Lithuanian Mathematical Journal, Vol. 44*, No.* 1*,* 2004

A WEIGHTED GOODNESS-OF-FIT TEST FOR GARCH(1, 1) SPECIFICATION

I. Berkes

A. Rényi Institute of Mathematics, Hungarian Academy of Sciences,
P.O. Box 127, H-1364 Budapest, Hungary
(e-mail: berkes@stat.tu-graz.ac.at)

L. Horváth

Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112-0090, USA
(e-mail: horvath@math.utah.edu)

P. Kokoszka

Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA
(e-mail: piotr@math.usu.edu)

Abstract. Suppose that $\hat{r}_n(j)$ is the lag-j autocorrelation of the squared residuals computed from a realization of length n under the assumption that the observations follow a GARCH(1, 1) model. We study the asymptotic distribution of the statistics of the form $Q_n = n \sum_{1 \leqslant i \leqslant j \leqslant m(n)} \lambda_i^{1/2} \hat{r}_n(i) \gamma_m(i, j) \lambda_j^{1/2} \hat{r}_n(j)$, where the λ_j are nonnegative summable weights and the matrix $\gamma_m(i, j)$, $1 \leqslant i, j \leqslant m$, can be estimated from the data. We show that, under weak assumptions on model errors, the statistic Q_n converges in distribution to $Q = \sum_{1 \leqslant i < \infty} \lambda_i N_i^2$, where the N_i are iid standard normal. We discuss choices of the weights λ_j for which the distribution of Q is tabulated. Our results lead to and provide a rigorous justification for Portmanteau goodness-of-fit tests for GARCH(1, 1) specification.

Keywords: GARCH(1, 1) sequence, quasi-maximum likelihood estimator, squared residuals, asymptotic normality, χ^2-tests.

1. INTRODUCTION

This paper is concerned with developing the asymptotic theory for goodness-of-fit tests based on weighted sums of autocorrelations of squared GARCH(1, 1) residuals. It extends our research presented in Horváth and Kokoszka [12] and Berkes *et al.* [3] in several directions. In Horváth and Kokoszka [12], we considered the original ARCH(p) model of Engle [10] and assumed that the model errors have finite fourth moments. In Berkes *et al.* [3], we considered the more general GARCH(p, q) model of Bollerslev [6] and imposed much weaker assumptions on model errors, but focused only on a standard sum of squares statistic. The reason for considering a narrower class of statistics in the GARCH(p, q) case was that the GARCH specification is mathematically more difficult to deal with in many respects relevant to the problems considered here. The present paper shows how these problems can be addressed in the setting of the popular GARCH(1, 1) model. It can be expected that our theory can be extended to the general GARCH(p, q) specification but such an extension would potentially be

Research supported by NSF grant INT-0223262, NATO grant PST.CLG.977607 and Hungarian National Foundation for Scientific Research Grant T 29621.

Published in Lietuvos Matematikos Rinkinys, Vol. 44, No. 1, pp. 3–22, January–March, 2004. Original article submitted May 14, 2003.

very technical and is not undertaken here. We would like to point out that the GARCH(1, 1) model often exhibits features qualitatively different from those of the ARCH(p) model, and many authors focused on GARCH(1, 1) to develop new insights (see, e.g., Nelson [22], Lumsdaine [18], Teräsvirta [28], Maercker and Moser [19], Mikosch and Stărică [20], Ling and Li [17], and Berkes *et al.* [4], among others). The GARCH(1, 1) model is often sufficient to capture the main features of the volatility process and for its prediction (see, e.g., Chapter 15 of Hull [13] and Chapter 7 of Zivot and Wang [30]).

We now briefly explain the main idea of the present paper. For linear time series models, including the popular ARMA models, the usual practice to assess the goodness-of-fit of a model is to examine the plot of the autocorrelations of the residuals and check whether it looks approximately like a plot of the autocorrelations of a white noise sequence. In addition, several formal statistical tests are applied. The most popular of these, the Ljung–Box and McLeod–Li tests, are based on sums of autocorrelations of, respectively, residuals and squared residuals and have the asymptotic χ^2 distribution (see, e.g., Brockwell and Davis [7] for the details). Suppose now that $\hat{r}_n(k)$ is the lag-k autocorrelation of the squared residuals of a GARCH(1, 1) sequence (precise definitions are given in Section 2). According to Theorem 3.1, the statistic $\sum_{k=1}^{K} \hat{r}_n^2(k)$ does not converge to a χ^2 distribution. In order to obtain an asymptotic χ^2 distribution, one has to construct a quadratic form as in relation (3.15) in Theorem 3.2. Results of this type have been known some time. Using heuristic derivations, Li and Mak [15] obtained such results under the assumption of normal errors. In this paper, we provide rigorous mathematical arguments valid under very weak conditions on the errors which are not even assumed to have a density (see conditions (2.6)–(2.10)). The χ^2-statistic in (3.15) is based on K lags, where K is a fixed constant. All lags contribute essentially the same amount to the value of the test statistic. However, we may not wish to specify an *a priori* upper limit on the number of the lags and may want to give different weights to different lags. We show in Theorem 3.3 that this is possible for the GARCH(1, 1) model.

The paper is organized as follows. In Section 2, we collect the necessary definitions and state the assumptions. In Section 3, we recall the relevant results of our previous work and state the main result of the paper in Theorem 3.3, which is followed by a number of examples. Section 4 contains the proof of Theorem 3.3. At the beginning of that section, we outline the main new ideas developed in order to prove Theorem 3.3.

2. DEFINITIONS AND ASSUMPTIONS

A GARCH(1, 1) sequence $\{y_k, -\infty < k < \infty\}$ satisfies the equations

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

and

$$\sigma_k^2 = \omega + \alpha y_{k-1}^2 + \beta \sigma_{k-1}^2. \tag{2.2}$$

We assume that

$$\omega > 0, \quad \alpha \geqslant 0, \quad \text{and} \quad \beta \geqslant 0, \tag{2.3}$$

where $\boldsymbol{\theta} = (\omega, \alpha, \beta)$ is the parameter of the GARCH(1, 1) process. We also assume that

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

Nelson [22] proved that (2.1) and (2.2) have a unique stationary solution if and only if

$$E \log(\beta + \alpha \varepsilon_0^2) < 0. \tag{2.5}$$

Throughout this paper, we assume that (2.1)–(2.5) hold, since these are a minimal set of conditions for the existence of a GARCH(1, 1) sequence.

Since the parameter $\boldsymbol{\theta}$ is unknown, it must be estimated from the observations $y_1, y_2, \ldots, y_n$. Lee and Hansen [14] and Lumsdaine [18] studied the asymptotic properties of the quasi-maximum likelihood estimator of $\boldsymbol{\theta}$. Berkes *et al.* [5] weakened the conditions assumed by Lee and Hansen [14] and Lumsdaine [18] and proved the almost sure consistency and asymptotic normality of the quasi-maximum likelihood estimator in more general

GARCH(p, q) models. The efficiency of the quasi-maximum likelihood estimator is discussed in Berkes and Horváth [2]. Their conditions are very weak, so we follow their approach.

The logarithm of the quasi-likelihood function of the sample $y_1, y_2, \dots, y_n$ is given by

$$\tilde{L}_n(\mathbf{u}) = \sum_{1<k\leqslant n} -\frac{1}{2}\left\{\log \widetilde{w}_k(\mathbf{u}) + \frac{y_k^2}{\widetilde{w}_k(\mathbf{u})}\right\},$$

where

$$\widetilde{w}_k(\mathbf{u}) = \frac{x}{1-t} + s\sum_{1\leqslant i\leqslant k-1} t^{i-1}y_{k-i}^2, \quad \mathbf{u} = (x, s, t).$$

The quasi-maximum likelihood estimator is defined as

$$\tilde{\boldsymbol{\theta}}_n = \arg\max_{\mathbf{u}\in U} \tilde{L}_n(\mathbf{u}),$$

where U is given by

$$U = \left\{\mathbf{u} = (x, s, t)\colon\ t \leqslant \rho_0 \text{ and } u_1 \leqslant \min(x, s, t) \leqslant \max(x, s, t) \leqslant u_2\right\}$$

with some $0 < u_1 < u_2$ and $0 < \rho_0 < 1$.

If the innovations $\varepsilon_1, \dots, \varepsilon_n$ were observable, the estimator for the autocovariances would be

$$r_n(k) = \frac{1}{n}\sum_{k<i\leqslant n} (\varepsilon_i^2 - 1)(\varepsilon_{i-k}^2 - 1),$$

assuming that $E\varepsilon_i^2 = 1$.

Since $\varepsilon_1^2, \dots, \varepsilon_n^2$ are not observable, we replace them with the squared residuals

$$\hat{\varepsilon}_i^2 = y_i^2/\widetilde{w}_i\big(\tilde{\boldsymbol{\theta}}_n\big), \quad 1 \leqslant i \leqslant n,$$

and we use

$$\hat{r}_n(k) = \frac{1}{n}\sum_{k<i\leqslant n} (\hat{\varepsilon}_i^2 - 1)(\hat{\varepsilon}_{i-k}^2 - 1).$$

Next we state a set of conditions which is shown in Berkes *et al.* [5] to be sufficient for the asymptotic normality of $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$:

$$\varepsilon_0^2 \text{ is a nondegenerate random variable,} \tag{2.6}$$

$$\boldsymbol{\theta} \text{ is in the interior of } U, \tag{2.7}$$

$$\lim_{t\to 0} t^{-\mu} P\{\varepsilon_0^2 \leqslant t\} = 0 \quad \text{with some } \mu > 0, \tag{2.8}$$

$$E\varepsilon_0^2 = 1, \tag{2.9}$$

and

$$E|\varepsilon_0^2|^{2+\delta} < \infty \quad \text{with some } \delta > 0. \tag{2.10}$$

As mentioned in Introduction, Li and Mak [15] and Ling and Li [16] developed portmanteau tests based on the autocorrelations of the squared residuals. These authors assumed the conditional normality and finite fourth moments of the observations $y_1, y_2, \ldots, y_n$. As argued by Lee and Hansen [14] (cf. also Guillaume *et al.* [11] for an empirical study), these assumptions need not necessarily be satisfied by models and time series encountered in applications. We allow the observations to have essentially arbitrarily heavy tails. The asymptotic normality of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$, $1 \leqslant i_1 < \ldots < i_K$, under conditions (2.6)–(2.10) is a special case of Theorem 2.2 of Berkes *et al.* [3] (cf. Theorem 3.1 below). To state their result, which is used to derive a χ^2 test, we need further notation.

Let

$$d_0^2 = E(\varepsilon_0^2 - 1)^2,$$

$$\mathbf{A}_0 = \operatorname{cov}(\ell_0'(\boldsymbol{\theta})), \tag{2.11}$$

and

$$\mathbf{B}_0 = E(\ell_0''(\boldsymbol{\theta})), \tag{2.12}$$

where

$$\ell_0(\mathbf{u}) = -\frac{1}{2}\left\{\log w_0(\mathbf{u}) + \frac{y_0^2}{w_0(\mathbf{u})}\right\}$$

and

$$w_k(\mathbf{u}) = \frac{x}{1-t} + s \sum_{1 \leqslant i < \infty} t^{i-1} y_{k-i}^2, \quad -\infty < k < \infty.$$

We note that the matrices in (2.11) and (2.12) appeared in the asymptotic covariance matrix of $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ (cf. Berkes *et al.* [5]) and that they reflect the estimation of $\boldsymbol{\theta}$ in the definition of the squared residuals. Let

$$\mathbf{c}_k = E\left[(\varepsilon_{-k}^2 - 1)\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right]. \tag{2.13}$$

Berkes *et al.* [3] showed that $\mathbf{c}_k$ is finite. We use $\mathbf{u}^T$ to denote the transpose of vectors and matrices.

3. MAIN RESULTS

Let $1 \leqslant K < \infty$ and $1 \leqslant i_1 < i_2 < \cdots < i_K$. The asymptotic covariance matrix of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$ is

$$\mathbf{D}_{i_1,i_2,\ldots,i_K} = d_0^4\big(\mathbf{I}_K + \mathbf{C}_{i_1,i_2,\ldots,i_K}\mathbf{S}\mathbf{C}_{i_1,i_2,\ldots,i_K}^T\big),$$

where $\mathbf{I}_K$ denotes the $K \times K$ identity matrix, $\mathbf{C}_{i_1,i_2,\ldots,i_K}$ is the $K \times 3$ matrix defined as

$$\mathbf{C}_{i_1,i_2,\ldots,i_K} = \begin{pmatrix} \mathbf{c}_{i_1} \\ \mathbf{c}_{i_2} \\ \vdots \\ \mathbf{c}_{i_K} \end{pmatrix},$$

and

$$\mathbf{S} = \frac{1}{d_0^2}\mathbf{B}_0^{-1} + \frac{1}{d_0^4}\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}.$$

Theorems 3.1 and 3.2 were obtained by Berkes *et al.* [3].

THEOREM 3.1. *If* (2.6)–(2.10) *hold, then*

$$n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K)) \xrightarrow{\mathcal{D}} \mathbf{N}(\mathbf{0}, \mathbf{D}_{i_1,i_2,\ldots,i_K}),$$

where $\mathbf{N}(\mathbf{0}, \mathbf{D}_{i_1,i_2,\ldots,i_K})$ *stands for a* K*-variate normal random variable with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{D}_{i_1,i_2,\ldots,i_K}$.

THEOREM 3.2. *If* (2.6)–(2.10) *hold, then*

$$\mathbf{D}_{i_1,i_2,\ldots,i_K} \text{ is a positive definite matrix,}$$

$$\mathbf{D}^{-1}_{i_1,i_2,\ldots,i_K} = d_0^{-4}\{\mathbf{I}_K - \mathbf{C}_{i_1,i_2,\ldots,i_K}(\mathbf{I}_3 + \mathbf{S}^{1/2}\mathbf{C}^T_{i_1,i_2,\ldots,i_K}\mathbf{C}_{i_1,i_2,\ldots,i_K}\mathbf{S}^{1/2})^{-1}\mathbf{S}^{1/2}\mathbf{C}^T_{i_1,i_2,\ldots,i_K}\} \tag{3.14}$$

and

$$n(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))\mathbf{D}^{-1}_{i_1,i_2,\ldots,i_K}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))^T \xrightarrow{\mathcal{D}} \chi^2(K), \tag{3.15}$$

where $\chi^2(K)$ *denotes a chi-square random variable with* K *degrees of freedom.*

Theorem 3.2 provides a very simple way to compute $\mathbf{D}^{-1}_{i_1,i_2,\ldots,i_K}$. Note that, in (3.14), we need to compute the inverse of a 3×3 matrix.

The focus of this paper is on weighted statistics in which the number of lags, $m(n)$, at which the sample autocorrelations of the squared residuals are computed, is allowed to increase to infinity with the sample size n. We assume that

$$1 \leqslant m(n) \leqslant n \quad \text{and} \quad m(n) \to \infty. \tag{3.16}$$

The weights $\lambda_1, \lambda_2, \ldots$ are nonnegative and satisfy the condition

$$\sum_{1 \leqslant k < \infty} \lambda_k < \infty. \tag{3.17}$$

We use the notation Γ_m for $\mathbf{D}_{1,2,\ldots,m}$, and $\{\gamma_m(i, j)\colon 1 \leqslant i, j \leqslant m\}$ will denote the elements of Γ_m^{-1}. In Lemmas 4.7 and 4.8, we discuss the existence and some properties of Γ_m^{-1}. We consider the asymptotic behavior of

$$Q_n = n \sum_{1 \leqslant i, j \leqslant m} \lambda_i^{1/2}\hat{r}_n(i)\gamma_m(i, j)\lambda_j^{1/2}\hat{r}_n(j). \tag{3.18}$$

THEOREM 3.3. *If* (2.6)–(2.10), (3.16), *and* (3.17) *hold, then*

$$Q_n \xrightarrow{\mathcal{D}} Q, \tag{3.19}$$

where $Q = \sum_{1 \leqslant i < \infty} \lambda_i N_i^2$ *and* $N_1, N_2, \ldots$ *are independent, identically distributed standard normal random variables.*

Horváth and Kokoszka [12] found some special cases where the formulas for the distribution function of Q are known.

Example 2.1. Let $\lambda_j = 1/(\pi^2 j^2)$. Then Q equals in distribution to the square-integral of a Brownian bridge on [0, 1]. So a formula for the distribution function of Q can be found in Smirnov [27] and Anderson and Darling [1]. A table is provided by Shorack and Wellner [26, p. 147].

Example 2.2. If $\lambda_{2j-1} = \lambda_{2j} = 1/(4\pi^2 j^2)$, then Q is just the weak limit of the Watson [29] statistic.

Example 2.3. If $\lambda_j = 1/(\pi^2(j - 1/2)^2)$, then Q equals in distribution to the square-integral of a Brownian motion on [0, 1]. Formulas for the distribution function of Q are given in Cameron and Martin [8] and Rothman and Woodroofe [24]. Csörgő and Horváth [9] provide a table for the distribution function of Q.

4. PROOF OF THEOREM 3.3

The proof of Theorem 3.3 is constructed from a number of technical lemmas. Lemma 4.1 is well known and is stated here for easy reference. Lemmas 4.2 and 4.3 are borrowed from the earlier work of the authors and are also stated here, as they are extensively used in the proofs of the other lemmas. Lemma 4.4, which is the most technical result of the paper, establishes an upper bound on the rate of decay of the vectors $\mathbf{c}_k$ defined in (2.13) as $k \to \infty$. A result of this type is needed in the present context, since, unlike Berkes *et al.* [3], we are now dealing with infinitely many lags for the sample autocorrelations. After stating Lemma 4.5, which is borrowed from Berkes *et al.* [3], and providing a technical bound in Lemma 4.6, we establish in Lemmas 4.7 and 4.8, respectively, the nonsingularity of the matrix Γ_m appearing in (3.18) and the existance and nonsingularity of appropriate limit matrices which are needed to establish the convergence of the statistic Q_n. Lemma 4.8, in particular, relies on new arguments, which were not needed in Berkes *et al.* [3], where only a fixed number of lags was considered. The proof of Theorem 3.3 then follows from Lemmas 4.4–4.8 and fairly standard arguments.

Let

$$\xi_i = \beta + \alpha\varepsilon_i^2, \quad -\infty < i < \infty.$$

LEMMA 4.1 (Nelson [22]). *If* $E|\log \xi_0| < \infty$ *and* $E \log \xi_0 < 0$, *then*

$$\sigma_k^2 = \omega\left\{1 + \sum_{1\leqslant \ell<\infty} \prod_{1\leqslant i\leqslant \ell} \xi_{k-i}\right\}.$$

LEMMA 4.2 (Berkes *et al.* [5]). *If* $E|\varepsilon_0^2|^\kappa < \infty$ *with some* $\kappa > 0$, *then there is* $\kappa^* > 0$ *such that*

$$E|y_0^2|^{\kappa^*} < \infty \quad \text{and} \quad E|\sigma_0^2|^{\kappa^*} < \infty.$$

Let $\|\cdot\|$ denote the maximum norm of vectors and matrices.

LEMMA 4.3. *If* (2.6)–(2.10) *hold*, *then*, *for any* $\nu > 0$,

$$E\left(\frac{\sum_{1\leqslant i<\infty} i\beta^{i-2}y_{-i}^2}{1+\sum_{1\leqslant i<\infty}\beta^{i-1}y_{-i}^2}\right)^\nu < \infty \tag{4.1}$$

and

$$E\left(\sup_{\mathbf{u}\in U}\frac{\|w_0'(\mathbf{u})\|}{w_0(\mathbf{u})}\right)^\nu < \infty. \tag{4.2}$$

Moreover, *for any* $-\infty < \nu^* < \infty$, *there is* $\gamma > 0$ *such that*

$$E\left(\sup_{\mathbf{u}\in U(\gamma)}\frac{w_0(\boldsymbol{\theta})}{w_0(\mathbf{u})}\right)^{\nu^*} < \infty, \tag{4.3}$$

where $U(\gamma) = \{\mathbf{u}\in U\colon\ \|\boldsymbol{\theta}-\mathbf{u}\| \leqslant \gamma\}$.

Proof. Relations (4.1) and (4.2) are proved by Berkes *et al.* [5], while (4.3) by Berkes and Horváth [2].

LEMMA 4.4. *If* (2.6)–(2.10) *hold*, *then*, *for any* $\nu > 0$, $\|\mathbf{c}_k\| = \mathrm{O}(k^{-\nu})$ *as* $k \to \infty$.

Proof. Using the notation $c_0(\mathbf{u}) = x/(1-t)$ and $c_i(\mathbf{u}) = st^{i-1}$, $1 \leqslant i < \infty$, $\mathbf{u} = (x, s, t)$, we can write

$$\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} = \frac{c_0'(\boldsymbol{\theta}) + \sum_{1\leqslant i<\infty} c_i'(\boldsymbol{\theta}) y_{-i}^2}{c_0(\boldsymbol{\theta}) + \sum_{1\leqslant i<\infty} c_i(\boldsymbol{\theta}) y_{-i}^2}.$$

For any $k > 0$, we define

$$\bar{y}_{-i,k}^2 = \begin{cases} y_{-i}^2 & \text{if } i > k, \\ \sigma_{-k}^2 \varepsilon_{-k+1}^2 & \text{if } i = k, \\ \bar{\sigma}_{-i,k}^2 \varepsilon_{-i}^2 & \text{if } 1 \leqslant i < k, \end{cases}$$

where

$$\begin{aligned} \bar{\sigma}_{-i,k}^2 = \omega(1 &+ \xi_{-i-1} + \xi_{-i-1}\xi_{-i-2} + \cdots + \xi_{-i-1}\xi_{-i-2}\cdots\xi_{-k+2} \\ &+ \xi_{-i-1}\xi_{-i-2}\cdots\xi_{-k+1}X_{-k}) \end{aligned}$$

and

$$X_{-k} = 1 + \xi_{-k-1} + \xi_{-k-1}\xi_{-k-2} + \cdots.$$

This means that $\bar{\sigma}_{-i,k}^2$ is defined as σ_{-i}^2 in Lemma 4.1 but with ξ_{-k} replaced by 1. Let

$$\mathbf{G}_k = \frac{c_0'(\boldsymbol{\theta}) + \sum_{1\leqslant i<\infty} c_i'(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}{c_0(\boldsymbol{\theta}) + \sum_{1\leqslant i<\infty} c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}.$$

Since $c_0(\boldsymbol{\theta}) > 0$ and

$$\frac{\|c_i'(\boldsymbol{\theta})\|}{c_i(\boldsymbol{\theta})} = \mathrm{O}(i) \quad \text{as } i \to \infty \tag{4.4}$$

(cf. Lemma 3.2 by Berkes *et al.* [5]), we get that

$$\|\mathbf{G}_k\| \leqslant C_1 \left(\frac{\sum_{1\leqslant i<\infty} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}{1 + \sum_{1\leqslant i<\infty} c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2} + 1 \right) \tag{4.5}$$

with some constant C_1. Moreover,

$$\frac{\sum_{1\leqslant i<\infty} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}{1 + \sum_{1\leqslant i<\infty} c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2} \leqslant k + \frac{\sum_{1\leqslant i<\infty,\, i\neq k} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}{1 + \sum_{1\leqslant i<\infty,\, i\neq k} c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}. \tag{4.6}$$

For any $M \geqslant 1$, we have

$$\frac{\sum_{1\leqslant i<\infty,\, i\neq k} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2}{1 + \sum_{1\leqslant i<\infty,\, i\neq k} c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2} \leqslant M + \sum_{M\leqslant i<\infty,\, i\neq k} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2.$$

Clearly, $\xi_{-k} \geqslant \beta$ and $\beta < 1$ by (2.7), hence, $\bar{y}_{-i,k}^2 \leqslant y_{-i}^2/\beta$ if $1 \leqslant i < k$, resulting in

$$\sum_{M\leqslant i<\infty,\, i\neq k} i c_i(\boldsymbol{\theta}) \bar{y}_{-i,k}^2 \leqslant \frac{1}{\beta} \sum_{M\leqslant i<\infty} i c_i(\boldsymbol{\theta}) y_{-i}^2.$$

By Lemma 3.1 of Berkes *et al.* [3] there are C_2 and $0 < \vartheta < 1$ such that

$$ic_i(\boldsymbol{\theta}) \leqslant C_2\vartheta^i \quad \text{for all } 1 \leqslant i < \infty.$$

Let $\bar{\vartheta} > 1$ be such that $\vartheta\bar{\vartheta} < 1$. By Lemma 4.2 and the Markov inequality we have

$$\begin{aligned} P\left\{\sum_{M\leqslant i<\infty} \vartheta^i y_{-i}^2 \geqslant t\right\} &\leqslant \sum_{M\leqslant i<\infty} P\left\{y_{-i}^2 \geqslant t\vartheta^{-i}\bar{\vartheta}^{-i}\left(\frac{\bar{\vartheta}}{\bar{\vartheta}-1}\right)^{-1}\right\} \\ &\leqslant \sum_{M\leqslant i<\infty} P\left\{(y_{-i}^2)^{\kappa^*} \geqslant t^{\kappa^*}\vartheta^{-i\kappa^*}\bar{\vartheta}^{-i\kappa^*}\left(\frac{\bar{\vartheta}}{\bar{\vartheta}-1}\right)^{-\kappa^*}\right\} \\ &\leqslant \frac{E(y_0^2)^{\kappa^*}}{1-(\vartheta\bar{\vartheta})^{\kappa^*}}\left(\frac{\bar{\vartheta}}{\bar{\vartheta}-1}\right)^{\kappa^*} t^{-\kappa^*}(\vartheta\bar{\vartheta})^{M\kappa^*}. \end{aligned}$$

Choosing $M = t/2$, it follows that, for any $\nu > 0$, there is a constant $C_3 = C_3(\nu)$ such that

$$P\left\{\frac{\sum_{1\leqslant i<\infty,\, i\neq k} ic_i(\boldsymbol{\theta})\bar{y}_{-i,k}^2}{1+\sum_{1\leqslant i<\infty,\, i\neq k} c_i(\boldsymbol{\theta})\bar{y}_{-i,k}^2} > t\right\} \leqslant C_3 t^{-\nu} \tag{4.7}$$

for all $t \geqslant 2$. Hence, using (4.5)–(4.7), we get that

$$E\|\mathbf{G}_k\|^\eta = \mathrm{O}(k^\eta) \quad \text{for any } \eta > 0. \tag{4.8}$$

Thus, by (2.9) and the independence of ε_{-k}^2 and $\mathbf{G}_k$ we have

$$E(\varepsilon_{-k}^2 - 1)\mathbf{G}_k = \mathbf{O}. \tag{4.9}$$

Next we write

$$\left\|\mathbf{G}_k - \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right\| \leqslant \|G_{k,1}\| + |G_{k,2}|\,\|\mathbf{G}_k\|,$$

where

$$\mathbf{G}_{k,1} = \frac{\sum_{1\leqslant i\leqslant k} c_i'(\boldsymbol{\theta})(y_{-i}^2 - \bar{y}_{-i,k}^2)}{w_0(\boldsymbol{\theta})}$$

and

$$G_{k,2} = \frac{\sum_{1\leqslant i\leqslant k} c_i(\boldsymbol{\theta})(y_{-i}^2 - \bar{y}_{-i,k}^2)}{w_0(\boldsymbol{\theta})}.$$

Let $0 < \rho < 1$ be any number satisfying $E\log\xi_0 < \log\rho < 0$, and let F_k be the event that

$$\max_{0\leqslant i\leqslant k/2} \prod_{i+1\leqslant j\leqslant k-1} \xi_{-j} \leqslant \rho^{k/2}.$$

We show that

$$P\{F_k\} \geqslant 1 - C_4 k^{-\nu} \tag{4.10}$$

for all $\nu > 0$ with some $C_4 = C_4(\nu)$. Since $E|\log \xi_0|^{2\zeta} < \infty$ for all $\zeta > 0$, by the Rosenthal inequality (cf. Petrov [23, p. 59]) we have that

$$E\left|\sum_{i+1\leqslant j\leqslant k-1}(\log \xi_{-j} - E\log \xi_0)\right|^{2\zeta} \leqslant C_5(k-i)^{\zeta} \quad \text{if } \zeta \geqslant 1.$$

Thus, by the Markov inequality we get

$$\begin{aligned}
&P\left\{\max_{0\leqslant i\leqslant k/2}\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j} \geqslant \rho^{k/2}\right\}\\
&\quad\leqslant P\left\{\max_{0\leqslant i\leqslant k/2}\sum_{i+1\leqslant j\leqslant k-1}(\log \xi_{-j} - E\log \xi_0) \geqslant \frac{k}{4}(\log\rho - E\log\xi_0)\right\}\\
&\quad\leqslant \sum_{0\leqslant i\leqslant k/2} P\left\{\sum_{i+1\leqslant j\leqslant k-1}(\log \xi_{-j} - E\log \xi_0) \geqslant \frac{k}{4}(\log\rho - E\log\xi_0)\right\}\\
&\quad\leqslant C_6\sum_{0\leqslant i\leqslant k/2} k^{-2\zeta}(k-i)^{\zeta} \leqslant C_6 k^{-(\zeta-1)},
\end{aligned}$$

as claimed. As noted above, $\bar{y}^2_{-i,k} \leqslant y^2_{-i}/\beta$ if $1 \leqslant i < k$ and, therefore,

$$\begin{aligned}
\left\|\sum_{1\leqslant i\leqslant k} c_i'(\boldsymbol{\theta})(y^2_{-i} - \bar{y}^2_{-i,k})\right\| \leqslant &\sum_{1\leqslant i<\infty}\|c_i'(\boldsymbol{\theta})\|y^2_{-i} + \|c_k'(\boldsymbol{\theta})\|\sigma^2_{-k}(\varepsilon^2_{-k} + \varepsilon^2_{-k+1})\\
&+ \frac{1}{\beta}\sum_{1\leqslant i\leqslant k-1}\|c_i'(\boldsymbol{\theta})\|y^2_{-i}.
\end{aligned} \tag{4.11}$$

Here, in the second term, we have $\sigma^2_{-k}\varepsilon^2_{-k} = y^2_{-k}$, and this term can be incorporated into the third one when we sum up to k. Clearly, $w_0(\boldsymbol{\theta}) \geqslant c_{k-1}(\boldsymbol{\theta})y^2_{-k+1}$, $y^2_{-k+1} = \sigma^2_{-k+1}\varepsilon^2_{-k+1} = (\omega + \alpha y^2_{-k} + \beta\sigma^2_{-k})\varepsilon^2_{-k+1} \geqslant \beta\sigma^2_{-k}\varepsilon^2_{-k+1}$, and, therefore,

$$\frac{\|c_k'(\boldsymbol{\theta})\|\sigma^2_{-k}\varepsilon^2_{-k+1}}{w_0(\boldsymbol{\theta})} \leqslant \frac{1}{\beta}\frac{\|c_k'(\boldsymbol{\theta})\|}{c_{k-1}(\boldsymbol{\theta})}.$$

Since $c_k(\boldsymbol{\theta}) = \alpha\beta^{k-1}$, we have

$$\frac{c_k(\boldsymbol{\theta})}{c_{k-1}(\boldsymbol{\theta})} = \mathrm{O}(1),$$

and, thus, by (4.4)

$$\frac{\|c_k'(\boldsymbol{\theta})\|\sigma^2_{-k}\varepsilon^2_{-k+1}}{w_0(\boldsymbol{\theta})} = \mathrm{O}(k). \tag{4.12}$$

Let $I\{\cdot\}$ denote the indicator function, and let $\overline{F}_k$ be the complement of F_k. By (4.4), the first relation in Lemma 4.3, and (4.10)–(4.12) we have that, for all $\nu > 1$,

$$E\big(\|\mathbf{G}_{k,1}\|^3 I\{\overline{F}_k\}\big) \leqslant \big(E(\|\mathbf{G}_{k,1}\|^6)\big)^{1/2}\big(P\{\overline{F}_k\}\big)^{1/2} = \mathrm{O}(k^{-\nu}). \tag{4.13}$$

Next we note that

$$\sigma^2_{-i} - \bar{\sigma}^2_{-i,k} = \omega\xi_{-i-1}\cdots\xi_{-k+1}(1-\xi_{-k})X_{-k}$$

and, thus, for any $1 < \gamma < \rho^{-1/2}$, using (4.4) and Lemma 3.1 of Berkes *et al.* [3], we have

$$\begin{aligned}
&\frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})(y^2_{-i} - \bar{y}^2_{-i,k})\|}{w_0(\boldsymbol{\theta})} I\{F_k\} \\
&\quad = \frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega(1-\xi_{-k})X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} I\{F_k\} \\
&\quad = \frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega(1-\xi_{-k})X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} \\
&\qquad \times I\{F_k\} I\{|1-\xi_{-k}|X_{-k} \leqslant \gamma^k\} \\
&\qquad + \frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega(1-\xi_{-k})X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} \\
&\qquad \times I\{F_k\} I\{|1-\xi_{-k}|X_{-k} > \gamma^k\} \\
&\quad \leqslant C_7\rho^{k/2}\gamma^k \sum_{1\leqslant i\leqslant k/2} i\beta^{i-2}\varepsilon^2_{-i} + C_8 k I\{|1-\xi_k|X_{-k} > \gamma^k\}
\end{aligned} \tag{4.14}$$

with some constants C_7 and C_8, where the second term is obtained by noting that

$$\begin{aligned}
&\frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega(1-\xi_{-k})X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} \\
&\quad \leqslant \frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} \\
&\qquad + \frac{|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})\big(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\big)\omega\xi_{-k}X_{-k}\varepsilon^2_{-i}\|}{w_0(\boldsymbol{\theta})} \\
&\quad \leqslant C_8 k,
\end{aligned}$$

which, in turn, easily follows from

$$\sigma^2_{-i} \geqslant \omega\left(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\right)\xi_{-k}X_{-k},$$

$$\sigma^2_{-i} \geqslant \omega\beta\left(\prod_{i+1\leqslant j\leqslant k-1}\xi_{-j}\right)X_{-k},$$

and the fact that

$$\sum_{1\leqslant i\leqslant k/2}\|c_i'(\boldsymbol{\theta})\|y^2_{-i} \leqslant C_9 \sum_{1\leqslant i\leqslant k/2} i c_i(\boldsymbol{\theta})y^2_{-i} \leqslant C_9 k w_0(\boldsymbol{\theta}).$$

By the Minkowski inequality and (2.10) we have that

$$E\left(\sum_{1\leqslant i\leqslant k/2} i\beta^i \varepsilon_{-i}^2\right)^{2+\delta} = \mathrm{O}(1).$$

Since $\sigma_{-k}^2 = \omega X_{-k}$, (1.10) and Lemma 4.2 imply the existence of $\zeta > 0$ such that $E|X_{-k}|^\zeta < \infty$ and $E|X_{-k}\xi_{-k}|^\zeta < \infty$ and, consequently,

$$EI\{|1-\xi_{-k}|X_{-k} > \gamma^k\} = \mathrm{O}\big(\mathrm{e}^{-\hat{\delta}k}\big) \quad \text{with some } \hat{\delta} > 0.$$

Thus, we showed that

$$E\left(\frac{\|\sum_{1\leqslant i\leqslant k/2} c_i'(\boldsymbol{\theta})(y_{-i}^2 - \bar{y}_{-i,k}^2)\|}{w_0(\boldsymbol{\theta})} I\{F_k\}\right)^{2+\delta} = \mathrm{O}(k^{-\nu}) \tag{4.15}$$

for all $\nu > 0$.

Let $\eta > 0$. From (4.4) it follows that $\|c_i'(\boldsymbol{\theta})\| \leqslant C_{10}(1+\eta)^i c_i(\boldsymbol{\theta})$ for all $i \geqslant 1$ and, thus,

$$\begin{aligned}\left\|\sum_{k/2<i\leqslant k} c_i'(\boldsymbol{\theta}) y_{-i}^2\right\| &\leqslant C_{10} \sum_{k/2\leqslant i\leqslant k} c_i(\boldsymbol{\theta})(1+\eta)^i y_{-i}^2 \\ &\leqslant C_{10}(1+\eta)^{-k/2} \sum_{k/2\leqslant i\leqslant k} (1+\eta)^{2i} c_i(\boldsymbol{\theta}) y_{-i}^2 \\ &\leqslant C_{10}(1+\eta)^{-k/2} \sum_{1\leqslant i<\infty} (1+\eta)^{2i} c_i(\boldsymbol{\theta}) y_{-i}^2.\end{aligned} \tag{4.16}$$

Since $\bar{y}_{-i,k}^2 \leqslant y_{-i}^2/\beta$, from (4.16) we get that

$$\left\|\sum_{k/2<i\leqslant k} c_i'(\boldsymbol{\theta}) \bar{y}_{-i,k}^2\right\| \leqslant \frac{C_{10}}{\beta}(1+\eta)^{-k/2} \sum_{1\leqslant i<\infty} (1+\eta)^{2i} c_i(\boldsymbol{\theta}) y_{-i}^2. \tag{4.17}$$

Minor modifications of the proof of Lemma 5.2 of Berkes *et al.* [5] yield

$$E\left(\frac{\sum_{1\leqslant i<\infty}(1+\eta)^{2i} c_i(\boldsymbol{\theta}) y_{-i}^2}{w_0(\boldsymbol{\theta})}\right)^3 < \infty \tag{4.18}$$

if $\eta > 0$ is small enough. Putting together (4.13)–(4.18), we conclude that

$$E\|\mathbf{G}_{k,1}\|^{2+\delta} = \mathrm{O}(k^{-\nu}) \quad \text{for all } \nu > 0. \tag{4.19}$$

Similarly to (4.19), one can show that

$$E|G_{k,2}|^{2+\delta} = \mathrm{O}(k^{-\nu}) \quad \text{for all } \nu > 0. \tag{4.20}$$

Lemma 4.4 now follows from (4.8), (4.9), and (4.13)–(4.20).

Let

$$a_k(\mathbf{u}) = \sum_{k<i\leqslant n}\left(\frac{y_i^2}{w_i(\mathbf{u})} - 1\right)\left(\frac{y_{i-k}^2}{w_{i-k}(\mathbf{u})} - 1\right)$$

and

$$\tilde{a}_k(\mathbf{u}) = \sum_{k<i\leqslant n} \left(\frac{y_i^2}{\widetilde{w}_i(\mathbf{u})} - 1\right)\left(\frac{y_{i-k}^2}{\widetilde{w}_{i-k}(\mathbf{u})} - 1\right).$$

For any $L > 0$, we define

$$U_n(L) = \left\{\mathbf{u} \in U\colon \ \|\boldsymbol{\theta} - \mathbf{u}\| \leqslant Ln^{-1/2}\right\}.$$

The following result is a special case of Lemma 5.1 of Berkes *et al.* [3].

LEMMA 4.5. *If* (2.6)–(2.10) *hold, then, for any* $0 < L < \infty$,

$$\sup_{0\leqslant k<n} \sup_{\mathbf{u}\in U_n(L)} |\tilde{a}_k(\mathbf{u}) - a_k(\mathbf{u})| = \mathrm{o}(n^{1/2}) \quad a.s.$$

LEMMA 4.6. *If* (2.6)–(2.10) *hold, then, for any* $0 < L < \infty$,

$$E\Big(\sup_{\mathbf{u}\in U_n(L)} |a_k(\mathbf{u})|\Big)^2 \leqslant C_{11} n$$

with some constant C_{11} *for all* $1 \leqslant k < \infty$.

Proof. Elementary algebra gives

$$\begin{aligned} a_k(\mathbf{u}) &= \sum_{k<i\leqslant n} (\varepsilon_i^2 - 1)(\varepsilon_{i-k}^2 - 1) + \sum_{k<i\leqslant n} \varepsilon_i^2 \left(\frac{\sigma_i^2}{w_i(\mathbf{u})} - 1\right)(\varepsilon_{i-k}^2 - 1) \\ &\quad + \sum_{k<i\leqslant n} (\varepsilon_i^2 - 1)\varepsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{w_{i-k}(\mathbf{u})} - 1\right) + \sum_{k<i\leqslant n} \varepsilon_i^2 \left(\frac{\sigma_i^2}{w_i(\mathbf{u})} - 1\right)\varepsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{w_{i-k}(\mathbf{u})} - 1\right) \\ &=: A_{k,1} + A_{k,2}(\mathbf{u}) + A_{k,3}(\mathbf{u}) + A_{k,4}(\mathbf{u}). \end{aligned}$$

Since the terms of the sum defining $A_{k,1}$ are orthogonal, we have

$$EA_{k,1}^2 = (n-k)d_0^4.$$

Moreover,

$$\begin{aligned} E \sup_{\mathbf{u}\in U_n(L)} A_{k,2}^2(\mathbf{u}) &\leqslant E \sum_{k<i,j\leqslant n} \varepsilon_i^2 \varepsilon_j^2 |\varepsilon_{i-k}^2 - 1||\varepsilon_{j-k}^2 - 1| \\ &\qquad \times \sup_{\mathbf{u}\in U_n(L)} \left|\frac{\sigma_i^2}{w_i(\mathbf{u})} - 1\right| \sup_{\mathbf{u}\in U_n(L)} \left|\frac{\sigma_j^2}{w_j(\mathbf{u})} - 1\right|. \end{aligned}$$

Since $\sigma_0^2 = w_0(\boldsymbol{\theta})$, the mean-value theorem yields

$$\sup_{\mathbf{u}\in U_n(L)} \left|\frac{\sigma_0^2}{w_0(\mathbf{u})} - 1\right| \leqslant \frac{L}{n^{1/2}} \sup_{\mathbf{u}\in U} \frac{\|w_0'(\mathbf{u})\|}{w_0(\mathbf{u})} \sup_{\mathbf{u},\mathbf{v}\in U_n(L)} \frac{w_0(\mathbf{v})}{w_0(\mathbf{u})},$$

and, therefore, Lemma 4.3 implies

$$E\left(n^{1/2}\sup_{\mathbf{u}\in U_n(L)}\left|\frac{\sigma_0^2}{w_0(\mathbf{u})}-1\right|\right)^{\nu}\leqslant C_{12}\quad \text{for all } \nu>0$$

with some $C_{12}(\nu, L)$. Also using (2.4), (2.10) and Hölder's inequality, it is easily seen that, for any $k < i,\ j \leqslant n$, we have

$$E\big(\varepsilon_i^2\varepsilon_j^2\big|\varepsilon_{i-k}^2-1\|\varepsilon_{j-k}^2-1\big|\big)^{1+\eta}\leqslant C_{13}$$

for some constants $\eta > 0$ and $C_{13} > 0$. Hence, the above estimates and another application of Hölder's inequality imply that

$$E\sup_{\mathbf{u}\in U_n(L)}A_{k,2}^2(\mathbf{u})\leqslant C_{14}n.$$

Similar arguments give

$$E\sup_{\mathbf{u}\in U_n(L)}A_{k,3}^2(\mathbf{u})+E\sup_{\mathbf{u}\in U_n(L)}A_{k,4}^2(\mathbf{u})\leqslant C_{15}n,$$

completing the proof of Lemma 4.6.

LEMMA 4.7. *If* (2.6)–(2.10) *hold*, *then*

$$\Gamma_m=d_0^4\left(\mathbf{I}_m-\frac{1}{4}\mathbf{C}_{1,2,\ldots,m}\mathbf{A}_0^{-1}\mathbf{C}_{1,2,\ldots,m}^T\right)\tag{4.21}$$

and

$$\Gamma_m^{-1}=d_0^{-4}\left(\mathbf{I}_m+\frac{1}{4}\mathbf{C}_{1,2,\ldots,m}\left(\mathbf{A}_0-\frac{1}{4}\mathbf{C}_{1,2,\ldots,m}^T\mathbf{C}_{1,2,\ldots,m}\right)^{-1}\mathbf{C}_{1,2,\ldots,m}^T\right).\tag{4.22}$$

Proof. By Remark 4.5 of Berkes *et al.* [5] we have that

$$\mathbf{B}_0=-\frac{2}{d_0^2}\mathbf{A}_0$$

and, therefore,

$$\mathbf{S}=-\frac{1}{4}\mathbf{A}_0^{-1},$$

completing the proof of (4.21). Now Theorem A.5.1 of Muirhead [21] yields (4.22).

LEMMA 4.8. *If* (2.6)–(2.10) *hold*, *then*

$$\lim_{m\to\infty}\mathbf{C}_{1,2,\ldots,m}^T\mathbf{C}_{1,2,\ldots,m}=\widetilde{\mathbf{C}}\quad \textit{exists}\tag{4.23}$$

and

$$\mathbf{A}_0-\frac{1}{4}\widetilde{\mathbf{C}}\quad \textit{is nonsingular.}\tag{4.24}$$

Proof. Lemma 4.4 yields that $\widetilde{\mathbf{C}}$ exists. By Berkes *et al.* [5] we have that

$$\mathbf{A}_0 = \frac{1}{4} d_0^2 E\left[\left(\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right)^T - \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right].$$

It is easy to see that

$$E\left[\left(d_0 \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} - \frac{1}{d_0} \sum_{1 \leqslant k < \infty} (1 - \varepsilon_{-k}^2)\mathbf{c}_k\right)^T \left(d_0 \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} - \frac{1}{d_0} \sum_{1 \leqslant k < \infty} (1 - \varepsilon_{-k}^2)\mathbf{c}_k\right)\right] = 4\left(\mathbf{A}_0 - \frac{1}{4}\widetilde{\mathbf{C}}\right).$$

Let $w_0'(\boldsymbol{\theta}) = (w_{01}', w_{02}', w_{03}')$ and $\mathbf{c}_k = (c_{k1}, c_{k2}, c_{k3})$. The matrix in (4.24) is singular if and only if there is $(\tau_1, \tau_2, \tau_3) \neq \mathbf{0}$ such that

$$\tau_1 w_{01}' + \tau_2 w_{02}' + \tau_3 w_{03}' - \sum_{1 \leqslant k < \infty} (1 - \varepsilon_{-k}^2) c_k^* w_0(\boldsymbol{\theta}) = 0, \tag{4.25}$$

where $c_k^* = (\tau_1 c_{k1} + \tau_2 c_{k2} + \tau_3 c_{k3})/d_0^2$. We next note that

$$w_0(\boldsymbol{\theta}) = \frac{\omega}{1 - \beta} + \alpha \sum_{1 \leqslant k < \infty} \beta^{k-1} y_{-k}^2$$

and

$$w_{01}' = \frac{1}{1 - \beta}, \qquad w_{02}' = \sum_{1 \leqslant k < \infty} \beta^{k-1} y_{-k}^2,$$

$$w_{03}' = \frac{\omega}{(1 - \beta)^2} + \alpha \sum_{2 \leqslant k < \infty} (k - 1)\beta^{k-2} y_{-k}^2.$$

So, using the equations above, we get that

$$\tau_1 \frac{1}{1 - \beta} + \tau_2 \sum_{1 \leqslant k < \infty} \beta^{k-1} y_{-k}^2 + \tau_3 \alpha \sum_{2 \leqslant k < \infty} (k - 1)\beta^{k-2} y_{-k}^2 + \tau_3 \frac{\omega}{(1 - \beta)^2}$$
$$+ \left(\sum_{1 \leqslant \ell < \infty} c_\ell^* \varepsilon_{-\ell}^2 - \sum_{1 \leqslant \ell < \infty} c_\ell^*\right)\left(\frac{\omega}{1 - \beta} + \alpha \sum_{1 \leqslant k < \infty} \beta^{k-1} y_{-k}^2\right) = 0.$$

Separating those terms which depend only on ε_{-1}^2, we arrive at

$$\begin{aligned}
&\varepsilon_{-1}^4 \alpha c_1^* \sigma_{-1}^2 + \varepsilon_{-1}^2 \Bigg(c_1^* \Bigg(\frac{\omega}{1-\beta} + \alpha \sum_{2\leqslant k<\infty} \beta^{k-1} y_{-k}^2 \Bigg) + \tau_2 \sigma_{-1}^2 \\
&\quad - \Bigg(\sum_{1\leqslant \ell<\infty} c_\ell^* \Bigg) \alpha \sigma_{-1}^2 + \Bigg(\sum_{2\leqslant \ell<\infty} c_\ell^* \varepsilon_{-\ell}^2 \alpha \Bigg) \sigma_{-1}^2 \Bigg) \\
&\quad + \sum_{2\leqslant k<\infty} \Bigg(\tau_2 \beta + \tau_3 \alpha (k-1) - \alpha\beta \sum_{1\leqslant \ell<\infty} c_\ell^* \Bigg) \beta^{k-2} y_{-k}^2 \\
&\quad + \sum_{2\leqslant \ell<\infty} c_\ell^* \varepsilon_{-\ell}^2 \Bigg(\frac{\omega}{1-\beta} + \alpha \sum_{2\leqslant k<\infty} \beta^{k-1} y_{-k}^2 \Bigg) = \tau^*
\end{aligned}$$

with some constant τ^*. The last equality is a linear or quadratic equation for ε_{-1}^2 with coefficients measurable with respect to the σ-algebra $\mathcal{F}_{-2} = \sigma\{\varepsilon_i,\ i \leqslant -2\}$. Clearly, the coefficients of ε_{-1}^4 and ε_{-1}^2 must be both 0, since, otherwise, the solution of the equation, i.e., ε_{-1}^2, would also be measurable with respect to $\mathcal{F}_{-2}$, an impossibility. Noting that $\sigma_{-1}^2 \geqslant \omega > 0$, we obtain that $c_1^* = 0$ and

$$\tau_2 - \Bigg(\sum_{1\leqslant \ell<\infty} c_\ell^* \Bigg) \alpha + \Bigg(\sum_{2\leqslant \ell<\infty} c_\ell^* \varepsilon_{-\ell}^2 \Bigg) \alpha = 0. \tag{4.26}$$

Since $\alpha > 0$ and $\varepsilon_i, -\infty < i < \infty$, are independent, (4.26) can hold only if $c_\ell^* = 0$ for all $\ell \geqslant 2$. Hence, if (4.25) holds, then $c_\ell^* = 0$ for all $1 \leqslant \ell < \infty$ and, therefore, (4.25) becomes

$$\tau_1 \frac{w_1'}{w_{01}(\boldsymbol{\theta})} + \tau_2 \frac{w_2'}{w_{02}(\boldsymbol{\theta})} + \tau_3 \frac{w_3'}{w_{03}(\boldsymbol{\theta})} = 0,$$

contradicting that $\mathbf{A}_0$ is nonsingular.

Proof of Theorem 3.3. By the Cholesky decomposition (cf. Seber [25, p. 388]) and Theorem 3.1 we have

$$n \sum_{1\leqslant i,j\leqslant K} \lambda_i^{1/2} \hat{r}_n(i) \gamma_m(i,j) \lambda_j^{1/2} \hat{r}_n(j) \xrightarrow{\mathcal{D}} \sum_{1\leqslant i\leqslant K} \lambda_i N_i^2 \tag{4.27}$$

for all $1 \leqslant K < \infty$, where $N_1, N_2, \ldots$ are independent standard normal random variables.

Berkes *et al.* [5] proved that

$$\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = \mathrm{O}_P(n^{-1/2}),$$

and, therefore, it suffices to show that, for any $0 < L < \infty$ and $x > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\left\{ \sup_{\mathbf{u}\in U_n(L)} n \left| \sum_{K<i\leqslant m} \sum_{1\leqslant j\leqslant m} \lambda_i^{1/2} \tilde{a}_i(\mathbf{u}) \gamma_m(i,j) \lambda_j^{1/2} \tilde{a}_j(\mathbf{u}) \right| > x \right\} = 0 \tag{4.28}$$

and

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\left\{ \sup_{\mathbf{u}\in U_n(L)} n \left| \sum_{1\leqslant i\leqslant m} \sum_{K<j\leqslant m} \lambda_i^{1/2} \tilde{a}_i(\mathbf{u}) \gamma_m(i,j) \lambda_j^{1/2} \tilde{a}_j(\mathbf{u}) \right| > x \right\} = 0.$$

Due to symmetry, we need to establish only (4.28).

We recall that Γ_m^{-1} satisfies (3.14). Let

$$\mathbf{C}_{1,2,\dots,m}\left(\mathbf{A}_0-\frac{1}{4}\mathbf{C}^T_{1,2,\dots,m}\mathbf{C}_{1,2,\dots,m}\right)^{-1}\mathbf{C}^T_{1,2,\dots,m}=\{\gamma^*_{i,j},\, 1\leqslant i,\, j\leqslant m\}.$$

From Lemmas 4.7 and 4.8 it follows that there is a constant C_{16} such that $|\gamma^*_{i,j}|\leqslant C_{16}/(ij)^2$ for all $1\leqslant i,\, j\leqslant m$ and $1\leqslant m<\infty$. Hence, by Lemmas 4.5 and 4.4, for any $\Delta>0$, there is a random variable n_0 such that

$$\sup_{\mathbf{u}\in U_n(L)}\left|\sum_{K<i\leqslant m}\sum_{1\leqslant j\leqslant m}\lambda_i^{1/2}\tilde{a}_i(\mathbf{u})\gamma_m(i,\, j)\lambda_j^{1/2}\tilde{a}_j(\mathbf{u})-\sum_{K<i\leqslant m}\sum_{1\leqslant j\leqslant m}\lambda_i^{1/2}a_i(\mathbf{u})\gamma_m(i,\, j)\lambda_j^{1/2}\tilde{a}_j(\mathbf{u})\right|$$
$$\leqslant\Delta\left\{\sum_{K<i\leqslant m}\lambda_i\sup_{\mathbf{u}\in U_n(L)}|a_i(\mathbf{u})|+\sum_{K<i\leqslant m}\sum_{1\leqslant j\leqslant m}\lambda_i^{1/2}\lambda_j^{1/2}\frac{1}{i^2}\frac{1}{j^2}\Big(\sup_{\mathbf{u}\in U_n(L)}|a_i(\mathbf{u})|+\sup_{\mathbf{u}\in U_n(L)}|a_j(\mathbf{u})|\Big)\right\}$$

if $n\geqslant n_0$. Using condition (3.17) and Lemma 4.6, we get

$$\sum_{1\leqslant i\leqslant m}\lambda_i\sup_{\mathbf{u}\in U_n(L)}\big|a_i(\mathbf{u})\big|+\sum_{1\leqslant i,j\leqslant m}\lambda_i^{1/2}\lambda_j^{1/2}\frac{1}{i^2}\frac{1}{j^2}\Big(\sup_{\mathbf{u}\in U_n(L)}|a_i(\mathbf{u})|+\sup_{\mathbf{u}\in U_n(L)}|a_j(\mathbf{u})|\Big)=\mathrm{O}_P(1).$$

Thus, the proof of (4.28) will be complete if we prove that

$$\lim_{K\to\infty}\limsup_{n\to\infty}P\left\{\sup_{\mathbf{u}\in U_n(L)}n\left|\sum_{K<i\leqslant m}\sum_{1\leqslant j\leqslant m}\lambda_i^{1/2}a_i(\mathbf{u})\gamma_m(i,\, j)\lambda_j^{1/2}a_j(\mathbf{u})\right|>x\right\}=0. \tag{4.29}$$

The result in (4.29) is an immediate consequence of (3.17) and Lemmas 4.4–4.6.

REFERENCES

1. T. W. Anderson and D. A. Darling, Asymptotic theory for certain ‘goodness-of-fit’ criteria based on stochastic processes, *Ann. Math. Statist.*, **23**, 193–212 (1952).
2. I. Berkes and L. Horváth, The efficiency of the estimators of the parameters in GARCH processes, *Annals of Statistics*, **32** (2004).
3. I. Berkes, L. Horváth, and P. S. Kokoszka, Asymptotics for GARCH squared residual correlations, *Econometric Theory*, **19**, 515–540 (2003).
4. I. Berkes, L. Horváth, and P. S. Kokoszka, Estimation of the maximal moment exponent of a GARCH(1, 1) sequence, *Econometric Theory*, **19**, 565–586 (2003).
5. I. Berkes, L. Horváth, and P. S. Kokoszka, GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–227 (2003).
6. T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327 (1986).
7. P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Second edn., Springer, New York (1996).
8. R. H. Cameron and W. T. Martin, The Wiener measure of Hilbert neighborhoods in the space of real continuous functions, *J. Math. Phys.*, **23**, 195–209 (1944).
9. S. Csörgő and L. Horváth, On the Koziol–Green model for random censorship, *Biometrika*, **68**, 391–401 (1981).
10. R. F. Engle, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007 (1982).
11. D. M. Guillaume, M. M. Dacorogna, R. D. Dave, U. A. Müller, R. B. Olsen, and O. V. Pictet, 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 (1997).
12. L. Horváth and P. S. Kokoszka, Large sample distribution of ARCH(p) squared residual correlations, *Econometric Theory*, **17**, 283–295 (2001).

13. J. C. Hull, *Options*, *Futures & other Derivatives*, Prentice Hall, Upper Saddle River, NJ (2000).
14. S. W. Lee and B. E. Hansen, Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator, *Econometric Theory*, **10**, 29–52 (1994).
15. W. K. Li and T. K. Mak, On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity, *J. Time Series Anal.*, **15**, 627–636 (1994).
16. S. Ling and W. K. Li, On fractionally integrated autoregressive moving-average timeseries models with conditional heteroscedasticity, *J. Amer. Statist. Assoc.*, **92**, 1184–1194 (1997).
17. S. Ling and W. K. Li, Asymptotic inference for unit root processes with GARCH(1, 1) errors, *Econometric Theory*, **19**, 541–564 (2003).
18. R. L. Lumsdaine, Consistency and asymptotic normality of the quasi-likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models, *Econometrica*, **64**, 575–596 (1996).
19. G. Maercker and M. Moser, Yule–Walker estimators in GARCH(1, 1) models: asymptotic normality and bootstrap, Beiträge zur Statistik 58. Universität Heidelberg, Available at http://www.statlab.uni-heidelberg.de/reports/www.html (1999).
20. T. Mikosch and C. Stărică, Limit theory for the sample autocorrelation and extremes of a GARCH(1, 1) process, *The Annals of Statistics*, **28**, 1427–1451 (2000).
21. R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York (1982).
22. D. B. Nelson, Stationarity and persistence in the GARCH(1, 1) model, *Econometric Theory*, **6**, 318–334 (1990).
23. V. V. Petrov, *Limit Theorems of Probability Theory*: *Sequences of Independent Random Varianles*, Clarendon Press (1995).
24. E. D. Rothman and M. Woodroofe, A Cramér–von Mises type statistic for testing symmetry, *Ann. Math. Statist.*, **43**, 2035–2038 (1972).
25. G. A. F. Seber, *Linear Regression Analysis*, Wiley, New York (1977).
26. G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, New York (1986).
27. N. V. Smirnov, On the distribution of the von Mises ω^2-criterion, *Matem. Sbornik*, **5**, 973–993 (1937).
28. T. Teräsvirta, Two stylized facts and the GARCH(1, 1) model. Working Paper Series in Economics and Finance 96. Stockholm School of Economics (1996).
29. G. S. Watson, Goodness-of-fit tests on a circle, *Biometrika*, **48**, 109–114 (1961).
30. E. Zivot and J. Wang, *Modeling Financial Time Series with S-Plus*, Springer, Berlin (2003).