PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 25, Fasc. 2 (2005), pp. 331–362

DATA-DRIVEN SCORE TEST OF FIT FOR CONDITIONAL DISTRIBUTION IN THE GARCH(1,1) MODEL

BY

TADEUSZ INGLOT AND BARTOSZ STAWIARSKI (WROCŁAW)

Abstract. A data-driven score test for a conditional distribution in the GARCH(1,1) model is proposed. Conditional distribution assumption is verified by a score test, obtained from nesting the null density into an exponential family and then choosing the dimension of this exponential family by a score-based selection rule. A simulation study, which is provided, shows good empirical behaviour of the proposed test, outperforming in most cases the behaviour of competitive tests.

2000 AMS Mathematics Subject Classification: Primary: 62M10, 91B84; Secondary: 62G10, 62E20, 60F05, 65C05.

Key words and phrases: GARCH(1,1) model, noise distribution, efficient score vector, score test, BIC Schwarz selection rule, martingale difference array, central limit theorem, ergodic theorem, Monte Carlo simulations.

1. INTRODUCTION

One of the most celebrated breakthroughs in econometric modelling methods was the conception of Autoregressive Conditionally Heteroscedastic (ARCH) time series by Engle (1982) and more general GARCH by Bollerslev (1986). Nowadays, econometrics and finance are still the most relevant fields of practical research standing behind this theory. In contrast to classical ARMA models, these nonlinear ones adopt a time-changing conditional variance.

Since Bollerslev (1986), a large number of papers concerning the conditional Gaussian GARCH series has been published. However, the normality assumption was found incompatible with real data. Vast empirical evidence (see e.g. Diebold (1988)) against conditional normality revealed the necessity of considering other distributions. Observed excessive kurtosis, heavier tails of the rescaled residuals were the major arguments justifying the usage of t-distributions in econometric modelling of innovations (noise) in GARCH time series, as shown in Bollerslev (1987) and Baillie and Bollerslev (1989). Hence, testing assumptions for the noise distribution has become one of the important issues in fitting the proper model to the data at hand. Several solutions to this problem have been proposed in the literature, some of them extending classical results for i.i.d. observations. Chen and Kuan (2002) derive a modification of the original Jarque–Bera (J–B) normality test and apply it in GARCH framework. Third and fourth sample moments play an essential role in their test statistic. Monte Carlo simulations proved good performance in detecting some departures from normality. Fiorentini et al. (2004), in turn, conclude that a closely related Kiefer–Salmon normality test slightly outperforms the J–B test when applied to time-varying variance models. A different approach was proposed in Bai (2003). With the aid of Khmaladze martingale transformation he elaborated a flexible asymptotically distribution free Kolmogorov-type goodness of fit test. That test can be applied in a general context of dynamic models. Chen (2002) proposed a test based on the characteristic function. However, a choice of weight functions appearing in integral transformations essentially influences the sensitivity of the test.

The aim of this paper is to propose a data-driven test of fit for the noise distribution in GARCH(1,1) model. Although this model is the simplest member of the GARCH(p,q) class, it retains well heteroscedastic properties and exhibits an exponentially vanishing lag structure, absent in ARCH. Our construction matches some score tests and some selection rule. Note also that score tests are often called *smooth tests*.

The concept of smooth goodness of fit tests dates back to 1937, when Neyman introduced a locally optimal test intended to detect departures from the null distribution in many directions equally well. In contrast, classical Kolmogorov–Smirnov or Cramér–von Mises tests downweight successive directions in some sense, and therefore are able to detect very few deviations from the null distributions, only. The numerical evidence and some discussion of this fact can be found e.g. in Milbrodt and Strasser (1990) and Inglot and Ledwina (2001). The idea adopted by Neyman (1937) was to embed the null density into an exponential family and then consider the equivalent parametric testing. The choice of a dimension of this exponential family has substantial influence on the behaviour of the resulting test. This was shown e.g. in Inglot, Kallenberg and Ledwina (1994) and in Kallenberg and Ledwina (1995). To overcome this problem Ledwina (1994) proposed a data-driven smooth test of fit, where the dimension of the exponential family is estimated from the data by Schwarz's BIC selection rule. Inglot et al. (1997) showed that this construction can be adopted to testing composite hypotheses, where the null density depends on some nuisance parameters. Extensive simulations performed in Kallenberg and Ledwina (1997a) confirmed that this data-driven smooth test of fit compares very well even to competitors like Shapiro–Wilk normality test. The ability to detect a wide variety of deviations from the null distribution accounts for an omnibus character of the test.

Recently, the problem of smooth testing was tackled in the context of dependent random variables. Ducharme and Lafaye de Micheaux (2004) derived a data-driven goodness of fit test for normality of innovations in causal, invertible ARMA(p,q) models. The numerical simulations performed there show that the test is more powerful than the classical Anderson–Darling or J–B tests. Therefore, it seems promising to make a step forward by considering a data-driven smooth test for noise distribution in nonlinear GARCH time series. Simulation study presented herein proves that such a test is indeed competitive to the ones discussed above.

The paper is organized as follows. In Section 2 we define GARCH(1,1) model and formulate our assumptions. In Section 3 we derive efficient score vector and establish its asymptotic behaviour; then we define a score-based selection rule and corresponding data-driven statistic and, finally, propose a test statistic. In Section 4 we specify estimators of nuisance parameters and other quantities appearing in our test statistic. We also present a simulation study in which we compare our test to J–B test for normality as well as to the test proposed in Bai (2003). Sections 5 and 6 contain proofs of main and auxiliary results. In the Appendix we derive some properties of GARCH(1,1) series needed in proofs in Sections 5 and 6.

2. THE MODEL AND BASIC ASSUMPTIONS

Throughout the paper we shall consider the GARCH(1,1) model proposed by Bollerslev (1986) as a generalization of the ARCH family introduced in Engle (1982). The GARCH(1,1) time series $\{X_t, t \in Z\}$ (sometimes called *strong-GARCH*, as in Gouriéroux (1997)), defined on a space $(\Omega, \mathscr{F}, P)$, is given by the following relations:

$$\begin{aligned} &X_t = \sqrt{h_t}\,\varepsilon_t, \\ &h_t = \alpha_0 + \alpha X_{t-1}^2 + \beta h_{t-1}, \qquad \alpha_0, \alpha, \beta > 0,\ \alpha+\beta < 1,\ t \in Z, \end{aligned} \tag{2.1}$$

where Z is the set of integers, while $\{\varepsilon_t, t \in Z\}$ is a sequence of i.i.d. random variables having a density $f(x)$ with respect to the Lebesgue measure and $E\varepsilon_t = 0$, $\operatorname{Var}\varepsilon_t = 1$. So, the model we consider is indexed by $\vartheta = (\alpha_0, \alpha, \beta)$ from $\Theta = \{\vartheta = (\alpha_0, \alpha, \beta)\colon \alpha_0, \alpha, \beta > 0,\ \alpha+\beta < 1\}$. The assumption $0 < \alpha+\beta < 1$ ensures strict stationarity and ergodicity of the series $\{X_t\}$. These properties of GARCH time series were intensively investigated by many authors (e.g. Engle (1982), Milhøj (1985) and Bollerslev (1986)). Nelson (1990) and Bougerol and Picard (1992) formulated necessary and sufficient conditions for ergodicity and strict stationarity. For further references and detailed overview of results on this subject we send the reader to Li et al. (2002).

Denote by $\mathscr{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$, $t \in Z$, the σ-field of the process history up to time t. Clearly, for any t, the conditional variance h_t is $\mathscr{F}_{t-1}$-measurable. For each $t \in Z$ put

$$e_t^2 = \alpha \varepsilon_t^2 + \beta. \tag{2.2}$$

The formula (2.1) yields

$$h_t = \alpha_0 + (\alpha \varepsilon_{t-1}^2 + \beta) h_{t-1} = \alpha_0 + e_{t-1}^2 h_{t-1}. \tag{2.3}$$

Successively iterating (2.3) we obtain after n steps

$$h_t = \alpha_0 (1 + e_{t-1}^2 + e_{t-1}^2 e_{t-2}^2 + \ldots + e_{t-1}^2 \ldots e_{t-n+1}^2) + e_{t-1}^2 \ldots e_{t-n}^2 h_{t-n}.$$

Passing n to infinity we get, due to the model assumption $\alpha + \beta < 1$, an explicit almost sure (and L_1) series representation

$$h_t = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \prod_{s=1}^{j} e_{t-s}^2, \qquad t \in Z. \tag{2.4}$$

It is a special case of a matrix formula for h_t in the general GARCH(p,q) model derived by Li and Ling (1997). They proved that the representation is unique, $\mathcal{F}_t$-adaptive, strictly stationary and ergodic. Note that from (2.1) and (2.4) it is easily seen that α_0 is a quadratic scale coefficient, i.e. multiplying X_t by $\sqrt{c}$, $c > 0$, is equivalent to replacing α_0 with $c\alpha_0$. The conditional variance h_t can also be expressed in another form more useful for our purposes. Namely, iterating the second formula in (2.1) we get for $t > 1$

$$h_t = \alpha_0 \sum_{s=0}^{t-2} \beta^s + \alpha \sum_{s=0}^{t-2} \beta^s X_{t-1-s}^2 + \beta^{t-1} h_1. \tag{2.5}$$

The equality (2.5) describes the conditional variance h_t in terms of the observed time series $\{X_t, 1 \leqslant t \leqslant n\}$.

Now, we shall introduce our main assumptions on the model given by (2.1). Whenever it does not lead to ambiguity, to omit the time subscript, we shall denote by ε an r.v. distributed as ε_t. Below $\mathbf{1}_A$ denotes the indicator of a set A. The assumptions (A1)–(A3), listed below, will be valid throughout the rest of the paper.

(A1) The noise density f is absolutely continuous on the real line and has finite Fisher information

$$I_f = E\left(\frac{f'(\varepsilon)}{f(\varepsilon)}\right)^2 = \int_{\{f>0\}} \frac{(f'(y))^2}{f(y)} dy < \infty.$$

(A2) The function

$$\zeta(y) = \left(y \frac{f'(y)}{f(y)} + 1\right) \mathbf{1}_{\{f>0\}}(y)$$

is not an almost everywhere constant function.

(A3) $E|\zeta(\varepsilon)|^3 < \infty$.

Remark 2.1. The assumption (A2) will be used to prove linear independence of components of the score vector (cf. Proposition 3.3). It is mild and holds for standard densities. An example, when (A2) is not satisfied, is as follows. Take for some $C > 1$ and $y_0 > 0$

$$f(y) = \begin{cases} (2|y|\log C)^{-1} & \text{for } |y| \in [y_0, Cy_0], \\ 0 & \text{otherwise.} \end{cases}$$

Then it is immediately seen that $\zeta(y) = 0$ almost everywhere. The same holds for convex combinations of such functions.

Remark 2.2. The assumption (A3) will be used in the proof of Theorem 3.6 to show that a Lyapunov-type condition for normalized efficient score vector holds true. This assumption is a technical one and could be weakened. Since it is satisfied in most cases important for applications, we decided, for the sake of simplicity, to impose this stronger form.

Observe that $E\zeta(\varepsilon) = 0$ and put

$$J_f = \operatorname{Var}\zeta(\varepsilon) = E\left(\varepsilon \frac{f'(\varepsilon)}{f(\varepsilon)}\right)^2 - 1. \tag{2.6}$$

3. CONSTRUCTION OF THE TEST STATISTIC

Suppose we observe a stochastic process $\{X_t, t \in Z\}$ obeying the model (2.1). Given observations $X_1, \ldots, X_n$ of $\{X_t\}$, we would like to verify the null hypothesis asserting

H_0: ε_t's in (2.1) have the density f, $\vartheta \in \Theta$,

where $f(y)$ is a completely specified density function (with respect to the Lebesgue measure) on the real line. The GARCH(1,1) coefficient $\vartheta = (\alpha_0, \alpha, \beta) \in \Theta$ of our model constitutes the 3-dimensional nuisance parameter in this testing problem.

At the first step, applying the Neyman idea, we fix k, $k \geqslant 1$, and embed the null density f into an auxiliary k-dimensional exponential family of the form

$$\exp\{\theta^T \Phi(F(y)) - C_k(\theta)\} f(y), \tag{3.1}$$

where $F(y)$ denotes the cdf of $f(y)$, $\Phi(y) = [\Phi_1(y), \ldots, \Phi_k(y)]^T$ is a vector of bounded orthonormal functions in $L_2[0, 1]$, $\theta \in R^k$, and $C_k(\theta)$ is the normalizing constant. All vectors appearing throughout the paper are column ones. Assuming that the unknown noise density belongs to this family, verifying H_0 is equivalent to testing

H_0^*: $\theta = 0$, $\vartheta \in \Theta$.

For a fixed $k \geqslant 1$ testing such a hypothesis can be performed by the efficient score test. Vast literature on this subject is available, see e.g. Cox and Hinkley (1974), Thomas and Pierce (1979) and Inglot et al. (1997). Note that related terminology is rich. Efficient score tests are called also *score*, *smooth* or *efficient tests*. Following Cox and Hinkley (1974), we shall use below the name *score test*.

Finally, at the second step, we adopt the idea proposed by Ledwina (1994) and extended to the case when nuisance parameters are present in Inglot et al. (1997) to choose the dimension k using the data at hand. Such a strategy results in a data-driven score statistic.

In consecutive subsections we shall follow the standard way (in i.i.d. case) of deriving efficient score vector and constructing score statistic.

3.1. The score vector. From (2.1) we have $X_t = \sqrt{h_t}\,\varepsilon_t$ with h_t given by (2.5). Therefore, one can express $(\varepsilon_1, \ldots, \varepsilon_n)$ in terms of $(X_1, \ldots, X_n)$ and h_1. To this end take $h > 0$, put $q_1 = q_1(h, \vartheta) = h$ and for $t > 1$ set

$$(3.2) \qquad q_t = q_t(x_1, \ldots, x_{t-1}; h, \vartheta) = \alpha_0 \sum_{s=0}^{t-2} \beta^s + \alpha \sum_{s=0}^{t-2} \beta^s x_{t-1-s}^2 + \beta^{t-1} h,$$

where $\vartheta = (\alpha_0, \alpha, \beta) \in \Theta$. Then, by (2.5), we have

$$h_t = q_t(X_1, \ldots, X_{t-1}; h_1, \vartheta),$$

where ϑ denotes an unknown true value of the nuisance parameter of our model. Consequently,

$$(3.3) \qquad \varepsilon_t = \frac{X_t}{\sqrt{h_t}} = \frac{X_t}{\sqrt{q_t(X_1, \ldots, X_{t-1}; h_1, \vartheta)}}, \qquad t = 1, 2, \ldots$$

From (3.1) we infer that the joint density of the vector $(\varepsilon_1, \ldots, \varepsilon_n)$ has the form

$$g_k(y_1, \ldots, y_n) = \prod_{t=1}^{n} \exp\{\theta^T \Phi(F(y_t)) - C_k(\theta)\} \cdot f(y_t).$$

Hence and from (3.3) we get, by standard calculations, the explicit formula for the logarithm of conditional density of $(X_1, \ldots, X_n)$ given $h_1 = h$:

$$(3.4) \qquad L_k = L_k(x_1, \ldots, x_n; \theta, \vartheta \mid h_1 = h)$$
$$= \sum_{t=1}^{n} \{\theta^T \Phi(F(x_t q_t^{-1/2})) - C_k(\theta)\} + \sum_{t=1}^{n} \log f(x_t q_t^{-1/2})$$
$$+ \sum_{t=1}^{n} \log(q_t^{-1/2}).$$

Later on (cf. (3.7) and Theorem 3.6) it will be shown that the influence of the unknown value of h is asymptotically negligible due to the exponential decay of the multiplier β^{t-2} standing at it. So, we do not need to consider h as an additional nuisance parameter. In some papers it is suggested that h should be the stationary solution of Nelson (1990) (see Lumsdaine (1996)), whereas e.g. Drost and Klaassen (1997) consider it as a certain "starting value".

Now, we define components of the score vector l as derivatives of $L_k(X_1, \ldots, X_n; \theta, \vartheta \mid h_1 = h)$ with respect to all parameters involved, calculated at the point $(0, \vartheta)$, i.e. under the null hypothesis. Namely,

$$l^T = l^T(\vartheta) = [l_\theta^T, l_\vartheta^T] = \left[\left(\frac{\partial L_k}{\partial \theta}\right)^T, \left(\frac{\partial L_k}{\partial \vartheta}\right)^T\right]\Bigg|_{\theta=0}. \tag{3.5}$$

The first k components of l correspond to the parameter of interest θ while the next three components concern the nuisance parameter ϑ. Set

$$Q_t = q_t(X_1, \ldots, X_{t-1}; h, \vartheta).$$

Routine calculations give

$$l_\theta = \frac{\partial L_k}{\partial \theta}\bigg|_{\theta=0} = \left(\frac{\partial L_k(X_1, \ldots, X_n; \theta, \vartheta \mid h_1 = h)}{\partial \theta}\right)\bigg|_{\theta=0} = \sum_{t=1}^{n} \Phi(F(X_t Q_t^{-1/2})) \tag{3.6}$$

and

$$\begin{aligned} l_\vartheta = \frac{\partial L_k}{\partial \vartheta}\bigg|_{\theta=0} &= \left(\frac{\partial L_k(X_1, \ldots, X_n; \theta, \vartheta \mid h_1 = h)}{\partial \vartheta}\right)\bigg|_{\theta=0} \\ &= -\frac{1}{2}\sum_{t=2}^{n} \frac{\zeta(X_t Q_t^{-1/2})}{Q_t}\frac{\partial Q_t}{\partial \vartheta} = -\sum_{t=2}^{n} \frac{\zeta(X_t Q_t^{-1/2})}{2Q_t} \\ &\times \Big[\sum_{s=0}^{t-2} \beta^s, \sum_{s=0}^{t-2} \beta^s X_{t-1-s}^2, \sum_{s=0}^{t-2} s\beta^{s-1}(\alpha_0 + \alpha X_{t-1-s}^2) + (t-1)\beta^{t-2} h\Big]^T. \end{aligned} \tag{3.7}$$

To formulate our basic results concerning the score vector l let us introduce some further notation. Observe that from (2.2) and (2.4) it follows that $h_t > \alpha_0/(1-\beta)$ a.s. for every $t \in Z$. For a future use set $\alpha_0/(1-\beta) = \kappa_0$. Here and in the sequel we shall denote by P_h the probability on the σ-field $\sigma(X_1, X_2, \ldots)$ induced by the family of conditional densities of $(X_1, \ldots, X_n)$ given $h_1 = h$, $h > \kappa_0$, under the null hypothesis. These densities have the form (cf. (3.4) with $\theta = 0$)

$$\prod_{t=1}^{n} f\left(\frac{x_t}{\sqrt{q_t}}\right)\frac{1}{\sqrt{q_t}}, \qquad n = 1, 2, \ldots \tag{3.8}$$

Accordingly, E_h will stand for expectation with respect to this probability.

The following fact plays a crucial role in the proofs of properties of l.

PROPOSITION 3.1. *Suppose* $\{X_t, t \in Z\}$ *obeys the model* (2.1). *For any* $h > \kappa_0$ *and* $n \geq 1$ *the random variables*

$$\tilde{\varepsilon}_1 = X_1 Q_1^{-1/2},\ \tilde{\varepsilon}_2 = X_2 Q_2^{-1/2}, \ldots, \tilde{\varepsilon}_n = X_n Q_n^{-1/2}$$

are independent under P_h *and have the same distribution as* ε_t*'s.*

The proof of Proposition 3.1 as well as the proofs of all other results of this section are given in Section 5. As a corollary of Proposition 3.1 we get square integrability of l. Namely, we have

PROPOSITION 3.2. *Assume that* (A3) *is fulfilled. Then for any* $\vartheta \in \Theta$, *any* $h > \kappa_0$ *and every* $n \geq 1$ $E_h l(\vartheta) = 0$ *and* $E_h \|l(\vartheta)\|^2 < \infty$ *hold true, where* $\|\bullet\|$ *denotes the Euclidean norm.*

An important property of the score vector l, needed to construct a score statistic, is nonsingularity of its covariance matrix. This is guaranteed by the following proposition.

PROPOSITION 3.3. *Suppose* $\{X_t, t \in Z\}$ *follows the model* (2.1), *and* (A2) *is satisfied. Then for any* $\vartheta \in \Theta$, *any* $h > \kappa_0$ *and every* $n \geq 1$ *the components of the score vector* $l(\vartheta)$ *given in* (3.5)–(3.7) *are linearly independent random variables.*

3.2. Efficient score vector and related results. Denote by $B^{(n)}(\vartheta)$ the covariance matrix of the normalized score vector $n^{-1/2} l(\vartheta)$, i.e.

$$B^{(n)}(\vartheta) = n^{-1} E_h\big(l(\vartheta)\, l^T(\vartheta)\big).$$

According to (3.5) divide $B^{(n)}(\vartheta)$ into four blocks putting

$$B^{(n)}(\vartheta) = \begin{bmatrix} B_{11}^{(n)}(\vartheta) & B_{12}^{(n)}(\vartheta) \\ B_{21}^{(n)}(\vartheta) & B_{22}^{(n)}(\vartheta) \end{bmatrix}. \tag{3.9}$$

Note that $B_{22}^{(n)}(\vartheta)$ is a 3×3 matrix of covariances of $n^{-1/2} l_\vartheta$, $B_{12}^{(n)}(\vartheta) = [B_{21}^{(n)}(\vartheta)]^T = n^{-1} E_h\big(l_\theta(\vartheta)\, l_\vartheta^T(\vartheta)\big)$ is a $k \times 3$ matrix of mixed covariances while, by Proposition 3.1 and orthonormality of Φ_j's, $B_{11}^{(n)}(\vartheta) = I$ is the $k \times k$ identity matrix. From Proposition 3.3 we infer that $B^{(n)}(\vartheta)$ and $B_{22}^{(n)}(\vartheta)$ are nonsingular for every n, $\vartheta \in \Theta$, and $h > \kappa_0$.

Recall that components of the efficient score vector $l_k^*(\vartheta)$ are defined to be residuals of the orthogonal projections (in $L_2(P_h)$ for our case) of components of the vector $n^{-1/2} l_\theta$ onto the subspace spanned by the components of $n^{-1/2} l_\vartheta$. Consequently (cf., e.g., Cox and Hinkley (1974)),

$$l_k^*(\vartheta) = n^{-1/2}\big(l_\theta - B_{12}^{(n)}(\vartheta)\,[B_{22}^{(n)}(\vartheta)]^{-1}\, l_\vartheta\big).$$

Thus we have got the following theorem.

THEOREM 3.4. *Suppose the observed time series* $\{X_t\}$ *obeys the model* (2.1), *and* (A1)–(A3) *are satisfied. Let* $\Phi(y) = [\Phi_1(y), \ldots, \Phi_k(y)]^T$ *be a vector of bounded orthonormal functions in* $L_2[0, 1]$. *Then for any* $h > \kappa_0 = \alpha_0/(1-\beta)$ *the efficient score vector* $l_k^*(\vartheta)$ *for testing* H_0^* *has the form*

$$(3.10)\qquad l_k^*(\vartheta) = n^{-1/2}\sum_{t=1}^{n}\Phi(F(\tilde{\varepsilon}_t)) + n^{-1/2}\sum_{t=2}^{n}\frac{\zeta(\tilde{\varepsilon}_t)}{2Q_t}B_{12}^{(n)}(\vartheta)[B_{22}^{(n)}(\vartheta)]^{-1}\frac{\partial Q_t}{\partial\vartheta},$$

where $\tilde{\varepsilon}_t$'s are defined in Proposition 3.1. *Moreover, the covariance matrix of $l_k^*(\vartheta)$ under P_h is given by*

$$(3.11)\qquad \mathcal{M}^{(n)}(\vartheta) = I - B_{12}^{(n)}(\vartheta)[B_{22}^{(n)}(\vartheta)]^{-1}B_{21}^{(n)}(\vartheta).$$

The formula (3.10) describing the efficient score vector reveals its martingale structure. Indeed, for $t \geqslant 1$ let us write $\sigma_t = \sigma\{X_1, \ldots, X_t\}$, set $Y_{1n} = n^{-1/2}\Phi(F(\tilde{\varepsilon}_1))$ and

$$(3.12)\qquad Y_{tn} = n^{-1/2}\left(\Phi(F(\tilde{\varepsilon}_t)) + \frac{1}{2}\frac{\zeta(\tilde{\varepsilon}_t)}{Q_t}B_{12}^{(n)}(\vartheta)[B_{22}^{(n)}(\vartheta)]^{-1}\frac{\partial Q_t}{\partial\vartheta}\right), \quad t = 2, \ldots, n,$$

the consecutive summands in (3.10). Clearly, for each $t = 1, \ldots, n$, Y_{tn} is σ_t-measurable. We have

PROPOSITION 3.5. *For any $n \geqslant 1$ and $h > \kappa_0$ the sequence $\{Y_{1n}, \ldots, Y_{nn}\}$ is a martingale difference array with respect to $\{\sigma_1, \ldots, \sigma_n\}$ under P_h.*

We omit an elementary proof of Proposition 3.5. Let us put $X_{tn} = (\mathcal{M}^{(n)}(\vartheta))^{-1/2}Y_{tn}$, $t = 1, \ldots, n$. Then

$$(3.13)\qquad \sum_{t=1}^{n} X_{tn} = (\mathcal{M}^{(n)}(\vartheta))^{-1/2} l_k^*(\vartheta)$$

is the standardized efficient score vector. The martingale structure of $l_k^*(\vartheta)$ allows us to apply the result of Kundu et al. (2000) to establish limit behaviour of (3.13). The main theorem of this paper is as follows.

THEOREM 3.6. *Suppose the observed time series $\{X_t\}$ obeys the model* (2.1), *and* (A1)–(A3) *are satisfied. Let $\Phi(y) = [\Phi_1(y), \ldots, \Phi_k(y)]^T$ be a vector of bounded orthonormal functions in $L_2[0, 1]$. Then for any $k \geqslant 1$ and almost every $h > \kappa_0 = \alpha_0/(1-\beta)$ (with respect to the Lebesgue measure) we have in R^k, under P_h,*

$$(3.14)\qquad (\mathcal{M}^{(n)}(\vartheta))^{-1/2} l_k^*(\vartheta) \xrightarrow{\mathcal{D}} N(0, I) \quad \text{as } n \to \infty,$$

where I is the $k\times k$ identity matrix.

Theorem 3.6 is proved in Section 5. Now, for a given $k \geqslant 1$, set

$$(3.15)\qquad W_k(\vartheta) = \left\|(\mathcal{M}^{(n)}(\vartheta))^{-1/2} l_k^*(\vartheta)\right\|^2 = (l_k^*(\vartheta))^T(\mathcal{M}^{(n)}(\vartheta))^{-1} l_k^*(\vartheta),$$

where $\|\bullet\|$ again denotes the Euclidean norm in R^k. From Theorem 3.6 we immediately get

COROLLARY 3.7. *Under the assumptions of Theorem* 3.6 *it follows that*

$$(3.16)\qquad W_k(\vartheta) \xrightarrow{\mathcal{D}} \chi_k^2 \quad \text{as } n \to \infty$$

under P_h, where χ_k^2 is a chi-square random variable with k degrees of freedom.

3.3. Data-driven score statistic. Up to now we focused on the fixed dimension k of exponential family built on the null distribution. As mentioned in the Introduction, the choice of k strongly influences the sensitivity of the corresponding score test. Therefore, following the construction proposed by Ledwina (1994) and developed in Kallenberg and Ledwina (1997a, b) we suggest the adaptive choice of the dimension k based on the data. Assume for a moment that ϑ is known and consider a score-based selection rule defined as

$$S1(\vartheta) = \min\{k\colon\ 1 \leqslant k \leqslant K,\ W_k(\vartheta) - ck\log n \geqslant W_j(\vartheta) - cj\log n \quad \text{for all } j = 1, \ldots, K\}. \tag{3.17}$$

Here $K < \infty$ is a fixed, but arbitrarily chosen, maximal dimension we allow while c is some positive constant controlling the magnitude of the penalty. $S1(\vartheta)$ is intended to simplify the original Schwarz BIC criterion. The choice $c = 1$ corresponds to BIC penalty. We have

$$\{S1(\vartheta) > 1\} = \bigcup_{k=2}^{K} \{S1(\vartheta) = k\} \subset \sum_{k=2}^{K} \{W_k(\vartheta) \geqslant c(k-1)\log n\}.$$

In consequence, Corollary 3.7 implies

$$P_h\big(S1(\vartheta) > 1\big) \leqslant \sum_{k=2}^{K} P_h\big(W_k(\vartheta) \geqslant c(k-1)\log n\big) \to 0 \quad \text{as } n \to \infty. \tag{3.18}$$

This together with Corollary 3.7 determines the asymptotic behaviour of $W_{S1(\vartheta)}(\vartheta)$.

COROLLARY 3.8. *Under the assumptions of Theorem* 3.6 *it follows that*

$$W_{S1(\vartheta)}(\vartheta) \xrightarrow{\mathscr{D}} \chi_1^2 \quad \textit{as } n \to \infty \tag{3.19}$$

under P_h.

Obviously, the statistic $W_{S1(\vartheta)}(\vartheta)$, depending on the unknown nuisance parameter ϑ, cannot be directly used as a test statistic. A natural and standard solution is to insert into the two items (i.e. $W_k(\vartheta)$ and $S1(\vartheta)$) an estimate of ϑ. Below, we describe our implementation which patterns standard approach in the case of independent observations; cf. also Bühler and Puri (1966) and Inglot et al. (1997), Section 3.

Suppose $\hat{\vartheta} = (\hat{\alpha}_0, \hat{\alpha}, \hat{\beta})$ is a square-root consistent estimator of $\vartheta = (\alpha_0, \alpha, \beta)$ while $\hat{B}_{12}^{(n)}$ and $\hat{B}_{22}^{(n)}$ are consistent estimators of the matrices $B_{12}^{(n)}(\vartheta)$ and $B_{22}^{(n)}(\vartheta)$ defined in (3.9). The consistency is required to hold under the null model, only. Put

$$\hat{Q}_t = \hat{\alpha}_0 \sum_{s=0}^{t-2} \hat{\beta}^s + \hat{\alpha} \sum_{s=0}^{t-2} \hat{\beta}^s X_{t-1-s}^2 + \hat{\beta}^{t-1} h \quad \text{and} \quad \hat{\varepsilon}_t = X_t \hat{Q}_t^{-1/2}.$$

Then the estimated efficient score vector takes the form

$$(3.20)\quad \hat{l}_k^*(\hat{\vartheta}) = n^{-1/2}\sum_{t=1}^{n}\Phi\big(F(\hat{\varepsilon}_t)\big)+n^{-1/2}\sum_{t=2}^{n}\frac{\zeta(\hat{\varepsilon}_t)}{2\hat{Q}_t}\hat{B}_{12}^{(n)}[\hat{B}_{22}^{(n)}]^{-1}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}},$$

and consequently we get the score statistic

$$\hat{W}_k(\hat{\vartheta}) = (\hat{l}_k^*(\hat{\vartheta}))^T(I-\hat{B}_{12}^{(n)}[\hat{B}_{22}^{(n)}]^{-1}\hat{B}_{21}^{(n)})^{-1}\hat{l}_k^*(\hat{\vartheta}).$$

Now, (3.17) allows us to introduce the selection rule $\hat{S}$ by

$$(3.21)\quad \hat{S} = \min\{k:\ 1\leqslant k\leqslant K,\ \hat{W}_k(\hat{\vartheta})-ck\log n \geqslant \hat{W}_j(\hat{\vartheta})-cj\log n \quad \text{for all } j=1,\ldots,K\}.$$

Finally, we propose $\hat{W}_{\hat{S}}(\hat{\vartheta})$ as a data-driven score test statistic for testing H_0^*.

In the case of i.i.d. observations, the above assumptions on estimators along with a counterpart of (3.19) imply that data-driven score statistic tends, under the null hypothesis, to χ_1^2 random variable. Such an asymptotic result rather shows disappearing influence of nuisance parameters upon the null distribution than provides a way to calculate critical values. The reason is the rate of convergence of the selection rule to the dimension 1 influences the correct critical values for moderate n. Typically, the asymptotic critical value strongly underestimates the actual ones. For more discussion and some approximation see Kallenberg and Ledwina (1997b).

For dependent data the picture may be different as shown e.g. in Ignaccolo (2004). However, one may conjecture that $\hat{W}_{\hat{S}}(\hat{\vartheta})$ shall asymptotically stabilize when $n\to\infty$. Formal derivation of the asymptotic distribution of $\hat{W}_{\hat{S}}(\hat{\vartheta})$ needs some additional work and will be a subject of a future paper. Here, we restrict ourselves to showing empirical behaviour of the test based on $\hat{W}_{\hat{S}}(\hat{\vartheta})$. The simulation study presented in the next section nicely confirms our conjecture.

4. SIMULATION STUDY

4.1. Specification of estimators and other quantities. In our implementation a Quasi Maximum Likelihood Estimator $\hat{\vartheta}$ (QMLE) of ϑ constructed on the basis of Gaussian likelihood was used. It has been well established in the literature and popular in GARCH modelling. Weiss (1986) and Lee and Hansen (1994) proved that this estimator is square-root consistent and asymptotically normal if the noise distribution has finite fourth moment. Moreover, Ling and McAleer (2003) proved the consistency of QMLE under the second moment condition, only. Note that Drost and Klaassen (1997) proposed semiparametric methods for efficient estimation of model parameters that allow for unknown form of a noise density. The widely accepted iterative BHHH algorithm (see Berndt et al. (1974)) was applied here to obtain $\hat{\vartheta}$. One drawback is that the BHHH algorithm encounters stability problems if the starting parameter is too distant from the true value, as was indicated in Mak et al. (1997). The magnitude of the effect was especially cumbersome when $\beta > 0.8$, i.e. when the true model is close to the nonstationary IGARCH(1,1) model. Such cases were discarded from our simulations so as not to disturb the stability of critical values. Anyway, most attention was paid to the case when $\alpha+\beta$ is larger than 0.8, which accounts for longer (but still exponential) "memory" of the time series, observed in high-frequency financial data. For empirical confirmation of this phenomenon see e.g. papers by Mittnik et al. (1998) or Brooks et al. (2001).

Another crucial step was to estimate the covariance matrices $B_{12}^{(n)}(\vartheta)$ and $B_{22}^{(n)}(\vartheta)$ appearing in $W_k(\vartheta)$ (cf. (3.15) and (3.11)). To avoid estimating further unknown parameters we decided to use moment estimators appearing in (5.19) and (5.20) with estimated $\hat{\vartheta}$ plugged into these formulae. Namely, we set

(4.1) $$\hat{B}_{12}^{(n)} = -\Delta\left[\frac{1}{2n}\sum_{t=2}^{n}\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right]^T$$

and

(4.2) $$\hat{B}_{22}^{(n)} = \frac{J_f}{4n}\sum_{t=2}^{n}\left[\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right]\left[\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right]^T,$$

where

(4.3) $$\Delta = E\left[\Phi(F(\varepsilon))\zeta(\varepsilon)\right]$$

and J_f is given by (2.6). Stationarity and ergodicity of the GARCH(1,1) process $\{X_t\}$, leading to (5.19) and (5.20) in Section 5, argue for the consistency of $\hat{B}_{12}^{(n)}$ and $\hat{B}_{22}^{(n)}$. We state the consistency of (4.1) and (4.2) in Lemma 5.4.

From Theorem 3.6 it follows that unknown value of h has no influence upon the asymptotic behaviour of W_S provided $h > \kappa_0 = \alpha_0/(1-\beta)$. Since we considered $\beta < 0.8$ and $\alpha_0 = 0.001$, we chose a nonrandom fixed value $h = 0.1$ which is considerably greater than actual κ_0 in all presented cases.

As an orthonormal basis we took the cosine system $\Phi_j(x) = \sqrt{2}\cos(j\pi x)$, $j = 1, 2, \ldots,$ on $[0, 1]$.

By taking various penalty coefficients $c \in (0, 1]$ in (3.21), sensitivity of the score test can be regulated. For some evidence see Kallenberg and Ledwina (1997b). In general, too small c lowers the penalty impact and, in consequence, one loses some power for smooth deviations and gains some power for highly oscillating ones. Obviously, the choice of c considerably influences also the critical values. Passing c below 0.5 resulted in large critical values exceeding 10 for $c = 0.3$. As there is no analytic answer which penalty coefficient to choose, we suggested searching for a reasonable computational compromise. Therefore we took $c = 0.5$ together with $c = 1$ as two competing penalty coefficients.

In our empirical study, the following two standardized null distributions were considered:

- *normal* with the density $f(y) = (2\pi)^{-1/2}\exp(-y^2/2)$,
- *Laplace* with the density $f(y) = 2^{-1/2}\exp(-\sqrt{2}\,|y|)$.

Testing the conditional GARCH normality is highly desirable as this assumption has been constantly questioned since Diebold (1988) provided strong evidence on favour of heavier tails in the noise distribution. The standard Laplace distribution, in turn, allows for heavier tails and may serve as competitor to e.g. more complex t-distributions.

Below, we describe all remaining quantities and functions (depending on the null distribution) which appear in the explicit form of $\hat{l}_k^*(\hat{\vartheta})$ in (3.20), (4.1) and (4.2). In the *normal* case we have $\zeta(y) = 1 - y^2$ and $J_f = 2$ while for the *Laplace* case $\zeta(y) = 1 - \sqrt{2}\,|y|$ and $J_f = 1$. For the cosine system the components of the vector Δ defined in (4.3) equal: $\Delta_j = 0$ for odd j's and $\Delta_j = -\sqrt{8}\int_0^{1/2}\cos(j\pi u)(\Phi^{-1}(u))^2\,du$ for even j's for the *normal* case, while $\Delta_j = 0$ for odd j's and $\Delta_j = \sqrt{8}\int_0^{1/2}\cos(j\pi u)(\log 2u)\,du$ for even j's for the *Laplace* case. Here Φ denotes the standard normal distribution function. In our simulations the non-zero integrals were calculated numerically.

4.2. Null behaviour of $\hat{W}_{\hat{S}}(\hat{\vartheta})$. Table 1 contains simulated critical values at 0.05 significance level, determined separately for the *normal* and *Laplace* cases. Corresponding average critical values are presented in the last row. Our choices $\beta > \alpha$ were motivated by empirical justification of Brooks et al. (2001), since it is β that governs the memory of the process by means of inclusion of the past conditional variances in h_t.

The influence of α and β on the critical value is highly limited. The maximal oscillations around the average critical values do not exceed 4% of this average in each case.

TABLE 1. Simulated critical values for $\hat{W}_{\hat{S}}$ for *normal* and *Laplace* null hypotheses for various GARCH(1,1) parameters. Significance level 0.05, $\alpha_0 = 0.001$, $h = 0.1$, $K = 10$, $n = 500$, $M = 5000$ Monte Carlo runs

α	β	*normal*		*Laplace*	
		$c = 0.5$	$c = 1$	$c = 0.5$	$c = 1$
0.3	0.4	7.600	4.335	7.385	4.430
0.3	0.5	7.435	4.418	7.380	4.219
0.2	0.7	7.944	4.536	7.372	4.633
0.25	0.65	7.413	4.406	7.702	4.342
0.4	0.5	7.345	4.511	7.536	4.507
0.3	0.65	7.974	4.294	7.471	4.260
Average critical values		7.618	4.417	7.474	4.398

One may also ask about possible influence of the maximal dimension K upon the critical values. Earlier experiences with data-driven tests argue for stable behaviour of the critical values when $K > 5$ (cf. Table 1 in Kallenberg and Ledwina (1997b)).

In Table 2 we show critical values for three various K in *normal* case for $c = 0.5$, four pairs of (α, β) and α_0, n and M as in Table 1. Indeed, the results presented in Table 2 confirm the expected stability of critical values. Observed small fluctuations are probably caused by poor stability of the estimator $\hat{\vartheta}$.

In all cases the selection rule $\hat{S}$ chooses the dimension $k = 1$ at a stable frequency 88–90%. This stands in accordance with (3.18).

In the context of financial time series, $n = 500$ observations correspond roughly to 2 years of daily quotations. Still, in empirical economics there is a vast research done with larger data sets. From this point of view it is reasonable to examine changes of critical values under varying n. Stabilization with a growing sample size is expected. This is confirmed by Table 3. As previously, we present here only the *normal* case, but the same behaviour was observed also for the *Laplace* case. To show the stability of the procedure, this time we took some different pairs of GARCH parameters. Even though the stabilization of critical values is not too fast, the oscillations essentially decrease with growing n. This justifies the use of average critical values for $n \geqslant 500$, which we recommend for practical implementation.

TABLE 2. *Normal* case. Simulated critical values of $\hat{W}_{\hat{S}}$ for various K. Significance level 0.05, $\alpha_0 = 0.001$, $h = 0.1$, $c = 0.5$, $n = 500$, $M = 5000$ Monte Carlo runs

α	β	$K = 6$	$K = 10$	$K = 15$
0.3	0.5	7.246	7.435	7.276
0.2	0.7	8.017	7.944	7.838
0.4	0.5	7.504	7.345	7.378
0.25	0.65	7.279	7.413	7.721

TABLE 3. *Normal* case. Simulated critical values of $\hat{W}_{\hat{S}}$ for various n. Significance level 0.05, $\alpha_0 = 0.001$, $h = 0.1$, $K = 10$, $c = 0.5$, $M = 5000$ Monte Carlo runs

α	β	$n = 300$	$n = 500$	$n = 800$
0.3	0.4	8.752	7.600	7.057
0.3	0.5	8.104	7.435	6.976
0.2	0.7	10.941	7.944	7.199
0.3	0.65	8.285	7.974	7.112

4.3. Power behaviour and comparison of powers. In the power-behaviour study two competing tests were taken into account. The first one was proposed by Bai (2003) and is obtained by Khmaladze transformation of the empirical process and an application of Kolmogorov–Smirnov statistic. The second is the Jarque–Bera (J–B) normality test adopted for the GARCH case by Fiorentini et al. (2004). It is constructed on the basis of the third and fourth sample moments and provides an asymptotically chi-square distributed test statistic with 2 degrees of freedom under H_0.

For both null hypotheses we considered the following alternatives:

- the family of standardized Generalized Error Distributions (GED(ν) for short) with $\nu \in (1, 2)$, see Remark 4.1 below;

• standardized t-Student distribution with 5 degrees of freedom, considered also in Bai (2003) (t-Student(5) for short);

• standardized χ^2 distribution with 5 degrees of freedom ($\chi^2(5)$ for short); and, additionally, standard Laplace for the *normal* null hypothesis and vice versa.

Remark 4.1. Recall that Generalized Error Distribution (or Generalized Exponential Distribution, see Mittnik et al. (1998)) with parameter ν is defined by the density

$$f(y) = \tfrac{1}{2}\nu C(\nu)[\Gamma(1/\nu)]^{-1}\exp\{-(C(\nu)|y|)^{\nu}\},$$

where $\nu > 0$ and $C(\nu) = \sqrt{\Gamma(3/\nu)[\Gamma(1/\nu)]^{-1}}$ while $\Gamma(p)$ denotes the Euler gamma function. The standard normal distribution corresponds to the case $\nu = 2$ whereas $\nu = 1$ gives the standard Laplace distribution. By changing ν inside $(1, 2)$ departures from our both null hypotheses can be modelled in a continuous manner.

In order to inspect in a continuous way the sensitivity of compared tests we considered the family of contaminated alternatives of the form

$$f_{\varrho}(y) = (1-\varrho)f(y)+\varrho f_1(y) = f(y)+\varrho(f_1(y)-f(y)), \quad \text{where } \varrho\in[0, 1]. \tag{4.4}$$

As previously, f is the null density while f_1 denotes one of the alternative densities listed above.

For the two competing tests, the corresponding critical values were calculated by $M = 5000$ Monte Carlo runs implementing appropriate algorithms evaluating the respective test statistics. We obtained the following critical values at 0.05 significance level for $n = 500$:

normal case – $CV_{\mathrm{BAI}} = 5.305$ (asymptotic critical value 2.22) and $CV_{\mathrm{J-B}} = 5.612$ (asymptotic critical value 5.991);

Laplace case – $CV_{\mathrm{BAI}} = 2.580$ (asymptotic critical value 2.22).

For our test (denoted by $\hat{W}_{\hat{S}}$) we used the average critical values from Table 1.

The results shown in Tables 4 and 5 were calculated by $M = 2000$ Monte Carlo runs for the GARCH(1,1) model with $\alpha_0 = 0.001$, $\alpha = 0.3$, $\beta = 0.5$ and with alternative densities given by (4.4).

Comparing two first columns in Table 4 it can be observed that changing c below the value $c = 1$ can sometimes improve the power of $\hat{W}_{\hat{S}}$ e.g. for t-Student and chi-square distributions. Bai test scarcely manages to compete with $\hat{W}_{\hat{S}}$ and J–B tests in the case of GED alternatives. On the other hand, $\hat{W}_{\hat{S}}$ competes well with J–B test and often outperforms it.

TABLE 4. *Normal* case. Empirical powers of $\hat{W}_{\hat{S}}$, Bai and J–B tests. Significance level 0.05, $\alpha_0 = 0.001$, $h = 0.1$, $\alpha = 0.3$, $\beta = 0.5$, $K = 10$, $M = 2000$ Monte Carlo runs

Alternative density f_1	ϱ	Empirical powers (in %)			
		$\hat{W}_{\hat{S}}$ $c = 0.5$	$\hat{W}_{\hat{S}}$ $c = 1$	BAI	J–B
Laplace	0.5	93	93	55	88
	0.8	100	100	80	95
GED(1.25)	0.5	60	60	34	58
	0.9	97	98	55	95
GED(1.5)	0.8	52	52	26	50
	1.0	69	72	32	65
t-Student(5)	0.8	90	88	69	95
	1.0	98	98	80	100
$\chi^2(5)$	0.4	77	40	71	93
	0.6	99	84	90	100

As seen in Table 5, Bai test fails to detect departures from the non-Gaussian null distribution. In contrast, $\hat{W}_{\hat{S}}$ retains its good performance against various types of alternatives, which confirms its omnibus character. For the *Laplace* case the gain of the power when choosing $c = 0.5$ is more visible than in the *normal* case.

Taking into account the above results we recommend our data-driven score test with $c = 0.5$ as a sensitive tool for testing the noise distribution in GARCH(1,1) model.

TABLE 5. *Laplace* case. Empirical powers of $\hat{W}_{\hat{S}}$ and Bai tests. Significance level 0.05, $\alpha_0 = 0.001$, $h = 0.1$, $\alpha = 0.3$, $\beta = 0.5$, $K = 10$, $M = 2000$ Monte Carlo runs

Alternative density f_1	ϱ	Empirical powers (in %)		
		$\hat{W}_{\hat{S}}$ $c = 0.5$	$\hat{W}_{\hat{S}}$ $c = 1$	BAI
normal	0.5	75	66	10
	0.9	100	100	27
GED(1.75)	0.5	66	51	8
	0.9	100	100	17
GED(1.5)	0.8	73	71	9
	1.0	92	92	13
t-Student(5)	0.8	47	44	8
	1.0	69	66	10
$\chi^2(5)$	0.4	81	61	5
	0.6	100	97	5

5. PROOFS

Proof of Proposition 3.1. Observe that from the definition of q_t given in (3.2) we obtain the recurrence formula

$$Q_t = \alpha_0 + \alpha X_{t-1}^2 + \beta Q_{t-1} = \alpha_0 + (\alpha \tilde{\varepsilon}_{t-1}^2 + \beta) Q_{t-1}, \quad t = 2, \ldots, n.$$

Thus $(X_1, \ldots, X_n) = (\tilde{\varepsilon}_1 \sqrt{Q_1}, \ldots, \tilde{\varepsilon}_n \sqrt{Q_n})$ can be explicitly expressed as a measurable function of $(\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n)$. Hence and by the conditional density of $(X_1, \ldots, X_n)$, given $h_1 = h$, under the null hypothesis (see (3.8)) we see, by standard calculations, that the joint density of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ is simply $f(y_1) \ldots f(y_n)$. This completes the proof. ∎

Proof of Proposition 3.2. Introduce further notation putting

$$(5.1) \qquad \tilde{e}_t^2 = \alpha \tilde{\varepsilon}_t^2 + \beta, \quad t = 1, 2, \ldots$$

By Proposition 3.1 and the square integrability of Φ_j's it follows that components of l_θ, given in (3.6), have finite variances under P_h. On the other hand, the assumption (A3) guarantees that $\zeta(\tilde{\varepsilon}_t)$ has a finite second moment. As $\zeta(\tilde{\varepsilon}_t)$ is independent of $Q_t^{-1} \partial Q_t / \partial \vartheta$, finiteness of the second moment of l_ϑ, given by (3.7), is implied by (5.3) of the following lemma.

LEMMA 5.1. *Let Q_t, $t = 2, \ldots, n$, be defined as in* (3.7). *Then for every $s = 0, 1, \ldots, t-2$ and $t = 2, 3, \ldots$ it follows that*

$$(5.2) \qquad \frac{X_{t-1-s}^2}{Q_t} \leqslant \frac{\tilde{\varepsilon}_{t-s-1}}{\tilde{e}_{t-1}^2 \ldots \tilde{e}_{t-s-1}^2} \leqslant \frac{1}{\alpha \beta^s} \text{ a.s. under } P_h.$$

Moreover, for every $r \geqslant 1$

$$(5.3) \qquad \sup_{t \geqslant 2} E_h \left\| \frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right\|^r < \infty.$$

Lemma 5.1 can be proved in a similar way as (A.9) in the Appendix and goes along the lines of the proof of Lemma 3 in the Appendix in Lumsdaine (1996), so we omit it. Finally, the relation $E_h l(\vartheta) = 0$ is a general property of a score vector. ∎

Proof of Proposition 3.3. First we shall prove that the components of $l_\vartheta = \partial L_k / \partial \vartheta$, given by (3.7), are linearly independent random variables. On the contrary, suppose that there exist real numbers C_1, C_2, C_3 such that under P_h

$$(5.4) \qquad \sum_{t=2}^{n} \frac{\zeta(\tilde{\varepsilon}_t)}{Q_t} \Big[\sum_{s=0}^{t-2} \beta^s \big(C_1 + C_2 X_{t-1-s}^2 + C_3 s \beta^{-1} (\alpha_0 + \alpha X_{t-1-s}^2)\big) + C_3 (t-1) \beta^{t-2} h \Big] = 0 \text{ a.s.}$$

Let us put

$$U_t = \sum_{s=0}^{t-2} \left(C_1 \beta^s + C_2 \beta^s X_{t-1-s}^2 + C_3 s\beta^{s-1} (\alpha_0 + \alpha X_{t-1-s}^2)\right) + C_3 (t-1) \beta^{t-2} h.$$

Then transposing the last summand in (5.4) to the right-hand side we get

$$\sum_{t=2}^{n-1} \frac{\zeta(\tilde{\varepsilon}_t)}{Q_t} U_t = -\zeta(\tilde{\varepsilon}_n) \frac{U_n}{Q_n} \text{ a.s.} \tag{5.5}$$

Both the left-hand side and the second factor on the right-hand side are σ_{n-1}-measurable. Since, by the assumption (A2), $\zeta(\tilde{\varepsilon}_n)$ is not a constant random variable, it follows that there exists a measurable set $A \subset R$ such that

$$0 < P_h\left(\zeta(\tilde{\varepsilon}_n) \in A\right) < 1. \tag{5.6}$$

From (5.5) we infer that

$$\mathcal{N}_0 = \{U_n \neq 0, \zeta(\tilde{\varepsilon}_n) \in A\} = \left\{U_n \neq 0, -\frac{Q_n}{U_n} \sum_{t=2}^{n-1} \frac{\zeta(\tilde{\varepsilon}_t) U_t}{Q_t} \in A\right\} \in \sigma_{n-1}.$$

By Proposition 3.1 the event $\{\zeta(\tilde{\varepsilon}_n) \in A\}$ is independent of σ_{n-1}, and hence independent of $\mathcal{N}_0$. This implies

$$P_h(\mathcal{N}_0) = P_h\left(\mathcal{N}_0 \cap \{\zeta(\tilde{\varepsilon}_n) \in A\}\right) = P_h(\mathcal{N}_0) P_h\left(\zeta(\tilde{\varepsilon}_n) \in A\right).$$

Therefore, by (5.6), we get $P_h(\mathcal{N}_0) = 0$, and hence $U_n = 0$ a.s., which reads as

$$C_1 + C_2 Q_{n-1} \tilde{\varepsilon}_{n-1}^2 + \sum_{s=1}^{n-2} \left(C_1 \beta^s + C_2 \beta^s X_{n-1-s}^2 + C_3 s\beta^{s-1} (\alpha_0 + \alpha X_{n-1-s}^2)\right) + C_3 (n-1) \beta^{n-2} h = 0 \text{ a.s.} \tag{5.7}$$

This implies $C_2 = 0$ because $\tilde{\varepsilon}_{n-1}^2$ is a nonconstant random variable independent of σ_{n-2} and Q_{n-1} is positive. Using the relation $X_{n-2}^2 = Q_{n-2} \tilde{\varepsilon}_{n-2}^2$, we obtain (5.7) in the form

$$C_1 \sum_{s=0}^{n-2} \beta^s + C_3 \left\{ \sum_{s=2}^{n-2} s\beta^{s-1} (\alpha_0 + \alpha X_{n-1-s}^2) + (n-1) \beta^{n-2} h + \alpha_0 \right\} + C_3 \alpha Q_{n-2} \tilde{\varepsilon}_{n-2}^2 = 0 \text{ a.s.}$$

Exploiting again the property that $\tilde{\varepsilon}_{n-2}$ is independent of σ_{n-3}, we get $C_3 = 0$ similarly as previously, and consequently $C_1 = 0$.

Now, to prove linear independence of all components of $l(\vartheta)$ suppose, on the contrary, that there exist scalars $D_1, \ldots, D_k$ and C_1, C_2, C_3 such that

$$\sum_{s=1}^{n} \sum_{j=1}^{k} D_j \Phi_j(F(\tilde{\varepsilon}_s)) = \sum_{t=2}^{n} \frac{\zeta(\tilde{\varepsilon}_t)}{Q_t} \left(C_1 \frac{\partial Q_t}{\partial \alpha_0} + C_2 \frac{\partial Q_t}{\partial \alpha} + C_3 \frac{\partial Q_t}{\partial \beta}\right) \text{ a.s.} \tag{5.8}$$

Taking the conditional expectation with respect to σ_{n-1} over both sides of (5.8), remembering that $\partial Q_t/\partial\vartheta$ is σ_{n-1}-measurable for $t \leqslant n$ and that $E_h\Phi(F(\tilde{\varepsilon}_t)) = 0$, $E_h\zeta(\tilde{\varepsilon}_t) = 0$, we obtain

$$(5.9)\qquad \sum_{s=1}^{n-1}\sum_{j=1}^{k} D_j\Phi_j(F(\tilde{\varepsilon}_s)) = \sum_{t=2}^{n-1}\frac{\zeta(\tilde{\varepsilon}_t)}{Q_t}\left(C_1\frac{\partial Q_t}{\partial\alpha_0}+C_2\frac{\partial Q_t}{\partial\alpha}+C_3\frac{\partial Q_t}{\partial\beta}\right) \text{ a.s.}$$

Repeating this step by taking successive conditional expectations with respect to $\sigma_{n-2}, \ldots, \sigma_1$ we finally get $\sum_{j=1}^{k} D_j\Phi_j(F(\tilde{\varepsilon}_1)) = 0$ a.s. By the orthogonality of Φ_j's the last equality holds if and only if $D_1 = \ldots = D_k = 0$. This completes the proof. ■

Proof of Theorem 3.6. The proof is based on the following theorem of Kundu et al. (2000).

THEOREM (Kundu et al. (2000)). *Let for every $n \geqslant 1$, $\{X_{tn}\}_{t=1,\ldots,n}$, defined on $(\Omega, \mathscr{F}, P)$, be an R^k-valued martingale difference array with respect to an increasing sequence $\{\sigma_{t,n}\}_{t=1,\ldots,n}$ of sub-σ-fields of $\mathscr{F}$. Assume $E(\|X_{tn}\|^2) < \infty$ for every $1 \leqslant t \leqslant n$ and $n \geqslant 1$, where $\|\cdot\|$ denotes the Euclidean norm. Further assume that*

(i) *there exists a positive definite matrix Σ such that for every $b \in R^k$*

$$\sum_{t=1}^{n} E\left((b^T X_{tn})^2 \mid \sigma_{t-1,n}\right) \xrightarrow{P} b^T\Sigma b;$$

(ii) *for some basis $\phi = (\phi_1, \ldots, \phi_k)$ in R^k and any $\delta > 0$ the following Lindeberg-type condition holds*:

$$\sum_{t=1}^{n} E\left((\phi_j^T X_{tn})^2 \mathbf{1}_{\{|\phi_j^T X_{tn}| > \delta\}} \mid \sigma_{t-1,n}\right) \xrightarrow{P} 0 \qquad \text{for } j = 1, \ldots, k.$$

Then $\sum_{t=1}^{n} X_{tn} \xrightarrow{\mathscr{D}} N(0, \Sigma)$.

It is easily seen that if $E\|X_{tn}\|^3 < \infty$ for all $1 \leqslant t \leqslant n$ and $n \geqslant 1$, then (ii) may be replaced by a stronger Lyapunov-type condition

$$(5.10)\qquad \sum_{t=1}^{n} E(\|X_{tn}\|^3 \mid \sigma_{t-1,n}) \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

In our application of the above theorem we take X_{tn} as given in (3.13), considered as random variables on $(\Omega, \mathscr{F}_0, P_h)$, where $\mathscr{F}_0 = \sigma(X_1, X_2, \ldots)$. To check (i) and (5.10) we shall need the following lemma proved in Section 6.

LEMMA 5.2. *For almost every $h > \kappa_0$ and under the assumptions of Theorem 3.6 matrices $B_{12}^{(n)}(\vartheta)$ and $B_{22}^{(n)}(\vartheta)$ defined in (3.9) converge, as $n \to \infty$, to the limiting matrices $B_{12}^{\infty}(\vartheta)$ and $B_{22}^{\infty}(\vartheta)$ given explicitly in (6.2) and (6.3), respectively. Moreover, $B_{22}^{\infty}(\vartheta)$ is nonsingular for any $\vartheta \in \Theta$.*

Checking (5.10). As, by Lemma 5.2, the normalizing matrix in the formula defining X_{tn} converges to some limiting one, to have (5.10) it is enough to prove

$$(5.11) \qquad \sum_{t=1}^{n} E_h(\|Y_{tn}\|^3 \mid \sigma_{t-1,n}) \xrightarrow{P_h} 0.$$

By the triangle inequality and the elementary inequality $(x+y)^3 \leqslant 4x^3+4y^3$ we get

$$(5.12) \qquad \|Y_{tn}\|^3 \leqslant 4n^{-3/2}\|\Phi(F(\tilde{\varepsilon}_t))\|^3 + \frac{1}{2}n^{-3/2}|\zeta(\tilde{\varepsilon}_t)|^3 \left\| B_{12}^{(n)}(\vartheta)[B_{22}^{(n)}(\vartheta)]^{-1}\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\|^3.$$

Taking the conditional expectation on both sides of (5.12) and remembering that $\tilde{\varepsilon}_t$ is independent of $\sigma_{t-1,n}$ while Q_t is measurable with respect to $\sigma_{t-1,n}$, we estimate the sum in (5.11) by

$$4n^{-1/2}E\|\Phi(F(\varepsilon))\|^3 + \frac{1}{2}n^{-3/2}(E|\zeta(\varepsilon)|^3)\|B_{12}^{(n)}(\vartheta)[B_{22}^{(n)}(\vartheta)]^{-1}\|^3 \sum_{t=1}^{n}\left\|\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\|^3 \quad \text{a.s.},$$

where for a matrix $A = [a_{ij}]$ we put $\|A\|^2 = \sum_{i,j} a_{ij}^2$. Using boundedness of Φ, (5.3) of Lemma 5.1 and again Lemma 5.2 we see that the sum in (5.11) is $O_{P_h}(n^{-1/2})$, which proves (5.11).

Checking (i). We shall show that (i) holds with $\Sigma = I$. In fact, we shall show even a little more proving that

$$(5.13) \qquad \sum_{t=1}^{n} E_h(X_{tn}X_{tn}^T \mid \sigma_{t-1,n}) \xrightarrow{P_h} I,$$

where I stands for the $k\times k$ identity matrix. From Lemma 5.2 we have

$$\mathscr{M}^{(n)}(\vartheta) \to I - B_{12}^{\infty}(\vartheta)[B_{22}^{\infty}(\vartheta)]^{-1}B_{21}^{\infty}(\vartheta) = \mathscr{M}^{\infty}(\vartheta) \quad \text{as } n\to\infty.$$

So, by the definition of X_{tn}, to get (5.13) it is enough to prove that

$$(5.14) \qquad \sum_{t=1}^{n} E_h(Y_{tn}Y_{tn}^T \mid \sigma_{t-1,n}) \xrightarrow{P_h} \mathscr{M}^{\infty}(\vartheta).$$

Using (3.12) we get

$$(5.15) \qquad \sum_{t=1}^{n} E_h(Y_{tn}Y_{tn}^T \mid \sigma_{t-1,n}) = \mathscr{M}_{1n} + \mathscr{M}_{2n} + \mathscr{M}_{3n},$$

where

$$(5.16) \qquad \mathscr{M}_{1n} = \frac{1}{n}\sum_{t=1}^{n} E_h\big(\Phi(F(\tilde{\varepsilon}_t))[\Phi(F(\tilde{\varepsilon}_t))]^T \mid \sigma_{t-1,n}\big),$$

$$\mathscr{M}_{2n} = \frac{1}{n}\sum_{t=1}^{n} E_h\left(\Phi(F(\tilde{\varepsilon}_t))\zeta(\tilde{\varepsilon}_t)\left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right]^T[B_{22}^{(n)}(\vartheta)]^{-1}B_{21}^{(n)}(\vartheta)\,\middle|\,\sigma_{t-1,n}\right),$$

$$\mathcal{M}_{3n} = \frac{1}{4n} \sum_{t=1}^{n} E_h \Bigg([\zeta(\tilde{\varepsilon}_t)]^2 B_{12}^{(n)}(\vartheta)$$

$$\times [B_{22}^{(n)}(\vartheta)]^{-1} \left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]\left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]^T [B_{22}^{(n)}(\vartheta)]^{-1} B_{21}^{(n)}(\vartheta) \Bigg| \sigma_{t-1,n} \Bigg).$$

Since $\tilde{\varepsilon}_t$ is independent of $\sigma_{t-1,n}$, we have (by orthonormality of Φ)

$$\mathcal{M}_{1n} = \frac{1}{n} \sum_{t=1}^{n} I = I,$$

$$\mathcal{M}_{2n} = \Delta \left(\frac{1}{n} \sum_{t=1}^{n} \frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta} \right)^T [B_{22}^{(n)}(\vartheta)]^{-1} B_{21}^{(n)}(\vartheta), \tag{5.17}$$

where Δ is defined in (4.3), and from (2.6) we obtain

$$\mathcal{M}_{3n} = \frac{J_f}{4} B_{12}^{(n)}(\vartheta) [B_{22}^{(n)}(\vartheta)]^{-1} \times \left(\frac{1}{n} \sum_{t=1}^{n} \left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]\left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]^T \right) [B_{22}^{(n)}(\vartheta)]^{-1} B_{21}^{(n)}(\vartheta). \tag{5.18}$$

Now, if we shall show that for almost every $h > \kappa_0$ it follows that

$$\Delta \left(\frac{1}{n} \sum_{t=1}^{n} \frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta} \right)^T \xrightarrow{P_h} -2B_{12}^{\infty}(\vartheta) \tag{5.19}$$

and

$$\frac{1}{n} \sum_{t=1}^{n} \left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]\left[\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right]^T \xrightarrow{P_h} \frac{4}{J_f} B_{22}^{\infty}(\vartheta), \tag{5.20}$$

then Lemma 5.2 together with (5.17) and (5.18) will imply (5.14). This, however, is stated in the following lemma which is proved in Section 6.

LEMMA 5.3. *Under the assumptions of Theorem* 3.6, (5.19) *and* (5.20) *are fulfilled for almost every* $h > \kappa_0$.

The proof of Theorem 3.6 is complete. ■

LEMMA 5.4. *Suppose* $\hat{\vartheta} = (\hat{\alpha}_0, \hat{\alpha}, \hat{\beta})$ *is a consistent estimator of* ϑ *and* $E\varepsilon^6 < \infty$. *Then for almost every* $h > \kappa_0$, $\hat{B}_{12}^{(n)}$ *and* $\hat{B}_{22}^{(n)}$, *given by* (4.1) *and* (4.2), *are consistent estimators of* $B_{12}^{(n)}(\vartheta)$ *and* $B_{22}^{(n)}(\vartheta)$, *respectively, under* P_h, *i.e.*

$$\hat{B}_{12}^{(n)} - B_{12}^{(n)}(\vartheta) \xrightarrow{P_h} 0 \quad \text{and} \quad \hat{B}_{22}^{(n)} - B_{22}^{(n)}(\vartheta) \xrightarrow{P_h} 0.$$

Proof. By the assumption $E\varepsilon^6 < \infty$ and the argument used in Lemma 2 of Lumsdaine (1996) it follows that for $t > 20$

$$\frac{Q_t}{\hat{Q}_t} \leqslant \frac{1}{\hat{\alpha}\hat{\beta}^{19}} H_t, \qquad \sup_{t>20} E_h H_t^{9/4} < \infty, \tag{5.21}$$

where $H_t = (1+\hat{\varepsilon}_{t-1}^2)\dots(1+\hat{\varepsilon}_{t-20}^2)(\hat{\varepsilon}_{t-1}^2+\dots+\hat{\varepsilon}_{t-20}^2)^{-1}$. Now introduce some additional auxiliary notation setting for $b\in(0,1)$ and $t\geqslant 4$

$$d_{1t}(b)=\sum_{s=1}^{t-2} sb^{s-1}\frac{X_{t-1-s}^2}{Q_t},\qquad d_{2t}(b)=\sum_{s=2}^{t-2} s(s-1)b^{s-2}\frac{X_{t-1-s}^2}{Q_t}.$$

By a similar reasoning to that in (A.10) but using (6.1) instead of (2.3) we prove that, for any $r\geqslant 1$ and $b\leqslant(\xi_r^*)^{-1}$ with ξ_r^* defined in (A.8),

$$\sup_{t\geqslant 4} E_h d_{jt}^r(b)<\infty,\qquad j=1,2. \tag{5.22}$$

Take any $h>\kappa_0$ for which (5.19) and (5.20) hold. Observe that, by Lemma 5.2, to show the consistency of $\hat{B}_{12}^{(n)}$ and $\hat{B}_{22}^{(n)}$ it is enough to prove that

$$\frac{1}{n}\sum_{t=2}^{n}\left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\|\xrightarrow{P_h}0 \tag{5.23}$$

and

$$\frac{1}{n}\sum_{t=2}^{n}\left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)^T-\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right)\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right)^T\right\|\xrightarrow{P_h}0. \tag{5.24}$$

We have

$$\begin{aligned}\left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\| &\leqslant\frac{1}{\hat{Q}_t}\left\|\left(\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{\partial Q_t}{\partial\vartheta}\right\|+\left\|\left(\frac{1}{\hat{Q}_t}-\frac{1}{Q_t}\right)\frac{\partial Q_t}{\partial\vartheta}\right\|\\ &\leqslant\frac{1}{\hat{Q}_t}\left\|\left(\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{\partial Q_t}{\partial\vartheta}\right\|+\left\|\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\|\left\|\frac{1}{\hat{Q}_t}D_t\right\|\|\hat{\vartheta}-\vartheta\|,\end{aligned} \tag{5.25}$$

where $D_2^T=[1,\ X_1^2,\ h]$ and for $t>2$

$$\begin{aligned}D_t^T=\Big[&\sum_{s=0}^{t-2}\hat{\beta}^s,\ \sum_{s=0}^{t-2}\hat{\beta}^s X_{t-1-s}^2,\\ &\sum_{s=1}^{t-2}(\hat{\beta}^{s-1}+\dots+\beta^{s-1})(\alpha_0+\alpha X_{t-1-s}^2)+h(\hat{\beta}^{t-2}+\dots+\beta^{t-2})\Big].\end{aligned}$$

An elementary calculation shows that, using (5.21), for $t>20$ and on the event $\{\hat{\beta}<\beta_0\}$ with $\beta_0\in(\beta,1)$ we have the following estimates (cf. Lemmas 1, 2 and 3 in Lumsdaine (1996) for similar considerations):

$$\left\|\frac{1}{Q_t}\frac{\partial Q_t}{\partial\vartheta}\right\|\leqslant C_1+\alpha d_{1t}(\beta_0),\qquad\left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial\vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)\right\|\leqslant C_2+\hat{\beta}^{-19}H_t d_{1t}(\beta_0),$$

$$\left\|\frac{1}{\hat{Q}_t}D_t\right\|\leqslant C_3+\alpha\hat{\alpha}^{-1}\hat{\beta}^{-19}H_t d_{1t}(\beta_0)$$

and

$$\frac{1}{\hat{Q}_t}\left\|\left(\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{\partial Q_t}{\partial \vartheta}\right\| \leqslant (C_4+2\hat{\alpha}^{-1}\hat{\beta}^{-19}H_t d_{1t}(\beta_0)+\hat{\beta}^{-19}H_t d_{2t}(\beta_0))\|\hat{\vartheta}-\vartheta\|,$$

where C_j, $j = 1, 2, 3, 4$, depend on ϑ, $\hat{\vartheta}$, h and β_0 but do not depend on t. Hence and from (5.25) we get

$$(5.26)\quad \left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right\| \leqslant \{C_4+C_1 C_3+C_3\alpha d_{1t}(\beta_0)$$
$$+(2+C_1\alpha)\hat{\alpha}^{-1}\hat{\beta}^{-19}H_t d_{1t}(\beta_0)+\hat{\beta}^{-19}H_t d_{2t}(\beta_0)+\alpha^2\hat{\alpha}^{-1}\hat{\beta}^{-19}H_t d_{1t}^2(\beta_0)\}$$
$$\times\|\hat{\vartheta}-\vartheta\| = \mathscr{H}_t\|\hat{\vartheta}-\vartheta\|.$$

Choose $\beta_1 > \beta$ such that $\beta_1 < (\xi_4^*)^{-1}$, where ξ_r^* is defined in (A.8) Then by (5.22) and the Schwarz inequality it follows that $d_{jt}(\beta_1)$, $H_t d_{jt}(\beta_1)$ and $H_t d_{jt}^2(\beta_1)$ are bounded in $L_1(P_h)$ uniformly with respect to t, $t > 20$. This together with (5.26) and consistency of $\hat{\vartheta}$ proves (5.23).

To prove (5.24) write for $t > 20$ and $\beta_0 \in (\beta, 1)$

$$(5.27)\quad \left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)^T-\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right)\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right)^T\right\|$$
$$\leqslant \left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)-\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right\|\left(\left\|\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right\|+\left\|\left(\frac{1}{\hat{Q}_t}\frac{\partial Q_t}{\partial \vartheta}\bigg|_{\vartheta=\hat{\vartheta}}\right)\right\|\right)$$
$$\leqslant \mathscr{H}_t(C_1+C_2+\alpha d_{1t}(\beta_0)+\hat{\beta}^{-19}H_t d_{1t}(\beta_0))\|\hat{\vartheta}-\vartheta\|.$$

Now, choose $\beta_2 > \beta$ such that $\beta_2 < (\xi_{27}^*)^{-1}$. Then from (5.26) and (5.27) we see that all expressions on the right-hand side of (5.27) that depend on t are of the form $H_t^{s_1} d_{jt}^{s_2}(\beta_2)$ with $s_1 = 1, 2$ and $s_2 = 1, 2, 3$. By Hölder's inequality and (5.21) we infer that they are bounded in $L_1(P_h)$ uniformly with respect to t, $t > 20$. So, consistency of $\hat{\vartheta}$ implies (5.24). ∎

6. PROOFS OF AUXILIARY LEMMAS

It is more convenient to prove first Lemma 5.3.

Proof of Lemma 5.3. We shall infer (5.19) and (5.20) from (A.12) and (A.13) in the Appendix by a measure theoretic considerations.

Let us put $Z^+ = \{1, 2, \ldots\}$ and $Z^- = \{0, -1, -2, \ldots\}$, write $R^Z = R^{Z^-}\times R^{Z^+}$, and let μ be the distribution of the i.i.d. sequence $\{e_t^2\}$ on R^Z. If μ^+ denotes the distribution of $\{e_1^2, e_2^2, \ldots\}$ on R^{Z^+} and μ^- the distribution of $\{e_0^2, e_{-1}^2, \ldots\}$ on R^{Z^-}, then $\mu = \mu^- \times \mu^+$. Observe that h_1 is a measurable function of $e_0^2, e_{-1}^2, \ldots$ So, denoting by ν_1 its distribution, we see that the representation (2.4) implies that ν_1 is absolutely continuous with respect to the Lebesgue measure and is supported on (κ_0, ∞). In consequence, the distribution of $\{h_1, e_1^2, e_2^2, \ldots\}$ on $R \times R^{Z^+}$ is $\nu_1 \times \mu^+$.

Now, inspecting the formula (A.11) and using again (2.4) and (2.5), we see that elements of matrices

$$\frac{1}{n}\sum_{s=0}^{t-2}\frac{1}{h_s}\frac{\partial h_s}{\partial \vartheta} \quad \text{and} \quad \frac{1}{n}\sum_{s=0}^{t-2}\left(\frac{1}{h_s}\frac{\partial h_s}{\partial \vartheta}\right)\left(\frac{1}{h_s}\frac{\partial h_s}{\partial \vartheta}\right)^T$$

are images of the random vector $\{h_1, e_1^2, e_2^2, \ldots\}$ by measurable functions. So, almost-sure convergence in (A.12) and (A.13) means almost sure convergence with respect to $\nu_1 \times \mu^+$. By the Fubini theorem for almost every $h > \kappa_0$ (with respect to ν_1 or, equivalently, with respect to the Lebesgue measure) these functions converge μ^+ almost surely, and hence in the measure μ^+, to the same limits.

On the other hand, iterating the recurrence formula

$$Q_t = \alpha_0 + \alpha X_{t-1}^2 + \beta Q_{t-1}, \; t \geqslant 2, \quad Q_1 = h, \tag{6.1}$$

we see, as previously, that for any fixed $h > \kappa_0$ and every $n \geqslant 1$, $\{X_1, \ldots, X_n\}$ is the image of $\{\tilde{e}_1^2, \tilde{e}_2^2, \ldots\}$ by a measurable mapping and the induced distribution is the conditional distribution of $\{X_1, \ldots, X_n\}$ given $h_1 = h$. This and the above imply

$$\frac{1}{n}\sum_{t=1}^{n}\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta} \xrightarrow{P_h} E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right)\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right)^T \xrightarrow{P_h} E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T[\Upsilon_1, \Upsilon_2, \Upsilon_3],$$

where Υ_i, $i = 1, 2, 3$, are defined in (A.6). Setting

$$B_{12}^{\infty}(\vartheta) = -\frac{1}{2}\Delta E[\Upsilon_1, \Upsilon_2, \Upsilon_3] \tag{6.2}$$

and

$$B_{22}^{\infty}(\vartheta) = \frac{J_f}{4}E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T[\Upsilon_1, \Upsilon_2, \Upsilon_3], \tag{6.3}$$

we infer that the relations (5.19) and (5.20) hold true. ∎

Proof of Lemma 5.2. From (3.6), (3.7) and (3.9) we have

$$nB_{12}^{(n)}(\vartheta) = E_h l_\theta l_\vartheta^T \tag{6.4}$$

$$= -\frac{1}{2}\sum_{t=2}^{n} E_h\left(\Phi(F(\tilde{\varepsilon}_t))\zeta(\tilde{\varepsilon}_t)\left(\frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}\right)^T\right) - \sum_{1 \leqslant t < r \leqslant n} E_h\left(\Phi(F(\tilde{\varepsilon}_t))\zeta(\tilde{\varepsilon}_r)\left(\frac{1}{Q_r}\frac{\partial Q_r}{\partial \vartheta}\right)^T\right).$$

By Proposition 3.1, $\tilde{\varepsilon}_r$ is independent of $\tilde{\varepsilon}_t$ for $t < r$ and independent of Q_r. Moreover, $E_h \zeta(\tilde{\varepsilon}_r) = 0$. Consequently, the second sum on the right-hand side of (6.4) vanishes. Therefore, from (5.17) we get

$$B_{12}^{(n)}(\vartheta) = -\frac{1}{2n} \Delta \sum_{t=2}^{n} E_h \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right)^T. \tag{6.5}$$

Similarly, we obtain (cf. (2.6))

$$nB_{22}^{(n)}(\vartheta) = E_h l_\vartheta l_\vartheta^T = \frac{1}{4} \sum_{t=2}^{n} E_h \left([\zeta(\tilde{\varepsilon}_t)]^2 \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right) \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right)^T \right) \tag{6.6}$$
$$+\frac{1}{2} \sum_{2 \leqslant t < r \leqslant n} E_h \left(\zeta(\tilde{\varepsilon}_t) \zeta(\tilde{\varepsilon}_r) \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right) \left(\frac{1}{Q_r} \frac{\partial Q_r}{\partial \vartheta} \right)^T \right) = \frac{J_f}{4} \sum_{t=2}^{n} E_h \left(\left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right) \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right)^T \right).$$

Hence, by the definitions (6.2) and (6.3) of matrices $B_{12}^{\infty}(\vartheta)$ and $B_{22}^{\infty}(\vartheta)$, we see that it is enough to prove (cf. (A.12) and (A.13))

$$\frac{1}{n} \sum_{t=2}^{n} E_h \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right) \to E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T \quad \text{as } n \to \infty \tag{6.7}$$

and

$$\frac{1}{n} \sum_{t=2}^{n} E_h \left(\left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right) \left(\frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \right)^T \right) \to E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T [\Upsilon_1, \Upsilon_2, \Upsilon_3] \quad \text{as } n \to \infty. \tag{6.8}$$

It will be done by showing that each element of matrices in (6.7) and (6.8) tends, when $n \to \infty$, to the corresponding element of the limiting matrices.

First, we shall show that

$$E_h \frac{1}{Q_t} \to E \frac{1}{h_1} \quad \text{as } t \to \infty \tag{6.9}$$

and for each $s \geqslant 0$

$$E_h \frac{X_{t-1-s}^2}{Q_t} \to E \frac{X_{-s}^2}{h_1} \quad \text{as } t \to \infty. \tag{6.10}$$

Fix $s \geqslant 0$. By Proposition 3.1, random variables $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_t$ under P_h have the same joint distribution as $\varepsilon_{-t+1}, \ldots, \varepsilon_0$ under P. This and (6.1) allow us to write for $t \geqslant s+4$

$$E_h \frac{X_{t-1-s}^2}{Q_t} = E_h \frac{\alpha_0 (1 + \tilde{e}_{t-2-s}^2 + \ldots + \tilde{e}_{t-2-s}^2 \ldots \tilde{e}_2^2) + \tilde{e}_{t-2-s}^2 \ldots \tilde{e}_1^2 h}{\alpha_0 (1 + \tilde{e}_{t-1}^2 + \ldots + \tilde{e}_{t-1}^2 \ldots \tilde{e}_2^2) + \tilde{e}_{t-1}^2 \ldots \tilde{e}_1^2 h} \tilde{\varepsilon}_{t-1-s}^2 \tag{6.11}$$
$$= E \frac{\alpha_0 (1 + e_{-1-s}^2 + \ldots + e_{-1-s}^2 \ldots e_{-t+3}^2) + e_{-1-s}^2 \ldots e_{-t+2}^2 h}{\alpha_0 (1 + e_0^2 + \ldots + e_0^2 \ldots e_{-t+3}^2) + e_0^2 \ldots e_{-t+2}^2 h} \varepsilon_{-s}^2.$$

Since $Ee_t^2 = \alpha+\beta < 1$ and $h > \kappa_0$ is fixed, it follows from (2.4) that for given s the numerator of the expression standing under the expectation in (6.11) tends, as $t \to \infty$, in $L_1(P)$ to h_{-s} while the denominator tends to h_1. So, the expression standing under the expectation in (6.11) tends in probability to X_{-s}^2/h_1. On the other hand, for every $t \geqslant s+4$

$$(6.12)\quad \frac{\alpha_0(1+e_{-1-s}^2+\ldots+e_{-1-s}^2\ldots e_{-t+3}^2)+e_{-1-s}^2\ldots e_{-t+2}^2 h}{\alpha_0(1+e_0^2+\ldots+e_0^2\ldots e_{-t+3}^2)+e_0^2\ldots e_{-t+2}^2 h}\varepsilon_{-s}^2$$

$$\leqslant \frac{\alpha_0(1+e_{-1-s}^2+\ldots+e_{-1-s}^2\ldots e_{-t+3}^2)}{\alpha_0(e_0^2\ldots e_{-s}^2+\ldots+e_0^2\ldots e_{-t+3}^2)}\varepsilon_{-s}^2+\frac{e_{-1-s}^2\ldots e_{-t+2}^2 h}{e_0^2\ldots e_{-t+2}^2 h}\varepsilon_{-s}^2$$

$$\leqslant \frac{2}{\alpha e_0^2\ldots e_{-s+1}^2}\leqslant \frac{2}{\alpha\beta^s}.$$

Hence (6.10) follows by the Lebesgue bounded convergence theorem.

Similarly,

$$E_h\frac{1}{Q_t} = E\frac{1}{\alpha_0(1+e_0^2+\ldots+e_0^2\ldots e_{-t+3}^2)+e_0^2\ldots e_{-t+2}^2 h},$$

where the expression in the denominator tends in probability to h_1 and is a.s. bounded from below by α_0. Therefore, (6.9) follows again by the Lebesgue theorem.

Going back to the proof of (6.7), for the first element on the left-hand side we have, by the definition of Υ_1,

$$E_h\frac{1}{n}\sum_{t=2}^{n}\sum_{s=0}^{t-2}\beta^s\frac{1}{Q_t} = \frac{1}{n}\sum_{t=2}^{n}\frac{1-\beta^{t-1}}{1-\beta}E_h\frac{1}{Q_t}\to\frac{1}{1-\beta}E\frac{1}{h_1} = E\Upsilon_1$$

as $n\to\infty$. As for the second element, let us write

$$(6.13)\quad E_h\frac{1}{n}\sum_{t=2}^{n}\sum_{s=0}^{t-2}\beta^s\frac{X_{t-1-s}^2}{Q_t}$$

$$= \sum_{s=0}^{n-2}\beta^s\left(\frac{1}{n-s-1}\sum_{t=s+2}^{n}E_h\frac{X_{t-1-s}^2}{Q_t}\right)-\frac{1}{n}\sum_{s=0}^{n-2}(s+1)\beta^s\left(\frac{1}{n-s-1}\sum_{t=s+2}^{n}E_h\frac{X_{t-1-s}^2}{Q_t}\right).$$

By (6.10), the relation

$$E_h\frac{X_{t-1-s}^2}{Q_t}\leqslant\frac{2}{\alpha}\xi_1^s$$

(cf. (A.9) in the Appendix) and $\beta\xi_1 < 1$ we infer that the first term in (6.13) converges to $E\Upsilon_2$ while the second is negligible.

The same argument works for the third element of the matrix in (6.7) because, obviously,

$$\frac{1}{n}\sum_{t=2}^{n}(t-1)\beta^{t-2}h\to 0\quad\text{as } n\to\infty.$$

The convergence in (6.8) follows by a quite similar argument, so we omit it. Nonsingularity of $B_{22}^{\infty}(\vartheta)$ is a consequence of Lemma A.2 in the Appendix. Thus Lemma 5.2 is proved. ∎

APPENDIX

In this section we prove some limit theorems being straightforward consequences of the ergodicity of the innovation sequence $\{e_t^2\}$ defined in (2.2).

Let Z denote the set of integers, R^Z be the space of doubly infinite real sequences endowed with a product σ-field $\mathscr{A}$, and let μ be the distribution of the i.i.d. sequence $\{e_t^2\}_{t\in Z}$. Note that μ depends on GARCH parameters α and β which satisfy (2.1) and are fixed from now on.

Suppose τ: $R^Z \to R$ is a measurable functional such that $E_\mu|\tau| < \infty$. Then, by the classical Birkhoff ergodic theorem (see e.g. Friedman (1970)),

$$\frac{1}{n}\sum_{j=1}^{n} \tau(\{e_{j+t-1}^2\}) \to E_\mu \tau \tag{A.1}$$

almost surely and the convergence is also in $L_1(P)$.

In order to apply (A.1) define measurable mappings ψ and ψ_1 on R^Z with values in R^Z by

$$\begin{aligned} \psi(\{z_j\}) &= \{\alpha_0+\alpha_0\sum_{s=2}^{\infty} z_{j-1}\ldots z_{j-s}\}, \\ \psi_1(\{z_j\}) &= \left\{(\alpha_0+\alpha_0\sum_{s=2}^{\infty} z_{j-1}\ldots z_{j-s})\frac{z_j-\beta}{\alpha}\right\}, \end{aligned} \tag{A.2}$$

whenever the right-hand sides of these formulae have sense and the zero sequence otherwise. Here α_0 is the third GARCH parameter appearing in (2.1). From (2.4) it follows that the formulae in (A.2) are well defined for μ-almost all sequences $\{z_j\}$, $\psi(\{e_j^2\}) = \{h_j\}$ a.s. and $\psi_1(\{e_j^2\}) = \{X_j^2\}$ a.s.

Applying mappings ψ and ψ_1 introduce three measurable functionals τ_i, $i = 1, 2, 3$, on R^Z which are, in a natural way, related to the GARCH(1,1) model, and particularly to the part l_ϑ of the score vector given in (3.7). For a sequence $\{z_j\}$ and $s \in Z$ write a coordinate selector as $\{z_j\}_s = z_s$, and set

$$\tau_1(\{z_j\}) = \sum_{s=0}^{\infty} \beta^s \frac{1}{\psi(\{z_j\})_1}, \tag{A.3}$$

$$\tau_2(\{z_j\}) = \sum_{s=0}^{\infty} \beta^s \frac{\psi_1(\{z_j\})_{-s}}{\psi(\{z_j\})_1}, \tag{A.4}$$

$$\tau_3(\{z_j\}) = \sum_{s=0}^{\infty} s\beta^{s-1} \frac{\alpha_0+\alpha\psi_1(\{z_j\})_{-s}}{\psi(\{z_j\})_1}, \tag{A.5}$$

whenever the expressions on the right-hand sides of (A.3)–(A.5) have sense and 0 otherwise. We shall show that formulae (A.3)–(A.5) are well defined for μ almost all $\{z_j\}$ and that τ_i as well as their twofold products $\tau_i \tau_k$, $i, k = 1, 2, 3$, are integrable with respect to μ. To this end define

$$\Upsilon_i = \tau_i(\{e_j^2\}), \quad i = 1, 2, 3. \tag{A.6}$$

By (A.3) we have

$$\Upsilon_1 = \frac{1}{h_1} \sum_{s=0}^{\infty} \beta^s,$$

so Υ_1 is a.s. bounded by α_0^{-1}, and hence has all moments. As to

$$\Upsilon_2 = \sum_{s=0}^{\infty} \beta^s \frac{X_{-s}^2}{h_1},$$

set $\xi_0 = 1$ and

$$\xi_r = E\left(\frac{1}{e_t^2}\right)^r = \int_R \frac{1}{(\alpha y^2 + \beta)^r} f(y)\, dy, \quad r = 1, 2, \ldots \tag{A.7}$$

Obviously, $\beta \xi_r < \xi_{r-1}$. Consequently,

$$\xi_r^* = \max_{1 \leqslant j \leqslant r} \frac{\xi_j}{\xi_{j-1}} < \frac{1}{\beta}. \tag{A.8}$$

Iterating (2.3) and using (2.2) we get

$$\frac{X_{-s}^2}{h_1} = \frac{h_{-s} \varepsilon_{-s}^2}{h_1} \leqslant \frac{h_{-s} \varepsilon_{-s}^2}{e_0^2 \ldots e_{-s}^2 h_{-s}} \leqslant \frac{1}{\alpha e_0^2 \ldots e_{-s+1}^2} \quad \text{a.s.} \tag{A.9}$$

and, consequently, by mutual independence of e_t^2's for each $r \geqslant 1$ we have

$$E\left(\sum_{s=0}^{\infty} \beta^s \frac{X_{-s}^2}{h_1}\right)^r = \sum_{s_1=0}^{\infty} \ldots \sum_{s_r=0}^{\infty} \beta^{s_1 + \ldots + s_r} E\left(\frac{X_{-s_1}^2}{h_1} \ldots \frac{X_{-s_r}^2}{h_1}\right) \tag{A.10}$$

$$\leqslant \frac{r!}{\alpha^r} \sum_{0 \leqslant s_1 \leqslant \ldots \leqslant s_r < \infty} \beta^{s_1 + \ldots + s_r} E \frac{1}{e_0^{2r} \ldots e_{-s_1+1}^{2r} e_{-s_1}^{2r-2} \ldots e_{-s_2+1}^{2r-2} e_{-s_2}^{2r-4} \ldots e_{-s_r+1}^2}$$

$$\leqslant \frac{r!}{\alpha^r} \sum_{0 \leqslant s_1 \leqslant \ldots \leqslant s_r < \infty} \beta^{s_1 + \ldots + s_r} \left(\frac{\xi_r}{\xi_{r-1}}\right)^{s_1} \left(\frac{\xi_{r-1}}{\xi_{r-2}}\right)^{s_2} \ldots (\xi_1)^{s_r}$$

$$\leqslant \frac{r!}{\alpha^r} \Big(\sum_{s=0}^{\infty} (\beta \xi_r^*)^s\Big)^r < \infty,$$

which proves that Υ_2 has also all moments. Finally, an analogous argument shows that

$$\Upsilon_3 = \sum_{s=0}^{\infty} s\beta^{s-1} \frac{\alpha_0 + \alpha X_{-s}^2}{h_1}$$

has also all moments. Thus, in particular, we have shown that τ_i and $\tau_i \tau_k$, $i, k = 1, 2, 3$, are integrable with respect to μ. This allows us to apply (A.1) to all of these functionals.

However, for further applications we need rather finite sums instead of infinite series. Note that, by the above reasoning, we have already proved that the remainders in all series defining Υ_i and $\Upsilon_i \Upsilon_k$ tend to zero in $L_1(P)$ norm at exponential rates, so, by the Borel–Cantelli lemma, also almost surely. To make this clearer consider, as an example, a functional τ_2.

Applying (A.1) to τ_2 we get

$$\frac{1}{n}\sum_{t=1}^{n} \tau_2(\{e_{j+t-1}^2\}) = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{\infty} \beta^s \frac{X_{t-1-s}^2}{h_t}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \beta^s \frac{X_{t-1-s}^2}{h_t} + \frac{1}{n}\sum_{t=1}^{n} \mathcal{R}_{2t} \to E_\mu \tau_2 = E\Upsilon_2$$

almost surely and in $L_1(P)$, where we have put

$$\mathcal{R}_{2t} = \sum_{s=t-1}^{\infty} \beta^s \frac{X_{t-1-s}^2}{h_t}.$$

By (A.9), (A.10) and stationarity of $\{X_t\}$ and $\{h_t\}$ we obtain

$$E\mathcal{R}_{2t} = \sum_{s=t-1}^{\infty} \beta^s E\frac{X_{-s}^2}{h_1} \leqslant \frac{1}{\alpha}\sum_{s=t-1}^{\infty} \beta^s \xi_1^s = \frac{1}{\alpha}\frac{(\beta\xi_1)^{t-1}}{1-\beta\xi_1} \to 0 \quad \text{as } t \to \infty$$

and for any $\eta > 0$, by the Markov inequality,

$$\sum_{t=1}^{\infty} P(\mathcal{R}_{2t} \geqslant \eta) \leqslant \frac{1}{\eta}\sum_{t=1}^{\infty} E\mathcal{R}_{2t} \leqslant \frac{1}{\eta\alpha(1-\beta\xi_1)}\sum_{t=1}^{\infty} (\beta\xi_1)^{t-1} < \infty,$$

which, by the Borel–Cantelli lemma, implies $\mathcal{R}_{2t} \to 0$ a.s. when $t \to \infty$. Hence

$$\frac{1}{n}\sum_{t=1}^{n} \mathcal{R}_{2t} \to 0$$

almost surely and in $L_1(P)$ as $n \to \infty$. Consequently, we have proved that

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{t-2} \beta^s \frac{X_{t-1-s}^2}{h_t} \to E\Upsilon_2.$$

As said before, the same statements hold for all Υ_i's and $\Upsilon_i \Upsilon_k$'s as well. Taking formula (2.5), differentiating both sides with respect to $\vartheta = (\alpha_0, \alpha, \beta)$ and dividing by h_t we get the random vector

$$\text{(A.11)} \quad \left(\frac{1}{h_t}\frac{\partial h_t}{\partial \vartheta}\right)^T = \left[\sum_{s=0}^{t-2} \beta^s \frac{1}{h_t}, \ \sum_{s=0}^{t-2} \beta^s \frac{X_{t-1-s}^2}{h_t}, \ \sum_{s=0}^{t-2} s\beta^{s-1} \frac{\alpha_0+\alpha X_{t-1-s}^2}{h_t} + (t-1)\beta^{t-2}\frac{h_1}{h_t}\right].$$

Taking into account that the random variable $(t-1)\beta^{t-2} h_1 h_t^{-1}$ tends to 0 as $t \to \infty$ almost surely and in $L_1(P)$ we summarize the above reasoning in the following proposition.

PROPOSITION A.1. *The following asymptotic results hold true almost surely and in* $L_1(P)$:

$$\text{(A.12)} \quad \frac{1}{n}\sum_{t=1}^{n} \frac{1}{h_t}\frac{\partial h_t}{\partial \vartheta} \to E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T,$$

$$\text{(A.13)} \quad \frac{1}{n}\sum_{t=1}^{n} \left(\frac{1}{h_t}\frac{\partial h_t}{\partial \vartheta}\right)\left(\frac{1}{h_t}\frac{\partial h_t}{\partial \vartheta}\right)^T \to E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T[\Upsilon_1, \Upsilon_2, \Upsilon_3],$$

where ^{T} *denotes the transposition.*

We end this section showing linear independence of Υ_1, Υ_2 and Υ_3.

LEMMA A.2. *Random variables* $\Upsilon_i = \tau_i(\{e_t^2\})$, $i = 1, 2, 3$, *defined in* (A.6) *are linearly independent. Consequently, the matrix* $E[\Upsilon_1, \Upsilon_2, \Upsilon_3]^T[\Upsilon_1, \Upsilon_2, \Upsilon_3]$ *is nonsingular.*

Proof. Suppose, on the contrary, that there exist real numbers C_i, $i = 1, 2, 3$, such that

$$\text{(A.14)} \quad C_1\Upsilon_1 + C_2\Upsilon_2 + C_3\Upsilon_3 = \frac{1}{h_1}\Big(C_1\sum_{s=0}^{\infty}\beta^s + C_2\sum_{s=0}^{\infty}\beta^s X_{-s}^2 + C_3\sum_{s=0}^{\infty} s\beta^{s-1}(\alpha_0+\alpha X_{-s}^2)\Big) = 0 \text{ a.s.}$$

Since $h_1 > \kappa_0 > 0$ a.s., we can omit this factor and write (A.14) as

$$\text{(A.15)} \quad \frac{1}{h_0}\Big(C_1\sum_{s=0}^{\infty}\beta^s + C_2\sum_{s=1}^{\infty}\beta^s X_{-s}^2 + C_3\sum_{s=1}^{\infty} s\beta^{s-1}(\alpha_0+\alpha X_{-s}^2)\Big) = -C_2\varepsilon_0^2 \text{ a.s.}$$

The left-hand side of (A.15) is $\mathscr{F}_{-1}$-measurable, hence independent of ε_0^2, which implies that $C_2\varepsilon_0^2$ is independent of itself. Since ε_0^2 is a.s. positive, we get $C_2 = 0$. Now, (A.15) can be written as

$$\text{(A.16)} \quad C_1\sum_{s=0}^{\infty}\beta^s + C_3\sum_{s=2}^{\infty} s\beta^{s-1}(\alpha_0+\alpha X_{-s}^2) = -C_3(\alpha_0+\alpha X_{-1}^2) \text{ a.s.}$$

Repeating the above argument to (A.16) and remembering that $\alpha_0, \alpha > 0$ we infer that $C_3 = 0$. Hence $C_1 = 0$ and the proof is complete. ■

Acknowledgements. The authors would like to thank T. Ledwina for recommending the paper of Ducharme and Lafaye de Micheaux and for useful and constructive comments to the first draft of this paper which helped them to improve the presentation of the results. The authors are also grateful to the referee for useful remarks. The second-named author expresses his thanks and gratitude to K. Sheppard for providing the GED distribution generator.

REFERENCES

[1] J. Bai, *Testing parametric conditional distributions of dynamic models*, Rev. Econom. Statist. 85 (2003), pp. 531–549.

[2] R. T. Baillie and T. Bollerslev, *The message in daily exchange rates: a conditional-variance tale*, J. Bus. Econom. Statist. 7 (1989), pp. 297–305.

[3] E. K. Berndt, B. H. Hall, R. E. Hall and J. A. Hausman, *Estimation inference in nonlinear structural model*, Ann. Econom. Social Measurement 4 (1974), pp. 653–665.

[4] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, J. Econometrics 31 (1986), pp. 307–327.

[5] T. Bollerslev, *A conditionally heteroskedastic time-series model for speculative prices and rates of return data*, Rev. Econometrics Statist. 69 (1987), pp. 542–547.

[6] P. Bougerol and N. M. Picard, *Stationarity of GARCH processes and of some nonnegative time series*, J. Econometrics 52 (1992), pp. 115–127.

[7] C. Brooks, S. P. Burke and G. Persand, *Benchmarks and the accuracy of GARCH model estimation*, Internat. J. Forecasting 17 (2001), pp. 45–56.

[8] W. J. Bühler and P. S. Puri, *On optional asymptotic tests of composite hypotheses with several constraints*, Z. Wahrsch. Verw. Gebiete 5 (1966), pp. 71–88.

[9] Y.-T. Chen, *A new class of characteristic-function-based distribution tests and its application to GARCH model*, Working paper, ISSP, Academia Sinica, 2002.

[10] Y.-T. Chen and C.-M. Kuan, *Generalized Jarque–Bera test of conditional normality*, Working paper, Academia Sinica, 2002.

[11] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London 1974.

[12] F. X. Diebold, *Empirical Modeling of Exchange Rate Dynamics*, Springer, New York 1988.

[13] F. C. Drost and C. A. Klaassen, *Efficient estimation in semiparametric GARCH models*, J. Econometrics 81 (1997), pp. 183–221.

[14] G. R. Ducharme and P. Lafaye de Micheaux, *Goodness-of-fit tests of normality for the innovation in ARMA models*, J. Time Ser. Anal. 25 (2004), pp. 373–395.

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

[16] G. Fiorentini, E. Sentana and G. Calzolari, *On the validity of the Jarque–Bera normality test in conditionally heteroskedastic dynamic regression models*, Econom. Lett. 83 (2004), pp. 307–312.

[17] N. Friedman, *Introduction to Ergodic Theory*, Princeton, New Jersey, 1970.

[18] C. Gouriéroux, *ARCH Models and Financial Applications*, Springer, New York 1997.

[19] R. Ignaccolo, *Goodness-of-fit tests for dependent data*, Nonparametric Statistics 16 (2004), pp. 19–38.

[20] T. Inglot, W. C. M. Kallenberg and T. Ledwina, *Power approximations to and power comparison of smooth goodness-of-fit tests*, Scand. J. Statist. 21 (1994), pp. 131–145.

[21] T. Inglot, W. C. M. Kallenberg and T. Ledwina, *Data driven smooth tests for composite hypotheses*, Ann. Statist. 25 (1997), pp. 1222–1250.

[22] T. Inglot and T. Ledwina, *Intermediate approach to comparison of some goodness-of-fit tests*, Ann. Inst. Statist. Math. 53 (2001), pp. 810–834.

[23] W. C. M. Kallenberg and T. Ledwina, *Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests*, Ann. Statist. 23 (1995), pp. 1594–1608.

[24] W. C. M. Kallenberg and T. Ledwina, *Data driven tests for composite hypotheses: Comparison of powers*, J. Stat. Comput. Simul. 59 (1997a), pp. 101–121.

[25] W. C. M. Kallenberg and T. Ledwina, *Data driven smooth tests when the hypothesis is composite*, J. Amer. Statist. Assoc. 92 (1997b), pp. 1094–1104.

[26] S. Kundu, S. Majumdar and K. Mukherjee, *Central limit theorems revisited*, Statist. Probab. Lett. 47 (2000), pp. 265–275.

[27] T. Ledwina, *Data driven version of Neyman's smooth test of fit*, J. Amer. Statist. Assoc. 89 (1994), pp. 1000–1005.

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

[29] W. K. Li and S. Ling, *On fractionally integrated autoregressive moving average time series models with conditional heteroskedasticity*, J. Amer. Statist. Assoc. 92 (1997), pp. 1184–1194.

[30] W. K. Li, S. Ling and M. McAleer, *A survey of recent theoretical results for time series models with GARCH errors*, J. Econom. Survey 16 (2002), pp. 245–269.

[31] S. Ling and M. McAleer, *Asymptotic theory for a vector ARMA–GARCH model*, Econometric Theory 19 (2003), pp. 278–308.

[32] R. L. Lumsdaine, *Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models*, Econometrica 64 (1996), pp. 575–596.

[33] T. K. Mak, H. Wong and W. K. Li, *Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares*, Comput. Statist. Data Analysis 24 (1997), pp. 169–178.

[34] H. Milbrodt and H. Strasser, *On the asymptotic power of the two-sided Kolmogorov–Smirnov test*, J. Statist. Plann. Inference 26 (1990), pp. 1–23.

[35] A. Milhøj, *The moment structure of ARCH processes*, Scand. J. Statist. 12 (1985), pp. 281–292.

[36] S. Mittnik, M. S. Paolella and S. T. Rachev, *Unconditional and conditional distributional models for the Nikkei index*, Asia-Pacific Financial Markets 5 (1998), pp. 99–128.

[37] D. B. Nelson, *Stationarity and persistence in the GARCH(1,1) model*, Econometric Theory 6 (1990), pp. 318–334.

[38] J. Neyman, *Smooth tests for goodness of fit*, Skand. Aktuarietidskr. 20 (1937), pp. 149–199.

[39] D. R. Thomas and D. A. Pierce, *Neyman's smooth goodness-of-fit test when the hypothesis is composite*, J. Amer. Statist. Assoc. 74 (1979), pp. 441–445.

[40] A. A. Weiss, *Asymptotic theory for ARCH models: Estimation and testing*, Econometric Theory 2 (1986), pp. 107–131.

Tadeusz Inglot
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
and
Polish Academy of Sciences
E-mail: inglot@im.pwr.wroc.pl

Bartosz Stawiarski
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: stawiar@im.pwr.wroc.pl

Received on 2.11.2005;
revised version on 1.12.2005