_{\Upsilon} \left| \widehat{\gamma}_{j,w}(\mathbf{x}, \boldsymbol{\theta}_n) \right|^2 W(d\mathbf{x}),$$

for a fixed $m \in \mathbb{N}$ and some choice of $\{a_{n,j}\}_{j=1}^{m}$. One can argue that our tests naturally discount higher order lags, which is consistent with the stylized fact that many underlying real variables are more affected by the recent past events than by the remote past events. The latter problem is much more difficult to solve without altering the other appealing properties of the tests. Other alternatives proposed in the literature such as the martingale transformation used in Koul and Stute (1999), cf. Khamaladze (1981), are difficult in our context. The main reason is that, unlike in Koul and Stute (1999), the dependence structure of the regressors plays a crucial role in the covariance operator of our null limit process. To end this section, we would like to stress the generality of the approach proposed here. Many important testing problems that arise in time series can be expressed as conditional moment restrictions. Higher conditional moments modeling, e.g conditional variance, or testing for conditional symmetry are examples. These hypotheses are fundamental in financial and economic applications and can be dealt with similar techniques to those proposed here.

APPENDIX: PROOFS

Proof of Lemma 1: Let $m(\mathbf{y}) = E[e_t(\boldsymbol{\theta}_0) \mid \mathbf{I}_{t-1} = \mathbf{y}]$ a.s.. Let define the finite Borel measure

$$\nu(B) = \int_B m(\mathbf{y}) P_I(d\mathbf{y}),$$

where P_I is the probability measure associated to $\mathbf{I}_{t-1}$ and B is a Borel set of $\mathbb{R}^\infty$, i.e., $B \in \mathcal{B}$ say. The proof of (i) and (ii) is a direct application of Lemma 1.3.12 (b) in van der Vaart and Wellner (1996), after noting that the argument used there can be extended to any finite measure and not only a probability measure. Hence, we deduce that $\nu(B) = 0$, $\forall B \in \mathcal{B}$, and then $m = 0$ a.s.. The proof for (iii) follows from Theorem 2.3 in Stinchcombe and White (1998). The proof for (iv) follows because if $\nu(B) = 0$ for all Borel sets B in a separating class, then it holds for all Borel sets, by definition.

Lemma A1: Under A4 and A5 the effect of estimating the conditioning set $\mathbf{I}_{t-1}$ by $\widehat{\mathbf{I}}_{t-1}$ in $f(\mathbf{I}_{t-1}, \boldsymbol{\theta}_n)$ has no effect on the asymptotic theory. That is,

$$\left\| S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) - \widetilde{S}_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) \right\|^2 \xrightarrow{P} 0.$$

where $\widetilde{S}_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ is the same process as $S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ but with $\mathbf{I}_{t-1}$ replacing $\widehat{\mathbf{I}}_{t-1}$.

Proof of Lemma A1: Write

$$\begin{aligned}
& E \left\| S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) - \widetilde{S}_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) \right\|^2 \\
= & \sum_{j=1}^{n} \frac{1}{(j\pi)^2} n_j^{-1} E \left(\sum_{t=j}^{n} (f(\mathbf{I}_{t-1}, \boldsymbol{\theta}_n) - f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}_n)) w_{t-j}(\mathbf{x}) \right)^2 \\
\leq & C \sum_{j=1}^{n} \frac{1}{(j\pi)^2} n_j^{-1} \left(\sum_{t=j}^{n} \left(E \sup_{\boldsymbol{\theta} \in \Theta} (f(\mathbf{I}_{t-1}, \boldsymbol{\theta}) - f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}))^2 \right)^{1/2} \right)^2 \\
\leq & C \left(\sum_{t=1}^{n} \left(E \sup_{\boldsymbol{\theta} \in \Theta} (f(\mathbf{I}_{t-1}, \boldsymbol{\theta}) - f(\widehat{\mathbf{I}}_{t-1}, \boldsymbol{\theta}))^2 \right)^{1/2} \right)^2 \left(\sum_{j=1}^{n} \frac{1}{(j\pi)^2} n_j^{-1} \right),
\end{aligned}$$

where the first inequality is due to the Minkowski's inequality. Then, the Lemma follows from A5. For simplicity, we rename $\widetilde{S}_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ again as $S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$. The next Lemma establishes the asymptotic expansion of the process $S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ under the null.

Lemma A2: *Under (2) and the assumptions A1-A5,*

$$\| S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) - S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_0) + \mathbf{G}_w'(\boldsymbol{\eta}, \boldsymbol{\theta}_0) \mathbf{V} \|^2 \xrightarrow{P} 0.$$

Proof of Lemma A2: By the Mean Value Theorem and A1-A5

$$S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) = S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_0) + \frac{\partial S_{n,w}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0), \tag{15}$$

where $\widetilde{\boldsymbol{\theta}}_n$ is a mean value satisfying $\left|\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right| \leq |\boldsymbol{\theta}_n - \boldsymbol{\theta}_0|$ a.s.. Note that the process $S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n)$ can be written as

$$S_{n,w}(\boldsymbol{\eta}, \boldsymbol{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\boldsymbol{\theta}_n)\sum_{j=1}^{t} n^{1/2} n_j^{-1/2} w_{t-j}(\mathbf{x})\frac{\sqrt{2}\sin j\pi\lambda}{j\pi} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\boldsymbol{\theta}_n) Q_{t,w}(\boldsymbol{\eta}),$$

where $Q_{t,w}(\boldsymbol{\eta})$ is implicitly define. Hence,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial S_{n,w}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} &= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial e_t(\widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} Q_{t,w}(\boldsymbol{\eta}) \\
&= -\sum_{j=1}^{n}\frac{1}{n}\sum_{t=j}^{n} n^{1/2} n_j^{-1/2}\mathbf{g}_t(\widetilde{\boldsymbol{\theta}}_n) w_{t-j}(\mathbf{x})\Psi_j(\lambda) \\
&= -\sum_{j=1}^{n}\mathbf{b}_{j,n}(\mathbf{x}, \widetilde{\boldsymbol{\theta}}_n)\Psi_j(\lambda),
\end{aligned}$$

where $\mathbf{b}_{j,n}(\mathbf{x}, \widetilde{\boldsymbol{\theta}}_n) = n^{-1}\sum_{t=j}^{n} n^{1/2} n_j^{-1/2}\mathbf{g}_t(\widetilde{\boldsymbol{\theta}}_n) w_{t-j}(\mathbf{x})$. Hence, Assumptions A1-A5, the uniform argument of Jennrich (1969, Theorem 2) and applying Lemma 1 in Escanciano and Velasco (2003, hereafter EV) yield

$$\left\| \frac{1}{\sqrt{n}}\frac{\partial S_{n,w}(\boldsymbol{\eta}, \widetilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} + \sum_{j=1}^{n}\mathbf{b}_j(\mathbf{x}, \boldsymbol{\theta}_0)\Psi_j(\lambda) \right\| \xrightarrow{P} 0.$$

The last display, Assumption A3, Theorem 1 in EV and (15) imply the result.

Proof of Theorem 1: We apply Lemma A2 and Theorem 1 in EV but with $w_{t-j}(\mathbf{x})$ replacing $exp(ixY_{t-j})$ there.

Proof of Corollary 1: By A5, Theorem 1 and the Continuous Mapping Theorem (see e.g. Billingsley 1999) the result holds.

Proof of Corollary 2: Note that

$$\int_{\Pi} (S_{n,I}(\boldsymbol{\eta}, \boldsymbol{\theta}_n))^2\{F_n(d\mathbf{x}) - F_n(d\mathbf{x})\}d\lambda = \sum_{j=1}^{n}\frac{1}{(j\pi)^2}\int_{\Upsilon}(n_j^{1/2}\widehat{\gamma}_{j,I}(\mathbf{x}, \boldsymbol{\theta}_n))^2\{F_n(d\mathbf{x}) - F(d\mathbf{x})\}.$$

From Corollary 1 in Escanciano (2004), using the continuity of $F(\mathbf{x})$, we have that $n_j^{1/2}\widehat{\gamma}_{j,I}(\mathbf{x}, \boldsymbol{\theta}_n)$ is tight in $\ell^\infty(\Upsilon_c)$, the metric space of all uniformly bounded real-valued functions on Υ_c endowed with the supremum norm. Also we have the Glivenko-Cantelli's Theorem for stationary and ergodic sequences, see Dehling and Philipp (2002, p. 4). Hence, applying Lemma 3.1 in Chang (1990), we can conclude that

$$\int_{\Upsilon}(n_j^{1/2}\widehat{\gamma}_{j,I}(\mathbf{x}, \boldsymbol{\theta}_n))^2\{F_n(d\mathbf{x}) - F(d\mathbf{x})\} \xrightarrow{P} 0, \ \forall j \geq 1.$$

Applying a partition argument as in the proof of Lemma 1 in EV the corollary is proved.

Proof of Theorem 2: By Lemma 1 in EV it is sufficient to consider elements of the form

$$\widehat{\gamma}_{j,w}(\mathbf{x}, \boldsymbol{\theta}_n) = \frac{1}{n_j}\sum_{t=j}^{n} e_t(\boldsymbol{\theta}_n) w_{t-j}(\mathbf{x}) = \frac{1}{n_j}\sum_{t=j}^{n}\{\varepsilon_t + a_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n)\} w_{t-j}(\mathbf{x}), \ j \geq 1.$$

We apply a linearization argument for $\{f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n)\}$ as in Theorem 1 and use A4 to obtain

$$\sup_{\mathbf{x}\in\Upsilon_c}\left|\frac{1}{n_j}\sum_{t=j}^{n}\{\varepsilon_t + a_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n)\}w_{t-j}(\mathbf{x}) - E[a_t w_{t-j}(\mathbf{x})]\right| = o_P(1),\ j \geq 1.$$

Proof of Theorem 3: Again, we write for fix $j \geq 1$

$$\begin{aligned}\widehat{\gamma}_{j,w}(\mathbf{x},\boldsymbol{\theta}_n) &= \frac{1}{n_j}\sum_{t=j}^{n}\{\varepsilon_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n) + \frac{a_t}{\sqrt{n}}\}w_{t-j}(\mathbf{x}) \\ &= \frac{1}{n_j}\sum_{t=j}^{n}\{\varepsilon_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n)\}w_{t-j}(\mathbf{x}) + \frac{1}{n_j}\sum_{t=j}^{n}\frac{a_t}{\sqrt{n}}w_{t-j}(\mathbf{x}) = \widehat{\gamma}_{1j}(\mathbf{x},\boldsymbol{\theta}_n) + \widehat{\gamma}_{2j}(\mathbf{x},\boldsymbol{\theta}_n),\end{aligned}$$

where by $\widehat{\gamma}_{1j}(\mathbf{x},\boldsymbol{\theta}_n)$ and $\widehat{\gamma}_{2j}(\mathbf{x},\boldsymbol{\theta}_n)$ are implicitly defined. Let use the same arguments as in Theorem 1 for $\widehat{\gamma}_{1j}(\mathbf{x},\boldsymbol{\theta}_n)$, but now with A7, to show that

$$\sup_{\mathbf{x}\in\Upsilon_c}\left|\frac{1}{n_j}\sum_{t=j}^{n}\{\varepsilon_t + f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}_n)\}w_{t-j}(\mathbf{x}) - \boldsymbol{\xi}_a'\mathbf{b}_j(\mathbf{x},\boldsymbol{\theta}_0)\right| = o_P(1),\ j \geq 1.$$

Also, from A4

$$\left|n_j^{1/2}\widehat{\gamma}_{2j}(\mathbf{x},\boldsymbol{\theta}_n) - E[a_t w_{t-j}(\mathbf{x})]\right| = o_P(1),\ j \geq 1,$$

uniformly in $\mathbf{x} \in \Upsilon_c$. Then, the latter displays jointly with Lemma 1 in EV yield the result.

Proof of Theorem 4: We write similarly to Lemma A2

$$\begin{aligned}S_{n,w}^*(\boldsymbol{\eta}) &= n^{-1/2}\sum_{t=1}^{n}e_t^*(\boldsymbol{\theta}_n)Q_{t,w}(\boldsymbol{\eta}) - n^{-1/2}\sum_{t=1}^{n}\{f(\mathbf{I}_{t-1},\boldsymbol{\theta}_n^*) - f(\mathbf{I}_{t-1},\boldsymbol{\theta}_n)\}Q_{t,w}(\boldsymbol{\eta}) \\ &= S_{n,w,L}^*(\boldsymbol{\eta}) - I^* - II^* - III^*,\end{aligned}$$

where

$$I^* = n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)n^{-1}\sum_{t=1}^{n}\{\mathbf{g}(\mathbf{I}_{t-1},\widetilde{\boldsymbol{\theta}}_n^*) - \mathbf{g}(\mathbf{I}_{t-1},\boldsymbol{\theta}_n)\}Q_{t,w}(\boldsymbol{\eta}),$$

$$II^* = n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)n^{-1}\sum_{t=1}^{n}[\mathbf{g}(\mathbf{I}_{t-1},\boldsymbol{\theta}_n)Q_{t,w}(\boldsymbol{\eta}) - \mathbf{G}_w'(\boldsymbol{\eta},\boldsymbol{\theta}_*)]$$

and $III^* = n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)\mathbf{G}_w'(\boldsymbol{\eta},\boldsymbol{\theta}_*)$, with $\widetilde{\boldsymbol{\theta}}_n^*$ satisfying $\left|\widetilde{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_n\right| \leq |\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n|$ a.s.(conditionally on the sample). Under our assumptions is easy to show that conditionally on the sample, $\|I^*\| = o_P(1)$ and $\|II^*\| = o_P(1)$ with probability one. Therefore, in $L_2(\Pi,\nu)$

$$S_{n,w}^*(\boldsymbol{\eta}) = S_{n,w,L}^*(\boldsymbol{\eta}) - n^{1/2}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n)'\mathbf{G}_w'(\boldsymbol{\eta},\boldsymbol{\theta}_*) + o_P(1)\ a.s..$$

The convergence of the finite-dimensional distributions follows from the last expression, A7, Theorem 1 in EV and from the Cramér-Wold device. The tightness (a.s.) follows from Theorem 2.5.2 in van der Vaart and Wellner (1996). The proof is finished.

REFERENCES

Bierens, H. J. (1982), "Consistent Model Specification Tests," *Journal of Econometrics,* 20, 105-134.

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

Billingsley, P. (1999), *Convergence of Probability Measures.* Second Edition. Wiley, New York.

Box, G., and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autorregressive Integrated Moving Average Time Series Models," *Journal of American Statistical Association,* 65, 1509-1527.

Brock, W., Dechert, W.D., and Sheinkman, J. (1987), "A Test for Independence Based on the Correlation Dimension," unpublished manuscript. Madison, University of Wisconsin.

Chang, N. M. (1990), "Weak convergence of a self-consistent estimator of a survival function with doubly censored data," *Annals of Statistics,* 18, 391-404.

de Jong, R. M. (1996), "The Bierens' Tets Under Data Dependence," *Journal of Econometrics,* 72, 1-32.

Dehling, H., and Philipp, W. (2002), "Empirical Process Techniques for Dependent Data". In H. Dehling, T. Mikosch and M. Sorensen (Eds.) *Empirical processes techniques for dependent data*, Birkhauser Boston, p. 1-115.

Delgado M., Hidalgo J., and Velasco, C. (2003), "Distribution Free Goodness-of-Fit for Linear Models," Forthcoming in the *Annals of Statistics.*

Deo, R. S. (2000), "Spectral Tests of the Martingale Hypothesis under Conditional Heteroskedasticity," *Journal of Econometrics,* 99, 291-315.

Escanciano, J. C. (2004), "Model Checks Using Residual Marked Empirical Processes," Working Paper n^o 13/04 Facultad de Económicas y Empresariales, Universidad de Navarra.

Escanciano, J. C. and Velasco, C. (2003), "Generalized Spectral Tests for the Martingale Difference Hypothesis," preprint.

Eubank, R., and Hart, J. (1992), "Testing goodness-of-fit in regression via order selection criteria," *Annals of Statistics,* 20, 1412-1425.

Fan, J., and Huang, L. (2001), "Goodness-of-fit tests for parametric regression models," *Journal of the American Statistical Association,* 96, 640-652.

Fan, J., and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York.

Franke, J., Kreiss, J. P., and Mammen, E. (2002), "Bootstrap of Kernel Smoothing in Nonlinear Time Series," *Bernoulli,* 8, 1-37.

Finkenstädt, B. (1995), *Nonlinear Dynamics in Economics.* Springer, Berlin.

Ginè, E., and J. Zinn (1990), "Bootstraping General Empirical Measures," *Annals of Probability,* 18, 851-869.

Hall, P., and Heyde, C.C. (1980), *Martingale Limit Theorems and Its Application.* Academic Press, New York.

Härdle, W., Horowitz, J. and Kreiss, J. P. (2003), "Bootstrap Methods for Time Series," *International Statistical Review,* 71, 435-459.

Härdle, W., and Mammen, E. (1993), "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics,* 21, 1926-1974.

Heyde, C. C. (1997), *Quasi-Likelihood and its applications.* Springer, Berlin.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function," *Journal of American Statistical Association,* 84, 1201-1220.

Hong, Y., and Lee, T.H. (2003), "Diagnostic Checking for Adequacy of Nonlinear Time Series Models," *Econometric Theory*, 19, 1065-1121.

Hong, Y., and Lee, Y.J. (2004), "Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form," Forthcoming in *Review of Economic Studies.*

Jakubowski, A. (1980), "On Limit Theorems for Sums of Dependent Hilbert Space Valued Random Variables," in *Lecture Notes in Statistics*, 2, 178-187.

Jennrich, R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics,* 40, 633-643.

Khamaladze, E. V. (1981), "Martingale approach in the theory of Goodness-of-fit tests," *Theory of Probability and its Applications* 26, 240-257.

Koul, H. L. (2002), *Weighted Empirical Processes in Dynamic Nonlinear Models.* 2nd ed, Lecture Notes in Statistics, Vol. 166, Springer.

Koul H.L., and Stute W. (1999), "Nonparametric Model Checks for Time Series," *Annals of Statistics,* 27, 204-236.

Li, Q., Hsiao, C., and 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," *Annals of Statistics,* 16, 1696-1708.

Ljung, G.M., and Box, G.E.P. (1978), "A Measure of Lack of Fit in Time Series Models," *Biometrika,* 65, 297-303.

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

Paparoditis, E. (2000), "Spectral Density Based Goodness-of-Fit Tests for Time Series Models," *Scandinavian Journal of Statistics,* 27, 143-176.

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

Peña, D. and Rodriguez, J. (2001), "A Powerful Portmanteau Test of Lack of Fit for Time Series," *Journal of the American Statistical Association, 97, 601-610.*

Politis, D. and J. Romano (1994), "Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap," *Statistica Sinica,* 4, 461-476.

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

Stute, W. (1997), "Nonparametic Model Checks for Regression," *Annals of Statistics,* 25, 613-641.

Stute, W., Gonzalez-Manteiga, W., and Presedo-Quindimil, M. (1998), "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association,* 93, 141-149.

Tong, H. (1990), *Nonlinear Time Series, A Dynamic system Approach.* Clarendon Press, Oxford.

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

Wu, C.F.J. (1986), "Jacknife, Bootstrap and other Resampling Methods in Regression Analysis (with discussion)," *Annals of Statistics,* 14, 1261-1350

Figure 1. Size and Power at 5%

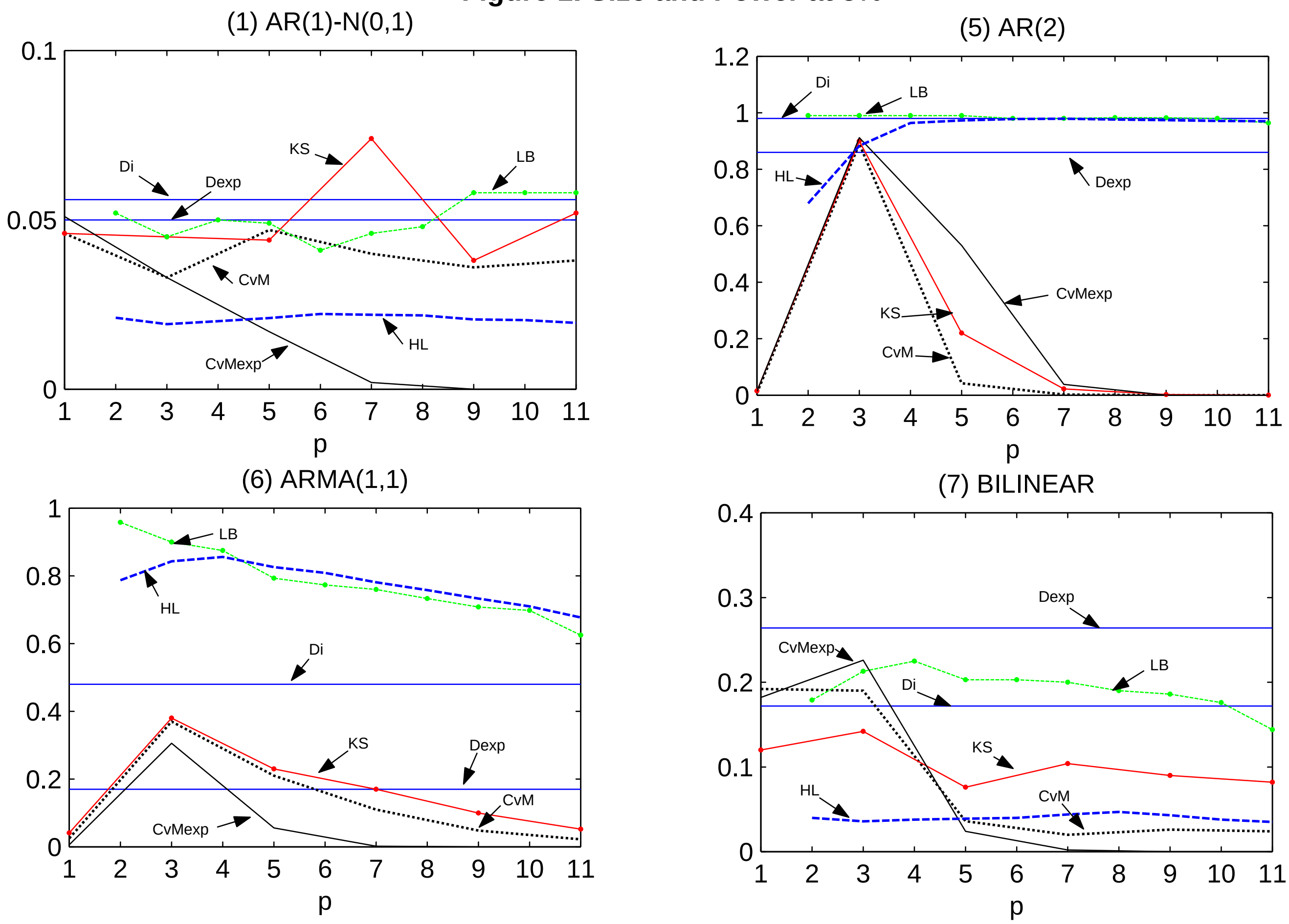

Figure 2. Size and Power at 5%