Asymptotics for GARCH Squared Residual Correlations

István Berkes*
A. Rényi Institute of Mathematics, Hungarian Academy of Sciences,
P.O. Box 127, H-1364 Budapest, Hungary

Lajos Horváth†
Department of Mathematics, University of Utah, 155 South 1440 East
Salt Lake City, UT 84112-0090 USA

Piotr Kokoszka†
Department of Mathematics and Statistics, Utah State University,
3900 Old Main Hill, Logan, UT 84322-3900 USA

We develop an asymptotic theory for quadratic forms of the autocorrelations of squared residuals from a GARCH(p,q) model. Denoting by $\hat{r}_n(k),\ k \geq 1$, these autocorrelations computed from a realization of length n, we show that the statistic $n(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))\hat{\mathbf{D}}_n^{-1}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))^T$, where $\hat{\mathbf{D}}_n$ is a matrix computed from the data, converges to the chi–square distribution with K degrees of freedom for any $1 \leq i_1 < \ldots < i_K$. Our results are valid under weak assumptions on the innovations and model coefficients which admit that arbitrary low order moments of the observations can be infinite. The matrix $\hat{\mathbf{D}}_n$ and its asymptotic limit $\mathbf{D}$ depend on the distribution of the innovations. A small simulation study illustrates the theory and shows, in particular, that using the matrix $\mathbf{D}$ computed under the assumption of normal innovations may lead to incorrect conclusions if the innovations have a different distribution.

AMS 2000 Subject Classification: Primary 62M10, Secondary 60F05
Key words and phrases: GARCH(p,q) sequence, quasi–maximum likelihood estimator, squared residuals, asymptotic normality, χ^2-tests.

*Supported by the Hungarian National Foundation for Scientific Research, Grant T 29621
†Supported by the grant NATO PST.CLG.977607

1 Introduction

The present paper extends earlier results presented in Horváth and Kokoszka (2001), where the asymptotic distribution of sums of squared residual autocorrelations in an ARCH(p) model was considered under the assumption that the realizations have finite fourth moment. In practical applications, however, the most frequently used model is GARCH$(1, 1)$, see e.g. Chapter 15 of Hull (2001) for an introduction and McNeil and Frey (2000) for a recent application. We consider the more general GARCH(p, q) model. If the parameters of a GARCH(p, q) are estimated by the quasi–maximum likelihood method (cf. Lee and Hansen (1994), Lumsdaine (1996) and Berkes, Horváth and Kokoszka (2001) for the general results), it is implicitly assumed that the innovations have at least four finite moments.

The statistical methodology investigated in this paper was proposed by Li and Mak (1994) and Ling and Li (1997) who aimed at developing portmanteau tests based on autocorrelations of squared residuals in a large class of heteroskedastic time series models. Using Gaussian formalism, they derived the asymptotic distributions of analogues of the Box–Pierce statistic. These authors assumed the conditional normality and finite fourth moments of the observations. As argued by Lee and Hansen (1994), see also Guillaume et al. (1997) for an empirical study, these assumptions need not necessarily be satisfied by models and time series encountered in econometric applications, and so an analogous theory under more general assumptions should be developed. As in Lee and Hansen (1994) and Berkes, Horváth and Kokoszka (2001) we do not assume that the innovations have a density and we allow the observations to have arbitrarily heavy tails, so that even the expected value need not exist. We show that under these more general assumptions the estimated autocorrelations are asymptotically normal. We then apply our results to construct a GARCH(p, q) specification test. Goodness–of–fit tests based on model residuals have long been an established tool in diagnostic checking of time series models with additive noise, see e.g. Chapter 8 of Box *et al* (1994) for an introduction and Hong (1996), de Jong (1996), Bierens and Ploberber (1999) for some new approaches. In this paper we focus on a model with multiplicative noise.

Horváth *et al.* (2001) obtained the asymptotic distribution of the sequential empirical process of the squared residuals of ARCH(p) processes and showed that it does not converge to a standard Gaussian process, so, in contrast to the approach considered in the present paper, it is unlikely that asymptotically pivotal goodness-of-fit tests based on the empirical process can be developed.

Berkes and Horváth (2002) extended the results of Horváth *et al.* (2001) to the setting considered in this paper while Horváth *et al.* (2002) studied by means of simulations the corresponding bootstrap tests.

A GARCH(p, q) process is defined by the equations

$$y_k = \sigma_k \epsilon_k \tag{1.1}$$

and

$$\sigma_k^2 = \omega + \sum_{1 \le i \le p} \alpha_i y_{k-i}^2 + \sum_{1 \le j \le q} \beta_j \sigma_{k-j}^2, \tag{1.2}$$

where

$$\omega > 0, \quad \alpha_i \ge 0, \quad 1 \le i \le p, \quad \beta_j \ge 0, \quad 1 \le j \le q \tag{1.3}$$

are constants. We also assume that

$$\{\epsilon_i, -\infty < i < \infty\} \text{ are independent, identically distributed random variables.} \tag{1.4}$$

Nelson (1990) showed that in the case of GARCH$(1, 1)$ sequence, (1.1) and (1.2) have a unique stationary solution if and only if $E \log(\beta_1 + \alpha_1 \epsilon_0^2) < 0$. Bougerol and Picard (1992a,b) found necessary and sufficient conditions for the existence of a unique stationary solution of (1.1) and (1.2) in the case of general GARCH(p, q). To state their condition we must introduce further notation. Let

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

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

and

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

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

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

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

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

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

$$\gamma_L = \inf_{0\le n<\infty} \frac{1}{n+1} E \log ||A_0 A_1 \dots A_n||,$$

assuming that

$$E(\log ||A_0||) < \infty. \tag{1.5}$$

(We note that $||A_0|| \geq 1$; cf. Berkes, Horváth and Kokoszka (2001)). Bougerol and Picard (1992a,b) showed that if (1.5) holds, then (1.1) and (1.2) have a unique stationary solution if and only if

$$\gamma_L < 0. \tag{1.6}$$

Throughout this paper we will assume that conditions (1.1)–(1.6) hold. Clearly, they are a minimal set of conditions for the existence and stationarity of the GARCH(p,q) sequence.

Lumsdaine (1996) studied the quasi–maximum likelihood estimator of $\boldsymbol{\theta} = (\omega, \alpha_1, \alpha_2, \cdots, \alpha_p, \beta_1, \beta_2, \cdots, \beta_q)$ from $y_1, y_2, \cdots, y_n$ and proved its (weak) consistency and asymptotic normality for the GARCH$(1,1)$ model. We wish to note that the consistency (and therefore the asymptotic normality as well) is local, i.e. in the definition of the quasi–maximum likelihood estimator, the maximum of the likelihood function is computed in a small neighborhood of $\boldsymbol{\theta}$. Berkes, Horváth and Kokoszka (2001) removed some unnecessary conditions assumed by Lumsdaine (1996) and extended the almost sure consistency and asymptotic normality to GARCH(p,q) sequences. The conditions in Berkes, Horváth and Kokoszka (2001) are essentially optimal, so we follow their approach.

We continue with the formal definition of the quasi–maximum likelihood estimator for $\boldsymbol{\theta}$. First we define a sequence of functions by a recursion. Let $\mathbf{u} = (x, \mathbf{s}, \mathbf{t}) \in \mathbb{R}^{p+q+1}$, $\mathbf{s} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$. If $q \geq p$, then

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

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

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

If $i > \max(p, q)$, then

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

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

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

where

$$\tilde{w}_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\leq i\leq k-1} c_i(\mathbf{u}) y_{k-i}^2. \tag{1.7}$$

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

$$\begin{aligned}
U = \{\mathbf{u} : t_1 + t_2 + \cdots + t_q \leq \rho_0 \text{ and } u_1 \leq \min(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \\
\leq \max(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \leq u_2\},
\end{aligned}$$

$\mathbf{u} = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$, with some $0 < u_1 < u_2$, $0 < \rho_0 < 1$ and $qu_1 < \rho_0$.

2 Main Results

If the innovations $\epsilon_1, \ldots, \epsilon_n$ were observable, an estimator for the autocovariance function of the squared innovations would be

$$r_n(k) = \frac{1}{n}\sum_{k<i\leq n}(\epsilon_i^2 - 1)(\epsilon_{i-k}^2 - 1).$$

(We assume in (2.5) below that $E\epsilon_0^2 = 1$.) Since $\epsilon_1^2, \ldots, \epsilon_n^2$ are not observable, we replace them with the squared residuals

$$\hat{\epsilon}_i^2 = y_i^2/\tilde{w}_i(\tilde{\boldsymbol{\theta}}_n), \quad 1 \le i \le n$$

and we shall use

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

Our first result compares $\hat{r}_n(k)$ to $r_n(k)$. First we list the conditions which are shown in Berkes, Horváth and Kokoszka (2001) to be sufficient for the asymptotic normality of $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$:

$$\epsilon_0^2 \text{ is a nondegenerate random variable,} \tag{2.1}$$

$$\text{the polynomials } \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_p x^p \text{ and } 1 - \beta_1 x - \beta_2 x^2 - \cdots - \beta_q x^q \text{ are coprimes in the set of polynomials with real coefficients,} \tag{2.2}$$

$$\boldsymbol{\theta} \text{ is in the interior of } U, \tag{2.3}$$

$$\lim_{t\to 0} t^{-\mu} P\{\epsilon_0^2 \le t\} = 0 \text{ with some } \mu > 0, \tag{2.4}$$

$$E\epsilon_0^2 = 1 \tag{2.5}$$

and

$$E|\epsilon_0^2|^{2+\delta} < \infty \text{ with some } \delta > 0. \tag{2.6}$$

If k gets large then $\tilde{w}_k(\mathbf{u})$ of (1.7) can be replaced with

$$w_k(\mathbf{u}) = c_0(\mathbf{u}) + \sum_{1\le i<\infty} c_i(\mathbf{u}) y_{k-i}^2. \tag{2.7}$$

Here the infinite series converges with probability one by Lemmas 4.1 and 4.4 and relation (5.4) below. Let

$$\mathbf{c}_k = E\left[(\epsilon_{-k}^2 - 1)\frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\right]. \tag{2.8}$$

We show in Section 4 that $\mathbf{c}_k$ is finite. We use $\mathbf{u}^T$ to denote transpose of a vector or a matrix.

Theorem 2.1 *If (2.1)–(2.6) hold, then for any* $0 \leq k < \infty$

$$\hat{r}_n(k) = r_n(k) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n)\mathbf{c}_k^T + o_P(n^{-1/2}). \tag{2.9}$$

According to Theorem 2.1 the first order properties (i.e. the mean, law of large numbers) are the same for $\hat{r}_n(k)$ as for $r_n(k)$. However, they differ in the second order properties (i.e. variance, central limit theorem).

Theorem 2.1 implies the asymptotic normality of $\hat{r}_k(k)$ for any $1 \leq k < \infty$. In order to get χ^2–statistics we shall prove the joint asymptotic normality of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$ for any $1 \leq i_1 < i_2 < \cdots < i_K$. We start with the description of the asymptotic covariance matrix of $n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K))$. Let

$$d_0^2 = E(\epsilon_0^2 - 1)^2,$$

$$\mathbf{A}_0 = \operatorname{cov}(\ell_0'(\boldsymbol{\theta})) \tag{2.10}$$

and

$$\mathbf{B}_0 = E(\ell_0''(\boldsymbol{\theta})), \tag{2.11}$$

where

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

We note that the matrices in (2.10) and (2.11) have already appeared in the asymptotic covariance matrix of $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ (cf. Berkes, Horváth and Kokoszka (2001)) and they were proved to be nonsingular. Their appearance in our theorems is due to the presence of $\tilde{\boldsymbol{\theta}}_n$ in the definition of the squared residuals. For any $1 \leq K < \infty$ and $1 \leq i_1 < i_2 < \ldots < i_K$ let $\mathbf{D}$ be the matrix with elements

$$\mathbf{D}(k,k) = d_0^4 + d_0^2 \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{c}_{i_k}^T + \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}\mathbf{c}_{i_k}^T$$

and

$$\mathbf{D}(k,l) = d_0^2 \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{c}_{i_\ell}^T + \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}\mathbf{c}_{i_\ell}^T, \quad \text{if } k \neq \ell.$$

Theorem 2.2 *If (2.1)–(2.6) hold, then*

$$n^{1/2}(\hat{r}_n(i_1), \ldots, \hat{r}_n(i_K)) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{D}), \tag{2.12}$$

where $N(\mathbf{0}, \mathbf{D})$ *stands for a multivariate normal random variable with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{D}$.

Remark 1.1. Berkes, Horváth and Kokoszka (2001) proved that $\mathbf{A}_0 = -\mathbf{B}_0$, if $E\epsilon_0^4 = 3$ which is satisfied, e.g. if ϵ_0 is standard normal. In this case $\mathbf{D}(k,k) = 4 + \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{c}_{i_k}^T$ and $\mathbf{D}(k,\ell) = \mathbf{c}_{i_k}\mathbf{B}_0^{-1}\mathbf{c}_{i_\ell}^T$, if $k \neq \ell$.

Li and Mak (1994) and Ling and Li (1997) assumed that conditionally on the past the y's have normal distribution. In our setting, this reduces to standard normal innovations ϵ_i. In this case, our results accord with those of these authors.

The covariance matrix in Theorem 2.2 has a very special structure. Let $\mathbf{C}$ be the $K \times (p+q+1)$ matrix defined as

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

and let $\mathbf{I}_K$ denote the $K \times K$ identity matrix. Using this notation we can write

$$\mathbf{D} = d_0^4(\mathbf{I}_K + \mathbf{C}\mathbf{S}\mathbf{C}^T), \tag{2.13}$$

where

$$\mathbf{S} = \frac{1}{d_0^2}\mathbf{B}_0^{-1} + \frac{1}{d_0^4}\mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{B}_0^{-1}. \tag{2.14}$$

Theorem 2.3 *If (2.1)–(2.6) hold and*

$$P\{\epsilon_0^2 \leq t\} \ \textit{ is continous,} \tag{2.15}$$

then

$$\mathbf{D} \ \textit{ is a positive definite matrix.} \tag{2.16}$$

Theorem 2.3 yields that $\mathbf{D}^{-1}$ exists. The existence of $\mathbf{D}^{-1}$ was used in Li (1992), Li and Mak (1994) and Ling and Li (1997) without proving it.

Combining Theorems 2.2 and 2.3 we get the convergence in distribution of quadratic forms of $\hat{r}_n(i_1), \hat{r}_n(i_2), \ldots, \hat{r}_n(i_K)$.

Theorem 2.4 *If (2.1)–(2.6) hold, then*

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

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

The matrix $\mathbf{D}$ is unknown in Theorems 2.2 and 2.3 so these results cannot be applied immediately. However, $\mathbf{D}$ can be easily estimated from the data. Let

$$\tilde{\ell}_k(\mathbf{u}) = -\frac{1}{2}\left\{\log \tilde{w}_k(\mathbf{u}) + \frac{y_k^2}{\tilde{w}_k(\mathbf{u})}\right\}.$$

We note that $\sum_{1<k\leq n} \tilde{\ell}'_k(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0}$ if n is large enough. Hence

$$\hat{\mathbf{A}}_n = \frac{1}{n}\sum_{1<i\leq n} (\tilde{\ell}'_i(\tilde{\boldsymbol{\theta}}_n))^T \tilde{\ell}'_i(\tilde{\boldsymbol{\theta}}_n)$$

can be used to estimate $\mathbf{A}_0$. We use

$$\hat{\mathbf{B}}_n = \frac{1}{n}\sum_{1<i\leq n} \tilde{\ell}''_i(\tilde{\boldsymbol{\theta}}_n),$$

$$\hat{\mathbf{c}}_{k,n} = \frac{1}{n-k}\sum_{1\leq i\leq n-k} (\hat{\epsilon}_i^2 - 1)\frac{\tilde{w}'_{k+i}(\tilde{\boldsymbol{\theta}}_n)}{\tilde{w}_{k+i}(\tilde{\boldsymbol{\theta}}_n)}$$

and

$$\hat{d}_n^2 = \frac{1}{n}\sum_{1\leq i\leq n} (\hat{\epsilon}_i^2 - 1)^2$$

to estimate $\mathbf{B}_0$, $\mathbf{c}_k$ and d_0^2, respectively. Let

$$\hat{\mathbf{C}}_n = \begin{pmatrix} \hat{\mathbf{c}}_{i_1,n} \\ \hat{\mathbf{c}}_{i_2,n} \\ \vdots \\ \hat{\mathbf{c}}_{i_K,n} \end{pmatrix}.$$

Replacing d_0, $\mathbf{C}$ and $\mathbf{S}$ with the corresponding estimators in (2.13) and (2.14) we get $\hat{\mathbf{D}}_n^{-1}$, our estimator for $\mathbf{D}^{-1}$.

Theorem 2.5 *If (2.1)–(2.6) hold, then*

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

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

The proofs of Theorems 2.1–2.4 will be postponed until Section 5. Next we discuss a small simulation study to illustrate the applicability of our results.

3 A simulation study

The goal of the small simulation study which we performed in conjunction with the theoretical research presented in this paper was to develop some idea on how the goodness–of–fit procedures we propose might perform in finite samples. The results presented here are intended only to highlight some points of interest and are by no means intended as a guide for practitioners. If anything, we hope that our research and the simulation results presented in this section will encourage further investigation of these tests, and this should lead to a better fine-tuning of the finite sample procedures. It is also possible that a more extensive simulation study would modify our tentative conclusions.

In all our simulations we estimated the GARCH(1,1) model, which is most frequently used in practice. Under the null hypothesis the simulated series follow the GARCH(1,1) model and under the alternative they do not follow GARCH(1,1). Ling and Li (1997) performed a similar simulation study assuming that the errors are normal. For this reason we focus here on non-normal errors. We wanted to shed some light on the following questions:
1) If the innovations are very different from normal, but the underlying model is correctly specified, is there a distinct possibility of arriving at incorrect conclusions if one uses the statitistic proposed by Ling and Li (1997) (which was derived under the assumption of Gaussian errors and should be used only for such errors)?
2) Ling and Li (1997) assumed finite fourth moment of the observations, but our theoretical results are valid without this assumption. Do the tests have correct size in finite samples when the fourth moment is infinite?
Beyond these two questions, we would like to see how much is gained by applying the procedures proposed by Ling and Li (1997) and developed in this paper rather than using the "standard" statistic $n\sum_{1\le i\le K}\hat{r}_n^2(i)$, and to have some idea about the power of the test against some typical alternatives.

With these questions in mind we considered the following test statistics:

$$Q(K) = n(\hat{r}_n(1),\ldots,\hat{r}_n(K))\hat{\mathbf{D}}_n^{-1}(\hat{r}_n(1),\ldots,\hat{r}_n(K))^T,$$

where $\hat{\mathbf{D}}_n$ is the estimator from Theorem 2.5. The analogue of the statistic $Q(K)$ in situations in which we *know* that the errors are standard normal (or at least have the same first four moments as a standard normal random variable), c.f. Remark 1.1, is

$$Q^{(G)}(K) = n(\hat{r}_n(1),\ldots,\hat{r}_n(K))\{\hat{\mathbf{D}}_n^{(G)}\}^{-1}(\hat{r}_n(1),\ldots,\hat{r}_n(K))^T,$$

where $\hat{\mathbf{D}}_n^{(G)}$ is the estimator of

$$\mathbf{D}^{(G)} = 4\mathbf{I}_K + \mathbf{C}\mathbf{B}_0^{-1}\mathbf{C}^T,$$

obtained by replacing $\mathbf{C}$ and $\mathbf{B}_0$ with the estimates proposed in Section 2. We also take a look at the statistic

$$Q^*(K) = n \sum_{1 \le i \le K} [\hat{r}_n(i)]^2$$

to see what conclusions a researcher not familiar with the theory developed in this paper and the work of Ling and Li (1997) might arrive at if she/he incorrectly used the "usual" statistic $Q^*(K)$. In our simulations, we treated $Q^*(K)$ as if it had the asymptotic chi-squared distribution with K degrees of freedom (in analogy to $Q(K)$). If we treated it, as is usual in case of ARIMA models, as if it had less than K degrees of freedom, its performance would be even worse.

We present here detailed results for two models:

N1) $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \omega = 0.01, \alpha = 0.1, \beta = 0.8,$

N2) $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \omega = 0.01, \alpha = 0.5, \beta = 0.5.$

The parameters of the model N1 are typical for daily stock returns, whereas model N2 has very heavy tails: $P(|y| > u) \sim cu^{-2}$, so that even the variance of the observations is infinite.

We considered sample of sizes 250, 500. These are natural sample sizes if one wants to model daily returns on speculative assets because there are about 250 trading days per year. The empirical sizes are based on one thousand replications and so the standard error of the size estimates is about 0.005.

Comparing the entries in Tables 1 and 2, we see that using the "Gaussian" matrix $\hat{\mathbf{D}}_n^{(G)}$ instead of the matrix $\hat{\mathbf{D}}_n$ can lead to enormous size distortions if the errors are far from normal, double exponential in this case. In such cases using the statistic $Q^*(K)$ instead of the correct statistic $Q(K)$ also causes a very large size distortion, see Table 1. The latter effect is not so noticeable if the innovations *are* normal, as can be gleaned from Table 3 which also shows that in such cases using the matrix $\hat{\mathbf{D}}_n^{(G)}$ gives better results than using $\hat{\mathbf{D}}_n$. In order to estimate $\mathbf{D}$ one has to estimate d_0^2 and $\mathbf{A}_0$, and this is not required when estimating $\mathbf{D}^{(G)}$.

The empirical sizes of tests based on the statistic $Q(K)$ see Table 1, are smaller than the nominal size of .05, even though the sample size of $n = 500$ is quite large. The situation here appears to be quite similar to what has been observed for the statistic $Q^*(K)$ in the context of ARIMA models, see Section 8.2.2 of Box *et al*

		$Q(4)$	$Q^*(4)$	$Q(6)$	$Q^*(6)$
Model N1	n=250	.037	.447	.044	.596
Model N1	n=500	.027	.538	.032	.684
Model N2	n=250	.020	.474	.023	.624
Model N2	n=500	.037	.576	.035	.707

TABLE 1. Frequencies of rejection at five per cent significance level of the null hypothesis of a GARCH(1,1) specification. The errors are double exponential and the matrix $\hat{\mathbf{D}}_n$ is used.

		$Q^{(G)}(4)$	$Q^*(4)$	$Q^{(G)}(6)$	$Q^*(6)$
Model N1	n=250	.412	.447	.550	.596
Model N1	n=500	.479	.538	.642	.684
Model N2	n=250	.427	.474	.553	.624
Model N2	n=500	.441	.576	.576	.707

TABLE 2. Frequencies of rejection at five per cent significance level of the null hypothesis of a GARCH(1,1) specification. The errors are double exponential but the matrix $\hat{\mathbf{D}}_n^{(G)}$ is used.

(1994) where a discussion and further references are provided. To address this problem Ljung and Box (1978) modified the statistic $Q^*(K)$ by using a more accurate approximation for the variance of the white noise autocorrelations. It might be interesting to see if such a finite sample correction is possible in the present setting. For samples of size $n = 10^3$ and the statistic $Q(K)$ the empirical size is around .04, but in applications stationarity of such long samples may be questionable. Observe that for model N1, see Tables 1 and 2, the empirical size appears to be closer to the nominal size for $n = 250$ than for $n = 500$. We repeated these simulations with different sets of random numbers and slightly different models and this anomaly appears to exist when β is close to 1 and α close to zero. It does not appear for other choices of parameters (e.g model N2).

To investigate the power of the tests we focused on two alternatives

A1) $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2$, $\omega = 0.01, \alpha = 0.1, \beta_1 = 0.1, \beta_2 = 0.7$,

A2) $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta(t)\sigma_{t-1}^2$, $\omega = 0.01, \alpha = 0.1, \beta(t) = 0.8, t \leq n/2$, $\beta(t) = 0.0, t > n/2$.

		$Q(4)$	$Q^*(4)$	$Q(6)$	$Q^*(6)$
Model A1	n=250	.21	.23	.22	.27
Model A1	n=500	.16	.26	.20	.27
		$Q^{(G)}(4)$	$Q^*(4)$	$Q^{(G)}(6)$	$Q^*(6)$
Model A2	n=250	.43	.23	.41	.27
Model A2	n=500	.56	.26	.74	.31

TABLE 3. Frequencies of rejection at five per cent significance level of the null hypothesis of a GARCH(1,1) specification. The errors are standard normal.

For sample sizes of 250 and 500 the power for the statistic $Q(K)$ ranges from 0.1 to 0.3 what accords with the low empirical size reported above.

In conclusion, our limited simulation study indicates that the goodness–of–fit tests based on the statistics studied in this paper may be quite useful when finite sample approximations to their distributions are better understood, but even in the present form they may serve as preliminary diagnostic tools. Our study clearly establishes that it is not acceptable to use tests derived under the assumption of the normality of errors if these are far from normal.

4 Some technical results

We start with a partial sum representation of $\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$.

Theorem 4.1 *If (2.1)–(2.6) hold, then*

$$\mathbf{A}_0 \text{ is non-singular,} \tag{4.1}$$

$$\mathbf{B}_0 \text{ is non-singular,} \tag{4.2}$$

and

$$\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \frac{1}{n} \sum_{1 \le k \le n} \frac{1}{2}(1 - \epsilon_k^2) \frac{w_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})} \mathbf{B}_0^{-1} + o_P(n^{-1/2}). \tag{4.3}$$

PROOF: Lee and Hansen (1994) and Lumsdaine (1996) obtained (4.1)–(4.3) for GARCH$(1,1)$ under stronger conditions. Theorem 4.1 was proven by Berkes, Horváth and Kokoszka (2001) under the present conditions for general GARCH(p,q) processes.

Lemma 4.1 *If (2.1)–(2.6) hold, then there is* $\delta^* > 0$ *such that*

$$E|y_0^2|^{\delta^*} < \infty \quad and \quad E|\sigma_0^2|^{\delta^*} < \infty. \tag{4.4}$$

Proof: Lemma 2.3 of Berkes, Horváth and Kokoszka (2001) yields (4.4).

For any $\gamma > 0$ let $U(\gamma) = \{\mathbf{u} \in U : ||\boldsymbol{\theta} - \mathbf{u}|| \le \gamma\}$.

Lemma 4.2 *If (2.1)–(2.6) hold, then for any* $-\infty < \nu < \infty$ *there is a* $0 < \gamma < \infty$ *such that*

$$E\left(\sup_{\mathbf{u}\in U(\gamma)} \frac{w_0(\boldsymbol{\theta})}{w_0(\mathbf{u})}\right)^{\nu} < \infty. \tag{4.5}$$

Also, for any $\kappa > 0$

$$E\left(\sup_{\mathbf{u}\in U} \frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})}\right)^{\kappa} < \infty, \tag{4.6}$$

and

$$E\left(\sup_{\mathbf{u}\in U} \frac{||w_0''(\mathbf{u})||}{w_0(\mathbf{u})}\right)^{\kappa} < \infty. \tag{4.7}$$

Proof: The result in (4.5) is a special case of Lemma 3.7 in Berkes and Horváth (2001).

Combining Lemmas 3.2, 3.3 and 5.2 in Berkes, Horváth and Kokoszka (2001) and Lemma 4.1 we obtain immediately (4.6) and (4.7) (cf. also Lemma 3.4 in Berkes and Horváth (2001)).

The next lemma wil be used to show that $\mathbf{D}^{-1}$ exists. Let $1 \le i_1 < i_2 < \ldots, i_K$ and define

$$\boldsymbol{\mu} = \left(\epsilon_{-i_1}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\mathbf{z}_1^T, \ldots, \epsilon_{-i_K}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\mathbf{z}_K^T\right),$$

where $\mathbf{z}_1 \in \mathbb{R}^{p+q+1}, \ldots, \mathbf{z}_K \in \mathbb{R}^{p+q+1}$ are constants.

Lemma 4.3 *If (2.1)–(2.6) and (2.15) hold, then* $\mathbf{D}_{\boldsymbol{\mu}}^* = E\boldsymbol{\mu}^T\boldsymbol{\mu}$, *is non–singular and positive definite.*

PROOF: Condition (2.6) and Lemma 4.2 imply that $E||\boldsymbol{\mu}||^2 < \infty$ and therefore $\mathbf{D}^*_{\boldsymbol{\mu}}$ exists. Thus it is enough to prove that $\mathbf{D}^*_{\boldsymbol{\mu}}$ is non–singular.

We assume that $\mathbf{D}^*_{\boldsymbol{\mu}}$ is singular and show that this assumption leads to contradiction. If $\mathbf{D}^*_{\boldsymbol{\mu}}$ is singular, then there are λ_0 and $(\lambda_1, \dots, \lambda_K) \neq \mathbf{0}$ such that

$$\sum_{1 \le j \le K} \lambda_j \left(\epsilon^2_{-i_j} + \frac{w_0'(\boldsymbol{\theta})}{w(\boldsymbol{\theta})} \mathbf{z}_j^T \right) = \lambda_0.$$

Hence by (2.7) and Berkes, Horváth and Kokoszka (2001) there are constants $\gamma_0, \gamma_1, \dots$ satisfying $|\gamma_i| = O(\rho_*^i)$ with some $0 < \rho_* < 1$ such that

$$\left(\lambda_0 - \sum_{1 \le j \le K} \lambda_j \epsilon^2_{-i_j} \right) w_0(\boldsymbol{\theta}) = \sum_{1 \le j < \infty} \gamma_j y^2_{-j} + \gamma_0. \tag{4.8}$$

By (1.1) and (1.2) we have that

$$w_0(\boldsymbol{\theta}) = \sigma_0^2 = \alpha_1 \epsilon^2_{-1} \sigma^2_{-1} + \omega + \sum_{2 \le i \le p} \alpha_i y^2_{-i} + \sum_{1 \le j \le q} \beta_j \sigma^2_{-j}. \tag{4.9}$$

Replacing $w_0(\boldsymbol{\theta})$ in (4.8) with the expression in (4.9) we obtain that

$$\alpha_1 \left(\lambda_0 - \sum_{2 \le j \le K} \lambda_j \epsilon^2_{-i_j} \right) \epsilon^2_{-1} \sigma^2_{-1} - \lambda_1 \epsilon^2_{-i_1} \sigma_0^2 - \gamma_1 \epsilon^2_{-1} \sigma^2_{-1} = \mathbf{Z}, \tag{4.10}$$

where

$$\mathbf{Z} = \sum_{2 \le j < \infty} \gamma_j y^2_{-j} + \gamma_0 + \left(\sum_{2 \le j \le K} \lambda_j \epsilon^2_{-i_j} - \lambda_0 \right) \left(\omega + \sum_{2 \le i \le p} \alpha_i y^2_{-i} + \sum_{1 \le j \le q} \beta_j \sigma^2_{-j} \right).$$

It follows from the definition of the GARCH sequence that $\mathbf{Z}$ is measurable with respect to $\sigma(\epsilon^2_{-2}, \epsilon^2_{-3}, \dots)$, the σ–algebra generated by $\epsilon^2_{-2}, \epsilon^2_{-3}, \dots$. Equation (4.10) yields that

$$\epsilon^2_{-1} = \frac{\mathbf{Z} + \left(\omega + \sum_{2 \le i \le p} \alpha_i y^2_{-i} + \sum_{1 \le j \le q} \beta_j \sigma^2_{-j} \right) \lambda_1 \epsilon^2_{-i_1}}{\sigma^2_{-1} \left(\alpha_1 \left(\lambda_0 - \sum_{1 \le j \le K} \lambda_j \epsilon^2_{-i_j} \right) - \gamma_1 \right)}. \tag{4.11}$$

Condition (2.15) implies that

$$P\left\{\alpha_1\left(\lambda_0-\sum_{1\le j\le K}\lambda_j\epsilon^2_{-i_j}\right)-\gamma_1=0\right\}=0$$

and since $\sigma^2_{-1}\ge\omega>0$, the right hand side of (4.11) is finite almost surely. If $i_1>1$ or $i_1=1$ and $\lambda_1=0$, then the right hand side of (4.11) is measurable with respect to the $\sigma(\epsilon^2_{-2},\epsilon^2_{-3},\ldots)$. By (2.1) this contradicts (1.4).

Now we assume that $i_1=1$ and $\lambda_1\ne 0$. We can rewrite (4.11) as

$$-\lambda_1\alpha_1\sigma^2_{-1}\epsilon^4_{-1}-\epsilon^2_{-1}\left[\lambda_1\left(\omega+\sum_{2\le i\le p}\alpha_i y^2_{-i}+\sum_{1\le j\le q}\beta_j\sigma^2_{-j}\right)\right.$$
$$\left.-\alpha_1\left(\lambda_0-\sum_{2\le j\le K}\lambda_j\epsilon^2_{-i_j}\right)\sigma^2_{-1}+\gamma_1\sigma^2_{-1}\right]=\mathbf{Z}.$$

The equation above is a quadratic function of ϵ^2_{-1}. Since the coefficients are measurable with respect to $\sigma(\epsilon^2_{-2},\epsilon^2_{-3},\ldots)$, the solution, ϵ^2_{-1}, must be measurable with respect to $\sigma(\epsilon^2_{-2},\epsilon^2_{-3},\ldots)$, contradicting (1.4) and (2.1).

The last lemma will be used to estimate the difference between $w_k(\mathbf{u})$ and $\tilde{w}_k(\mathbf{u})$. Let $\log^+ x=\log x$, if $x\ge 1$ and 0 otherwise. We cite Lemma 2.2 of Berkes, Horváth and Kokoszka (2001) for further reference.

Lemma 4.4 *If $\{\xi_k, 0\le k<\infty\}$ is a sequence of identically distributed random variables satisfying*

$$E\log^+|\xi_0|<\infty,$$

then $\sum_{0\le k<\infty}\xi_k z^k$ *converges with probability one for any* $|z|<1$.

5 Proof of Theorems 2.1–2.4

Let

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

and

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

Elementary algebra gives

$$
\begin{aligned}
a_k(\mathbf{u}) = \; & \sum_{k<i\le n} (\epsilon_i^2 - 1)(\epsilon_{i-k}^2 - 1) \\
& + \sum_{k<i\le n} \epsilon_i^2 \left(\frac{\sigma_i^2}{w_i(\mathbf{u})} - 1 \right) (\epsilon_{i-k}^2 - 1) \\
& + \sum_{k<i\le n} (\epsilon_i^2 - 1)\epsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{w_{i-k}(\mathbf{u})} - 1 \right) \\
& + \sum_{k<i\le n} \epsilon_i^2 \left(\frac{\sigma_i^2}{w_i(\mathbf{u})} - 1 \right) \epsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{w_{i-k}(\mathbf{u})} - 1 \right) \\
= \; & A_{k,1} + A_{k,2}(\mathbf{u}) + A_{k,3}(\mathbf{u}) + A_{k,4}(\mathbf{u})
\end{aligned} \tag{5.1}
$$

and similarly

$$
\tilde{a}_k(\mathbf{u}) = A_{k,1} + \tilde{A}_{k,2}(\mathbf{u}) + \tilde{A}_{k,3}(\mathbf{u}) + \tilde{A}_{k,4}(\mathbf{u}), \tag{5.2}
$$

where

$$
\tilde{A}_{k,2}(\mathbf{u}) = \sum_{k<i\le n} \epsilon_i^2 \left(\frac{\sigma_i^2}{\tilde{w}_i(\mathbf{u})} - 1 \right) (\epsilon_{i-k}^2 - 1),
$$

$$
\tilde{A}_{k,3}(\mathbf{u}) = \sum_{k<i\le n} (\epsilon_i^2 - 1)\epsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{\tilde{w}_{i-k}(\mathbf{u})} - 1 \right)
$$

and

$$
\tilde{A}_{k,4}(\mathbf{u}) = \sum_{k<i\le n} \epsilon_i^2 \left(\frac{\sigma_i^2}{\tilde{w}_i(\mathbf{u})} - 1 \right) \epsilon_{i-k}^2 \left(\frac{\sigma_{i-k}^2}{\tilde{w}_{i-k}(\mathbf{u})} - 1 \right).
$$

Lemma 5.1 *If (2.1)–(2.6) hold, then there is $\gamma > 0$ such that*

$$
\max_{0\le k<n} \sup_{\mathbf{u}\in U(\gamma)} |a_k(\mathbf{u}) - \tilde{a}_k(\mathbf{u})| = o(n^{1/(2+\delta')}) \;\; a.s. \tag{5.3}
$$

for all $\delta' < \delta$, where δ is defined in (2.6).

PROOF: By Lemma 3.1 in Berkes, Horváth and Kokoszka (2001) there is a constant C_1 such that

$$
\sup_{\mathbf{u}\in U} c_i(\mathbf{u}) \le C_1 \rho_0^{i/q}, \qquad 0 \le i < \infty. \tag{5.4}
$$

Hence,

$$|w_j(\mathbf{u}) - \tilde{w}_j(\mathbf{u})| = \sum_{j \le \ell < \infty} c_\ell(\mathbf{u}) y_{j-\ell}^2 \le C_1 \sum_{j \le \ell < \infty} \rho_0^{\ell/q} y_{j-\ell}^2 = C_1 \rho_0^{j/q} \zeta, \tag{5.5}$$

where $\zeta = \sum_{0 \le i < \infty} \rho_0^{i/q} y_{-i}^2$. By Lemmas 4.1 and 4.4 ζ is finite with probability one.

Since $c_0(\mathbf{u}) \ge u_1 > 0$ and $c_i(\mathbf{u}) \ge 0$ for all $\mathbf{u} \in U$, we get

$$\begin{aligned}
&\max_{0 \le k < n} \sup_{\mathbf{u} \in U} |A_{k,2}(\mathbf{u}) - \tilde{A}_{k,2}(\mathbf{u})| \\
&\le \max_{0 \le k < n} \sum_{k < i \le n} |\epsilon_{i-k}^2 - 1| \epsilon_i^2 \sigma_i^2 \sup_{\mathbf{u} \in U} \frac{|w_i(\mathbf{u}) - \tilde{w}_i(\mathbf{u})|}{w_i(\mathbf{u}) \tilde{w}_i(\mathbf{u})} \\
&\le \frac{C_1}{u_1^2} \zeta \max_{0 \le k < n} \sum_{k < i \le n} |\epsilon_{i-k}^2 - 1| y_i^2 \rho_0^{i/q} \\
&\le \frac{C_1}{u_1^2} \zeta \max_{1 \le k < n} |\epsilon_k^2 - 1| \sum_{0 \le i < \infty} \rho_0^{i/q} y_i^2
\end{aligned} \tag{5.6}$$

on account of (5.5). Using again Lemmas 4.1 and 4.4 we conclude that

$$\sum_{0 \le i < \infty} \rho_0^{i/q} y_i^2 \text{ converges with probability one.} \tag{5.7}$$

By (2.6) we have that

$$|\epsilon_j^2 - 1| = o(j^{1/(2+\delta)}) \text{ a.s. as } j \to \infty \tag{5.8}$$

and therefore by (5.6) we have

$$\max_{0 \le k < n} \sup_{\mathbf{u} \in U} |A_{k,2}(\mathbf{u}) - \tilde{A}_{k,2}(\mathbf{u})| = o(n^{1/(2+\delta)}) \text{ a.s.} \tag{5.9}$$

Proceeding as in (5.6), by (5.5), (5.7) and (5.8) we have

$$\begin{aligned}
&\max_{0\le k<n}\sup_{\mathbf{u}\in U}|A_{k,3}(\mathbf{u})-\tilde{A}_{k,3}(\mathbf{u})| \\
&\le \max_{0\le k<n}\sum_{k<i\le n}|\epsilon_i^2-1|y_{i-k}^2\sup_{\mathbf{u}\in U}\frac{|w_{i-k}(\mathbf{u})-\tilde{w}_{i-k}(\mathbf{u})|}{w_{i-k}(\mathbf{u})\tilde{w}_{i-k}(\mathbf{u})} \\
&\le \frac{C_1}{u_1^2}\zeta\max_{0\le k<n}\sum_{k<i\le n}|\epsilon_i^2-1|y_{i-k}^2\rho_0^{(i-k)/q} \\
&\le \frac{C_1}{u_1^2}\zeta\max_{1\le i<n}|\epsilon_i^2-1|\sum_{0<\ell<\infty}\rho_0^{\ell/q}y_\ell^2 \\
&\overset{a.s.}{=} o(n^{1/(2+\delta)}).
\end{aligned} \tag{5.10}$$

Next we write

$$\begin{aligned}
&|A_{k,4}(\mathbf{u})-\tilde{A}_{k,4}(\mathbf{u})| \\
&\le \sum_{k<i\le n}\epsilon_i^2\left|\frac{\sigma_i^2}{\tilde{w}_i(\mathbf{u})}-1\right|\epsilon_{i-k}^2\sigma_{i-k}^2\left|\frac{1}{w_{i-k}(\mathbf{u})}-\frac{1}{\tilde{w}_{i-k}(\mathbf{u})}\right| \\
&\quad+\sum_{k<i\le n}\epsilon_{i-k}^2\left|\frac{\sigma_{i-k}^2}{w_{i-k}(\mathbf{u})}-1\right|\epsilon_i^2\sigma_i^2\left|\frac{1}{w_i(\mathbf{u})}-\frac{1}{\tilde{w}_i(\mathbf{u})}\right| \\
&= Z_{k,n}^{(1)}(\mathbf{u})+Z_{k,n}^{(2)}(\mathbf{u}).
\end{aligned} \tag{5.11}$$

By (5.5) we conclude

$$Z_{k,n}^{(1)}(\mathbf{u})\le\frac{C_1}{u_1^2}\zeta\sum_{k<i\le n}\rho_0^{(i-k)/q}y_{i-k}^2\epsilon_i^2\left|\frac{\sigma_i^2}{\tilde{w}_i(\mathbf{u})}-1\right|.$$

Using the definitions of $w_i(\mathbf{u})$ and $\tilde{w}_i(\mathbf{u})$ with (5.4) and (5.5) we get for any $\mathbf{u}\in U$

$$\begin{aligned}
\left|\frac{\sigma_i^2}{\tilde{w}_i(\mathbf{u})}-1\right| &\le 1+\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\frac{w_i(\mathbf{u})}{\tilde{w}_i(\mathbf{u})}\\
&\le 1+\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\left(1+\frac{1}{u_1}\sum_{i\le j<\infty}c_j(\mathbf{u})y_{i-j}^2\right)\\
&\le 1+\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\left(1+\frac{C_1}{u_1}\sum_{i\le j<\infty}\rho_0^{j/q}y_{i-j}^2\right)\\
&\le 1+\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\left(1+\frac{C_1}{u_1}\rho_0^{i/q}\zeta\right)\le 1+\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\left(1+\frac{C_1}{u_1}\zeta\right).
\end{aligned}$$

Hence (5.7) implies

$$\begin{aligned}
Z_{k,n}^{(1)}(\mathbf{u}) &\le \frac{C_1}{u_1^2}\zeta\sum_{k<i\le n}y_{i-k}^2\rho_0^{(i-k)/q}\epsilon_i^2\\
&\quad+\frac{C_1}{u_1^2}\zeta\left(1+\frac{C_1}{u_1}\zeta\right)\sum_{k<i\le n}y_{i-k}^2\rho_0^{(i-k)/q}\epsilon_i^2\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\\
&\le \left\{\frac{C_1}{u_1^2}\zeta\sum_{1\le j<\infty}\rho_0^{j/q}y_j^2\right\}\max_{1\le i\le n}\epsilon_i^2\\
&\quad+\left\{\frac{C_1}{u_1^2}\zeta\left(1+\frac{C_1}{u_1}\zeta\right)\sum_{1\le j<\infty}\rho_0^{j/q}y_j^2\right\}\max_{1\le i\le n}\left(\epsilon_i^2\sup_{\mathbf{u}\in U(\gamma)}\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\right)\\
&= Z_n^{(1,1)}+Z_n^{(1,2)}.
\end{aligned}$$

By (5.7) and (5.8) we have

$$\max_{0\le k\le n} Z_n^{(1,1)} = o(n^{1/(2+\delta)}) \ \ a.s.$$

Let $0<\delta'<\delta$ and choose δ^* so that $\delta'<\delta^*<\delta$. Using (4.5) we see that there are $\gamma>0$ and C_2 such that

$$E\left(\sup_{\mathbf{u}\in U(\gamma)}\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})}\right)^{2+\delta^*}\le C_2. \tag{5.12}$$

Since ϵ_i^2 and $(\sigma_i^2, \tilde{w}_i(\mathbf{u}))$ are independent, by (5.12) and the Markov inequality we have

$$P\left\{\sup_{\mathbf{u}\in U(\gamma)} \epsilon_i^2 \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})} \geq x i^{1/(2+\delta')}\right\} \leq C_2 E|\epsilon_0^2|^{2+\delta^*} x^{-(2+\delta^*)} i^{-(2+\delta^*)/(2+\delta')}$$

and therefore the Borel–Cantelli lemma implies

$$\sup_{\mathbf{u}\in U(\gamma)} \epsilon_i^2 \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{u})} = o(i^{1/(2+\delta')}) \quad a.s.$$

Thus we showed that

$$\max_{0\leq k<n} \sup_{\mathbf{u}\in U(\gamma)} Z_n^{(1,2)}(\mathbf{u}) = o(n^{1/(2+\delta')}) \quad a.s.$$

resulting in

$$\max_{0\leq k<n} \sup_{\mathbf{u}\in U(\gamma)} Z_{k,n}^{(1)}(\mathbf{u}) = o(n^{1/(2+\delta')}) \quad a.s. \tag{5.13}$$

Similar arguments yield

$$\max_{0\leq k<n} \sup_{\mathbf{u}\in U(\gamma)} Z_{k,n}^{(2)}(\mathbf{u}) = o(n^{1/(2+\delta')}) \quad a.s. \tag{5.14}$$

Lemma 5.1 now follows from (5.9)–(5.11), (5.13) and (5.14).

For any $0 < L < \infty$ let

$$U_n(L) = \{\mathbf{u} \in U : ||\boldsymbol{\theta} - \mathbf{u}|| \leq L/n^{1/2}\}.$$

Lemma 5.2 *If (2.1)–(2.6) hold, then for any $0 \leq k < \infty$ and $0 < L < \infty$ we have*

$$\sup_{\mathbf{u}\in U_n(L)} |a_k(\mathbf{u}) - (A_{k,1} + n(\boldsymbol{\theta} - \mathbf{u})\mathbf{c}_k^T)| = o(n^{1/2}) \quad a.s., \tag{5.15}$$

as $n \to \infty$.

PROOF: Using Taylor expansion for $w_i^{-1}(\mathbf{u})$ we get

$$
\begin{aligned}
A_{k,2}(\mathbf{u}) &= \sum_{k<i\le n} \epsilon_i^2(\epsilon_{i-k}^2-1)\sigma_i^2\left(\frac{1}{w_i(\mathbf{u})}-\frac{1}{w_i(\boldsymbol{\theta})}\right)\\
&= \sum_{k<i\le n} \epsilon_i^2(\epsilon_{i-k}^2-1)(\boldsymbol{\theta}-\mathbf{u})\left(\frac{w_i'(\boldsymbol{\theta})}{w_i(\boldsymbol{\theta})}\right)^T\\
&\quad -\frac{1}{2}\sum_{k<i\le n} \epsilon_i^2\sigma_i^2(\epsilon_{i-k}^2-1)(\boldsymbol{\theta}-\mathbf{u})\left(\frac{w_i''(\boldsymbol{\xi}_i)}{w_i^2(\boldsymbol{\xi}_i)}\right)(\boldsymbol{\theta}-\mathbf{u})^T \qquad (5.16)\\
&\quad +\sum_{k<i\le n} \epsilon_i^2\sigma_i^2(\epsilon_{i-k}^2-1)(\boldsymbol{\theta}-\mathbf{u})\frac{(w_i'(\boldsymbol{\xi}_i))^T w_i'(\boldsymbol{\xi}_i)}{w_i^3(\boldsymbol{\xi}_i)}(\boldsymbol{\theta}-\mathbf{u})^T\\
&= A_{k,2}^{(1)}(\mathbf{u})+A_{k,2}^{(2)}(\mathbf{u})+A_{k,2}^{(3)}(\mathbf{u}),
\end{aligned}
$$

where $\boldsymbol{\xi}_i$ satisfies $||\boldsymbol{\xi}_i-\boldsymbol{\theta}||\le||\mathbf{u}-\boldsymbol{\theta}||$. The Hölder inequality, (2.6) and (4.6) yield $E||(\epsilon_{i-k}^2-1)w_i'(\boldsymbol{\theta})/w_i(\boldsymbol{\theta})||<\infty$. So by (2.5) and the independence of ϵ_i^2 and $\epsilon_{i-k}^2 w_i'(\mathbf{u})/w_i(\mathbf{u})$ we have

$$E\epsilon_i^2(\epsilon_{i-k}^2-1)\left(\frac{w_i'(\boldsymbol{\theta})}{w_i(\boldsymbol{\theta})}\right)^T = E\epsilon_i^2 E(\epsilon_{i-k}^2-1)\left(\frac{w_i'(\boldsymbol{\theta})}{w_i(\boldsymbol{\theta})}\right)^T=\mathbf{c}_k^T.$$

For any $1\le k<\infty$, $\{\epsilon_i^2(\epsilon_{i-k}^2-1)w_i'(\boldsymbol{\theta})/w_i(\boldsymbol{\theta}),\ -\infty<i<\infty\}$ is a stationary and ergodic sequence (cf. Theorem 3.5.8 of Stout (1974)) with finite mean, so by the ergodic theorem we conclude

$$\frac{1}{n}\sum_{k<i\le n}\epsilon_i^2(\epsilon_{i-k}^2-1)\left(\frac{w_i'(\boldsymbol{\theta})}{w_i(\boldsymbol{\theta})}\right)^T\to\mathbf{c}_k^T \ \text{ a.s.}$$

Hence

$$\sup_{\mathbf{u}\in U_n(L)}\|A_{k,2}^{(1)}(\mathbf{u})-n(\boldsymbol{\theta}-\mathbf{u})\mathbf{c}_k^T\|=o(n^{1/2}) \ \text{ a.s.} \qquad (5.17)$$

Next we write

$$\sup_{\mathbf{u}\in U_n(L)}|A_{k,2}^{(2)}(\mathbf{u})| \le \frac{L^2}{n}\sum_{k<i\le n}\epsilon_i^2|\epsilon_{i-k}^2-1|\sup_{\mathbf{v}\in U_n(L)}\frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{v})}\sup_{\mathbf{u}\in U}\left\|\frac{w_i''(\mathbf{u})}{w_i(\mathbf{u})}\right\|.$$

By the Hölder inequality for all large enough n we have

$$
\begin{aligned}
& E\epsilon_i^2|\epsilon_{i-k}^2-1| \sup_{\mathbf{v}\in U_n(L)} \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{v})} \sup_{\mathbf{u}\in U} \frac{||w_i''(\mathbf{u})||}{w_i(\mathbf{u})} \\
&= E\epsilon_i^2 E|\epsilon_{i-k}^2-1| \sup_{\mathbf{v}\in U_n(L)} \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{v})} \sup_{\mathbf{u}\in U} \frac{||w_i''(\mathbf{u})||}{w_i(\mathbf{u})} \\
&\le (E|\epsilon_{i-k}^2-1|^{1+\delta})^{1/(1+\delta)} \left(E\left(\sup_{\mathbf{v}\in U_n(L)} \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{v})} \sup_{\mathbf{u}\in U} \frac{||w_i''(\mathbf{u})||}{w_i(\mathbf{u})} \right)^{(1+\delta)/\delta} \right)^{\delta/(1+\delta)} \\
&\le (E|\epsilon_{i-k}^2-1|^{1+\delta})^{1/(1+\delta)} \left(E\left(\sup_{\mathbf{v}\in U_n(L)} \frac{w_i(\boldsymbol{\theta})}{w_i(\mathbf{v})} \right)^{2(1+\delta)/\delta} \right)^{\delta/(2(1+\delta))} \\
&\qquad \times \left(E\left(\sup_{\mathbf{u}\in U} \frac{||w_i''(\mathbf{u})||}{w_i(\mathbf{u})} \right)^{2(1+\delta)/\delta} \right)^{\delta/(2(1+\delta))} < \infty,
\end{aligned}
$$

where in the last step we used the Cauchy inequality, (2.6) and Lemma 4.2. Using again the ergodic theorem we get

$$
\sup_{\mathbf{u}\in U_n(L)} |A_{k,2}^{(2)}(\mathbf{u})| = O(1) \ \text{ a.s.,} \tag{5.18}
$$

as $n\to\infty$. Similarly to (5.18) one can verify that

$$
\sup_{\mathbf{u}\in U_n(L)} |A_{k,2}^{(3)}(\mathbf{u})| = O(1) \ \text{ a.s.,} \tag{5.19}
$$

as $n\to\infty$. Putting together (5.16)–(5.19) we obtain that

$$
\sup_{\mathbf{u}\in U_n(L)} |A_{k,2}(\mathbf{u}) - n(\boldsymbol{\theta}-\mathbf{u})\mathbf{c}_k^T| = o(n^{1/2}) \ \text{ a.s.,} \tag{5.20}
$$

as $n\to\infty$.

Repeating the arguments leading to (5.16) we conclude

$$\begin{aligned}
A_{k,3}(\mathbf{u}) &= \sum_{k<i\le n} \epsilon_{i-k}^2(\epsilon_i^2-1)\sigma_{i-k}^2\left(\frac{1}{w_{i-k}(\mathbf{u})}-\frac{1}{w_{i-k}(\boldsymbol{\theta})}\right)\\
&= \sum_{k<i\le n} \epsilon_{i-k}^2(\epsilon_i^2-1)(\boldsymbol{\theta}-\mathbf{u})\left(\frac{w'_{i-k}(\boldsymbol{\theta})}{w_{i-k}(\boldsymbol{\theta})}\right)^T\\
&\quad -\frac{1}{2}\sum_{k<i\le n} \epsilon_{i-k}^2\sigma_{i-k}^2(\epsilon_i^2-1)(\boldsymbol{\theta}-\mathbf{u})\left(\frac{w''_{i-k}(\boldsymbol{\xi}_{i-k})}{w^2_{i-k}(\boldsymbol{\xi}_{i-k})}\right)(\boldsymbol{\theta}-\mathbf{u})^T\\
&\quad +\sum_{k<i\le n} \epsilon_{i-k}^2\sigma_{i-k}^2(\epsilon_i^2-1)(\boldsymbol{\theta}-\mathbf{u})\frac{(w'_{i-k}(\boldsymbol{\xi}_{i-k}))^T w'_{i-k}(\boldsymbol{\xi}_{i-k})}{w^3_{i-k}(\boldsymbol{\xi}_{i-k})}(\boldsymbol{\theta}-\mathbf{u})^T\\
&= A_{k,3}^{(1)}(\mathbf{u})+A_{k,3}^{(2)}(\mathbf{u})+A_{k,3}^{(3)}(\mathbf{u}),
\end{aligned}$$

where $\boldsymbol{\xi}_{i-k}$ satisfies $||\boldsymbol{\xi}_{i-k}-\boldsymbol{\theta}||\le||\mathbf{u}-\boldsymbol{\theta}||$. Since ϵ_i^2 and $\{\epsilon_{i-k}^2, w'_{i-k}(\boldsymbol{\theta})/w_{i-k}(\boldsymbol{\theta})\}$ are independent, by (2.5) and Lemma 4.2 we have

$$E(\epsilon_i^2-1)\epsilon_{i-k}^2\left(\frac{w'_{i-k}(\boldsymbol{\theta})}{w_{i-k}(\boldsymbol{\theta})}\right)^T = E(\epsilon_i^2-1)E\epsilon_{i-k}^2\left(\frac{w'_{i-k}(\boldsymbol{\theta})}{w_{i-k}(\boldsymbol{\theta})}\right)^T=\mathbf{0}.$$

So the ergodic theorem yields

$$\sup_{\mathbf{u}\in U_n(L)}|A_{k,3}^{(1)}(\mathbf{u})|=o(n^{1/2})\ \text{ a.s.},$$

as $n\to\infty$ for any $0<L<\infty$. Similarly to (5.18) and (5.19) we have

$$\sup_{\mathbf{u}\in U_n(L)}|A_{k,3}^{(2)}(\mathbf{u})|=O(1)\ \text{ a.s.}$$

and

$$\sup_{\mathbf{u}\in U_n(L)}|A_{k,3}^{(3)}(\mathbf{u})|=O(1)\ \text{ a.s.}$$

Thus we have established that

$$\sup_{\mathbf{u}\in U_n(L)}|A_{k,3}(\mathbf{u})|=o(n^{1/2})\ \text{ a.s.} \tag{5.21}$$

Using the mean value theorem we conclude

$$\sup_{\mathbf{u}\in U_n(L)} |A_{k,4}(\mathbf{u})| \le \frac{L^2}{n} \sum_{k<i\le n} \epsilon_i^2 \epsilon_{i-k}^2 \sup_{\mathbf{u}\in U} \frac{||w_i'(\mathbf{u})||}{w_i(\mathbf{u})} \sup_{\mathbf{u},\mathbf{v}\in U_n(L)} \frac{w_i(\mathbf{u})}{w_i(\mathbf{v})}$$

$$\times \sup_{\mathbf{u}\in U} \frac{||w_{i-k}'(\mathbf{u})||}{w_{i-k}(\mathbf{u})} \sup_{\mathbf{u},\mathbf{v}\in U_n(L)} \frac{w_{i-k}(\mathbf{u})}{w_{i-k}(\mathbf{v})},$$

so following the proof of (5.18), using Lemma 4.2 and the ergodic theorem we have

$$\sup_{\mathbf{u}\in U_n(L)} |A_{k,4}(\mathbf{u})| = O(1) \ \ a.s. \tag{5.22}$$

Lemma 5.2 now follows from (5.1) and (5.20)–(5.22).

Proof of Theorem 2.1: Putting together Lemmas 5.1 and 5.2 we get that for any $0 \le k < \infty$ and $0 < L < \infty$

$$\sup_{\mathbf{u}\in U_n(L)} |\tilde{a}_k(\mathbf{u}) - (A_{k,1} + n(\boldsymbol{\theta} - \mathbf{u})\mathbf{c}_k^T)| = o(n^{1/2}) \ \ a.s., \tag{5.23}$$

as $n \to \infty$. We note that the random vectors $(1-\epsilon_i^2)w_i'(\boldsymbol{\theta})/w_i(\boldsymbol{\theta})$, $-\infty < i < \infty$ are uncorrelated and have bounded second moments, so by the Chebisev inequality we have

$$\left|\left| \sum_{1\le i\le n} (1-\epsilon_i^2) \frac{w_i'(\boldsymbol{\theta})}{w_i(\boldsymbol{\theta})} \mathbf{B}_0^{-1} \right|\right| = O_P(n^{1/2}).$$

Thus by (4.3) we get

$$||\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}|| = O_P(n^{-1/2}), \tag{5.24}$$

and therefore Theorem 2.1 follows from (5.23).

Proof of Theorem 2.2: By Theorems 2.1 and 4.1 it is enough to prove that

$$n^{-1/2} \sum_{1\le j\le n} \boldsymbol{\eta}_j \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{D}), \tag{5.25}$$

where

$$\boldsymbol{\eta}_j = \left((\epsilon_j^2 - 1)(\epsilon_{j-i_1}^2 - 1) + \frac{1}{2}(\epsilon_j^2 - 1)\frac{w_j'(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta})} \mathbf{B}_0^{-1} \mathbf{c}_{i_1}^T, \ldots, \right.$$

$$\left. (\epsilon_j^2 - 1)(\epsilon_{j-i_K}^2 - 1) + \frac{1}{2}(\epsilon_j^2 - 1)\frac{w_j'(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta})} \mathbf{B}_0^{-1} \mathbf{c}_{i_K}^T \right).$$

We note that $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots$ are uncorrelated, stationary random vectors. It is easy to see that $E\boldsymbol{\eta}_j = \mathbf{0}$ and since

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

(cf. the proof of Remark 4.5 in Berkes, Horváth, and Kokoszka (2001)) we get $E\boldsymbol{\eta}_j^T \boldsymbol{\eta}_j = \mathbf{D}$.

Next we observe that $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots$ is a stationary, ergodic process. It is also a martingale difference sequence with respect to the σ-algebras generated by $\{\epsilon_j, -\infty < j < i\}$ with finite second moments. Using the Cramér–Wold theorem (cf. Billingsley (1968, p. 49)) and Theorem 23.1 of Billingsley (1968, p. 206) we obtain (5.25).

Proof of Theorem 2.3: We note that

$$\mathbf{D} = \operatorname{cov}(\boldsymbol{\eta}_0) = E(\epsilon_0^2 - 1)^2 E \Bigg\{ \left(\epsilon_{-i_1}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} \mathbf{z}_1^T, \ldots, \epsilon_{-i_K}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} \mathbf{z}_K^T \right)^T$$
$$\times \left(\epsilon_{-i_1}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} \mathbf{z}_1^T, \ldots, \epsilon_{-i_K}^2 - 1 + \frac{w_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})} \mathbf{z}_K^T \right) \Bigg\},$$

where $\mathbf{z}_j = \mathbf{c}_{i_j} (\mathbf{B}_0^{-1})^T / 2$. Thus (2.6) follows from Lemma 4.3.

Proof of Theorem 2.4: It is an immediate consequence of Theorem 2.2 and (2.16).

Let

$$\hat{\mathbf{A}}_n(\mathbf{u}) = \frac{1}{n} \sum_{1<i\le n} (\tilde{\ell}_i'(\mathbf{u}))^T \tilde{\ell}_i'(\mathbf{u}),$$

$$\hat{\mathbf{B}}_n(\mathbf{u}) = \frac{1}{n} \sum_{1<i\le n} \tilde{\ell}_i''(\mathbf{u}),$$

$$\hat{\mathbf{c}}_{k,n}(\mathbf{u}) = \frac{1}{n-k} \sum_{1\le i\le n-k} (\hat{\epsilon}_i^2 - 1) \frac{\tilde{w}_{i+k}'(\mathbf{u})}{\tilde{w}_{i+k}(\mathbf{u})}$$

and

$$\mathbf{A}_0(\mathbf{u}) = E(\ell_0'(\mathbf{u}))^T \ell_0'(\mathbf{u}), \ \ \mathbf{B}_0(\mathbf{u}) = E\ell_0''(\mathbf{u}), \ \ \mathbf{c}_k(\mathbf{u}) = E\left[(\epsilon_{-k}^2 - 1) \frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right].$$

Lemma 5.3 *If (2.1)–(2.6) hold, then there is* $\gamma > 0$ *such that* $\mathbf{A}_0(\mathbf{u}), \mathbf{B}_0(\mathbf{u})$ *and* $\mathbf{c}_k(\mathbf{u})$ *exist for all* $\mathbf{u} \in U(\gamma)$*,*

$$\sup_{\mathbf{u} \in U(\gamma)} |\hat{\mathbf{A}}_n(\mathbf{u}) - \mathbf{A}_0(\mathbf{u})| = o_P(1) \tag{5.26}$$

$$\sup_{\mathbf{u} \in U(\gamma)} |\hat{\mathbf{B}}_n(\mathbf{u}) - \mathbf{B}_0(\mathbf{u})| = o_P(1) \tag{5.27}$$

$$\sup_{\mathbf{u} \in U(\gamma)} |\hat{\mathbf{c}}_{k,n}(\mathbf{u}) - \mathbf{c}_k(\mathbf{u})| = o_P(1) \tag{5.28}$$

and

$$\hat{d}_n^2 \to d_0^2 \ \text{a.s.}, \tag{5.29}$$

as $n \to \infty$.

PROOF: It is easy to see that

$$\ell_0'(\mathbf{u}) = -\frac{1}{2} \left\{ \frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} - \epsilon_0^2 \frac{\sigma_0^2}{w_0(\mathbf{u})} \frac{w_0'(\mathbf{u})}{w_0(\mathbf{u})} \right\}.$$

Using the Hölder inequality, (2.6), (4.5) and (4.6) we conclude

$$\begin{aligned} E||(\ell_0'(\mathbf{u}))^T \ell_0'(\mathbf{u})|| &\le C_3 E \left(\frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} + \epsilon_0^2 \frac{\sigma_0^2}{w_0(\mathbf{u})} \frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} \right)^2 \\ &\le 4C_3 \left\{ E \left(\frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} \right)^2 + E \left(\epsilon_0^2 \frac{\sigma_0^2}{w_0(\mathbf{u})} \frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} \right)^2 \right\} \\ &< \infty \end{aligned}$$

for all $\mathbf{u} \in U(\gamma)$ assuming that $\gamma > 0$ is small enough. The existence of $\mathbf{B}_0(\mathbf{u})$ is proven in Lemma 5.6 of Berkes, Horváth and Kokoszka (2001). By the Cauchy inequality, (2.6) and (4.6) we have

$$E|\epsilon_{-k}^2 - 1| \frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} \le (E(\epsilon_{-k}^2 - 1)^2)^{1/2} \left(E \left(\frac{||w_0'(\mathbf{u})||}{w_0(\mathbf{u})} \right)^2 \right)^{1/2} < \infty,$$

establishing the existence of $\mathbf{c}_k(\mathbf{u})$ for all $\mathbf{u} \in U$.

Observing that $n\hat{\mathbf{B}}_n(\mathbf{u}) = \tilde{L}_n''(\mathbf{u})$, (5.27) is proven in Berkes, Horváth and Kokoszka (2001). The proofs of (5.26) and (5.28) are similar so we provide an outline for (5.28) only. Following the proof of Lemma 5.1 one can verify that with a small enough $\gamma > 0$

$$\sup_{\mathbf{u}\in U(\gamma)} |\hat{\mathbf{c}}_{k,n}(\mathbf{u}) - \tilde{\mathbf{c}}_{k,n}(\mathbf{u})| = o(1) \quad a.s.,$$

where

$$\tilde{\mathbf{c}}_{k,n}(\mathbf{u}) = \frac{1}{n-k} \sum_{1\le i\le n-k} (\hat{\epsilon}_i^2 - 1)\frac{w'_{i+k}(\mathbf{u})}{w_{i+k}(\mathbf{u})}.$$

Introducing

$$\mathbf{c}_{k,n}(\mathbf{u}) = \frac{1}{n-k} \sum_{1\le i\le n-k} (\epsilon_i^2 - 1)\frac{w'_{i+k}(\mathbf{u})}{w_{i+k}(\mathbf{u})}$$

we write

$$\begin{aligned}
\tilde{\mathbf{c}}_{k,n}(\mathbf{u}) - \mathbf{c}_{k,n}(\mathbf{u}) &= \frac{1}{n-k} \sum_{1\le i\le n-k} y_i^2 \left(\frac{1}{\tilde{w}_i(\tilde{\boldsymbol{\theta}}_n)} - \frac{1}{w_i(\boldsymbol{\theta})}\right) \frac{w'_{i+k}(\mathbf{u})}{w_{i+k}(\mathbf{u})} \\
&= \frac{1}{n-k} \sum_{1\le i\le n-k} y_i^2 \left(\frac{1}{\tilde{w}_i(\tilde{\boldsymbol{\theta}}_n)} - \frac{1}{w_i(\tilde{\boldsymbol{\theta}}_n)}\right) \frac{w'_{i+k}(\mathbf{u})}{w_{i+k}(\mathbf{u})} \\
&\quad + \frac{1}{n-k} \sum_{1\le i\le n-k} y_i^2 \left(\frac{1}{w_i(\tilde{\boldsymbol{\theta}}_n)} - \frac{1}{w_i(\boldsymbol{\theta})}\right) \frac{w'_{i+k}(\mathbf{u})}{w_{i+k}(\mathbf{u})} \\
&= g_{k,n}^{(1)}(\mathbf{u}) + g_{k,n}^{(2)}(\mathbf{u}).
\end{aligned}$$

Applying (5.5) there is a random variable ζ^* such that

$$\sup_{\mathbf{u}\in U} ||g_{k,n}^{(1)}(\mathbf{u})|| \le \zeta^* \frac{1}{n-k} \sum_{1\le i\le n-k} y_i^2 \sup_{\mathbf{u}\in U} \frac{||w'_{i+k}(\mathbf{u})||}{w_{i+k}(\mathbf{u})} \rho_0^{i/q} = O\left(\frac{1}{n}\right) \quad a.s.$$

by (4.6) and Lemmas 4.1 and 4.4. Using the mean value theorem we get with a small enough γ that

$$\begin{aligned}
&\sup_{\mathbf{u}\in U(\gamma)} ||g_{k,n}^{(2)}(\mathbf{u})|| \\
&\quad \le C_4 ||\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}|| \frac{1}{n-k} \sum_{1\le i\le n-k} \epsilon_i^2 \sup_{\mathbf{u},\mathbf{v}\in U(\gamma)} \frac{w_i(\mathbf{u})}{w_i(\mathbf{v})} \sup_{\mathbf{u}\in U} \frac{||w'_i(\mathbf{u})||}{w_i(\mathbf{u})} \sup_{\mathbf{u}\in U} \frac{||w'_{i+k}(\mathbf{u})||}{w_{i+k}(\mathbf{u})}.
\end{aligned}$$

By (2.6) and Lemma 4.2 we obtain

$$E\epsilon_0^2 \sup_{\mathbf{u},\mathbf{v}\in U(\gamma)} \frac{w_0(\mathbf{u})}{w_0(\mathbf{v})} \sup_{\mathbf{u}\in U} \frac{||w'_0(\mathbf{u})||}{w_0(\mathbf{u})} \sup_{\mathbf{u}\in U} \frac{||w'_k(\mathbf{u})||}{w_k(\mathbf{u})} < \infty$$

and therefore by the ergodic theorem we have

$$\frac{1}{n-k}\sum_{1\le i\le n-k}\epsilon_i^2 \sup_{\mathbf{u},\mathbf{v}\in U(\gamma)}\frac{w_i(\mathbf{u})}{w_i(\mathbf{v})}\sup_{\mathbf{u}\in U}\frac{||w_i'(\mathbf{u})||}{w_i(\mathbf{u})}\sup_{\mathbf{u}\in U}\frac{||w_{i+k}'(\mathbf{u})||}{w_{i+k}(\mathbf{u})}=O(1)\ \ a.s.$$

So (5.24) yields

$$\sup_{\mathbf{u}\in U(\gamma)}||g_{k,n}^{(2)}(\mathbf{u})||=o_P(1)$$

with some $\gamma>0$. The ergodic theorem implies that

$$\mathbf{c}_{k,n}(\mathbf{u})\ \stackrel{a.s.}{\rightarrow}\ \mathbf{c}_k(\mathbf{u})\ \text{ for any }\ \mathbf{u}\in U \tag{5.30}$$

and

$$\sup_{\mathbf{u}\in U(\gamma)}||\mathbf{c}_{k,n}'(\mathbf{u})||=O(1)\ \ a.s. \tag{5.31}$$

Putting together (5.30) and (5.31) we conclude by standard arguments

$$\sup_{\mathbf{u}\in U(\gamma)}||\mathbf{c}_{k,n}(\mathbf{u})-\mathbf{c}_k(\mathbf{u})||=o(1)\ \ a.s.$$

completing the proof of (5.28).

Since by the independence of ϵ_0 and $w_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta})$ we have that $\mathbf{c}_0=\mathbf{0}$, (5.29) follows from Theorem 2.1 and the strong law of large numbers.

Proof of Theorem 2.5: Lemma 5.3 and (5.24) imply

$$\hat{\mathbf{D}}_n^{-1}\ \stackrel{P}{\rightarrow}\ \mathbf{D}^{-1}$$

and therefore the result follows from the central limit theorem in Theorem 2.4.

Acknowledgements. We would like to thank the two referees for their comments and Professor Bruce E. Hansen, the co–editor in charge, for his sound advice on how to improve the paper.

References

Berkes, I. and Horváth, L. (2001). The efficiency of the estimators of the parameters in GARCH processes. Preprint.

Berkes, I. and Horváth, L. (2002) Limit results for the empirical process of squared residuals in GARCH models. Preprint.

Berkes, I., Horváth, L. and Kokoszka, P. (2001). GARCH processes: structure and estimation. *Bernoulli* To appear.

Bierens, H. J. and Ploberger, W. (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica.* **65**, 1129–1151.

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

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

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

Box, G. E. P, Jenkins, G. M. and Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control.* Prentice Hall, Englewood Cliffs (Third Edition).

de Jong, R.M. (1996) The Bierens test under data dependence. *Journal of Econometrics.* **72**, 1–32.

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

Horváth, L., Kokoszka, P. (2001). Large sample distribution of weighted sums of ARCH(p) square residual correlations. *Econometric Theory.* **17** 283–295.

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

Horváth, L., Kokoszka, P. and Teyssiére, G. (2002) Bootstrap specification testsfor ARCH based on the empirical process of the squared residuals. Preprint.

Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica.* **64**, 837–864.

Hull, J.C. (2000). *Options, Futures & other Derivatives.* Prentice Hall, Upper Saddle River, NJ.

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

Li, W. K. (1992). On the asymptotic distribution of residual autocorrelations in nonlinear time series modelling. *Biometrika.* **79**, 435–437.

Li, W. K. and Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis.* **15**, 627–636.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving–average time series models with conditional heteroscedasticity. *J. Amer. Stat. Assoc.* **92**, 1184–1194.

Ljung G. and Box G. (1978) On a measure of lack of fit in time series models. *Biometrika.* **66**, 67–72.

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

McNeill, A. J. and Frey, R. (2000). Estimation of tail–related risk measures for heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance.* **7**, 271–300.

Nelson, D. B. (1990). Stationarity and persistence in GARCH(1, 1) model. *Econometric Theory.* **6**, 318-334.

Stout, W. F. (1974). *Almost Sure Convergence.* Academic Press, New York.