because $S_t = \sum_{j=1}^{t} y_j$ is an I(1) process. Combining the all the results yields

$$T^{-2/3}R \xrightarrow{p} 2k^{-1} > 0 \quad \text{for C0,}$$

$$T^{-1}R \xrightarrow{p} 2mk^{-1}\sigma_y^2\omega_y^{-2} > 0 \quad \text{for CC,}$$

$$T^{-5/3}R \Rightarrow 2k^{-1}\sigma_y^2\omega_y^{-2}\left\{2\int W^2 + W(1)^2 - 2W(1)\int W\right\}^{-1} > 0 \quad \text{for CI, and}$$

$$T^{-1}R \Rightarrow (2\sigma_y^2 + y_\infty^2)\left\{2\int W^2 + W(1)^2 - 2W(1)\int W\right\}^{-1} > 0 \quad \text{for II.}$$

Therefore, all the test statistics diverge in the positive direction as required. ■

Econometric Theory, **22**, 2006, 373–402. Printed in the United States of America.
DOI: 10.1017/S0266466606060191

MONITORING CONSTANCY OF VARIANCE IN CONDITIONALLY HETEROSKEDASTIC TIME SERIES

LAJOS HORVÁTH
University of Utah

PIOTR KOKOSZKA
Utah State University

AONAN ZHANG
Bank of America

We propose several methods of on-line detection of a change in unconditional variance in a conditionally heteroskedastic time series. We follow the paradigm of Chu, Stinchcombe, and White (1996, *Econometrica* 64, 1045–1065) in which the first m observations are assumed to follow a stationary process and the monitoring scheme has asymptotically controlled probability of falsely rejecting the null hypothesis of no change. Our theory is applicable to broad classes of GARCH-type time series and relies on a strong invariance principle that holds for the squares of observations generated by such models. Practical implementation of the procedures, which uses a bandwidth selection procedure of Andrews (1991, *Econometrica* 59, 817–858), is proposed, and the performance of the methods is investigated by a simulation study.

1. INTRODUCTION

This paper is concerned with on-line detection of a change in unconditional variance in a conditionally heteroskedastic time series. Our sequential testing procedures are similar in spirit to the procedures developed by Chu, Stinchcombe, and White (1996) and Horváth, Hušková, Kokoszka, and Steinebach (2004), who considered monitoring for changes in linear regression models. To explain the idea, suppose r_t, $t = 1,2,\ldots$ are returns on a speculative asset. We assume that the first m observations $r_1,\ldots,r_m$ are a realization of a stationary process and, in particular, the unconditional variance does not change up to time m. We then monitor the observations $r_{m+1}, r_{m+2},\ldots$ as they arrive, and at each time $n > m$ we have to make a decision whether $r_1,\ldots,r_n$ are a realization of the same process as $r_1,\ldots,r_m$. The detection algorithms are designed in such

This research was partially supported by NSF grants INT-0223262 and DMS-0413653 and NATO grant PST.EAP.CLG 980599. Address correspondence to Piotr Kokoszka, Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA; e-mail: piotr@stat.usu.edu.

© 2006 Cambridge University Press 0266-4666/06 $12.00

a way that they will detect changes that lead to a change in the unconditional variance.

The problem of detecting structural changes in conditionally heteroskedastic time series has received increased attention in recent years. Most work has, however, focused on the so-called a posteriori change-point problem in which a historical sample of fixed size is given and a decision has to be made whether one model is suitable for the whole sample. There are many variants of this problem depending on whether one is interested in the constancy of the parameters of a parametric model or merely in the constancy of specified moments. Contributions in this direction have been made by Inclán and Tiao (1994), Chu (1995), Kokoszka and Leipus (1998, 1999, 2000, 2002), Kim, Cho, and Lee (2000), Inoue (2001), Andreou and Ghysels (2002, 2003), and Kulperger and Yu (2005), among others. By contrast, very few contributions have been made to the problem of sequential testing for a structural change in conditionally heteroskedastic time series. Mikosch and Stărică (1999, 2002) suggested a sequential change-point detection method based on the periodogram, whereas Berkes, Gombay, Horváth, and Kokoszka (2004) developed a method based on likelihood scores that applies to the same testing setting as in Chu et al. (1996) but focuses on GARCH(p,q) models rather than linear regression models.

The methods proposed in this paper are essentially nonparametric, even though their specific implementations may require estimating an approximate parametric model for the data. This is because the detectors have a general form $D(X_1,\ldots,X_n)/\hat{\sigma}$, where $\hat{\sigma}^2$ is an estimator of the variance of the sample mean of dependent stationary observations. Whereas the function $D(\cdot)$ is fully specified and depends only on the observations, some estimators $\hat{\sigma}^2$ may rely on model assumptions. In our simulation study, we use a kernel estimator and the data-driven bandwidth selection procedure of Andrews (1991) that relies on postulating and estimating an approximate parametric model for the observations.

Theoretical justification for the procedures proposed here applies to broad classes of conditionally heteroskedastic time series. The main assumption is that the squared returns obey a strong invariance principle; see equation (2.4). Such a very general assumption is possible because recent research has established that practically all heteroskedastic models of importance in econometrics obey the strong invariance principle (2.4); see Carasco and Chen (2002) and further references in Section 2. Using a strong approximation leads to straightforward proofs that avoid the often very intricate arguments used in the work of Chu et al. (1996), who based their theory on the weak convergence of measures in the Skorokhod space.

The paper is organized as follows. After formulating the monitoring problem and presenting some further background in Section 2, we describe the detection schemes and establish their asymptotic properties in Section 3. Section 4 describes the variance estimator used in simulations presented in Section 5. Appendixes A and B contain the proofs of theorems stated in Section 3.

2. PROBLEM FORMULATION AND ASSUMPTIONS

Assume that the returns r_t have mean zero and denote $X_t = r_t^2$, $t = 1,2,\ldots$, so that $\omega_t = EX_t$ is the variance of the tth return.

We assume that

$$\omega_1 = \omega_2 = \cdots = \omega_m =: \omega_0. \tag{2.1}$$

We wish to test

$$H_0 : \omega_n = \omega_0 \quad \text{for } n \geq m+1 \tag{2.2}$$

against

$$H_A : \exists\, n^* \geq m+1 \quad \omega_n = \omega_0 \quad \text{for } n \leq n^*-1 \quad \text{and}$$
$$\omega_n = \omega_* \neq \omega_0 \quad \text{for } n \geq n^*. \tag{2.3}$$

We assume that under H_0 the X_t satisfy the following strong invariance principle:

$$\sum_{1\leq i\leq n} (X_i - \omega_0) - \sigma W(n) \overset{a.s.}{=} o(n^\alpha) \tag{2.4}$$

with some $0 < \alpha < \frac{1}{2}$, where $W(\cdot)$ is a Wiener process.

The following theorem gives sufficient conditions for (2.4) to hold.

THEOREM 2.1. *Let $\{X_k\}$ be a weakly stationary sequence of random variables with mean zero and uniformly bounded $(2+\delta)$th moments for some $0 < \delta \leq 2$. Assume that $\{X_k\}$ satisfies the strong mixing condition*

$$\sup_{A\in\mathcal{F}_1^m, B\in\mathcal{F}_{m+n}^\infty} |P(A\cap B) - P(A)P(B)| \leq n^{-\gamma} \quad \text{for all } m,n \geq 1 \tag{2.5}$$

with $\gamma \geq 300(1+2/\delta)$, where $\mathcal{F}_k^\ell$ denotes the σ-field generated by $X_k, X_{k+1}, \ldots, X_\ell$. Then letting $S_n = X_1 + \cdots + X_n$, the limit

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} ES_n^2$$

exists, and if $\sigma > 0$, then there exists a Wiener process $\{W(t), 0 \leq t < \infty\}$ such that

$$S_n - \sigma W(n) \overset{a.s.}{=} O(n^{1/2-\varrho}),$$

where $\varrho = \delta/600$.

There are several theorems of this type. Theorem 2.1 is due to Phillip and Stout (1975); see their Theorem 8.1 on p. 96. The constants in the theorem are

far from optimal; however, the theorem of Phillip and Stout (1975) covers most applications.

Recently, Carasco and Chen (2002) established easily verifiable sufficient conditions for exponential β-mixing (absolute regularity) and the existence of finite moments in several important models extending the standard GARCH(1,1). Because exponential β-mixing implies exponential strong mixing (see, e.g., Bradley, 1986, eqn. (1.7)), these conditions imply (2.5) whenever $X_k = f(r_k)$ and the r_k follow one of the GARCH(1,1)-type models considered by Carasco and Chen (2002), $f(\cdot)$ being any measurable function. Necessary and sufficient conditions for the existence of higher order moments in these models were established by He and Teräsvirta (1999). The results of Carasco and Chen (2002) also imply that condition (2.5) holds if $X_k = f(r_k)$, the r_k follow a GARCH(p,q) process that satisfies $\sum_i \alpha_i + \sum_j \beta_j < 1$, and the innovations have a density that is continuous and positive on the whole line; see their Proposition 12.

As a corollary of the preceding discussion, we conclude that the theorems in Section 3 hold, in particular, if $X_t = r_t^2$ and the r_t follow a strictly stationary process that is strongly mixing with exponential rate and has finite $(4 + \delta)$th moment.

3. DETECTION ALGORITHMS

We now introduce four sequential monitoring methods considered in this paper. These methods are formulated in terms of the observations X_t, which can be interpreted as the squares of mean zero returns. With such an interpretation, the methods are designed to detect a change in unconditional variance of returns. Even though this has been our primary motivation, the proposed algorithms have a much wider applicability. In fact, the X_t can be viewed as a sequence of observations satisfying a week dependence condition (invariance principle (2.4)), and the monitoring is then intended to detect a change in the mean of the X_t.

The monitoring schemes considered in Section 3.1 are motivated by the methods proposed by Chu et al. (1996) in the context of detecting parameter changes in linear regression models. The two methods described in Section 3.2 are based on procedures proposed in Horváth et al. (2004), also in the context of linear regression models.

In the following discussion, we denote by $\hat{\sigma}$ an estimator of the parameter σ appearing in Theorem 2.1 that is consistent under the null. We use $\hat{\sigma} = \hat{\sigma}_m$, i.e., an estimator computed from the noncontaminated initial m observations, or $\hat{\sigma} = \hat{\sigma}_n$, i.e., an estimator that is sequentially updated as new observations arrive. Specific estimators and their asymptotic properties are discussed in Section 4. The use of the estimator $\hat{\sigma}_m$ is justified asymptotically, but no corresponding theory is yet available for the estimator $\hat{\sigma}_n$, which is however recommended, as it gives much better empirical results; see Section 5.

3.1. CUSUM and Fluctuation Monitoring Schemes

Cumulative sum (CUSUM) monitoring is based on the detector

$$S_n = \sum_{i=2}^{n} \nu_i^{-1/2}(X_i - \bar{X}_{i-1}), \qquad n \geq 2, \tag{3.1}$$

where

$$\nu_i = \frac{i}{i-1}, \qquad n \geq 2 \tag{3.2}$$

and

$$\bar{X}_{i-1} = \frac{1}{i-1}\sum_{j=1}^{i-1} X_j, \qquad i \geq 2. \tag{3.3}$$

(The weights ν_i are chosen so that $\mathrm{Var}[S_n] = (n-1)\mathrm{Var}[X_1]$.)

The conditions imposed on the boundary function $g(\cdot)$ appearing in Theorems 3.1 and 3.2, which follow, are collected in the following assumption.

Assumption 3.1. The function $g : [1,\infty) \to R$ satisfies

$$g(\cdot) \text{ is continuous on } [1,\infty); \tag{3.4}$$

$$g(t) > ct^{1/2}, \qquad 1 \leq t < \infty, \quad \text{for some } c > 0; \tag{3.5}$$

and

$$\limsup_{t\to\infty} \frac{|W(t)|}{g(t)} < \infty \quad a.s. \tag{3.6}$$

Sufficient conditions for (3.6) to hold can be obtained from the results formulated in Section 4.1 of Csörgő and Horváth (1993). For convenience, we state them in the following proposition.

PROPOSITION 3.1. *Suppose $g : [0,\infty) \to R$ satisfies the following conditions:*

$$\inf_{1\leq t\leq c} g(t) > 0 \quad \text{for all } c > 1, \tag{3.7}$$

$$g(t)/t \quad \text{is nonincreasing on } [c,\infty) \quad \text{for some } c > 0. \tag{3.8}$$

Then

$$\int_1^\infty \frac{1}{t}\exp(-cg^2(t)/t)\,dt < \infty \quad \text{for some } c > 0 \tag{3.9}$$

and (3.6) are equivalent.

The assumptions we impose on the function $g(\cdot)$ are different from those used by Chu et al. (1996), who required that $g(\cdot)$ be regular (in the sense defined in Chu et al., 1996, p. 1050), and $t^{-1/2}g(t)$ be eventually nondecreasing. Assumption (3.4) implies regularity whereas, in the class of eventually positive functions, if $t^{-1/2}g(t)$ is eventually nondecreasing, then assumption (3.5) holds. We point out that (3.6) (and hence (3.9)) is equivalent with the existence of $\sup_{1\le t<\infty}|W(t)|/g(t)$, which appears as the limit in Theorems 3.1 and 3.2. Also, the discussion in Csörgő and Horváth (1993, pp. 190–195) shows that (3.8) is weaker than assuming that $t^{-1/2}g(t)$ is eventually nondecreasing.

Proposition 3.1 is an integral test that provides an analytic expression for functions satisfying (3.6). If a function $g(\cdot)$ satisfies (3.6), it is called an upper class function. For discussion of upper class functions of the Wiener process we refer to Itô and McKean (1965, pp. 33–36) and Csörgő and Horváth (1993).

In the empirical applications, we will work with the function

$$g_a(t) = [t(a^2 + \ln t)]^{1/2}, \qquad t \ge 1, \tag{3.10}$$

which meets Assumption 3.1. In (3.10) and throughout the paper, ln denotes the natural logarithm.

Conditions (3.4), (3.5), (3.7), and (3.8) are obviously satisfied by the function $g_a(t)$. Because

$$\int_1^\infty \frac{1}{t}\exp(-cg_a^2(t)/t)\,dt = \int_1^\infty e^{-ca^2}\frac{1}{t^{1+c}}\,dt,$$

assumption (3.9) is also met.

It is known that (see Chu et al., 1996, p. 1052)

$$P\{|W(t)| \ge g_a(t), \text{ for some } t \ge 1\} = 2[1 - \Phi(a) + a\phi(a)], \tag{3.11}$$

where Φ and ϕ are, respectively, the cumulative distribution function (c.d.f.) and probability density function (p.d.f.) of a standard normal random variable. For boundary crossing probabilities (3.11) of 5% and 10%, a^2 equals 7.78 and 6.25, respectively.

Theorem 3.1, which is proved in Appendix A, justifies a monitoring scheme based on the detector (3.1).

THEOREM 3.1. *Suppose assumption (2.4) holds and the function $g(\cdot)$ satisfies Assumption 3.1. Then, under H_0,*

$$\sup_{m\le n<\infty}\frac{|S_n|}{m^{1/2}g(n/m)} \xrightarrow{d} \sup_{1\le t<\infty}\frac{\sigma|W(t)|}{g(t)}. \tag{3.12}$$

Using the function $g_a(\cdot)$ in Theorem 3.1, we obtain the following rejection rule: reject H_0 if

$$\frac{1}{\hat\sigma}\left|\sum_{i=2}^{n}\nu_i^{-1/2}(X_i - \bar X_{i-1})\right| > m^{1/2}\left[\frac{n}{m}\left(a^2 + \ln\frac{n}{m}\right)\right]^{1/2}, \quad \text{for some } n > m. \tag{3.13}$$

By (3.12) and (3.11), as $m \to \infty$, the probability of falsely rejecting H_0 thus tends to a prescribed significance level α that is controlled by the constant a.

We now turn to the fluctuation detector defined as

$$Z_n = n\sigma^{-1}(\bar X_n - \bar X_m), \qquad n > m, \tag{3.14}$$

with $\bar X_n$ as defined in (3.3).

The following theorem provides a justification for the monitoring scheme based on the fluctuation detector (3.14).

THEOREM 3.2. *Suppose assumption (2.4) holds and the function $g(\cdot)$ satisfies Assumption 3.1. Then, under H_0,*

$$\sup_{m<n<\infty}\frac{|Z_n|}{m^{1/2}((n-m)/m)g(n/(n-m))} \xrightarrow{d} \sup_{1<t<\infty}\frac{|W(t)|}{g(t)}.$$

Theorem 3.2 is proved in Appendix A.

Similarly as in CUSUM monitoring, Theorem 3.2 leads to the following rejection rule: reject H_0 if for some $n > m$

$$\frac{n}{\hat\sigma}|\bar X_n - \bar X_m| > m^{1/2}\left(\frac{n-m}{m}\right)\left[\frac{n}{n-m}\left(a^2 + \ln\frac{n}{n-m}\right)\right]^{1/2}. \tag{3.15}$$

3.2. Monitoring Schemes Based on Partial Sums of Residuals and Recursive Residuals

The following two kinds of residuals are used to construct detectors:

$$\hat\xi_i = X_i - \bar X_m, \qquad i > m \tag{3.16}$$

and

$$\tilde\xi_i = X_i - \bar X_{i-1}, \qquad i \ge 2. \tag{3.17}$$

First, define the detector

$$\hat S(m,k) = \sum_{i=m+1}^{m+k}\hat\xi_i, \qquad k \ge 1 \tag{3.18}$$

and consider the boundary function

$$g_\gamma(m,k) = \left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma, \qquad 0 \le \gamma < \frac{1}{2} - \alpha. \tag{3.19}$$

Define the critical value $c_\alpha(\gamma)$ by

$$P\left\{\sup_{0\le t\le 1}\frac{|W(t)|}{t^\gamma} \ge c_\alpha(\gamma)\right\} = \alpha.$$

The critical values $c_\alpha(\gamma)$ are tabulated in Table 1 of Horváth et al. (2004).

THEOREM 3.3. *If (2.4) holds, then under* H_0

$$\lim_{m\to\infty} P\left\{\sup_{1\le k<\infty}\frac{|\hat S(m,k)|}{\sqrt{m}g_\gamma(m,k)} \ge c_\alpha(\gamma)\right\} = P\left\{\sup_{0\le t\le 1}\frac{\sigma|W(t)|}{t^\gamma} \ge c_\alpha(\gamma)\right\}.$$

Theorem 3.3, which is proved in Appendix B, leads to the following rejection rule: reject H_0 if

$$\frac{1}{\hat\sigma}\left|\sum_{i=m+1}^{m+k}(X_i - \bar X_m)\right| \ge c_\alpha(\gamma)m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma, \quad \text{for some } k \ge 1, \tag{3.20}$$

or, by letting $n = m + k$, reject if

$$\frac{1}{\hat\sigma}\left|\sum_{i=m+1}^{n}(X_i - \bar X_m)\right| \ge c_\alpha(\gamma)m^{1/2}\left(\frac{n}{m}\right)\left(\frac{n-m}{n}\right)^\gamma, \quad \text{for some } n > m. \tag{3.21}$$

Finally, define the detector

$$\tilde S(m,k) = \sum_{i=m+1}^{m+k}\tilde\xi_i, \qquad k \ge 1. \tag{3.22}$$

We now denote the boundary function by $h(t)$, $t \ge 0$, and assume that $h(t) = g(t+1)$ for a function $g(u)$, $u \ge 1$ that satisfies Assumption 3.1. Thus $h(\cdot)$ satisfies the following assumption.

Assumption 3.2. The function $h(\cdot)$ satisfies the following conditions:

$h(\cdot)$ is continuous on $[0,\infty)$; (3.23)

$h(t) > c(t+1)^{1/2}, \qquad 0 \le t < \infty, \quad$ for some $c > 0$ (3.24)

and

$$\limsup_{t\to\infty}\frac{|W(t)|}{h(t)} < \infty \quad a.s. \tag{3.25}$$

Similarly to Proposition 3.1, if (3.23) and (3.24) are satisfied, then (3.25) holds if and only if

$$\int_0^\infty \frac{1}{t}\exp\left(\frac{-ch^2(t)}{t}\,dt\right) < \infty \quad \text{for some } c > 0.$$

THEOREM 3.4. *Suppose assumption (2.4) holds and the function* $h(\cdot)$ *satisfies Assumption 3.2. Then, under* H_0,

$$\sup_{1\le k<\infty}\frac{|\tilde S(m,k)|}{m^{1/2}h(k/m)} \xrightarrow{d} \sup_{0<t<\infty}\frac{\sigma|W(t)|}{h(t)}.$$

Theorem 3.4 is proved in Appendix B.

The discussion in Section 3.1 shows that the boundary function

$$h_a(t) = (t+1)^{1/2}[a^2 + \ln(t+1)]^{1/2} \tag{3.26}$$

satisfies Assumption 3.2. It is known that (see Chu et al., 1996, eqn. (8))

$$P\left\{\sup_{0<t<\infty}\frac{|W(t)|}{h_a(t)} \ge 1\right\} = \exp(-a^2/2). \tag{3.27}$$

For the asymptotic false alarm rate (3.27) of 5% and 10%, a^2 equals 6.0 and 4.6, respectively.

Theorem 3.4 and (3.27) lead to the rejection rule: reject H_0 if for some $k \ge 1$

$$\frac{1}{\hat\sigma}\left|\sum_{i=m+1}^{m+k}(X_i - \bar X_{i-1})\right| \ge \sqrt{m}\left(1+\frac{k}{m}\right)^{1/2}\left(a^2 + \ln\left(1+\frac{k}{m}\right)\right)^{1/2} \tag{3.28}$$

or, by letting $n = m + k$, reject if

$$\frac{1}{\hat\sigma}\left|\sum_{i=m+1}^{n}(X_i - \bar X_{i-1})\right| \ge \sqrt{m}\left(\frac{n}{m}\right)^{1/2}\left(a^2 + \ln\left(\frac{n}{m}\right)\right)^{1/2}, \quad \text{for some } n > m. \tag{3.29}$$

4. ESTIMATION OF THE ASYMPTOTIC VARIANCE

The on-line change-point detection procedures discussed in this paper rely on the consistent estimation of the asymptotic variance σ^2 given by

$$\sigma^2 = \sum_{l=-\infty}^{+\infty} \gamma_l, \qquad \gamma_l = \mathrm{Cov}(X_k, X_{k+l}).$$

This estimation problem has been extensively studied; see Bartlett (1950), Grenander and Rosenblatt (1957), Parzen (1957), Anderson (1971), Andrews (1991), Andrews and Monahan (1992), and Haan and Levin (1997).

Because σ^2 is the value of the spectral density at 0, we focus on estimators of the form

$$\hat{\sigma}_n^2 = \frac{n}{n-1} \sum_{l=-n+1}^{n-1} k\left(\frac{l}{W_m}\right) \hat{\gamma}_l(n), \qquad n \geq m, \tag{4.1}$$

where $k(\cdot)$ is a real-valued kernel, W_m is the bandwidth parameter, and $\hat{\gamma}_l(n)$ is the σth sample autocovariance of $\{X_t\}_1^n$.

We also discuss the results obtained by using the estimator

$$\hat{\sigma}_m^2 = \frac{m}{m-1} \sum_{l=-m+1}^{m-1} k\left(\frac{l}{W_m}\right) \hat{\gamma}_l(m). \tag{4.2}$$

We emphasize that the optimal bandwidth is obtained using the fixed initial m data points, and so it does not change as we proceed to monitor the data, whereas the sample autocovariance function $\hat{\gamma}_l$ is either sequentially updated in $\hat{\sigma}_n$ or computed only once in $\hat{\sigma}_m$.

We restrict ourselves to the following two kernel functions:

(1) Bartlett: $$k_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

(2) Quadratic Spectral (QS): $$k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

According to Andrews (1991), the optimal bandwidths W_m^* for the two kernels are

$$\text{Bartlett: } W_m^* = 1.1447(a(1)m)^{1/3}, \tag{4.3}$$

$$\text{QS: } W_m^* = 1.3221(a(2)m)^{1/5}, \tag{4.4}$$

where $a(i)$, $i = 1$ or 2, is a function of the unknown spectral density function $f(\lambda)$ of the process $\{X_t\}$. In the applications and simulations discussed in the following material, we assume that the returns follow a GARCH(1,1) model

before a change has occurred. We now explain how the constants $a(i)$ can be found under this assumption.

Assume then that $X_t = r_t^2$ and

$$r_t = h_t Z_t, \tag{4.5}$$

where $\{Z_t\}$ is an i.i.d. sequence with zero mean and unit variance and h_t evolves according to

$$h_t^2 = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}^2. \tag{4.6}$$

Following Hamilton (1994, pp. 665–666), we obtain

$$X_t = \omega + (\alpha + \beta)X_{t-1} + \nu_t - \beta\nu_{t-1}, \tag{4.7}$$

where $\nu_t \equiv X_t^2 - h_t^2$ is a white noise, i.e., a second-order stationary sequence of uncorrelated random variables. Expression (4.7) will then be recognized as an ARMA(1,1) process for X_t, in which the autoregressive coefficient is $\alpha + \beta$ and the moving average coefficient is $-\beta$.

For ARMA(1,1) models with autoregressive parameter ρ and moving average parameter ψ, estimates of $a(1)$ and $a(2)$ in (4.3) and (4.4) are given, respectively, in equations (6.6) and (6.5) in Andrews (1991). These equations involve an integer parameter p, which we set equal to 1. We thus obtain

$$\hat{a}(1) = \frac{4(1 + \hat{\rho}\hat{\psi})^2(\hat{\rho} + \hat{\psi})^2}{(1 - \hat{\rho}^2)^2(1 + \hat{\psi})^4} \tag{4.8}$$

and

$$\hat{a}(2) = \frac{4(1 + \hat{\rho}\hat{\psi})^2(\hat{\rho} + \hat{\psi})^2}{(1 - \hat{\rho})^4(1 + \hat{\psi})^4}, \tag{4.9}$$

where $\hat{\rho}$ and $\hat{\psi}$ are appropriate estimates. In this paper, using the m historical observations, we compute the quasi-maximum likelihood estimates (QMLE) $\hat{\alpha}$ and $\hat{\beta}$ of the GARCH(1,1) model and set $\hat{\rho} = \hat{\alpha} + \hat{\beta}$ and $\hat{\psi} = -\hat{\beta}$.

We conclude this section by discussing the asymptotic properties of the estimators $\hat{\sigma}_n^2$ and $\hat{\sigma}_m^2$ introduced previously.

The theory underlying the use of the estimator $\hat{\sigma}_m^2$ is fully developed. Theorem 3.1(i) of Giraitis, Kokoszka, Leipus, and Teyssière (2003) implies that if W_m is a deterministic function such that $W_m \to \infty$ and $W_m/m \to 0$, then

$$\hat{\sigma}_m^2 \xrightarrow{P} \sigma^2, \quad \text{as } m \to \infty \tag{4.10}$$

under both the null and the alternative. Relation (4.10) also holds for random bandwidths (4.3) and (4.4); see Theorem 3(a) of Andrews (1991).

The theory underlying the use of the estimator $\hat{\sigma}_n^2$ is not fully developed yet. Berkes, Horváth, Kokoszka, and Shao (2005, 2006) considered the estimator

$$\hat{s}_n^2 = \frac{n}{n-1} \sum_{l=-n+1}^{n-1} k\left(\frac{l}{W_n}\right) \hat{\gamma}_l(n), \qquad n \geq 1, \tag{4.11}$$

with a deterministic bandwidth W_n rather than W_m as in (4.1). Under the null hypothesis of no change in the parameters, $\hat{s}_n^2 \xrightarrow{a.s.} \sigma^2$, as $n \to \infty$, by Theorem 1.1(i) of Berkes et al. (2005). This almost sure convergence implies that

$$\sup_{m<n<\infty} |\hat{s}_n^2 - \sigma^2| \xrightarrow{a.s.} 0, \quad \text{as } m \to \infty. \tag{4.12}$$

It can be expected that for deterministic bandwidth W_m relation (4.12) continues to hold with $\hat{s}_n^2$ replaced by $\hat{\sigma}_n^2$, but this has not been verified yet. A much more difficult problem is to show that

$$\sup_{m<n<\infty} |\hat{\sigma}_n^2 - \sigma^2| \xrightarrow{P} 0, \quad \text{as } m \to \infty \tag{4.13}$$

for random W_m as in (4.3) and (4.4).

Relation (4.13) is needed to asymptotically justify the use of the estimator $\hat{\sigma}_n^2$. However, even though the proof of (4.13) is not available yet, we recommend using $\hat{\sigma}_n^2$ as it gives much better empirical results than $\hat{\sigma}_m^2$.

5. SIMULATION STUDY

The objective of this section is to compare the finite-sample performance of the four monitoring schemes introduced in Section 3. We highlight only the most important findings; detailed simulation results are presented in Zhang (2005).

For ease of reference, the following table lists the schemes and the abbreviations that will be used in the discussion that follows.

Scheme	Rejection rule	Abbreviation
CUSUM	(3.13) (Theorem 3.1)	CS
Fluctuation	(3.15) (Theorem 3.2)	FL
Partial sum of residuals	(3.21) (Theorem 3.3)	PS
Partial sum of recursive residuals	(3.29) (Theorem 3.4)	RR

Recall that the boundary function $g_a(t) = [t(a^2 + \ln t)]^{1/2}$, $t \geq 1$ is used for both CS and FL monitoring and, for asymptotic controlled sizes of 5% and 10%, a^2 equals 7.78 and 6.25, respectively. The rejection condition for the method PS involves $c_\alpha(\gamma)$; see (3.21). If $\gamma = 0.25$, the value used in our simulations,

critical value of $c_\alpha(\gamma) = 2.386$ (2.106) gives 5% (10%) asymptotic false alarm rate. The boundary function is $h_a(t) = (t+1)^{1/2}[a^2 + \ln(t+1)]^{1/2}$ in the RR method, setting a^2 equal to 6.0 (4.6) gives α of 5% (10%) asymptotically.

We consider the two kernels, Bartlett and QS, for each of the four schemes. We use autocovariances computed either from all observations available up to the current time n or from the initial m observations. We focus mainly on the former way of computing the autocovariances because it will be seen to give better results.

We evaluate the four methods and their several variants on time series of squares of simulated GARCH(1,1) processes with mean zero. In the following discussion, the observations X_t are said to follow Model i, $i = 1,2,3,4$, if $X_t = r_t^2$ and the r_t follow (4.5) and (4.6) with standard normal Z_t and the parameters displayed in Table 1 (with $\mu = 0$). The pair of Models 1 and 2 reflects a possible change point in the Dow Jones index, and the pair of Models 3 and 4 reflects a possible change point in the NASDAQ index. A detailed justification for the choice of these change-point models is presented in Zhang (2005). Here we simply view them as examples of potential typical change points that can be encountered in econometric practice.

The primary concern regarding the performance of the four sequential monitoring procedures is the false alarm rate α, namely, the probability of falsely rejecting a true null hypothesis of no change in variance. To evaluate this probability, we generate GARCH(1,1) time series according to Models 1 and 3, the two typical prechange specifications in Table 1. With historical sample sizes $m = 100,200,300,400,500$, we begin to monitor the squared process from the $(m+1)$th observation. The monitoring horizon q is set to be one, two, and three times m. We replicate the monitoring procedures on a given series of length $(q+1)m$ for a large number of times, $R = 1{,}000$ in our simulations. The empirical sizes can then be computed by dividing by R the number of times a boundary is crossed. Theoretically, they should become close to the asymptotic size when m and q approach infinity.

Table 2 presents the empirical sizes for the four algorithms when $m = 500$, with autocovariances sequentially computed. We first notice that the sizes produced by methods CS and RR are below the target levels. Almost all overrejections are in the cells of methods FL and PS, and those of FL are more severe: the worst is 9.9% (13.7%) for the controlled size of 5% (10%) and they equal or surpass the nominal levels even when the monitoring horizon is only one m long, which equals 500 in Table 2. The four methods give nearly equal performance with respect to the two different models. The empirical sizes, however, do depend on model specification to some degree. Our simulations, whose details are not reported here, show that, when the sum of α and β is smaller and not as close to 1 as in Models 1 and 3, the problem of overrejection is much less severe even for the FL scheme. This is probably due to the fact that the greater $\alpha + \beta$ is, the less accurate estimation of the variance of squared GARCH(1,1) observations we can get.

TABLE 1. Change-point models of the return data

Stock index	Model	Fitting period	Number of observations	μ	ω	α	β	Sample variance	Implied variance
DJIA	Model 1	(1/1/92–12/31/95)	1,010	0.070	0.028	0.064	0.869	0.409	0.416
	Model 2	(1/1/96–12/31/99)	1,011	0.081	0.082	0.092	0.853	1.427	1.495
NASDAQ	Model 3	(7/1/94–6/30/97)	757	0.130	0.131	0.136	0.749	1.140	1.139
	Model 4	(7/1/97–12/31/98)	382	0.324	0.534	0.186	0.661	3.431	3.492

TABLE 2. Empirical sizes (in percentages) for the four monitoring methods applied to simulated series of squared GARCH(1,1) observations following Models 1 and 3 in Table 1

			Model 1				Model 3			
			Bartlett		QS		Bartlett		QS	
Method	q		5%	10%	5%	10%	5%	10%	5%	10%
CS	m		2.8	5.1	2.4	4.0	2.8	5.7	4.3	6.3
	$2m$		3.8	7.0	3.5	6.2	3.6	7.2	4.9	8.0
	$3m$		4.7	8.0	4.1	7.5	4.3	7.9	5.5	8.3
FL	m		6.9	10.2	5.0	7.8	8.6	10.6	6.2	9.8
	$2m$		8.2	12.3	6.3	10.4	9.3	11.6	6.8	11.1
	$3m$		9.1	13.7	6.7	11.6	9.9	12.3	7.4	11.8
PS	m		3.6	6.5	2.3	5.1	2.9	5.8	2.6	5.5
	$2m$		6.0	10.5	5.1	9.1	4.5	9.0	3.9	8.4
	$3m$		7.3	13.0	6.8	11.1	5.6	10.8	4.5	10.8
RR	m		0.9	1.6	0.4	1.2	0.2	0.8	0.2	1.2
	$2m$		2.0	4.2	1.4	3.2	1.2	2.4	0.8	2.4
	$3m$		3.1	6.1	2.6	4.7	1.6	3.9	1.0	3.1

Note: The number of replications is $R = 1{,}000$ with historical size $m = 500$ and autocovariances computed sequentially.

To save space, we do not present tables with the empirical sizes for the other four values of m, 100, 200, 300, and 400, considered in our study. To illustrate the overall conclusions that can be drawn from these tables, we present a representative graphical comparison in Figure 1. We focus on the results for Model 1 with controlled size 10% and $q = 3$. Plots of other cases support our overall conclusions. Figure 1 supports the observation made earlier that methods CS and RR are more conservative than methods PS and FL. In addition, empirical size basically decreases as we extend the historical sample size m. This is especially true for methods FL and RR with the Bartlett window, whose sizes fall monotonically as m increases from 100 to 500. When the Bartlett kernel is used, it seems that $m = 200$ is long enough for methods CS and RR to secure a size close to the nominal level of 10%, whereas for methods PS and FL even $m = 500$ can only yield sizes that are about 14%. As for the QS kernel, the sizes decrease consistently when m increases from 100 to 300 for all four algorithms, and they roughly level off for longer historical samples, meaning that extending m beyond 300 does not appreciably improve the performance. As we are more concerned about overrejections, the QS window is recommended, because it works better for methods PS or FL and brings their empirical sizes down to less than 12%, as opposed to 16% for the Bartlett window, when $m = 300$. Overall, however, the difference between the two windows is not large.

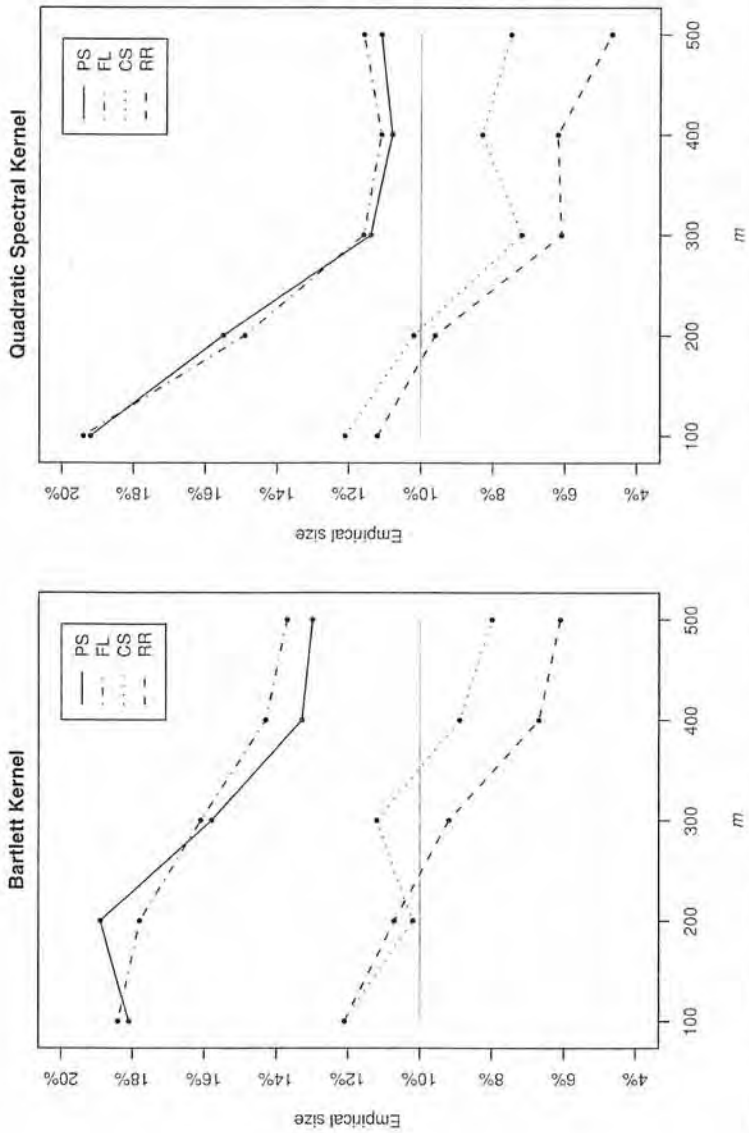

FIGURE 1. Comparison of empirical sizes. Results are reported for Model 1 with either Bartlett kernel or QS kernel. The monitoring horizon is $q = 3$. The straight line marks the target size of 10%.

We next examine the power of the tests. We consider two typical variance changes that are represented by the passages from Model 1 to Model 2 and from Model 3 to Model 4. Focusing again on the five values of m equal to 100, 200, 300, 400, and 500, we generate the data and let the model transitions happen at $t = 1.1m + 1$. The monitoring starts from the $(m + 1)$th observation, but the monitoring horizon q is fixed at 500, instead of varying with m. The empirical power of the tests is the percentage of rejections in R replications. We used $R = 1{,}000$ in our simulations. A commonly used criterion of evaluating sequential procedures is the average run length (ARL) defined as the average of detection delays in the presence of a real break. Empirical ARL can be computed by subtracting the time of real break from the average of the alarm times in the R replications. To save space, we only report in Table 3 the results for $m =$ 100, 300, 500 with controlled size 10%. The empirical ARL is shown in parentheses next to the power.

The general relation between size and power in hypothesis testing is that a test with a smaller size, i.e. a lower probability of type I error, tends to have a higher probability of type II error, i.e. lower power. With this in mind, it is not surprising to see that methods CS and RR have less power than methods FL and PS, because, as shown in Table 2 and Figure 1, their sizes are smaller and do not suffer from the problem of overrejection. This difference is most obvious when $m = 100$ and becomes negligible when $m \geq 300$ and the transition is from Model 1 to Model 2, in which case the empirical powers are all more than

TABLE 3. Empirical power and ARL of the four methods applied to the two types of variance changes

			Model 1 → Model 2				Model 3 → Model 4			
			Bartlett		QS		Bartlett		QS	
Method	m	Break	Power	(ARL)	Power	(ARL)	Power	(ARL)	Power	(ARL)
CS	100	111	94.3	(89)	91.7	(103)	69.8	(116)	61.3	(121)
	300	331	98.9	(102)	98.1	(117)	86.4	(155)	81.9	(169)
	500	551	98.5	(115)	98.7	(129)	90.0	(189)	87.8	(193)
FL	100	111	98.1	(38)	97.6	(43)	83.2	(42)	79.4	(47)
	300	331	99.9	(36)	100	(38)	97.4	(47)	96.7	(57)
	500	551	100	(37)	100	(39)	99.1	(53)	98.7	(60)
PS	100	111	99.0	(40)	99.2	(46)	88.9	(48)	87.1	(59)
	300	331	100	(46)	100	(49)	98.2	(64)	98.0	(77)
	500	551	100	(51)	100	(54)	99.3	(74)	98.6	(81)
RR	100	111	97.8	(76)	96.2	(83)	79.8	(98)	74.7	(110)
	300	331	99.7	(88)	99.8	(99)	93.2	(139)	92.2	(160)
	500	551	99.9	(105)	100	(113)	96.6	(166)	95.4	(175)

Note: The number of replications is $R = 1{,}000$, with controlled size of 10%, and $\hat{\gamma}_l$ computed sequentially. The monitoring horizon is $q = 500$.

98%. The powers corresponding to the transition from Model 3 to Model 4 are generally lower, with longer ARL. This is may be because the variance increases about 3.5 times in the transition from Model 1 to Model 2 and about 3 times in the transition from Model 3 to Model 4. The difference is however practically negligible for methods FL and PS with $m \geq 300$. An appealing property of the tests is that greater power comes with shorter ARL. In particular, the average detection delays of methods FL and PS, the two tests with greater power, are around 40 and less than 55 for the first type of change, but are more than double for the other two algorithms. As regards the choice of kernel function, the Bartlet kernel gives higher power and shorter ARL than the QS kernel. Overall, we may expect to get higher power by using more observations as a historical sample, but the gains, in terms of both power and ARL, of extending m beyond 300 are not significant.

Using box plots, we present the distributions of the first hitting time in Figure 2 with $m = 300$. The plots for methods FL and PS are almost identical, except that the former has a slightly smaller median and interquartile range (the difference between third and first quartiles). The distributions for methods CS and RR are clearly more spread out. All distributions have elongated upper tails, i.e., are positively skewed. For the transition from Model 3 to Model 4, all summary statistics, including first quartile, median, third quartile, and interquartile range, increase, meaning that all methods are less effective in detecting this transition. Hitting time distributions for $m = 100$ and $m = 500$ have similar shapes. Perhaps somewhat counterintuitively, increasing m causes the center of the distributions to move to the right (longer delay time).

To investigate the effect of the location of the break point on the power and the distribution of the detection time, we run simulations with the structural change occurring at $1.2m + 1$, with the length of the monitoring horizon remaining equal to 500. There are barely any differences in terms of the power of the tests. The ARL, however, increases; it prolongs by around 20 for the transition from Model 1 to Model 2 and by about 30 for the other transition. The increases of the order statistics, such as median and the quartiles, are small.

We now discuss the effect of using covariances computed from the initial m observations rather than from the observations up to the current time n. The simulation results show that all four monitoring procedures suffer from the problem of overrejection; their relative performance, however, does not change. The methods CS and RR are still more conservative, producing smaller false alarm rates. Nearly all sizes are at least doubled; some are even more than ten times greater than those in Table 2. For the monitoring horizon $q = 3m$, with $m = 500$, method FL gives, as observed before, the worst results: 20.3% (23.9%) for nominal size 5% (10%). The power of the tests exceeds 97% in all cases, with much shorter ARL, basically halved, than those in Table 3. Because methods with covariances computed sequentially also have, however, satisfactory power, we recommend this way of computing the autocovariances.

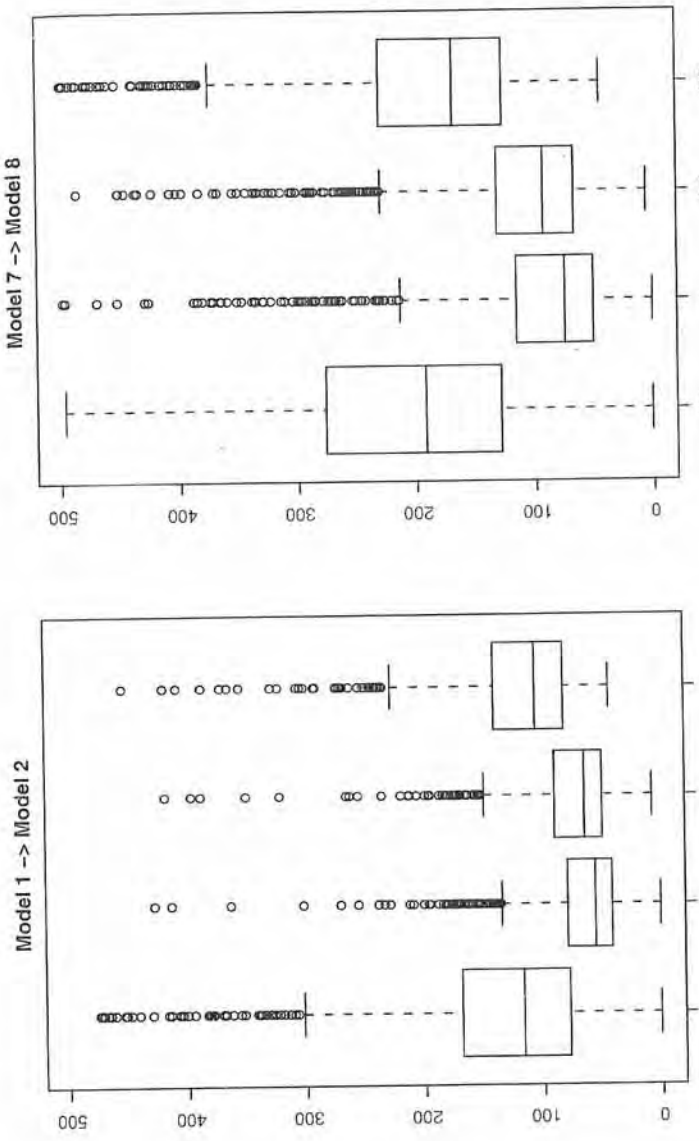

FIGURE 2. Comparison of the distributions of the first hitting time. Results are reported for $m = 300$ with break at $n = 331$, 10% controlled size, and the Bartlett kernel. The number of replications is $R = 1{,}000$. The historical sample size $m = 300$ is subtracted from the hitting times for ease of visual comparison.

In conclusion, we can say that if the primary concern is to control the false alarm rate under the nominal level, the methods CS and RR with sequentially computed covariances and the QS kernel are recommended. The choice of the algorithm is especially important for $m = 100$ and $m = 200$. The false alarm rates of methods PS and FL are much higher, but if we use the QS kernel and historical samples longer than 300, the problem becomes less severe. Generally, the QS kernel gives somewhat more conservative sizes than the Bartlett kernel. Methods PS and FL have only somewhat higher power than the other two methods, but the detection delay of methods CS and RR is basically two times larger than that of methods PS and FL.

REFERENCES

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. Wiley.

Andreou, E. & E. Ghysels (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600.

Andreou, E. & E. Ghysels (2003) Tests for breaks in the conditional co-movements of asset returns. *Statistica Sinica* 13, 1045–1073.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Bartlett, M.S. (1950) Periodogram analysis of continuous spectra. *Biometrika* 37, 1–16.

Berkes, I., E. Gombay, L. Horváth, & P. S. Kokoszka (2004) Sequential change-point detection in GARCH(p,q) models. *Econometric Theory* 20, 1140–1167.

Berkes, I., L. Horváth, P. Kokoszka, & Q.-M. Shao (2005) Almost sure convergence of the Bartlett estimator. *Periodica Mathematica Hungarica* 51, 11–25.

Berkes, I., L. Horváth, P. Kokoszka, & Q.-M. Shao (2006) On discriminating between long-range dependence and changes in mean. *Annals of Statistics* 34(3), 2395–2422.

Bradley, R.C. (1986) Basic properties of strong mixing conditions. In E. Eberlein & M.S. Taqqu (eds.), *Dependence in Probability and Statistics*, pp. 165–192. Boston: Birkhaüser.

Carasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chu, C.-S.J. (1995) Detecting parameter shift in GARCH models. *Econometric Reviews* 14, 241–266.

Chu, C.-S.J., M. Stinchcombe, & H. White (1996) Monitoring structural change. *Econometrica* 64, 1045–1065.

Csörgő, M. & L. Horváth (1993) *Weighted Approximations in Probability and Statistics*. Wiley.

Csörgő, M. & P. Révész (1981) *Strong Approximations in Probability and Statistics*. Academic Press.

Giraitis, L., P.S. Kokoszka, R. Leipus, & G. Teyssière (2003) Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 112, 265–294.

Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. Wiley.

Haan, W.J. Den & A. Levin (1997) A practitioner's guide to robust covariance matrix estimation. In C.R. Rao & G.S. Maddala (eds.), *Handbook of Statistics*, vol. 15, pp. 291–341.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

He, C. & T. Teräsvirta (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.

Horváth, L., M. Hušková, P. Kokoszka, & J. Steinebach (2004) Monitoring changes in linear models. *Journal of Statistical Planning and Inference* 126, 225–251.

Inclán, C. & G.C. Tiao (1994) Use of cumulative sums of squares for retrospective detection of change of variance. *Journal of the American Statistical Association* 89, 913–923.

Inoue, A. (2001) Testing for distributional change in time series. *Econometric Theory* 17, 156–187.

Itô, K. & H.P. McKean (1965) *Diffusion Processes and Their Sample Paths*. Springer.

Kim, S., S. Cho, & S. Lee (2000) On the CUSUM test for parameter changes in GARCH(1,1) model. *Communications in Statistics—Theory and Methodology* 29, 445–462.

Kokoszka, P.S. & R. Leipus (1998) Change-point in the mean of dependent observations. *Statistics and Probability Letters* 40, 385–393.

Kokoszka, P.S. & R. Leipus (1999) Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal* 39, 231–247.

Kokoszka, P.S. & R. Leipus (2000) Change-point estimation in ARCH models. *Bernoulli* 6, 513–539.

Kokoszka, P.S. & R. Leipus (2002) Detection and estimation of changes in regime. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 325–337. Boston: Birkhaüser.

Kulperger, R. & H. Yu (2005) High moment partial sum processes of residuals in GARCH models and their applications. *Annals of Statistics* 33, 2395–2422.

Mikosch, T. & C. Stărică (1999) Change of Structure in Financial Time Series, Long Range Dependence and the GARCH Model. Technical report, University of Groningen.

Mikosch, T. & C. Stărică (2002) Long-range dependence effects and ARCH modeling. In P. Doukhan, G. Oppenheim, & M.S. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 439–459. Boston: Birkhäuser.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.

Phillip, W. & W. Stout (1975) *Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables*. Memoirs of the American Mathematical Society 161.

Zhang, A. (2005) Estimation, testing and monitoring of GARCH models. Ph.D. dissertation, Utah State University.

APPENDIX A: Proofs of Theorems 3.1 and 3.2

The proof of Theorem 3.1 relies on several lemmas and Proposition A.1.

LEMMA A.1. *If (2.4) holds, then*

$$\sum_{1\le i\le n}(X_i-\omega_0)=O((n\log\log n)^{1/2}) \quad a.s.$$

Proof. The lemma follows from assumption (2.4) and the law of the iterated logarithm for the Wiener process. ■

LEMMA A.2. *If (2.4) holds, then*

$$|X_n|=o(n^{\alpha}) \quad a.s.$$

Proof. Observe that

$$\begin{aligned}|X_n| &\le |X_n-\omega_0|+|\omega_0|\\ &\le \left|\sum_{i=1}^{n}(X_i-\omega_0)-\sigma W(n)\right|+\left|\sum_{i=1}^{n-1}(X_i-\omega_0)-\sigma W(n-1)\right|\\ &\quad +\sigma|W(n)-W(n-1)|+|\omega_0|\\ &\overset{a.s.}{=} o(n^{\alpha})+\sigma|W(n)-W(n-1)|.\end{aligned}$$

Because $W(n) - W(n-1)$ is a standard normal random variable,

$$P[|W(n) - W(n-1)| \geq (10 \log n)^{1/2}] = O(n^{-2}),$$

so by the Borel–Cantelli lemma

$$|W(n) - W(n-1)| \overset{a.s.}{=} O((\log n)^{1/2}).$$

This completes the proof of Lemma A.2. ∎

We will use the decomposition

$$S_n = \sum_{2 \leq i \leq n} (X_i - \bar{X}_{i-1}) + \sum_{2 \leq i \leq n} \left[\left(1 - \frac{1}{i}\right)^{1/2} - 1 \right] (X_i - \bar{X}_{i-1}). \tag{A.1}$$

LEMMA A.3. *If (2.4) holds, (A.1) satisfies*

$$\sum_{2 \leq i \leq n} \left[\left(1 - \frac{1}{i}\right)^{1/2} - 1 \right] (X_i - \bar{X}_{i-1}) \overset{a.s.}{=} o(n^{\alpha}).$$

Proof. By the mean value theorem, we have

$$\begin{aligned} &\left| \sum_{2 \leq i \leq n} \left[\left(1 - \frac{1}{i}\right)^{1/2} - 1 \right] (X_i - \bar{X}_{i-1}) \right| \\ &\quad = \left| \sum_{2 \leq i \leq n} \left[\left(1 - \frac{1}{i}\right)^{1/2} - 1 \right] (X_i - \omega_0 - (\bar{X}_{i-1} - \omega_0)) \right| \\ &\quad \leq C \left[\sum_{2 \leq i \leq n} \frac{1}{i} |X_i - \omega_0| + \sum_{2 \leq i \leq n} \frac{1}{i} |\bar{X}_{i-1} - \omega_0| \right]. \end{aligned}$$

By Lemma A.2,

$$\sum_{2 \leq i \leq n} \frac{1}{i} |X_i - \omega_0| = o(n^{\alpha}).$$

By Lemma A.1,

$$|\bar{X}_{i-1} - \omega_0| \overset{a.s.}{=} O\left(\frac{\log \log i}{i} \right)^{1/2},$$

and so

$$\sum_{2 \leq i \leq n} \frac{1}{i} |\bar{X}_{i-1} - \omega_0| \overset{a.s.}{=} O(1) \sum_{2 \leq i \leq n} \left(\frac{\log \log i}{i^3} \right)^{1/2} = O(1).$$

This completes the proof of Lemma A.3. ∎

Introduce the process

$$W^*(t) = W(t) - \int_1^{t+1} \frac{1}{x} W(x)\, dx, \qquad t \geq 0. \tag{A.2}$$

Computing the covariances shows that $W^*(\cdot)$ is a standard Wiener process.

PROPOSITION A.1. *If condition (2.4) holds, then*

$$S_n - \sigma W^*(n) \overset{a.s.}{=} o(n^{\alpha}).$$

Proof. Using decomposition (A.1) and Lemma A.3, we observe that

$$S_n \overset{a.s.}{=} \sum_{2 \leq i \leq n} (X_i - \omega_0) - \sum_{2 \leq i \leq n} \frac{1}{i-1} \sum_{1 \leq j \leq i-1} (X_j - \omega_0) + o(n^{\alpha}).$$

Note that

$$\sum_{2 \leq i \leq n} \frac{1}{i-1} \sum_{1 \leq j \leq i-1} (X_j - \omega_0) = \int_1^n \frac{1}{[x]} U(x)\, dx,$$

where $[x]$ is the integer part of x and

$$U(x) = \sum_{1 \leq j \leq x} (X_j - \omega_0).$$

It is easy to see that

$$\left| \frac{1}{x} - \frac{1}{[x]} \right| = O\left(\frac{1}{x^2} \right), \qquad x \geq 1.$$

Therefore by Lemma A.1,

$$\begin{aligned} &\left| \int_1^{n-1} \frac{1}{x} U(x)\, dx - \int_1^n \frac{1}{[x]} U(x)\, dx \right| \\ &\quad \leq \left| \int_1^n \left(\frac{1}{x} - \frac{1}{[x]} \right) U(x)\, dx \right| + \left| \int_n^{n+1} \frac{1}{x} U(x)\, dx \right| \\ &\quad \overset{a.s.}{=} O(1) \left\{ \int_1^n \frac{(x \log \log x)^{1/2}}{x^2}\, dx + \frac{1}{n} (n \log \log n)^{1/2} \right\} = O(1). \end{aligned}$$

We can thus conclude that

$$S_n \overset{a.s.}{=} \sigma W(n) - \int_1^{n+1} \frac{1}{x} U(x)\, dx + o(n^{\alpha}).$$

It remains to note that by (2.4) and the modulus of continuity of $W(\cdot)$ (cf. Csörgő and Révész, 1981, Lem. 1.1.1)

$$U(x) - \sigma W(x) \overset{a.s.}{=} o(x^{\alpha}), \quad \text{as } x \to \infty$$

and so

$$\left| \int_1^{n+1} \frac{1}{x} (U(x) - \sigma W(x))\, dx \right| \overset{a.s.}{=} o(1) \int_1^{n+1} \frac{1}{x} x^{\alpha} = o(n^{\alpha}).$$

This completes the proof of Proposition A.1. ∎

Proof of Theorem 3.1. First we verify that

$$\sup_{m \le n < \infty} \frac{|S_n|}{m^{1/2} g(n/m)} \overset{a.s.}{=} \sup_{m \le n < \infty} \frac{\sigma |W^*(n)|}{m^{1/2} g(n/m)} + o(1) \qquad (m \to \infty). \tag{A.3}$$

Relation (A.3) follows from Proposition A.1 and (3.5) because

$$\sup_{m \le n < \infty} \frac{|S_n - \sigma W^*(n)|}{m^{1/2} g(n/m)} \overset{a.s.}{=} o(1) \sup_{m \le n < \infty} \frac{n^{\alpha - 1/2}}{(n/m)^{-1/2} g(n/m)} = o(1).$$

By the scale transformation of the Wiener process

$$\sup_{m \le n < \infty} \frac{|W^*(n)|}{m^{1/2} g(n/m)} \overset{d}{=} \sup_{m \le n < \infty} \frac{|W^*(n/m)|}{g(n/m)}, \tag{A.4}$$

so it suffices to verify that

$$\sup_{m \le n < \infty} \frac{|W^*(n/m)|}{g(n/m)} \xrightarrow{a.s.} \sup_{1 \le t < \infty} \frac{|W^*(t)|}{g(t)}. \tag{A.5}$$

The remainder of the proof consists of the verification of (A.5). To lighten the notation, denote

$$U(t) = \frac{|W^*(t)|}{g(t)}.$$

Because $U(\cdot)$ is continuous with probability one, we assume in the following discussion that we consider a continuous realization for which (3.6) holds. Fix $\epsilon > 0$.

By (3.6), there is a constant $0 \le c < \infty$ such that

$$\limsup_{t \to \infty} U(t) = c. \tag{A.6}$$

By (A.6), there is $T^* > 1$ such that

$$c \le \sup_{T^* \le t < \infty} U(t) \le c + \epsilon. \tag{A.7}$$

Because for any $T > 0$,

$$\sup_{1 \le t < \infty} U(t) = \max \left\{ \sup_{1 \le t \le T} U(t), \sup_{T < t < \infty} U(t) \right\},$$

we conclude that

$$\max \left\{ \sup_{1 \le t \le T^*} U(t), c \right\} \le \sup_{1 \le t < \infty} U(t) \le \max \left\{ \sup_{1 \le t \le T^*} U(t), c + \epsilon \right\}. \tag{A.8}$$

By the first inequality in (A.7), there is $M > T^*$ such that

$$c - \epsilon \le \sup_{T^* \le t \le M} U(t).$$

Because for any $T > 1$,

$$\sup_{m \le n \le mT} U(n/m) \to \sup_{1 \le t \le T} U(t), \quad \text{as } m \to \infty, \tag{A.9}$$

we have, as $m \to \infty$,

$$\begin{aligned} \sup_{m \le n < \infty} U(n/m) &\ge \max \left\{ \sup_{m \le n \le mT^*} U(n/m), \sup_{mT^* < n \le mM} U(n/m) \right\} \\ &\to \max \left\{ \sup_{1 \le t \le T^*} U(t), \sup_{T^* < t \le M} U(t) \right\} \\ &\ge \max \left\{ \sup_{1 \le t \le T^*} U(t), c - \epsilon \right\}. \end{aligned}$$

Hence, there is m^* such that for $m \ge m^*$,

$$\sup_{m \le n < \infty} U(n/m) \ge \max \left\{ \sup_{1 \le t \le T^*} U(t), c \right\} - \epsilon. \tag{A.10}$$

By (A.10) and the second inequality in (A.8), for $m \ge m^*$,

$$\max \left\{ \sup_{1 \le t \le T^*} U(t), c \right\} - \epsilon \le \sup_{m \le n < \infty} U(n/m) \le \max \left\{ \sup_{1 \le t \le T^*} U(t), c \right\} + \epsilon. \tag{A.11}$$

Relation (A.5) follows by combining (A.8) and (A.11). ∎

Proof of Theorem 3.2. By Assumption (2.4),

$$Z_n = \sigma^{-1} \left[\sum_{1 \le i \le n} (X_i - \omega_0) - \frac{n}{m} \sum_{1 \le i \le m} (X_i - \omega_0) \right] \overset{a.s.}{=} W(n) - \frac{n}{m} W(m) + o(n^{\alpha}).$$

By assumption (3.5)

$$\frac{n^{\alpha}}{m^{1/2} \left(\frac{n-m}{m} \right) g\left(\frac{n}{n-m} \right)} = O(1) \frac{n^{\alpha - 1/2}}{(n-m)^{1/2}} = O(n^{\alpha - 1/2}) = O(1). \tag{A.12}$$

Using the scale transformation of the Wiener process and the fact that

$$\{W(t) - tW(1), t \in (1,\infty)\} \stackrel{d}{=} \left\{(t-1)W\left(\frac{t}{t-1}\right), t \in (1,\infty)\right\}$$

we conclude that

$$\sup_{m<n<\infty} \frac{\left|W(n) - \frac{n}{m}W(m)\right|}{m^{1/2}\left(\frac{n-m}{m}\right)g\left(\frac{n}{n-m}\right)} \stackrel{d}{=} \sup_{m<n<\infty} \frac{\left|W(n/m) - \frac{n}{m}W(1)\right|}{\left(\frac{n-m}{m}\right)g\left(\frac{n}{n-m}\right)} \tag{A.13}$$

$$\stackrel{d}{=} \sup_{m<n<\infty} \frac{\left|W\left(\frac{n/m}{n/m-1}\right)\right|}{g\left(\frac{n/m}{n/m-1}\right)}.$$

By (A.12) and (A.13), it remains to show that

$$\sup_{m<n<\infty} \frac{\left|W\left(\frac{n/m}{n/m-1}\right)\right|}{g\left(\frac{n/m}{n/m-1}\right)} \xrightarrow{a.s.} \sup_{1<t<\infty} \frac{\left|W\left(\frac{t}{t-1}\right)\right|}{g\left(\frac{t}{t-1}\right)}. \tag{A.14}$$

Note that

$$\sup_{1<t<\infty} \frac{\left|W\left(\frac{t}{t-1}\right)\right|}{g\left(\frac{t}{t-1}\right)} \stackrel{d}{=} \sup_{1<t<\infty} \frac{|W(t)|}{g(t)}.$$

Introduce the map $u(t) = t/(t-1)$, $t > 1$. Because $|u'(t)| \le (c-1)^{-2}$ for $t \ge c > 1$, for any fixed $c > 1$

$$\sup_{n \ge cm} \frac{|W(u(n/m))|}{g(u(n/m))|} \xrightarrow{a.s.} \sup_{t \ge c} \frac{|W(u(t))|}{g(u(t))}. \tag{A.15}$$

Because

$$\sup_{1<t<c} \frac{|W(u(t))|}{g(u(t))} = \sup_{u>c/(c-1)} \frac{|W(u)|}{g(u)},$$

assumption (3.6) implies that

$$\lim_{c \to 1} \sup_{1<t<c} \frac{|W(u(t))|}{g(u(t))} \stackrel{a.s.}{=} 0. \tag{A.16}$$

Relation (A.14) follows from (A.15) and (A.16). This completes the proof of Theorem 3.2. ∎

APPENDIX B: Proofs of Theorems 3.3 and 3.4

Proof of Theorem 3.3. Observe that under H_0, by (2.4),

$$\hat{S}(m,k) = \sum_{m<i\le m+k} (X_i - \omega_0) - \frac{k}{m} \sum_{1\le i\le m} (X_i - \omega_0)$$

$$\stackrel{a.s.}{=} \sigma[W(m+k) - W(m) - (k/m)W(m)] + o((m+k)^\alpha + (k/m)m^\alpha).$$

Therefore,

$$\sup_{1\le k<\infty} \frac{|\hat{S}(m,k)|}{m^{1/2}g_\gamma(m,k)} \stackrel{a.s.}{=} \sigma \sup_{1\le k<\infty} \frac{|W(m+k) - W(m) - (k/m)W(m)|}{m^{1/2}g_\gamma(m,k)}$$

$$+ o(1) \sup_{1\le k<\infty} \frac{(m+k)^\alpha + (k/m)m^\alpha}{m^{1/2}g_\gamma(m,k)}.$$

Elementary verification shows that

$$\sup_{1\le k<\infty} \frac{(m+k)^\alpha + (k/m)m^\alpha}{m^{1/2}g_\gamma(m,k)} = O(m^{\alpha+\gamma-1/2}),$$

so the assumption $\gamma < \frac{1}{2} - \alpha$ yields

$$\sup_{1\le k<\infty} \frac{|\hat{S}(m,k)|}{m^{1/2}g_\gamma(m,k)} \stackrel{a.s.}{=} \sigma \sup_{1\le k<\infty} \frac{|W(m+k) - W(m) - (k/m)W(m)|}{m^{1/2}g_\gamma(m,k)} + o(1).$$

It thus remains to verify that

$$\sup_{1\le k<\infty} \frac{|W(m+k) - W(m) - (k/m)W(m)|}{m^{1/2}g_\gamma(m,k)} \xrightarrow{d} \sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma}. \tag{B.1}$$

Notice that if $k/m \to t$, as $m \to \infty$, then

$$g_\gamma(m,k) \to h_\gamma(t) := (1+t)\left(\frac{t}{1+t}\right)^\gamma, \qquad t \in [0,\infty). \tag{B.2}$$

Note also that (see Horváth et al., 2004, proof of Thm. 2.1)

$$\sup_{0\le t<\infty} \frac{|W(1+t) - W(1) - tW(1)|}{h_\gamma(t)} \stackrel{d}{=} \sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma}. \tag{B.3}$$

Using the scale transformation of the Wiener process, (B.2) and the modulus of continuity of the Wiener process, and finally (B.3), we obtain

$$\begin{aligned}\sup_{1\le k<\infty} \frac{|W(m+k) - W(m) - (k/m)W(m)|}{m^{1/2} g_\gamma(m,k)} &\stackrel{d}{=} \sup_{1\le k<\infty} \frac{|W(1+k/m) - W(1) - (k/m)W(1)|}{h_\gamma(k/m)} \\ &\xrightarrow{a.s.} \sup_{0\le t<\infty} \frac{|W(1+t) - W(1) - tW(1)|}{h_\gamma(t)} \\ &\stackrel{d}{=} \sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma}.\end{aligned}$$

This completes the verification of (B.1) and the proof of Theorem 3.3. ∎

Proof of Theorem 3.4. Observe that under H_0

$$\begin{aligned}\tilde{S}(m,k) = &\sum_{m<i\le m+k} (X_i - \omega_0) - \sum_{m<i\le m+k} \frac{1}{i-1} \sum_{j=1}^{m} (X_j - \omega_0) \\ &- \sum_{m<i\le m+k} \frac{1}{i-1} \sum_{j=m+1}^{i-1} (X_j - \omega_0).\end{aligned}$$

Using (2.4) we obtain

$$\sup_{1\le k<\infty} \frac{\sum_{m<i\le m+k} (X_i - \omega_0) - \sigma[W(m+k) - W(m)]}{(m+k)^\alpha} \stackrel{a.s.}{=} o(1) \qquad (m \to \infty)$$

and

$$\begin{aligned}&\left| \sum_{m<i\le m+k} \frac{1}{i-1} \sum_{j=1}^{m} (X_j - \omega_0) - \sum_{m<i\le m+k} \frac{1}{i-1} \sigma W(m) \right| \\ &\quad \le \left| \sum_{j=1}^{m} (X_j - \omega_0) - \sigma W(m) \right| \sum_{m<i\le m+k} \frac{1}{i-1} = \xi_m \ln\left(\frac{m+k}{m}\right),\end{aligned}$$

where $\xi_m \stackrel{a.s.}{=} o(m^\alpha)$, as $m \to \infty$. Because

$$\sup_{1\le k<\infty} \left| \sum_{m<i\le m+k} \frac{1}{i-1} - \ln\left(\frac{m+k}{m}\right) \right| = O\left(\frac{1}{m}\right), \quad m \to \infty,$$

we see that there is a random variable $\xi_m^* \stackrel{a.s.}{=} o(m^\alpha)$ $(m \to \infty)$, which does not depend on k, such that

$$\left| \sum_{m<i\le m+k} \frac{1}{i-1} \sum_{j=1}^{m} (X_j - \omega_0) - \sigma W(m) \ln\left(\frac{m+k}{m}\right) \right| \le \xi_m^* \ln\left(\frac{m+k}{m}\right).$$

Similarly, as $m \to \infty$

$$\sup_{1\le k<\infty} \frac{1}{(m+k)^\alpha} \left| \sum_{m<i\le m+k} \sum_{j=m+1}^{i-1} (X_j - \omega_0) - \sigma \int_m^{m+k} [W(x-1) - W(m)]\,dx \right| \stackrel{a.s.}{=} o(1).$$

We note that

$$\sup_{1\le k<\infty} \frac{m^\alpha \ln\left(\dfrac{m+k}{m}\right)}{(m+k)^\alpha} = o(1), \qquad (m \to \infty)$$

and conclude that, as $m \to \infty$,

$$\begin{aligned}\sup_{1\le k<\infty} \frac{1}{(m+k)^\alpha} \Bigg| \frac{1}{\sigma} \tilde{S}(m,k) - \Bigg\{ &[W(m+k) - W(m)] \\ &- W(m) \ln\left(\frac{m+k}{m}\right) - \int_m^{m+k} \frac{1}{x-1} \\ &\times [W(x-1) - W(m)]\,dx \Bigg\} \Bigg| \stackrel{a.s.}{=} o(1).\end{aligned} \tag{B.4}$$

By (3.24),

$$\sup_{1\le k<\infty} \frac{(m+k)^\alpha}{m^{1/2} h(k/m)} = O(1), \qquad (m \to \infty). \tag{B.5}$$

Combining (B.4) and (B.5), we obtain

$$\begin{aligned}\frac{1}{\sigma} \sup_{1\le k<\infty} \frac{|\tilde{S}(m,k)|}{m^{1/2} h(k/m)} &\stackrel{a.s.}{=} \sup_{1\le k<\infty} \frac{\left| [W(m+k) - W(m)] - \log((m+k)/m) W(m) - \displaystyle\int_m^{m+k} \frac{1}{x-1} [W(x-1) - W(m)]\,dx \right|}{m^{1/2} h(k/m)} \\ &\quad + o(1).\end{aligned} \tag{B.6}$$

By the scale transformation and the modulus of continuity of the Wiener process,

$$\sup_{1\le k<\infty}\frac{\left|[W(m+k)-W(m)]-\log((m+k)/m)W(m)-\displaystyle\int_m^{m+k}\frac{1}{x-1}[W(x-1)-W(m)]\,dx\right|}{m^{1/2}h(k/m)}$$

$$\stackrel{d}{=}\sup_{1\le k<\infty}\frac{\left|[W(1+k/m)-W(1)]-\log(1+k/m)W(1)-\displaystyle\int_{1-1/m}^{1+k/m}\frac{1}{s}[W(s)-W(1)]\,ds\right|}{h(k/m)}$$

$$\stackrel{a.s.}{=}\sup_{1\le k<\infty}\frac{|\Gamma(k/m)|}{h(k/m)}+o(1),$$

where

$$\Gamma(t)=W(1+t)-W(1)-\log(1+t)W(1)-\int_1^{1+t}\frac{1}{s}[W(s)-W(1)]\,ds.$$

As observed in Section 6 of Horváth et al. (2004), computing the covariances of the process $\Gamma(\cdot)$ shows that it is a Wiener process. Therefore, it remains to show that

$$\sup_{1\le k<\infty}\frac{|\Gamma(k/m)|}{h(k/m)}\xrightarrow{a.s.}\sup_{0<t<\infty}\frac{|\Gamma(t)|}{h(t)}.$$

This can be done by repeating the corresponding argument used in the proof of Theorem 3.1. Thus the proof of Theorem 3.4 is complete. ■

Econometric Theory, **22**, 2006, 403–428. Printed in the United States of America.
DOI: 10.1017/S0266466606060208

EMPIRICAL LIKELIHOOD FOR GARCH MODELS

NGAI HANG CHAN
Chinese University of Hong Kong

SHIQING LING
Hong Kong University of Science and Technology

This paper develops an empirical likelihood approach for regular generalized autoregressive conditional heteroskedasticity (GARCH) models and GARCH models with unit roots. For regular GARCH models, it is shown that the log empirical likelihood ratio statistic asymptotically follows a χ^2 distribution. For GARCH models with unit roots, two versions of the empirical likelihood methods, the least squares score and the maximum likelihood score functions, are considered. For both cases, the limiting distributions of the log empirical likelihood ratio statistics are established. These two statistics can be used to test for unit roots under the GARCH framework. Finite-sample performances are assessed through simulations for GARCH models with unit roots.

1. INTRODUCTION

For independent and identically distributed (i.i.d.) random variables $y_1,\ldots,y_n$, the distribution F of y_i is usually estimated by the empirical distribution function $F_n=\sum_{i=1}^n p_i I_{\{y_i\le x\}}$ subject to the constraints $\sum_{i=1}^n p_i=1$ and $p_i\ge 0$. This empirical distribution is maximized at $p_i=1/n$, $i=1,\ldots,n$. Owen (1988, 1990) defines the empirical likelihood ratio as

$$R(F)=\prod_{i=1}^n np_i,$$

where $p_i=dF(y_i)=P(Y=y_i)$. If the distribution function F is characterized by an unknown parameter λ, then the probabilities p_i should be subject to some restrictions on the parameter λ. For example, when F is characterized by the mean μ, then p_i needs to satisfy the first-order restriction $\sum_{i=1}^n p_i(x_i-\mu)=0$.

This research was supported in part by Hong Kong Research grants Council Grants CUHK4043/02P and HKUST6273/03H. The authors thank two referees and the Co-Editor, Bruce Hansen, for insightful and helpful comments about the relationship between QMLE and MELE, which led to substantial improvement of the presentation. Computational assistance from Jerry Wong and Chun-Yip Yau is also gratefully acknowledged. Address correspondence to Ngai Hang Chan, Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong; e-mail: nhchan@sta.cuhk.edu.hk.