



CAEPR Working Paper  
#2007-009

# Joint and Marginal Diagnostic Tests for Conditional Mean and Variance Specifications

Juan Carlos Escanciano  
Indiana University Bloomington

June 12, 2007

This paper can be downloaded without charge from the Social Science Research Network electronic library at: <http://ssrn.com/abstract=993205>.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at [caepr@indiana.edu](mailto:caepr@indiana.edu) or via phone at 812-855-4050.

©2007 by Juan Carlos Escanciano. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

# JOINT AND MARGINAL DIAGNOSTIC TESTS FOR CONDITIONAL MEAN AND VARIANCE SPECIFICATIONS\*

J. CARLOS ESCANCIANO<sup>†</sup>

*Indiana University*

## Abstract

This article proposes a general class of joint and marginal diagnostic tests for parametric conditional mean and variance models of possibly nonlinear non-Markovian time series sequences. The use of joint and marginal tests is motivated from the fact that marginal tests for the conditional variance may lead misleading conclusions when the conditional mean is misspecified. 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, the proposed tests are robust to higher order dependence of unknown form, in particular to conditional skewness and kurtosis. It turns out that the asymptotic null distributions of the new tests depend on the data generating process, so a new 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 modeling. Finally, an application to the S&P 500 highlights the merits of our approach.

---

\**JEL Classification:* C12, C14, C52.

*Key words and phrases.* Diagnostic tests; Model checks; Generalized spectral analysis; Nonlinear time series; Wild bootstrap; Volatility model.

<sup>†</sup>Corresponding address: Indiana University, Department of Economics, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA, e-mail: jescanci@indiana.edu.

## 1. INTRODUCTION

Specification analysis of volatility models is of major interest in empirical finance and economics. Volatility modeling is an essential factor in asset pricing and market risk management, see e.g. the Value at Risk methodology in Jorion (1997). It also plays an important role in asset allocation under the mean-variance framework and in studies of the intertemporal relation between risk and return (see e.g. Glosten, Jagannathan and Runkle, 1993.) In econometrics, many inference procedures proposed in the literature depend crucially on the correct joint specification of the conditional mean and variance<sup>1</sup>. Despite the significant empirical and theoretical importance of this testing problem, very few conditional variance specification tests have been proposed. Furthermore, the proposed tests may not be adequate for the kind of time series data arising in financial applications, which may possess time-varying higher order conditional moments of unknown form and highly persistent nonlinear dependence. The aim of this paper is to construct new joint and marginal specification tests especially convenient for financial and economic applications. The use of joint and marginal tests pursued in this article is motivated from the fact that the conditional variance specification may be misleading when the conditional mean is misspecified.

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  $(\Omega, \mathcal{F}, P)$ . Henceforth  $A'$  denotes the matrix transpose of  $A$ . Here  $Y_t$  is the dependent random variable (r.v.) and let  $Z_{t-1} = (Y_{t-1}, X'_t)' \in \mathbb{R}^{1+m}$ ,  $m \in \mathbb{N}$ , be the explanatory random vector containing the lagged value of the dependent variable and other explanatory variables  $X_t$ , say. In this paper we are mainly concerned with the case in which the conditioning set at time  $t - 1$  is given by  $I_{t-1} = (Z'_{t-1}, Z'_{t-2}, \dots)'$ . 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  $f(I_{t-1}) = E[Y_t | I_{t-1}]$  is almost surely (a.s.) the conditional mean,  $h^2(I_{t-1}) = Var[Y_t | I_{t-1}]$  is a.s. the conditional variance, and  $u_t = (Y_t - f(I_{t-1}))/h(I_{t-1})$ ,  $t \in \mathbb{Z}$ , are standardized errors. Let  $\mathcal{M} = \{f(\cdot, \theta), h^2(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$  be a given parametric family of functions and consider the 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 specifications for  $f(I_{t-1})$  and  $h(I_{t-1})$ , respectively, and  $\{u_t(\theta)\}$

---

<sup>1</sup>In estimation theory, consistency of classical estimators such as the Quasi-Maximum Likelihood Estimator (QMLE), or efficiency improvements in Wefelmeyer (1996), depend crucially on the correct joint specification. In testing theory, tests for conditional symmetry in Bai and Ng (2001), goodness-of-fit tests for conditional distributions in Bai (2003) and Koul and Ling (2006) or goodness-of-fit test for copulas in Chen, Fan and Patton (2003), among many others, also rely on the correct specification of a mean-variance model.

is a sequence of disturbances of the model, defined implicitly from (1). Following the financial literature parlance we also refer to  $h(I_{t-1}, \theta)$  as a *volatility model* for  $h(I_{t-1})$ .

Since the seminal work on the ARCH model by Engle (1982) there has been a vast quantity of research uncovering the properties of competing volatility models. As pointed out by Engle (2002), the number of new models proposed, estimated and analyzed has been dramatic. The most influential models were the first: the GARCH model of Bollerslev (1986) and the EGARCH model of Nelson (1991). There is already a well-developed theory for many aspects of the aforementioned models, including theorems for stationarity and ergodicity, moments, and estimation. See Li, Ling and McAleer (2002) and Straumann (2005) for recent surveys. Surprisingly enough, specification analysis for these models is less elaborated, with very few proposals available in the literature. This paper contributes to this literature by proposing a general methodology for testing the adequacy of a possibly nonlinear mean-variance model under fairly general regularity conditions. Given the large supply of existing competing models a specification test for evaluation of a mean-variance model is very welcome.

The specification in (1) covers the well-known classes of linear ARMA-ARCH, ARMA-(E)GARCH models, as well as many other nonlinear conditional mean and variance models, see, e.g., Fan and Yao (2003) for some review. In empirical finance an important class is the ARCH in mean (ARCH-M) models of Engle, Lilien and Robins (1987) and its generalizations, where the conditional variance enters in the specification of the conditional mean in order to assess return-risk relationships. Our parametrization in (1) is general enough to allow for (G)ARCH-M models.

Our tests are constructed from the fact that under the correct specification of  $f(I_{t-1}, \theta_0)$  and  $h(I_{t-1}, \theta_0)$ ,  $\{u_t(\theta_0)\}$  becomes a martingale difference sequence with respect to  $\mathcal{F}_t$ , the  $\sigma$ -field generated by  $I_t$ , with zero mean and unit conditional variance. That is, the correct joint specification is tantamount to

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

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 in  $H_0$  is responsible for the correct specification of the conditional mean, whereas both conditional moment restrictions are necessary for the adequacy of the conditional variance. Notice that the null hypothesis  $H_0$  is less stringent than the classical independent and identically distributed (i.i.d.) assumption on standardized errors  $\{u_t(\theta_0)\}$ , which is systematically assumed in the literature. There is now, however, a growing econometrics and financial 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 Siddique (1999, 2000) and Jondeau and Rockinger (2003). In this paper we are consistent with this existing literature and propose tests for  $H_0$  allowing for unknown serial dependence and time-varying higher order conditional moments in  $u_t(\theta_0)$  given  $I_{t-1}$ .

Neglected higher order dependence can cause misleading conclusions in existing diagnostic tests for volatility models, see our application to the S&P 500.

There is already an extensive literature on testing the correct specification of a parametric dynamic conditional mean model, see Escanciano (2006a) for a recent review. On the contrary, the literature on joint specification tests of conditional mean and variance functions is very scarce<sup>2</sup>. The existing proposals are extensions of mean's specification tests to joint specification tests for mean and variance. Ngatchou-Wandji (2005) proposed  $\chi^2$ -discrepancy measures that, although being simple, fail to be consistent against a large class of alternatives of the correct specification. 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.

The most used diagnostic tools in empirical research are 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) and Hidalgo and Zaffaroni (2006). 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  $I_{t-1}^u = (u_{t-1}(\theta_0), u_{t-2}(\theta_0), \dots)'$ , and thus, condition (2) yields

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

The latter point motivates some authors to consider specification tests 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). As a result, tests based on correlation or autocorrelation measures are not consistent for any misspecified model with uncorrelated errors (centered square errors). These tests may incur in an increase of the Type II error probability.

There are at least two limitations of the aforementioned works that one may consider important. First, the proposed tests only allow for a finite-dimensional conditioning set  $I_{t-1}$ . Notice that common models for the conditional variance are non-Markovian, e.g. the popular GARCH models or the ARCH( $\infty$ ) model of Robinson (1991). See also Linton and Mammen (2005). Markovian models, such as the ARCH(d) model, are known not to capture the dynamics well<sup>3</sup>. In particular, a well-known “stylized fact” in financial data is a highly persistent volatility, which is consistent with significative conditioning variables in the variance specification at long lags<sup>4</sup>.

<sup>2</sup>The problem of testing simultaneously many conditional moment restrictions has already been considered in, e.g. Chen and Fan (1999), under mixing data, or in Delgado, Dominguez and Lavergne (2006) for independent data. However, none of these tests have considered the problem we deal with here.

<sup>3</sup>See e.g., Ding et al. (1993), Baillie et al. (1996), Ding and Granger (1996), Breidt et al. (1998), Andersen et al. (2001) and Mikosch and Starica (2003) for evidence of high persistence in financial data.

<sup>4</sup>Nelson (1991) argues that as the frequency at which the data are sampled becomes very high, persistence should

Second, a distinguished feature of the simultaneous testing problem in (2) is that the correct specification of the conditional variance generally depends on the correct specification of the conditional mean. A possible solution to this problem requires nonparametric estimation of the conditional mean, which involves the choice of bandwidth parameters, see e.g. Blake and Kapetanios (2006). This problem becomes relevant if the focus is in the conditional variance and neglected nonlinearity in the conditional mean is a likely event. This is the case in most financial applications. In this paper we recommend the use of marginal and joint tests to identify the cause of the rejection of the joint test. Our application to the S&P 500 stock index nicely illustrates this point.

In this paper we extend the generalized spectral distribution approach in Escanciano (2006a) from mean diagnostics to simultaneous and marginal mean-variance specifications. The generalized spectrum is especially convenient when dealing with infinite-dimensional information sets and non-Markovian processes. A novel contribution of the paper is the development and theoretical justification of a bootstrap procedure to approximate the critical values of tests and the use of joint and marginal tests. 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 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  $H_0$ , while being robust to higher order (unknown) conditional dependence; (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  $n^{-1/2}$ , with  $n$  the sample size; and (vi) the tests are simple to compute and are valid under fairly general regularity conditions on the underlying data generating process (DGP).

The remainder of this paper is organized as follows. In Section 2 we present the generalized spectral distribution tests for testing  $H_0$ . In Section 3 we study the asymptotic distribution of our tests under the null. We propose and justify theoretically a bootstrap method to implement the tests in Section 4. Finally, we make an extensive simulation exercise and an empirical application to the S&P 500 stock index in Section 5. All proofs are gathered in an appendix. Throughout,  $A^c$  and  $|A|$  denote the complex conjugate 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  $n \rightarrow \infty$ . In the sequel  $C$  is a generic constant that may change from one expression to another.

---

become larger.

## 2. THE INTEGRATED GENERALIZED SPECTRAL TESTS

Our methodology for testing  $H_0$  relies on a pairwise generalized spectrum approach that has been shown to be very useful in a variety of testing problems, see, e.g., Hong (1999), Escanciano and Velasco (2006) and Escanciano (2006a), among others. The rationale for our test is as follows. Under  $H_0$ ,

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

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

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

See Stinchcombe and White (1998) and Escanciano (2006a) for primitive conditions on the family  $\mathcal{W}$  to satisfy the equivalence between (4) and (5). The nuisance parameter space  $\Upsilon$  and its dimension  $s$  depend on the particular family  $\mathcal{W}$  used. Common examples of weight functions  $w$  are  $w(Z_{t-j}, x) = 1(Z_{t-j} \leq x)$ , with  $x \in \Upsilon_{ind} \equiv [-\infty, \infty]^{m+1}$ , so  $s = m+1$  here, and where  $1(A)$  denotes the indicator of the event  $A$ , or  $w(Z_{t-j}, x) = \exp(ix'Z_{t-j})$ , with  $i = \sqrt{-1}$  and  $x \in \Upsilon_{exp} \equiv \mathbb{R}^{m+1}$ . The weight  $w(Z_{t-j}, x) = 1(\beta'Z_{t-j} \leq u)$ , where  $x = (\beta', u)' \in \Upsilon_{pro} \equiv \mathbb{S}^{m+1} \times [-\infty, \infty] \subset [-\infty, \infty]^s$ , with  $s = m+2$  and  $\mathbb{S}^{m+1} = \{\beta \in \mathbb{R}^{m+1} : |\beta| = 1\}$ , was proposed in Escanciano (2006b) and leads to powerful tests. The present tests are based on a generalized spectral distribution function approach using the measures  $\{\gamma_{j,w}(\cdot, \theta_0)\}_{j=1}^{\infty}$  for a generic  $w$ .

We consider simultaneously all the dependence measures  $\{\gamma_{j,w}(\cdot, \theta_0)\}_{j=1}^{\infty}$  in (5) as follows. Define  $\gamma_{-j,w}(\cdot, \theta_0) := \gamma_{j,w}(\cdot, \theta_0)$  for  $j \geq 1$  and  $\gamma_{0,w}(\cdot, \theta_0) := E[e_t(\theta_0)w(Z_t, x)]$ , and denote the Fourier transform of the functions  $\{\gamma_{j,w}(\cdot, \theta_0)\}_{j=-\infty}^{\infty}$  by

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

which contains the same information about  $H_0$  as the whole sequence  $\{\gamma_{j,w}(x, \theta_0)\}_{j=0}^{\infty}$ .

Note that under  $H_0$ ,  $f_w(u, x, \theta_0) \equiv f_{0,w}(x, \theta_0) = (2\pi)^{-1}\gamma_{0,w}(x, \theta_0)$ , which can serve as the basis upon which a test for (2) is constructed. Nonparametric smoothed estimation is necessary for estimating  $f_w(u, x, \theta_0)$ . This can be avoided by means of 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.,

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

which after some manipulation can be written as

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

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  $(Z'_{t-1}, Z'_{t-2}, \dots, Z'_0)'$  and that may contain some initial values. We obtain residuals

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

where  $\theta_n$  is a  $\sqrt{n}$ -consistent estimator for  $\theta_0$ , e.g., the Quasi-Maximum Likelihood Estimator (QMLE). The sample version of  $\gamma_{j,w}(x, \theta_0)$  for  $j \geq 0$  is then given by

$$\widehat{\gamma}_{j,w}(x, \theta_n) = \frac{1}{n_j} \sum_{t=j}^n \widehat{e}_t w(Z_{t-j}, x), \quad n_j = n - j + 1, \quad \widehat{e}_t = (\widehat{e}_{1t}, \widehat{e}_{2t})'. \quad (9)$$

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 leads to 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,

$$J_{n,w}^2(\theta_n) = \int_{\Pi} |S_{n,w}(\lambda, x, \theta_n)|_M^2 W(dx) d\lambda = \sum_{j=1}^n \frac{n_j}{(j\pi)^2} \int_{\Upsilon} |\widehat{\gamma}_{j,w}(x, \theta_n)|_M^2 W(dx), \quad (10)$$

where  $W(\cdot)$  is an integrating function depending on the weight family  $\mathcal{W}$  and satisfying some mild conditions (see Assumption A5 below) and  $M$  is a  $2 \times 2$  positive semidefinite matrix. Our tests reject  $H_0$  for “large” values of  $J_{n,w}^2(\theta_n)$ . Note that  $J_{n,w}^2(\theta_n)$  uses all lags contained in the sample, it 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  $J_{n,w}^2(\theta_n)$  in directing the power against some desired directions (see Escanciano (2007b) for discussion regarding  $w$  and  $W$ .) In the simulations we choose  $M$  with rows  $(m_1, 0)$  and  $(0, m_2)$ . The nonnegative weights  $m_1$  and  $m_2$  represent the contribution of the marginal components (i.e. mean and variance) to the joint test. For instance, the choice  $m_1 = 1$  and  $m_2 = 0$  leads to the marginal mean test which is “tilted” against mean misspecifications. In this paper we advocate for the use of joint tests based on  $J_{n,w}^2(\theta_n)$  and marginal tests derived from the choices  $m_1 = 1$  and  $m_2 = 0$  (mean component) and  $m_1 = 0$  and  $m_2 = 1$  (variance component.) This constitutes an interesting situation that marks departure from other approaches that exist in the literature. See Section 5 for the use of joint and marginal tests.



### 3. ASYMPTOTIC NULL DISTRIBUTION

To elaborate the asymptotic theory we consider the following assumptions. Recall that  $e_{1t}(\theta) = Y_t - f(I_{t-1}, \theta)$  and  $e_{2t}(\theta) = e_{1t}^2(\theta) - h^2(I_{t-1}, \theta)$ , with  $\theta \in \Theta \subset \mathbb{R}^p$ . We define the score  $g(I_{t-1}, \theta)$  with rows  $g'_{1t}(\theta)$  and  $g'_{2t}(\theta)$ , given by

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

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

*Assumption A1:*

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

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

*Assumption A2:* Let  $\Theta_0$  be a small convex neighborhood of  $\theta_0$ . 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  $\Theta$  is compact in  $\mathbb{R}^p$ . The parameter  $\theta_0$  belongs to the interior of  $\Theta$ .

A3(b): The estimator  $\theta_n$  satisfies the asymptotic Bahadur 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  $L(\theta_0) = E[m(I_{t-1}, \theta_0) e_t(\theta_0) e_t'(\theta_0) m'(I_{t-1}, \theta_0)]$  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  $2 \times 2$  positive semidefinite 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

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

whenever  $\{(\zeta_t, \xi_t), t = 0, \pm 1, \dots\}$  is a strictly stationary and ergodic process with  $\zeta_t \in \mathbb{R}$ ,  $\xi_t \in \mathbb{R}^{1+m}$ ,  $E|\zeta_1| < \infty$ , and  $\Upsilon_c$  is any compact subset of  $\Upsilon \subset [-\infty, \infty]^s$ .

*Assumption A5:* Let  $\Theta_0$  be a small neighborhood around  $\theta_0$ . The observed information set available at period  $t$ ,  $\widehat{I}_t$ , may contain some assumed initial values and satisfies

$$\left( \sum_{t=1}^n \left[ E \sup_{\theta \in \Theta_0} \left| \left( f(\widehat{I}_{t-1}, \theta) - f(I_{t-1}, \theta) \right) \right|^2 \right]^{1/2} \right)^2 = o(n),$$

$$\left( \sum_{t=1}^n \left[ E \sup_{\theta \in \Theta_0} \left| \left( Y_t - f(\widehat{I}_{t-1}, \theta) \right)^2 - \left( Y_t - f(I_{t-1}, \theta) \right)^2 \right|^2 \right]^{1/2} \right)^2 = o(n)$$

and

$$\left( \sum_{t=1}^n \left[ E \sup_{\theta \in \Theta_0} \left| \left( h^2(\widehat{I}_{t-1}, \theta) - h^2(I_{t-1}, \theta) \right) \right|^2 \right]^{1/2} \right)^2 = o(n).$$

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. Note that our conditions permit long memory processes. Assumption A2 is on the model and is standard in the literature, see, e.g., Escanciano (2006a). Assumption A3 is satisfied under mild conditions for most estimators. Conditions for A3 for the local QMLE under martingale conditions have been established in Lee and Hansen (1994). See also, among others, Hall and Heyde (1980, Chapter 6), Horváth et al. (2001), Francq and Zakořan (2004) and Straumann (2005). As shown in e.g. Francq and Zakořan (2004) the QMLE of pure GARCH models,  $\theta_n$  say, satisfies A3 under mild conditions and  $H_0$ , with

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n J^{-1} \frac{1}{h^4(I_{t-1}, \theta_0)} g_{2t}(\theta_0) e_{2t}(\theta_0) + o_P(1),$$

where  $J \equiv J(\theta_0) = E[h^{-4}(I_{t-1}, \theta_0) g_{2t}(\theta_0) g'_{2t}(\theta_0)]$ . Extensions to ARMA-GARCH models can be also found in Francq and Zakořan (2004).

Examples of  $W(\cdot)$  include the cumulative distributions functions (cdf) of a  $N(0,1)$ , Double Exponential or the Student's  $t_\nu$  distribution. The reader is referred to Escanciano and Velasco (2006) for discussions on the choice of  $W$ . All previous examples of functions  $w$  satisfy A4, see the discussion in the previous section. 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  $\eta = (\lambda, x)' \in \Pi$ . We consider  $S_n(\lambda, x, \theta_n) \equiv S_n(\eta, \theta_n)$  as a random element on the Hilbert space  $L_2(\Pi, \nu, M)$  of all bivariate complex-valued and square  $\nu$ -integrable functions on  $\Pi$ , where  $\nu$  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  $\implies$  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.,  $Z_n \xrightarrow{L_2} Z \iff \|Z_n - Z\| \xrightarrow{P} 0$ . Let us define  $\Psi_j(\lambda) = \sqrt{2}(\sin j\pi\lambda)/j\pi$ ,  $b_j(x, \theta_0) = E[w_{t-j}(x)g_t(\theta_0)]$ ,  $G_w(\eta) = \sum_{j=1}^{\infty} b_j(x, \theta_0)\Psi_j(\lambda)$ ,  $Q_{\infty, w}(\eta) = \sum_{j=1}^{\infty} w_{t-j}(x)\Psi_j(\lambda)$ , and

$$\sigma_h^2 = E[\langle e_t(\theta_0)Q_{\infty, w}, h \rangle^2], \quad h \in L_2(\Pi, \nu, M). \quad (11)$$

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  $\sigma_h^2$  is defined in (11). 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*

$$\text{Cov}(S_w^0(\eta)', V) = E[m(I_{t-1}, \theta_0)e_t(\theta_0)e_t'(\theta_0)Q_{\infty, w}(\eta)].$$

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 (local) power properties of  $J_{n, w}^2(\theta_n)$  can be studied along the arguments in Escanciano (2006a). 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  $J_{n, w}^2(\theta_n)$  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 serial 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, and Escanciano (2007a) for time series sequences. It is shown in these papers that the most relevant bootstrap method for regression problems is the wild bootstrap introduced in Wu (1986) and Liu (1988). Here we extend the wild bootstrap to our present context. We approximate the asymptotic null distribution of  $S_{n,w}$  by that of

$$S_{n,w}^*(\lambda, x, \theta_n^*) \equiv (S_{n,w,m}^*(\lambda, x, \theta_n^*), S_{n,w,v}^*(\lambda, x, \theta_n^*))' = \sum_{j=1}^n n_j^{1/2} \hat{\gamma}_j^*(x) \Psi_j(\lambda),$$

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

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

and

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

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

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

**Step 2:** Generate wild bootstrap residuals according to  $\hat{\varepsilon}_{1t}^* = \hat{e}_{1t}(\theta_n) V_t$  and  $\hat{\varepsilon}_{2t}^* = \hat{e}_{2t}(\theta_n) V_t$  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 independent of the sequence  $\{(Y_t, \hat{I}_{t-1})'\}_{t=1}^n$ .

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

$$Y_{1t}^* = f(\hat{I}_{t-1}, \theta_n) + \hat{\varepsilon}_{1t}^* \text{ for } 1 \leq t \leq n, \quad (12)$$

and

$$Y_{2t}^* = h^2(\hat{I}_{t-1}, \theta_n) + \hat{\varepsilon}_{2t}^* \text{ for } 1 \leq t \leq n. \quad (13)$$

**Step 4:** Compute  $\theta_n^*$  from the bootstrap data  $\{Y_{1t}^*, Y_{2t}^*, \hat{I}_{t-1}\}_{t=1}^n$  to construct  $\hat{e}_{1t}^* = Y_{1t}^* - f(\hat{I}_{t-1}, \theta_n^*)$  and  $\hat{e}_{2t}^* = Y_{2t}^* - h^2(\hat{I}_{t-1}, \theta_n^*)$ , for  $t = 1, \dots, n$ .

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

$$P(V_t = 0.5(1 - \sqrt{5})) = b \text{ and } P(V_t = 0.5(1 + \sqrt{5})) = 1 - b, \quad (14)$$

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

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

That is, the bootstrap distribution

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

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  $F_J^*(c_{n,\alpha}^*) = 1 - \alpha$  (a.s.). Also, we can use the bootstrap p-values,  $p_n^*$  say, rejecting  $H_0$  when  $p_n^* < \alpha$ , where  $p_n^* = P\left(J_{n,w}^{2*} \geq J_{n,w}^2(\theta_n) \mid \{Y_t, \hat{I}_{t-1}\}_{t=1}^n\right)$ . 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  $\eta_n^*$  converges in probability a.s. to  $\eta_n$  if for all  $\delta > 0$ ,  $P\left(|\eta_n^* - \eta_n| \geq \delta \mid \{Y_t, \hat{I}_{t-1}\}_{t=1}^n\right) \rightarrow 0$  a.s., which is expressed as  $\eta_n^* = \eta_n + o_P(1)$  a.s. Define weak convergence almost surely under the bootstrap law, denoted by  $\xrightarrow{*}$  a.s., as

$$R_n^* \xrightarrow{*} R \text{ a.s.} \iff E\left(f(R_n^*) \mid \{Y_t, \hat{I}_{t-1}\}_{t=1}^n\right) \rightarrow E\left(f(R) \mid \{Y_t, \hat{I}_{t-1}\}_{t=1}^n\right) \text{ a.s.}$$

for any continuous and bounded real valued function  $f$  on  $L_2(\Pi, \nu, M)$ .

In order to show that the bootstrap assisted tests are valid, we need to assume that the bootstrap analog of  $\theta_n$ ,  $\theta_n^*$  in step 4 above, satisfies an asymptotic expansion like A3(b) in the bootstrap world. Conditions for the satisfaction of A6 below have to be studied on a case-by-case basis. See Dominguez (2004) for an example of analysis with the nonlinear least squares estimator in conditional mean models.

*Assumption A6:*

A6(a): The estimator  $\theta_n^*$  satisfies the asymptotic expansions

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

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

A6(b): There exists a unique  $\theta_1 \in \Theta$  such that  $|\theta_n - \theta_1| = o_P(1)$ . Moreover, A5 is satisfied replacing  $\Theta_0$  by  $\Theta_1$ , a small neighborhood around  $\theta_1$ .

A6(c):  $E\left[\sup_{\theta \in \Theta_1} |m(I_{t-1}, \theta)|\right] \leq C$ ,  $E\left[\sup_{\theta \in \Theta_1} |m(I_{t-1}, \theta) e_t(\theta)|^2\right] \leq C$ ,  $E\left[\sup_{\theta \in \Theta_1} |e_{ht}(\theta)|^2\right] \leq C$  and  $E\left[\sup_{\theta \in \Theta_1} |m(I_{t-1}, \theta) e_t(\theta) e_{ht}(\theta)|\right] \leq C$  for  $h = 1, 2$ .

**Theorem 2** *Assume A1-A6. Then*

$$S_{n,w}^* \xrightarrow[*]{\Rightarrow} \tilde{S}_w, \text{ a.s.},$$

where  $\tilde{S}_w$  is the same Gaussian process of Theorem 1 but with  $\theta_1$  replacing  $\theta_0$ .

There is a particular case of Theorem 2 that is worth to be mentioned. Consider the situation where the conditional mean is misspecified, so  $P(E[e_{1t}(\theta) | I_{t-1}] = 0) < 1$  for all  $\theta \in \Theta$ , but the second conditional moment restriction in (2) holds, i.e.,  $E[e_{2t}(\theta_0) | I_{t-1}] = 0$ . From the latter restriction, one can usually obtain an estimator  $\theta_n$  such that  $|\theta_n - \theta_0| = o_P(1)$ . One can prove that under such circumstances, the bootstrap marginal test for testing  $E[e_{2t}(\theta_0) | I_{t-1}] = 0$  still yields valid inferences. Of course, in such a situation we are not modeling the conditional variance but the dispersion from other location parameter  $f(I_{t-1}, \theta_0)$ . For instance, econometricians in financial applications often specified  $f(I_{t-1}, \theta_0) \equiv 0$ . Our previous discussion guarantee that as long as  $E[Y_t^2 | I_{t-1}] = h^2(I_{t-1}, \theta)$  and  $|\theta_n - \theta_0| = o_P(1)$  hold, our bootstrap conditional variance test yields valid inference, so it is not affected by the conditional mean misspecification. This is a robust property of our bootstrap tests not shared by other existing testing methods.

Theorem 2 shows that the bootstrap assisted tests lead to correct critical values under the null hypothesis. Following Escanciano (2006a), it can be shown that the bootstrap tests are consistent and preserve the asymptotic local power properties of the tests based on  $S_{n,w}$ . Next section investigates the finite sample performance of the proposed bootstrap procedures.

## 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) ( $M_{n,p}$ ), the Portmanteau tests of Li and Mak's (1994) ( $LM_m$ ) and the joint mean-variance extensions of the some tests considered in Escanciano (2007a), see also Koul and Stute (1997). 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.,

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

where the subindex  $I$  in  $\hat{\gamma}_{I,j} = (\hat{\gamma}_{I,j,m}, \hat{\gamma}_{I,j,v})'$  corresponds to the use of  $w(Y_{t-j}, x) = 1(Y_{t-j} \leq x)$  in (9) and

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

Note that the use of the empirical cdf does not affect the asymptotic theory, see Escanciano (2006a). 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  $J_{n,C}^2$ ,  $D_{n,C,m}^2$  and  $D_{n,C,v}^2$  based on  $w(Y_{t-j}, x) = \exp(ixY_{t-j})$  and the integrating function  $\phi$ , the density function of the standard normal random variable, which yields the test statistic

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

Our joint test statistics  $J_{n,I}^2$  and  $J_{n,C}^2$  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  $(0, m_2)$ .

For a null ARMA-ARCH( $r$ ) model, the simpler version of Li and Mak's test statistic can be written as

$$LM_m = n(n+2) \sum_{j=r+1}^m (n-j)^{-1} \hat{\rho}_2^2(j) \rightarrow_d \chi_{m-r}^2, \quad m > r,$$

where  $\hat{\rho}_2^2(j)$  is the sample autocorrelation function of  $\{u_t^2(\theta_0)\}$ . The null limit distribution depends on  $r$ , the order of the ARCH model. When other conditional variance models are estimated, the test statistic itself has to be modified as suggested by Li and Mak (1994) or by Lundbergh and Teräsvirta (2002). Li and Mak's test  $LM_m$  is checking for uncorrelation in  $\{u_t^2(\theta_0)\}$  up to order  $m$ , see the discussion around (3).

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. It is worth to stress that 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. Hong and Lee's (2003) test is given by

$$M_{n,p} = \left[ HL_{n,p} - \hat{C}_0 K_2 \right] / \left[ 2 \hat{D}_0 K_4 \right], \quad (15)$$

with  $HL_{n,p}$  defined by

$$HL_{n,p} = \int \sum_{j=1}^{n-1} k^2(j/p) (n-j) |\hat{\sigma}_j(y, x, \theta_n)|^2 W(dy) W(dx), \quad (16)$$

where  $\hat{\sigma}_j(y, x, \theta_n)$  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

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

and

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

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  $k(z) = \sin(\pi z)/\pi z$ .

Escanciano (2007a) proposed bootstrap specification tests based on a finite dimensional conditioning set, in the spirit of Koul and Stute (1999). Here we consider extensions to joint tests for mean and variance specifications. Though such extensions have not been justified yet, the present paper provides evidence of their theoretical validity. We denote by  $CvM_P$  the Cramér-von Mises statistic, with  $P$  as the number of lags used. These statistics are based on the multivariate integrated regression functions, i.e.

$$CvM_P = \frac{m_1}{\widehat{\sigma}_{1e}^2 n^2} \sum_{j=1}^n \left[ \sum_{t=1}^n \widehat{e}_{1t}(\theta_n) 1(I_{t-1,P} \leq I_{j-1,P}) \right]^2 + \frac{m_2}{\widehat{\sigma}_{2e}^2 n^2} \sum_{j=1}^n \left[ \sum_{t=1}^n \widehat{e}_{2t}(\theta_n) 1(I_{t-1,P} \leq I_{j-1,P}) \right]^2,$$

where  $I_{t-1,P} = (Y_{t-1}, \dots, Y_{t-P})$  is the  $P$ -lagged values of the series. Again,  $CvM_{P,m}$  and  $CvM_{P,v}$  correspond to the respective choices of  $(m_1, m_2) = (1, 0)$  and  $(m_1, m_2) = (0, 1)$  in  $CvM_P$ .

Throughout  $\varepsilon_t$  and  $v_t$  are independent sequences of i.i.d.  $N(0, 1)$ . We consider the nominal level 5% in all tests and simulations. 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 to eliminate the effect on initial values in the generation of the data. For the bootstrap approximation we employ a sequence  $\{V_t\}$  of i.i.d Bernoulli variates given in (14). The power in the non-bootstrap cases is size-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  $p$  from  $p = 2$  to 11. For Li and Mak's (1994) ( $LM_m$ ) test we use  $m$  from  $m = 2$  to 11, and for  $CvM_P$  we use  $P = 1, 3, 5, 7$  and 9. For simplicity, we only present in tables the values  $m, p = 2, 6$  and 10 and  $P = 1, 3$  and 5.

## 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  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$  with  $M_{n,p}$ ,  $CvM_{P,v}$  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:

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

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 for a better discrimination among the tests). To compute the statistics  $D_{n,I,v}^2$ ,  $D_{n,C,v}^2$  and  $CvM_{P,v}$  we use the residuals  $\hat{e}_{2t}(\theta_n) := Y_t^2 - h^2(Y_{t-1}, \theta_n)$  where  $h^2(Y_{t-1}, \theta_n) = \hat{a} + \hat{b}Y_{t-1}^2$ , and  $\theta_n = (\hat{a}, \hat{b})$  is the least squares estimators 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  $\hat{u}_t(\theta_n) = Y_t/h(Y_{t-1}, \theta_n)$ .

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 sample size is  $n = 100$ . The tests  $D_{n,I,v}^2$ ,  $D_{n,C,v}^2$ ,  $CvM_{P,v}$ ,  $LM_m$  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  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$  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  $D_{n,C,v}^2$  outperforms  $D_{n,I,v}^2$  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. The generalized Koul and Stute (1999) test  $CvM_{P,v}$  is very sensitive to the lag order  $P$ , see the results for ARCH(2), EGARCH, LM and NLMA. The best choice for  $P$  in  $CvM_{P,v}$  depends on the alternative at hand and leads to a test comparable to  $D_{n,I,v}^2$ , but worse than  $D_{n,C,v}^2$ . Hong and Lee's (2003) test  $M_{n,p}$

has good empirical power against ARCH(2), EGARCH, BIL and LOGISTIC MAP 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.

---

TABLE 1 ABOUT HERE

---

It is shown in these simulations that  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$  have omnibus power against all linear and nonlinear alternatives considered. Notably, the exponential based test  $D_{n,C,v}^2$  has excellent empirical power properties against all considered alternatives, being the best test in many cases. These simulations confirm that in practice it could be better to consider a pairwise spectral approach than a joint approach as in  $CvM_{P,v}$ . Even for the optimal choice of  $P$ ,  $CvM_{P,v}$  leads to a test which is less powerful than  $D_{n,C,v}^2$ . Also our tests are in many cases more powerful than Hong and Lee's (2003) test  $M_{n,p}$ , even when the latter take into account additional information from the serial dependence of the standardized innovations in higher moments than those considered in (2). Now, we consider joint conditional mean and conditional variance models.

## 5.2 Joint Specifications of Conditional Mean and Variance

In this subsection we first 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  $J_{n,I}^2$  and  $J_{n,C}^2$ , with the marginal tests  $D_{n,I,m}^2$ ,  $D_{n,C,m}^2$ ,  $D_{n,I,v}^2$ ,  $D_{n,C,v}^2$ , the tests based on a finite-dimensional information set  $CvM_P$ ,  $CvM_{P,m}$  and  $CvM_{P,v}$ , and with  $M_{n,p}$ . The user-chosen parameters are  $p = 2, 6$  and  $10$  and  $P = 3$ . To save space we have not reported results for the Portmanteau test  $LM_m$ . The simulation design is the same as before.

Our second null model is:

$$Y_t = aY_{t-1} + h_t\varepsilon_t, \quad a \neq 0, \quad h_t^2 = b + cY_{t-1}^2.$$

In the second block of simulations 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. GARCH-M model:  $Y_t = 2.5h_t^2 + u_t$ ,  $u_t = h_t\varepsilon_t$ ,  $h_t^2 = 0.001 + 0.29u_{t-1}^2 + 0.7h_{t-1}^2$ .

4. AR(2)-CH(1) model:  $Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + h_t\varepsilon_t$ ,  $h_t^2 = 0.9 + 0.1Y_{t-1}^2$ .
5. 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  $|Y_{t-1}| > 1$ .

These are well known linear and nonlinear models. We report the RP for  $J_{n,I}^2$ ,  $J_{n,C}^2$ ,  $D_{n,I,m}^2$ ,  $D_{n,C,m}^2$ ,  $D_{n,I,v}^2$ ,  $D_{n,C,v}^2$ ,  $CvM_P$ ,  $CvM_{P,m}$ ,  $CvM_{P,v}$  and  $M_{n,p}$  in Tables 2 and 3. They correspond respectively to samples sizes  $n = 50$  and  $200$ . The empirical size is satisfactory for all tests against the AR(1)-CH(1) model. 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 the marginal and joint tests. Hong and Lee's test has reasonable empirical power. Tests based on a finite-dimensional conditioning set have less power than our spectral-based tests and they are comparable to  $M_{n,p}$ . Among all statistics, our tests  $J_{n,I}^2$  and  $J_{n,C}^2$  have the highest empirical powers against the AR(1)-BIL alternative. The GARCH-M is very popular in financial applications. The coefficient of  $h_t^2$  is called the risk premium parameter is taken to be 2.5, which is reasonable value for this parameter, see e.g. Glosten, Jagannathan and Runkle (1993). The GARCH-M model has a misspecified conditional mean and conditional variance. Hong and Lee's test is the most powerful for this alternative. Our tests and  $CvM_P$ ,  $CvM_{P,m}$ , and  $CvM_{P,v}$  are comparable, and rather surprisingly, marginal tests for the conditional variance have no power against this alternative.

For the AR(2)-CH(1) model both joint tests, those based on the spectral approach and those based on a finite-dimensional conditioning set, have a good empirical power. Again marginal tests for the conditional variance have no power against this alternative. The empirical power of  $M_{n,p}$  is more or less satisfactory but very sensitive to  $p$ . It is important to remark that for this alternative the use of  $P = 1$  in  $CvM_P$  and  $CvM_{P,m}$  leads to tests with no power, with RP of 0.01 and 0.08, respectively. This shows again the importance of the choice of the lag order  $P$  in tests based on finite-dimensional conditioning sets. A practitioner using  $CvM_P$   $P = 1$  may conclude that both the conditional mean and variance are correctly specified. For the TAR model our test statistics  $J_{n,I}^2$ ,  $J_{n,C}^2$ ,  $D_{n,I,m}^2$  and  $D_{n,C,m}^2$  outperform  $M_{n,p}$ . The test statistics  $CvM_3$ ,  $CvM_{3,m}$  and  $CvM_{3,v}$  have very low power against the TAR alternative. Marginal tests for the conditional variance have low power. As in the previous two models, this may be a consequence of the effect of the misspecified conditional mean on the marginal variance test and shows the importance of considering joint and marginal tests. The power increases as the sample size increases for all tests, as expected.

---

TABLES 2 AND 3 ABOUT HERE

---

Finally, we consider the Exponential Autoregressive ARCH(1) model (EXPAR-ARCH(1)):

$$Y_t = 0.9Y_{t-1} + \theta_{01}Y_{t-1} \exp(-0.05Y_{t-1}^2) + h_t\varepsilon_t, h_t^2 = 0.1 + \theta_{02}Y_{t-1}^2.$$

This model was considered in Ngatchou-Wandji (2005). The condition for stationarity and ergodicity of this model is  $|\theta_{01}| + |\theta_{02}| + 1.8\theta_{01} < 0.1$ . The null AR(1) model corresponds to  $\theta_{01} = \theta_{02} = 0$ . We consider in this simulations all combinations among the values  $\theta_{01} = 0$  and  $-2.0$  and  $\theta_{02} = 0$ ,  $0.4$  and  $0.8$ . The Monte Carlo set-up is as in previous simulations. Table 4 reports the RP for this DGP for sample sizes  $n = 50$  and  $n = 100$ .

---

TABLES 4 ABOUT HERE

---

From the results reported in Table 4 we conclude that the empirical size performance of tests is satisfactory for as small sample size as  $n = 50$ , although some underrejection is observed for mean marginal tests. The size already improves for  $n = 100$ . As for the power, we see again that  $D_{n,C,v}^2$  and its components have excellent empirical power properties against these alternatives, being the best tests in all cases. In particular,  $D_{n,C,v}^2$  and its marginals have better empirical power properties than  $M_{n,p}$ , even uniformly in  $p$ , see for instance the case  $\theta_{01} = -2.0$  and  $\theta_{02} = 0$ . The tests based on a finite dimensional conditioning set have low power against these alternatives.

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 is a useful inference procedure for detecting whether the misspecification is in the conditional mean, in the conditional variance or in both. Some examples have shown that some caution is necessary when interpreting the marginal conditional variance tests in the case of a misspecified conditional mean. This motivates our use of marginal and joint tests. The new proposed tests, specially those based on exponential functions, compare very well to the competing tests even in situations favoring the latter ones (e.g. optimal choice of  $P$  in  $CvM_P$  and  $p$  in  $M_{n,p}$ ).

### 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. 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 1138 observations. Bera and Higgins (1997) try to discriminate between an AR(1)-GARCH(1,1) model and a bilinear specification. Their results are inconclusive.

On the other hand, Lumsdaine and Ng (1999) conducted a standard ARCH(1) test to this data set, after fitting an AR(1) model to the conditional mean. The resulting ARCH effects test rejects the null hypothesis of conditional homoskedasticity. These authors proved that if recursive residuals and its squares are included in the mean equation, the ARCH(1) test then fails to reject the null hypothesis of constant conditional variance, in favor of nonlinearity in the conditional mean. These authors conclude: “perhaps some other nonlinear model would be preferred to the GARCH(1,1) specification; accounting for this nonlinearity weakens the evidence in favor of conditional heteroskedasticity”.

We have applied our marginal and joint tests to this data set. Our findings are summarized as follows. A linear AR(1) model for the conditional mean fits the data well, in contrast to the possible nonlinearity (in the mean function) supported in Lumsdaine and Ng (1999). Our marginal tests for the conditional variance favor a conditional homoskedastic model, even in cases with misspecified conditional mean. The AR(1)-GARCH(1,1) model fitted in Bera and Higgins (1997) is highly rejected for this data set.

More concretely, we have fitted a linear AR(1) model with conditional homoskedastic errors to the log differences of the S&P500 ( $Y_t$ ), such as

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \alpha_0 u_t. \quad (17)$$

The bootstrap p-values for  $J_{n,I}^2$ ,  $J_{n,C}^2$ ,  $D_{n,I,m}^2$ ,  $D_{n,C,m}^2$ ,  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$  are, respectively, 0.70, 0.41, 0.33, 0.14, 0.30 and 0.30. Results with  $CvMP$ ,  $CvMP_m$  and  $CvMP_v$  with several values of  $P$  support the correct specification of (17). If the AR(1) component of the mean is neglected, the p-values for  $J_{n,I}^2$ ,  $J_{n,C}^2$ ,  $D_{n,I,m}^2$ ,  $D_{n,C,m}^2$ ,  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$  become, respectively, 0.01, 0.02, 0.01, 0.01, 0.38 and 0.33. From the marginal mean tests  $D_{n,I,m}^2$  and  $D_{n,C,m}^2$  we conclude that the constant conditional mean is rejected at 5%. If the AR(1)-GARCH(1,1) model is fitted, we obtain that the conditional mean is well specified with a p-value of 0.416 for  $D_{n,I,m}^2$  and 0.233 for  $D_{n,C,m}^2$ , whereas, the conditional variance is misspecified, as can be deduced from the zero p-value of  $J_{n,I}^2$ ,  $J_{n,C}^2$ ,  $D_{n,I,v}^2$  and  $D_{n,C,v}^2$ . These results are confirmed with Hong and Lee’s (2003) test.

Therefore, we find that a linear AR(1) model with conditional homoskedastic martingale difference errors fits the S&P500 in this period well, thereby rejecting the AR(1)-GARCH(1,1) model advocated in Bera and Higgins (1997). Our results support that the findings in Lumsdaine and Ng (1999) are perhaps a consequence of the lack of robustness of standard ARCH(1) effects tests to higher order conditional dependence such as conditional kurtosis rather than a consequence of neglected nonlinearity in the conditional mean. Further investigation on this issue is referred for future research. This application highlights the merits of the use of marginal tests and the robustness of our procedures to higher order conditional dependence of unknown form.

## APPENDIX: PROOFS

First, consider three useful lemmas. Lemma 1 is a trivial multivariate extension of Lemma 1 in Escanciano and Velasco (2006). Throughout the proofs

$$Q_{t,w}(\eta) = \sum_{j=1}^t n^{1/2} n_j^{-1/2} w_{t-j}(x) \frac{\sqrt{2} \sin j\pi\lambda}{j\pi}, \quad (18)$$

and

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

Define  $L_2(\Pi, \nu)$  as the Hilbert space of all univariate complex-valued and square  $\nu$ -integrable functions on  $\Pi$ . In  $L_2(\Pi, \nu)$  we define the inner product

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

**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)  $\int_{\mathbb{R}^s} E |h_{j,n}(x)|_M^2 W(dx) < C$  uniformly in  $j \geq 1$ .
- (ii)  $\sup_{x \in \Upsilon_c} |h_{j,n}(x)| = o_p(1) \forall j, 1 \leq j \leq n$ , 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:** Escanciano and Velasco (2006).

**Lemma 2:** *Under A3-A5 the effect of estimating the information set  $I_{t-1}$  by  $\hat{I}_{t-1}$  in  $f(I_{t-1}, \theta_n)$  and  $h(I_{t-1}, \theta_n)$  has no effect on the asymptotic theory. That is,*

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

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

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

Write,

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

where the last equality is due to Minkowski's inequality and A5. Apply Chebyshev's inequality to conclude.

For simplicity, we rename  $\tilde{S}_{n,w}(\eta, \theta_n)$  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,*

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

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

$$S_{n,w}(\eta, \theta_n) = S_{n,w}(\eta, \theta_0) + \frac{\partial S_{n,w}(\eta, \tilde{\theta}_n)}{\partial \theta'}(\theta_n - \theta_0), \quad (19)$$

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

$$S_{n,w}(\eta, \theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t(\theta_n) Q_{t,w}(\eta),$$

where  $Q_{t,w}(\eta)$  is defined in (18). Hence,

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

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

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

The last display, Assumption A3 and (19) imply the result.

**Proof of Theorem 1:** We apply Lemma A2 here and Theorem 1 in Escanciano and Velasco (2006), but with  $w_{t-j}(x)$  replacing  $\exp(ixY_{t-j})$  there, to show that the marginal components of  $S_{n,w} \equiv (S_{n,w,m}, S_{n,w,v})'$  converge weakly in  $L_2(\Pi, \nu)$ . Now, marginal tightness implies joint tightness. Notice that the convergence of the finite dimensional distributions is characterized by the convergence of  $\langle S_{n,w,m}, h \rangle$ ,  $\langle S_{n,w,v}, h \rangle$  and  $\langle S_{n,w,m}, S_{n,w,v} \rangle$ , for all  $h \in L_2(\Pi, \nu)$ . It is an easy matter to verify that these correspond, respectively, to the distributions of  $\langle S_{w,m}, h \rangle$ ,  $\langle S_{w,v}, h \rangle$  and  $\langle S_{w,m}, S_{w,v} \rangle$ , for all  $h \in L_2(\Pi, \nu)$ , where  $S_w = (S_{w,m}, S_{w,v})'$ .

**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:** We need to show that the process  $S_{n,w}^*(\lambda, x, \theta_n^*)$  (conditionally on the sample) has the same asymptotic finite projections as the process  $S_{n,w}(\eta)$ , and that  $S_{n,w}^*(\lambda, x, \theta_n^*)$  is tight. Write  $\widehat{\varepsilon}_t^* = (\widehat{\varepsilon}_{1t}^*, \widehat{\varepsilon}_{2t}^*)'$ , and denote

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

and

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

The tightness of  $S_{n,w}^*(\lambda, x, \theta_n^*)$  follows from the marginal tightness of  $S_{n,w,m}^*(\lambda, x, \theta_n^*)$  and  $S_{n,w,v}^*(\lambda, x, \theta_n^*)$ .

We shall prove the marginal tightness. We write

$$S_{n,w,m}^*(\eta, \theta_n^*) = n^{-1/2} \sum_{t=1}^n \widehat{\varepsilon}_{1t}^* Q_{t,w}(\eta) - n^{-1/2} \sum_{t=1}^n \{f(\widehat{I}_{t-1}, \theta_n^*) - f(\widehat{I}_{t-1}, \theta_n)\} Q_{t,w}(\eta).$$

By Lemma 1 and A5,

$$\left\| n^{-1/2} \sum_{t=1}^n \{f(\widehat{I}_{t-1}, \theta_n^*) - f(\widehat{I}_{t-1}, \theta_n)\} Q_{t,w}(\eta) - I^* - II^* - III^* \right\| = o_P(1),$$

where

$$I^* = n^{1/2} (\theta_n^* - \theta_n)' n^{-1} \sum_{t=1}^n \{g_1(I_{t-1}, \widetilde{\theta}_n^*) - g_1(I_{t-1}, \theta_n)\} Q_{t,w}(\eta),$$

$$II^* = n^{1/2} (\theta_n^* - \theta_n)' n^{-1} \sum_{t=1}^n [g_1(I_{t-1}, \theta_n) Q_{t,w}(\eta) - G_{1w}(\eta, \theta_n)],$$

$$III^* = n^{1/2} (\theta_n^* - \theta_n)' G_{1w}(\eta, \theta_1),$$

$G_{1w}(\eta, \theta_1)$  is the first component of  $G_w(\eta)$ , and  $\widetilde{\theta}_n^*$  satisfies  $|\widetilde{\theta}_n^* - \theta_n| \leq |\theta_n^* - \theta_n|$  a.s. (conditionally on the sample). Under our assumptions it is easy to show that, conditionally on the sample,  $\|I^*\| = o_P(1)$  and  $\|II^*\| = o_P(1)$  with probability one. Therefore, in  $L_2(\Pi, \nu)$

$$\begin{aligned} S_{n,w,m}^*(\eta, \theta_n^*) &= n^{-1/2} \sum_{t=1}^n \widehat{\varepsilon}_{1t}^* Q_{t,w}(\eta) - n^{1/2} (\theta_n^* - \theta_n)' G_{1w}(\eta, \theta_1) + o_P(1) \text{ a.s.} \\ &\equiv \widetilde{S}_{n,w,m}^*(\eta, \theta_n^*) + o_P(1) \text{ a.s.} \end{aligned} \quad (20)$$

Similarly, one can prove that

$$\begin{aligned} S_{n,w,v}^*(\eta, \theta_n^*) &= n^{-1/2} \sum_{t=1}^n \widehat{\varepsilon}_{2t}^* Q_{t,w}(\eta) - n^{1/2} (\theta_n^* - \theta_n)' G_{3w}(\eta, \theta_1) + o_P(1) \text{ a.s.} \\ &\equiv \widetilde{S}_{n,w,v}^*(\eta, \theta_n^*) + o_P(1) \text{ a.s.} \end{aligned} \quad (21)$$



where

$$g'_{3t}(\theta) = 2h(I_{t-1}, \theta) \frac{\partial h(I_{t-1}, \theta)}{\partial \theta'},$$

$b_{3j}(x, \theta) = E[w_{t-j}(x)g_{3t}(\theta)]$  and  $G_{3w}(\eta, \theta) = \sum_{j=1}^{\infty} b_{3j}(x, \theta)\Psi_j(\lambda)$ . It is worth to remark that under the null hypothesis  $G_{3w}(\eta, \theta_0) = G_{2w}(\eta, \theta_0)$ .

Define

$$S_{n,w,m}^*(\eta, \theta_n) = n^{-1/2} \sum_{t=1}^n \widehat{e}_{1t}(\theta_n) V_t Q_{t,w}(\eta) \quad S_{n,w,v}^*(\eta, \theta_n) = n^{-1/2} \sum_{t=1}^n \widehat{e}_{2t}(\theta_n) V_t Q_{t,w}(\eta).$$

Now, using standard bootstrap notation, we denote by  $E^*$  and  $Var^*$  the expectation and the variance, respectively, given the sample. Define  $s_{n,t,m}^*(\eta) = n^{-1/2} \widehat{e}_{1t}(\theta_n) V_t Q_{t,w}(\eta)$ , and note that  $s_{n,t,m}^*(\eta)$  and  $s_{n,s,m}^*(\eta)$  are independent given the sample for  $s \neq t$ . Thus, it is sufficient for the tightness of the first summand in (20) that  $E^* \|s_{n,t,m}^*\|^2 < \infty$  a.s. for all samples, which is trivially satisfied, see example 1.8.5 in van der Vaart and Wellner (1996). This jointly with A6 imply the asymptotic tightness of  $S_{n,w,m}^*(\eta, \theta_n^*)$ . The tightness of  $S_{n,w,v}^*(\eta, \theta_n^*)$  follows from the same arguments.

We now show that the process  $S_{n,w}^*(\lambda, x, \theta_n^*)$  (conditionally on the sample) has the same asymptotic finite projections as the process  $S_{n,w}(\eta)$ . More concretely, we need to prove that (conditional on the sample)  $\langle S_{n,w,m}^*(\cdot, \theta_n^*), h \rangle$ ,  $\langle S_{n,w,v}^*(\cdot, \theta_n^*), h \rangle$  and  $\langle S_{n,w,m}^*(\cdot, \theta_n^*), S_{n,w,v}^*(\cdot, \theta_n^*) \rangle$  converge to the same distribution as  $\langle S_{n,w,m}, h \rangle$ ,  $\langle S_{n,w,v}, h \rangle$  and  $\langle S_{n,w,m}, S_{n,w,v} \rangle$ , respectively.

We shall prove first that (conditional on the sample)  $\langle S_{n,w,m}^*(\cdot, \theta_n^*), h \rangle$  converges to the same distribution as  $\langle S_{n,w,m}, h \rangle$ . The proof with  $\langle S_{n,w,v}^*(\cdot, \theta_n^*), h \rangle$  and  $\langle S_{n,w,v}, h \rangle$  is similar and hence, omitted. To that end, we write

$$\langle S_{n,w,m}^*(\cdot, \theta_n), h \rangle = \sum_{t=2}^n \zeta_{nt}^*,$$

where  $\zeta_{nt}^* = n^{-1/2} \widehat{e}_{1t}(\theta_n) V_t \langle Q_{t,w}(\cdot), h \rangle$ . Then

$$E^* \left( \sum_{t=1}^n \zeta_{nt}^* \right) = \sum_{t=1}^n n^{-1/2} \widehat{e}_{1t}(\theta_n) \langle Q_{t,w}(\cdot), h \rangle E[V_t] = 0,$$

while

$$Var^* \left( \sum_{t=1}^n \zeta_{nt}^* \right) = \sum_{t=1}^n n^{-1} \widehat{e}_{1t}^2(\theta_n) \langle Q_{t,w}(\cdot), h \rangle^2 Var[V_t] := \sigma_{n,h,m}^2.$$

Note that conditional on the original data,  $\zeta_{nt}^*$  is an independent (not identically distributed) array of random variables. By a strong law of large numbers for stationary and ergodic sequences and A2, it is easy to show that  $\sigma_{n,h,m}^2 \rightarrow \sigma_{h,m}^2 = E[e_{1t}^2(\theta_0) \langle Q_{\infty,w}(\cdot), h \rangle^2]$  a.s.

Then, we will verify a Lindeberg-Feller's condition. Note that  $\|Q_{t,w}\|$  and  $V_t$  are bounded, hence

$$\sum_{t=2}^n E^* [|\zeta_{nt}^*|^2 1(|\zeta_{nt}^*| > \delta)] \leq \frac{C}{n} \sum_{t=2}^n \widehat{e}_{1t}^2(\theta_n) 1(|\widehat{e}_{1t}(\theta_n)| > \delta' \sqrt{n}) \text{ a.s.}$$

for some positive constants  $\delta$  and  $\delta'$ . By our assumptions the last expression converges almost surely to zero, see Stute, Gonzalez-Manteiga and Presedo-Quindimil (1998, p. 149) for a similar situation. Then the triangular array  $\{\zeta_{nt}^*\}$  satisfies the conditions of the Lindeberg-Feller's CLT, conditionally on almost all samples, so that  $\sum_{t=1}^n \zeta_{nt}^* \implies_* N(0, \sigma_{h,m}^2)$  a.s. The latter convergence, (20), A6(a), and A3(b) imply that  $\langle S_{n,w,m}^*(\cdot, \theta_n^*), h \rangle$  converges to the same distribution as  $\langle S_{n,w,m}, h \rangle$ .

Now, write  $\langle \tilde{S}_{n,w,m}^*(\cdot, \theta_n^*), \tilde{S}_{n,w,v}^*(\cdot, \theta_n^*) \rangle$  as

$$\begin{aligned} & n^{-1} \sum_{t=1}^n \sum_{s=1}^n \hat{e}_{1t} \hat{e}_{2s} V_t V_s \langle Q_{t,w}(\cdot), Q_{s,w}(\cdot) \rangle - n^{1/2} (\theta_n^* - \theta_n)' n^{-1/2} \sum_{t=1}^n \hat{e}_{1t} V_t \langle Q_{t,w}(\cdot), G_{3w}(\cdot) \rangle \\ & - n^{1/2} (\theta_n^* - \theta_n)' n^{-1/2} \sum_{t=1}^n \hat{e}_{2t} V_t \langle Q_{t,w}(\cdot), G_{1w}(\cdot) \rangle + \langle G_{1w}(\cdot), G_{3w}(\cdot) \rangle n^{1/2} (\theta_n^* - \theta_n)' n^{1/2} (\theta_n^* - \theta_n) \\ & : = A_{1n} - A_{2n} - A_{3n} + A_{4n}. \end{aligned}$$

Following previous arguments it can be shown

$$A_{1n} \implies_* E[e_{1t}(\theta_1) e_{2t}(\theta_1) \langle Q_{\infty,w}(\cdot), Q_{\infty,w}(\cdot) \rangle] \text{ a.s.},$$

$$A_{2n} \implies_* E[m(I_{t-1}, \theta_1) e_t(\theta_1) e_{1t}(\theta_1) \langle Q_{\infty,w}(\cdot), G_{3w}(\cdot) \rangle] \text{ a.s.},$$

$$A_{3n} \implies_* E[m(I_{t-1}, \theta_1) e_t(\theta_1) e_{2t}(\theta_1) \langle Q_{\infty,w}(\cdot), G_{1w}(\cdot) \rangle] \text{ a.s.}$$

and

$$A_{4n} \implies_* E[e_t'(\theta_1) m(I_{t-1}, \theta_1) m(I_{t-1}, \theta_1) e_t(\theta_1) \langle G_{1w}(\cdot), G_{3w}(\cdot) \rangle] \text{ a.s.}$$

Hence,  $\langle \tilde{S}_{n,w,m}^*(\cdot, \theta_n^*), \tilde{S}_{n,w,v}^*(\cdot, \theta_n^*) \rangle$  converges to the same limit as  $\langle S_{n,w,m}(\cdot, \theta_n), S_{n,w,v}(\cdot, \theta_n) \rangle$ . The rest of the proof follows similar arguments and hence, it is omitted.

## Acknowledgements

We would like to thank two referees and Prof. Miguel A. Delgado for very helpful comments and suggestions which have improved a previous version of the paper. 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.

## References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. *Journal of American Statistical Association* 96, 42–55.
- Bai, J., 2003. Testing parametric conditional distributions of dynamic models. *Review of Economic and Statistics* 85, 531-549.
- Bai, J., Ng, S., 2001. A test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225-258.
- Baillie, R. T., Bollerslev, T., Mikkelsen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.
- Blake, A.P., Kapetanios, G., 2006. Testing for ARCH in the presence of nonlinearity of unknown form in the conditional mean. Forthcoming in the *Journal of Econometrics*.
- Bera, A. K., 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., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., Chou, R. Y., Kroner, K. F., 1992. ARCH modelling in Finance. *Journal of Econometrics* 52, 5–59.
- Box, G., Pierce, D., 1970. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of American Statistical Association* 65, 1509-1527.
- Breidt, F. J., Crato, N., de Lima, P., 1998. On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325–348.
- Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.
- Chen, X., Fan, Y., Patton, A., 2003. Simple tests for models of dependence between multiple financial time series: with applications to U.S. equity returns and exchange rates. Unpublished manuscript.
- Ding, Z., Granger, C. W. J., 1996. Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73, 185–215.
- Ding, Z., Granger, C. W. J., Engle, R. F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Delgado, M. A, Dominguez, M. A., Lavergne, P., 2006. Consistent tests of conditional moment restrictions. *Annales d’Economie et Statistique* 81, 33-67.

- Dominguez, M. A., 2004. On the power of bootstrapped specification tests. *Econometric Reviews* 23, 215-228.
- Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987-1008.
- Engle, R., 2002. New frontiers in ARCH models. *Journal of Applied Econometrics* 17, 425-446.
- Engle, R., Lilien, D., Robins, R., 1987. Estimation of time varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391-407.
- Escanciano, J. C., 2006a. Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* 101, 531-541.
- Escanciano, J. C., 2006b. Consistent diagnostic test for regression models using projections. *Econometric Theory* 22, 1030-1051.
- Escanciano, J. C., 2007a. Model checks using residual marked empirical processes. *Statistica Sinica* 17, 115-138.
- Escanciano, J. C., 2007b. On the lack of power of omnibus specification tests. Second revision in *Econometric Theory*.
- Escanciano, J. C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* 134, 151-185.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods* (Springer-Verlag, New York).
- Feigin, P. D., Heathcote, C. R., 1976. The Empirical Characteristic Function and the Cramér-von Mises Statistic. *Sankhya* 38, Series A, 309-325.
- Francq, C., Zakořan, J. M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH. *Bernoulli* 10, 605-637.
- Gallant, A. R., Hsieh, D. A., 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., King, M., 2004. Model specification testing in nonparametric and semiparametric time series econometrics. Working paper.
- Giné, E., Zinn, J. 1990. Bootstrapping general empirical measures. *Annals of Probability* 18, 851-869.
- Glosten, L.R., Jannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1801.
- Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and Its Application* (Academic Press, New

- York).
- Härdle, W., 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.
- Harvey, C., A. Siddique, 1999. Autoregressive conditional skewness. *Journal of Financial y Quantitative Analysis* 34, 465-487.
- Harvey, C., A. Siddique, 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263-1296.
- Hidalgo, J., Zaffaroni, P., 2006. A goodness of fit test for ARCH( $\infty$ ). Forthcoming in *The Journal of Econometrics*
- Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function. *Journal of American Statistical Association* 84, 1201-1220.
- Hong, Y., Lee, T. H., 2003. Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory* 19, 1065-1121.
- 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., M. Rockinger, 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27, 1699-1737.
- Jorion, P. 1997. *Value at Risk, the New Benchmark for Controlling Market Risk*, McGraw-Hill.
- Horváth, L., Kokoszka, P., 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., Ling, S., 2006. Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994-1012.
- Koul, H. L., Stute W., 1999. Nonparametric model checks for time series. *Annals of Statistics* 27, 204-236.

- Lee, S., Hansen, B., 1994. Asymptotic theory for the GARCH(1,1) quasimaximum likelihood estimator. *Econometric Theory* 10, 29–52.
- Li, W. K., Mak, T. K., 1994. On the squared residual autocorrelation in nonlinear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627-636.
- Linton, O., Mammen, E., 2005. Estimating semiparametric ARCH models by kernel smoothing methods. *Econometrica* 73, 771-836
- Liu, R. Y., 1988. Bootstrap procedures under some non-i.i.d models. *Annals of Statistics* 16, 1696-1708.
- Ljung, G. M., Box, G. E. P., 1978. A measure of lack of fit in time series models. *Biometrika* 65, 297-303.
- Lundbergh, S., T. Teräsvirta, 2002. Evaluating GARCH models. *Journal of Econometrics* 110, 417-435.
- Lumsdaine, R. L., 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.
- Mikosch, T., Starica, C., 2003. Long-range dependence effects and ARCH modeling. In *Theory and Applications of Long-Range Dependence*, Boston (eds P. Doukhan, G. Oppenheim and M. S. Taqqu), pp. 439–459. Birkhäuser.
- Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.
- Ngatchou-Wandji, J., 2005. Checking nonlinear heteroscedastic time series models. *Journal of Statistical Planning and Inference* 133, 33-68.
- Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295-325.
- Straumann, D., 2005. *Estimation in Conditionally Heteroscedastic Time Series Models*. Lectures notes in Statistics 181, Springer-Verlag, Berlin-Heidelberg.
- Stute, W., Gonzalez-Manteiga, W., Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* 93, 141-149.
- Wefelmeyer, W., 1996. Quasi-likelihood models and optimal inference. *Annals of Statistics* 24, 405-422.
- Wu, C. F. J., 1986. Jackknife, 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.

$n = 100$	<b>ARCH(1)</b>	<b>ARCH(2)</b>	<b>GARCH(1,1)</b>	<b>EGARCH</b>	<b>SV</b>	<b>BIL</b>	<b>LM</b>	<b>NLMA</b>
$D_{n,I}^2$	4.5	18.0	19.8	82.8	54.2	88.0	99.7	21.6
$D_{n,C}^2$	4.6	52.5	31.2	90.8	67.0	96.7	99.3	77.5
$CvM_{1,v}$	4.8	7.5	13.2	54.6	39.9	87.1	99.7	20.8
$CvM_{3,v}$	5.0	21.6	20.8	36.8	48.7	70.0	18.6	16.2
$CvM_{5,v}$	3.6	18.8	20.8	14.2	53.0	53.9	11.0	8.0
$M_{n,2}$	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.

$n = 50$	<b>AR(1)-CH(1)</b>	<b>AR(1)-BIL</b>	<b>GARCH-M</b>	<b>AR(2)-CH(1)</b>	<b>TAR</b>
$D_{n,I,m}^2$	3.8	4.0	13.3	76.1	27.5
$D_{n,I,v}^2$	6.8	77.7	3.3	4.2	9.6
$J_{n,I}^2$	5.2	68.7	7.0	36.1	18.1
$D_{n,C,m}^2$	3.3	5.9	11.0	34.9	40.2
$D_{n,C,v}^2$	4.8	65.5	3.3	4.9	8.5
$J_{n,C}^2$	4.4	51.7	2.0	20.7	31.0
$CvM_{3,m}$	3.8	3.4	15.3	62.6	6.6
$CvM_{3,v}$	4.6	42.5	3.3	2.8	3.4
$CvM_3$	4.6	26.6	7.0	50.5	4.6
$M_{n,2}$	3.3	42.4	13.3	16.0	21.6
$M_{n,6}$	3.9	35.7	22.3	62.0	17.1
$M_{n,10}$	5.7	27.9	23.0	62.7	14.9

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

$n = 200$	AR(1)-CH(1)	AR(1)-BIL	GARCH-M	AR(2)-CH(1)	TAR
$D_{n,I,m}^2$	5.3	7.5	66.6	100.0	91.5
$D_{n,I,v}^2$	6.3	97.5	2.7	3.5	39.0
$J_{n,I}^2$	5.9	94.0	41.6	100.0	75.0
$D_{n,C,m}^2$	6.9	7.9	51.6	99.5	97.0
$D_{n,C,v}^2$	6.3	92.0	4.0	1.0	37.0
$J_{n,C}^2$	6.6	85.0	30.3	74.0	92.5
$CvM_{3,m}$	5.2	4.2	52.6	99.7	22.6
$CvM_{3,v}$	7.5	98.0	12.3	3.4	5.4
$CvM_3$	6.7	95.2	37.6	99.1	12.8
$M_{n,2}$	4.0	99.0	65.6	87.0	80.0
$M_{n,6}$	3.3	99.5	80.6	100.0	69.0
$M_{n,10}$	4.6	95.5	84.6	100.0	57.5

**Table 4.** Empirical Size and Power of Tests at 5% for EXPAR-ARCH(1).

<b>n</b>	<b>50</b>						<b>100</b>					
	<b>0.0</b>			<b>-2.0</b>			<b>0.0</b>			<b>-2.0</b>		
<b><math>\theta_{01}</math></b>	<b>0.0</b>	<b>0.4</b>	<b>0.8</b>	<b>0.0</b>	<b>0.4</b>	<b>0.8</b>	<b>0.0</b>	<b>0.4</b>	<b>0.8</b>	<b>0.0</b>	<b>0.4</b>	<b>0.8</b>
$D_{n,I,m}^2$	2.0	3.0	5.9	32.6	47.2	53.2	4.0	4.0	7.0	67.2	86.0	79.8
$D_{n,I,v}^2$	5.2	36.0	42.8	2.6	13.0	23.0	5.9	48.6	55.8	5.0	23.8	50.4
$J_{n,I}^2$	4.0	27.8	36.6	13.2	34.8	52.4	5.7	37.6	48.8	41.4	81.0	84.6
$D_{n,C,m}^2$	1.4	3.0	6.8	<b>56.2</b>	47.2	53.6	3.8	4.2	7.6	<b>82.4</b>	79.8	77.6
$D_{n,C,v}^2$	4.8	<b>55.0</b>	<b>76.0</b>	2.4	70.0	85.2	5.0	<b>83.2</b>	<b>97.2</b>	4.8	98.2	<b>100</b>
$J_{n,C}^2$	2.6	47.2	72.8	36.6	<b>75.8</b>	<b>89.8</b>	4.0	78.8	93.4	68.2	<b>99.0</b>	<b>100</b>
$CvM_{3,m}$	2.8	3.0	6.4	9.6	10.6	12.6	4.2	3.2	7.8	24.6	32.2	27.2
$CvM_{3,v}$	5.2	31.8	33.0	3.0	10.4	20.4	5.6	39.8	40.0	1.8	40.4	51.8
$CvM_3$	5.0	24.8	26.8	6.2	10.4	19.2	5.4	29.4	33.2	16.2	30.8	42.8
$M_{n,2}$	2.6	34.8	71.2	8.2	58.2	82.0	4.2	60.2	91.4	14.8	84.8	99.0
$M_{n,6}$	3.2	36.4	70.2	8.6	55.6	80.2	3.4	60.6	92.6	14.0	87.4	98.8
$M_{n,10}$	4.6	35.0	65.4	8.4	50.4	74.0	3.6	59.6	90.4	12.6	86.2	98.0