Econometric Theory, **20**, 2004, 1140–1167. Printed in the United States of America.
DOI: 10.1017/S0266466604206041

SEQUENTIAL CHANGE-POINT DETECTION IN GARCH(p,q) MODELS

ISTVÁN BERKES
Hungarian Academy of Sciences

EDIT GOMBAY
University of Alberta

LAJOS HORVÁTH
University of Utah

PIOTR KOKOSZKA
Utah State University

We suggest a sequential monitoring scheme to detect changes in the parameters of a GARCH(p,q) sequence. The procedure is based on quasi-likelihood scores and does not use model residuals. Unlike for linear regression models, the squared residuals of nonlinear time series models such as generalized autoregressive conditional heteroskedasticity (GARCH) do not satisfy a functional central limit theorem with a Wiener process as a limit, so its boundary crossing probabilities cannot be used. Our procedure nevertheless has an asymptotically controlled size, and, moreover, the conditions on the boundary function are very simple; it can be chosen as a constant. We establish the asymptotic properties of our monitoring scheme under both the null of no change in parameters and the alternative of a change in parameters and investigate its finite-sample behavior by means of a small simulation study.

1. INTRODUCTION

The assumption that the parameters remain stable over time plays a crucial role in statistical modeling and inference. If the parameters have changed within the observed sample, then, e.g., forecasts lose accuracy and the parameter estimates no longer provide meaningful information. Because of the importance of

This research was partially supported by NSF grant INT-0223262 and NATO grant PST.CLG.977607. The work of the first author was supported by the Hungarian National Foundation for Scientific Research, grants T 29621, 37886; the work of the second author was supported by NSERC Canada. Address correspondence to István Berkes, A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, P.O. Box 127, H-1364 Budapest, Hungary. Edit Gombay, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. Lajos Horváth, Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112-0090, USA. 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.

 © 2004 Cambridge University Press 0266-4666/04 $12.00

parameter stability, the detection of possible changes in the data generating process has become an active area of research. There are hundreds of papers studying various aspects of change-point detection in econometric settings. We list those most closely related to the present work later in this section. For surveys concerned with the general statistical methodology for change-point detection, we refer to Besseville and Nikifirov (1993), Brodsky and Darkhovsky (1993, 2000), and Csörgő and Horváth (1997).

The present paper is motivated by the work of Chu, Stinchcombe, and White (1996), who pose the following problem: "Given a previously estimated model, the arrival of new data invites the question: is yesterday's model capable of explaining today's data?" These authors develop fluctuation and cumulative sum (CUSUM) monitoring procedures for linear regression models. Our paper focuses on generalized autoregressive conditional heteroskedasticity (GARCH) models, which are important in financial applications (see, e.g., Gouriéroux, 1997) and, as explained later, differ from linear models in ways that make a direct application of the approach of Chu et al. (1996) not readily possible.

Before outlining the central ideas of Chu et al. (1996) and contrasting our approach with theirs, we note that procedures for detecting parameter changes in a GARCH specification may have a number of useful applications. For example, risk managers use GARCH for calculating portfolio risk measures such as value at risk. GARCH parameters are typically estimated over a rolling window of returns, and change points in this window will introduce bias. Similarly, option traders use GARCH to make up for the well-known biases in Black–Scholes option prices. A detection of parameter changes may lead to a more cautious interpretation of the calculated option prices. Although our paper is not concerned with such applied issues, we do hope that it will make a contribution to the important problem of monitoring for changes in GARCH models.

As argued convincingly in Chu et al. (1996), the sequential analysis of economic and financial data is somewhat different from engineering applications. The sampling is costless under the no change null hypothesis, and no action is required if the observed processes is "in control," i.e., there is no change in the parameters of the data generating process. Because it is impossible to eliminate false alarm due to chance, the probability of stopping under the no change null hypothesis should be less than a given level $0 < \alpha < 1$. On the other hand, it is desirable to stop with probability one if a change occurs. With these goals in mind, Chu et al. (1996) propose to consider decisions functions of the form

$$Q_t^{(m)} = \sum_{m+1 \le i < m+mt} \hat{\omega}_i, \qquad t \in [0,\infty), \tag{1.1}$$

where m denotes the number of initial observations used to estimate the model and the $\hat{\omega}_i$ are model residuals that are estimated sequentially. The idea of their procedure is as follows: denoting by $W(\cdot)$ the standard Wiener process, suppose it can be shown that, as $m \to \infty$,

$$m^{-1/2}Q_t^{(m)} \to W(t) \quad \text{in } D[0,\infty) \tag{1.2}$$

and that a function g can be found such that

$$\lim_{m\to\infty} P(|S_n| \geq \sqrt{m}g(n/m), \quad \text{for some } n \geq m)$$
$$= P(|W(t)| \geq g(t) \quad \text{for some } t \geq 1), \tag{1.3}$$

provided the sequence S_n satisfies a functional central limit theorem. It can then be concluded from (1.2) and (1.3) that

$$\lim_{m\to\infty} P(|Q_t^{(m)}| \geq \sqrt{m}g(n/m), \quad \text{for some } n \geq m)$$
$$= P(|W(t)| \geq g(t) \quad \text{for some } t \geq 1). \tag{1.4}$$

For several judicious choices of g, the probability on the right-hand side of (1.4) can be computed analytically, and consequently a monitoring scheme can be developed such that asymptotically, as $m \to \infty$, the decision function $|Q_t^{(m)}|$ crosses the boundary $\sqrt{m}g(n/m)$ at some future date n with a prescribed probability α, provided the parameters have not changed.

The present paper is concerned with the detection of changes in GARCH(p,q) processes that are used to model volatility. Statistics designed for detecting changes in volatility are typically based on the squares of observations or the squares of residuals. The squares of residuals from a GARCH(p,q) process do not satisfy a functional limit theorem with the Wiener process in the limit because of the presence of extra terms in their covariance structure (see, e.g., Boldin, 1998; Horváth, Kokoszka, and Teyssière, 2001; Koul, 2002, Ch. 8; Berkes and Horváth, 2003; Kulperger and Yu, 2003). Thus the approach of Chu et al. (1996) cannot be readily applied. In this paper, we put forward a different approach that does not use model residuals but relies more directly on the quasi-likelihood function. Suppose we have observed $y_1, \dots, y_m$ for which we have postulated a model with d parameters. Denote by $\mathbf{u}$ a generic element of the parameter space and by $\ell_i(\mathbf{u})$ the conditional quasi-likelihood of y_i given $y_{i-1}, \dots, y_1$, so that the quasi-likelihood function is $L_m(\mathbf{u}) = \sum_{1\leq i\leq m} \ell_i(\mathbf{u})$. For time series models, the $\ell_i(\mathbf{u})$ typically cannot be computed exactly because of the dependence on the unobserved $y_k, k \leq 0$, and some approximations $\hat{\ell}_i(\mathbf{u})$ must be used instead. Denote by $\hat{\ell}_i'(\mathbf{u})$ the d-dimensional row vector of partial derivatives with respect to the model parameters and consider the matrix

$$\hat{\mathbf{D}}_m = \frac{1}{m} \sum_{1<i\leq m} (\hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m))^T \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m), \tag{1.5}$$

where $\hat{\boldsymbol{\theta}}_m$ is the quasi–maximum likelihood parameter estimate. We can now construct the d-dimensional process

$$G_m(t) = \sum_{m<i\leq mt} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m)\hat{\mathbf{D}}_m^{-1/2}, \qquad t \in [0,\infty). \tag{1.6}$$

Our approach relies on the realization that the process $G_m(\cdot)$ can be well approximated by the process

$$\left((1+t)W_1\left(\frac{t}{1+t}\right), \dots, (1+t)W_d\left(\frac{t}{1+t}\right)\right), \qquad t \in [0,\infty), \tag{1.7}$$

where $W_j(\cdot), j = 1, \dots d$ are independent standard Wiener processes. By taking appropriate functionals, it is then easy to construct a monitoring scheme with a controlled probability of a false rejection of the null hypothesis. The details are presented in Section 3.

Although the preceding idea is applicable to essentially any parametric model for which a reasonable approximate quasi-likelihood function can be found, a rigorous verification for a complex nonlinear model requires some effort. First of all, appropriate approximations $\hat{\ell}_i'(\mathbf{u})$ must be used. To obtain them, we use an expansion developed in Berkes, Horváth, and Kokoszka (2003), which is described in Section 2. Second, asymptotics for the matrix $\hat{\mathbf{D}}_m$ in (1.5) must be established that allow us to approximate the process $G(\cdot)$ by a bridge-type multidimensional Gaussian process (see Lemma 6.6 in Section 6). The corresponding result is stated in Proposition 3.1, which is proved in Section 5. Finally, the partial sum process $\sum_{m<i\leq mt} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m)$ must be carefully approximated, which is accomplished in several stages presented in Section 6.

By considering the quasi-likelihood scores, our approach is related to that of Nyblom (1989), but it is different in that our focus is on controlling the probability of false alarms, whereas Nyblom (1989) concentrates on constructing locally most powerful tests against alternatives that assume that the parameter changes $\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}$ from time $k-1$ to time k form a martingale sequence with known covariance matrices $\mathbf{G}_k = E[(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1})(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1})^T]$. Such assumptions are not appropriate in the context of on-line monitoring for a change in the GARCH(p,q) parameters. Moreover, to prove a functional limit theorem similar to our relation (6.53), Nyblom (1989) needs to impose a number of additional technical assumptions on the asymptotic properties of the likelihood scores. Our assumptions pertain only to model parameters and the distribution of the model errors and are very weak, essentially necessary for the consistency and asymptotic normality of the quasi–maximum likelihood estimator. Another important contribution to the theory of optimal a posteriori (in-sample) change-point tests under local alternatives is made by Sowell (1996), who considers a much more general setting. His assumptions, however, also impose various conditions on scorelike objects rather than on the model parameters and errors.

There has been a growing literature concerned with the change-point problem specifically in GARCH setting or with a more general problem of detecting changes in volatility. Lamoureux and Lastrapes (1990) and Mikosch and Stărică (2002), among others, show that change points in a GARCH specification may lead to the presence of spurious persistence in volatility. Mikosch and

Stărică (2002) also propose a periodogram-based on-line change-point detection procedure but do not develop it theoretically. The following papers deal with change-point detection in a historical sample (a posteriori methods). Inclan and Tiao (1994), Kokoszka and Leipus (1999, 2000), Kim, Cho, and Lee (2000), and Andreou and Ghysels (2002) study CUSUM type tests based on squared observations. Tests based on the empirical distribution function of the squared residuals are studied by Horváth et al. (2001) and Kokoszka and Teyssiere (2002). Inoue (2001) proposes a simulation-based method based on the empirical distribution function of the observations and applies it to stochastic volatility models. CUSUM tests based on the partial sums of powers of residuals are developed by Kulperger and Yu (2003). Kokoszka and Teyssière (2002) also study a generalized likelihood ratio test. Chu (1995) and Lundbergh and Teräsvirta (2002) investigate Lagrange multiplier tests. The preceding list is not meant to be exhaustive but is intended to indicate a recent interest in and a variety of approaches to the problem of change-point detection in GARCH processes. For a review of sequential testing strategies in the case of independent observations, see, e.g., Gombay (1996, 2002).

The paper is organized as follows. In Section 2 we present the necessary technical background on GARCH processes, which includes very weak assumptions on the model parameters and errors. In particular, we do not require that the innovations ϵ_i have a smooth density or that the observations y_i have finite expected value. Section 3 describes the monitoring scheme and contains results establishing its asymptotic behavior under the null and under the alternative. Results of a small simulation study are presented in Section 4. Proofs are collected in Sections 5–7.

2. DEFINITIONS AND ASSUMPTIONS

In this section we present the general framework used throughout the paper. We recall the definition of a GARCH(p,q) process and present recursions needed to define the quasi–maximum likelihood estimator. We state the conditions for the existence of a stationary solution to the GARCH(p,q) equations and for the consistency and asymptotic normality of the maximum likelihood estimator. The section is based primarily on the results of Bougerol and Picard (1992a, 1992b) and Berkes et al. (2003).

We assume that

$$y_k = \sigma_k \epsilon_k \tag{2.8}$$

and under the no change null hypothesis

$$\sigma_k^2 = \omega + \sum_{1\le i\le p} \alpha_i y_{k-i}^2 + \sum_{1\le j\le q} \beta_j \sigma_{k-j}^2, \tag{2.9}$$

where $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ is the parameter of the process. Under the alternative

$$\sigma_k^2 = \begin{cases} \omega + \sum_{1\le i\le p} \alpha_i y_{k-i}^2 + \sum_{1\le j\le q} \beta_i \sigma_{k-j}^2, & -\infty < k \le k^* \\ \omega^* + \sum_{1\le i\le p} \alpha_i^* y_{k-i}^2 + \sum_{1\le j\le q} \beta_i^* \sigma_{k-j}^2, & k^* < k < \infty, \end{cases} \tag{2.10}$$

i.e., a change in the parameters occurred at time k^* and the new parameter is $\boldsymbol{\theta}^* = (\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*)$. In the following discussion, we refer to the specification (2.9) as the null hypothesis H_0 and to (2.10) for some integer k^* as the alternative hypothesis H_A.

Throughout this paper we assume that

$$\epsilon_k, -\infty < k < \infty \text{ are independent, identicallydistributed random variables.} \tag{2.11}$$

Note that like Bougerol and Picard (1992a) we do not assume that the errors ϵ_k have mean zero. Our procedure is not based on the deviations of residuals from zero. Additional assumptions on the distribution of the ϵ_k are stated in conditions (2.17)–(2.19).

Our procedure is based on the quasi–maximum likelihood estimator of the parameters of a GARCH process developed by Lee and Hansen (1994), Lumsdaine (1996), and Berkes et al. (2003). (For a more general method we refer to Berkes and Horváth, 2004.) To define this estimator for general GARCH(p,q) processes, denote by $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$ the generic element of the parameter space U, which, following Berkes et al. (2003), is defined as follows: let $0 < \underline{u} < \bar{u}$, $0 < \rho_0 < 1$, $q\underline{u} < \rho_0$. Then

$$U = \{\mathbf{u} : t_1 + t_2 + \cdots + t_q \le \rho_0 \text{ and } \underline{u} < \min(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \le \max(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \le \bar{u}\}.$$

We assume that

$$\boldsymbol{\theta} \in U. \tag{2.12}$$

Define now the log quasi-likelihood function as

$$\hat{L}_m(\mathbf{u}) = \frac{1}{m} \sum_{1<k\le m} -\frac{1}{2}\left\{\log \hat{w}_k(\mathbf{u}) + \frac{y_k^2}{\hat{w}_k(\mathbf{u})}\right\},$$

where

$$\hat{w}_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le i<k} c_i(\mathbf{u}) y_{k-i}^2.$$

The functions $c_i(\mathbf{u})$, $0 \le i < \infty$ are defined by recursion. If $q \ge p$, then

$$\begin{aligned}
c_0(\mathbf{u}) &= x/(1-(t_1+\cdots+t_q)),\\
c_1(\mathbf{u}) &= s_1,\\
c_2(\mathbf{u}) &= s_2 + t_1 c_1(\mathbf{u}),\\
&\vdots\\
c_p(\mathbf{u}) &= s_p + t_1 c_{p-1}(\mathbf{u}) + \cdots + t_{p-1} c_1(\mathbf{u}),\\
c_{p+1}(\mathbf{u}) &= t_1 c_p(\mathbf{u}) + \cdots + t_p c_1(\mathbf{u}),\\
&\vdots\\
c_q(\mathbf{u}) &= t_1 c_{q-1}(\mathbf{u}) + \cdots + t_{q-1} c_1(\mathbf{u}),
\end{aligned}$$

and if $q < p$, the preceding equations are replaced with

$$\begin{aligned}
c_0(\mathbf{u}) &= x/(1-(t_1+\cdots+t_q)),\\
c_1(\mathbf{u}) &= s_1,\\
c_2(\mathbf{u}) &= s_2 + t_1 c_1(\mathbf{u}),\\
&\vdots\\
c_{q+1}(\mathbf{u}) &= s_{q+1} + t_1 c_q(\mathbf{u}) + \cdots + t_q c_1(\mathbf{u}),\\
&\vdots\\
c_p(\mathbf{u}) &= s_p + t_1 c_{p-1}(\mathbf{u}) + \cdots + t_q c_{p-q}(\mathbf{u}).
\end{aligned}$$

In general, if $i > R = \max(p,q)$, then

$$c_i(\mathbf{u}) = t_1 c_{i-1}(\mathbf{u}) + t_2 c_{i-2}(\mathbf{u}) + \cdots + t_q c_{i-q}(\mathbf{u}).$$

The preceding recursions ensure that for the true value $\boldsymbol{\theta}$, $\sigma_k^2 = c_0(\boldsymbol{\theta}) + \sum_{1\le i<\infty} c_i(\boldsymbol{\theta}) y_{k-i}^2$.

To formulate a necessary and sufficient condition for the existence of a unique stationary sequence satisfying (2.8) and (2.9) we must introduce further notation. Let

$$\boldsymbol{\tau}_n = (\beta_1 + \alpha_1 \epsilon_n^2, \beta_2, \ldots, \beta_{q-1}) \in \mathbb{R}^{q-1},$$

$$\boldsymbol{\xi}_n = (\epsilon_n^2, 0, \ldots, 0) \in \mathbb{R}^{q-1},$$

and

$$\boldsymbol{\alpha} = (\alpha_2, \ldots, \alpha_{p-1}) \in \mathbb{R}^{p-2}.$$

(Clearly, without loss of generality we may and shall assume $\min(p,q) \ge 2$.) Define the $(p+q-1)\times(p+q-1)$ matrix A_n, written in block form, by

$$A_n = \begin{bmatrix} \boldsymbol{\tau} & \beta_q & \boldsymbol{\alpha} & \alpha_p \\ \mathbf{I}_{q-1} & 0 & 0 & 0 \\ \boldsymbol{\xi}_n & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I}_{p-2} & 0 \end{bmatrix},$$

where $\mathbf{I}_{q-1}$ and $\mathbf{I}_{p-2}$ are the identity matrices of size $q-1$ and $p-2$, respectively. The norm of any $d \times d$ matrix M is defined by

$$\|M\| = \sup\{\|M\mathbf{x}\|_d / \|\mathbf{x}\|_d : \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \ne 0\},$$

where $\|\cdot\|_d$ is the euclidean norm in $\mathbb{R}^d$. The top Lyapunov exponent γ_L associated with the sequence $\{A_n, -\infty < n < \infty\}$ is

$$\gamma_L = \inf_{1\le n<\infty} \frac{1}{n} E \log\|A_0 A_1 \ldots A_n\|,$$

assuming that

$$E(\log\|A_0\|) < \infty. \tag{2.13}$$

(We note that $\|A_0\| \ge 1$; cf. Berkes et al., 2003.) Bougerol and Picard (1992a, 1992b) show that if (2.13) holds, then (2.8) and (2.9) have a unique stationary solution if and only if

$$\gamma_L < 0. \tag{2.14}$$

We note that (2.14) implies $\beta_1 + \cdots + \beta_q < 1$ (cf. Bougerol and Picard, 1992b).

The next two conditions are needed to uniquely identify the parameter $\boldsymbol{\theta}$:

the polynomials $\alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_p x^p$ and

$1 - \beta_1 x - \beta_2 x^2 - \cdots - \beta_q x^q$ are coprimes on the set of polynomials with real coefficients, **(2.15)**

and

ϵ_0^2 is a nondegenerate random variable. **(2.16)**

We also assume

$$\lim_{t\to 0} t^{-\mu} P\{\epsilon_0^2 \le t\} = 0 \quad \text{with some } \mu > 0, \tag{2.17}$$

which will be needed to estimate the moments of $w_0(\mathbf{u})/w_0(\boldsymbol{\theta})$ (cf. Berkes et al., 2003).

Finally, the last set of conditions concerns the moments of ϵ_0:

$$E\epsilon_0^2 = 1 \tag{2.18}$$

and

$$E|\epsilon_0^2|^{2+\delta} < \infty \quad \text{with some } \delta > 0. \tag{2.19}$$

Berkes et al. (2003) show that $m^{1/2}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})$ is asymptotically normal if (2.8), (2.9), and (2.13)–(2.19) hold.

3. MAIN RESULTS

Suppose we have observed $y_1, \ldots, y_m$, which represent available historical data. The estimator for the unknown parameter $\boldsymbol{\theta}$ based on these data is defined by

$$\hat{\boldsymbol{\theta}}_m = \arg\max\{\hat{L}_m(\mathbf{u}) : \mathbf{u} \in U\}, \tag{3.20}$$

where U is a suitably chosen compact set defined in Section 2. Using the notation introduced in Section 2, we also define the conditional likelihoods

$$\hat{\ell}_i(\mathbf{u}) = -\frac{1}{2}\left\{\log \hat{w}_i(\mathbf{u}) + \frac{y_i^2}{\hat{w}_i(\mathbf{u})}\right\}$$

and the matrix

$$\hat{\mathbf{D}}_m = \frac{1}{m} \sum_{1<i\le m} (\hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m))^T \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m).$$

(Here ^{T} denotes the transpose of vectors and matrices.)

Let $|\cdot|$ denote the maximum norm of vectors and matrices. We can now define the stopping time k_m as

$$k_m = \min\left\{k : \left|\sum_{m<i\le m+k} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m)\hat{\mathbf{D}}_m^{-1/2}\right| > m^{1/2}\left(1 + \frac{k}{m}\right) b\left(\frac{k}{m}\right)\right\}.$$

If $k_m < \infty$, we say that a change occurred. We choose the boundary function $b(t)$ so that

$$\lim_{m\to\infty} P_{H_0}\{k_m < \infty\} = \alpha, \tag{3.21}$$

where $0 < \alpha < 1$ is a prescribed number and

$$\lim_{m\to\infty} P_{H_A}\{k_m < \infty\} = 1. \tag{3.22}$$

Recall that H_0 is defined by (2.8) and (2.9) and H_A by (2.8) and (2.10). Conditions on the boundary function $b(\cdot)$ are specified in the following two sections, which study the asymptotic behavior of the monitoring scheme, respectively, under H_0 and H_A. Unlike for the scheme of Chu et al. (1996), which requires more complex boundary functions, in our setting, the simplest choice satisfying all the assumptions is to take a constant $b(t) = b > 0$.

3.1. Monitoring Scheme under H_0

Throughout this section we assume that (2.8) and (2.9) hold.

It follows from the definition that the matrix $\hat{\mathbf{D}}_m$ is nonnegative definite. We show in Proposition 3.1, which follows, that $\hat{\mathbf{D}}_m$ is nonsingular with probability tending to one as $m \to \infty$. Hence $\hat{\mathbf{D}}_m^{-1/2}$ exists with probability tending to one as $m \to \infty$.

To formulate Proposition 3.1, we define

$$w_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le i<\infty} c_i(\mathbf{u}) y_{k-i}^2$$

and

$$\ell_k(\mathbf{u}) = -\frac{1}{2}\left(\log w_k(\mathbf{u}) + \frac{y_k^2}{w_k(\mathbf{u})}\right).$$

Finally, define

$$\mathbf{D}(\mathbf{u}) = E[(\ell_0'(\mathbf{u}))^T \ell_0'(\mathbf{u})]$$

and

$$\mathbf{D} = \mathbf{D}(\boldsymbol{\theta}).$$

PROPOSITION 3.1. *If (2.8), (2.9), and (2.13)–(2.19) are satisfied, then*

$$\hat{\mathbf{D}}_m \to \mathbf{D} \ a.s.$$

Also, $\mathbf{D}$ *is a positive definite, nonsingular matrix.*

The proof of Proposition 3.1 is presented in Section 5.

We impose the following conditions on the boundary function $b(t)$:

$$b(t) \quad \text{is continuous on } (0,\infty) \tag{3.23}$$

and

$$\inf_{0<t<\infty} b(t) > 0. \tag{3.24}$$

THEOREM 3.1. *If (2.8), (2.9), (2.13)–(2.19), and conditions (3.23) and (3.24) are satisfied, then*

$$\lim_{m\to\infty} P_{H_0}\{k_m < \infty\} = 1 - \left(P\left\{\sup_{0\le t<1} \frac{|W(t)|}{b(t/(1-t))} \le 1\right\}\right)^{p+q+1},$$

where $\{W(t), 0 \le t \le 1\}$ *denotes a Wiener process.*

The proof of Theorem 3.1 is presented in Section 6.

Choosing $b(t) = b$, a constant function, and using the well-known formula for the distribution function of $\sup_{0\le t\le 1}|W(t)|$ (cf. Csörgő and Révész, 1981, p. 43) we obtain the following corollary.

COROLLARY 3.1. *If (2.8), (2.9), (2.13)–(2.19), and conditions (3.23) and (3.24) are satisfied, then*

$$\lim_{m\to\infty} P_{H_0}\{k_m < \infty\} = 1 - \left(P\left\{\sup_{0\le t\le 1}|W(t)| \le b\right\}\right)^{p+q+1}$$
$$= 1 - \left\{\frac{4}{\pi}\sum_{0\le k<\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{\pi^2(2k+1)^2}{8b^2}\right)\right\}^{p+q+1}.$$

Corollary 3.1 allows us to specify the critical level b for any significance level α in (3.21).

3.2. Monitoring Scheme under H_A

Under the alternative (2.10) the parameter changes from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$ at time $k^* > m$. We define the sequence $\{z_k, -\infty < k < \infty\}$, which represents the model after a change in parameters, by the equations

$$z_k = \gamma_k \epsilon_k \tag{3.25}$$

and

$$\gamma_k^2 = \omega^* + \sum_{1\le i\le p}\alpha_i^* z_{k-i}^2 + \sum_{1\le j\le q}\beta_j^*\gamma_{k-j}^2. \tag{3.26}$$

The error sequence ϵ_k has not changed.

Our first condition is that the parameter $\boldsymbol{\theta}^*$ defines a GARCH(p,q) process. The vectors $\boldsymbol{\tau}_n^* = (\beta_1^* + \alpha_1^*\epsilon_n^2, \beta_2^*, \ldots, \beta_{q-1}^*) \in \mathbb{R}^{q-1}$ and $\boldsymbol{\alpha}^* = (\alpha_2^*, \ldots, \alpha_{p-1}^*) \in \mathbb{R}^{p-2}$ are defined as $\boldsymbol{\tau}_n$ and $\boldsymbol{\alpha}$ in Section 2, but α_i and β_i are replaced by α_i^* and β_j^*. Similarly to A_n we define

$$A_n^* = \begin{bmatrix} \boldsymbol{\tau}_n^* & \beta_q^* & \alpha^* & \alpha_p^* \\ \mathbf{I}_{q-1} & 0 & 0 & 0 \\ \boldsymbol{\xi}_n & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I}_{p-2} & 0 \end{bmatrix}.$$

The top Lyapunov exponent of the sequence A_n^*, $-\infty < n < \infty$ is denoted by γ_L^*. As in Section 2, the equations (3.25) and (3.26) have a unique stationary solution if and only if

$$\gamma_L^* < 0, \tag{3.27}$$

assuming that

$$E(\log\|A_0^*\|) < \infty. \tag{3.28}$$

Similarly to $\ell_i(\mathbf{u})$ we define

$$g_i(\mathbf{u}) = -\frac{1}{2}\left\{\log v_i(\mathbf{u}) + \frac{z_i^2}{v_i(\mathbf{u})}\right\},$$

where

$$v_i(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le j<\infty} c_j(\mathbf{u}) z_{j-i}^2.$$

Hence

$$g_i'(\mathbf{u}) = \frac{1}{2}\left(\epsilon_i^2\frac{\gamma_i^2}{v_i(\mathbf{u})}\frac{v_i'(\mathbf{u})}{v_i(\mathbf{u})} - \frac{v_i'(\mathbf{u})}{v_i(\mathbf{u})}\right).$$

Let

$$\bar{g}(\mathbf{u}) = Eg_0(\mathbf{u}).$$

We assume that

$$\bar{g}'(\boldsymbol{\theta}) \ne \mathbf{0} \tag{3.29}$$

and

there is U^*, a neighborhood of $\boldsymbol{\theta}$, such that $\bar{g}(\mathbf{u})$, $\bar{g}'(\mathbf{u})$, $\bar{g}''(\mathbf{u})$ exist and are continuous for each $\mathbf{u} \in U^*$. (3.30)

We note that under some regularity conditions $\bar{g}'(\mathbf{u}) = \mathbf{0}$ if and only if $\mathbf{u} = \boldsymbol{\theta}^*$ (cf. Berkes et al., 2003, Lemma 5.5). We also assume that

$$k^* = k^*(m) \quad \text{and} \quad \limsup_{m\to\infty}(k^*/m) < \infty \tag{3.31}$$

and

$$\sup_{0<t<\infty} b(t) < \infty. \tag{3.32}$$

THEOREM 3.2. *If (2.8), (2.10), (2.13)–(2.19), (3.23), (3.24), and (3.25)–(3.32) hold, then*

$$\lim_{m\to\infty} P_{H_A}\{k_m < \infty\} = 1.$$

The proof of Theorem 3.2 is presented in Section 7.

4. SIMULATIONS

In this section we report the results of a simulation study intended to assess the behavior of the procedure in finite samples. Even before we commenced the numerical experiments, it was clear to us that much larger sample sizes than those considered by Chu et al. (1996) would be needed for the asymptotic behavior to manifest itself. There are two main reasons for this. First, the required model estimation in Chu et al. consists essentially in estimating the mean of normal variables, and it is well known that already for samples of size 50 such estimates are very accurate. In our context, we have to estimate the parameters of a GARCH process via nonlinear optimization. It is well known that although these estimates are optimal when the innovations are normal (the case we considered in the simulation study), they may have large biases and standard errors even for samples as large as 1,000. For a sample size of 1,000, these estimators are accurate for a wide range of parameters, but for some choices of parameters they exhibit large biases. Second, our procedure requires the estimation of the covariance matrix $\hat{\mathbf{D}}_m$. Because there is no close formula for $\hat{\mathbf{D}}_m$, it would have to be estimated even if the parameters were known. We conducted a number of experiments, not reported here in a systematic way, in which we used the exact values of the GARCH parameters rather than their estimates. Even in this situation, samples of size about 1,000 are required to obtain relatively stable estimates of $\hat{\mathbf{D}}_m$.

We now proceed with a detailed description of our simulation study and the conclusions it leads to. We focused on the popular GARCH(1,1) models and considered a wide range of the parameters ω, α_1, and β_1. The GARCH models were simulated and estimated using the S+ module GARCH, whereas for the estimation of $\hat{\mathbf{D}}_m$ it was necessary to write a much faster C++ code and interface it with S+.

We report the results for three GARCH(1,1) models:

Model I: $\omega = 0.05$, $\alpha_1 = 0.4$, $\beta_1 = 0.3$;
Model II: $\omega = 0.05$, $\alpha_1 = 0.5$, $\beta_1 = 0.0$;
Model III: $\omega = 1.0$, $\alpha_1 = 0.3$, $\beta_1 = 0.2$.

The results for these three models are fairly representative of the overall conclusions.

To facilitate the graphical presentation of the results, we work with the normalized decision function

$$C(k) = \left(1 + \frac{k}{m}\right)^{-1} \frac{1}{\sqrt{m}} \left| \left[\sum_{m<i\le m+k} l_i'(\hat{\theta}_m)\hat{\mathbf{D}}_m^{-1/2} \right] \right|, \qquad 1 \le k < \infty.$$

A change in parameters is signaled if $C(k) > c(\alpha)$, where $c(\alpha)$ satisfies

$$1 - P\left(\sup_{0<t<1} |W(t)| \le c(\alpha) \right)^3 = \alpha. \tag{4.33}$$

TABLE 1. Critical values calculated according to relation (4.1)

α	0.10	0.05	0.01
$c(\alpha)$	2.382	2.632	3.150

Table 1 gives the critical values for the conventional significance levels.

In Table 2 we report the empirical rejection probabilities of the null hypothesis of no change in the model parameters assuming this hypothesis is true. It is seen that for models with pronounced GARCH characteristics, i.e., parameters α_1 and β_1 much larger than ω, the procedure has correct size for monitoring horizons of about 500 time units with the 10% bound being somewhat more reliable. By a monitoring horizon we understand here the length of time we are willing to use the procedure without updating the parameters. We note that the theory developed in this paper shows that the empirical size tends to the nominal size as $m \to \infty$, so for any finite m size distortions will be present. This is particularly visible if the GARCH parameters are difficult to estimate, as in Model III (the process looks more like a white noise); the procedure has a high probability of type I error. We conjecture that in such situations m much larger than 1,000 would be required to obtain empirical size close to the nominal size. Using the true values of α_1, β_1, and ω, leads to entries about half the size of those reported in Table 2. With $m = 1{,}000$, the method is not accurate for monitoring horizons longer than 500 and cannot be used in an automatic way. As we mention in the discussion toward the end of this section, a visual real-time inspection of the graph of $C(k)$ following an alarm (critical level exceeded) might indicate that there is no reason to suspect a change in model parameters

TABLE 2. Empirical sizes for monitoring horizons k

k	100	200	300	400	500	600	700	800	900	1,000
Model I: $\omega = 0.05$, $\alpha_1 = 0.4$, $\beta_1 = 0.3$										
$\alpha = 0.10$	0.001	0.008	0.021	0.045	0.072	0.099	0.132	0.146	0.165	0.183
$\alpha = 0.05$	0.001	0.003	0.009	0.028	0.052	0.070	0.095	0.112	0.127	0.139
Model II: $\omega = 0.05$, $\alpha_1 = 0.5$, $\beta_1 = 0.0$										
$\alpha = 0.10$	0.013	0.027	0.041	0.054	0.068	0.080	0.099	0.116	0.129	0.137
$\alpha = 0.05$	0.013	0.026	0.038	0.051	0.066	0.076	0.090	0.109	0.120	0.128
Model III: $\omega = 1.0$, $\alpha_1 = 0.3$, $\beta_1 = 0.2$										
$\alpha = 0.10$	0.047	0.146	0.237	0.329	0.407	0.465	0.519	0.561	0.593	0.616
$\alpha = 0.05$	0.029	0.101	0.190	0.256	0.334	0.390	0.440	0.483	0.516	0.539

Note: Simulations were done with $m = 1{,}000$ and are based on 2,000 replications.

(see Figure 3). Alternatively, finite-sample corrections could be obtained by simulation for specific values of m and monitoring horizons of interest. Such simulations would be specific to a problem at hand and have not been conducted.

The power of the procedure for three change-point scenarios is reported in Table 3. As can be expected, large changes in parameters are detected more reliably.

From a practical point of view, it is more useful to study the distribution of the detection delay time or, equivalently, the distribution of the random time when the decision function $C(k)$ first exceeds a critical level. In Table 4, we report selected descriptive statistics for such distributions. The estimated densities are depicted in Figure 1.

Focusing first on the first three change-point models reported in Table 4, we note that the distribution of the delay time is fairly symmetric but its spread increases as the change point moves further away from the point where the monitoring was initiated. Similar findings were reported in Chu et al. (1996). However, unlike for the fluctuation monitoring scheme investigated in Chu et al., the average delay time does not appear to increase with the distance of the change point from the initiation point, and it is about 20 for a change from Model I to Model III. For relatively less significant changes in parameters, such as the change from Model I to Model II, the delay time is much longer. Even in such situations, however, a visual real-time inspection of the graph of $C(k)$ may suggest that something is happening to the parameters of the model. In the panel in the right-bottom corner of Figure 2, five randomly selected trajectories of $C(k)$ for the change from Model I to Model II are shown. A picture of this type may be fairly typical in real-data applications as the parameters need

TABLE 3. Empirical power of the test

k	100	200	300	400	500	600	700	800	900	1,000
Model I changes to Model III at $k^* = 50$										
$\alpha = 0.10$	0.975	1	1	1	1	1	1	1	1	1
$\alpha = 0.05$	0.962	1	1	1	1	1	1	1	1	1
Model I changes to Model III at $k^* = 250$										
$\alpha = 0.10$	0.002	0.004	0.947	1	1	1	1	1	1	1
$\alpha = 0.05$	0.000	0.003	0.932	1	1	1	1	1	1	1
Model I changes to Model III at $k^* = 500$										
$\alpha = 0.10$	0.000	0.009	0.020	0.037	0.055	0.999	1	1	1	1
$\alpha = 0.05$	0.000	0.005	0.012	0.024	0.034	0.998	1	1	1	1
Model I changes to Model II at $k^* = 250$										
$\alpha = 0.10$	0.002	0.004	0.017	0.071	0.252	0.523	0.735	0.859	0.927	0.966
$\alpha = 0.05$	0.000	0.003	0.008	0.045	0.177	0.407	0.646	0.796	0.892	0.946

Note: For each k, the fraction of trajectories crossing the critical levels from Table 1 is reported. Simulations were done with $m = 1{,}000$ and are based on 1,000 replications.

TABLE 4. Elementary statistics for the distribution of the first exceedance of the 10% critical level

Mean	SE	Min	Q1	Med	Q3	Max
Model I changes to Model III at $k^* = 50$						
71.4	12.4	50	63	71	79	135
Model I changes to Model III at $k^* = 250$						
272.3	18.1	89	262	271	282	338
Model I changes to Model III at $k^* = 500$						
516.4	54.7	121	511	523	538	618
Model I changes to Model II at $k^* = 250$						
612.9	166.5	89	498	589	710	1,000

Note: The estimated densities are graphed in Figure 1. Simulations were done with $m = 1{,}000$ and are based on 1,000 replications.

not switch immediately into a new regime but may evolve gradually through a number of smaller changes. In contrast, as shown in Figure 3, if there is no change, the trajectory of $C(k)$ may occasionally exceed the critical value, but it will not show a pronounced upward trend such as that manifest in Figure 2.

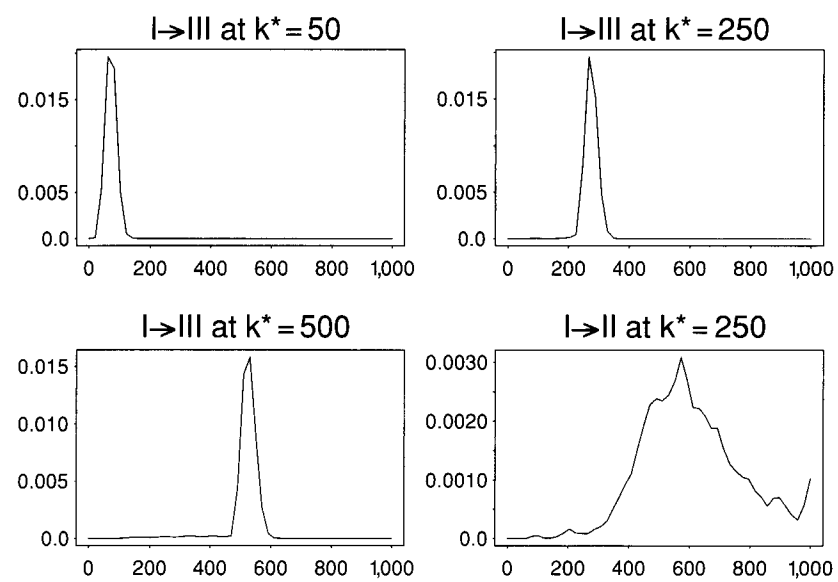

FIGURE 1. Estimated densities of the first exceedance of the 10% critical level. Estimates were obtained using the cosine kernel with support of length 70. Simulations were done with $m = 1{,}000$ and are based on 1,000 replications.

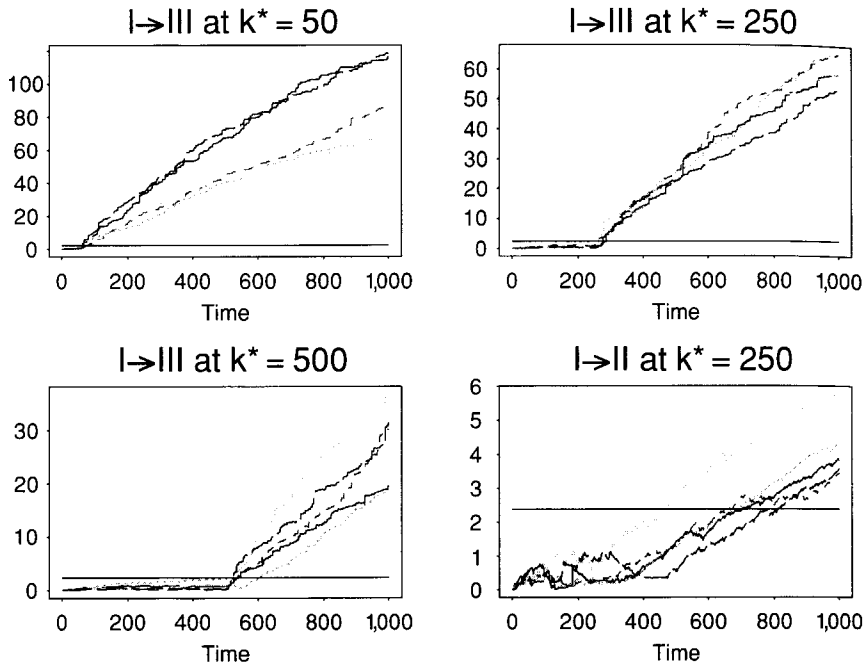

FIGURE 2. Five randomly selected realizations of the sequence $C(k)$ for the data summarized in Table 4. The vertical lines correspond to 10% critical values. The corresponding estimated densities of the first hitting time are depicted in Figure 1.

5. PROOF OF PROPOSITION 3.1

Let

$$\hat{\mathbf{D}}_m(\mathbf{u}) = \frac{1}{m} \sum_{1<i\le m} (\hat{\ell}_i'(\mathbf{u}))^T \hat{\ell}_i'(\mathbf{u})$$

and

$$\mathbf{D}_m(\mathbf{u}) = \frac{1}{m} \sum_{1\le i\le m} (\ell_i'(\mathbf{u}))^T \ell_i'(\mathbf{u}).$$

In the proof of their Lemma 5.8, Berkes et al. (2003) show that there is a constant $0 < \varrho < 1$ and a positive random variable ξ such that

$$\sup_{\mathbf{u}\in U} |\hat{\ell}_k'(\mathbf{u}) - \ell_k'(\mathbf{u})| \le \xi \varrho^k. \tag{5.34}$$

Berkes et al. (2003) also show that $\sup_{\mathbf{u}\in U} |\ell_i(\mathbf{u})|$ is a stationary sequence and

$$E \sup_{\mathbf{u}\in U} |\ell_0(\mathbf{u})| < \infty.$$

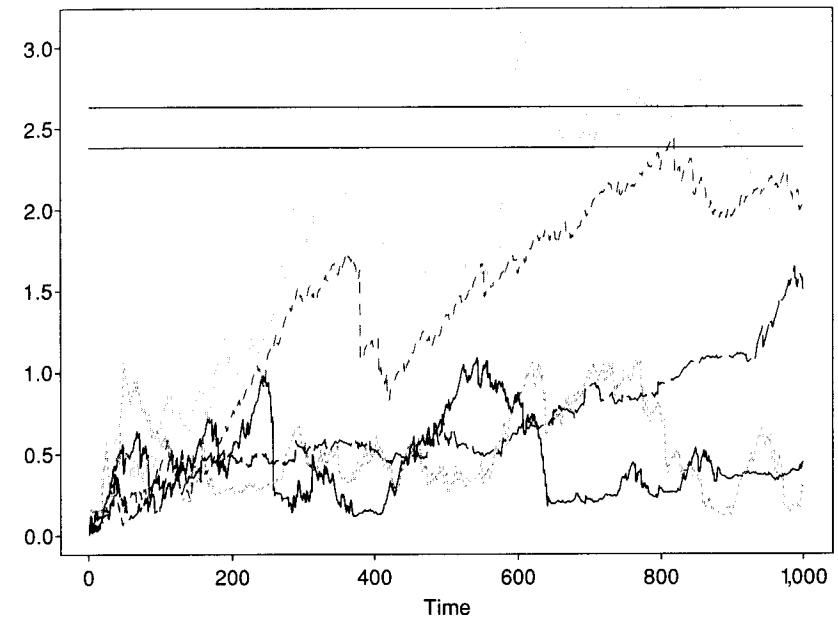

FIGURE 3. Five randomly selected realizations of the sequence $C(k)$ for Model 1 and $m = 1{,}000$. The two horizontal lines correspond to 5 and 10% critical values from Table 1. The fractions of trajectories crossing these lines before time k are reported in Table 2.

Hence

$$\sup_{\mathbf{u}\in U} |\hat{\mathbf{D}}_m(\mathbf{u}) - \mathbf{D}_m(\mathbf{u})| \le \frac{1}{m}\left(2\xi \sum_{1<i\le m} \rho^i \sup_{\mathbf{u}\in U} |\ell_i(\mathbf{u})| + \xi^2 \sum_{1<i\le m} \rho^{2i}\right)$$
$$\overset{\text{a.s.}}{=} O\left(\frac{1}{m}\right) \tag{5.35}$$

by Lemma 2.2 of Berkes et al. (2003).

Next we show that

$$E \sup_{\mathbf{u}\in U} |(\ell_0'(\mathbf{u}))^T \ell_0'(\mathbf{u})| < \infty. \tag{5.36}$$

By (6.49) we have that

$$\sup_{\mathbf{u}\in U} |(\ell_0'(\mathbf{u}))^T \ell_0(\mathbf{u})| \le \sup_{\mathbf{u}\in U} |\ell'(\mathbf{u})|^2$$
$$\le \epsilon_0^4 \left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})}\right)^2 \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})}\right)^2 + \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})}\right)^2.$$

Using the independence of ϵ_0 and $(w_0(\mathbf{u}), \sigma_0^2)$, $\mathbf{u} \in U$ we get that

$$E\epsilon_0^4 \left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})} \right)^2 \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} \right)^2$$
$$= E\epsilon_0^4 E \left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})} \right)^2 \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} \right)^2$$
$$\le E\epsilon_0^4 E \left(E \left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})} \right)^{2+\delta/2} \right)^{2/(2+\delta/2)} \left(E \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} \right)^{2(4+\delta)/\delta} \right)^{\delta/(4+\delta)}.$$

Lemma 5.1 of Berkes et al. (2003) yields that

$$E \left(\sup_{\mathbf{u}\in U} \frac{\sigma_0^2}{w_0(\mathbf{u})} \right)^{2+\delta/2} < \infty, \tag{5.37}$$

and Lemma 3.6 of Berkes and Horváth (2004) gives

$$E \left(\sup_{\mathbf{u}\in U} \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} \right)^{2(4+\delta)/\delta} < \infty.$$

Hence the proof of (5.36) is complete. For each $\mathbf{u} \in U$, $\ell_i'(\mathbf{u})$ is a stationary and ergodic sequence. So by (5.36) we can use the ergodic theorem, resulting in

$$\mathbf{D}_m(\mathbf{u}) \to \mathbf{D}(\mathbf{u}) \quad \text{a.s. for any } \mathbf{u} \in U. \tag{5.38}$$

Next we show that there are a constant C_1 and $U^* \subseteq U$, a neighborhood of $\boldsymbol{\theta}$, such that

$$|E(\ell_0'(\mathbf{u}))^T \ell_0'(\mathbf{u}) - E(\ell_0'(\mathbf{v}))^T \ell_0'(\mathbf{v})| \le C_1 |\mathbf{u} - \mathbf{v}|, \quad \text{if } \mathbf{u}, \mathbf{v} \in U^*. \tag{5.39}$$

Using (6.49) we can write

$$E \sup \left\{ \frac{1}{|\mathbf{u} - \mathbf{v}|} |(\ell_0'(\mathbf{u}))^T (\ell_0'(\mathbf{u}) - \ell_0'(\mathbf{v}))| : \mathbf{u}, \mathbf{v} \in U^* \right\} \le A_1 + \cdots + A_4, \tag{5.40}$$

where

$$A_1 = E\epsilon_0^4 E \sup \left\{ \frac{1}{|\mathbf{u} - \mathbf{v}|} \frac{\sigma_0^2}{w_0(\mathbf{u})} \left| \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right)^T \sigma_0^2 \left(\frac{w_0'(\mathbf{u})}{w_0^2(\mathbf{u})} - \frac{w_0'(\mathbf{v})}{w_0^2(\mathbf{v})} \right) \right| : \mathbf{u}, \mathbf{v} \in U^* \right\},$$

$$A_2 = E\epsilon_0^2 E \sup \left\{ \frac{1}{|\mathbf{u} - \mathbf{v}|} \left| \frac{\sigma_0^2}{w_0(\mathbf{u})} \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right)^T \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} - \frac{w_0'(\mathbf{v})}{w_0(\mathbf{v})} \right) \right| : \mathbf{u}, \mathbf{v} \in U^* \right\},$$

$$A_3 = E\epsilon_0^2 E \sup \left\{ \frac{1}{|\mathbf{u} - \mathbf{v}|} \left| \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right)^T \sigma_0^2 \left(\frac{w_0'(\mathbf{u})}{w_0^2(\mathbf{u})} - \frac{w_0'(\mathbf{v})}{w_0^2(\mathbf{v})} \right) \right| : \mathbf{u}, \mathbf{v} \in U^* \right\},$$

and

$$A_4 = E \sup \left\{ \frac{1}{|\mathbf{u} - \mathbf{v}|} \left| \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right)^T \left(\frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} - \frac{w_0'(\mathbf{v})}{w_0(\mathbf{v})} \right) \right| : \mathbf{u}, \mathbf{v} \in U^* \right\}.$$

Using the mean value theorem coordinate-wise we get that

$$A_1 \le C_2 E \left[\left(\sup \left\{ \frac{\sigma_0^2}{w_0(\mathbf{u})}, \mathbf{u} \in U \right\} \right)^2 \right.$$
$$\times \sup \left\{ \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} : \mathbf{u} \in U \right\} \sup \left\{ \frac{w_0(\mathbf{v})}{w_0(\mathbf{u})} ; \mathbf{u}, \mathbf{v} \in U^* \right\}$$
$$\left. + \left(\sup \left\{ \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} : \mathbf{u} \in U \right\} \right)^2 \left(\sup \left\{ \frac{w_0(\mathbf{v})}{w_0(\mathbf{u})} : \mathbf{u}, \mathbf{v} \in U^* \right\} \right)^2 \right].$$

By the Hölder inequality we have

$$A_1 \le C_2 \left\{ E \left(\sup \left\{ \frac{\sigma_0^2}{w_0(\mathbf{u})} \mathbf{u} \in U \right\} \right)^{2+\delta/2} \right\}^{2/(4+\delta)}$$
$$\times \left(E \left\{ \left(\sup \left\{ \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} : \mathbf{u} \in U \right\} \right)^2 \right.\right.$$
$$\left.\left. \times \left(\sup \frac{w_0(\mathbf{v})}{w_0(\mathbf{u})} : \mathbf{u}, \mathbf{v} \in U^* \right)^2 + 1 \right\}^{(4+\delta)/\delta} \right)^{\delta/(4+\delta)}. \tag{5.41}$$

By (5.37), the first expected value is finite in (5.41). Using the Cauchy inequality we get that

$$E \left[\left\{ \sup \left(\frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} : \mathbf{u} \in U \right) \right\}^{(8+2\delta)/\delta} \left\{ \sup \left(\frac{w_0(\mathbf{v})}{w_0(\mathbf{u})} : \mathbf{u}, \mathbf{v} \in U^* \right) \right\}^{(8+2\delta)/\delta} \right]$$
$$\le \left\{ E \left(\sup \frac{|w_0'(\mathbf{u})|}{w_0(\mathbf{u})} : \mathbf{u} \in U \right)^{(16+4\delta)/\delta} \right.$$
$$\left. \times E \left(\sup \left(\frac{w_0(\mathbf{v})}{w_0(\mathbf{u})} : \mathbf{u}, \mathbf{v} \in U^* \right) \right)^{(16+4\delta)/\delta} \right\}^{1/2}. \tag{5.42}$$

The first expected value on the right-hand side of (5.42) is finite according to Lemma 3.6 of Berkes and Horváth (2004). The second expected value on the right-hand side of (5.42) is finite by Lemma 3.7 of Berkes and Horváth (2004) assuming that U^* is a small enough neighborhood of $\boldsymbol{\theta}$. Hence

$$A_1 < \infty. \tag{5.43}$$

Similar arguments show that

$$A_i < \infty, \qquad i = 2,3,4, \tag{5.44}$$

implying that

$$E \sup\left\{ \frac{1}{|\mathbf{u}-\mathbf{v}|} |(\ell_0'(\mathbf{u}))^T(\ell_0'(\mathbf{u}) - \ell_0'(\mathbf{v}))| : \mathbf{u},\mathbf{v} \in U^* \right\} < \infty, \tag{5.45}$$

assuming that U^* is a small enough neighborhood of $\boldsymbol{\theta}$. By symmetry, we have that

$$E \sup\left\{ \frac{1}{|\mathbf{u}-\mathbf{v}|} |(\ell_0'(\mathbf{u}) - \ell_0'(\mathbf{v}))^T\ell_0'(\mathbf{v})| : \mathbf{u},\mathbf{v} \in U^* \right\} < \infty. \tag{5.46}$$

Hence the proof of (5.39) is complete.

We note that

$$\sup_{\mathbf{u},\mathbf{v}\in U^*} \frac{1}{|\mathbf{u}-\mathbf{v}|} |\mathbf{D}_m(\mathbf{u}) - \mathbf{D}_m(\mathbf{v})|$$
$$\le \frac{2}{m} \sum_{1<i\le m} \sup_{\mathbf{u},\mathbf{v}\in U^*} \frac{1}{|\mathbf{u}-\mathbf{v}|} |(\ell_i'(\mathbf{u}))^T\ell_i'(\mathbf{u}) - (\ell_i'(\mathbf{v}))^T\ell_i'(\mathbf{v})|.$$

Because $\sup\{|(\ell_i'(\mathbf{u}))^T\ell_i'(\mathbf{u}) - (\ell_i'(\mathbf{v}))^T\ell_i'(\mathbf{v})|/|\mathbf{u}-\mathbf{v}| : \mathbf{u},\mathbf{v} \in U^*\}$ is a stationary and ergodic sequence with finite mean by (5.39), the ergodic theorem implies that

$$\lim_{m\to\infty} \frac{1}{|\mathbf{u}-\mathbf{v}|} |\mathbf{D}_m(\mathbf{u}) - \mathbf{D}_m(\mathbf{v})| = c^*$$

with some constant c^*. Hence by (5.38) we have that

$$\sup_{\mathbf{u}\in U^*} |\mathbf{D}_m(\mathbf{u}) - \mathbf{D}(\mathbf{u})| \to 0 \quad \text{a.s.} \tag{5.47}$$

Berkes et al. (2003) show that

$$\hat{\boldsymbol{\theta}}_m \to \boldsymbol{\theta} \quad \text{a.s.}, \tag{5.48}$$

and therefore the first part of Proposition 3.1 follows from (5.47).

The nonsingularity of $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$ is proved by Berkes et al. (2003), and the positive definiteness of $\mathbf{D}$ is obvious.

6. PROOF OF THEOREM 3.1

The proof of Theorem 3.1 is based on several lemmas, which we present after introducing some additional notation.

Let

$$\ell_i(\mathbf{u}) = -\frac{1}{2}\left\{\log w_i(\mathbf{u}) + \frac{y_i^2}{w_i(\mathbf{u})}\right\},$$

where

$$w_i(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le j<\infty} c_j(\mathbf{u}) y_{j-i}^2.$$

We note that

$$\ell_i'(\mathbf{u}) = \frac{1}{2}\left(\epsilon_i^2 \frac{\sigma_i^2}{w_i(\mathbf{u})} \frac{w_i'(\mathbf{u})}{w_i(\mathbf{u})} - \frac{w_i'(\mathbf{u})}{w_i(\mathbf{u})}\right), \tag{6.49}$$

and we define

$$T_m(k,\mathbf{u}) = \sum_{m<i\le m+k} \ell_i'(\mathbf{u}), \qquad 1 \le k < \infty.$$

LEMMA 6.1. *If the conditions of Theorem 3.1 are satisfied, then*

$$\sup_{1\le k<\infty} \left| \sum_{m<i\le m+k} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m) - T_m(k,\hat{\boldsymbol{\theta}}_m) \right| \Big/ \left(m^{1/2}\left(1+\frac{k}{m}\right) b\left(\frac{k}{m}\right)\right) \overset{a.s.}{=} O(m^{-1/2}),$$

as $m \to \infty$.

Proof. By Lemmas 5.8 and 5.9 of Berkes et al. (2003) we have that

$$\sup_{\mathbf{u}\in U} \left| \sum_{1\le i\le n} (\hat{\ell}_i'(\mathbf{u}) - \ell_i'(\mathbf{u})) \right| \overset{a.s.}{=} O(1),$$

implying Lemma 6.1. ■

Let

$$\mathbf{Z}(\mathbf{u}) = E\ell_0''(\mathbf{u}).$$

LEMMA 6.2. *If the conditions of Theorem 3.1 are satisfied, then there is* U^*, *a neighborhood of* $\boldsymbol{\theta}$, *such that*

$$\sup_{\mathbf{u}\in U^*} \left| \frac{1}{n} \sum_{1\le i\le n} \ell_i''(\mathbf{u}) - \mathbf{Z}(\mathbf{u}) \right| \to 0 \quad a.s.$$

Proof. This is Lemma 5.6 in Berkes et al. (2003). ■

LEMMA 6.3. *If the conditions of Theorem 3.1 are satisfied, then*

$$\sup_{1\le k<\infty} \frac{\left|\sum_{m<i\le m+k} \ell_i'(\hat{\boldsymbol{\theta}}_m) - \left(\sum_{m<i\le m+k} \ell_i'(\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})k\mathbf{Z}(\boldsymbol{\theta})\right)\right|}{m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)} = o_P(1),$$

as $m \to \infty$.

Proof. First we show that there is a neighborhood of $\boldsymbol{\theta}$, say, U^*, such that

$$\sup_{1\le k<\infty} \sup_{\mathbf{u}\in U^*} \frac{\left|\sum_{m<i\le m+k} \ell_i''(\mathbf{u}) - k\mathbf{Z}(\mathbf{u})\right|}{m\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)} = o_P(1), \tag{6.50}$$

as $m \to \infty$. Because $\ell_i''(\mathbf{u})$, $\mathbf{u} \in U^*$ is a stationary sequence, by (3.24) it is enough to prove that

$$\sup_{1\le k<\infty} \sup_{\mathbf{u}\in U^*} \frac{\left|\sum_{k<i\le m+k} \ell_i''(\mathbf{u}) - k\mathbf{Z}(\mathbf{u})\right|}{m+k} = o_P(1), \quad \text{as } m \to \infty. \tag{6.51}$$

However, (6.51) is an immediate consequence of Lemma 6.2. Theorem 4.4 of Berkes et al. (2003) implies that

$$|\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}| = O_P(m^{-1/2}). \tag{6.52}$$

Using the mean value theorem coordinate-wise for $\sum \ell_i'(\hat{\boldsymbol{\theta}}_m)$ and then (6.50), (6.52) for the coordinates of $\sum \ell_i''(\mathbf{u})$ we get Lemma 6.3. ■

LEMMA 6.4. *If (2.8), (2.9), and (2.13)–(2.19) are satisfied, then*

$$\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta} = -\frac{1}{m}\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\mathbf{Z}^{-1}(\boldsymbol{\theta})(1+o(1)) \quad a.s.$$

as $m \to \infty$.

Proof. Lemma 6.4 follows from Theorem 4.4 of Berkes et al. (2003). ■

LEMMA 6.5. *If the conditions of Theorem 3.1 are satisfied, then*

$$\sup_{1\le k<\infty} \frac{\left|\sum_{m<i\le m+k} \ell_i'(\hat{\boldsymbol{\theta}}_m) - \left(\sum_{m<i\le m+k} \ell_i'(\boldsymbol{\theta}) - \frac{k}{m}\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\right)\right|}{m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)} = o_P(1),$$

as $m \to \infty$.

Proof. We note that

$$\max_{1<k\le m} \frac{\left|\frac{k}{m}\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\right|}{m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)}$$

$$\le \sup_{1\le k<\infty} \frac{\frac{k}{m}}{1+\frac{k}{m}} \sup_{1\le j<\infty} \frac{1}{b\left(\frac{j}{k}\right)} m^{-1/2}\left|\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\right| = O_P(1),$$

by conditions (3.23), (3.24), and the asymptotic normality of $m^{-1/2} \times \sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})$ (cf. Berkes et al., 2003). Hence putting together Lemmas 6.3 and 6.4 we get the result in Lemma 6.5. ■

LEMMA 6.6. *If the conditions of Theorem 3.1 are satisfied, then*

$$\sup_{1\le k<\infty} \frac{\left|\sum_{m<i\le m+k} \ell_i'(\boldsymbol{\theta}) - \frac{k}{m}\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\right|}{m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)} \xrightarrow{\mathcal{D}} \sup_{0<t<\infty} \frac{|W_{\mathbf{D}}(1+t) - (1+t)W_{\mathbf{D}}(1)|}{(1+t)b(t)},$$

as $m \to \infty$, *where* $W_{\mathbf{D}}(s)$ *is a Gaussian process with* $EW_{\mathbf{D}}(s) = 0$ *and* $EW_{\mathbf{D}}^T(s)W_{\mathbf{D}}(s') = \min(s,s')\mathbf{D}$.

Proof. As is shown in Berkes et al. (2003), $\ell_i'(\boldsymbol{\theta})$ is a stationary ergodic martingale difference sequence; clearly $\operatorname{cov}(\ell_i'(\boldsymbol{\theta})) = \mathbf{D}$. Hence the Cramér–Wold device (cf. Billingsley, 1968, p. 206) yields that for any $T > 0$

$$m^{-1/2}\sum_{1\le i\le mt} \ell_i'(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}[0,T]} W_{\mathbf{D}}(t) \quad \text{as } m \to \infty. \tag{6.53}$$

Hence

$$m^{-1/2}\left\{\sum_{m<i\le m(1+t)} \ell_i'(\boldsymbol{\theta}) - t\sum_{1\le i\le m} \ell_i'(\boldsymbol{\theta})\right\} \xrightarrow{\mathcal{D}[0,T]} W_{\mathbf{D}}(1+t) - (1+t)W_{\mathbf{D}}(1) \tag{6.54}$$

for any $T > 0$ as $m \to \infty$. By the Hájek–Rényi–Chow inequality (cf. Chow, 1960) we have

$$\lim_{T\to\infty} \limsup_{m\to\infty} P\left\{\sup_{mT\le k<\infty} \left|\sum_{1\le i\le m+k} \ell_i'(\boldsymbol{\theta})\right| \Big/ \left(m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)\right) \ge x\right\} = 0 \tag{6.55}$$

for any $x > 0$. The coordinates of $W_{\mathbf{D}}(t)$ are Brownian motions, so by the law of the iterated logarithm and (3.24) we have

$$\sup_{T\le t<\infty} \frac{|W_{\mathbf{D}}(1+t)|}{(1+t)b(t)} \to 0 \quad \text{a.s. } (T\to\infty). \tag{6.56}$$

Lemma 6.6 now follows from (6.53)–(6.56). ■

Proof of Theorem 3.1. Putting together Lemmas 6.1–6.6 we get that

$$\sup_{1\le k<\infty} \frac{\left|\sum_{m<i\le m+k} \ell_i'(\hat{\boldsymbol{\theta}}_m)\hat{\mathbf{D}}_m^{-1/2}\right|}{m^{1/2}\left(1+\frac{k}{m}\right)b\left(\frac{k}{m}\right)} \xrightarrow{\mathcal{D}} \sup_{0<t<\infty} \frac{|(W_{\mathbf{D}}(1+t)-(1+t)\mathbf{W}_{\mathbf{D}}(1))\mathbf{D}^{-1/2}|}{(1+t)b(t)}.$$

Elementary arguments show that

$$E\{(W_{\mathbf{D}}(1+t)-(1+t)\mathbf{W}_D(1))\mathbf{D}^{-1/2}\}^T\{(W_{\mathbf{D}}(1+s)-(1+s)W_{\mathbf{D}}(1))D^{-1/2}\}$$
$$= t(1+s)\mathbf{I}_{p+q+1}, \qquad t\le s,$$

where $\mathbf{I}_{p+q+1}$ is the identity matrix in $\mathbb{R}^{p+q+1}$. Computing the covariances one can verify that

$$\{(W_{\mathbf{D}}(1+t)-(1+t)\mathbf{W}_{\mathbf{D}}(1))\mathbf{D}^{-1/2}, t\ge 0\}$$
$$\stackrel{\mathcal{D}}{=} \left\{(1+t)W_1\left(\frac{t}{1+t}\right),\dots,(1+t)W_{p+q+1}\left(\frac{t}{1+t}\right), t\ge 0\right\},$$

where $W_1, W_2, \dots, W_{p+q+1}$ are independent Wiener processes. Hence

$$\sup_{0<t<\infty} \frac{|(W_{\mathbf{D}}(1+t)-(1+t)W_{\mathbf{D}}(1))\mathbf{D}^{-1/2}|}{(1+t)b(t)} \stackrel{\mathcal{D}}{=} \max_{1\le i\le p+q+1} \sup_{0<s<1} \frac{|W_i(s)|}{b\left(\frac{s}{1-s}\right)},$$

completing the proof of Theorem 3.1. ■

7. PROOF OF THEOREM 3.2

By Proposition 3.1 it is enough to show that

$$\sup_{1\le k<\infty} \left|\sum_{m<i\le m+k} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m)\mathbf{D}^{-1/2}\right| \Big/ \left(m^{1/2}\left(1+\frac{k}{m}\right)\right) \xrightarrow{P} \infty, \tag{7.57}$$

as $m\to\infty$. Theorem 3.1 yields that

$$\sup_{1\le k\le k^*-m} \left|\sum_{m<i\le m+k} \hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m)\mathbf{D}^{-1/2}\right| \Big/ \left(m^{1/2}\left(1+\frac{k}{m}\right)\right) = O_P(1),$$

as $m\to\infty$. Let $\mathbf{X}_n = (\sigma_{n+1}^2,\dots,\sigma_{n-p+2}^2, y_n^2,\dots,y_{n-q+2}^2)^T \in \mathbb{R}^{p+q-1}$ and $\mathbf{d} = (\omega^*,0,\dots,0)^T$. Using (2.9) and (2.10) we get that

$$\mathbf{X}_{n+1} = A_{n+1}^*\mathbf{X}_n + \mathbf{d} \quad \text{if } n\ge k^*,$$

and induction yields

$$\mathbf{X}_{k^*+k} = A_{k^*+k}^*\dots A_{k^*+1}^*\mathbf{X}_{k^*} + \left(\sum_{0\le i\le k-1} A_{k^*+k}^*A_{k^*+k-1}^*\dots A_{k^*+k-i}^* + \mathbf{I}_{p+q-1}\right)\mathbf{d}.$$

Condition (3.25) and the independence of the matrices A_j^* yield that there is a constant $0<\varrho_*<1$ such that

$$|A_{k^*+k}^*\dots A_{k^*+1}^*\mathbf{X}_{k^*}| = O_P(\varrho_*^k).$$

Thus

$$|y_{k^*+k}^2 - z_{k^*+k}^2| = O_P(\varrho_*^k) \tag{7.58}$$

and

$$|\sigma_{k^*+k}^2 - \gamma_{k^*+k}^2| = O_P(\varrho_*^k). \tag{7.59}$$

Hence following the proof of Lemma 6.1, one can easily derive from (7.58) and (7.59) that

$$\sup_{k^*-m<k<\infty} \left|\sum_{m<i\le m+k} (\hat{\ell}_i'(\hat{\boldsymbol{\theta}}_m) - g_i'(\hat{\theta}_m))\right| \Big/ \left(m^{1/2}\left(1+\frac{k}{m}\right)\right) = o_P(1).$$

Using the mean value coordinate-wise and the ergodic theorem we get that

$$\sum_{m<i\le m+k} g_i'(\hat{\boldsymbol{\theta}}_m) = \sum_{m<i\le m+k} g_i'(\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})k(\bar{g}''(\boldsymbol{\theta})^T + o_P(1))$$
$$= k(\bar{g}'(\boldsymbol{\theta}) + o_P(1)). \tag{7.60}$$

Choosing any sequence $\tilde{k} = \tilde{k}(m)$ satisfying $\tilde{k}/m^{1/2}\to\infty$ and $\tilde{k} = O(m)$ we get from (7.60)

$$\left|\sum_{m<i\le m+\tilde{k}} g_i'(\hat{\boldsymbol{\theta}}_m)\mathbf{D}^{-1/2}\right| \Big/ \left(m^{1/2}\left(1+\frac{\tilde{k}}{m}\right)\right)$$
$$= |\bar{g}'(\boldsymbol{\theta})\mathbf{D}^{-1/2} + o_P(1)| \frac{\tilde{k}}{m^{1/2}\left(1+\frac{\tilde{k}}{m}\right)} \xrightarrow{P} \infty,$$

because $|\bar{g}'(\boldsymbol{\theta})\mathbf{D}^{-1/2}| \ne 0$ by (3.29).

REFERENCES

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

Berkes, I. & L. Horváth (2004) The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* 32, 633–655.

Berkes, I. & L. Horváth (2003b) Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and Their Applications* 105, 279–298.

Berkes, I., L. Horváth, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.

Besseville, M. & I.V. Nikifirov (1993) *Detection of Abrupt Changes: Theory and Applications.* Prentice Hall.

Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley.

Boldin, M.V. (1998) On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitteilungen aus dem Mathematischen Seminar Giessen* 235, 49–66.

Bougerol, P. & N. Picard (1992a) Strict stationary of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

Bougerol, P. & N. Picard (1992b) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

Brodsky, B.E. & B.S. Darkhovsky (1993) *Nonparametric Methods in Change-Point Problems.* Kluwer.

Brodsky, B.E. & B.S. Darkhovsky (2000) *Non-Parametric Statistical Diagnosis.* Kluwer.

Chow, Y.S. (1960) A martingale inequality and the law of large numbers. *Proceedings of the American Mathematics Society* 11, 107–111.

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 (1997) *Limit Theorems in Change-Point Analysis.* Wiley.

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

Gombay, E. (1996) The weighted sequential likelihood ratio. *Canadian Journal of Statistics* 24, 229–239.

Gombay, E. (2003) Sequential change-point detection and estimation. *Sequential Analysis* 22, 203–222.

Gouriéroux, C. (1997) *ARCH Models and Financial Applications.* Springer-Verlag.

Horváth, L., P. Kokoszka, & G. Teyssière (2001) Empirical process of the squared residuals of an ARCH sequence. *Annals of Statistics* 29, 445–469.

Inclan, C. & G. Tiao (1994) Use of cumulative sums of squares for retrospective detection of change in variance. *Journal of the American Statistical Association* 89, 913–929.

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

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

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

Kokoszka, P. & G. Teyssière (2002) Change Point Detection in GARCH Models: Asymptotic and Bootstrap Tests. Preprint available at http://stat.usu.edu/~piotr/research.html. Utah State University.

Koul, H. (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models.* Springer.

Kulperger, R. & H. Yu (2003) High Moment Partial Sum Processes of Residuals in GARCH Models. Preprint, University of Western Ontario.

Lamoureux, C. & W.D. Lastrapes (1990) Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics* 8, 225–234.

Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 6, 575–596.

Lundbergh, S. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417–435.

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

Nyblom, Y. (1989) Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 6, 223–230.

Sowell, F. (1996) Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 64, 1085–1107.