Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 01/07

Data-Driven Smooth Tests for the Martingale Difference Hypothesis

Juan Carlos Escanciano
Department of Economics
Indiana University

Silvia Mayoral
Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Data-Driven Smooth Tests for the Martingale Difference Hypothesis
Juan Carlos Escanciano and Silvia Mayoral
Working Paper No.01/07
January 2007

ABSTRACT

A general method for testing the martingale difference hypothesis is proposed. The new tests are data-driven smooth tests based on the principal components of certain marked empirical processes that are asymptotically distribution-free, with critical values that are already tabulated. The data-driven smooth tests are optimal in a semiparametric sense discussed in the paper, and they are robust to conditional heteroskedasticity of unknown form. A simulation study shows that the smooth tests perform very well for a wide range of realistic alternatives and have more power than the omnibus and other competing tests. Finally, an application to the S&P 500 stock index and some of its components highlights the merits of our approach.

Juan Carlos Escanciano
Department of Economics, Indiana University,
100 S. Woodlawn, Wylie Hall, Bloomington,
IN 47405-7104, USA.
jescanci@indiana.edu

Silvia Mayoral
Universidad de Navarra
Depto. Métodos Cuantitativos
Campus Universitario
31080 Pamplona
smayoral@unav.es

Data-Driven Smooth Tests for the Martingale Difference Hypothesis[1]

Juan Carlos Escanciano[2] and Silvia Mayoral

January 25, 2007.

Abstract: A general method for testing the martingale difference hypothesis is proposed. The new tests are data-driven smooth tests based on the principal components of certain marked empirical processes that are asymptotically distribution-free, with critical values that are already tabulated. The data-driven smooth tests are optimal in a semiparametric sense discussed in the paper, and they are robust to conditional heteroskedasticity of unknown form. A simulation study shows that the smooth tests perform very well for a wide range of realistic alternatives and have more power than the omnibus and other competing tests. Finally, an application to the S&P 500 stock index and some of its components highlights the merits of our approach.

Keywords and Phrases: Nonlinear time series; Empirical processes; Pivotal tests; Neyman's tests; Semiparametric efficiency; Market efficiency; Testing for no effect.

[1] J. Carlos Escanciano is Assistant Professor at Indiana University, Department of Economics, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA, e-mail: jescanci@indiana.edu. Research funded by the Spanish Ministerio de Educación y Ciencia, reference numbers SEJ2004-04583/ECON and SEJ2005-07657/ECON. Silvia Mayoral is Assistant Professor at Universidad de Navarra, Facultad de Económicas, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain, e-mail: smayoral@unav.es. Research funded by *Comunidad Autónoma de Madrid* Grant s-0505/tic/000230, and the Spanish Ministerio de Educación y Ciencia, reference number BEC2000-1388-C04-03.

[2] Corresponding address: Department of Economics, Indiana University, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA.

1 Introduction

Testing for the martingale difference hypothesis (MDH) of a linear or nonlinear time series is central in many areas such as statistics, economics and finance. In particular, many economic theories in a dynamic context, including the market efficiency hypothesis, rational expectations or optimal asset pricing, lead to such dependence restrictions on the underlying economic variables, see e.g. Cochrane (2001). Moreover, testing for the MDH seems to be the first natural step in modeling the conditional mean of a time series and it has important consequences in modeling higher order conditional moments. This article proposes data-driven smooth tests for the MDH based on the principal components of certain marked empirical processes having the following attributes: (i) they are asymptotically distribution-free, with critical values from a χ^2-distribution, (ii) they are robust to second and higher order conditional moments of unknown form, in particular, to conditional heteroscedasticity (iii) in contrast to omnibus tests, smooth tests possess good local power properties and are optimal in a semiparametric sense to be discussed below, and (iv) they are very simple to compute, without resorting to nonparametric smoothing estimation.

More precisely, let $\{Y_t\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$. The MDH states that the best predictor, in a mean square error sense, of Y_t given $\mathbf{I}_{t-1} := (Y_{t-1}, Y_{t-2}, ...)'$ is just the unconditional expectation, which is zero for a martingale difference sequence (mds). In other words, the MDH states that $Y_t = X_t - X_{t-1}$, where X_t is a martingale process with respect to the σ-field generated by $\mathbf{I}_{t-1}$, i.e., $\mathcal{F}_{t-1} := \sigma(\mathbf{I}_{t-1})$.

The classical procedure for testing the MDH in statistical applications is to assume that the data generating process (DGP) belongs to a parametric family, and proceeds with a standard parametric test such as the $t-$test. For instance, in financial econometrics, it is common to assume that the DGP follows a linear autoregressive model of order one with generalized conditionally heteroscedastic errors of order (1,1) (in short AR(1)-GARCH(1,1) model), where

$$Y_t = cY_{t-1} + \varepsilon_t, \tag{1}$$

$|c| < 1$, $\varepsilon_t = \sigma(\mathbf{I}_{t-1},\theta_0)u_t$, $\{u_t\}$ is a sequence of independent and identically distributed (iid) disturbances, independent of $\mathbf{I}_{t-1}$, and the conditional variance is given by

$$\sigma^2(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0) \equiv \sigma_t^2 = \theta_{01} + \theta_{02}\varepsilon_{t-1}^2 + \theta_{03}\sigma_{t-1}^2.$$

$\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \theta_{03})' \in \Theta \subset \mathbb{R}^3$, with $\Theta = \{(\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3 : \theta_1 > 0,\ \theta_j \geq 0,\ j = 2,\ 3,$ and $\theta_2 + \theta_3 < 1\}$. Then one proceeds to test within the model (1) for

$$\widetilde{H}_0 : c = 0 \text{ against } \widetilde{H}_1 : c \neq 0.$$

To that end, standard $t-$tests are commonly used. However, parametric tests such as the $t-$test are in general not robust to misspecifications in the parametric conditional variance. Moreover, although robust versions are available in the literature, see e.g. Deo (2000), tests based on correlations, like the $t-$tests, are only able to detect very few alternatives. In particular, these classical tests fail to detect many nonlinear alternatives, which are likely to occur in financial applications, see Hsieh (1989), Gallant, Hsieh and Tauchen (1991) and Escanciano and Velasco (2006), among others. See also Section 5 for some evidence of this "lack of power" with stocks returns.

Nonparametric tests for the MDH vary from classical tests based on correlations or periodograms, such as Box and Pierce (1970) or Durlauf (1991), to the more sophisticated tests based on the generalized spectral approach, e.g. Escanciano and Velasco (2006), and empirical processes theory in Domínguez and Lobato (2003). Tests based on the generalized spectral approach and empirical processes theory are more powerful than correlation-based tests for nonlinear alternatives, but they usually involve bootstrap approximations, hampering their use in statistical applications. In this paper we consider simple and powerful tests, which are especially suited for practitioners since they are valid under fairly weak regularity conditions on the DGP and do not need of resampling methods.

Our null hypothesis is that Y_t is a mds, i.e.

$$H_0 : E[Y_t \mid \mathbf{I}_{t-1}] = 0 \text{ almost surely (a.s.)}$$

The alternative H_1 is the negation of the null, i.e., Y_t is not a mds.

The rationale for our tests follows from the asymptotic properties of a marked empirical process (cf. Koul and Stute, 1999), which for a sample $\{Y_t\}_{t=0}^n$ of size $n+1$ is given by

$$R_n(x) := \frac{1}{\widehat{\sigma}\sqrt{n}} \sum_{t=1}^{n} Y_t 1(Y_{t-1} \leq x), \qquad x \in \mathbb{R}, \tag{2}$$

where $\widehat{\sigma}^2 = n^{-1} \sum_{t=1}^n Y_t^2$.

Under the null H_0, the process R_n is centered, but under the alternative H_1, it is expected to be not centered anymore, allowing us to base the tests on suitable functionals of R_n. More concretely, under H_0 a suitable standardization of the limit process of R_n is a standard Brownian motion in proper time (cf. Theorem 1), so suitable functionals of the limit will be distribution-free. When the functionals are appropriate norms, the resulting tests are omnibus. Although considering an omnibus test is naturally the first idea when there is no a priori preference of directions in the alternative hypothesis, it is worth noting that there is an important limitation of omnibus tests: despite the capability of an omnibus test to detect the deviations in all the directions, it is well-known that they have reasonable nontrivial local power against very few orthogonal directions, see

Janssen (2000) and Escanciano (2005) for theoretical explanations.

In this paper we construct data-driven smooth tests based on the principal components of R_n that overcome the "lack of local power" of the omnibus tests. Omnibus tests down weight the contribution of the principal components whereas our new smooth tests give the same weight to the number of components used, which is the optimal weighting scheme. The number of components is chosen following a data-driven selection rule that combines the two most popular selection rules, Akaike and Schwarz selection criteria. As we shall show, these data-driven smooth tests are more powerful than omnibus tests for a wide class of realistic alternatives and they are optimal in a semiparametric sense discussed below. They are able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$. Moreover, they are robust to second and higher order time-varying conditional moments of unknown form and, unlike the omnibus tests, they provide information on the possible alternative in case of rejection. All these appealing properties make of the new smooth tests an attractive testing procedure for the MDH.

The remainder of this paper is organized as follows. Section 2 discusses asymptotically distribution-free omnibus tests for H_0 based on R_n. In Section 3 we develop data-driven smooth tests from the omnibus tests by means of the principal components decomposition of R_n. Section 4 considers some Monte Carlo experiments to study the finite sample performance of the proposed tests. In Section 5 we apply our methodology to the S&P 500 stock index and some of its components. Section 6 discusses extensions of the basic framework and concludes. Mathematical proofs are gathered in the Appendix.

2 Omnibus tests

This section deals with omnibus tests for H_0 based on continuous functionals of R_n. Let F be the cumulative distribution function (cdf) of Y_t. The symbol $\Longrightarrow$ denotes weak convergence in the metric space $D([-\infty,\infty])$ of the *cadlag* (right-continuous with left limits) functions on $[-\infty,\infty]$, endowed with the Skorohod metric, see Billingsley (1999). Notice that R_n belongs to $D([-\infty,\infty])$ after defining $R_n(-\infty) := 0$ and $R_n(+\infty) := n^{-1/2}\sum_{t=1}^{n} Y_t$. The following regularity condition is necessary for the subsequent asymptotic analysis.

A1*:* (a) $\{Y_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process with $0 < E[Y_t^2] < \infty$; (b) F is an absolutely continuous cdf; (c) $E[Y_t^4 \left|Y_{t-1}\right|^{1+\delta}] < \infty$, for some $\delta > 0$. Also, the conditional density of Y_t given $\mathbf{I}_{t-1}$ is bounded and continuous.

Assumption A1 is a condition on the DGP and it is sufficient for the weak convergence of R_n in $D([-\infty,\infty])$, see Koul and Stute (1999) for similar assumptions. A1 is rather weak and permits a large class of nonlinear time series, including heteroskedastic ones. Let B be a standard Brownian

motion on $[0,1]$, and define

$$\tau^2(x) := \sigma^{-2} E[Y_t^2 1(Y_{t-1} \leq x)],$$

with $0 < \sigma^2 := E[Y_t^2] < \infty$. Notice that $\tau^2(-\infty) = 0$, $\tau^2(+\infty) = 1$, and $\tau^2(\cdot)$ is non-decreasing and continuous if F is continuous. Next theorem establishes the weak convergence of R_n.

THEOREM 1: *Under A1 and H_0,*

$$R_n \Longrightarrow B(\tau^2(\cdot)).$$

An immediate consequence of Theorem 1 is that, under A1 and H_0, and using the scaling properties of the Brownian motion,

$$KS_n := \sup_{x\in\mathbb{R}} |R_n(x)| \xrightarrow{d} \sup_{x\in\mathbb{R}} \left|B(\tau^2(x))\right| = \sup_{t\in[0,1]} |B(t)|,$$

(where the equality is in distribution.) And similarly, from Theorem 1 and Lemma 3.1 in Chang (1990),

$$CvM_n := \int_{\mathbb{R}} |R_n(x)|^2 \tau_n^2(dx) \xrightarrow{d} CvM_\infty := \int_{\mathbb{R}} \left|B(\tau^2(x))\right|^2 \tau^2(dx) = \int_{[0,1]} |B(u)|^2 du,$$

where $\tau_n^2(x) := \widehat{\sigma}^{-2} n^{-1} \sum_{t=1}^n Y_t^2 1(Y_{t-1} \leq x)$.

Norms of R_n, such as KS_n or CvM_n, constitute omnibus tests for H_0 with power against a large class of alternatives in H_1, see Section 6 for a characterization of such alternatives. A similar test to KS_n has been considered in Koul and Stute (1999), see also Domínguez and Lobato (2003). The test based on CvM_n is a variation of the standard Cramér-von Mises (CvM) test, which usually uses the empirical cdf of $\{Y_t\}_{t=1}^n$ replacing τ_n^2. The use of τ_n^2 is motivated from the pivotal property of the limit distribution of CvM_n.

3 From omnibus tests to data-driven smooth tests

There has been some recent theoretical evidence that omnibus tests, such as those based on KS_n and CvM_n, have local power against very few orthogonal directions in the alternative hypothesis, see Escanciano (2005). This theoretical finding is supported by our empirical results in the Monte Carlo experiments and the application in Section 5. The purpose of this paper is to introduce a new class of test for the MDH solving this deficiency. In this section we develop data-driven smooth tests as a solution to the lack of local power of the CvM tests. Our construction relies on the principal component decomposition of the CvM test based on CvM_n as in Durbin and Knott (1972).

Define

$$\lambda_j := \frac{1}{(j-1/2)^2\pi^2} \qquad \psi_j(t) := \sqrt{2}\sin((j-1/2)\pi t), \qquad t \in [0,1],\ j = 1, 2....$$

Notice that $\{\psi_j(\cdot)\}_{j=1}^{\infty}$ constitutes an orthonormal basis of $L_2[0,1]$, the Hilbert space of all square-integrable functions (with respect to Lebesgue measure.) Let $L_2(\mathbb{R}, \tau^2)$ be the Hilbert space of all τ^2-square-integrable functions on $\mathbb{R}$, endowed with the inner product

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)\tau^2(dx)$$

Hence, the basis defined by

$$\varphi_j(x) := \psi_j(\tau^2(x)), \qquad x \in \mathbb{R},\ j = 1, 2...,$$

is an orthonormal complete basis of $L_2(\mathbb{R}, \tau^2)$, since

$$\left\langle \varphi_j, \varphi_h \right\rangle = \int_{\mathbb{R}} \psi_j(\tau^2(x))\psi_h(\tau^2(x))\tau^2(dx) = \int_{[0,1]} \psi_j(u)\psi_h(u)du = \begin{cases} = 1 & j = h \\ = 0 & j \neq h. \end{cases}$$

Moreover, $\{\varphi_j\}_{j=1}^{\infty}$ are the eigenfunctions of the covariance operator of $B(\tau^2(\cdot))$ with associated eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$, i.e.,

$$\int_{\mathbb{R}} E[B(\tau^2(x))B(\tau^2(y))]\varphi_j(x)\tau^2(dx) = \lambda_j\varphi_j(y) \qquad \text{for all } j = 1, 2...$$

Hence, both $R_n(x)$ and $B(\tau^2(x))$ can be expanded using the basis $\{\varphi_j\}_{j=1}^{\infty}$, to obtain the so-called Karhunen-Loève representations (in distribution), see e.g. Bosq (2000),

$$R_n(\cdot) = \sum_{j=1}^{\infty} \lambda_j^{1/2}\epsilon_{n,j}\varphi_j(\cdot)$$

and

$$B(\tau^2(\cdot)) = \sum_{i=1}^{\infty} \lambda_j^{1/2}\epsilon_j\varphi_j(\cdot),$$

where $\epsilon_j := \lambda_j^{-1/2}\left\langle B(\tau^2(\cdot)), \varphi_j\right\rangle$ and $\epsilon_{n,j} := \lambda_j^{-1/2}\left\langle R_n, \varphi_j\right\rangle$ are, respectively, the principal components and sample principal components of $B(\tau^2(\cdot))$ and $R_n(\cdot)$. Two important properties are worth to be mentioned:

(i) From Theorem 1 and the fact that $\{\psi_j, \lambda_j\}_{j=1}^{\infty}$ are the eigenelements of the covariance operator of the standard Brownian motion, it follows that $\{\epsilon_j\}_{j=1}^{\infty}$ are iid $N(0,1)$ random variables

(r.v's) and $\{\epsilon_{n,j}\}_{j=1}^{\infty}$ are, at least, uncorrelated with unit variance. To prove that, write

$$\begin{aligned} E[\epsilon_j \epsilon_h] &= \lambda_j^{-1/2}\lambda_h^{-1/2} \int_{\mathbb{R}} \int_{\mathbb{R}} E[B(\tau^2(x))B(\tau^2(y))]\varphi_j(x)\varphi_h(y)\tau^2(dx)\tau^2(dy) \\ &= \lambda_j^{-1/2}\lambda_h^{-1/2} \int_{[0,1]} \int_{[0,1]} E[B(u))B(v)]\psi_j(u)\psi_h(v)dudv \\ &= \lambda_j^{1/2}\lambda_h^{-1/2} \int_{[0,1]} \psi_j(v)\psi_h(v)dv. \end{aligned}$$

(ii) Second, Parseval's identity yields

$$CvM_{\infty} = \sum_{j=1}^{\infty} \lambda_j \epsilon_j^2. \tag{3}$$

Therefore, from (i) and (ii) it follows that the asymptotic null distribution of CvM_n can be expressed as a weighted sum of independent χ_1^2 r.v's with weights λ_j. From (3) it can be immediately seen that alternatives for which the first components are significatively zero (i.e. those where $\epsilon_j^2 \approx 0$ for $j = 1, ..., m$, for a moderate m) are heavily down weighted by λ_j. These alternatives are called *high-frequency* alternatives and they are difficult to be detected by CvM_n. In other words, the CvM test based on CvM_n, although being omnibus, is only able to detect "in practice" (i.e. in terms of local power) those alternatives where the first components are significatively different from zero (i.e. low-frequency alternatives). See Janssen (2000) for further theoretical support on this "lack" of power for general functionals in the context of goodness-of-fit tests of distributions and Escanciano (2005) for theoretical evidence in model checks.

A possible solution to overcome the previous problem is offered by the so-called smooth tests. They were first proposed by Neyman (1937) in the context of goodness-of-fit of distributions, and since then, there have been a plethora of researches documenting their theoretical and empirical properties. Many authors, including Eubank and LaRiccia (1992), Ledwina (1994), Fan (1996), Inglot and Ledwina (1996) and Kallenberg and Ledwina (1997), among others, have shown theoretical and empirical evidence that smooth tests outperform omnibus test over a wide range of realistic alternatives, see, e.g., Eubank and LaRiccia (1992, p. 2072). All these proposals are devoted to goodness-of-fit tests of distribution functions. See Rayner and Best (1989) for a monograph on smooth tests in the latter framework and Koziol (1980) for the problem of testing for symmetry.

There have been some contributions of the smooth approach in regression problems. Fan and Huang (2001) consider data-driven Neyman's tests using Fourier transforms for linear models with iid observations and Gaussian errors, extending previous work by Fan (1996) to regressions. Aerts, Claeskens and Hart (1999) considered a general methodology for parametric models for iid data,

extending previous work by Eubank and Hart (1992). Eubank (2000) has compared, theoretically and by simulations, the test proposed in Eubank and Hart (1992) and a data-driven smooth test using the Schwarz criterion, as in Ledwina (1994), for the problem of testing for no effect, which is the iid version of the problem considered in the present paper. To the best of our knowledge, our tests provide the first (data-driven) smooth tests in a semiparametric time series framework under general serial dependence. Following the results in Eubank (2000) and in Inglot and Ledwina (2006a), we propose a smooth test coupled with a data-driven choice for the number of principal components, which combines the advantages of the Schwarz and Akaike criteria.

To avoid the down weighting due to the $\lambda_j^{'}$s we construct the test statistic

$$T_{n,m} = \sum_{j=1}^{m} \widehat{\epsilon}_{j,n}^2, \tag{4}$$

where

$$\begin{aligned} \widehat{\epsilon}_{j,n} &= \lambda_j^{-1/2} \int_{\mathbb{R}} \psi_j(\tau_n^2(x)) R_n(x) \tau_n^2(dx) \\ &= \lambda_j^{-1/2} \frac{\sqrt{2}}{\widehat{\sigma}^2 n} \sum_{s=1}^{n} \sin((j-1/2)\widehat{\sigma}^{-2}\pi\tau_n^2(Y_{s-1})) Y_s^2 R_n(Y_{s-1}), \end{aligned}$$

estimates ϵ_j. Under the null H_0, the asymptotic distribution of $T_{n,m}$ for a fixed m is a χ_m^2-distribution, see Theorem 2. For each fixed $m \in \mathbb{N}$, the test based on rejecting H_0 when $\rho_{n,m,\alpha} := 1(T_{n,m} > \chi_{m,\alpha}^2)$ takes the value one, where $\chi_{m,\alpha}^2$ is the $(1-\alpha)-$quantile of the $chi-$square distribution with m degrees of freedom, is called a *smooth test*.

Examples in the literature of goodness-of-fit tests for distributions show that a considerable loss of power may occur when a wrong choice of m is made, see e.g. Kallenberg and Ledwina (1997) and Section 5 below. This illustrates that a good procedure for choosing m based on the data is very welcome. Here, we adopt the combination rule of the Schwarz's and Akaike's selection rules of Inglot and Ledwina (2006a) for the choice of m, i.e., we define

$$\widetilde{m} = \min\{m : 1 \le m \le d; L_m \ge L_h, h = 1, 2..., d\}, \tag{5}$$

where

$$L_m = T_{n,m} - \pi(m, n, q), \tag{6}$$

and d is an upper bound that can be arbitrarily large but fixed, and

$$\pi(j, n, c) = \begin{cases} j \log n, & \text{if} \quad \max_{1 \le j \le d} |\widehat{\epsilon}_{j,n}| \le \sqrt{q \log n} \\ 2j, & \text{if} \quad \max_{1 \le j \le d} |\widehat{\epsilon}_{j,n}| > \sqrt{q \log n}, \end{cases}$$

where q is some fixed positive number. Our choice of q is 2.4 and is motivated from an extensive simulation study in Inglot and Ledwina (2006b) and from simulations in the present paper. Small values of q result in the Akaike's criterion choice, while large $q's$ lead to the choice of the Schwarz's criterion. Moderate values, such as 2.4, provide a "switching effect" in which one combines the advantages of the two selection rules, that is, when the alternative is of high frequency Akaike is used whereas if the alternative is low-frequency Schwarz is chosen.

Our final test is the data-driven smooth test

$$T_{n,\widetilde{m}} = \sum_{j=1}^{\widetilde{m}} \widehat{\epsilon}_{j,n}^2.$$

Other penalization terms different from the one used here are also valid under mild conditions on the penalization, as shown by Kallemberg (2002). We shall show in Section 4 and Section 5 that our combination rule works quite well for moderate sample sizes as those encountered in financial applications. See Section 6 for further motivation and variations of the selection rule. Next theorem establishes the asymptotic null distribution of smooth tests.

THEOREM 2: *Under A1 and* H_0*,*

(i)
$$T_{n,m} \xrightarrow{d} \chi_m^2.$$

(ii) *Furthermore,*
$$T_{n,\widetilde{m}} \xrightarrow{d} \chi_1^2.$$

As with other smooth tests, our test can be interpreted as an optimal test for a "general parametric model" that nests the null hypothesis as a particular case. This is the original and fundamental idea of Neyman (1937). Unlike in other smooth tests, in our case the general model is semiparametric and involves infinite-dimensional nuisance parameters in a time series framework. Under this interpretation, the optimality of the smooth test can be formally formulated using the theory of semiparametric efficient tests in Choi, Hall, and Schick (1996). This theory parallels the efficient estimation theory of semiparametric and nonparametric models as discussed in the excellent monograph by Bickel, Klaassen, Ritov and Wellner (1993).

To that end, define the conditional variance $\sigma^2(x) := E\left[Y_t^2 \mid Y_{t-1} = x\right]$ and the functions

$$h_j(x) = \sqrt{2}\lambda_j^{-1/2}\sigma^2(x)\cos\left((j-1/2)\sigma^{-2}\pi\tau^2(x)\right), \qquad j = 1, 2...,$$

and consider the semiparametric models

$$\begin{aligned} Y_t &= c_1 h_1(Y_{t-1}) + \cdots + c_m h_m(Y_{t-1}) + \varepsilon_t, \\ &= \mathbf{c}'\mathbf{h}(Y_{t-1}) + \varepsilon_t, \end{aligned} \tag{7}$$

where $\varepsilon_t = Y_t - E[Y_t \mid \mathbf{I}_{t-1}]$, $\mathbf{c} = (c_1, ..., c_m)'$ and $\mathbf{h}(Y_{t-1}) = (h_1(Y_{t-1}), ..., h_m(Y_{t-1}))'$.

Model (7) is semiparametric in the sense that we do not know neither the distribution of the lagged variable Y_{t-1} nor the (conditional) distribution of errors ε_t given Y_{t-1}, which are infinite-dimensional nuisance parameters. Choi, Hall, and Schick (1996) have introduced the concept of an asymptotically uniformly most powerful invariant and unbiased (AUMPIU) test in a semiparametric framework where the parameter of interest is finite-dimensional, as is our case with $\mathbf{c} \in \mathbb{R}^m$. The next theorem proves the asymptotic efficiency of smooth tests for the case of Markov processes of order one. Extensions to higher order Markov processes are trivial, and hence, omitted. The Markov property is not necessary but facilitates the application of the existing semiparametric efficiency theory (cf. Wefelmeyer, 1997.) It is expected that the optimality result can be extended to non-Markovian processes as long as a local asymptotic normality property of the nonparametric model is guaranteed.

A2*:* $\{Y_t\}_{t\in\mathbb{Z}}$ is a Markov process of order one.

THEOREM 3: *Under A1 and A2, the smooth test* $\rho_{n,m,\alpha}$ *based on rejecting when* $T_{n,m}$ *is large is an AUMPIU test for testing* $H_{s0} : \mathbf{c} = \mathbf{0}$ *against* $H_{s1} : \mathbf{c} \neq \mathbf{0}$ *in model (7), in the sense discussed in Choi, Hall, and Schick (1996).*

The optimality result of our smooth tests in Theorem 3 complements other optimality properties that have been obtained in the context of goodness-of-fit tests of distributions, see, for instance, the intermediate efficiency concept in Inglot and Ledwina (1996). In principle, these alternative concepts might be extended to our semiparametric framework considered here. See Eubank (2000) for such an extensions for iid data and fixed design.

An attractive feature of our optimal smooth tests is that when H_{s0} is rejected, $\widehat{E}[Y_t \mid Y_{t-1}] = \widehat{\mathbf{c}}'\widehat{\mathbf{h}}(Y_{t-1})$ provides an alternative model for the conditional mean $E[Y_t \mid Y_{t-1}]$, where $\widehat{\mathbf{c}}$ is the least squares estimator in (7) and $\widehat{\mathbf{h}}(Y_{t-1})$ replaces $\sigma^2(x)$ and $\tau^2(x)$ in $\mathbf{h}(Y_{t-1})$ by nonparametric estimators $\sigma_n^2(x)$ and $\tau_n^2(x)$, respectively. The estimator $\widehat{E}[Y_t \mid Y_{t-1}]$ can be interpreted as a series expansion estimator. In this sense, smooth tests are more informative than omnibus tests when the null hypothesis is rejected. See our application in Section 5 for an example of estimate of alternative models for the conditional mean of some stock returns.

4 A simulation study

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. We compare our data-driven smooth test with the standard $t-$test, the omnibus tests proposed in Section 2, KS_n and CvM_n, and the smooth test with a fixed number of components $T_{n,m}$. The number of Monte Carlo experiments in all simulations is 1000. We report empirical size and power at 5% nominal level, results with other nominal levels are similar and hence, omitted. The bound for the number of components in $T_{n,\widetilde{m}}$ is chosen to be $d = 10$ in all simulations. Unreported simulations here and simulations in related literature, see e.g. Kallenberg and Ledwina (1997), show that the choice of d is not as important as the choice of m. Selection rules, such as the one considered here, are stable as a function of d. In these models the innovations $\{u_t\}$ are iid distributed as $N(0,1)$.

The models used in the simulations include the following:

1. An AR(1)-GARCH(1,1) model as in (1), where

$$Y_t = cY_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \sigma(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0)u_t,$$

$$\sigma^2(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0) \equiv \sigma_t^2 = \theta_{01} + \theta_{02}\varepsilon_{t-1}^2 + \theta_{03}\sigma_{t-1}^2,$$

   with $(\theta_{01}, \theta_{02}, \theta_{03}) = (0.025, 0.25, 0.5)$.

2. An AR(1)-EGARCH(1,1) model, where

$$Y_t = cY_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \sigma(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0)u_t,$$

$$\sigma^2(\mathbf{I}_{t-1}, \boldsymbol{\theta}_0) \equiv ln(\sigma_t^2) = \theta_{01} + \theta_{02}(|\varepsilon_{t-1}| - E|\varepsilon_{t-1}| + \theta_{03}\varepsilon_{t-1}) + \theta_{04}ln(\sigma_{t-1}^2),$$

   with $(\theta_{01}, \theta_{02}, \theta_{03}, \theta_{04}) = (0.2, 0.1, 0.98, 0.01)$.

3. A nonlinear autoregressive process

$$Y_t = ch_j(Y_{t-1}) + u_t,$$

$$h_j(x) = \cos\left((j - 1/2)\pi x\right),$$

   with $j = 2$, 3.

4. Non-linear Moving Average (NLMA) model :

$$Y_t = u_{t-1}u_{t-2}(u_{t-2} + u_t + 1).$$

In models (1-3) the null hypothesis corresponds to $c = 0$. We have considered c from -0.9 to 0.9 in intervals of length 0.1 to study power performance for models (1-2). The AR(1)-GARCH(1,1) and AR(1)-EGARCH(1,1) models are both standard and the most used models in financial applications. In Figure 1 we report the empirical size and power for the model AR(1)-GARCH(1,1) and AR(1)-EGARCH(1,1) at 5% level of the test statistics KS_n, CvM_n, $t-$test, $T_{n,3}$ and $T_{n,\widetilde{m}}$. The sample sizes for models (1-2) are $n = 500$ and $n = 1000$.

FIGURE 1 ABOUT HERE

It can be seen from Figure 1 that all tests present a good size performance. The data-driven smooth test has empirical sizes of 0.064 (GARCH, $n = 500$), 0.048 (GARCH, $n = 1000$), 0.060 (EGARCH, $n = 500$) and 0.053 (EGARCH, $n = 1000$). The empirical size with Gaussian innovations is satisfactory for all test statistics and both models. Unreported simulations with $t-$distributed innovations with 5 degrees of freedom show some overrejection of the data-driven test (e.g. for GARCH 0.112 with $n = 500$ and 0.079 with $n = 1000$). This overrejection for non-Gaussian innovations is not specific of our data-driven test and appears in other data-driven smooth tests proposed in the literature, see e.g. Ledwina (1994). Kallenberg and Ledwina (1997) recommend to use an improved approximation to the asymptotic critical value. Their idea, however, cannot be directly applied to our present framework. As expected, the $t-$test presents the best power against these linear alternatives, followed by the data-driven and fixed smooth tests. Both smooth tests perform similarly in terms of empirical power. The omnibus tests have low power in comparison with the other competing tests, and in agreement with the theoretical results shown in Escanciano (2005).

Table 1 reports the empirical power and size of tests for model 3 for some values of c and the sample size $n = 500$.

TABLE 1 ABOUT HERE

As can be seen from Table 1, model 3 with $j = 3$ is an example of a high-frequency alternative. It is shown that the unique test detecting these alternatives is the data-driven smooth test $T_{n,\widetilde{m}}$. This example illustrates that a wrong choice in the number of components may lead to a considerable loss of power, see the results for $T_{n,3}$. For $j = 2$ the model yields an alternative of intermediate frequency. The smooth test with a fixed number of components detects this alternative but with

much less power than the data-driven smooth test. Omnibus tests as well as the $t-$test have very low power against this alternative.

In Table 2 we report the empirical power of the tests for the NLMA model and sample sizes $n = 300$ and $n = 500$. We can observe that the omnibus tests and the $t-$test have no power against this alternative. The NLMA model generates uncorrelated data, so this alternative is not detected by tests based on correlations, and in particular by the $t-$test. Among the tests considered, only smooth tests are able to detect this alternative, with $T_{n,\widetilde{m}}$ being the best.

TABLE 2 ABOUT HERE

The conclusions from these simulations and other simulations in Section 6 are the following. The data-driven smooth test has a reasonable size performance for moderate sample sizes and presents excellent power properties against the alternatives considered, being in many cases the unique consistent test. These properties make of the data-driven test a convenient test procedure for financial applications where the sample size is usually moderate or large, meaning $n \geq 500$, where we recommend to use the asymptotic critical values. Simulated critical values are, of course, also possible and are recommended for small sample sizes or for fat-tailed distributions. Smooth tests with a fixed number of components maintain an excellent size performance even for very small sample sizes and have reasonable power for a large class of alternatives. In particular, they have more power than omnibus tests. Only high-frequency alternatives are hardly detected by smooth tests with a fixed number of components.

5 An application to economic time series

In this section, we present applications of our tests to some daily closed stock prices. We consider the S&P 500 stock index and ten of its components: Ameriprice Financial (AF), Bank of America CP. (BA), Citigroup Inc. (Cit), Eaton CP. (Ea), Ecolab Inc. (Ec), Exxon Mobil CP. (Ex), General Electric Co. (GE), General Motors. (GM), Pepsi Bottling Grp. (Pep) and Starbucks CP. (Sta). Some of these stocks are within the five most important components of the S&P 500 and different sectors are considered, such as Financial, Services, Industrial Goods, Consumer Goods, Healthcare and Basic Materials sectors. We study the period from 2th January 2003 until 30th December 2005, with a total of 755 observations. The prices are obtained from www.finance.yahoo.com.

We consider the returns of each stocks, obtained as $100 \log(S_t/S_{t-1})$, where S_t is the index's price at day t. The following table shows the summary statistics of the returns.

TABLE 3 ABOUT HERE

We can observe that the sample kurtosis coefficients are large for all series, compared to the kurtosis coefficient of the standard normal distribution which is 3. In the application the returns have been centered before applying the test, although this does not make any difference for the results below.

Table 4 indicates the p-values for the tests and Monte Carlo setup of Section 5.

TABLE 4 ABOUT HERE

The data-driven smooth test rejects the MDH for all stocks with exception of Cit and GE, for all reasonable nominal levels. Omnibus tests are unable to reject these alternatives (with the exception of the KS test for Sta at 5%, with a p-value of 0.038) and the $t-$test only does for Ec and is doubtful for the S&P 500. The smooth test $T_{n,3}$ rejects the MDH at 5% for all stocks but for AF, Cit, GE and GM. Thus, it seems that AF and GM are high-frequency alternatives. To gain some insight in the character (low- or high- frequency) of the alternatives we report in Table 5 the choice of $\widetilde{m}(d)$ and the corresponding squared component $\widehat{\epsilon}^2_{\widetilde{m}(d),n}$ for $d = 10$. We also note that the first significative component of AF is the 5th, with $\widehat{\epsilon}^2_{5,n} = 15.2$. Notice that under the null, $\widehat{\epsilon}^2_{5,n}$ is distributed as a χ^2_1 distribution, so this value is significatively different from zero. For GM the first significative component is the 4th, with $\widehat{\epsilon}^2_{4,n} = 4.4$. It is worth to remark that for GM $\widehat{\epsilon}^2_{10,n} = 27.97$. Then, it is confirmed that AF and GM are high-frequency alternatives. The omnibus and the smooth tests with a fixed number of components are silent with respect to these alternatives, whereas the new data-driven smooth test clearly rejects the MDH. These examples highlight the properties of the new tests.

Our new smooth tests statistic find nonlinear dependence in the conditional mean of these stocks, in contrast to most of the theoretical and applied literature which assume no structure in the conditional mean of financial data. The nonlinearity in the conditional mean suggests that additional effort has to be dedicated to investigate the form of such nonlinearity before modeling the conditional variance.

TABLE 5 ABOUT HERE

To gain some insights in the nonlinearity of the S&P 500 we plot in Figure 2 the fitted regression from

$$Y_t = c_1 \cos\left(0.5\sigma^{-2}\pi\tau^2(Y_{t-1})\right) + c_2 \cos\left(1.5\sigma^{-2}\pi\tau^2(Y_{t-1})\right) + \varepsilon_t. \tag{8}$$

We plot the least squares fitted values $\widehat{Y}_t$ from the previous regression against the lagged values of Y_{t-1}. Figure 2 reveals that the conditional mean at lag one of the S&P 500 is nonlinear and non-decreasing around zero, a feature which is consistent with the well-known fact that the sample autocorrelation at lag $j = 1$ of stock returns is usually positive. We observe an asymmetric effect in the fitted conditional mean with variations in negative values of Y_{t-1} larger than variations in positive values. This is consistent with the well-known "Leverage effect" in stocks returns, in which volatility is higher when past stock returns are negative. The positive correlation effect is reversed for "large" absolute values of the returns. More concretly, for values larger than 1.2 the fitted regression is decreasing, reflecting the expectations of investors that after a large positive return foresee a decay in the stock price return.

FIGURE 2 ABOUT HERE

---

6 Extensions, modifications and conclusions

This section discusses extensions and modifications of the basic setup considered in the paper. The major limitation of the proposed methodology is that tests based on R_n in (2) only have power against alternatives satisfying $E[Y_t 1(Y_{t-1} \leq x)] \neq 0$ in a set with positive Lebesgue measure. These alternatives correspond to those such that $E[Y_t \mid Y_{t-1}] \neq 0$. The motivation for the use of just Y_{t-1} as the conditioning variable is practical, one expects that the most important lag is the first one in real data, but mostly theoretical, since the principal components and eigenvalues associated to R_n are only known for this univariate case.

Here, we discuss two alternatives for applying the methodology of this paper to a more general multivariate case. We can consider the situation where the conditioning set is a $d-$variate random vector, say $\mathbf{X}_t = (Y_t, \mathbf{Z}_t', , ..., Y_{t-P+1}, \mathbf{Z}_{t-P+1}')'$ where $\mathbf{Z}_t$ is a vector of explanatory random variables, and the mean of Y_t is different from zero. That is, we are now concerned with testing the hypothesis

$$H_0^* : E[Y_t \mid \mathbf{X}_{t-1}] = \mu \ a.s. \qquad \mu \in \mathbb{R}. \tag{9}$$

This is the set-up considered in Domínguez and Lobato (2003).

The first possibility we mention consists in (nonparametrically) estimating the principal com-

ponents of the marked empirical process

$$R_{n,w}(\mathbf{x}) := n^{-1/2}\sum_{t=1}^{n}(Y_t - \overline{Y})w(\mathbf{X}_{t-1}, \mathbf{x}) \qquad \mathbf{x} \in \mathbb{R}^d,$$

for a suitable weight function w and where $\overline{Y}$ is the usual sample mean. See Escanciano (2006) for possible functions w. The principal components of $R_{n,w}$ can be estimated consistently along the lines suggested in Escanciano (2005) by solving the eigenvalues and eigenvectors of an $n \times n$ matrix.

The second possibility is based on a marked process based on projections

$$R_{n,ind}(\boldsymbol{\beta}_0, x) := n^{-1/2}\sum_{t=1}^{n}(Y_t - \overline{Y})1(\boldsymbol{\beta}_0'\mathbf{X}_{t-1} \leq x) \qquad x \in \mathbb{R}.$$

The direction of projection $\boldsymbol{\beta}_0$ can be computed from projection pursuit techniques (c.f. Huber 1985) or from dimension reduction techniques as in Cook and Li (2002). Although, how to choose the projection direction is important, it will not be discussed here for the sake of space. In the context of Generelized Linear models, Stute, Presedo-Quindimil, González-Manteiga and Koul (2006) advocate for the use of $R_{n,ind}(\boldsymbol{\beta}_n, x)$, with $\boldsymbol{\beta}_n$ an estimator of the Generalized Linear model parameter. Here we assume that if $\boldsymbol{\beta}_0$ is unknown, it can be estimated by a $\sqrt{n}-$consistent estimator $\boldsymbol{\beta}_n$, without restricting ourselves to a particular estimator.

As shown by Stute et al. (2006), under mild conditions, it follows that

$$\sup_{x\in\mathbb{R}} |R_{n,ind}(\boldsymbol{\beta}_n, x) - R_{n,ind}(\boldsymbol{\beta}_0, x)| \xrightarrow{P} 0.$$

Moreover, the asymptotic distribution of $R_{n,ind}(\boldsymbol{\beta}_n, u)$ under H_0^* is a Gaussian process with covariance function

$$K(x_1, x_2) = E[(Y_t - \mu)^2 \left\{1(\boldsymbol{\beta}_0'\mathbf{X}_{t-1} \leq x_1) - F_{\boldsymbol{\beta}_0}(x_1)\right\}\left\{1(\boldsymbol{\beta}_0'\mathbf{X}_{t-1} \leq x_2) - F_{\boldsymbol{\beta}_0}(x_2)\right\}],$$

where $F_{\boldsymbol{\beta}_0}$ is the cdf of $\boldsymbol{\beta}_0'\mathbf{X}_{t-1}$.

Let $\psi(\cdot)$ be the function defined by

$$\psi(u) = \int_{-\infty}^{u} \sigma^2_{\boldsymbol{\beta}_0}(x) F_{\boldsymbol{\beta}_0}(dx),$$

with $\sigma^2_{\boldsymbol{\beta}_0}(x) := E\left[(Y_t - \mu)^2 \mid \boldsymbol{\beta}_0'\mathbf{X}_{t-1} = x\right]$ the conditional variance. We shall transform the limit process of $R_{n,ind}(\boldsymbol{\beta}_n, u)$ to a Brownian motion in proper time. It is not only the asymptotic distribution freeness of the transformation which makes this approach attractive. Rather the transformed process may be used to construct smooth tests through a principal component decomposition as

discussed above. Let

$$A(u) := \int_u^\infty \sigma^{-2}_{\boldsymbol{\beta}_0}(x) F_{\boldsymbol{\beta}_0}(dx).$$

Assume throughout that $A(u) \neq 0 \ \forall u \in \mathbb{R}$, and consider the linear operator $T(\cdot)$ defined by

$$Tf(u) := f(u) - \int_{-\infty}^u A^{-1}(x) \int_x^\infty \sigma^{-2}_{\boldsymbol{\beta}_0}(y) f(dy) F_{\boldsymbol{\beta}_0}(dx),$$

where $f(\cdot)$ is either of bounded total variation or a Brownian motion $B \circ \psi$. In the latter case, the integral needs to be interpreted as a stochastic integral. Such transformations have been considered in the goodness-of-fit literature by Khmaladze (1981, 1988), see also Koul and Stute (1999), Stute and Zhu (2002) and Delgado, Hidalgo and Velasco (2005), among others. Note that $T(\cdot)$ depends on unknown quantities. A natural estimator of $T(\cdot)$ is

$$T_n f(u) = f(u) - \int_{-\infty}^u A_n^{-1}(x) \int_x^\infty \sigma^{-2}_{n,\boldsymbol{\beta}_n}(y) f(dy) F_{n,\boldsymbol{\beta}_n}(dx),$$

where

$$A_n(u) = \int_u^\infty \sigma^{-2}_{n,\boldsymbol{\beta}_n}(x) F_{n,\boldsymbol{\beta}_n}(dx),$$

and $\sigma^{-2}_{n,\boldsymbol{\beta}_n}(x)$ is any nonparametric consistent estimator of the conditional variance, e.g. a Nadaraya-Watson estimator. Then the transformed process can be written as

$$\begin{aligned} T_n R_{n,ind}(\boldsymbol{\beta}_n, u) &= R_{n,ind}(\boldsymbol{\beta}_n, u) - \frac{1}{n^{3/2}} \sum_{t=1}^n \sum_{s=1}^n 1(\boldsymbol{\beta}_n' \mathbf{X}_{t-1} \leq u) A_n^{-1}(\boldsymbol{\beta}_n' \mathbf{X}_{t-1}) \\ \times 1(\boldsymbol{\beta}_n' \mathbf{X}_{t-1} &\leq \boldsymbol{\beta}_n' \mathbf{X}_{s-1})(Y_s - \overline{Y}) \sigma^{-2}_{n,\boldsymbol{\beta}_n}(\boldsymbol{\beta}_n' \mathbf{X}_{s-1}). \end{aligned}$$

It can be shown that under the null hypothesis H_0^* and some mild conditions, including that $\sigma^2_{n,\boldsymbol{\beta}_n}$ is a uniformly consistent estimator of $\sigma^2_{\boldsymbol{\beta}_0}$, we have that

$$T_n R_{n,ind}(\boldsymbol{\beta}_n, u) \Longrightarrow B \circ \psi(u) \text{ in } D([-\infty, \infty)).$$

See Stute, Thies and Zhu (1998).

In particular from the scaling properties of the Brownian motion and the Continuous Mapping Theorem we have that

$$\int_{-\infty}^a \left| \psi_n^{-1}(a) T_n R_{n,ind}(\boldsymbol{\beta}_n, x) \right|^2 \psi_n(dx) \xrightarrow{d} \int_0^1 |B(u)|^2 \, du, \tag{10}$$

where

$$\psi_n(x) = n^{-1}\sum_{t=1}^{n}(Y_t - \overline{Y})^2 1(\boldsymbol{\beta}_n'\mathbf{X}_{t-1} \leq x).$$

At this point, the data-driven smooth test can be computed in exactly the same manner as in Section 3. Formal details are omitted for the sake of space.

Now, we discuss different selection rules for the data-driven smooth test. In this paper we have adopted the combination rule of Inglot and Ledwina (2006a), but other rules are available in the literature. See the aforementioned references. Among other rules, the most popular choice in the context of smooth tests is the Schwarz's selection rule of Ledwina (1994). This corresponds to

$$m_{BIC} = \min\{m : 1 \leq m \leq d; L_m \geq L_h, h = 1, 2..., d\},$$

where

$$L_m = T_{n,m} - m\log n.$$

Also other choices of q in our combination rule can be considered. Inglot and Ledwina (2006b) provided simulations with different choices of q for data-driven smooth tests for testing uniformity in the context of goodness-of-fit tests for distributions. Here we run a small Monte Carlo experiment and consider the AR(1)-GARCH(1,1) model as in the Monte Carlo section for $c = -0.1$, 0 and 0.1. We recall that small values of q in the combination rule results in the Akaike's criterion, while large $q's$ lead to the choice of the Schwarz's criterion. This is confirmed in the simulations. In Table 6 we report the empirical rejection probabilities for the AR(1)-GARCH(1,1) model. $T_{n,m_{BIC}}$ stands for the data-driven smooth test with the Schwarz's selection rule whereas $T_{n,\widetilde{m}(q)}$ denotes the data-driven test with our combination rule using q. We observe that different values of q lead to small variations in the rejection probabilities.

TABLE 6 ABOUT HERE

To conclude, we have proposed a new data-driven smooth test for the MDH with excellent power properties, comparing well to other competing tests. Theoretical results such as the lack of power of omnibus tests or the inability of the $t-$test to detect certain nonlinear alternatives have been confirmed also empirically. The new smooth tests provide a compromise between the omnibus tests, which are consistent against all alternatives, and directional tests, which are optimal in a given (unique) univariate direction. Optimality, in a semiparametric sense, of our test has been shown for a class of Markov processes. We have demonstrated that high-frequency alternatives are likely to appear in financial applications. Our data-driven smooth tests are especially convenient to

detect such alternatives, while being able to detect also low-frequency alternatives. An important extension of our tests would be to consider the bound d tending to infinity with the sample size. This extension is difficulted by the general serial dependence allowed in our framework. It is expected that under suitable mixing conditions such an extension can be accomplished. This challenging problem is deferred for future research.

7 Appendix: Mathematical Proofs

PROOF OF THEOREM 1: The proof follows easily from Lemma 2 in Domínguez and Lobato (2004), and hence, it is omitted. ■

PROOF OF THEOREM 2: (i) The Uniform Ergodic Theorem, see e.g. Dehling and Philipp (2002), and A1 yield

$$\sup_{x\in\mathbb{R}} \left|\tau_n^2(x) - \tau^2(x)\right| = o_P(1).$$

The last display, Theorem 1 and Lemma 3.1 in Chang (1990) imply, for $1 \leq j \leq d$,

$$\begin{aligned} \widehat{\epsilon}_{j,n} &= \lambda_j^{-1/2} \int_{\mathbb{R}} \psi_j(\tau^2(x)) R_n(x) \tau^2(dx) + o_P(1) \\ &:= \widetilde{\epsilon}_{j,n} + o_P(1), \end{aligned}$$

where $\widetilde{\epsilon}_{j,n}$ can be written as

$$\widetilde{\epsilon}_{j,n} := \frac{\sqrt{2}}{\sigma \lambda_j^{1/2} \sqrt{n}} \sum_{s=1}^{n} Y_s \cos((j-1/2)\sigma^{-2}\pi\tau^2(Y_{s-1})). \tag{11}$$

Notice that, $E[\widetilde{\epsilon}_{j,n}] = 0$ and

$$\begin{aligned} E[\widetilde{\epsilon}_{j,n}\widetilde{\epsilon}_{h,n}] &= 2\int_{\mathbb{R}} \cos((j-1/2)\pi\tau^2(x)) \cos((j-1/2)\pi\tau^2(x))\tau^2(dx) \\ &= \delta_{jh}, \end{aligned}$$

where $\delta_{jh} = 0$ if $j \neq h$, and $\delta_{jh} = 1$ otherwise.

Now, by the Cramer-Wold device, A1 and the Central Limit Theorem (CLT) for martingales with stationary and ergodic differences (Billingsley (1961)) it readily follows that the vector $(\widetilde{\epsilon}_{1,n}, ..., \widetilde{\epsilon}_{m,n})'$ converges to a $m-$variate standard normal random vector. This implies part (i).

As for part (ii). Denote the Schwarz's rule for the choice of m by

$$m_{BIC} = \min\{m : 1 \leq m \leq d; L_m \geq L_h, h = 1, 2..., d\},$$

where

$$L_m = T_{n,m} - m \log n.$$

We shall prove that under H_0, $\widetilde{m} = m_{BIC}$ with probability tending to one. To that end, define the event

$$A_n(q) = \left\{ \max_{1 \leq j \leq d} |\widehat{\epsilon}_{j,n}| > \sqrt{q \log n} \right\}.$$

From part (i) we have that $\max_{1 \leq j \leq d} |\widehat{\epsilon}_{j,n}| = O_P(1)$. Thus, it follows that $P(A_n(q)) = o(1)$, which in turn implies $\lim_{n \to \infty} P(\widetilde{m} = m_{BIC}) = 1$.

Now, we prove that, again under H_0,

$$\lim_{n \to \infty} P(m_{BIC} = 1) = 1. \tag{12}$$

First, notice that

$$P(m_{BIC} = 1) = 1 - \sum_{j=1}^{d} P(m_{BIC} = j) \leq 1 - \sum_{j=1}^{d} P(L_j \geq L_1). \tag{13}$$

Now,

$$P(L_j \geq L_1) \leq P(T_{n,j} \geq (j-1) \log(n)) \leq C n^{-\zeta}, \qquad \text{for some } \zeta > 0,$$

where the last inequality follows from the moderate deviation inequality for multivariate martingales in Grama and Haeusler (2006), see their Theorem 2 and Corollary 1. Therefore, Theorem 2 follows from (12) and application of the standard CLT for martingales. The theorem is proved. ■

Before proving Theorem 3 we need some notation and discussion. We proceed to investigate the Pitman asymptotic relative efficiency of tests in this semiparametric testing environment, along the lines of Choi, Hall, and Schick (1996). Write $(\varepsilon, X)'$ for a r.v. with the same distribution as $(\varepsilon_t, Y_{t-1})'$. Y has also the same distribution as X. Similarly, $\mathbf{Z}$ denotes a r.v. with the same probability distribution, say P, as $\mathbf{Z}_t = (Y_t, Y_{t-1})'$, $t \in \mathbb{Z}$. Let $L_2(P)$ be the space of square integrable random vectors with respect to P and let $||\cdot||_{2,P} \equiv ||\cdot||_2$ indicate the $L_2(P)$ norm. Likewise, define $L_2(P_n)$ and $||\cdot||_{2,P_n}$, where P_n is the empirical probability measure of $\{\mathbf{Z}_t\}_{t=1}^n$. Finally, let $\mathcal{P}$ be the set of probability measures for P for which the regularity conditions below hold.

The nuisance parameters in the model (7) are given by $\boldsymbol{\eta}_0 = (f_{\varepsilon|X}(\cdot), f(\cdot))'$, where $f_{\varepsilon|X}(\cdot)$ is the conditional density of errors ε given X, and $f(\cdot)$ is the density of X. The parameter of interest is $\mathbf{c} \in \mathbb{R}^m$. Let $\gamma_0 = (0, \boldsymbol{\eta}_0)$ and $\boldsymbol{\gamma} = (\mathbf{c}, \boldsymbol{\eta})$ with $\boldsymbol{\eta} = (h_{\varepsilon|X}(\cdot), h(\cdot))' \in \mathcal{H} = \mathcal{B}_1 \times \mathcal{B}_2$. Here $\mathcal{B}_1$ is the set of all conditional error densities consistent with the model (7) and $\mathcal{B}_2$ is the set of all densities for X, in both cases the densities are dominated by a particular σ-finite measure λ. Define the densities

$$g(y, x, \mathbf{c}, \boldsymbol{\eta}) = h_{\varepsilon|x}(y - \mathbf{c}'\mathbf{h}(x)) h_x(x)$$

and consider the family of probabilities: $\tilde{\mathcal{P}} := \{P \in \mathcal{P} : dP/d\lambda = g(y, x, \mathbf{c}, \boldsymbol{\eta})$, with $\int \varepsilon h_{\varepsilon|x}(\varepsilon|x)d\varepsilon = 0\}$. The family $\{P \in \tilde{\mathcal{P}} : dP/d\lambda = g(y, x, \mathbf{0}, \boldsymbol{\eta})\}$ represents the space of models under the null hypothesis. Then the whole class of semiparametric models under consideration are characterized by the family of distributions

$$\{P \in \tilde{\mathcal{P}} : dP/d\lambda = g(y, x, \mathbf{c}, \boldsymbol{\eta}), (\mathbf{c}, \boldsymbol{\eta}) \in \mathbb{R}^m \times \mathcal{H}\}.$$

The construction of the efficient score test proceeds as follows. Given the score $\dot{\mathbf{l}}_1$ in the marginal family $\mathcal{P}_1 = \{P \in \tilde{\mathcal{P}} : dP/d\lambda = g(y, x, \mathbf{c}, \boldsymbol{\eta}_0),\ \mathbf{c} \in \mathbb{R}^m\}$, one computes the tangent space $\dot{\mathcal{P}}_2$ of the nuisance parameter family $\mathcal{P}_2 = \{P \in \tilde{\mathcal{P}} : dP/d\lambda = g(y, x, \mathbf{0}, \boldsymbol{\eta}), \boldsymbol{\eta} \in \mathcal{H}\}$. Then the efficient score $\mathbf{l}_1^*$ can be constructed by the orthogonal projection of score $\dot{\mathbf{l}}_1$ on the orthocomplement of $\dot{\mathcal{P}}_2$, i.e., $\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - \Pi[\dot{\mathbf{l}}_1|\dot{\mathcal{P}}_2]$, where $\Pi[h|\dot{\mathcal{P}}_2]$ denotes the orthogonal projection in $L_2(P)$ of h on $\dot{\mathcal{P}}_2$. A score test using the efficient score along this least favorable direction will be an *asymptotically uniformly most powerful invariant and asymptotically unbiased* test at $\boldsymbol{\eta}_0$, in short AUMPIU $(\alpha, \boldsymbol{\eta}_0)$, as defined in Choi, Hall, and Schick (1996). Eventually it turns out that the test does not depend on $\boldsymbol{\eta}_0$, extending its uniformity over all alternatives with different $\boldsymbol{\eta}_0$'s. In this case, we say that the test is AUMPIU(α).

Wefelmeyer (1997) characterized the tangent space $\dot{\mathcal{P}}_2$ of the nuisance parameter family $\mathcal{P}_2$. The tangent space $\dot{\mathcal{P}}_2$ at $P \in \mathcal{P}_2$ is given by

$$\dot{\mathcal{P}}_2 = \{\mathbf{s} \in L_2(P) : E[\mathbf{s}(\mathbf{Z})] = 0, E[Y\mathbf{s}(\mathbf{Z})|X] = \mathbf{0}\}.$$

The following lemma establishes the projection operator $\Pi[h|\dot{\mathcal{P}}_2]$.

LEMMA A1: Under Assumptions A1-A2, $\dot{\mathcal{P}}_2$ is the tangent space of the nuisance parameter family $\mathcal{P}_2$, and

$$\Pi[\mathbf{s}|\dot{\mathcal{P}}_2](\mathbf{z}) = \mathbf{s}(\mathbf{z}) - E[\mathbf{s}(\mathbf{Z})] - y\sigma^{-2}(x)E[Y\mathbf{s}(\mathbf{Z})|X = x],\ \text{for } \mathbf{s} \in L_2(P),\ \mathbf{z} = (y, x).$$

PROOF OF LEMMA 1: That $\dot{\mathcal{P}}_2$ is the tangent space of the nuisance parameter family $\mathcal{P}_2$ follows from Wefelmeyer (1997). For the rest of the proof it suffices to show that (a) $\Pi[\mathbf{s}|\dot{\mathcal{P}}_2] \in \dot{\mathcal{P}}_2$ and (b) $\mathbf{s} - \Pi[\mathbf{s}|\dot{\mathcal{P}}_2] \perp \dot{\mathcal{P}}_2$. To show (a), notice that using the null restriction $E[Y|X] = 0$, we have $E(\Pi[h|\dot{\mathcal{P}}_2]) = 0$. Also

$$\begin{aligned} E[Y\Pi[\mathbf{s}|\dot{\mathcal{P}}_2]|X] &= E[Y\mathbf{s}(\mathbf{Z})|X] - E[Y\mathbf{s}(\mathbf{Z})|X = x] \\ &= \mathbf{0}. \end{aligned}$$

Hence (a) is proved. To show (b), notice that for $\mathbf{s} \in \dot{\mathcal{P}}_2$,

$$\begin{aligned} E[\Big(\mathbf{h} - \Pi[\mathbf{h}|\dot{\mathcal{P}}_2]\Big)\,\mathbf{s}(\mathbf{Z})] &= E[Y\sigma^{-2}(X)E[Y\mathbf{h}(\mathbf{Z})|X=x]\mathbf{s}(\mathbf{Z})] \\ &= \mathbf{0}. \end{aligned}$$

■

PROOF OF THEOREM 3: The marginal score $\dot{\mathbf{l}}_1$ is given by

$$\dot{\mathbf{l}}_1(\mathbf{z}) = -\mathbf{h}(x)l_{\varepsilon/x}(\mathbf{z}),$$

where $l_{\varepsilon/x}(\mathbf{z}) := \partial \ln f_{\varepsilon/x}(\mathbf{z})/\partial\varepsilon$. Notice that $E[l_{\varepsilon/x}(\mathbf{Z})Y \mid X] = -1$. Hence,

$$\begin{aligned} \mathbf{l}_1^* &= \dot{\mathbf{l}}_1 - \Pi[\dot{\mathbf{l}}_1|\dot{\mathcal{P}}_2] \\ &= y\sigma^{-2}(x)E[Y\dot{\mathbf{l}}_1(\mathbf{Z})|X=x] \\ &= y\sigma^{-2}(x)\mathbf{h}(x). \end{aligned}$$

An optimal (semiparametric) score test rejects the null hypothesis H_0 for large values of

$$T_{n,e}^*(a) = \left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{l}_1^*(\mathbf{Z}_t)\right\}' \mathbf{I}^{-1} \left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{l}_1^*(\mathbf{Z}_t)\right\},$$

where $\mathbf{I} := ||\mathbf{l}_1^*(\mathbf{Z}_t, a)||_{2,P}$. Similar arguments to those of Theorem 2 show that $\mathbf{I}$ is the identity matrix of order m and that

$$T_{n,e}^*(a) = \sum_{j=1}^{m}(\widetilde{\epsilon}_{j,n})^2 = \sum_{j=1}^{m}\widetilde{\epsilon}_{j,n}^2 + o_P(1),$$

where $\widetilde{\epsilon}_{j,n}$ is defined in (11). We have shown that the test based on $T_{n,e}(a)$ is AUMPIU(α) for testing $H_{s0} : \mathbf{c} = \mathbf{0}$ against $H_{s1} : \mathbf{c} \neq \mathbf{0}$. ■

REFERENCES

Aerts, M., Claeskens, G. and Hart, J. D. (1999), "Testing Lack of Fit in Multiple Regression," *Journal of the American Statistical Association,* 94, 869-879.

Bickel, P. J. , A.J. Klaassen, Y. Rikov, and J. A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models.* Springer Verlag, New York.

Billingsley, P. (1961), "The Lindeberg-Levy Theorem for Martingales," *Proceedings of the American Mathematical Society,* 12, 788-792.

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

Bosq, D. (2000), *Linear Processes in Function Spaces. Theory and Applications.* Springer, New York.

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

Chang, N. M. (1990), "Weak Convergence of a Self-Consistent Estimator of a Survival Function with Doubly Censored Data," *Annals of Statistics,* 18**,** 391-404.

Choi, S., W. J. Hall, and A. Schick (1996), "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *Annals of Statistics*, 24, 841-861.

Cochrane, J. (2001), *Asset Pricing.* Princeton University Press. Princeton, New Jersey.

Cook, R.D., and Li, B. (2002), "Dimension Reduction for Conditional Mean in Regression," *Annals of Statistics*, 30, 455-474.

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

Delgado, M. A., Hidalgo, J. and Velasco, C. (2005), "Distribution Free Goodness-of-fit Tests for Linear Models," *Annals of Statistics,* 33, 2568-2609.

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

Dominguez, M. A and Lobato, I. N. (2003), "A Consistent Test for the Martingale Difference Hypothesis," *Econometric Reviews*, 22, 351-377.

Durbin, J. and Knott, M. (1972), "Components of Cramer-von Mises Statistics. I," *Journal of the Royal Statistical Society. Ser. B*, 34, 290-307.

Durlauf, S. (1991), "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50, 355-376.

Escanciano, J. C. (2006), "Goodness-of-Fit Tests for Linear and Non-Linear Time Series Models," *Journal of the American Statistical Association,* 101, 531-541.

Escanciano, J. C. (2005), "On The Asymptotic Power Properties of Specification Tests for Dynamic Parametric Regressions," Working paper 07/05 Universidad de Navarra.

Escanciano, J. C., and Velasco, C. (2006), "Generalized Spectral Tests for the Martingale Difference Hypothesis," *Journal of Econometrics,* 134, 151-185.

Eubank, R. (2000), "Testing for No Effect by Cosine Series Methods," *Scandinavian Journal of*

Statistics, 27, 747-763.

Eubank, R.L. and LaRiccia, V.M. (1992), "Asymptotic Comparison of Cramer-von Mises and Nonparametric Function Estimation Techniques for Testing Goodness-of-Fit," *Annals of Statistics,* 20, 2071-2086.

Eubank, R., and Hart, J. (1992), "Testing Goodness-of-Fit in Regression via Order Selection Criteria," *Annals of Statistics,* 20, 1412-1425.

Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association,* 91, 674-688.

Fan, J. and Huang, L. (2001), "Goodness-of-Fit Tests for Parametric Regression Models," *Journal of the American Statistical Association,* 96, 640-652.

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, 199-240.

Grama, I. G., and Haeusler, E. (2006), "An Asymptotic Expansion for Probabilities of Moderate Deviations for Multivariate Martingales," *Journal of Theoretical Probability*, 19, 1-44.

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

Hsieh, D. A. (1989), "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *Journal of Business,* 62, 339-368.

Huber, P. J. (1985), "Projection Pursuit (with discussion)," *Annals of Statistics*, 13, 435-525.

Inglot, T. and Ledwina, T. (1996), "Asymptotic Optimality of Data-Driven Neyman's tests for Uniformity," *Annals of Statistics,* 24, 1982-2019.

Inglot, T. and Ledwina, T. (2006a), "Towards Data Driven Selection of a Penalty Function for Data Driven Neyman Tests," *Linear Algebra and its Applications*, 417, 124-133.

Inglot, T. and Ledwina, T. (2006b), "Data Driven Score Tests of Fit for Semiparametric Homoscedastic Linear Regression Model", Preprint 665, Institute of Mathematics Polish Academy of Sciences.

Kallemberg, W. C. M. (2002), "The Penalty in Data Driven Neyman's Tests," *Mathematical Methods of Statistics,* 11, 323-340.

Kallenberg, W. C. M. and Ledwina, T. (1997), "Data-Driven Smooth Tests When the Hypothesis is Composite," *Journal of the American Statistical Association,* 92, 1094-1104.

Khmaladze, E.V. (1981), "Martingale Approach in the Theory of Goodness-of-Fit Tests," *Theory of Probability and its Applications,* 26, 240-257.

Khmaladze, E.V. (1988), "An Innovation Approach to Goodness-of-Fit Tests in $\mathbb{R}^m$," *Annals of Statistics*, 16, 1503-1516.

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

Koziol, J. A. (1980), "On a Cramer-von Mises-Type Statistic for Testing Symmetry," *Journal of the American Statistical Association,* 75, 161-167.

Janssen, A. (2000), "Global Power Functions of Goodness of Fit Tests," *Annals of Statistics,* 28, 239-253.

Ledwina, T. (1994), "Data-Driven Version of Neyman's Smooth Test of Fit," *Journal of the American Statistical Association,* 89, 1000-1005.

Neyman, J. (1937), "Smooth Test for Goodness of Fit," *Scandinavian Aktuarietidskr,* 20, 149-199.

Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness of Fit.* London: Oxford University Press.

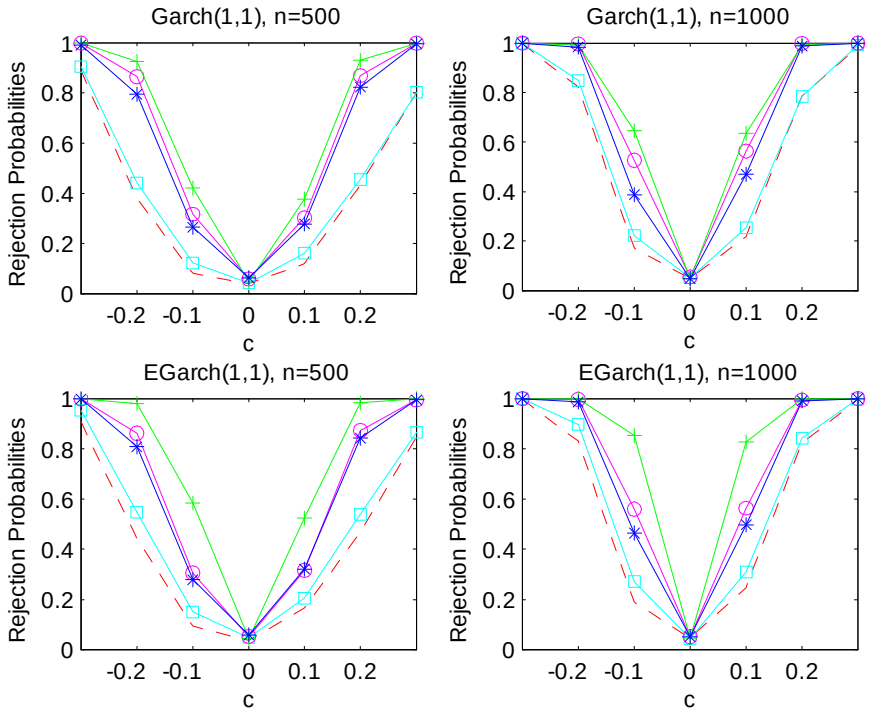
Stute, W., Presedo Quindimil, M., Gonzalez Manteiga, W. and Koul, H. L. (2006), "Model Checks of Higher Order Time Series," *Statistica Sinica*, forthcoming.

Stute, W., Thies, S., and Zhu, L.X. (1998), "Model Checks for Regression: an Innovation Process Approach," *Annals of Statistics,* 26, 1916-1934.

Stute, W. and Zhu, L. X. (2002), "Model Checks for Generalized Linear Models," *Scandinavian Journal of Statistics,* 29, 535-545.

Wefelmeyer, W. (1997), "Adaptive Estimators for Parameters of the Autoregression Function of a Markov Chain," *Journal of Statistical Planning and Inference,* 58, 389-398.

Figure 1: Size and Power of Tests at 5% for the model (1-2)



KS-dash, CvM-square, t-test–plus, $T_{n,\widetilde{m}}$-start, $T_{n,3}$-circle. 5% level, sample size $n = 500$ and $n = 1000$. Innovations distributed as $N(0, 1)$.

Table 1: Power and Size of Tests at 5% for the model 3.

n=500	$c = -0.6$		$c = -0.3$		$c = 0.0$		$c = 0.3$		$c = 0.6$	
	$j = 2$	$j = 3$	$j = 2$	$j = 3$	$j = 2$	$j = 3$	$j = 2$	$j = 3$	$j = 2$	$j = 3$
KS	0.123	0.059	0.077	0.067	0.038	0.036	0.067	0.052	0.131	0.058
CvM	0.067	0.042	0.068	0.070	0.039	0.044	0.062	0.048	0.068	0.038
$t-test$	0.058	0.042	0.051	0.049	0.035	0.060	0.047	0.048	0.038	0.047
$T_{n,3}$	0.179	0.056	0.095	0.063	0.051	0.038	0.094	0.058	0.155	0.053
$T_{n,\widetilde{m}}$	0.989	0.858	0.273	0.191	0.051	0.060	0.287	0.192	0.993	0.871

Table 2: Power of Tests at 5% for the model NLMA

	$n = 300$	$n = 500$
KS	0.039	0.074
CvM	0.042	0.051
$t-test$	0.065	0.073
$T_{n,3}$	0.480	0.642
$T_{n,\widetilde{m}}$	0.781	0.755

Table 3: Summary statistics of the returns

n=755	**S&P500**	**AF**	**BA**	**Cit**	**Ea**	**Ec**	**Ex**	**GE**	**GM**	**Pep**	**Sta**
Mean	-0.042	-0.269	-0.036	-0.039	-0.069	-0.052	-0.061	-0.043	0.091	-0.039	-0.137
Median	-0.071	0.013	-0.059	-0.025	-0.089	0.000	-0.150	0.000	0.129	-0.035	-0.031
SD	0.821	1.716	0.997	1.144	1.420	1.145	1.199	1.165	2.004	1.379	1.631
Skew.	-0.100	-0.768	1.800	-0.079	0.074	-0.201	0.339	-0.255	-0.045	0.028	-0.727
Kurt.	4.234	4.766	20.318	4.378	5.164	3.484	3.912	4.413	14.932	12.420	11.246
Maxi.	3.587	3.123	10.677	4.116	7.677	3.252	4.455	4.642	15.045	10.234	8.560
Minim.	-3.481	-6.518	-2.654	-5.320	-5.808	-5.103	-3.677	-5.756	-16.647	-9.481	-13.659

Table 4: p-values of the tests. S&P 500 and some of its components.

	S&P500	**AF**	**BA**	**Cit**	**Ea**	**Ec**	**Ex**	**GE**	**GM**	**Pep**	**Sta**
KS	0.136	0.150	0.299	0.424	0.088	0.170	0.170	0.597	0.256	0.299	0.038
CvM	0.113	0.136	0.248	0.447	0.127	0.137	0.166	0.566	0.227	0.248	0.103
$t-test$	0.045	0.462	0.285	0.624	0.788	0.000	0.299	0.700	0.745	0.395	0.057
$T_{n,3}$	0.000	0.405	0.012	0.227	0.004	0.000	0.000	0.699	0.406	0.001	0.018
$T_{n,\widetilde{m}}$	0.000	0.000	0.000	0.571	0.004	0.000	0.000	0.528	0.000	0.000	0.000

Table 5: Individual Principal Components. S&P 500 and some of its components.

	S&P500	**AF**	**BA**	**Cit**	**Ea**	**Ec**	**Ex**	**GE**	**GM**	**Pep**	**Sta**
$\widetilde{m}(10)$	2	10	10	1	2	2	2	1	10	2	10
$\widehat{\epsilon}^2_{\widetilde{m}(10),n}$	22.95**	9.05**	81.10**	0.32	6.73**	20.39**	16.29**	0.39	27.97**	15.14**	30.51**

Note: * Significative at 5%, ** Significative at 1%.

Figure 2: Fitted model (8) for S&P 500

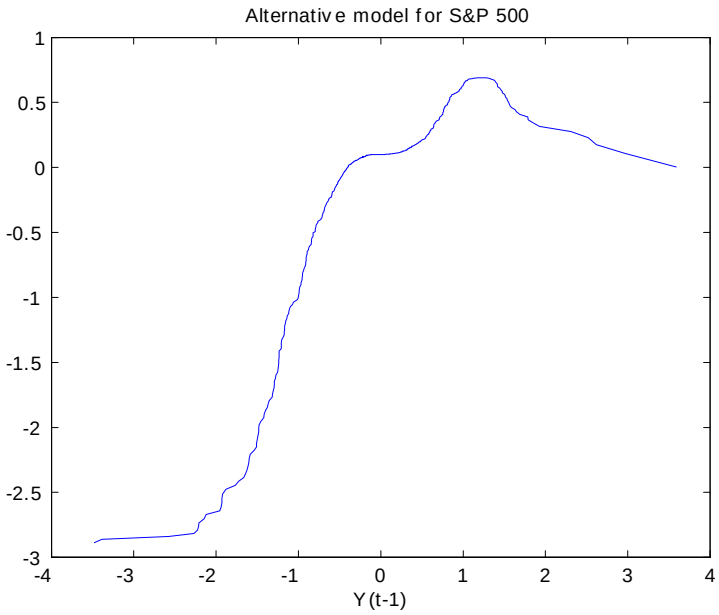


Table 6: GARCH(1,1) Gaussian-errors, $n = 500$

	$\mathbf{c = -0.1}$	$\mathbf{c = 0}$	$\mathbf{c = 0.1}$
$T_{n,m_{BIC}}$	0.259	0.060	0.273
$T_{n,\widetilde{m}(2.2)}$	0.267	0.067	0.279
$T_{n,\widetilde{m}(2.4)}$	0.264	0.064	0.277
$T_{n,\widetilde{m}(2.6)}$	0.264	0.063	0.276
$T_{n,\widetilde{m}(2.8)}$	0.264	0.062	0.276
$T_{n,\widetilde{m}(3)}$	0.262	0.061	0.275
$T_{n,\widetilde{m}(3.2)}$	0.261	0.060	0.274