Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 02/06

Joint Diagnostic Tests for Conditional Mean and Variance Specifications

J. Carlos Escanciano

Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Joint Diagnostic Tests for Conditional Mean and Variance Specifications
J. Carlos Escanciano
Working Paper No.02/06
February 2006
JEL No. C12, C14, C52.

ABSTRACT
This article proposes a general class of joint diagnostic tests for parametric conditional mean and variance models of possibly nonlinear and/or non-Markovian time series sequences. The new tests are based on a generalized spectral approach and, contrary to existing procedures, they do not need to choose a lag order depending on the sample size or to smooth the data. Moreover, they are robust to higher order dependence of unknown form. It turns out that the asymptotic null distributions of the new tests depend on the data generating process, so a bootstrap procedure is proposed and theoretically justified. 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 modelling. An application to the S&P500 highlights the merits of our approach.

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

ACKNOWLEDGEMENTS

Parts of this paper were written while I was visiting the Cowles Foundation at Yale University and the Economics Department at Cornell University, whose hospitality is gratefully acknowledged. Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001.

JOINT DIAGNOSTIC TESTS FOR CONDITIONAL MEAN AND VARIANCE SPECIFICATIONS[*†]

J. Carlos Escanciano[‡]

Universidad de Navarra

February 14, 2006

Abstract

This article proposes a general class of joint diagnostic tests for parametric conditional mean and variance models of possibly nonlinear and/or non-Markovian time series sequences. The new tests are based on a generalized spectral approach and, contrary to existing procedures, they do not need to choose a lag order depending on the sample size or to smooth the data. Moreover, they are robust to higher order dependence of unknown form. It turns out that the asymptotic null distributions of the new tests depend on the data generating process, so a bootstrap procedure is proposed and theoretically justified. 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 modelling. An application to the S&P500 highlights the merits of our approach.

[*]*JEL Classification:* C12, C14, C52.

Key words and phrases. Specification tests; Model adequacy; Generalized spectral analysis; Nonlinear time series; Wild bootstrap.

[†]Parts of this paper were written while I was visiting the Cowles Foundation at Yale University and the Economics Department at Cornell University, whose hospitality is gratefully acknowledged. Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001.

[‡]Corresponding address: Universidad de Navarra, Facultad de Económicas, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain, e-mail: jescanci@unav.es.

1. INTRODUCTION

The last decades have been witnessed of an increase interest in time series modelling. Much of the existing econometric and statistical literature has been concerned with the parametric time series modelling in terms of the conditional mean and variance functions given some information set. A large body of this literature has been devoted to the estimation of the parameters in such models and the associated inferences. On the contrary, the problem of testing the correct joint specification of the conditional mean and variance functions has been less elaborated. In this paper we consider this problem when the information set at time t is infinite-dimensional, thereby allowing for Markovian and non-Markovian time series sequences in the test procedure.

More precisely, let $\{(Y_t, Z'_{t-1})'\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$, where the real random variable (r.v.) Y_t is the dependent (predicted) variable and $Z_{t-1} = (Y_{t-1}, X'_{t-1})' \in \mathbb{R}^m$, $m \in \mathbb{N}$, is the explanatory random vector containing the lagged value of the dependent variable and other explanatory variables X_{t-1}, say. In this paper we are mainly concerned with the case in which the conditioning set at time $t-1$ is given by $I_{t-1} = (Z'_{t-1}, Z'_{t-2}, ...)'$. It is known that under square-integrability of Y_t we can write the tautological expression

$$Y_t = f(I_{t-1}) + h(I_{t-1})u_t,$$

where $f(I_{t-1}) = E[Y_t \mid I_{t-1}]$ is almost surely (a.s.) the conditional mean and $h^2(I_{t-1}) = Var[Y_t \mid I_{t-1}]$ is a.s. the conditional variance. Then, in parametric modelling one assumes the existence of a parametric family of functions $\mathcal{M} = \{f(\cdot,\theta),\ h^2(\cdot,\theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ and considers the following model

$$Y_t = f(I_{t-1},\theta) + h(I_{t-1},\theta)u_t(\theta) \tag{1}$$

where $f(I_{t-1},\theta)$ and $h(I_{t-1},\theta)$ are parametric specifications for $f(I_{t-1})$ and $h(I_{t-1})$, respectively, and $\{u_t(\theta)\}$ is a sequence of disturbances of the model. The specification (1) covers the well-known linear ARMA-ARCH, ARMA-GARCH models as well as nonlinear conditional mean and variance models, see, e.g., Fan and Yao (2003). When $f(I_{t-1},\theta_0)$ and $h(I_{t-1},\theta_0)$ are correctly specified for $f(I_{t-1})$ and $h(I_{t-1})$, $\{u_t(\theta_0)\}$ will be a zero mean and unit conditional variance martingale difference sequence with respect to $\mathcal{F}_{t-1}$, the σ-field generated by I_{t-1}. That is, the correct joint specification is tantamount to

$$H_0 : E[e_{1t}(\theta_0) \mid I_{t-1}] = 0 \text{ a.s. and } E[e_{2t}(\theta_0) \mid I_{t-1}] = 0 \text{ a.s. for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \tag{2}$$

where $e_{1t}(\theta) = Y_t - f(I_{t-1},\theta)$ and $e_{2t}(\theta) = e_{1t}^2(\theta) - h^2(I_{t-1},\theta)$. The first conditional moment restriction (CMR) in H_0 is responsible for the correct specification of the conditional mean whereas both CMR's are necessary for the adequacy of the conditional variance.

The main goal of this paper is to test for H_0 when the information set is infinite-dimensional. This problem is of certain relevancy in econometrics practice and, in particular, in financial econometrics modelling where parametric models such as (1) are commonly used, see, e.g., Straumann (2005) for a recent reference. A lack of fit in the postulated conditional mean and/or variance can lead to misleading conclusions and statistical inferences, and to suboptimal point forecasts. Therefore, in order to prevent wrong conclusions, every statistical inference that 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 vast amount of literature on testing the correct specification of a parametric dynamic conditional mean model, see Escanciano (2005) for an up-to-date list of references. On the contrary, the literature on joint specification tests of conditional mean and variance functions is very scarce. The problem of testing simultaneously many CMR's in a time series framework has already been considered in, e.g., Li (1999) and Chen and Fan (1999), under mixing data, or in Delgado, Dominguez and Lavergne (2005) for independent data. Ngatchou-Wandji (2005) considered joint specification tests for parametric conditional mean and variance functions. This author proposed χ^2-discrepancy measures that although being simple, fail to be consistent against a large class of alternatives of the correct specification. Moreover, Ngatchou-Wandji's (2005) test involves the critical choice of some subsets of a Euclidean space without any guidance for this choice. Recently, Gao and King (2004) have extended the initial smooth-based approach of Härdle and Mammen (1993) to tests for joint specifications of conditional mean and variance functions.

An important limitation of the aforementioned articles is that they consider a finite-dimensional information set I_{t-1}, and hence, they are not suitable for testing (2) here. Moreover, even for the case in which the information set is of finite dimension d say, most of the proposed tests deliver a poor power performance when d is large or moderate, due to the so-called "curse of dimensionality" problem.

To consider an infinite-dimensional information set and as an alternative approach to previous literature, Hong (1999) has introduced a generalized spectral density as a new tool for testing interesting hypotheses in a nonlinear time series framework. Using Hong's (1999) approach, Hong and Lee (2003) have proposed a diagnostic test for conditional mean and variance specifications based on checking the serial independence between $u_t(\theta_0)$ and $u_{t-j}(\theta_0)$ at all lags. However, the independence assumption on standardized errors is in general more restrictive than the null hypothesis (2) and, in particular, it is possible that their test rejects a correct null model because of higher order dependence, incurring in an increase of the Type I error probability. Moreover, the i.i.d assumption on standardized errors may contrast with the now growing econometric literature documenting time-varying conditional skewness and kurtosis in economic and financial time series, see e.g. Gallant, Hsieh and Tauchen (1991), Hansen (1994), Harvey and Siddque (1999, 2000) or

Jondeau and Rockinger (2003). On the contrary, the approach considered in this paper focus on the null hypothesis (2), allowing for higher conditional moments of unknown form.

An important part of the empirical econometric literature still uses as a diagnostic tool for testing the goodness-of-fit of model (1) the classical Portmanteau tests initially proposed by Box and Pierce (1970) and Ljung and Box (1978), and subsequently extended to some conditional variance models by Li and Mak (1994), see also Lundbergh and Teräsvirta (2002). The theoretical foundation of this approach is based on the fact that under our assumptions, $\sigma(I_{t-1}^u) \subset \sigma(I_{t-1})$, where $I_{t-1}^u = (u_{t-1}(\theta_0), u_{t-2}(\theta_0), ...)'$, and thus, condition (2) yields

$$E[u_t(\theta_0) \mid I_{t-1}^u] = 0 \text{ a.s. and } E[u_t^2(\theta_0) \mid I_{t-1}^u] = 1 \text{ a.s. for some } \theta_0 \in \Theta \subset \mathbb{R}^p. \quad (3)$$

The latter point motivates some authors to consider specification tests for the conditional mean and variance based on checking for serial dependence (or lack thereof) of the unobserved errors $\{u_t(\theta_0)\}$ and/or their centered squares. However, it is important to remark that the serial uncorrelatedness of standardized errors (or centered square errors) imply neither condition (3) nor (2). In other words, tests based on usual correlation or autocorrelation measures of errors (centered square errors) are not consistent in any misspecified model delivering uncorrelated errors (centered square errors), incurring in an increase of the Type II error probability.

The aim of this paper is to proposed a large class of joint diagnostic tests for testing H_0. We summarize the main characteristics of our tests as follows: (i) they are suitable for cases in which the information set is infinite-dimensional, allowing for Markovian as well as non-Markovian time series processes; (ii) they do not depend on any smoothing parameter or kernel; (iii) 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, while being robust to higher unknown conditional dependence such as conditional skewness or kurtosis; (iv) they 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; (v) they are consistent against pairwise Pitman's local alternatives converging at the parametric rate $n^{-1/2}$; (vi) they are valid under fairly general regularity conditions on the underlying data generating process (DGP); and (vii) they are simple to compute.

The rationale for our test is as follows. Under H_0,

$$\gamma_j(\theta_0) = E[e_t(\theta_0) \mid Z_{t-j}] = 0 \text{ a.s. } \forall j, j \geq 1, \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \quad (4)$$

where $e_t(\theta_0) = (e_{1t}(\theta_0), e_{2t}(\theta_0))'$. Then, by appropriately choosing a parametric family of functions $\{w(Z_{t-j}, x) : x \in \Upsilon \subset [-\infty, \infty]^s\}$ (cf. Lemma 1 in Escanciano 2005) condition (4) can be equivalently expressed as

$$\gamma_{j,w}(x, \theta_0) = E[e_t(\theta_0) w(Z_{t-j}, x)] = 0 \text{ almost everywhere (a.e.) in } \Upsilon \subset [-\infty, \infty]^s, \ j \geq 1. \quad (5)$$

Usual examples of weight functions w satisfying previous equivalence are $w(Z_{t-j}, x) = 1(Z_{t-j} \leq x)$ with $x \in [-\infty, \infty]^s$, where $1(A)$ denotes the indicator of the event A, or $w(Z_{t-j}, x) = \exp(ix'Z_{t-j})$ with $x \in \mathbb{R}^s$. Our tests are founded on a generalized spectral distribution function approach using the measures $\{\gamma_{j,w}(\cdot, \theta_0)\}_{j=1}^{\infty}$ and extend those considered in Escanciano (2005) to joint specifications of the conditional mean and variance.

The remainder of this paper is organized 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. In Section 4, we propose and justify theoretically a bootstrap method to implement the tests. Finally, we make an extensive simulation exercise and an empirical application in Section 5, comparing with competing tests. 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. Also $|A|_M$ denotes the weighted norm $A'MA^c$ for a positive definite matrix M and a complex vector A. Unless indicated, all limits 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 many procedures for testing the correct specification of a parametric conditional mean can be used to test H_0 in a two step procedure: first, apply any consistent diagnostic test for testing the correct specification of the conditional mean, and once it is accepted that $E[e_{1t}(\theta_0) \mid I_{t-1}] = 0$, i.e., the conditional mean is well-specified, one proceeds to test for the correct specification of the conditional variance. Notice that, in the first step it is important to have tests robust to conditional heteroskedasticity as well as tests consistent against any direction, because if the conditional mean is misspecified, then the inference on the conditional variance could give misleading results, see, e.g., Lumsdaine and Ng (1999). In the second step, one proceeds to test the conditional variance adequacy, i.e., $E[e_{2t}(\theta_0) \mid I_{t-1}] = 0$ a.s. However, the sequential use of those tests procedures requires some caution since they are in general mutually dependent, and hence, this sequential procedure may increase the probability of Type I error. To avoid this problem, we consider in this paper a joint test for the two CMR in (2).

Our methodology for testing H_0 relies on a pairwise approach that has been shown to be very useful in a variety of testing problems, see, e.g., Hong (1999), Escanciano and Velasco (2003), Hong and Lee (2003), Escanciano (2005) or Hong and Lee (2005). More concretely, we consider simultaneously all the dependence measures $\{\gamma_{j,w}(\cdot, \theta_0)\}$ in (5) and define $\gamma_{-j,w}(\cdot, \theta_0) = \gamma_{j,w}(\cdot, \theta_0)$ for $j \geq 1$, to write the Fourier transform of the functions $\{\gamma_{j,w}(\cdot, \theta_0)\}_{j=-\infty}^{\infty}$, i.e.,

$$f_w(u, x, \theta_0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{j,w}(x, \theta_0) e^{-iju} \qquad \forall u \in [-\pi, \pi], x \in \Upsilon, \tag{6}$$

which contains the same information about H_0 as the whole sequence $\{\gamma_{j,w}(x,\theta_0)\}_{j=0}^{\infty}$. Note that under H_0, $f_w(u,x,\theta_0) \equiv f_{0,w}(x,\theta_0) = (2\pi)^{-1}\gamma_{0,w}(x,\theta_0)$, and hence, a test can be based on a distance from the estimator of $f_w(u,x,\theta_0)$ under the null and under the alternative. However, to avoid smoothing estimation we consider a generalized spectral distribution function approach based on the dependence measures $\{\gamma_{j,w}(\cdot,\theta_0)\}_{j=-\infty}^{\infty}$. Our tests are then based on the integral of $f_w(u,x,\theta_0)$, i.e.,

$$H_w(\lambda,x,\theta_0) = 2\int_0^{\lambda\pi} f_w(u,x,\theta_0)du \qquad \forall\lambda\in[0,1], x\in\Upsilon,$$

which after some manipulation can be written as

$$H_w(\lambda,x,\theta_0) = \gamma_{0,w}(x,\theta_0)\lambda + 2\sum_{j=1}^{\infty}\gamma_{j,w}(x,\theta_0)\frac{\sin j\pi\lambda}{j\pi}. \tag{7}$$

Now, suppose we have a sample $\{Y_t,\widehat{I}_{t-1}\}_{t=1}^{n}$ of size n that is used to estimate the model (1). Here $\widehat{I}_{t-1}$ is the information set observed at time $t-1$ that contains $(Z'_{t-1},Z'_{t-2},...,Z'_0)'$ and that may contain some initial values. We obtain residuals

$$\widehat{e}_{1t} \equiv \widehat{e}_{1t}(\theta_n) = Y_t - f(\widehat{I}_{t-1},\theta_n) \qquad \widehat{e}_{2t} \equiv \widehat{e}_{2t}(\theta_n) = \left(Y_t - f(\widehat{I}_{t-1},\theta_n)\right)^2 - h^2(\widehat{I}_{t-1},\theta_n), \tag{8}$$

where θ_n is a $\sqrt{n}$-consistent estimator for θ_0, e.g., the Quasi-Maximum Likelihood Estimator (QMLE). The sample version of $\gamma_{j,w}(x,\theta_0)$ for $j\geq 1$ is then given by

$$\widehat{\gamma}_{j,w}(x,\theta_n) = \frac{1}{n_j}\sum_{t=j}^{n}\widehat{e}_t w(Z_{t-j},x), \qquad n_j = n-j+1,\ \widehat{e}_t = (\widehat{e}_{1t},\widehat{e}_{2t})'.$$

Hence, the sample analogue of (7) is

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

with $(n_j/n)^{1/2}$ a finite sample correction factor that 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, $H_w(\lambda,x,\theta_0) = \gamma_{0,w}(x,\theta_0)\lambda$, and therefore, tests can be based on the discrepancy between $\widehat{H}_w(\lambda,x,\theta_n)$ and $\widehat{H}_{0,w}(\lambda,x,\theta_n) = \widehat{\gamma}_{0,w}(x,\theta_n)\lambda$, i.e.,

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

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

$$J_{n,w}^2(\theta_n) = \int_{\Pi} |S_{n,w}(\lambda,x,\theta_n)|_M^2\, W(dx)d\lambda = \sum_{j=1}^{n}\frac{n_j}{(j\pi)^2}\int_{\Upsilon}\left|\widehat{\gamma}_{j,w}(x,\theta_n)\right|_M^2 W(dx), \tag{9}$$

where $W(\cdot)$ is an integrating function depending on the weight family $\{w(\cdot, x) : x \in \Upsilon \subset \mathbb{R}^s\}$ and satisfying some mild conditions (see Assumption A5 below) and M is a 2×2 positive definite matrix. Therefore, our tests consist in rejecting H_0 for "large" values of $J^2_{n,w}(\theta_n)$. Note that $J^2_{n,w}(\theta_n)$ uses all lags contained in the sample, does not depend on any lag order and is very simple to compute (see Section 5). On the other hand, the range of possibilities in the choice of w, M and W creates flexibility for $J^2_{n,w}(\theta_n)$ in directing the power against some desired directions. The next section justifies inferences based on the asymptotic theory.

3. ASYMPTOTIC NULL DISTRIBUTION

To elaborate the asymptotic theory we consider the following assumptions. We define the score $g(I_{t-1}, \theta)$ with rows $g'_{1t}(\theta)$ and $g'_{2t}(\theta)$, given by

$$g'_{1t}(\theta) = (\partial/\partial\theta')f(I_{t-1}, \theta),$$

and

$$g'_{2t}(\theta) = 2e_{1t}(\theta)g'_{1t}(\theta) + 2h(I_{t-1}, \theta)\frac{\partial h(I_{t-1}, \theta)}{\partial\theta'}.$$

To simplify notation write $w(Z_{t-j}, x) \equiv w_{t-j}(x)$.

Assumption A1:

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

A1(b): $E[e^2_{1t}(\theta_0)] < C$ and $E[e^2_{2t}(\theta_0)] < C$.

Assumption A2: Let Θ_0 be a small convex neighborhood of θ_0. The functions $f(I_{t-1}, \cdot)$ and $h(I_{t-1}, \cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with score $g_t(\theta_0) \equiv g(I_{t-1}, \theta)$ stationary, ergodic and $\mathcal{F}_{t-1}$-measurable. There exist functions $G_j(I_{t-1})$ with $\sup_{\theta\in\Theta_0} |g_{jt}(\theta)| \leq G_j(I_{t-1})$, with $E[G_j(I_{t-1})] < C$, for $j = 1, 2$.

Assumption A3:

A3(a): The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ. There exists a unique $\theta_1 \in \Theta$ such that $|\theta_n - \theta_1| = o_P(1)$.

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

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} m(I_{t-1}, \theta_0)e_t(\theta_0) + o_P(1),$$

where $m(\cdot)$ is such that $L(\theta_0) = E[m(Y_t, I_{t-1}, \theta_0)e_t(\theta_0)e'_t(\theta_0)m'(Y_t, I_{t-1}, \theta_0)]$ exists and is positive definite.

Assumption A4: The integrating function $W(\cdot)$ is a probability distribution function absolutely continuous with respect to Lebesgue measure. M is a 2×2 positive definite matrix. The weight function $w(\cdot)$ is such that the equivalence between (4) and (5) holds, and is uniformly bounded on compacta. Also, $w(\cdot)$ satisfies the uniform law of large numbers (ULLN)

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

whenever $\{(\zeta_t,\xi_t),\ t=0,\pm1,...\}$ is a strictly stationary and ergodic process with $\zeta_t\in\mathbb{R}$, $\xi_t\in\mathbb{R}^m$, $E|\zeta_1|<\infty$, and Υ_c is any compact subset of $\Upsilon\subset[-\infty,\infty]^s$.

Assumption A5: The observed information set available at period t, $\widehat{I}_t$, may contain some assumed initial values and satisfies

$$E\left(\sum_{t=1}^{n}(\widehat{e}_{1t}-\widetilde{e}_{1t})\right)^2=o(n)$$

and

$$E\left(\sum_{t=1}^{n}(\widehat{e}_{2t}-\widetilde{e}_{2t})\right)^2=o(n),$$

where $\widetilde{e}_{1t}$ and $\widetilde{e}_{2t}$ are computed as in (8) but with I_{t-1} replacing $\widehat{I}_{t-1}$.

Assumption A1 is a condition on the DGP. Note that we do not need any mixing or asymptotic independence assumption to derive the asymptotic theory, see, e.g., the mixing assumption A.1 in Hong and Lee (2003). 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 literature, see, e.g., Bierens and Ploberger (1997). Assumption A3 is satisfied under mild conditions for the NLSE (or its robust modifications, under further regularity assumptions) or for the QMLE, see Koul (2002, Chapters 5 and 8), Hall and Heyde (1980, Chapter 6), Horváth et al. (2001) or Straumann (2005). A3 implies that under H_0, $\theta_1=\theta_0$, but they may be different under the alternative. Examples of $W(\cdot)$ include the cumulative distributions functions (cdf) of a N(0,1), Double Exponential or the Student's t_ν distribution. 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 $\widehat{I}_{t-1}$ and is similar in spirit to Assumption A4 in Hong and Lee (2003). It is straightforward to show that A5 is satisfied for most standard examples, e.g., MA(1) and GARCH(1,1) models, under mild conditions on the conditional mean and variance parameters and some mild moment conditions.

To elaborate the asymptotic theory we need further notation. Let us define $\Pi=[0,1]\times\Upsilon$ and $\eta=(\lambda,x')'\in\Pi$. In this section we establish the null limit distribution of the process $S_n(\lambda,x,\theta_n)\equiv S_n(\eta,\theta_n)$ under H_0. We consider $S_n(\eta,\theta_n)$ as a random element on the Hilbert space $L_2(\Pi,\nu,M)$ of all bivariate 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], that is, $f \in L_2(\Pi, \nu, M)$ if

$$\|f\|^2 = \int_{\Pi} f'(\eta) M f^c(\eta) d\nu(\eta) = \int_{\Pi} f'(\eta) M f^c(\eta) W(dx) d\lambda < \infty.$$

In $L_2(\Pi, \nu, M)$ we define the inner product

$$\langle f, g \rangle = \int_{\Pi} f'(\eta) M g^c(\eta) W(dx) d\lambda.$$

If Z is an $L_2(\Pi, \nu, M)$-valued random variable, we say that Z has mean m if $E[\langle Z, h \rangle] = \langle m, h \rangle$ $\forall h \in L_2(\Pi, \nu, M)$. If $E\|Z\|^2 < \infty$ and Z has zero mean, then the covariance operator of Z, say C_Z, is defined by $C_Z(h) = E[\langle Z, h \rangle Z]$. Denote by $\Longrightarrow$ weak convergence in the Hilbert space $L_2(\Pi, \nu, M)$ endowed with the norm metric. Also, denote by $\xrightarrow{L_2}$ convergence in probability in $L_2(\Pi, \nu, M)$, i.e., $Z_n \xrightarrow{L_2} Z \Longleftrightarrow \|Z_n - Z\| \xrightarrow{P} 0$. Let us define $\Psi_j(\lambda) = \sqrt{2}(\sin j\pi\lambda)/j\pi$, $b_j(x,\theta_0) = E[w_{t-j}(x) g_t(\theta_0)]$, $G_w(\eta) \equiv G_w(\eta, \theta_0) = \sum_{j=1}^{\infty} b_j(x,\theta_0)\Psi_j(\lambda)$ and for $h \in L_2(\Pi, \nu, M)$,

$$\sigma_h^2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} E\left[e_t(\theta_0) M \int_{\Pi\times\Pi} h(\eta_1)(h^c(\eta_2))' w_{1-j}^c(x) w_{1-k}(y) \Psi_j(\lambda)\Psi_k(\varpi) d\nu(\eta_1) d\nu(\eta_2) M' e_t(\theta_0) \right], \tag{10}$$

with $\eta_1 = (\lambda, x')'$ and $\eta_2 = (\varpi, y')'$. Let V be a normal random vector with zero mean and variance-covariance matrix given by $L(\theta_0)$ (cf. A3(b)), and let $S_w^0(\cdot)$ be a Gaussian process in $L_2(\Pi, \nu, M)$ with zero mean and covariance operator $C_{S_w^0}$ satisfying $\sigma_h^2 = \langle C_{S_w^0}(h), h \rangle$, $\forall h \in L_2(\Pi, \nu, M)$, 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 $L_2(\Pi, \nu, M)$, where $S_w(\cdot)$ has the same distribution as $S_w^0(\cdot) - G_w(\cdot)V$, with*

$$Cov(S_w^0(\eta), V) = \sum_{j=1}^{\infty} E[e_t(\theta_0) \otimes m(I_{t-1}, \theta_0) e_t(\theta_0) w_{t-j}(x)] \Psi_j(\lambda),$$

where $\otimes$ stands for the Kronecker product.

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,*

$$J_{n,w}^2(\theta_n) \xrightarrow{d} J_{\infty,w}^2(\theta_0) = \int |S_w(\lambda, x, \theta_0)|_M^2 \, W(dx) d\lambda.$$

The asymptotic power properties of $J^2_{n,w}(\theta_n)$ can be studied using the arguments of Escanciano (2005). We do not discuss these issues here for the sake of space. To end this section, it is important to remark that the asymptotic null distribution of $J^2_{n,w}$ 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. Alternative solutions proposed in the literature, such as the martingale transformation used in Koul and Stute (1999) (cf. Khmaladze, 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.

4. BOOTSTRAP APPROXIMATION

Resampling methods have been extensively used in the model checks literature of regression time series models; see, e.g., Stute, Gonzalez-Manteiga and Presedo-Quindimil (1998) in an i.i.d context, or Escanciano (2006) 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. For simplicity, we shall assume throughout this section that the parameter θ can be partitioned into two parameters $\theta = (\alpha', \beta')'$ such that α only enters in the conditional mean and β in the conditional variance, that is, $f(\cdot, \theta) = f(\cdot, \alpha)$ and $h(\cdot, \theta) = h(\cdot, \beta)$. This situation covers most models of the literature and simplifies the bootstrap approach. Write $\theta_0 = (\alpha_0', \beta_0')'$ and $\theta_n = (\alpha_n', \beta_n')'$. Here we approximate the asymptotic null distribution of $S_{n,w}$ by that of

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

with $\widehat{\gamma}^*_j(x) = (\widehat{\gamma}^*_{j,m}(x), \widehat{\gamma}^*_{j,v}(x))'$,

$$\widehat{\gamma}^*_{j,m}(x) = \frac{1}{n_j} \sum_{t=j}^{n} \widehat{e}^*_{1t} w_{t-j}(x),$$

and

$$\widehat{\gamma}^*_{j,v}(x) = \frac{1}{n-j} \sum_{t=1+j}^{n} \widehat{e}^*_{2t} w_{t-j}(x),$$

and where $\widehat{e}^*_t = (\widehat{e}^*_{1t}, \widehat{e}^*_{2t})'$ are obtained from the following algorithm:

Step 1: Estimate the original model and obtain the residuals $\widehat{e}_t(\theta_n)$.

Step 2: Generate WB residuals according to $\widehat{\varepsilon}^*_{1t} = \widehat{e}_{1t}(\alpha_n) V_t$ and $\widehat{\varepsilon}^*_{2t} = \widehat{e}_{2t}(\theta_n) V_t$ for $1 \le t \le n$, with $\{V_t\}$ a sequence of i.i.d random variables with zero mean, unit variance, bounded support and independent of the sequence $\{(Y_t, \widehat{I}_{t-1})'\}_{t=1}^n$.

Step 3: Given θ_n and $\widehat{\varepsilon}_{1t}^*$ and $\widehat{\varepsilon}_{2t}^*$, generate bootstrap data according to

$$Y_{1t}^* = f(\widehat{I}_{t-1}, \alpha_n) + \widehat{\varepsilon}_{1t}^* \text{ for } 1 \leq t \leq n, \tag{11}$$

and

$$Y_{2t}^* = h^2(\widehat{I}_{t-1}, \beta_n) + \widehat{\varepsilon}_{2t}^* \text{ for } 1 \leq t \leq n. \tag{12}$$

Step 4: Compute $\theta_n^* = (\alpha_n^*, \beta_n^*)'$, where α_n^* is computed from the data $\{Y_{1t}^*, \widehat{I}_{t-1}\}_{t=1}^{\infty}$ in (11) and β_n^* is computed from the data $\{Y_{2t}^*, \widehat{I}_{t-1}\}_{t=1}^{\infty}$ in (12). Then, compute $\widehat{e}_t^* = (\widehat{e}_{1t}^*, \widehat{e}_{2t}^*)'$ according to $\widehat{e}_{1t}^* = Y_{1t}^* - f(\widehat{I}_{t-1}, \alpha_n^*)$ and $\widehat{e}_{2t}^* = Y_{2t}^* - h^2(\widehat{I}_{t-1}, \beta_n^*)$ for $t = 1, ..., n$.

Examples of $\{V_t\}$ sequences are i.i.d Bernoulli variates with

$$P(V_t = 0.5(1 - \sqrt{5})) = b \text{ and } P(V_t = 0.5(1 + \sqrt{5})) = 1 - b, \tag{13}$$

with $b = (1 + \sqrt{5})/2\sqrt{5}$, used in, e.g., 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). Other sequences can be found in Mammen (1993). The next theorem justifies theoretically the bootstrap approximation. The unknown limiting null distribution of $J_{n,w}^2(\theta_n)$, i.e., the distribution of $J_{\infty,w}^2(\theta_0)$, is approximated by the bootstrap distribution of

$$J_{n,w}^{2*} = \int \left|S_{n,w}^*(\lambda, x, \theta_n^*)\right|_M^2 W(dx)d\lambda.$$

That is, the bootstrap distribution

$$F_J^*(x) = P\left(J_{n,w}^{2*} \leq x \middle| \{Y_t, \widehat{I}_{t-1}\}_{t=1}^n\right)$$

estimates the asymptotic null distribution function

$$F_J(x) = P\left(J_{\infty,w}^2(\theta_0) \leq x\right).$$

Thus, H_0 will be rejected at the $100\alpha\%$ of significance when $J_{n,w}^2(\theta_n) \geq c_{n,\alpha}^*$, where $F_J^*\left(c_{n,\alpha}^*\right) = 1-\alpha$. Also, we can use the bootstrap $p - values$, p_n^* say, rejecting H_0 when $p_n^* < \alpha$, where $p_n^* = P\left(J_{n,w}^{2*} \geq J_{n,w}^2(\theta_n) \middle| \{Y_t, \widehat{I}_{t-1}\}_{t=1}^n\right)$. The bootstrap assisted test is valid if F_J^* is a consistent estimator of F_J at each continuity point of F_J. When consistency is a.s., it is expressed as $J_{n,w}^{2*} \rightarrow_d J_{\infty,w}^2(\theta_0)$ a.s. See Ginè and Zinn (1990) for discussion. Remark that we say that the bootstrap statistic η_n^* converges in probability a.s. to η_n if for all $\delta > 0$, $P\left(|\eta_n^* - \eta_n| \geq \delta \middle| \{Y_t, \widehat{I}_{t-1}\}_{t=1}^n\right) \rightarrow 0$ a.s., which is expressed as $\eta_n^* = \eta_n + o_P(1)$ a.s. In order to show that the bootstrap assisted tests are valid, we need to assume that the bootstrap analogs of θ_n satisfy an asymptotic expansion like A3(b) in the bootstrap world.

Assumption A6:

A6(a): The estimator $\theta_n^* = (\alpha_n^{*'}, \beta_n^{*'})'$ satisfies the asymptotic expansions

$$\sqrt{n}(\theta_n^* - \theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_t m(\widehat{I}_{t-1}, \theta_n) \widehat{e}_t(\theta_n) + o_P(1) \text{ a.s.},$$

where the function $m(\cdot)$ is as in A3.

A6(b): There exists an integrable function $K(I_{t-1})$ with $\sup_{\theta \in \Theta} |m(I_{t-1}, \theta)| \leq K(I_{t-1})$, with $E[K(I_{t-1})] < C$.

Theorem 2 *Assume A1-A6. Under the null hypothesis H_0 or under any fixed alternative hypothesis,*

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

where $\widetilde{S}_w$ is the same Gaussian process of Theorem 1 but with θ_1 replacing θ_0 and $\underset{*}{\Longrightarrow}$ denoting weak convergence almost surely under the bootstrap law; see Ginè 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 set $Z_t = Y_t$. We compare our tests with the generalized spectral test of Hong and Lee (2003) ($M_{n,p}$) and the Portmanteau tests of Li and Mak's (1994) (LM_m). We briefly describe our simulation setup. We denote by $J_{n,I}^2$ our new Cramér-von Mises test based on $w(Y_{t-j}, x) = 1(Y_{t-j} \leq x)$ and the empirical distribution function of $\{Y_{t-1}\}_{t=1}^n$ as the integrating measure, i.e.,

$$J_{n,I}^2 = \sum_{j=1}^{n} \frac{n_j}{n(j\pi)^2} \sum_{t=1}^{n} \left(m_1 \widehat{\gamma}_{I,j,m}^2 (Y_{t-1}, \theta_n) + m_2 \widehat{\gamma}_{I,j,v}^2 (Y_{t-1}, \theta_n) \right),$$

where

$$\widehat{\gamma}_{w,,j,m}(x, \theta_n) = \frac{1}{\widehat{\sigma}_{1e} n_j} \sum_{t=j}^{n} \widehat{e}_{1t} w(Y_{t-j}, x),$$

$$\widehat{\gamma}_{w,,j,v}(x, \theta_n) = \frac{1}{\widehat{\sigma}_{2e} n_j} \sum_{t=j}^{n} \widehat{e}_{2t} w(Y_{t-j}, x),$$

and

$$\widehat{\sigma}_{je}^2 = \frac{1}{n} \sum_{t=1}^{n} \widehat{e}_{jt}^2, \qquad j = 1, 2.$$

The subindex I in $\widehat{\gamma}_{I,,j,m}$ and $\widehat{\gamma}_{I,,j,v}$ corresponds to the use of $w(Y_{t-j}, x) = 1(Y_{t-j} \leq x)$. Note that the use of the empirical cdf does not affect the asymptotic theory, see Escanciano (2005). For the joint test we consider $(m_1, m_2) = (1, 1)$. The marginal tests $D_{n,I,m}^2$ and $D_{n,I,v}^2$ correspond to the choices $(m_1, m_2) = (1, 0)$ and $(m_1, m_2) = (0, 1)$, respectively.

Analogously, we define $J^2_{n,C}$, $D^2_{n,C,m}$ and $D^2_{n,C,v}$ based on $w(Y_{t-j}, x) = \exp(ixY_{t-j})$ and the integrating function ϕ, the density function of the standard normal random variable, which yields the test statistic

$$J^2_{n,C} = \sum_{j=1}^{n} \frac{1}{n_j (j\pi)^2} \sum_{t=j}^{n} \sum_{s=j}^{n} \left(m_1 \widehat{\sigma}^2_{1e} \widehat{e}_{1t} \widehat{e}_{1s} + m_2 \widehat{\sigma}^2_{2e} \widehat{e}_{2t} \widehat{e}_{2s} \right) \exp(-0.5 \cdot (Y_{t-j} - Y_{s-j})^2).$$

Our test statistics $J^2_{n,I}$ and $J^2_{n,C}$ are representatives of the CvM tests based on the most used weighting functions. These CvM tests are based on the choice M with rows $(m_1, 0)$ and $(0, m_2)$.

Hong and Lee's (2003) test is given by

$$M_{n,p} = \left[HL_{n,p} - \widehat{C}_0 K_2 \right] / \left[2\widehat{D}_0 K_4 \right], \tag{14}$$

where $HL_{n,p}$ is defined by

$$HL_{n,p} = \int \sum_{j=1}^{n-1} k^2(j/p)(n-j) \left| \widehat{\sigma}_j(y, x, \theta_n) \right|^2 W(dy) W(dx), \tag{15}$$

where $\widehat{\sigma}_j(y, x, \theta_n)$ is the sample covariance between $\exp(iyu_t(\theta_n))$ and $\exp(ixu_{t-j}(\theta_n))$, $k(\cdot)$ is a kernel function, p is a bandwidth and W is a weighting function.

Moreover, $K_2 = \sum_{j=1}^{n-1} k^2(j/p)$, $K_4 = \sum_{j=1}^{n-1} k^4(j/p)$ and the centering and scaling factors are, respectively

$$\widehat{C}_0 = \left[\int \widehat{\sigma}_0(y, -y, \theta_n) W(dy) \right]^2$$

and

$$\widehat{D}_0 = \left[\int \left| \widehat{\sigma}_0(y, x, \theta_n) \right|^2 W(dy) W(dx) \right]^2.$$

Under the null hypothesis of i.i.d standardized errors and some assumptions Hong and Lee (2003) showed that $M_{n,p}$ converges to a standard normal random variable. As in Hong and Lee (2003), we use the density function $W(\cdot) \equiv \phi(\cdot)$ and the Daniell kernel $k(z) = sin(\pi z)/\pi z$.

Throughout ε_t and v_t are independent sequences of i.i.d. $N(0, 1)$. We consider the nominal level 5%. The results with other significance levels are similar. 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. For the bootstrap approximation we employ a sequence $\{V_t\}$ of i.i.d Bernoulli variates given in (13). The power in the non-bootstrap cases is level-adjusted by using the empirical values obtained under the corresponding null hypothesis, although the difference is not substantial. To examine the impact of the bandwidth on $M_{n,p}$ we consider $p = 2$ to 11. For Li and Mak's (1994) (LM_m) test we use $m = 2$ to 11. For simplicity, we only present in tables the values $m, p = 2, 6$ and 10.

5.1 Conditional Variance Models

Now, we examine the adequacy of an ARCH(1) model against misspecifications in conditional mean, conditional variance and both conditional mean and variance. We compare our marginal tests $D^2_{n,I,v}$ and $D^2_{n,C,v}$ with $M_{n,p}$ and LM_m for linear and nonlinear conditional variance specifications. With the null ARCH(1) model, we examine the level and power against misspecifications in the conditional variance, their power against apparent ARCH structures and against chaotic processes with similar autocorrelations in squares to an ARCH(1). Our null model is an ARCH(1) model:

$$Y_t = h_t \varepsilon_t, \qquad h_t^2 = a + bY_{t-1}^2.$$

We examine the adequacy of this model under the following DGP:

1. ARCH(1) model: $Y_t = h_t\varepsilon_t$, $h_t^2 = 0.9 + 0.1Y_{t-1}^2$.
2. ARCH(2) model: $Y_t = h_t\varepsilon_t$, $h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.8Y_{t-2}^2$.
3. GARCH(1,1) model: $Y_t = h_t\varepsilon_t$, $h_t^2 = 0.01 + 0.29Y_{t-1}^2 + 0.7h_{t-1}^2$.
4. EGARCH(1,1) model: $Y_t = h_t\varepsilon_t$, $\ln h_t^2 = 0.01 + 0.9\ln h_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - (2/\pi)^{1/2}) - 0.8\varepsilon_{t-1}$.
5. Stochastic Volatility (SV) model: $Y_t = h_t\varepsilon_t$, $h_t^2 = 0.1Y_{t-1}^2 + \exp(0.98\ln h_{t-1}^2 + v_t)$.
6. Bilinear model (BIL): $Y_t = 0.8\varepsilon_{t-1}Y_{t-1} + \varepsilon_t$.
7. Logistic Map (LM): $Y_t = 4Y_{t-1}(1 - Y_{t-1})$, where Y_0 is generated from the uniform distribution on [0,1].
8. Non-Linear Moving Average model (NLMA): $Y_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$.

These models have been considered in Hong and Lee (2003) except for the parameter values of model 2 (we have changed the parameter values to a better discrimination among the tests). To compute the statistics $D^2_{n,I,v}$ and $D^2_{n,C,v}$, we use the residuals $\widehat{e}_{2t}(\beta_n) := Y_t^2 - h^2(Y_{t-1}, \beta_n)$ where $h^2(Y_{t-1}, \beta_n) = \widehat{a} + \widehat{b}Y_{t-1}^2$, and $\beta_n = (\widehat{a}, \widehat{b})$ is the least squares estimators (LSE) in the regression of Y_t^2 against a constant and Y_{t-1}^2. In $M_{n,p}$, and LM_m we use standardized residuals $\widehat{u}_t(\beta_n) = Y_t/h(Y_{t-1}, \beta_n)$. The sample size is $n = 100$. In Table 1 we report the empirical rejections probabilities (RP) associated with the models 1 to 8 to examine the empirical level and power of tests. The tests $D^2_{n,I,v}$, $D^2_{n,C,v}$, LM_m and $M_{n,p}$ show an excellent empirical level.

Table 1 also examines the empirical power of the tests against the conditional variance models 2 to 8. Our tests $D^2_{n,I,v}$ and $D^2_{n,C,v}$ have excellent empirical power against the EGARCH, SV, BILINEAR, LOGISTIC MAP and NLMA models, and moderate empirical power against ARCH(2) and GARCH(1,1). It is observed that $D^2_{n,C,v}$ outperforms $D^2_{n,I,v}$ for conditional variance models, this finding is similar to the well documented fact in the goodness-of-fit literature of distribution functions, see e.g. Feigin and Heathcote (1976), that indicator based tests have low power against changes in scale, whereas exponential functions have good power properties for changes in scale and mean. Hong and Lee's (2003) test $M_{n,p}$ has good empirical power against ARCH(2), EGARCH, BIL and LOGISTIC MAP models and moderate power against the rest of models. Notice that $M_{n,p}$ is very sensitive on p for ARCH(2) and SV models. Li and Mak's (1994) test LM_m has excellent empirical power against the models ARCH(2), GARCH(1,1) and SV, and has low power against BIL, LOGISTIC MAP and NLMA alternatives.

TABLES 1 AND 2 ABOUT HERE

It is shown in these simulations that $D^2_{n,I,v}$ and $D^2_{n,C,v}$ have omnibus power against all linear and nonlinear alternatives considered. Notably, the exponential based test $D^2_{n,C,v}$ has excellent empirical power properties against EGARCH, SV, BIL, LOGISTIC MAP and NLMA alternatives. Now, we consider joint conditional mean and conditional variance models.

5.2 Joint Specifications of Conditional Mean and Variance

In this subsection we examine the adequacy of an autoregressive conditional heteroskedastic model (AR(1)-CH(1)) against misspecifications in conditional mean, conditional variance and both conditional mean and variance. We compare our joint tests $J^2_{n,I}$ and $J^2_{n,C}$, with the marginal tests $D^2_{n,I,m}$, $D^2_{n,C,m}$, $D^2_{n,I,v}$, $D^2_{n,C,v}$, and $M_{n,p}$. The simulation design is the same as before. Our null model is:

$$Y_t = aY_{t-1} + h_t\varepsilon_t, \qquad h_t^2 = b + cY_{t-1}^2.$$

We examine the adequacy of this model under the following DGP:

1. AR(1)-CH(1) model: $Y_t = 0.6Y_{t-1} + h_t\varepsilon_t$, $h_t^2 = 0.9 + 0.1Y_{t-1}^2$.
2. AR(1)-BIL model: $Y_t = 0.6Y_{t-1} + 0.4Y_{t-1}\varepsilon_t + \varepsilon_t$.
3. AR(2)-CH(1) model: $Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + h_t\varepsilon_t$, $h_t^2 = 0.9 + 0.1Y_{t-1}^2$.
4. TAR model: $Y_t = 0.9Y_{t-1} + \varepsilon_t$ if $|Y_{t-1}| \leq 1$ and $Y_t = -0.3Y_{t-1} + \varepsilon_t$ if $|Y_{t-1}| > 1$.

We report the RP for $J^2_{n,I}$, $J^2_{n,C}$, $D^2_{n,I,m}$, $D^2_{n,C,m}$, $D^2_{n,I,v}$, $D^2_{n,C,v}$, and $M_{n,p}$ in Table 2. The samples sizes considered are $n = 50$, 100 and 200. The empirical level of the joint and marginal test statistics is excellent against the AR(1)-CH(1) model, and more or less satisfactory for $M_{n,p}$. For the AR(1)-BIL the conditional mean is well specified and the conditional variance is misspecified, this is reflected in the empirical powers of $J_{n,I}^2$, $J_{n,C}^2$, $D_{n,I,m}^2$, $D_{n,C,m}^2$, $D_{n,I,v}^2$ and $D_{n,C,v}^2$. Hong and Lee's test has reasonable empirical power. Among all statistics, our tests $J_{n,I}^2$ and $J_{n,C}^2$ have the highest empirical powers against the AR(1)-BIL. The AR(2)-CH(1) is a model with misspecified conditional mean and well specified conditional variance, again this is reflected in the empirical powers of $J_{n,I}^2$, $J_{n,C}^2$, $D_{n,I,m}^2$, $D_{n,C,m}^2$, $D_{n,I,v}^2$, $D_{n,C,v}^2$. The empirical powers of $J_{n,I}^2$ and $D_{n,I,m}^2$ are the highest for this alternative. The empirical power of $M_{n,p}$ is more or less good but very sensitive to p. Finally, the TAR model has misspecified conditional mean and variance. For this model, the marginal tests $D_{i,v}$ and $D_{\exp,v}$ are not able to detect the incorrect specification in the conditional variance. One possible reason that may explain this fact is that the marginal tests for the conditional mean and variance specifications might be negatively correlated, so a misspecification of the conditional mean is delivering a lack of power in the test for conditional variance misspecification. Nevertheless, we observe that for this alternative our test statistics $J_{n,I}^2$, $J_{n,C}^2$, $D_{n,I,m}^2$ and $D_{n,C,m}^2$ outperform $M_{n,p}$.

TABLE 2 ABOUT HERE

These simulations have confirmed the ability of our joint test to detect misspecifications in both the conditional mean and variance functions. Furthermore, we have shown that the use of the marginal and joint tests may be a useful inference procedure to detect if the misspecification is in the conditional mean, in the conditional variance or in both, although some caution is necessary for variance specifications when the conditional mean is misspecified.

5.3 Empirical Application: S&P500 Dynamics

We now apply our testing methodology to the well-known and extensively studied S&P500 daily stock index. The debate on whether the dynamics of economic and financial time series are determined by the conditional mean or the conditional variance has important implications on many other applications including portfolio selection and asset pricing. Model-based financial decisions such as hedging, risk management or option pricing rely on the correct specification of the dynamics of the underlying asset price process. The S&P500 daily stock index is a representative of the data for which the GARCH model has been extensively used, see e.g. Bollerslev, et al. (1992) and references therein. We consider a sample period from January 1, 1988 to May 28, 1993. The data are taken from Bera and Higgins (1997) and like they, we delete the last 10% observations, remaining 1210 observations. Bera and Higgins (1997) try to discriminate between a GARCH and a bilinear specification. Their results are inconclusive. Here, we only check if a GARCH(1,1) specification is adequate. As Bera and Higgins (1997) we specify an AR(1)-GARCH(1,1) model to the log differences of the S&P500 (Y_t), such as

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \\ h_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_3 h_{t-1}^2, \end{aligned}$$

where $\varepsilon_t = h_t u_t$ and h_t^2 is the conditional variance. The parameter estimation is by Gaussian Maximum Likelihood and the results are reported in Table 3, we also include the estimation results of Bera and Higgins (1997) for a better comparison, and as usual the standard errors are in parenthesis.

TABLE 5 ABOUT HERE

If we apply our tests to the S&P500 daily stock index with the same Monte Carlo setup as before we obtain that the conditional mean is well specified with a p-value of 0.416 for $D^2_{n,I,m}$ and 0.233 for $D^2_{n,C,m}$, whereas, the conditional variance is misspecified, as can be deduced from the zero p-value of $J^2_{n,I}$, $J^2_{n,C}$, $D^2_{n,I,v}$ and $D^2_{n,C,v}$. These results are confirmed with Hong and Lee's (2003) test, which rejects the correct specification for all values of p. Therefore, we find that the conditional mean of the S&P500 in this period is linear and that additional effort has to be dedicated to investigate the functional form of the conditional variance.

APPENDIX: PROOFS

First, consider three useful lemmas. Lemma 1 is a trivial multivariate extension of Lemma 1 in Escanciano and Velasco (2003, hereafter EV).

Lemma 1: *Suppose we have a random element in $L_2(\Pi, \nu, M)$ of the form $h_n(\eta) = \sum_{j=1}^{n} h_{j,n}(x) \frac{\sqrt{2}\sin j\pi\lambda}{j\pi}$. Assume that W is of bounded total variation and that*

(i) $\int_{\mathbb{R}^s} E\left|h_{j,n}(x)\right|_M^2 W(dx) < C$ *uniformly in* $j \geq 1$.

(ii) $\sup_{x\in\Upsilon_c} \left|h_{j,n}(x)\right| = o_p(1)\ \forall j, 1 \leq j \leq n$, *for all compact subsets* $\Upsilon_c \subset \Upsilon$.

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

Proof of Lemma 1: EV.

Lemma 2: Under A4 and A5 the effect of estimating the information set I_{t-1} by $\widehat{I}_{t-1}$ in $f(I_{t-1}, \theta_n)$ has no effect on the asymptotic theory. That is,

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

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

Proof of Lemma 2: Note that for any vector A there exists a constant C such that $|A|_M \leq C\,|A|$. Write

$$\begin{aligned} & E\left\|S_{n,w}(\eta, \theta_n) - \widetilde{S}_{n,w}(\eta, \theta_n)\right\|^2 \\ \leq\ & C\sum_{j=1}^{n} \frac{1}{(j\pi)^2} n_j^{-1} E\left(\sum_{t=j}^{n} (\widehat{e}_{1t} - \widetilde{e}_{1t})\right)^2 + C\sum_{j=1}^{n} \frac{1}{(j\pi)^2} n_j^{-1} E\left(\sum_{t=j}^{n} (\widehat{e}_{2t} - \widetilde{e}_{2t})\right)^2 \\ =\ & o(1), \end{aligned}$$

where the last equality is due to A5.

For simplicity, we rename $\widetilde{S}_{n,w}(\eta, \theta_n)$ again as $S_{n,w}(\eta, \theta_n)$. The next Lemma establishes the asymptotic linearization of the process $S_{n,w}(\eta, \theta_n)$ under the null.

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

$$\left\|S_{n,w}(\eta, \theta_n) - S_{n,w}(\eta, \theta_0) + G_w(\eta, \theta_0)V\right\|^2 \xrightarrow{P} 0.$$

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

$$S_{n,w}(\eta,\theta_n) = S_{n,w}(\eta,\theta_0) + \frac{\partial S_{n,w}(\eta,\widetilde{\theta}_n)}{\partial\theta'}(\theta_n - \theta_0), \tag{16}$$

where $\widetilde{\theta}_n$ is a mean value satisfying $\left|\widetilde{\theta}_n - \theta_0\right| \leq |\theta_n - \theta_0|$ a.s. Note that the process $S_{n,w}(\eta,\theta_n)$ can be written as

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

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

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

where $b_{j,n}(x,\widetilde{\theta}_n) = n^{-1}\sum_{t=j}^{n} n^{1/2} n_j^{-1/2} g_t(\widetilde{\theta}_n) w_{t-j}(x)$. Assumptions A1-A5, the uniform argument of Jennrich (1969, Theorem 2) and Lemma 1 in EV yield

$$\left\|\frac{1}{\sqrt{n}}\frac{\partial S_{n,w}(\eta,\widetilde{\theta}_n)}{\partial\theta} + \sum_{j=1}^{n} b_j(x,\theta_0)\Psi_j(\lambda)\right\| \xrightarrow{P} 0.$$

The last display, Assumption A3 and (16) imply the result.

Proof of Theorem 1: We apply Lemma A2 here and Theorem 1 in EV but with $w_{t-j}(x)$ 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 Theorem 2: Write $\widehat{\varepsilon}_t^* = (\widehat{\varepsilon}_{1t}^*, \widehat{\varepsilon}_{2t}^*)'$,

$$S_{n,w,m}^*(\lambda, x,\alpha_n^*) = \sum_{j=1}^{n} n_j^{1/2}\widehat{\gamma}_{j,m}^*(x)\Psi_j(\lambda),$$

and

$$S_{n,w,v}^*(\lambda, x,\beta_n^*) = \sum_{j=1}^{n} n_j^{1/2}\widehat{\gamma}_{j,v}^*(x)\Psi_j(\lambda).$$

Hence, from the arguments of Theorem 2 in Escanciano (2005) it can be shown that

$$S_{n,w,m}^*(\eta,\theta_n^*) = n^{-1/2}\sum_{t=1}^{n}\widehat{\varepsilon}_{1t}^* Q_{t,w}(\eta) - n^{1/2}(\alpha_n^* - \alpha_n)' G_{1w}(\eta,\theta_1) + o_P(1) \text{ a.s.}, \tag{17}$$

and

$$S^*_{n,w,v}(\eta,\theta^*_n) = n^{-1/2}\sum_{t=1}^{n}\widehat{\varepsilon}^*_{2t}Q_{t,w}(\eta) - n^{1/2}(\beta^*_n - \beta_n)'G_{2w}(\eta,\theta_1) + o_P(1) \text{ a.s.,} \tag{18}$$

where $b_{hj}(x,\theta_0) = E[w_{t-j}(x)g_{ht}(\theta_0)]$, $G_{hw}(\eta,\theta_0) = \sum_{j=1}^{\infty} b_{hj}(x,\theta_0)\Psi_j(\lambda)$ for $h = 1, 2$, and

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

From this point, the proof follows exactly the same steps as in Theorem 1 in Stute, González-Manteiga and Presedo-Quindimil (1998). The details are omitted for the sake of space.

Acknowledgements

Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001.

References

Bera, A. K., and Higgins, M. L., 1997, Arch and Bilinearity as Competing Models for Nonlinear Dependence, Journal of Business and Economic Statistics 15, 43-50.

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

Bollerslev, T. , Chou, R. Y., and Kroner, K. F., 1992, ARCH modelling in Finance, Journal of Econometrics 52, 5–59.

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

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

Delgado, M. A, Dominguez, M. and Lavergne, P., 2005, Consistent tests of conditional moment restrictions, forthcoming Annales d'Economie et Statistique

Escanciano, J. C., 2005, Goodness-of-fit tests for linear and non-linear time series models, forthcoming in Journal of the American Statistical Association.

Escanciano, J. C., 2006, Model Checks Using Residual Marked Empirical Processes, forthcoming in Statistica Sinica.

Escanciano, J. C., and Velasco, C., 2003, Generalized Spectral Tests for the Martingale Difference Hypothesis, forthcoming in Journal of Econometrics.

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

Feigin, P. D., and Heathcote, C. R., 1976, The Empirical Characteristic Function and the Cramér-von Mises Statistic, Sankhya 38, Series A, 309–325.

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

Gallant, A. R., Hsieh, D. A., and Tauchen, G. 1991, On Fitting 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, Cambridge, U.K. Cambridge University Press, pp. 199-240.

Gao, J., and King, M., 2004, Model Specification Testing in Nonparametric and Semiparametric Time Series Econometrics, working paper.

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

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

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

Hansen, B., 1994, Autoregressive Conditional Density Estimation, International Economic Review 35, 705-730.

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., 2005, Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form, Review of Economic Studies 72, 499-541.

Jakubowski, A., 1980, On Limit Theorems for Sums of Dependent Hilbert Space Valued Random Variables, 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.

Jondeau, E. and M. Rockinger, 2003, Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements, Journal of Economic Dynamics and Control 27, 1699-1737.

Horváth, L., Kokoszka, P. and Teyssiére, G., 2001, Empirical Process of the Squared Residuals of an Arch Sequence, Annals of Statistics 29, 445-469.

Khmaladze, 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., 1999, Consistent Model Specification Test for Time Series Econometric Models, Journal of Econometrics 92, 101-147.

Li, W.K. and Mak, T.K. 1994, On the squared residual autocorrelation in nonlinear time series with conditional heteroskedasticity, Journal of Time Series Analysis 15, 627-636.

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.

Lundbergh, S., and T. Teräsvirta, 2002, Evaluating GARCH Models, Journal of Econometrics 110, 417-435.

Lumsdaine, R. L., and Ng, S., 1999, Testing for ARCH in the Presence of a Possibly Misspecified Mean, Journal of Econometrics 93, 257-280.

Mammen, E., 1993, Bootstrap and Wild Bootstrap for High-Dimensional Linear Models, Annals of Statistics 21, 255-285.

Ngatchou-Wandji, J., 2005, Checking Nonlinear Heteroscedastic Time Series Models, Journal of Statistical Planning and Inference 133, 33-68.

Straumann, D., 2005, Estimation in Conditionally Heteroscedastic Time Series Models. Lectures notes in Statistics 181, Springer-Verlag, Berling-Heidelberg.

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.

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

Table 1. Empirical Size and Power of Tests at 5%. Conditional Variance Models.

$n = 100$	**ARCH(1)**	**ARCH(2)**	**GARCH(1,1)**	**EGARCH**	**SV**	**BIL**	**LM**	**NLMA**
$D^2_{n,I}$	4.5	18.0	19.8	82.8	54.2	88.0	99.7	21.6
$D^2_{n,C}$	4.6	52.5	31.2	90.8	67.0	96.7	99.3	77.5
$M_{n,2}$	5.3	11.2	14.4	89.0	30.4	93.6	100	65.2
$M_{n,6}$	4.9	63.0	28.4	95.0	49.4	90.0	100	51.0
$M_{n,10}$	6.3	60.2	31.0	94.0	52.6	82.0	100	33.2
LM_2	4.4	88.3	39.9	49.2	42.8	23.1	19.0	22.6
LM_6	3.8	77.7	60.6	53.4	59.8	16.5	14.8	11.9
LM_{10}	3.7	72.8	61.4	48.0	58.1	15.3	14.7	10.4

Table 2. Empirical Size and Power of Tests at 5%. Conditional Mean and Variance Models.

$n = 50$	**AR(1)-CH(1)**	**AR(1)-BIL**	**AR(2)-CH(1)**	**TAR**
$D^2_{n,I,m}$	3.8	4.0	76.1	27.5
$D^2_{n,I,v}$	6.8	77.7	4.2	9.6
$J^2_{n,I}$	5.2	68.7	36.1	18.1
$D^2_{n,C,m}$	3.3	5.9	34.9	40.2
$D^2_{n,C,v}$	4.8	65.5	4.9	8.5
$J^2_{n,C}$	4.4	51.7	20.7	31.0
$M_{n,2}$	3.3	42.4	16.0	21.6
$M_{n,6}$	3.9	35.7	62.0	17.1
$M_{n,10}$	5.7	27.9	62.7	14.9

Table 3. Empirical Size and Power of Tests at 5%. Conditional Mean and Variance Models.

$n = 100$	**AR(1)-CH(1)**	**AR(1)-BIL**	**AR(2)-CH(1)**	**TAR**
$D^2_{n,I,m}$	4.9	5.7	98.0	57.9
$D^2_{n,I,v}$	6.2	95.7	4.0	16.2
$J^2_{n,I}$	6.2	92.6	77.4	37.1
$D^2_{n,C,m}$	4.7	7.0	75.2	77.0
$D^2_{n,C,v}$	5.4	89.3	6.2	15.6
$J^2_{n,C}$	5.6	82.0	57.6	63.9
$M_{n,2}$	3.2	79.9	44.1	43.1
$M_{n,6}$	3.3	76.6	92.4	34.3
$M_{n,10}$	4.7	66.5	92.5	27.4

Table 4. Empirical Size and Power of Tests at 5%. Conditional Mean and Variance Models.

$n = 200$	**AR(1)-CH(1)**	**AR(1)-BIL**	**AR(2)-CH(1)**	**TAR**
$D^2_{n,I,m}$	5.3	7.5	100.0	91.5
$D^2_{n,I,v}$	6.3	97.5	3.5	39.0
$J^2_{n,I}$	5.9	94.0	100.0	75.0
$D^2_{n,C,m}$	6.9	7.9	99.5	97.0
$D^2_{n,C,v}$	6.3	92.0	1.0	37.0
$J^2_{n,C}$	6.6	85.0	74.0	92.5
$M_{n,2}$	4.0	99.0	87.0	80.0
$M_{n,6}$	3.3	99.5	100.0	69.0
$M_{n,10}$	4.6	95.5	100.0	57.5

Table 5. Estimates of AR(1)-GARCH(1,1) model for the S&P500 daily stock index.

Parameters	Our Estimate	Bera and Higgins (1997)'s Estimate)
β_0	0.059 (0.026)	0.052 (0.025)
β_1	0.080 (0.032)	0.066 (0.031)
α_1	0.049 (0.014)	0.011 (0.006)
α_2	0.026 (0.008)	0.013 (0.005)
α_3	0.890 (0.029)	0.968 (0.013)