Testing for parameter constancy in GARCH(p, q) models *

Istvan Berkes[†] Lajos Horváth[‡] Piotr Kokoszka[§]

Abstract

We propose a test for a change in the parameters of a GARCH(p, q) model. The test is based on approximate likelihood scores and does not require the observations to have finite variance. We show that the test has asymptotically correct size under weak assumptions on model errors.

Keywords: Change in parameters; GARCH model; Likelihood scores

1 Introduction

Detecting possible changes in the stochastic structure of a time series has become an important area of research in the last two decades. If the parameters of a model have changed within an observed realization, inference for these parameters will not be valid. Change point detection procedures are important in geophysical sciences, e.g. detection of changes in the structure of a seismogram or changes in climate variables. Detecting changes in the parameters of an industrial process has been another very important application since 1950's. Recently, there is a growing interest in testing for and estimating changes in the parameters of econometric models.

Johnson and Bagshaw (1974) and Bagshaw and Johnson (1975) were perhaps the first to investigate the performance of change point detection tests in the presence of serial correlation. More references can be found in Chapter 4 of Csörgő and Horváth (1997). In finance and economics, a class of ARCH models has gained prominence starting with the seminal work of Engle (1982). The ARCH models which are defined by the equation $y_k = \sigma_k \varepsilon_k$, where y_k are the observations. The innovations ε_k are typically assumed to be a sequence of independent identically distributed random variables. The random variable σ_k depends on the past observations $y_{k-1}, y_{k-2}, \ldots$ and model parameters, and is independent of ε_k. Typically, σ_k is thought of as the conditional variance of y_k given the past, but as we shall see in the following such an interpretation is not always possible. If the y_k follow an ARCH model, they exhibit no serial correlation but their squares or absolute values are correlated. For this reason, ARCH models are very different from the more traditional linear time series models like ARMA and call for different change point detection methods. There is a small but growing literature on change point analysis for ARCH processes. Kokoszka and Leipus (1998, 1999, 2000) and Lavielle

*Research supported by NSF grant INT-0223262 and NATO grant PST.EAP.CLG 980599.

[†] A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, P.O. Box 127, H-1364 Budapest, Hungary.

[‡] Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112-0090, USA.

[§] Corresponding Author. Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA. Partially supported by NSF grant DMS-0413653.

and Moulines (2000) developed procedures applicable to ARCH processes with at least finite second moment. These procedures have recently been refined and applied in more complex applied settings by Andreou and Ghysels (2002, 2003). Lundbergh and Teräsvirta (2002) studied Lagrange multiplier tests which can also be used for detecting a change in parameters of an ARCH model. Mikosch and Stărică (2002) proposed a periodogram-type statistic for testing goodness-of-fit.

The above procedures assume that the observations y_k have at least finite second moment. In many financial applications, however, such an assumption is not realistic, see e.g. Guillaume *et al.* (1997), Embrechts *et al.* (1997) and Rachev and Mittnik (2000). Kokoszka and Teyssière (2002) studied by means of simulations change-point tests which are applicable to ARCH processes which do not have finite second moments. In practice, however, there is also an obvious uncertainty about the distribution of the unobservable ε_k. For example, Horváth *et al.* (2004) analyzed daily returns on foreign exchange rates and confirmed that if a popular GARCH(1,1) model is assumed for the AR(1) residuals, its innovations cannot be normal. It is therefore important to develop change-point detection procedures which are robust to the assumptions of finite second moments of the observations y_k and the normality of the ε_k.

To explain the change point detection procedure we propose, we must first describe how the parameters of ARCH processes are estimated. Assuming for a moment that the density of the ε_k is known to be $h(x)$, the normalized log-likelihood function of the observations $y_1, \dots, y_n$ is

$$L_n(\mathbf{u}) = \frac{1}{n} \sum_{1 \le k \le n} \log \left(\frac{1}{\sigma_k(\mathbf{u})} h \left(\frac{y_k}{\sigma_k(\mathbf{u})} \right) \right),$$

where $\mathbf{u}$ is the vector of model parameters. In practice, $\sigma_k(\mathbf{u})$ is replaced by an approximation based on the available data $y_1, \dots, y_{k-1}$. Details are presented in Section 2. Since in reality the density h is unknown, it is replaced by the standard normal density and the resulting estimates are called pseudo- or quasi- maximum likelihood estimates. Berkes and Horváth (2004) showed that this procedure leads to a loss of efficiency if the distribution of the ε_k has heavier tails than the normal distribution and that using different functions h increases the efficiency. Moreover, Berkes and Horváth (2003) showed that the difference between the quasi maximum likelihood estimator (QMLE) and the true value is of order larger than $n^{-1/2}$ if $E|\varepsilon_0|^{4-\delta} = \infty$ for some $\delta > 0$. Thus, if this moment condition holds, change-point tests based on the differences of the QMLE's will have large type I error. The idea of using non-normal h was also explored by Drost and Klaassen (1997).

In this paper, we show that the theory developed in Berkes and Horváth (2004) allows to construct change-point tests which have asymptotically correct size under very weak assumptions on the function h and on the distribution of the unobservable errors ε_k. We focus on the popular class of GARCH(p, q) model. In Section 2 we state the assumptions, formulate the main result, Theorem 2.1, and provide several examples of useful functions h. The proof of Theorem 2.1 is presented in Section 3.

2 Assumptions and main results

We say that $\{y_k, -\infty < k < \infty\}$ is a GARCH(p, q) process if

$$y_k = \sigma_k \varepsilon_k \tag{2.1}$$

and

$$\sigma_k^2 = \omega + \sum_{1\le i\le p} \alpha_i y_{k-i}^2 + \sum_{1\le j\le q} \beta_j \sigma_{k-j}^2, \tag{2.2}$$

where

$$\omega > 0, \;\; \alpha_i \ge 0, \;\; 1 \le i \le p, \;\; \beta_j \ge 0, \;\; 1 \le j \le q \tag{2.3}$$

are constants and

$$\{\varepsilon_i, -\infty < i < \infty\} \text{ are independent, identically distributed random variables.} \tag{2.4}$$

Bougerol and Picard (1992a,1992b) found sufficient and necessary conditions for the existence of a unique stationary solution to equations (2.1) and (2.2) involving the top Lyapunov exponent. The definitions leading to these conditions are technical and are not used in our proofs, so we replace them by the following assumption

ASSUMPTION 2.1 *The innovations ε_k in (2.1) and the parameters in (2.2) are such that equations (2.1) and (2.2) have a unique strictly stationary solution.*

Under the null hypothesis of no change in the parameters the observations y_k follow (2.1), (2.2). Under the alternative, the parameter $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ changes at an unknown time k^* to the parameter $\boldsymbol{\theta}^* = (\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*)$. More specifically, under the alternative, (2.1) continues to hold, but (2.2) is replaced by

$$\sigma_k^2 = \begin{cases} \omega + \sum_{1\le i\le p} \alpha_i y_{k-i}^2 + \sum_{1\le j\le q} \beta_j \sigma_{k-j}^2, & \text{if } -\infty < k \le k^* \\[2ex] \omega^* + \sum_{1\le i\le p} \alpha_i^* y_{k-i}^2 + \sum_{1\le j\le q} \beta_j^* \sigma_{k-j}^2, & \text{if } k^* < k < \infty. \end{cases} \tag{2.5}$$

We now define an estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ which was shown by Berkes and Horváth (2004) to be strongly consistent and asymptotically normal. For this, we must first define the parameter space whose generic element will be denoted by $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$. If the null hypothesis is true, we denote the true parameter by $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$.

ASSUMPTION 2.2 *The parameter space U is*

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

where $0 < \underline{u} < \overline{u}$, $0 < \rho_0 < 1$, $q\underline{u} < \rho_0$.

We assume that under the null hypothesis

$$\boldsymbol{\theta} \text{ is in the interior of } U. \tag{2.6}$$

Assumptions (2.7) and (2.8) below are needed to uniquely identify the parameter $\boldsymbol{\theta}$.

$$\text{the polynomials } \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_p x^p \text{ and} \tag{2.7}$$
$$1 - \beta_1 x - \beta_2 x^2 - \cdots - \beta_q x^q \text{ are coprimes}$$
$$\text{in the set of polynomials with real coefficients}$$

$$\varepsilon_0^2 \text{ is a nondegenerate random variable.} \tag{2.8}$$

For $\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q) \in U$, define

$$w_k(\mathbf{u}) = x + \sum_{1 \le i \le p} s_i y_{k-i}^2 + \sum_{1 \le j \le q} t_j w_{k-j}(\mathbf{u}), \tag{2.9}$$

and its counterpart which can be computed from the available observations:

$$\hat{w}_k(\mathbf{u}) = x + \sum_{1 \le i \le p} s_i y_{k-i}^2 + \sum_{1 \le j \le q} t_j \hat{w}_{k-j}(\mathbf{u}), \quad k > \max(p, q). \tag{2.10}$$

For $k \le \max(p, q)$, $\hat{w}_k(\mathbf{u})$ can be defined as arbitrary constants. Note that $w_k(\boldsymbol{\theta}) = \sigma_k^2$.

For an arbitrary positive function h define

$$L_n(\mathbf{u}) = \frac{1}{n} \sum_{1 \le i \le n} l_i(\mathbf{u}), \tag{2.11}$$

where

$$l_i(\mathbf{u}) = \log \left\{ \frac{1}{w_i^{1/2}(\mathbf{u})} h(y_i / w_i^{1/2}(\mathbf{u})) \right\}. \tag{2.12}$$

Similarly, we define

$$\hat{L}_n(\mathbf{u}) = \frac{1}{n} \sum_{1 \le i \le n} \hat{l}_i(\mathbf{u}), \tag{2.13}$$

where

$$\hat{l}_i(\mathbf{u}) = \log \left\{ \frac{1}{\hat{w}_i^{1/2}(\mathbf{u})} h(y_i / \hat{w}_i^{1/2}(\mathbf{u})) \right\}. \tag{2.14}$$

The estimator $\hat{\boldsymbol{\theta}}_n$ is defined as follows:

$$\hat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\mathbf{u} \in U} \hat{L}_n(\mathbf{u}).$$

So far all our conditions are related to the structure of the GARCH(p, q) process. The following set of conditions concern the distribution of ε_0 and the smoothness of h. Towards the end of this section we provide several examples of the innovation distributions and appropriate functions h which satisfy our assumptions. We use $|\cdot|$ to denote the maximum norm of vectors and matrices. We assume

$$\lim_{t \to 0} t^{-\mu} P\{\varepsilon_0^2 \le t\} = 0, \qquad \text{with some } \mu > 0, \tag{2.15}$$

$$E|\varepsilon_0^2|^\kappa < \infty, \qquad \text{with some } \kappa > 0. \tag{2.16}$$

Assume also that there is $0 < C_0 < \infty$ such that

$$E|\log h(\varepsilon_0 t)| \le C_0(t^{\nu_0}+1) \text{ for all } t>0 \text{ with some } 0 \le \nu_0 < 2\kappa. \tag{2.17}$$

Let

$$g(y,t) = \log\{th(yt)\}, \qquad -\infty < y < \infty,\ t>0$$

and

$$g_1(y,t) = \frac{\partial}{\partial t} g(y,t), \qquad -\infty < y < \infty,\ t>0.$$

Assume that there is a function $C_1(y)$ such that

$$\begin{aligned} &|g_1(y,t)| \le C_1(y)(t^{\nu_1}+1)/t, \text{ for all } 0<t<\infty \text{ and } y \in \mathbb{R} \\ &\text{with some } 0 \le \nu_1 < 2\kappa \end{aligned} \tag{2.18}$$

and

$$EC_1(\varepsilon_0) < \infty. \tag{2.19}$$

If h is a density, then condition (2.18) means that the density function $th(yt)$ is smooth in the parameter t.

Berkes and Horváth (2004) showed that the condition

$$Eg(\varepsilon_0,t) < Eg(\varepsilon_0,1) \text{ for all } 0<t<\infty,\ t \ne 1. \tag{2.20}$$

implies that

$$L(\mathbf{u}) = E\log\left\{\frac{1}{(w_0(\mathbf{u}))^{1/2}} h(y_0/(w_0(\mathbf{u}))^{1/2})\right\}$$

has a unique maximum in U at $\boldsymbol{\theta}$.

To ensure the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$, Berkes and Horváth (2004) imposed further smoothness conditions on $th(yt)$. Let $g_2(y,t)$ and $g_3(y,t)$ be the second and third derivatives of $g(y,t)$ with respect to t. We assume that there are functions C_2 and C_3 such that

$$\begin{aligned} &|g_2(y,t)| \le C_2(y)(t^{\nu_2}+1)/t^2 \text{ for all } 0<t<\infty \text{ and } y \in \mathbb{R} \\ &\text{with some } 0 \le \nu_2 < \infty \end{aligned} \tag{2.21}$$

$$EC_2(\varepsilon_0) < \infty \tag{2.22}$$

$$\begin{aligned} &|g_3(y,t)| \le C_3(y)(t^{\nu_3}+1)/t^3 \text{ for all } 0<t<\infty \text{ and } y \in \mathbb{R} \\ &\text{with some } 0 \le \nu_3 < \infty \end{aligned} \tag{2.23}$$

and

$$EC_3(\varepsilon_0) < \infty. \tag{2.24}$$

We use $w_k'(\mathbf{u})$ to denote the row vector of the derivatives of $w_k(\mathbf{u})$ and $w_k''(\mathbf{u})$ the matrix of the second order partial derivatives of $w_k(\mathbf{u})$ (the Hessian matrix). Berkes *et al.* (2003) showed that

$$\mathbf{A} = E(w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta}))^T(w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta}))$$

exists and it is nonsingular (^{T} denotes the transpose). We also assume that

$$0 < Eg_1^2(\varepsilon_0, 1) < \infty, \tag{2.25}$$

$$E|g_2(\varepsilon_0, 1)| < \infty \text{ and } Eg_2(\varepsilon_0, 1) \neq 0. \tag{2.26}$$

THEOREM 2.1 *Suppose (2.1)–(2.4), Assumptions 2.1, 2.2 and (2.6)–(2.8), (2.15)–(2.26) are satisfied. Then*

$$\frac{1}{n}\left[\sum_{1\le i\le nt} \hat{l}_i'(\hat{\boldsymbol{\theta}}_n)\right]\mathbf{D}^{-1}\left[\sum_{1\le i\le nt} \hat{l}_i'(\hat{\boldsymbol{\theta}}_n)\right]^T \to \sum_{1\le i\le p+q+1} B_i^2(t) \text{ in } D[0,1],$$

where

$$\mathbf{D} = \mathrm{Cov}[\ l_0'(\boldsymbol{\theta})\ ] = E[\ l_0''(\boldsymbol{\theta})\ ]$$

and $\{B_i(t),\ 0 \le t \le 1\},\ \ i = 1, \ldots, p+q+1,$ *are independent Brownian bridges.*

Theorem 2.1 is proved in Section 3.

The existence of $\mathbf{D}^{-1}$ was established in Lemma 4.5 of Berkes and Horváth (2004). The matrix $\mathbf{D}$ can be estimated by $n^{-1}\sum_{1\le i\le n}[\hat{l}_i'(\hat{\boldsymbol{\theta}}_n)]^T[\hat{l}_i'(\hat{\boldsymbol{\theta}}_n)]$ or by $n^{-1}\sum_{1\le i\le n}\hat{l}_i''(\hat{\boldsymbol{\theta}}_n)$. The critical values for several functionals of $U_m(t) = \sum_{1\le i\le m} B_i^2(t)$ are known. For example, Table 4 on p. 444 of Kiefer (1959) reports the critical values for $\int_0^1 U_m(t)dt$.

We conclude this section by listing several examples of functions h satisfying (2.6)–(2.8) and (2.15)–(2.26), see Section 2 of Berkes and Horváth (2004) for the details. If $h(t) = (2\pi)^{-1/2}\exp(-t^2/2)$, we must assume that $E\varepsilon_0^4 < \infty$. If $h(t) = \exp(-|t|)/2$, then only $E\varepsilon_0^2 < \infty$ is needed. If $h(t) = 1/[2(1+|t|)^2]$, then only (2.16) is required.

3 Proof of Theorem 2.1

The proof of Theorem 2.1 follows from a number of lemmas, as will be explained towards the end of this section. We use the notation introduced in Section 2. By $|\cdot|$ we denote the max norm of vectors or matrices.

LEMMA 3.1 *Under the assumptions of Theorem 2.1,*

$$n^{-1/2}\max_{1\le k\le n}\left|\sum_{1\le i\le k}\hat{l}_i'(\hat{\boldsymbol{\theta}}_n) - \sum_{1\le i\le k} l_i'(\boldsymbol{\theta}) - k(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{D}\right| = o_P(1).$$

PROOF. Lemma 4.6 of Berkes and Horváth (2004) asserts that

$$\sup_{\mathbf{u}\in U}\left|\sum_{1\le i\le n}\hat{l}_i'(\mathbf{u}) - \sum_{1\le i\le n} l_i{}'(\mathbf{u})\right| = O(1) \quad \text{a.s.}$$

Therefore

$$\max_{1\le k\le n}\sup_{\mathbf{u}\in U}\left|\sum_{1\le i\le k}\hat{l}_i'(\mathbf{u}) - \sum_{1\le i\le k} l_i'(\mathbf{u})\right| = O(1) \quad \text{a.s.} \tag{3.1}$$

Hence it is enough to prove that

$$n^{-1/2} \max_{1\le k\le n} \left| \sum_{1\le i\le k} l_i{}'(\hat{\boldsymbol{\theta}}_n) - \sum_{1\le i\le k} l_i{}'(\boldsymbol{\theta}) - k(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{D} \right| = o_P(1). \tag{3.2}$$

To lighten the notation, set $d = p+q+1$, so that $\mathbf{u} = [u_1, u_2, \ldots, u_d]$. Observe that

$$l_i'(\mathbf{u}) = [l_{i,1}'(\mathbf{u}), l_{i,2}'(\mathbf{u}), \ldots, l_{i,d}'(\mathbf{u})]$$

and denote

$$l_i''(\mathbf{u}) = \begin{bmatrix} \frac{\partial l_{i,1}'(\mathbf{u})}{\partial u_1} & \frac{\partial l_{i,2}'(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial l_{i,d}'(\mathbf{u})}{\partial u_1} \\ \frac{\partial l_{i,1}'(\mathbf{u})}{\partial u_2} & \frac{\partial l_{i,2}'(\mathbf{u})}{\partial u_2} & \cdots & \frac{\partial l_{i,d}'(\mathbf{u})}{\partial u_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial l_{i,1}'(\mathbf{u})}{\partial u_d} & \frac{\partial l_{i,2}'(\mathbf{u})}{\partial u_d} & \cdots & \frac{\partial l_{i,d}'(\mathbf{u})}{\partial u_d} \end{bmatrix}.$$

Denote by $l_{i,j}''(\mathbf{u})$ the jth column of $l_i''(\mathbf{u})$. Applying the mean value theorem to each coordinate $\sum_{1\le i\le k} l_{i,j}'$ of $\sum_{1\le i\le k} l_i'$, we get

$$\sum_{1\le i\le k} l_{i,j}'(\hat{\boldsymbol{\theta}}_n) - \sum_{1\le i\le k} l_{i,j}'(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \sum_{1\le i\le k} l_{i,j}''(\boldsymbol{\xi}_{k,j}), \tag{3.3}$$

where each $\boldsymbol{\xi}_{k,j}$ satisfies $|\boldsymbol{\xi}_{k,j} - \boldsymbol{\theta}| \le |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}|$, $1 \le j \le d = p+q+1$. Theorem 1.2 of Berkes and Horváth (2004) (asymptotic normality) implies that

$$\left| \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right| = O_P(n^{-1/2}). \tag{3.4}$$

Therefore, denoting by $\mathbf{D}_j$ the jth column of $\mathbf{D}$ and using the definition of the max norm, we conclude from (3.3) and (3.4) that

$$\left| \sum_{1\le i\le k} l_i'(\hat{\boldsymbol{\theta}}_n) - \sum_{1\le i\le k} l_i'(\boldsymbol{\theta}) - k(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{D} \right| = O_P(n^{-1/2}) \max_{1\le j\le d} \left| \sum_{1\le i\le k} l_{i,j}''(\boldsymbol{\xi}_{k,j}) - k\mathbf{D}_j \right|. \tag{3.5}$$

According to Lemma 4.4 of Berkes and Horváth (2004), there is U^*, a neighborhood of $\boldsymbol{\theta}$, such that

$$\lim_{k\to\infty} \sup_{\mathbf{u}\in U^*} \left| \frac{1}{k} \sum_{1\le i\le k} l_i''(\mathbf{u}) - El_0''(\mathbf{u}) \right| = 0 \quad a.s. \tag{3.6}$$

Since $El_0''(\boldsymbol{\theta}) = \mathbf{D}$ and by Theorem 1.1 of Berkes and Horváth (2004) $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}$ a.s., we conclude from (3.6) and the relations $|\boldsymbol{\xi}_{k,j} - \boldsymbol{\theta}| \le |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}|$ that

$$\lim_{n\to\infty} \max_{1\le k\le n} \max_{1\le j\le d} \left| \frac{1}{k} \sum_{1\le i\le k} l_{i,j}''(\boldsymbol{\xi}_{k,j}) - \mathbf{D}_j \right| = 0 \quad a.s. \tag{3.7}$$

Using (3.5) and (3.7), we obtain

$$n^{-1/2} \max_{1\le k\le n} \left| \sum_{1\le i\le k} l_i'(\hat{\boldsymbol{\theta}}_n) - \sum_{1\le i\le k} l_i'(\boldsymbol{\theta}) - k(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{D} \right|$$

$$= O_P(n^{-1}) \max_{1\le k\le n} \max_{1\le j\le d} \left| \sum_{1\le i\le k} l_{i,j}''(\boldsymbol{\xi}_{k,j}) - k\mathbf{D}_j \right|$$

$$= O_P(1) \max_{1\le k\le n} \max_{1\le j\le d} \left| \frac{1}{k} \sum_{1\le i\le k} l_{i,j}''(\boldsymbol{\xi}_{k,j}) - \mathbf{D}_j \right| = o_P(1),$$

completing the proof of (3.2) and of Lemma 3.1.

Lemma 3.2 *Under the assumptions of Theorem 2.1,*

$$n^{-1/2} \max_{1\le k\le n} \left| \sum_{1\le i\le k} \hat{l}_i'(\hat{\boldsymbol{\theta}}_n) - \left[\sum_{1\le i\le k} l_i'(\boldsymbol{\theta}) + k(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{D} \right] \right| = o_P(1)$$

Proof. Use Lemma 3.1 and (3.1).

Lemma 3.3 *Under the assumptions of Theorem 2.1,* $\mathbf{D}^{-1}$ *exists and*

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = -\frac{1}{n} \sum_{1\le i\le n} l_i'(\boldsymbol{\theta})\mathbf{D}^{-1} + o_P(n^{-1/2}).$$

Proof. The existence of $\mathbf{D}^{-1}$ was established in Lemma 4.5 of Berkes and Horváth (2004). Thus, the lemma is an immediate consequence of Lemmas 4.5 and 4.7 of Berkes and Horváth (2004).

Lemma 3.4 *Under the assumptions of Theorem 2.1,*

$$n^{-1/2} \sum_{1\le i\le nt} l_i'(\boldsymbol{\theta}) \to W_{\mathbf{D}}(t), \text{ in } D\left([0,1], \mathbb{R}^{p+q+1}\right),$$

where $W_{\mathbf{D}}(t) \in \mathbb{R}^{p+q+1}$ *is a Gaussian process with zero mean and covariance matrix* $\min(s,t)\mathbf{D}$.

Proof. Recall from Section 2 that $l_i(\mathbf{u}) = g(y_i, w_i^{-1/2}(\mathbf{u}))$, where $g(y,t) = \log[th(yt)]$. Recall also that while we do not assume that h has a derivative, we do assume that the partial derivative $g_1(y,t)$ exists. Using the definition of g and a change of variables, it is easy to verify that

$$g_1(y,t) = g_1(yt,1)/t.$$

Since $w_i(\boldsymbol{\theta}) = \sigma_i^2$, it follows that

$$-2l_i'(\boldsymbol{\theta}) = g_1\left(y_i, w_i^{-1/2}(\boldsymbol{\theta})\right) w_i^{-3/2}(\boldsymbol{\theta}) w_i'(\boldsymbol{\theta})$$

$$= \sigma_i g_1(\sigma_i^{-1} y_i, 1) w_i^{-3/2}(\boldsymbol{\theta}) w_i'(\boldsymbol{\theta}) = g_1(\varepsilon_i, 1) w_i'(\boldsymbol{\theta}) / w_i(\boldsymbol{\theta}).$$

Since $w_i(\boldsymbol{\theta})$ is a function of $\varepsilon_j,\ j < i$,

$$-2E[l_i'(\boldsymbol{\theta})] = E[g_1(\varepsilon_i, 1)] E[w_i'(\boldsymbol{\theta})/w_i(\boldsymbol{\theta})] = 0$$

because by (2.20), $E[g_1(\varepsilon_i, 1)] = 0$. We conclude that

$$\left\{ \sum_{1 \le i \le k} l_i'(\boldsymbol{\theta}),\ \sigma(\varepsilon_i,\ i \le k) \right\} \tag{3.8}$$

is a martingale and that $E[l_i'(\boldsymbol{\theta})] = 0$.

By Lemma 3.6 of Berkes and Horváth (2004),

$$E|w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta})|^{\kappa} < \infty \tag{3.9}$$

for every $\kappa > 0$. Combining (3.9) with condition (2.25), and using the independence of $g_1(\varepsilon_i, 1)$ and $w_i^{-1}(\boldsymbol{\theta}) w_i'(\boldsymbol{\theta})$, shows that (3.8) is a square integrable martingale. Thus, the lemma follows from Theorem 23.1 of Billingsley (1968) and an application of the Cramér-Wald device, see e.g. Billingsley (1968), p. 49.

Lemma 3.5 *Under the assumptions of Theorem 2.1,*

$$n^{-1/2} \left[\sum_{1 \le i \le nt} l_i'(\boldsymbol{\theta}) - \frac{[nt]}{n} \sum_{1 \le i \le n} l_i'(\boldsymbol{\theta}) \right] \mathbf{D}^{-1/2} \tag{3.10}$$

$$\to (B_1(t), \ldots, B_{p+q+1}(t)) \text{ in } D\left([0,1], \mathbb{R}^{p+q+1}\right),$$

where $\{B_i(t),\ 0 \le t \le 1\},\ i = 1, \ldots, p+q+1$, *are independent Brownian bridges.*

Proof. By Lemma 3.4

$$n^{-1/2} \left[\sum_{1 \le i \le nt} l_i'(\boldsymbol{\theta}) - \frac{[nt]}{n} \sum_{1 \le i \le n} l_i'(\boldsymbol{\theta}) \right] \to W_{\mathbf{D}}(t) - t W_{\mathbf{D}}(1) \text{ in } D\left([0,1], \mathbb{R}^{p+q+1}\right). \tag{3.11}$$

Since the process $W_{\mathbf{D}}(\cdot)$ is Gaussian, computing the covariances shows that

$$\left\{ (W_{\mathbf{D}}(t) - t W_{\mathbf{D}}(1)) \mathbf{D}^{-1/2},\ 0 \le t \le 1 \right\} \stackrel{d}{=} \{B_1(t), \ldots, B_{p+q+1}(t),\ 0 \le t \le 1\}.$$

Proof of Theorem 2.1: By Lemmas 3.2 and 3.3, we get

$$n^{-1/2} \max_{1 \le k \le n} \left| \sum_{1 \le i \le k} \hat{l}_i'(\hat{\boldsymbol{\theta}}_n) - \left[\sum_{1 \le i \le k} l_i'(\boldsymbol{\theta}) - \frac{k}{n} \sum_{1 \le i \le n} l_i'(\boldsymbol{\theta}) \right] \right| = o_P(1).$$

Hence by Lemma 3.5,

$$n^{-1/2} \left[\sum_{1 \le i \le nt} \hat{l}_i'(\hat{\boldsymbol{\theta}}_n) \right] \mathbf{D}^{-1/2} \to (B_1(t), \ldots, B_{p+q+1}(t)) \text{ in } D\left([0,1], \mathbb{R}^{p+q+1}\right),$$

completing the proof of Theorem 2.1.

References

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

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

Bagshaw, M. and Johnson, R. A. (1975). The effect of serial correlation on the performance of CUSUM tests II. *Technometrics*, **17,** 73–80.

Berkes, I. and Horváth, L. (2003). The rate of consistency of of the quasi-maximum likelihood estimator. *Statistics and Probability Letters*, **61,** 133–143.

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics*, **32,** 633–665.

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

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

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

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

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Drost, F. C. and Klaassen, C. A. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, **81,** 193–221.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, Berlin.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50,** 987–1007.

Guillaume, D. M., Dacorogna, M. M., Dave, R. D., Müller, U. A., Olsen, R. B. and Pictet, O. V. (1997). From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, **1,** 95–129.

Horváth, L., Kokoszka, P. S. and Teyssière, G. (2004). Bootstrap misspecification tests for arch based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation*, **74,** 469–485.

Johnson, R. A. and Bagshaw, M. (1974). The effect of serial correlation on the performance of CUSUM tests. *Technometrics*, **16,** 103–112.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests. *Ann. Math. Statist.*, **30,** 420–447.

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

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

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

Kokoszka, P. S. and Teyssière, G. (2002). Change point detection in GARCH models: asymptotic and bootstrap tests. Technical Report. Utah State University, Logan. Available at `http://stat.usu.edu/~piotr/research.html`.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in time series. *Journal of Time Series Analysis*, **20**, 36–60.

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

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

Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* John Wiley & Sons Ltd.