Estimation of the maximal moment exponent of a GARCH(1, 1) sequence

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 propose an estimator for the maximal moment exponent of a GARCH(1, 1) sequence. We establish its consistency asymptotic normality with rate $n^{-1/2}$. Finite sample properties are investigated by means of a small simulation study.

1 Introduction

Suppose $\{y_k\}$ is a strictly stationary stochastic process such that for some $\nu_0 > 0$

$$Ey_k^{2\nu} < \infty \text{ for } \nu < \nu_0 \quad \text{and} \quad Ey_k^{2\nu} = \infty \text{ for } \nu > \nu_0. \tag{1.1}$$

We refer to $2\nu_0$ as the maximal moment exponent. For many time series models

$$P(y_k > x) \sim x^{-2\nu_0} L(x), \qquad \text{as } x \to \infty \tag{1.2}$$

for some slowly varying function L, and in this context $2\nu_0$ is called the tail index. The factor 2 in these definitions is convenient because for the GARCH(1,1) process considered in this paper ν_0 is the maximal moment exponent of the conditional variance process.

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

The knowledge of ν_0 is essential in many applications, see e.g. Chapter 6 of Embrechts *et al.* (1997) and references therein and Section 1 of Feuerverger and Hall (1999). In the context of actuarial data, the index ν_0 is related to several measures of risk all of which utilize information contained in the tails of distributions, see e.g. Dempster (2002) for a collection of articles studying such measures. It is also often useful to be able to estimate the value of ν_0 when analyzing returns on speculative assets. Quintos *et al.* (2001) use such estimates to construct tests for structural changes in stock prices. More generally, it is important to know whether $2\nu_0$ exceeds the values 1 or 2 to be able to use, respectively, the law of large numbers and the central limit theorem, and whether it exceeds 4 or 8 to be able to apply properly the asymptotic theory for sample autocovariances developed by Mikosch and Stărică (2000).

There is a large body of research on estimating ν_0, but most of it relies on the assumption that the y_k are independent and identically distributed. In many appliacations, most notably in those related to financial time series, such an assumption is unrealistic and methods robust to some dependence in the data are required, we review some recent developments in this area in the following. This paper focuses on estimating the maximal moment exponent of a GARCH(1,1) process which offers a parsimonious specification often used in practice, see e.g. Chapter 15 of Hull (2000). McNeil and Frey (2000) incorporate estimation of a GARCH(1,1) model in the evaluation of several risk measures for asset returns.

Before presenting our estimator, we briefly review some recent work on estimating ν_0 in the context of dependent observations. Most methods postulate (1.2) and are based on the upper order statistics with Hill's estimator being the most popular and most extensively studied. These methods are are described in Chapter 6 of Embrechts *et al.* (1997). The idea is to construct an estimator based on $k = k(n)$ upper order statistics, where $k \to \infty$ but $k/n \to 0$. For iid data these estimators are asymptotically normal but converge at the rate $k^{-1/2}$ which can be arbitrarily slow. The conditions for the asymptotic normality involve the higher order behavior of the underlying tail distribution. Hill's estimator is consistent for many classes of weakly dependent processes, see Hsing (1991), Rootzén *et al.* (1992) and Resnick and Stărică (1995, 1998). Hsing (1991) showed that the Hill's estimator is asymptotically normal for certain weakly dependent processes which need not be stationary. Resnick and Stărică (1997) showed that if Hill's estimator is based on residuals of an autoregressive process than its asymptotic variance is the same as in the corresponding iid case. Goldie and Smith (1987) proposed the *ratio estimator* which is an extension of Hill's estimator. Novak (2000) showed that this estimator is consistent for sationary ϕ-mixing sequences. As mentioned above, Quintos *et al.* (2001) used a form of Hill's estimator which is robust to weak dependence. Leadbetter (1991) studied a POT (peaks over threshhold) estimator in the context of dependent observations.

The approach proposed in this paper originated in the problem of estimating the adjustment coefficient in a model for claims arriving at an insurance company. This problem was considered by Pitts *et al.* (1996) whose goal was to estimate the solution R to the equation

$$M(r) = Ee^{rX} = 1$$

from a sample $X_1, \ldots, X_n$. The estimator is defined as the solution $\hat{R}_n$ to the empirical conterpart of the above equation:

$$\hat{M}(r) = \frac{1}{n} \sum_{i=1}^{n} e^{rX_i} = 1.$$

Pitts *et al.* (1996) showed that if the X_k are iid then

$$\sqrt{n}\left(\hat{R}_n - R\right) \to N(0, \sigma^2), \tag{1.3}$$

where

$$\sigma^2 = (M(2R) - 1)(M'(R))^{-2}. \tag{1.4}$$

Mathematical ideas related to this problem were earlier explored by Csörgő (1982) and Csörgő and Teugels (1990).

Mikosch and Stărică (2000) used the result of Pitts *et al.* (1996) with

$$X_i = \hat{\beta} + \hat{\alpha}\hat{\epsilon}_i^2 \tag{1.5}$$

to estimate the maximal index exponent of a GARCH(1,1) process defined by equations (2.1), (2.2) below. In (1.5), $\hat{\beta}, \hat{\alpha}$ and $\hat{\epsilon}_i$ are, respectively, maximum likelihood estimators and the sequence of residuals. Clearly the X_i in (1.5) are not independent and the result of Pitts *et al.* (1996) does not apply. In particular it may be misleading to use the asymptotic variance (1.4) to construct confidence intervals. In the present paper we define precisely the estimator arising from the above procedure and study its asymptotic properties. We note that the proof of Pitts *et al.* (1996) is based on a functional central limit theorem and the delta method. We use a different approach utilizing the ergodic theorem and the central limit theorem for martingales.

The paper is organized as follows. In Section 2, we present the definitions and the asymptotic theory. Section 3 describes the results of a small simulation study. Proofs are collected in Section 4.

2 Asymptotic theory

We consider the GARCH$(1,1)$ sequence $\{y_n, \ -\infty < n < \infty\}$ defined by the equations

$$y_k = \sigma_k \epsilon_k, \tag{2.1}$$

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

where $\boldsymbol{\theta} = (\omega, \alpha, \beta)$ is the parameter of the process, $\omega > 0, \alpha \geq 0, \beta \geq 0$ and

$$\{\epsilon_k, \ -\infty < k < \infty\} \text{ are independent, identically distributed random variables.} \tag{2.3}$$

Nelson (1990) showed that (2.1) and (2.2) have a unique stationary solution if and only if

$$E\log(\beta+\alpha\epsilon_0^2)<0. \tag{2.4}$$

The goal of this paper is to find a $n^{-1/2}$ consistent estimator for the number of moments of y_0.

Let

$$Z(t)=E(\beta+\alpha\epsilon_0^2)^t.$$

We are interested in the estimation of ν_0 satisfying

$$Z(\nu_0)=1 \ \text{ and } \ Z(t)<1 \ \text{ for all } \ 0<t<\nu_0. \tag{2.5}$$

because then by the Corollary on p. 322 of Nelson (1990) we have

$$E\sigma_0^{2\nu}<\infty \quad \text{if} \quad \nu<\nu_0 \quad \text{and} \quad E\sigma_0^{2\nu}=\infty \quad \text{if} \quad \nu>\nu_0. \tag{2.6}$$

Clearly, if $E\epsilon_0^{2\nu_0}<\infty$, then (2.6) implies (1.1).

If $Z_n(t)$ is an estimator for $Z(t)$, then the natural estimator for ν_0 is ν_n, the smallest positive random variable satisfying

$$Z_n(\nu_n)=1.$$

The random variables $\{\epsilon_i^2,\ -\infty<i<\infty\}$ cannot be observed and must be estimated from the sample. Using the recursion in (2.2) one can easily verify (cf. also Lee and Hansen (1994) and Lumsdaine (1996)) that

$$\sigma_k^2=\omega\frac{1-\beta^k}{1-\beta}+\alpha\sum_{0\le i\le k-1}\beta^i y_{k-i-1}^2+\beta^k\sigma_0^2. \tag{2.7}$$

Let $\hat{\boldsymbol{\theta}}_n$ be an estimator for $\boldsymbol{\theta}$. We estimate σ_k^2 with $\hat{w}_k(\hat{\boldsymbol{\theta}})$, where

$$\hat{w}_k(\mathbf{u})=x\frac{1-t^k}{1-t}+s\sum_{0\le i\le k-1}t^i y_{k-i-1}^2, \ \ \mathbf{u}=(x,s,t).$$

In light of (2.1),

$$\hat{\epsilon}_k^2=y_k^2/\hat{w}_k(\hat{\boldsymbol{\theta}}_n), \ \ 1\le k\le n,$$

can be used as an estimator for ϵ_k^2. We show that

$$Z_n(t)=\frac{1}{n}\sum_{1\le k\le n}(\hat{\beta}_n+\hat{\alpha}_n\hat{\epsilon}_k^2)^t$$

is a uniformly, weakly consistent estimator for $Z(t)$, where $\hat{\boldsymbol{\theta}}_n=(\hat{\omega}_n,\hat{\alpha}_n,\hat{\beta}_n)$. We assume that $\hat{\boldsymbol{\theta}}_n$ is weakly consistent with order $n^{-\gamma}$, i.e.

$$||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||=O_P(n^{-\gamma}) \ \text{ with some } \ 0<\gamma, \tag{2.8}$$

where $||\cdot||$ denotes the maximum norm of vectors. We note that (2.8) holds with $\gamma=1/2$ in case of quasi (pseudo) –likelihood estimators (cf. Lumsdaine (1996), Lee and Hansen (1994) and Berkes *et al.* (2002)) and in case of likelihood estimators under weak assumptions on the distribution of the innovations and model parameters, cf. Berkes and Horváth (2001).

THEOREM 2.1 *If (2.1)–(2.4), (2.8) hold and* $0 < \nu_0 < \infty$,

$$\omega > 0, \quad \alpha > 0, \quad 0 < \beta < 1, \tag{2.9}$$

and

$$E|\epsilon_0^2|^\delta < \infty \quad \textit{with some} \quad \delta > \max(\nu_0, 1/\gamma), \tag{2.10}$$

then there is a unique ν_0 *satisfying (2.5) and*

$$\nu_n \xrightarrow{P} \nu_0.$$

The main result of this paper is the asymptotic normality of $n^{1/2}(\nu_n - \nu_0)$, which will require stronger conditions than (2.8). We postulate a representation of the form

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \frac{1}{n} \sum_{1 \le k \le n} (\xi_1(k), \xi_2(k), \xi_3(k)) + o_P(n^{-1/2}), \tag{2.11}$$

where

$$\xi_i(k) = \ell_i(\epsilon_k^2) h_j(y_{k-1}, y_{k-2}, \ldots), \qquad i = 1, 2, 3,$$

with some functions ℓ_i and h_i satisfying

$$E\ell_i(\epsilon_0^2) = 0, E\ell_i^2(\epsilon_0^2) < \infty \;\text{ and }\; Eh_i^2(y_{-1}, y_{-2}, \ldots) < \infty, \quad i = 1, 2, 3. \tag{2.12}$$

Representation (2.11) was established by Horváth *et al.* (2001) in the context of ARCH(p) processes for several classes of estimators with the functions ℓ_i and h_i explicitely identified. It was extended to GARCH models in Berkes *et al.* (2002) and Berkes and Horváth (2001).

We now define the quantities needed to define the asymptotic variance. Let

$$q_1^*(t) = \frac{\alpha}{1-\beta} p_1 q_2(t), \qquad q_2^*(t) = -q_1(t) + \frac{\omega\alpha}{(1-\beta)^2} p_1 q_2(t) + \alpha^2 p_2 q_2(t)$$

and

$$q_3^*(t) = -\frac{\omega}{1-\beta} p_1 q_2(t),$$

where

$$q_1(t) = tE(\beta + \alpha\epsilon_0^2)^{t-1}, \qquad q_2(t) = tE\epsilon_0^2(\beta + \alpha\epsilon_0^2)^{t-1}$$

and

$$p_1 = E\frac{1}{\sigma_0^2}, \qquad p_2 = E\frac{1}{\sigma_0^2} \sum_{0 \le i < \infty} i\beta^{i-1} y_{-i-1}^2.$$

We note that condition (2.9) and (2.2) imply that $p_1 < \infty$, while the existence of p_2 will follow from Lemma 4.2.

The asymptotic variance of $n^{1/2}(\nu_n - \nu_0)$ is given by

$$d_0^2 = \{E(\beta + \alpha\epsilon_0^2)^{\nu_0} \log(\beta + \alpha\epsilon_0^2)\}^{-2} \Big\{ \sum_{1 \le i,\, j \le 3} q_i^*(\nu_0)\bar{h}_{ij}\bar{\ell}_{ij} q_j^*(\nu_0)$$

$$+ 2 \sum_{1 \le i \le 3} q_i^*(\nu_0)\bar{h}_i E\{\ell_i(\epsilon_0^2)(\beta + \alpha\epsilon_0^2)^{\nu_0}\} + Z(2\nu_0) - Z^2(\nu_0)\Big\},$$

where

$$\bar{h}_i = Eh_i(y_0, y_{-1}, \ldots), \qquad \bar{h}_{ij} = Eh_i(y_0, y_{-1}, \ldots)h_j(y_0, y_{-1}, \ldots)$$

and

$$\bar{\ell}_{ij} = E\ell_i(\epsilon_0^2)\ell_j(\epsilon_0^2) \quad (i, j = 1, 2, 3).$$

THEOREM 2.2 *We assume that (2.1)–(2.4), (2.9), (2.11), (2.12) hold,* $0 < \nu_0 < \infty$ *and*

$$E|\epsilon_0^2|^\delta < \infty \quad \textit{with some} \quad \delta > \max(\nu_0, 2). \tag{2.13}$$

Then

$$n^{1/2}(\nu_n - \nu_0) \xrightarrow{\mathcal{D}} N(0, d_0^2),$$

where $N(0, d_0^2)$ *stands for a normal random variable with mean* 0 *and variance* d_0^2.

3 Small sample properties

In this section we describe the results of a small simulation study which we conducted to assess the performance of the estimator in finite samples. We worked with time series of length $n = 1000$ which may be considered small for applications to financial data; Mikosch and Stărică (2000) considered a sample of 70 thousand observations and samples of this length are not unusual when analyzing intra-daily trading of speculative assets. The goal of our study was merely to gain some insight into the behavior of the estimator and we do not view the results presented in this section as a guide for practitioners, for this a more extensive study focusing on a specific application at hand would be required.

We considered several GARCH(1,1) models with standard normal innovations ϵ_k. Roughly speaking, the smaller the parameters α and β in (2.2), and consequently the higher the maximal moment exponent, the worse the performance of the estimator becomes. This is intuitively clear because if the parameters α and β are small, then the observations y_k look like the noise sequence ϵ_k which has finite moments of any order. This results in long right tails of the sampling distribution of the moment estimator. In the upper panels of Figures 1 and 2 we plotted the estimated sampling densities of the estimator ν_n. We also studied the effect of parameter estimation on the moment estimator. The lower panels of Figures 1 and 2 show the estimated sampling densities of ν_n when the estimators $\hat{\beta}_n$ and $\hat{\alpha}_n$ in the definition of $Z_n(t)$ are replaced by the true parameter values β and α. (We used the residuals as Mikosch and Stărică (2000), not the innovations.) It is seen that the variability of the estimators $\hat{\beta}_n$ and $\hat{\alpha}_n$ has a significant impact on the estimator ν_n and should be taken into account in empirical studies. This variability is described by (2.11) and the definitions leading to Theorem 2.2.

The true value of ν cannot be found analytically even when the values of α and β are known. The "true values" reported in the captions in Figures 1 and 2. were obtained by approximating $Z(t)$ by the mean of one million values of $(\beta + \alpha\epsilon_i^2)^t$ for iid standard normal ϵ_i.

We also investigated the effect of model misspecification on the sampling distribution of the estimator. As can be expected, the estimator is biased in such situations and the bias is probably not asymptotically negligible. In samples of length $n = 1000$ it appears however that for time series with heavy-tailed marginal distributions such that the dependence between the observations decays exponentially, the extremal behavior of the series can to some extend be approximated by fitting a GARCH(1,1) model. Figure 3 shows the sampling density of ν_n when the true DGP (data generating process) is AR(1) with stable innovations. It is well-known, see e.g. Chapter 7 of Embrechts *et al.* (1997), that for linear processes the extremal moment index of the observations is equal to the index of stability of the innovations, so in this case we know the true value of ν.

4 Proof of Theorems 1.1 and 1.2

The following lemma will be used in the proofs in Theorems 2.1 and 2.2.

LEMMA 4.1 *Let $0 < a < b < 1$, $d > 0$ and define*

$$\xi_k = \sum_{1 \le i \le k} b^i y_{k-i}^2 \bigg/ \left(d + \sum_{1 \le i \le k} a^i y_{k-i}^2 \right)$$

and

$$\xi_k' = \sum_{1 \le i < \infty} b^i y_{k-i}^2 \bigg/ \left(d + \sum_{1 \le i < \infty} a^i y_{k-i}^2 \right).$$

If

$$E|y_0|^{2\nu} < \infty, \tag{4.1}$$

then

$$P\{\xi_k \ge t\} \le Ct^{-\kappa} \tag{4.2}$$

and

$$P\{\xi_k' \ge t\} \le Ct^{-\kappa} \tag{4.3}$$

for all $t \ge 1$, where $\kappa = \nu - \nu(\log(bb_*))/\log(b/a)$ with any $1 < b_* < 1/b$ and C does not depend on k.

PROOF: For any $N \geq 1$ we have ($x \wedge y = \min(x,y)$)

$$\xi_k = \frac{\sum_{1\leq i\leq N\wedge k} b^i y_{k-i}^2}{d + \sum_{1\leq i\leq k} a^i y_{k-i}^2} + \frac{\sum_{N\wedge k<i\leq k} b^i y_{k-i}^2}{d + \sum_{1\leq i\leq k} a^i y_{k-i}^2} \leq \frac{\sum_{1\leq i\leq N\wedge k} b^i y_{k-i}^2}{\sum_{1\leq i\leq N\wedge k} a^i y_{k-i}^2} + \frac{1}{d}\sum_{N+1\leq i<\infty} b^i y_{k-i}^2.$$

By (4.1) and the Markov inequality we have

$$\begin{aligned} P\left\{\sum_{N+1\leq i<\infty} b^i y_{k-i}^2 \geq t\right\} &\leq \sum_{N+1\leq i<\infty} P\left\{y_{k-i}^2 \geq tb^{-i}\left(\frac{b_*}{b_*-1}\right)^{-1} b_*^{-i}\right\} \\ &\leq \sum_{N+1\leq i<\infty} P\left\{|y_0|^{2\nu} \geq t^{\nu}(bb_*)^{-i\nu}\left(\frac{b_*}{b_*-1}\right)^{-\nu}\right\} \\ &\leq E|y_0|^{2\nu} t^{-\nu}\left(\frac{b_*}{b_*-1}\right)^{\nu} \sum_{N+1\leq i<\infty}(bb_*)^{i\nu} \\ &\leq \frac{E|y_0|^{2\nu}}{1-(bb_*)^{\nu}}\left(\frac{b_*}{b_*-1}\right)^{\nu} t^{-\nu}(bb_*)^{N\nu}. \end{aligned} \tag{4.4}$$

It is easy to see that

$$\frac{\sum_{1\leq i\leq N\wedge k} b^i y_{k-i}^2}{\sum_{1\leq i\leq N\wedge k} a^i y_{k-i}^2} \leq \left(\frac{b}{a}\right)^{N\wedge k} \leq \left(\frac{b}{a}\right)^{N}. \tag{4.5}$$

Assume that $t > 1$ and let $N = c_1 \log t$ with $c_1 = 1/\log(b/a)$. Now (4.4) yields

$$P\left\{\sum_{N+1\leq i<\infty} b^i y_{k-i}^2 \geq t\right\} \leq c_2 t^{-\nu}((bb_*)^{\log t})^{\nu/\log(b/a)} = c_2 t^{-\kappa}$$

and by (4.5) we have

$$\frac{\sum_{1\leq i\leq N\wedge k} b^i y_{k-i}^2}{\sum_{1\leq i\leq N\wedge k} a^i y_{k-i}^2} \leq t,$$

completing the proof of (4.2).The proof of (4.3) is similar and hence is omitted.

The following result is an immediate consequence of Lemma 4.1.

LEMMA 4.2 *Let $0 < a < 1, d > 0$ and define*

$$\hat{\xi}_k = \sum_{1\leq i\leq k} i a^i y_{k-i}^2 \Big/ \left(d + \sum_{1\leq i\leq k} a^i y_{k-i}^2\right)$$

and

$$\tilde{\xi}_k = \sum_{1\le i<\infty} i a^i y_{k-i}^2 \Big/ \left(d + \sum_{1\le i<\infty} a^i y_{k-i}^2 \right).$$

If (4.1) holds with some $\nu > 0$, then

$$E\hat{\xi}_k^{\mu} \le C_1(\mu) \tag{4.6}$$

and

$$E\tilde{\xi}_k^{\mu} \le C_2(\mu) \tag{4.7}$$

for all $k \ge 1$ and $\mu \ge 0$, where $C_1(\mu)$ and $C_2(\mu)$ do not depend on k.

Proof: Let μ be given and choose $b > a$ so near to a that $\kappa > \mu + 2$. Let R be the smallest integer such that $ia^i \le b^i$ for $R \le i < \infty$. With this choice of b and R we have that

$$\frac{\sum\limits_{k\wedge R\le i\le k} i a^i y_{k-i}^2}{d + \sum\limits_{1\le i\le k} a^i y_{k-i}^2} \le \xi_k$$

and by Lemma 4.1 we have that $E\xi_k^{\mu} < \infty$. On the other hand,

$$\frac{\sum\limits_{1\le i\le k\wedge R} i a^i y_{k-i}^2}{d + \sum\limits_{1\le i\le k} a^i y_{k-i}^2} \le \frac{\sum\limits_{1\le i\le k\wedge R} i a^i y_{k-i}^2}{\sum\limits_{1\le i\le k} a^i y_{k-i}^2} \le k \wedge R \le R,$$

completing the proof of (4.6). Similar arguments give (4.7).

The proofs of Theorems 1.1 and 1.2 will use the following decomposition:

$$\begin{aligned} \hat{\beta}_n + \hat{\alpha}_n \hat{\epsilon}_k^2 - (\beta + \alpha\epsilon_k^2) &= \hat{\beta}_n - \beta + (\hat{\alpha}_n - \alpha)\epsilon_k^2 + \alpha\epsilon_k^2\sigma_k^{-2}(\sigma_k^2 - \hat{w}_k(\hat{\boldsymbol{\theta}}_n)) \\ &+ (\hat{\alpha}_n - \alpha)\epsilon_k^2(\sigma_k^2 - \hat{w}_k(\hat{\boldsymbol{\theta}}_n))/\hat{w}_k(\hat{\boldsymbol{\theta}}_n) + \alpha\epsilon_k^2(\sigma_k^2 - \hat{w}_k(\hat{\boldsymbol{\theta}}_n))^2/(\sigma_k^2\hat{w}_k(\hat{\boldsymbol{\theta}}_n)). \end{aligned} \tag{4.8}$$

Using the definition of $\hat{\sigma}_k^2(\hat{\boldsymbol{\theta}}_n)$ we write

$$\sigma_k^2 - \hat{w}_k(\hat{\theta}_n) = A_{1,k} + \cdots + A_{8,k}, \tag{4.9}$$

where

$$A_{1,k} = \frac{1}{1-\beta}(\omega - \hat{\omega}_n), \qquad A_{2,k} = \frac{\omega}{(1-\beta)^2}(\beta - \hat{\beta}_n),$$

$$A_{3,k} = (\alpha - \hat{\alpha}_n)\sum_{0\le i\le k-1} \beta^i y_{k-i-1}^2, \qquad A_{4,k} = \alpha \sum_{1\le i\le k-1} (\beta^i - \hat{\beta}_n^i) y_{k-i-1}^2,$$

$$A_{5,k} = (\beta - \hat{\beta}_n)(\hat{\omega}_n - \omega)\frac{1}{(1-\beta)(1-\hat{\beta}_n)},$$

$$A_{6,k} = -(\beta - \hat{\beta}_n)^2 \frac{\beta}{(1-\beta)^2(1-\hat{\beta}_n)}, \qquad A_{7,k} = (\hat{\alpha}_n - \alpha)\sum_{1\le i\le k-1} (\beta^i - \hat{\beta}_n^i) y_{k-i-1}^2,$$

and

$$A_{8,k} = \left(\sigma_0^2 - \frac{\omega}{1-\beta}\right)\beta^k.$$

Let $\underline{\omega}, \bar{\omega}, \underline{\alpha}, \bar{\alpha}, \underline{\beta}$ and $\bar{\beta}$ be arbitrary constants satisfying

$$0 < \underline{\omega} < \omega < \bar{\omega}, \quad 0 < \underline{\alpha} < \alpha < \bar{\alpha}, \quad 0 < \underline{\beta} < \beta < \bar{\beta} < 1$$

and define the event B_n on which $\underline{\omega} \le \hat{\omega}_n \le \bar{\omega}, \underline{\alpha} \le \hat{\alpha}_n \le \bar{\alpha}, \underline{\beta} \le \hat{\beta}_n \le \bar{\beta}$. Condition (2.8) yields that

$$\lim_{n\to\infty} P\{B_n\} = 1. \tag{4.10}$$

We continue with a technical lemma.

Lemma 4.3 *If (2.1)–(2.4), (2.8), (2.9) hold and*

$$E|\epsilon_0^2|^\delta < \infty \quad \textit{with some} \quad \delta > 1/\gamma, \tag{4.11}$$

then

$$\max_{1\le k\le n} \epsilon_k^2 A_{i,k}/\hat{w}_k(\hat{\boldsymbol{\theta}}_n) = o_P(1), \quad \textit{if} \quad i = 1, 2, \ldots, 7 \tag{4.12}$$

$$\max_{1\le k\le n} \{\epsilon_k^2 A_{8,k}/\hat{w}_k(\hat{\boldsymbol{\theta}}_n)\}/(k^{1/\delta}\bar{\beta}^k) = O_P(1), \tag{4.13}$$

$$\max_{1\le k\le n} \epsilon_k^2 A_{i,k}/\sigma_k^2 = o_P(1), \quad \textit{if} \quad i = 1, 2, \ldots, 7 \tag{4.14}$$

and

$$\max_{1\le k\le n} \{\epsilon_k^2 A_{8,k}/\sigma_k^2\}/(k^{1/\delta}\bar{\beta}^k) = O_P(1). \tag{4.15}$$

Proof: By (4.10) we can assume that B_n holds. On the event B_n

$$\hat{w}_k(\hat{\boldsymbol{\theta}}_n) \ge \underline{\omega}. \tag{4.16}$$

So using (2.8) we get

$$\max_{1\le k\le n} A_{i,k}/\hat{w}_k(\hat{\boldsymbol{\theta}}_n) = O_P(n^{-\gamma}), \quad \text{if} \quad i = 1, 2. \tag{4.17}$$

and

$$\max_{1\le k\le n} A_{i,k}/\hat{w}_k(\hat{\boldsymbol{\theta}}_n) = O_P(n^{-2\gamma}), \quad \text{if} \quad i = 5, 6. \tag{4.18}$$

Condition (4.11) yields

$$\epsilon_k^2 = o(k^{1/\delta}) \quad \text{a.s.}\ (k\to\infty), \tag{4.19}$$

and therefore (4.17), (4.18) and $\delta > 1/\gamma$ imply (4.12) when $i = 1, 2, 5$ and 6. On the event B_n we have

$$\frac{1}{\hat{w}_k(\hat{\boldsymbol{\theta}}_n)} \sum_{0\le i\le k-1} \beta^i y_{k-i-1}^2 \le \left\{\sum_{0\le i\le k-1} \beta^i y_{k-i-1}^2\right\} \Big/ \left\{\underline{\omega} + \underline{\alpha} \sum_{0\le i\le k-1} \underline{\beta}^i y_{k-i-1}^2\right\}.$$

By Lemma 2.3 in Berkes *et al.* (2002), there is ν^* such that $E|y_0|^{\nu^*} < \infty$. We can choose any $\underline{\beta} < \beta$ to have (4.10) so we can choose $\underline{\beta} < \beta$ so close to β such that

$$(\nu^*/2)(1 - \log \beta / \log(\beta/\underline{\beta})) \geq 4/(\gamma - 1/\delta)).$$

Hence by Lemma 4.1 we have

$$P\Big\{ \max_{1 \leq k \leq n} \frac{1}{\hat{w}_k(\hat{\boldsymbol{\theta}}_n)} \sum_{0 \leq i \leq k-1} \beta^i y_{k-i-1}^2 \geq n^{(\gamma - 1/\delta)/2} \Big\}$$

$$\leq \sum_{1 \leq k \leq n} P\Big\{ \sum_{0 \leq i \leq k-1} \beta^i y_{k-i-1}^2 \Big/ \Big(\underline{\omega} + \underline{\alpha} \sum_{0 \leq i \leq k-1} \underline{\beta}^i y_{k-i-1}^2 \Big) \geq n^{(\gamma - 1/\delta)/2} \Big\} \leq C/n. \qquad (4.20)$$

Using again (2.8), (4.19) and (4.20) we conclude

$$\max_{1 \leq k \leq n} \epsilon_k^2 A_{3,k} / \hat{w}_k(\hat{\boldsymbol{\theta}}_n) = o_P(n^{1/\delta}) O_P(n^{-\gamma}) O_P(n^{(\gamma - 1/\delta)/2}) = o_P(1). \qquad (4.21)$$

Using the mean-value theorem on the event B_n we have

$$\left| \sum_{1 \leq i \leq k-1} (\beta^i - \hat{\beta}_n^i) y_{k-i-1}^2 \right| \leq |\beta - \hat{\beta}_n| \sum_{1 \leq i \leq k-1} i \bar{\beta}^{i-1} y_{k-i-1}^2,$$

so following the proof of (4.21) one can easily verify that

$$\max_{1 \leq k \leq n} \epsilon_k^2 A_{4,k} / \hat{w}_k(\hat{\boldsymbol{\theta}}_n) = o_P(1).$$

The estimates in (4.13) are immediate consequences of (4.10), (4.16) and (4.19).

Following the proofs of (4.12) and (4.13) one can easily verify (3.13) and (4.15).

The next result shows that $Z_n(t)$ and $Z_n'(t)$ are uniformly consistent estimators for $Z(t)$ and $Z'(t)$.

LEMMA 4.4 *If the conditions of Lemma 4.3 are satisfied, then*

$$\sup_{0 \leq t \leq \delta} |Z_n(t) - Z(t)| = o_P(1) \qquad (4.22)$$

and

$$\sup_{0 \leq t \leq \delta} |Z_n'(t) - Z'(t)| = o_P(1). \qquad (4.23)$$

PROOF: Let $\eta_1 > 0$ and $\eta_2 > 0$. Putting together (4.8), (4.9) and Lemma 4.3 we get that there is an integer $N = N(\eta_1, \eta_2)$ satisfying

$$\liminf_{n\to\infty} P\{\beta + \alpha\epsilon_k^2 - \eta_1 \le \hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2 \le \beta + \alpha\epsilon_k^2 + \eta_1 \text{ for all } N \le k \le n\} \ge 1 - \eta_2. \tag{4.24}$$

Since the convergence of monotone functions to a limit is uniform over finite intervals, by the strong law of large numbers we have

$$\begin{aligned}\sup_{0\le t\le\delta}\Bigg|\frac{1}{n-N}\sum_{N\le k\le n}(\beta + \alpha\epsilon_k^2 + \eta_1)^t I\{\beta + \alpha\epsilon_k^2 + \eta_1 \le 1\}\\ - E((\beta + \alpha\epsilon_0^2 + \eta_1)^t I\{\beta + \alpha\epsilon_k^2 + \eta_1 \le 1\})\Bigg| = o(1) \quad \text{a.s.}\end{aligned} \tag{4.25}$$

$$\begin{aligned}\sup_{0\le t\le\delta}\Bigg|\frac{1}{n-N}\sum_{N\le k\le n}(\beta + \alpha\epsilon_k^2 + \eta_1)^t I\{\beta + \alpha\epsilon_k^2 + \eta_1 > 1\}\\ - E((\beta + \alpha\epsilon_0^2 + \eta_1)^t I\{\beta + \alpha\epsilon_k^2 + \eta_1 > 1\})\Bigg| = o(1) \quad \text{a.s.}\end{aligned} \tag{4.26}$$

Thus by (4.25) and (4.26) we have

$$\sup_{0\le t\le\delta}\left|\frac{1}{n-N}\sum_{N\le k\le n}(\beta + \alpha\epsilon_k^2 + \eta_1)^t - E(\beta + \alpha\epsilon_0^2 + \eta_1)^t\right| = o(1) \quad \text{a.s.}$$

Since

$$\lim_{\eta_1\to 0}\sup_{0\le t\le\delta}|E(\beta + \alpha\epsilon_0^2 + \eta_1)^t - Z(t)| = 0 \tag{4.27}$$

and

$$\sup_{0\le t\le\delta}\frac{1}{n}\sum_{1\le k\le N}(\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2)^t = o_P(1),$$

relation (4.22) is proved. Similar arguments give (4.23).

Proof of Theorem 1.1.: Let us observe first that $Z(0) = 1, Z'(t)$ and $Z''(t)$ exist on $[0, \delta]$. In (2.10) we assumed that $\nu_0 < \delta$. By (2.4) we have that $Z'(0) < 0$. Also $Z''(t) > 0$ and thus $Z(t)$ is convex on $[0, \delta]$. Corollary in Nelson (1990) yields that $Z(\tilde{\delta}) \ge 1$ for any $\nu_0 < \tilde{\delta} < \delta$. Since $Z(0) = 1$ and $Z'(0) < 0$ there is $\mu > 0$ such that $Z(t) < 1$ for all $0 < t \le \mu$. By $Z(\tilde{\delta}) \ge 1$ we conclude that there is a ν_0, $\mu < \nu_0 \le \tilde{\delta}$ such that $Z(\nu_0) = 1$. The observation that $Z''(t) > 0$ gives that $Z'(t)$ is strictly increasing. By $Z(0) = Z(\nu_0) = 1$ and $0 < \nu_0$, $Z(t)$ reaches its minumum at a unique point ν_* satisfying $0 < \nu_* < \nu_0$. Of course, $Z'(\nu_*) = 0$ and by the strict increase of Z' we obtain that $Z'(t) > 0$ for all $t > \nu_*$ and therefore ν_0 is unique.

Next we show that

$$\lim_{n\to\infty} P\{Z_n(t) < 1 \text{ for all } 0 < t < \infty\} = 0. \tag{4.28}$$

Let $\zeta \in (\tilde{\delta}, \delta)$. Then $Z(\zeta) > 1$ and by Lemma 4.4 we have

$$Z_n(\zeta) \quad \xrightarrow{P} \quad Z(\zeta) > 1,$$

establishing (4.28).

There are $\mu^* > 0$ and $\eta^* > 0$ such that $Z'(t) < -\mu^*$ on $0 \le t \le \eta^* < \tilde{\delta}$. Hence by Lemma 4.4 we obtain that

$$\lim_{n\to\infty} P\{Z_n'(t) \le -\mu^* \text{ for all } 0 \le t \le \eta^*\} = 1.$$

Since $Z'(\nu_*) = 0$, we get that $Z(t) < 1$ for all $0 < t \le \nu_*$. So by Lemma 4.4 we conclude that

$$\lim_{n\to\infty} P\{\nu_* \le \nu_n\} = 1.$$

Since $Z_n(t)$ is uniformly close to $Z(t)$ and ν_0 is the only point in $[\nu_*, \eta^*]$ where $Z(t) = 1$, standard arguments give Theorem 2.1.

We write

$$Z_n(t) - Z(t) = \frac{1}{n}\sum_{1\le k\le n}\{\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2)^t - (\beta + \alpha\epsilon_k^2)^t\} + \frac{1}{n}\sum_{1\le k\le n}\{(\beta+\alpha\epsilon_k^2)^t - Z(t)\}.$$

We show that the first term is essentially is a linear combination of $\hat{\omega}_n - \omega, \hat{\beta}_n - \beta$ and $\hat{\alpha}_n - \alpha$. A two-term Taylor expansion gives

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le k\le n}\{(\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2)^t - (\beta + \alpha\epsilon_k^2)^t\} \\
&= \frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\{(\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2) - (\beta+\alpha\epsilon_k^2)\} \\
&\quad + \frac{1}{2n}\sum_{1\le k\le n} t(t-1)\eta_k^{t-2}\{(\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2) - (\beta+\alpha\epsilon_k^2)\}^2 \\
&= V_n(t) + \bar{V}_n(t),
\end{aligned}$$

where η_k is between $\hat{\beta}_n + \hat{\alpha}_n\hat{\epsilon}_k^2$ and $\beta + \alpha\epsilon_k^2$. It is easy to see that

$$\begin{aligned}
V_n(t) &= (\hat{\beta}_n - \beta)\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1} + (\hat{\alpha}_n - \alpha)\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\epsilon_k^2 \\
&\quad + \alpha\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}(\hat{\epsilon}_k^2 - \epsilon_k^2) + (\hat{\alpha}_n - \alpha)\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}(\hat{\epsilon}_k^2 - \epsilon_k^2) \\
&= V_{n,1}(t) + \cdots + V_{n,4}(t).
\end{aligned}$$

Lemma 4.5 *If (2.1)–(2.4), (2.9), (2.11), (2.12) hold and*

$$E|\epsilon_0^2|^\delta < \infty \quad \textit{with some} \quad \delta > 2,$$

then

$$\sup_{0 \le t \le \delta} |V_{n,1}(t) - (\hat{\beta}_n - \beta) q_1(t)| = o_P(n^{-1/2}) \tag{4.29}$$

and

$$\sup_{0 \le t \le \delta} |V_{n,2}(t) - (\hat{\alpha}_n - \alpha) q_2(t)| = o_P(n^{-1/2}). \tag{4.30}$$

Proof: Since for any $i = 1, 2, 3$, the random variables $\ell_i(\epsilon_k^2) h_i(y_{k-1}, y_{k-2}, \ldots)$, $-\infty < k < \infty$ are stationary and uncorrelated with zero mean, we get

$$\operatorname{var}\left(\sum_{1 \le k \le n} \ell_i(\epsilon_k^2) h_i(y_{k-1}, y_{k-2}, \ldots) \right) = O(n),$$

so condition (2.11) yields

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| = O_P(n^{-1/2}). \tag{4.31}$$

By the law of large numbers for any fixed $t \le \delta$ we have

$$\frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} \longrightarrow q_1(t) \quad \text{a.s.}$$

and therefore the monotonicity argument leading to (4.25) yields

$$\sup_{0 \le t \le \delta} \left| \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} - q_1(t) \right| \overset{\text{a.s.}}{=} o(1). \tag{4.32}$$

Combining (4.31) and (4.32) we obtain (4.29). Similarly to (3.31) we have

$$\sup_{0 \le t \le \delta} \left| \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} \epsilon_k^2 - q_2(t) \right| \overset{\text{a.s.}}{=} o(1), \tag{4.33}$$

proving (4.30) on account of (4.31).

Using the definition of $\hat{\epsilon}_k^2$ we write

$$\begin{aligned}
V_{n,3}(t) &= \alpha \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} \epsilon_k^2 \left(\frac{\sigma_k^2}{\hat{w}_k(\hat{\boldsymbol{\theta}}_n)} - 1 \right) \\
&= \alpha \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} (\sigma_k^2 - \hat{w}_k(\hat{\boldsymbol{\theta}}_n)) \\
&\quad + \alpha \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha \epsilon_k^2)^{t-1} \epsilon_k^2 (\sigma_k^2 - \hat{w}_k(\hat{\boldsymbol{\theta}}_n))^2 / (\sigma_k^2 \hat{w}_k(\hat{\boldsymbol{\theta}}_n)) \\
&= V_{n,3}^*(t) + V_{n,3}^{**}(t).
\end{aligned}$$

Next we use (4.9) and get

$$V_{n,3}^*(t) = V_{n,3,1}^*(t) + \cdots + V_{n,3,9}^*(t),$$

where

$$V_{n,3,i}^*(t) = \alpha \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha\epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} A_{i,k}, \quad i = 1, 2, \ldots, 9.$$

LEMMA 4.6 *If the conditions of Lemma 4.5 are satisfied, then*

$$\sup_{0 \le t \le \delta} \left| V_{n,3,1}^*(t) - (\omega - \hat{\omega}_n) \frac{\alpha}{1-\beta} p_1 q_2(t) \right| = o_P(n^{-1/2})$$

and

$$\sup_{0 \le t \le \delta} \left| V_{n,3,2}^*(t) - (\beta - \hat{\beta}_n) \frac{\omega\alpha}{(1-\beta)^2} p_1 q_2(t) \right| = o_P(n^{-1/2}).$$

PROOF: Applying the ergodic theorem to the stationary ergodic sequence $t(\beta+\alpha\epsilon_k^2)^{t-1}\epsilon_k^2/\sigma_k^2$ and using the monotonicity argument leading to (4.25), we get

$$\sup_{0 \le t \le \nu_0 + \mu} \left| \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha\epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} - p_1 q_2(t) \right| \overset{\text{a.s.}}{=} o(1), \tag{4.34}$$

and therefore (4.31) implies Lemma 4.6.

LEMMA 4.7 *If the conditions of Lemma 4.5 are satisfied, then*

$$\sup_{0 \le t \le \delta} \left| V_{n,3,3}^*(t) - (\alpha - \hat{\alpha}_n) \left(q_2(t) - \frac{\omega}{1-\beta} p_1 q_2(t) \right) \right| = o_P(n^{-1/2}).$$

PROOF: By (2.7) we have

$$\alpha \sum_{0 \le i \le k-1} \beta^i y_{k-i-1}^2 = \sigma_k^2 - \frac{\omega}{1-\beta} + \frac{\omega\beta^k}{1-\beta} - \beta^k \sigma_0^2,$$

and therefore

$$\begin{aligned} V_{n,3,3}^*(t) &= (\alpha - \hat{\alpha}_n) \frac{1}{n} \sum_{1 \le k \le n} t(\beta + \alpha\epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} \left(\sigma_k^2 - \frac{\omega}{1-\beta} \right) \\ &\quad + (\alpha - \hat{\alpha}_n) \frac{1}{n} \sum_{1 \le k \le n} \beta^k t(\beta + \alpha\epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} \left(\frac{\omega}{1-\beta} - \sigma_0^2 \right). \end{aligned}$$

Since $0 < \beta < 1$, we conclude that

$$\sup_{0 \le t \le \delta} \sum_{1 \le k \le n} \beta^k t(\beta + \alpha\epsilon_k^2)^{t-1} \frac{\epsilon_k^2}{\sigma_k^2} = O_P(1),$$

and therefore (4.31) and (4.34) yield Lemma 4.7.

LEMMA 4.8 *If the conditions of Lemma 4.5 are satisfied, then*

$$\sup_{0 \le t \le \delta} |V_{n,3,4}^*(t) - (\beta - \hat{\beta}_n)\alpha^2 p_2 q_2(t)| = o_P(n^{-1/2}).$$

PROOF: Two-term Taylor expansion gives

$$V^*_{n,3,4}(t) = (\beta - \hat\beta_n)\alpha^2 \frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2}\sum_{1\le i\le k-1} i\beta^{i-1}y^2_{k-i-1}$$

$$+ (\beta - \hat\beta_n)^2\alpha^2 \frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2}\sum_{2\le i\le k-1} \frac{1}{2}i(i-1)\eta_i^{i-2}y^2_{k-i-1},$$

where η_i is between β and $\hat\beta_n$. For any $0<\Delta<\Delta'<1-\beta$ there is $C=C(\Delta,\Delta',\beta)$ such that

$$\sum_{k\le i<\infty} i\beta^{i-1}y^2_{k-i-1} \le C(\beta+\Delta)^k \sum_{k\le i<\infty}\left(\frac{\beta}{\beta+\Delta'}\right)^i y^2_{k-i-1}.$$

Since Δ, Δ' can be as small as we wish, Lemma 4.1 and our previous estimate imply that

$$\sup_{0\le t\le\delta}\left|\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2}\sum_{k\le i<\infty} i\beta^{i-1}y^2_{k-i-1}\right| = O_P(1).$$

Using again the ergodic theorem we conclude

$$\sup_{0\le t\le\delta}\left|\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2}\sum_{1\le i<\infty} i\beta^{i-1}y^2_{k-i-1} - p_2q_2(t)\right| = o_P(1)$$

and therefore

$$\sup_{0\le t\le\delta}\left|\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2(\boldsymbol\theta_0)}\sum_{1\le i\le k-1} i\beta^{i-1}y^2_{k-i-1} - \gamma_2\alpha_2(t)\right| = o_P(1). \tag{4.35}$$

By (4.31) we have for any $\Delta>0$, $\lim_{n\to\infty}P\{\hat\beta_n\le\beta+\Delta\}=1$ and therefore it is sufficient to show that for some $\Delta>0$

$$\sup_{0\le t\le\delta}\left|\frac{1}{n}\sum_{1\le k\le n} t(\beta+\alpha\epsilon_k^2)^{t-1}\frac{\epsilon_k^2}{\sigma_k^2}\sum_{2\le i\le k-1} \frac{1}{2}i(i-1)(\beta+\Delta)^{i-2}y^2_{k-i-1}\right| = O_P(1). \tag{4.36}$$

Following the proof of (4.35) one can derive (4.36) from the ergodic theorem and Lemma 4.1.

LEMMA 4.9 *If the conditions of lemma 4.5 are satisfied, then*

$$\sup_{0\le t\le\delta}|V^*_{n,3,i}(t)| = o_P(n^{-1/2}), \quad \text{if} \quad i=5,\dots,9.$$

PROOF: It is similar to the proofs of Lemmas 4.5–4.8, and hence omitted.

LEMMA 4.10 *If the conditions of Lemma 4.5 are satisfied,*

$$\sup_{0\le t\le\delta}\left|\frac{1}{n}\sum_{1\le k\le n}\{(\hat\beta_n+\hat\alpha_n\hat\epsilon_k^2)^t - (\beta+\alpha\epsilon_k^2)^t\}\right.$$

$$\left. - \{(\omega-\hat\omega_n)q_1^*(t) + (\beta-\hat\beta_n)q_2^*(t) + (\alpha-\hat\alpha_n)q_3^*(t)\}\right| = o_P(n^{-1/2}).$$

PROOF: Combining (4.31) and Lemmas 4.6–4.9 we get that

$$\sup_{0\le t\le\delta}|V^*_{n,3}(t)| = O_P(n^{-1/2}), \qquad \sup_{0\le t\le\delta}|V^{**}_{n,3}(t)| = o_P(n^{-1/2})$$

and therefore

$$\sup_{0\le t\le\delta}|V_{n,4}(t)| = O_P(1/n).$$

Thus we have

$$\sup_{0\le t\le\delta}|V_n(t) - \{(\omega-\hat\omega_n)q_1^*(t) + (\beta-\hat\beta_n)q_2^*(t) + (\alpha-\hat\alpha_n)q_3^*(t)\} = o_P(n^{-1/2}) \tag{4.37}$$

and therefore it suffices to show that

$$\sup_{0\le t\le\delta}|\bar V_n(t)| = o_P(n^{-1/2}). \tag{4.38}$$

The arguments leading to (4.37) can be used to establish (4.38) and therefore the details are omitted.

Next we show that $n^{1/2}(Z_n(t) - Z(t))$ converges in $\mathcal{C}[0,\delta]$ to a continuous Gaussian process. The covariance of the limit is

$$\begin{aligned} r(t,s) = &\sum_{1\le i,j\le 3} q_i^*(t)\bar h_{ij}\bar\ell_{ij}q_j^*(s) + \sum_{1\le i\le 3} q_i^*(t)\bar h_i E\{(\beta+\alpha\epsilon_0^2)^s\ell_i(\epsilon_0^2)\} \\ &+ \sum_{1\le i\le 3} q_i^*(s)\bar h_i E\{(\beta+\alpha\epsilon_0^2)^t\ell_i(\epsilon_0^2)\} + Z(t+s) - Z(t)Z(s). \end{aligned}$$

LEMMA 4.11 *If the conditions of Lemma 4.5 are satisfied, then*

$$n^{1/2}(Z_n(t) - Z(t)) \overset{\mathcal{C}[0,\delta]}{\longrightarrow} \Gamma(t),$$

where $\Gamma(t)$ is a Gaussian process with $E\Gamma(t) = 0$ and $E\Gamma(t)\Gamma(s) = r(t,s)$.

PROOF: By Lemma 4.10 and condition (2.11) it is enough to prove the weak convergence of $R_n(t) = n^{-1/2}\sum_{1\le i\le n} z_j(t)$, where

$$z_j(t) = \sum_{i=1,2,3} q_i^*\ell_i(\epsilon_j^2)h_i(y_{j-1}, y_{j-2},\ldots) + (\beta+\alpha\epsilon_j^2)^t - Z(t).$$

Let $\mathcal{F}_j$ be the σ-algebra generated by $\epsilon_i, -\infty < i \le j$. Let K be a positive intger, $\lambda_i, 0\le t_i\le\delta, 1\le i\le K$ be constants and define $z_j^* = \sum_{1\le i\le K}\lambda_i z_j(t_i)$. It is easy to see that z_j^* is a stationary and ergodic sequence and $(\sum_{1\le j\le \ell} z_j, \mathcal{F}_\ell,\ \ell = 1,2,\ldots)$ is a martingale.

Using Theorem 23.1 of Billingsley (1968) we conclude the asymptotic normality of $\sum_{1\le j\le n} z_j^*$ and therefore the Cramér–Wald device (cf. Billingsley (1968), p.49)) gives the convergence of the finite dimensional distributions of $R_n(t)$.

Next we show the tightness of $R_n(t)$. It is easy to see that $R_n(0)=0$ and

$$E\left(n^{-1/2}\sup_{0\le t,s\le \nu_0+\delta}\left|(q_i^*(t)-q_i^*(s))\sum_{0\le j\le n}\ell_i(\epsilon_j^2)h_i(y_{j-1},\ldots)\right|\Big/|t-s|\right)^2=O(1)$$

$i=1,2,3)$, since the variance of the sum above is $O(n)$. Next we note that

$$\begin{aligned}E&\left(\sum_{1\le k\le n}\{(\beta+\alpha\epsilon_k^2)^t-(\beta+\alpha\epsilon_k^2)^s-(Z(t)-Z(s))\}\right)^2\\&=nE((\beta+\alpha\epsilon_0^2)^t-(\beta+\alpha\epsilon_0^2)^s-(Z(t)-Z(s)))^2\end{aligned}$$

and

$$|Z(t)-Z(s)|\le C|t-s|$$

with some constant C. Also, by the mean-value theorem

$$|(\beta+\alpha\epsilon_0^2)^t-(\beta+\alpha\epsilon_0^2)^s|\le|t-s|(\beta+\alpha\epsilon_0^2)^\delta|\log(\beta+\alpha\epsilon_0^2)|$$

and therefore

$$E\left(n^{-1/2}\sum_{1\le k\le n}\{(\beta+\alpha\epsilon_k^2)^t-(\beta+\alpha\epsilon_k^2)^s-(Z(t)-Z(s))\}\right)^2\le\bar{C}|t-s|^2$$

with some constant $\bar{C}$. Now Theorem 12.3 in Billingsley (1968) gives tightness.

Proof of Theorem 1.2: Observing that

$$0=Z_n(\nu_n)-Z(\nu_0)=Z_n(\nu_n)-Z(\nu_n)+Z(\nu_n)-Z(\nu_0),$$

by the mean-value theorem we have

$$\nu_n-\nu_0=-\frac{1}{Z'(\eta_n)}(Z_n(\nu_n)-Z(\nu_n)),$$

where η_n is between ν_n and ν_0. Theorem 2.1 and the continuity of Z' give that $Z'(\eta_n)\xrightarrow{P}Z'(\nu_0)$. Theorem 2.1, Lemma 4.11 and the continuity of $\Gamma(t)$ yield $n^{1/2}(Z_n(\nu_n)-Z(\nu_n))\xrightarrow{\mathcal{D}}\Gamma(\nu_0)$, completing the proof of Theorem 2.2.

References

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

Berkes, I., Horváth, L. and Kokoszka, P. (2002). GARCH processes: structure and estimation. *Bernoulli*; forthcoming, preprint available at `http://math.usu.edu/~piotr/research.html`.

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

Csörgő, S. (1982). The empirical moment generating function. In *Nonparametric Statistical Inference (Coll. Math. Soc. J. Bolyai* **32***), Amsterdam* (eds B. V. Gnedenko, M. L. Puri and I. Vincze), pp. 139–159. North-Holland.

Csörgő, S. and Teugels, J. L. (1990). Empirical Laplace transform and approximation of compound distributions. *J. Appl. Probab.*, **27,** 88–101.

Dempster, M. (2002) (ed.). *Risk management: value at risk and beyond.* Cambridge University Press.

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

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27,** 760–781.

Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford*, **38,** 45–71.

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.

Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Statist.*, **19,** 1547–1569.

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

Leadbetter, M. R. (1991). On a basis for 'peaks over threshhold' modelling. *Statist. Probab. Letters*, **12,** 357–362.

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

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

McNeil, 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; Available at `http://www.math.ethz.ch/~mcneil/`.

Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelation and extremes of a GARCH(1, 1) process. *The Annals of Statistics*, **28,** 1427–1451.

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

Novak, S. Y. (2000). Confidence intervals for a tail index estimator. In *Proceedings Conf. "Measuring risk in comples stochastic systems*, *Berlin* (eds J. Franke, W. Haerdle and G. Stahl), pp. 229–236. Springer.

Pitts, S. M., Embrechts, P. and Grübel, R. (1996). Confidence bounds for the adjustment coefficient. *Adv. Appl. Probab.*, **28,** 802–827.

Quintos, C., Fan, Z. and Phillips, P. C. B. (2001). Structural change tests in tail behaviour and the Asian crisis. *The Review of Economic Studies*, **68,** 633–663.

Resnick, S. I. and Stărică, C. (1995). Consistency of Hill's estimator for dependent data. *J. Appl. Probab.*, **32,** 139–167.

Resnick, S. I. and Stărică, C. (1997). Asymptotic behaviour of Hill's estimator for autoregressive data. *Commun. Statist. - Stochastic Models*, **13,** 703–721.

Resnick, S. I. and Stărică, C. (1998). Tail estimation for dependent data. *Ann. Appl. Probability*, **8,** 1156–1183.

Rootzén, H., Leadbetter, M. R. and de Haan, L. (1992). Tail and quantile estimation for strongly mixing stationary processes. Technical Report, Department of Statistics, University of North Carolina.

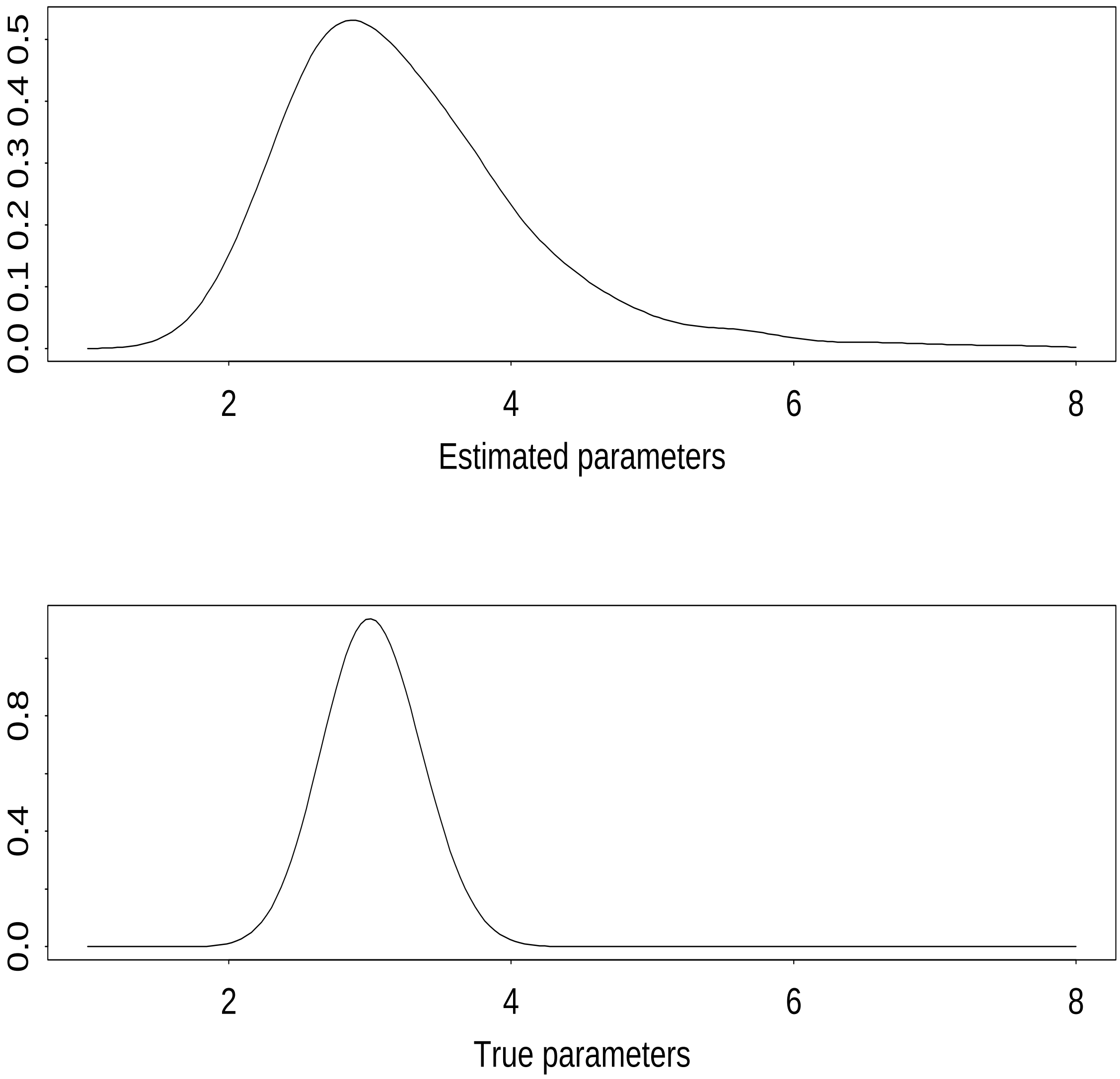

FIGURE 1. Top panel: Estimated sampling density of the estimator ν_n with $n = 1000$. The true DGP is GARCH(1,1) with $\alpha = 0.3, \beta = 0.3$ and standard normal innovations ϵ_k. The 2.5, 25, 50, 75 and 97.5 percentiles of the sampling distribution are respectively 2.06, 2.68, 3.12, 3.7, and 5.62. The approximate true value of ν is 3.436. Lower panel: Estimated sampling density of the estimator ν_n with $n = 1000$ but with the estimators $\hat{\beta}_n$ and $\hat{\alpha}_n$ replaced by the true values β and α. The estimated densities are based on 1000 replications of ν_n smoothed by a Gaussian kernel.

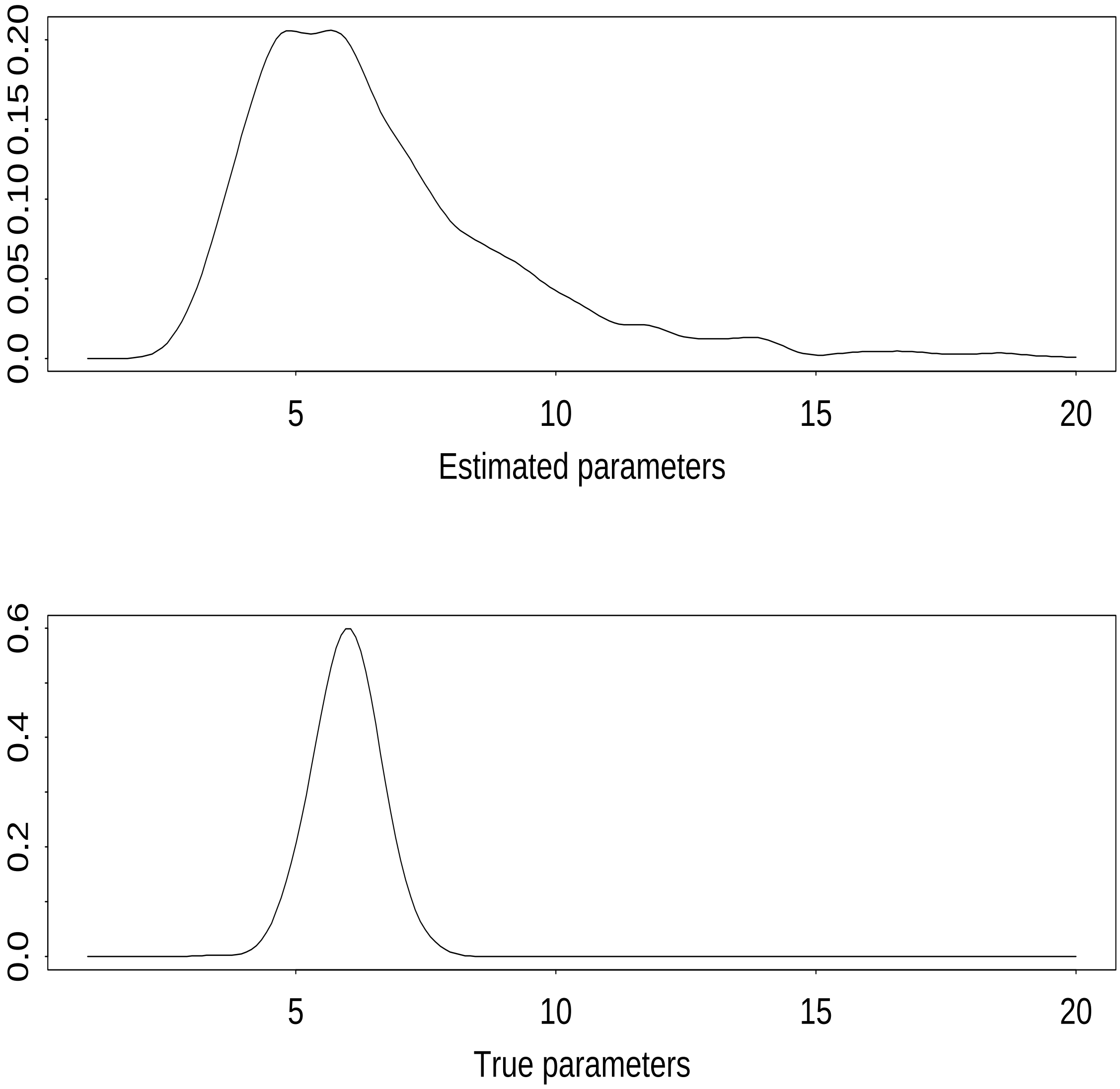

FIGURE 2. Top panel: Estimated sampling density of the estimator ν_n with $n = 1000$. The true DGP is GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$ and standard normal innovations ϵ_k. The 2.5, 25, 50, 75 and 97.5 percentiles of the sampling distribution are respectively 3.4, 4.82, 6.04, 7.78, and 14.04. The approximate true value of ν is 6.263. Lower panel: Estimated sampling density of the estimator ν_n with $n = 1000$ but with the estimators $\hat{\beta}_n$ and $\hat{\alpha}_n$ replaced by the true values β and α. The estimated densities are based on 1000 replications of ν_n smoothed by a Gaussian kernel.

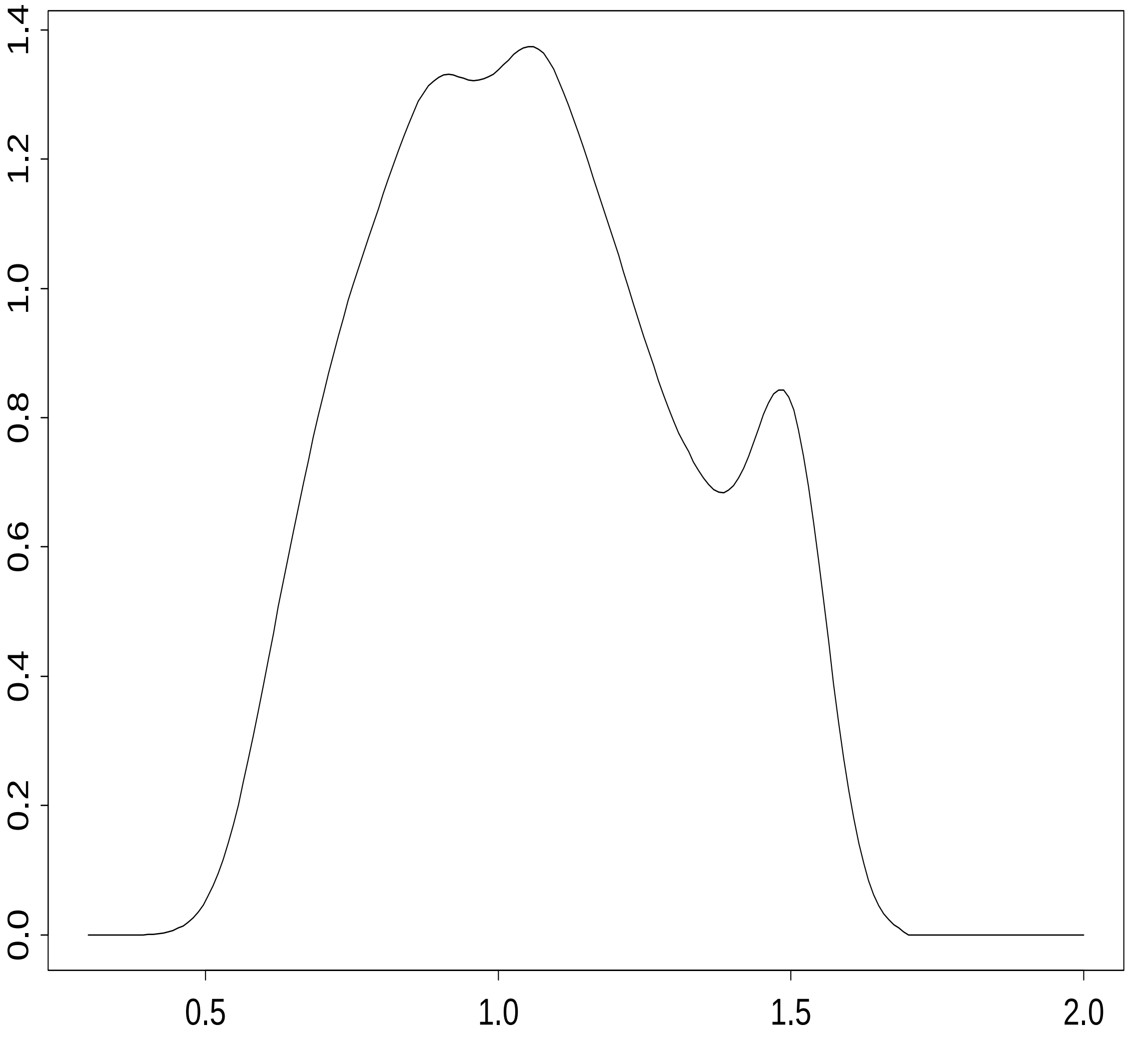

FIGURE 3. Estimated sampling density of the estimator ν_n with $n = 1000$. The true DGP is AR(1) with the autoregressive parameters equal to 0.5 and iid symmetric stable innovations with the index of stability equal to 1.75. The 2.5, 25, 50, 75 and 97.5 percentiles of the sampling distribution are respectively 0.63, 0.86, 1.05, 1.25 and 1.51. The exact true value of ν is 0.875. The estimated densities are based on 1000 replications of ν_n smoothed by a Gaussian kernel.