Horvath, Lajos; Kokoszka, Piotr; Teyssière, Gilles

Working Paper

Empirical process of the squared residuals of an ARCH sequence

SFB 373 Discussion Paper, No. 1999,87

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Horvath, Lajos; Kokoszka, Piotr; Teyssière, Gilles (1999) : Empirical process of the squared residuals of an ARCH sequence, SFB 373 Discussion Paper, No. 1999,87, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10046737

This Version is available at:
https://hdl.handle.net/10419/61773

Empirical process of the squared residuals of an ARCH sequence*

Lajos Horváth
University of Utah

Piotr Kokoszka
University of Liverpool

Gilles Teyssière
Humboldt Universität and GREQAM

17 August, 1999

Abstract

We show that the empirical process of the squared residuals of an ARCH(p) sequence converges in distribution to a Gaussian process $B(F(t)) + tf(t)\xi$, where F is the distribution function of the squared innovations, f its derivative, $\{B(t),\ 0 \le t \le 1\}$ a Brownian bridge and ξ a normal random variable.

Key words: ARCH model, empirical process, squared residuals.

Running head: Empirical process for ARCH

1 Introduction and results

Procedures based on the empirical distribution function of independent identically distributed observations occupy a central place in statistical inference, see Shorack and Wellner (1986). For time series data, residuals must be considered, and since these necessarily depend on parameter estimates, the asymptotic theory for the empirical distribution function is more comlex in such cases. Nevertheless, inference based on residuals, especially model goodness–of–fit tests and various diagnostic checks, is a fundamental tool in the statistical analysis of *linear* time series models, see Brockwell and Davis (1991). By contrast, large sample theory for the residuals of non–linear time series models is much less developed. Li and Mak (1994) and Horváth and Kokoszka (1999) study squared residual autocorrelations of ARCH sequences, whose importance in various specification tests was demonstrated by Lundbergh and Teräsvirta (1998). Tjøstheim (1999) considers non–parametric tests based on squared residuals.

*Research partially supported by EPSRC grant GR/M68879 at the University of Liverpool.

In this paper we consider the ARCH(p) model defined by the equations

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = b_0 + \sum_{j=1}^{p} b_j y_{t-j}^2,$$

where $\{\varepsilon_i, -\infty < i < \infty\}$ are independent identically distributed random variables with

$$E\varepsilon_0 = 0 \text{ and } \varepsilon_0^2 = 1.$$

We assume that $\mathbf{b} = (b_0, b_1, \ldots, b_p)$ is the parameter vector satisfying

$$b_0 > 0, \quad b_i \geq 0, \ 1 \leq i \leq p.$$

The distribution function of ε_0^2 will be denoted by F and we assume that

(1.1) $$f(t) = F'(t) \text{ exists and is continuous on } (0, \infty),$$

(1.2) $$\lim_{t \to 0} tf(t) = 0 \quad \text{and} \quad \lim_{t \to \infty} tf(t) = 0.$$

We assume that the parameter vector $\mathbf{b}$ is estimated by an estimator $\hat{\mathbf{b}}_n = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p)$ which admits the representation

(1.3) $$\hat{b}_i - b_i = \frac{1}{n} \sum_{1 \leq j \leq n} l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) + o_P(n^{1/2}), \quad 0 \leq i \leq p.$$

The functions l_i and f_i are regular in the sense that

(1.4) $$El_i(\varepsilon_0^2) = 0, \quad 0 \leq i \leq p,$$

(1.5) $$E\left[l_i(\varepsilon_0^2)\right]^2 < \infty, \quad 0 \leq i \leq p$$

and

(1.6) $$E\left[f_i(\varepsilon_0, \varepsilon_{-1}, \ldots)\right]^2 < \infty, \quad 0 \leq i \leq p.$$

We show later that commonly used estimators, see e.g. Chapter 4 of Gouriéroux (1997), admit representation (1.3).

The squared residuals are defined as

$$\hat{\varepsilon}_k^2 = \frac{y_k^2}{\hat{\sigma}_k^2}, \quad p < k \leq n,$$

where

$$\hat{\sigma}_k^2 = \hat{b}_0 + \sum_{1 \leq j \leq p} \hat{b}_j y_{k-j}^2, \quad p < k \leq n.$$

In this paper we study the weak convergence of the empirical process of the squared residuals

$$\hat{e}_n(t) = n^{1/2} \left(\hat{F}_n(t) - F(t) \right),$$

where

$$\hat{F}_n(t) = \frac{1}{n} \sum_{p < k \leq n} \mathbf{I}\left\{ \hat{\varepsilon}_k^2 \leq t \right\}$$

denotes the empirical distribution function of the squared residuals. Following Giraitis *et al.* (1999) we also assume that

$$E\varepsilon_0^4 < \infty \tag{1.7}$$

and

$$\left(E\varepsilon_0^4\right)^{1/2} \sum_{1\le j\le p} b_j < 1. \tag{1.8}$$

If condition (1.8) is satisfied, then the ARCH equations above have a unique strictly stationary solution such that $Ey_k^4 < \infty$ and the squares y_k^2 have a Volterra representation

$$y_k^2 = \sum_{l=0}^{\infty} \sum_{j_1,\ldots,j_l=1}^{p} b_{j_1} \ldots b_{j_l} \varepsilon_k^2 \varepsilon_{k-j_1}^2 \ldots \varepsilon_{k-j_1-\ldots-j_l}^2.$$

Thus y_k^2 is a function of $\varepsilon_k, \varepsilon_{k-1}, \ldots$, and so it follows form the standard theory, see e.g. Stout (1974) pp. 181–182, that the sequence $\{y_k^2\}$ is ergodic.

We note that condition (1.8) is not necessary for the covariance stationarity of the process $\{y_k^2\}$, but it is easy to verify. Necessary and sufficient conditions are more complex and difficult to state in a closed form for $p > 2$, see Section 3.4 of He and Teräsvirta (1998). It is also well–known that ARCH(p) and more general sequences are not only ergodic but also mixing with geometric rate, we refer to Lu and Cheng (1997), Remark 4.2, for futher references. The results of this paper remain valid if condition (1.8) is replaced by any assumption guaranteeing that the process $\{y_k\}$ is strictly stationary with $Ey_k^4 < \infty$ and ergodic. The theory developed here may not be valid if the assuption $Ey_k^4 < \infty$ is dropped as the results of Davis and Mikosch (1999) suggest. These authors consider however functions of the observations y_k rather than estimated residuals.

In order to state our main result we need further notation:

$$\alpha_i = Ef_i(\varepsilon_0, \varepsilon_{-1}, \varepsilon_{-2}, \ldots), \quad 0 \le i \le p,$$

$$g_i(t) = \alpha_i \int_0^t l_i(u) f(u) du, \quad 0 \le i \le p,$$

$$\beta_0 = E\left[\frac{1}{\sigma_0^2}\right], \quad \beta_i = E\left[\frac{y_{-i}^2}{\sigma_0^2}\right], \quad 1 \le i \le p,$$

$$\gamma_{ij} = E\left[l_i(\varepsilon_0^2) l_j(\varepsilon_0^2)\right] E\left[f_i(\varepsilon_{-1}, \varepsilon_{-2}, \ldots) f_j(\varepsilon_{-1}, \varepsilon_{-2}, \ldots)\right], \quad 0 \le i, j \le p,$$

$$\begin{aligned} r(t,s) &= F(t \wedge s) - F(t)F(s) \\ &+ tf(t) \sum_{0\le i\le p} \beta_i g_i(s) + sf(s) \sum_{0\le i\le p} \beta_i g(t) \\ &+ stf(t)f(s) \sum_{0\le i,j\le p} \beta_i \gamma_{ij} \beta_j. \end{aligned}$$

THEOREM 1.1 *If conditions* (1.1) – (1.8) *hold, then*

$$\hat{e}_n(t) \to \Gamma(t),$$

where the convergence is in the Skorokhod space $\mathcal{D}[0,\infty]$ *and* Γ *is a Gaussian proces with*

$$E\Gamma(t) = 0 \quad \text{and} \quad E\left[\Gamma(t)\Gamma(s)\right] = r(t,s).$$

The convergence in Theorem 1.1 is equivalent to the convergence $\hat{e}_n(F^{-1}(x)) \to \Gamma(F^{-1}(x))$, $0 \le x \le 1$, in $\mathcal{D}[0,1]$.

Note that the process Γ admits the representation

$$\Gamma(t) = B(F(t)) + tf(t)\xi, \tag{1.9}$$

where $\{B(t),\ 0 \le t \le 1\}$ is a Brownian bridge and ξ is a normal random variable. They are jointly Gaussian with covariance structure:

$$E\left[B(F(t))B(F(s))\right] = F(t \wedge s) - F(t)F(s),$$

$$E\xi^2 = \sum_{0 \le i,j \le p} \beta_i \gamma_{ij} \beta_j,$$

$$E\left[\xi B(F(t))\right] = \sum_{0 \le i \le p} \beta_i g_i(t).$$

Durbin (1973a, 1973b) was the first who considered the weak convergence of empirical process when parameters are estimated. He mainly studied the case when the parameters are estimated by the maximum likelihood method. Burke *et al.* (1979) and Csörgő and Révész (1981) considered the general case when it is assumed only that the difference between the estimator and the estimated parameter is approximately given by an integral with respect to the empirical process of the observations. The limit in their case has a representation like (1.9), but in the iid case, ξ is a stochastic integral of a deteministic function with respect to $B(F(\cdot))$ Our case is somewhat different. Before presenting detailed proofs, we outline the argument. Setting

$$\delta_i^2 = \frac{\hat{\sigma}_i^2}{\sigma_i^2}$$

we write

$$\hat{e}_n(t) = \hat{e}_{n,1}(t) + \hat{e}_{n,2}(t) + (p/\sqrt{n})F(t),$$

where

$$\hat{e}_{n,1}(t) = n^{-1/2} \sum_{p < i \le n} \left\{ \mathbf{I}\left\{\varepsilon_i^2 \le t\delta_i^2\right\} - F(t\delta_i^2) \right\}$$

and

$$\hat{e}_{n,2}(t) = n^{-1/2} \sum_{p < i \le n} \left\{ F(t\delta_i^2) - F(t) \right\}.$$

We will show in the proofs that δ_i^2 is so close to one that the difference between $\hat{e}_{n,1}(t)$ and

$$e_n(t) = n^{-1/2} \sum_{p < i \le n} \left(\mathbf{I}\left\{\varepsilon_i^2 \le t\right\} - F(t) \right)$$

is negligible. As for the second term, $\hat{e}_{n,2}(t)$ will be approximated by

$$h_n(t) = tf(t)n^{-1/2} \sum_{p < i \le n} \left(\delta_i^2 - 1\right). \tag{1.10}$$

Observe that

$$\sum_{p < i \le n} \left(\delta_i^2 - 1\right) = \sum_{p < i \le n} \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^2}$$

$$= (\hat{b}_0 - b_0) \sum_{p<i\le n} \frac{1}{\sigma_i^2} + (\hat{b}_1 - b_1) \sum_{p<i\le n} \frac{y_{i-1}^2}{\sigma_i^2} + \ldots + (\hat{b}_p - b_p) \sum_{p<i\le n} \frac{y_{i-p}^2}{\sigma_i^2}.$$

Therefore the joint convergence of $e_n(t)$ and $\sqrt{n}(\hat{\mathbf{b}}_n - \mathbf{b})$ will imply the result in Theorem 1.1.

2 Proofs

We start with several lemmas which establish the main steps outlined in the discussion following the statement of Theorem 1.1.

The maximum norm of a vector is denoted by $||\cdot||$.

LEMMA 2.1 *If conditions* (1.3) – (1.8) *hold, then*

$$\sqrt{n} \left|\left| \hat{\mathbf{b}}_n - \mathbf{b} \right|\right| = O_P(1). \tag{2.1}$$

PROOF. Let $\mathcal{F}_k$ be the σ–algebra generated by $\{\varepsilon_k, \varepsilon_{k-1}, \ldots\}$. Let

$$\eta_i(k) = \sum_{1\le j\le k} l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots), \quad 0 \le i \le p.$$

It is easy to see that for each $0 \le i \le p$, $\{\eta_i(k), \mathcal{F}_k\}$ is a martingale. We show that $n^{-1/2}\eta_i(n)$ is asymptotically normal, what will also establish (2.1). We will verify that conditions (3.33) – (3.35) of Theorem 3.5 of Hall and Heyde (1980) are satisfied. Setting

$$X_{j,i} = \eta_i(j) - \eta_i(j-1) = l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots),$$

we will thus verify that

$$E \left| \frac{1}{n} \sum_{1\le j\le n} E\left\{ X_{j,i}^2 | \mathcal{F}_{j-1} \right\} - \gamma_{ii} \right| \to 0, \tag{2.2}$$

$$\max_{1\le j\le n} \frac{1}{n} E\left\{ X_{j,i}^2 | \mathcal{F}_{j-1} \right\} \xrightarrow{P} 0 \tag{2.3}$$

and

$$E \left| \frac{1}{n} \sum_{1\le j\le n} X_{j,i}^2 - \gamma_{ii} \right| \to 0. \tag{2.4}$$

(Observe that $EX_{j,i} = El_i(\varepsilon_j^2) Ef_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) = 0$.) Relations (2.4) and (2.2) follow from the ergodic theorem because

$$EX_{j,i}^2 = El_i^2(\varepsilon_j^2) Ef_i^2(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) = \gamma_{ii} < \infty$$

and, by the independence of ε_j and $\mathcal{F}_{j-1}$,

$$E\left\{ X_{j,i}^2 | \mathcal{F}_{j-1} \right\} = f_i^2(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) El_i^2(\varepsilon_j^2).$$

To verify (2.3), note that

$$\max_{1\le j\le n} E\left\{ l_i^2(\varepsilon_j^2) f_i^2(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) | \mathcal{F}_{j-1} \right\} = El_i^2(\varepsilon_0^2) \max_{1\le j\le n} f_i^2(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots).$$

Observing that $f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)$ is a stationary sequence, the existence of a common second moment gives that

$$\sum_{1\le j<\infty} P\left\{|f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)| > \varepsilon j^{1/2}\right\}$$

$$= \sum_{1\le j<\infty} P\left\{|f_i(\varepsilon_0, \varepsilon_{-1}, \ldots)| > \varepsilon j^{1/2}\right\} < \infty$$

for all $\varepsilon > 0$. Hence

$$|f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)| \, / j^{1/2} \stackrel{a.s.}{\to} 0, \tag{2.5}$$

which implies that

$$\max_{1\le j\le n} |f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)| = o_P(n^{1/2}).$$

Thus we proved that for any $0 \le i \le p$ $E\left\{X_{j,i}^2 | \mathcal{F}_{j-1}\right\} = o_P(n^{1/2})$, which implies (2.3).

Of course, if $\gamma_{ii} = 0$, then the limit degenerates to 0. ■

Recall now the definition of the function h_n given in (1.10).

LEMMA 2.2 *If* (1.1) – (1.8) *hold, then*

$$\sup_{0\le t<\infty} |\hat{e}_{n,2}(t) - h_n(t)| = o_P(1).$$

PROOF. By the mean value theorem

$$F(t\delta_i^2) - F(t) = f(\xi_i)t(\delta_i^2 - 1) = tf(t)(\delta_i^2 - 1) + t(f(t) - f(\xi_i))(\delta_i^2 - 1), \tag{2.6}$$

where ξ_i is between t and $t\delta_i^2$. Hence

$$\begin{aligned} |\xi_i - t| &\le |t| \left\{ |\hat{b}_0 - b_0| \frac{1}{\sigma_i^2} + |\hat{b}_1 - b_1| \frac{y_{i-1}^2}{\sigma_i^2} + \ldots + |\hat{b}_p - b_p| \frac{y_{i-p}^2}{\sigma_i^2} \right\} \\ &\le \frac{|t|}{b_0} \left\{ |\hat{b}_0 - b_0| + |\hat{b}_1 - b_1| y_{i-1}^2 + \ldots + |\hat{b}_p - b_p| y_{i-p}^2 \right\}. \end{aligned}$$

Using the argument leading to (2.5), we obtain

$$\max_{1\le i\le n} y_i^2 = o_P(n^{1/2}) \tag{2.7}$$

and therefore Lemma 2.1 yields

$$\sup_{0\le t<\infty} \max_{p<i\le n} |\xi_i - t|/t = o_P(1). \tag{2.8}$$

Next we show that

$$\max_{p\le i\le n} \sup_{0\le t<\infty} t|f(t) - f(\xi_i)| = o_P(1). \tag{2.9}$$

Using condition (1.2) and relation (2.8), we obtain

$$\lim_{T\to\infty} \limsup_{n\to\infty} P\left\{ \max_{p\le i\le n} \sup_{T\le t<\infty} tf(\xi_i) > \varepsilon \right\} = 0,$$

$$\lim_{c\to 0}\limsup_{n\to\infty} P\left\{\max_{p\le i\le n}\sup_{0\le t\le c} tf(\xi_i) > \varepsilon\right\} = 0,$$

so it is enough to prove that

$$\max_{p\le i\le n}\sup_{c\le t<T} t|f(t) - f(\xi_i)| = o_P(1) \tag{2.10}$$

for any $0 < c < T < \infty$. Relation (2.10) follows from (1.1) and (2.8).

To complete the proof of Lemma 2.2, it is sufficient to establish that

$$\left|\sum_{p<i\le n}\left(\delta_i^2 - 1\right)\right| = O_P(n^{1/2}). \tag{2.11}$$

Using the definition of δ_i^2, we have that

$$\sum_{p<i\le n}\left(\delta_i^2 - 1\right) = \left(\hat{b}_0 - b_0\right)\sum_{p<i\le n}\frac{1}{\sigma_i^2} + \left(\hat{b}_1 - b_1\right)\sum_{p<i\le n}\frac{y_{i-1}}{\sigma_i^2} + \ldots + \left(\hat{b}_p - b_p\right)\sum_{p<i\le n}\frac{y_{i-p}}{\sigma_i^2}$$

and therefore by Lemma (2.1)

$$\left|\sum_{p<i\le n}\left(\delta_i^2 - 1\right)\right| \le \frac{n^{1/2}}{b_0} n^{1/2}\left|\left|\hat{\mathbf{b}}_n - \mathbf{b}\right|\right|\left\{1 + \frac{1}{n}\sum_{p<i\le n} y_{i-1}^2 + \ldots + \frac{1}{n}\sum_{p<i\le n} y_{i-p}^2\right\}$$

$$= O_P(n^{1/2})\left\{1 + \frac{1}{n}\sum_{p<i\le n} y_{i-1}^2 + \ldots + \sum_{p<i\le n} y_{i-p}^2\right\}.$$

By the pointwise ergodic theorem

$$\frac{1}{n}\sum_{p<i\le n} y_i^2 \overset{a.s.}{\to} EY_0^2, \tag{2.12}$$

which completes the proof of (2.11). ∎

Let $\mathbf{T} = [-T, T]^{p+1}$ and define

$$\gamma_i(\mathbf{s}) = s_0\frac{1}{\sigma_i^2} + s_1\frac{y_{i-1}^2}{\sigma_i^2} + \ldots + s_p\frac{y_{i-p}^2}{\sigma_i^2}, \quad \mathbf{s} = (s_0, s_1, \ldots, s_p),$$

so that $\delta_i = 1 + \gamma_i(\hat{\mathbf{b}}_\mathbf{n} - \mathbf{b})$.

LEMMA 2.3 *If conditions* (1.1) – (1.8) *hold, then for any* $T > 0$

$$\sup_{0\le t<\infty}\sup_{\mathbf{s}\in\mathbf{T}}\left|n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2 \le t + tn^{-1/2}\gamma_i(\mathbf{s})\right\} - F\left(t + tn^{-1/2}\gamma_i(\mathbf{s})\right)\right] - e_n(t)\right| = o_P(1).$$

PROOF. Fix $\varepsilon > 0$ and define

$$\varepsilon_i(k) = -T + k\varepsilon, \quad 0 \le k \le K = [2T/\varepsilon], \quad \varepsilon_i(K+1) = T, \quad 0 \le i \le p.$$

The $(p+1)(K+2)$ points $(\varepsilon_0(k_0), \varepsilon_1(k_1), \ldots, \varepsilon_p(k_p))$, $0 \le k_i \le K+1$, $0 \le i \le p$, define a grid in $\mathbf{T}$. If $\mathbf{T}^*$ is a cell of the grid, then the right–most point is defined as

$$\left(\max_{\mathbf{s}\in\mathbf{T}^*} s_0, \max_{\mathbf{s}\in\mathbf{T}^*} s_1, \ldots, \max_{\mathbf{s}\in\mathbf{T}^*} s_p\right), \quad (s_0, s_1, \ldots, s_p) = \mathbf{s},$$

the left–most point being defined similarly with max replaced by min. Set $\mathbf{k} = (k_0, k_1, \ldots, k_p)$ and let $\mathbf{T}(\mathbf{k})$ denote the cell of the grid with the right–most point

$$\boldsymbol{\varepsilon}(\mathbf{k}) := (\varepsilon_0(k_0), \varepsilon_1(k_1), \ldots, \varepsilon_p(k_p))$$

and the left–most point

$$\boldsymbol{\varepsilon}^*(\mathbf{k}) := \left(\varepsilon_0^*(k_0), \varepsilon_1^*(k_1), \ldots, \varepsilon_p^*(k_p)\right).$$

Let

$$\Omega_n^* = \{\omega : \max_{p\le i\le n} \sup_{\mathbf{s}\in\mathbf{T}} n^{-1/2}|\gamma_i(\mathbf{s})| > 1/2\}.$$

By (2.7) we have that

$$\lim_{n\to\infty} P\{\Omega_n^*\} = 0.$$

The mean value theorem yields that

$$\begin{aligned} &F\left(t + tn^{-1/2}\gamma_i(s_0, \ldots, s_{i-1}, \epsilon_i(k+1), s_{i+1}, \ldots, s_p)\right) \\ &\quad - F\left(t + tn^{-1/2}\gamma_i(s_0, \ldots, s_{i-1}, \epsilon_i(k), s_{i+1}, \ldots, s_p)\right) \\ &= f(\xi^*)t(\gamma_i(s_0, \ldots, s_{i-1}, \epsilon_i(k+1), s_{i+1}, \ldots, s_p) - \gamma_i(s_0, \ldots, s_{i-1}, \epsilon_i(k), s_{i+1}, \ldots, s_p)), \end{aligned}$$

where ξ^* is a point between the two arguments of F above. If $\omega \in \Omega_n^*$, then

$$tf(\xi^*) = \xi^* f(\xi^*)\frac{t}{\xi^*} \le 2C,$$

where $C = \sup_{0\le t<\infty} tf(t)$. Therefore

$$\begin{aligned} &F\left(t + tn^{-1/2}\gamma_i(s_0, \ldots, s_{i-1}, \varepsilon_i(k+1), s_{i+1}, \ldots, s_p)\right) \\ &\quad - F\left(t + tn^{-1/2}\gamma_i(s_0, \ldots, s_{i-1}, \varepsilon_i(k), s_{i+1}, \ldots, s_p)\right) \\ &\quad \le 2Cn^{-1/2}\varepsilon\frac{y_{j-i}^2}{\sigma_j^2} \end{aligned} \tag{2.13}$$

for all $p < j \le n$ and $0 \le i \le p$. Applying (2.13) consecutively to each coordinate, we get for any $\omega \in \Omega_n^*$

$$\sup_{0\le t<\infty} \max_{0\le k_0,\ldots,k_p\le K+1} \sup_{\mathbf{s}\in\mathbf{T}(\mathbf{k})} \left| \sum_{p<i\le n} \left[F\left(t + tn^{-1/2}\gamma_i(\mathbf{s})\right) - F\left(t + tn^{-1/2}\gamma_i\left(\boldsymbol{\varepsilon}(\mathbf{k})\right)\right)\right] \right| \tag{2.14}$$

$$\leq \varepsilon 2Cn^{-1/2}\left\{\sum_{p<i\leq n}\frac{1}{\sigma_i^2}+\sum_{p<i\leq n}\frac{y_{i-1}^2}{\sigma_i^2}+\ldots+\sum_{p<i\leq n}\frac{y_{i-p}^2}{\sigma_i^2}\right\}$$

$$\leq \varepsilon\frac{2C}{b_0}n^{1/2}\left\{1+\frac{1}{n}\sum_{p<i\leq n}y_{i-1}^2+\frac{1}{n}\sum_{p<i\leq n}y_{i-2}^2+\ldots+\frac{1}{n}\sum_{p<i\leq n}y_{i-p}^2\right\}.$$

Observe also that

$$\sup_{\mathbf{s}\in\mathbf{T}(\mathbf{k})}\left|\sum_{p<i\leq n}\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\mathbf{s})\right\}-\sum_{p<i\leq n}\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}^*(\mathbf{k}))\right\}\right|$$

$$\leq \sum_{p<i\leq n}\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}(\mathbf{k}))\right\}-\sum_{p<i\leq n}\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}^*(\mathbf{k}))\right\}.$$

Thus we have

$$(2.15)\ \sup_{0\leq t<\infty}\max_{0\leq k_0,\ldots,k_p\leq K+1}\sup_{\mathbf{s}\in\mathbf{T}(\mathbf{k})}\left|\sum_{p<i\leq n}\left[\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+tn^{-1/2}\gamma_i(\mathbf{s})\right)\right]\right.$$

$$\left.-\sum_{p<i\leq n}\left[\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}(\mathbf{k}))\right\}-F\left(t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}(\mathbf{k}))\right)\right]\right|$$

$$\leq \epsilon\frac{4C}{b_0}n^{1/2}\left\{1+\frac{1}{n}\sum_{p<i\leq n}y_{i-1}^2+\frac{1}{n}\sum_{p<i\leq n}y_{i-2}^2+\ldots+\frac{1}{n}\sum_{p<i\leq n}y_{i-p}^2\right\}$$

$$+\sup_{0\leq t<\infty}\max_{0\leq k_0,\ldots,k_p\leq K+1}\sup_{\mathbf{s}\in\mathbf{T}(\mathbf{k})}\left|\sum_{p<i\leq n}\left[\mathbf{I}\left\{\epsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}(\mathbf{k}))\right\}-F\left(t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}(\mathbf{k}))\right)\right]\right|$$

$$+\sup_{0\leq t<\infty}\max_{0\leq k_0,\ldots,k_p\leq K+1}\sup_{\mathbf{s}\in\mathbf{T}(\mathbf{k})}\left|\sum_{p<i\leq n}\left[\mathbf{I}\left\{\epsilon_i^2\leq t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}^*(\mathbf{k}))\right\}-F\left(t+tn^{-1/2}\gamma_i(\boldsymbol{\varepsilon}^*(\mathbf{k}))\right)\right]\right|.$$

Applying (2.12) we see from (2.15) that Lemma 2.3 will be proved once we have verified that for any fixed $\mathbf{s}\in\mathbf{T}$

$$(2.16)\quad \sup_{0\leq t<\infty}\left|n^{-1/2}\sum_{p<i\leq n}\left[\mathbf{I}\left\{\varepsilon_i^2\leq t+tn^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+tn^{-1/2}\gamma_i(\mathbf{s})\right)\right]\right|=o_P(1).$$

In the following $\mathbf{s}\in\mathbf{T}$ is therefore fixed.

Since F is continuous, $\xi_i=F(\varepsilon_i^2)$ is uniform on $[0,1]$. Also

$$(2.17)\qquad \left|F\left(t+tn^{-1/2}\gamma_i(\mathbf{s})\right)-F(t)\right|=tf(\xi_i(t))n^{-1/2}|\gamma_i(\mathbf{s})|,$$

where $\xi_i(t)$ is between t and $t+tn^{-1/2}\gamma_i(\mathbf{s})$. Using the definition of $\gamma_i(\mathbf{s})$ and (2.7), we obtain

$$n^{-1/2}\max_{p<i\leq n}|\gamma_i(\mathbf{s})|=o_P(1).$$

Hence for any $\delta > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left\{ \max_{p<i\le n} \sup_{M\le t<\infty} tf(\xi_i(t)) > \delta \right\} = 0, \tag{2.18}$$

and therefore the ergodic theorem for $(1/n)\sum_{p<i\le n} |\gamma_i(\mathbf{s})|$ gives

$$\lim_{M\to\infty} \limsup_{n\to\infty} P\left\{ \sup_{M\le t<\infty} \left| \sum_{p<i\le n} F\left(t + tn^{-1/2}\gamma_i(\mathbf{s}) - F(t)\right) \right| > \delta \right\} = 0 \tag{2.19}$$

for all $\delta > 0$. By (2.18), for any δ and η there are n_0 and M_0 such that

$$P\left\{ \max_{p<i\le n} \sup_{T\le t<\infty} tf(\xi_i(t)) > \delta \right\} \le \eta, \quad \text{if } M \ge M_0, n \ge n_0.$$

Hence by (2.17) we have

$$P\left\{ \sum_{p<i\le n} \mathbf{I}\left\{\xi_i \le F(t) - \delta n^{-1/2}|\gamma_i(\mathbf{s})|\right\} \right.$$

$$\left. \le \sum_{p<i\le n} \mathbf{I}\left\{\varepsilon_i^2 \le t + tn^{-1/2}\gamma_i(\mathbf{s})\right\} + \sum_{p<i\le n} \mathbf{I}\left\{\xi_i \le F(t) + \delta n^{-1/2}|\gamma_i(\mathbf{s})|\right\} \right\} \ge 1 - \eta,$$

if $M \ge M_0$ and $n \ge n_0$. By Theorem 1.1 of Koul (1991) we have that

$$\sup_{M_0\le t<\infty} \left| n^{-1/2} \sum_{p<i\le n} \left(\mathbf{I}\left\{\xi_i \le F(t) - \delta n^{-1/2}|\gamma_i(\mathbf{s})|\right\} - (F(t) - \delta n^{-1/2}|\gamma_i(\mathbf{s})|) \right) \right.$$

$$\left. - n^{-1/2} \sum_{p<i\le n} \left(\mathbf{I}\left\{\xi_i \le F(t)\right\} - F(t)\right) \right| = o_P(1)$$

and

$$\sup_{M_0\le t<\infty} \left(\mathbf{I}\left\{\xi_i \le F(t) + \delta n^{-1/2}|\gamma_i(\mathbf{s})|\right\} - (F(t) + \delta n^{-1/2}|\gamma_i(\mathbf{s})|) \right)$$

$$\left. - n^{-1/2} \sum_{p<i\le n} \left(\mathbf{I}\left\{\xi_i \le F(t)\right\} - F(t)\right) \right| = o_P(1).$$

By the ergodic theorem we have

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left\{ n^{-1/2}\delta \sum_{p<i\le n} n^{-1/2}|\gamma_i(\mathbf{s})| > c \right\} = 0$$

for all $c > 0$. Hence it is enough to show that for any $M > 0$

$$\sup_{0\le t\le M} \left| n^{-1/2} \sum_{p<i\le n} \left[\mathbf{I}\left\{\varepsilon_i^2 \le t + tn^{-1/2}\gamma_i(\mathbf{s})\right\} - F\left(t + tn^{-1/2}\gamma_i(\mathbf{s})\right) \right] \right. \tag{2.20}$$

$$-n^{-1/2}\sum_{p<i\le n}\left(\mathbf{I}\left\{\varepsilon_i^2\le t\right\}-F(t)\right)\Bigg| = o_P(1).$$

Instead of (2.20) we show that

$$\sup_{0\le t<\infty}\sup_{0\le u\le M}\Bigg|n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+un^{-1/2}\gamma_i(\mathbf{s})\right)\right] \tag{2.21}$$

$$-n^{-1/2}\sum_{p<i\le n}\left(\mathbf{I}\left\{\varepsilon_i^2\le t\right\}-F(t)\right)\Bigg| = o_P(1).$$

Write

$$\begin{aligned}
&n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+un^{-1/2}\gamma_i(\mathbf{s})\right)\right]\\
&=n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+un^{-1/2}\gamma_i(\mathbf{s})\right)\right]\mathbf{I}\left\{\gamma_i(\mathbf{s})\ge 0\right\}\\
&+n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+un^{-1/2}\gamma_i(\mathbf{s})\right)\right]\mathbf{I}\left\{\gamma_i(\mathbf{s})<0\right\}.
\end{aligned}$$

Suppose first that $\gamma_i(\mathbf{s})\ge 0$. Note that for any $u_j\le u\le u_{j+1}$

$$\begin{aligned}
0&\le\sum_{p<i\le n}\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-\sum_{p<i\le n}\mathbf{I}\left\{\varepsilon_i^2\le t+u_jn^{-1/2}\gamma_i(\mathbf{s})\right\} \qquad (2.22)\\
&\le\sum_{p<i\le n}\mathbf{I}\left\{\varepsilon_i^2\le t+u_{j+1}n^{-1/2}\gamma_i(\mathbf{s})\right\}-\sum_{p<i\le n}\mathbf{I}\left\{\varepsilon_i^2\le t+u_jn^{-1/2}\gamma_i(\mathbf{s})\right\}.
\end{aligned}$$

It is easy to see that there is a constant C such that

$$\begin{aligned}
0&\le\sum_{p<i\le n}\left[F(t+u_{j+1}n^{-1/2}\gamma_i(\mathbf{s}))-F(t+u_jn^{-1/2}\gamma_i(\mathbf{s}))\right] \qquad (2.23)\\
&\le C(u_{j+1}-u_j)n^{-1/2}\sum_{p<i\le n}|\gamma_i(\mathbf{s})|.
\end{aligned}$$

For any fixed $u\ge 0$, Theorem 1.1 of Koul (1991) yields

$$\sup_{0\le t<\infty}\Bigg|\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t+un^{-1/2}\gamma_i(\mathbf{s})\right\}-F\left(t+un^{-1/2}\gamma_i(\mathbf{s})\right)\right] \tag{2.24}$$

$$-n^{-1/2}\sum_{p<i\le n}\left[\mathbf{I}\left\{\varepsilon_i^2\le t\right\}-F(t)\right]\Bigg| = o_P(1).$$

In light of (2.22) and (2.23), relation (2.24) shows that (2.21) holds on the set $\{\gamma_i(\mathbf{s})\ge 0\}$. A similar argument applies to the set $\{\gamma_i(\mathbf{s})<0\}$. This completes the verification of (2.21). ∎

LEMMA 2.4 *If* (1.1) – (1.8) *hold, then*

$$\sup_{0\le t<\infty} |\hat{e}_n(t) - (e_n(t) + u_n(t))| = o_P(1),$$

where

$$u_n(t) = tf(t)n^{-1/2} \sum_{0\le k\le p} \beta_k \sum_{1\le j\le n} l_k(\varepsilon_j^2) f_k(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots).$$

PROOF. We use the inequality

$$|\hat{e}_n(t) - (e_n(t) + u_n(t))| \le |\hat{e}_{n,1}(t) - e_n(t)|$$

$$+ |\hat{e}_{n,2}(t) - h_n(t)| + |h_n(t) - u_n(t)|,$$

where h_n is given by (1.10).

Lemmas 2.1 and 2.3 yield

$$\sup_{0\le t<\infty} |\hat{e}_{n,1}(t) - e_n(t)| = o_P(1).$$

By Lemma 2.2

$$\sup_{0\le t<\infty} |e_{n,2}(t) - u_n^*(t)| = o_P(1).$$

To verify that

(2.25)
$$\sup_{0\le t<\infty} |u_n(t) - h_n(t)| = o_P(1)$$

observe that

$$\begin{aligned}
h_n(t) &= tf(t)n^{-1/2} \sum_{1\le j\le n} l_0(\varepsilon_j^2) f_0(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) \left(\frac{1}{n} \sum_{p<i\le n} \frac{1}{\sigma_i^2} \right) \\
&+ tf(t)n^{-1/2} \sum_{1\le j\le n} l_1(\varepsilon_j^2) f_1(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) \left(\frac{1}{n} \sum_{p<i\le n} \frac{y_{i-1}^2}{\sigma_i^2} \right) \\
&\vdots \\
&+ tf(t)n^{-1/2} \sum_{1\le j\le n} l_p(\varepsilon_j^2) f_p(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) \left(\frac{1}{n} \sum_{p<i\le n} \frac{y_{i-p}^2}{\sigma_i^2} \right).
\end{aligned}$$

Relation (2.25) now follows from (1.2), (1.4) – (1.6) and the relations $n^{-1} \sum_{p<i\le n} \sigma_i^{-2} y_{i-k}^2 \overset{a.s.}{\to} \beta_k$, $k = 1, \ldots p$ and an analogous relation for $k = 0$. ∎

PROOF OF THEOREM 1.1: In light of Lemma 2.4 it is enough to prove the weak convergence of $e_n(t) + u_n(t)$. We need tightness and the convergence of the finite dimensional distributions. The verification of tightness is easy. Since $e_n(t)$ is an empirical process of independent, identically distributed random variables, its tightness has been established, see e.g. Billingsley (1968). We showed in the proof of Lemma 2.1 that

$$\max_{0\le i\le p} n^{-1/2} \left| \sum_{1\le j\le n} l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) \right| = O_P(1),$$

which gives the tightness of $u_n(t)$.

The proof of the convergence of the finite dimensional distribution is based on the Cramér–Wold device. Let $\lambda_1, \lambda_2, \ldots, t_1, t_2, \ldots$ be constants and define

$$Z(k) = \sum_{1 \le i \le k} \zeta_i,$$

where

$$\zeta_i = \sum_{1 \le j \le M} \lambda_j \left[\left(\mathbf{I}\left\{ \varepsilon_i^2 \le t_j \right\} - F(t_j) \right) + t_j f(t_j) \sum_{0 \le k \le p} \beta_p l_k(\varepsilon_i^2) f_k(\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots) \right],$$

so that $n^{-1/2} Z(n) = \sum_{1 \le j \le M} \lambda_j \left[e_n(t_j) + u_n(t_j) \right]$. Recall that, as in the proof of Lemma 2.1, $\mathcal{F}_k$ is the σ–algebra generated by $\varepsilon_k, \varepsilon_{k-1}, \ldots$. We use again Theorem 3.5 of Hall and Heyde (1980). It is clear that $E\zeta_i = 0$. We will show that with $\eta^2 = E\zeta_1^2$

$$\max_{1 \le i \le n} E\left(\zeta_i^2 | \mathcal{F}_{i-1} \right) = o_P(n^{1/2}), \tag{2.26}$$

$$E \left| \sum_{1 \le i \le n} E\left(\zeta_i^2 | \mathcal{F}_{i-1} \right) - n\eta^2 \right| = o(n) \tag{2.27}$$

and

$$E \left| \sum_{1 \le i \le n} \zeta_i^2 - n\eta^2 \right| = o(n). \tag{2.28}$$

Since the first term in ζ_i is bounded by $2 \sum_{1 \le j \le M} |\lambda_j|$, (2.26) follows from (2.3).

Setting for brevity $f_k^-(i) = f_k(\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$, note that

$$\begin{aligned}
\zeta_j^2 &= \left(\sum_{1 \le j \le M} \lambda_j \left[\mathbf{I}\left\{ \varepsilon_i^2 \le t_j \right\} - F(t_j) \right] \right)^2 \\
&+ 2 \sum_{1 \le u,v \le M} \lambda_u \lambda_v \sum_{0 \le k \le p} \beta_k \left[\mathbf{I}\left\{ \varepsilon_i^2 \le t_u \right\} - F(t_u) \right] t_v f(t_v) l_k(\varepsilon_i^2) f_k^-(i) \\
&+ \sum_{1 \le u,v \le M} \lambda_u \lambda_v t_u t_v f(t_u) f(t_v) \sum_{0 \le k,m \le p} \beta_k \beta_m l_k(\varepsilon_i^2) l_m(\varepsilon_i^2) f_k^-(i) f_m^-(i)
\end{aligned}$$

and therefore

$$\begin{aligned}
E\left(\xi_i^2 | \mathcal{F}_{i-1} \right) &= E \left(\sum_{1 \le j \le M} \lambda_j \left[\mathbf{I}\left\{ \varepsilon_i^2 \le t_j \right\} - F(t_j) \right] \right)^2 \\
&+ 2 \sum_{1 \le u,v \le M} \lambda_u \lambda_v \sum_{0 \le k \le p} \beta_k f_k^-(i) t_v f(t_v) E\left(\left[\mathbf{I}\left\{ \varepsilon_i^2 \le t_u \right\} - F(t_u) \right] l_k(\varepsilon_i^2) \right) \\
&+ \sum_{1 \le u,v \le M} \lambda_u \lambda_v t_u t_v f(t_u) f(t_v) \sum_{0 \le k,m \le p} \beta_k \beta_m f_k^-(i) f_m^-(i) E\left(l_k(\varepsilon_i^2) l_m(\varepsilon_i^2) \right).
\end{aligned}$$

Relations (2.27) and (2.28) follow now from the ergodic theorem. By computing $E\zeta_i^2$ we obtain the covariance structure in Theorem 1.1. ∎

3 Asymptotic linearity of estimators

In this section we consider several examples of estimators satisfying (1.3). We would like to point out that, in general, asymptotic linearity like (1.3) is usually not difficult to establish whenever asymptotic normality holds.

3.1 Pseudo maximum likelihood estimation

Let

$$\mathcal{L}(\mathbf{s}) = -\frac{1}{2}\sum_{1\le i\le n}\log\sigma_i^2(\mathbf{s}) - \frac{1}{2}\sum_{1\le i\le n}\frac{y_i^2}{\sigma_i^2\mathbf{s}}$$

denote the log of the pseudo likelihood function. The estimator is the solution of the equation $\mathcal{L}'(\hat{\mathbf{b}}) = \mathbf{0}$, where

$$\mathcal{L}'(\mathbf{s}) = \frac{1}{2}\sum_{1\le i\le n}\left(\frac{y_i^2}{\sigma_i^2(\mathbf{s})} - 1\right)\frac{1}{\sigma_i^2(\mathbf{s})}\frac{\partial\sigma_i^2(\mathbf{s})}{\partial(\mathbf{s})}.$$

As pointed out in Section 4.1 of Gouriéroux (1997) $\hat{\mathbf{b}}$ is asymptotically normal under standard regularity conditions even if the ε_i are not standard normal, i.e. conditionally on the past, the observations are not necessarily normal. This holds true for time series models much more general than the ARCH(p) considered here. In the following we assume only that

$$\left|\left|\hat{\mathbf{b}} - \mathbf{b}\right|\right| = O_P(n^{-1/2}), \tag{3.1}$$

where $\mathbf{b}$ is the true value of the parameter vector. The second derivative of $\mathcal{L}(\mathbf{s})$ is the matrix $\mathcal{L}''(\mathbf{s})$. By the ergodic theorem

$$\left|\frac{1}{n}\mathcal{L}''(\mathbf{b}) - \mathcal{J}(\mathbf{b})\right| \xrightarrow{P} 0, \tag{3.2}$$

where $\mathcal{J}(\mathbf{b})$ is a deterministic matrix. Since $\mathcal{L}'(\hat{\mathbf{b}}) - \mathcal{L}'(\mathbf{b}) = -\mathcal{L}'(\mathbf{b})$, by the mean value theorem, we have

$$\mathcal{L}''(\boldsymbol{\beta})\left(\hat{\mathbf{b}} - \mathbf{b}\right) = -\mathcal{L}'(\mathbf{b}), \tag{3.3}$$

where $\boldsymbol{\beta}$ is a point between $\hat{\mathbf{b}}$ and $\mathbf{b}$. The matrix $\mathcal{J}(\mathbf{b})$ is invertible, see e.g. Gouriéroux (1997) pp. 50–51, so (3.3) yields

$$\hat{\mathbf{b}} - \mathbf{b} = -\frac{1}{n}\mathcal{J}^{-1}(\mathbf{b})\mathcal{L}'(\mathbf{b}) - \mathcal{J}^{-1}(\mathbf{b})\left(\frac{1}{n}\mathcal{L}''(\boldsymbol{\beta}) - \mathcal{J}(\mathbf{b})\right)(\hat{\mathbf{b}} - \mathbf{b}). \tag{3.4}$$

To see that the second term on the right–hand side of (3.4) is $o_P(n^{-1/2})$ use (3.1) and the decomposition

$$\frac{1}{n}\mathcal{L}''(\boldsymbol{\beta}) - \mathcal{J}(\mathbf{b}) = \frac{1}{n}\left(\mathcal{L}''(\boldsymbol{\beta}) - \mathcal{L}''(\mathbf{b})\right) + \left(\frac{1}{n}\mathcal{L}''(\mathbf{b}) - \mathcal{J}(\mathbf{b})\right). \tag{3.5}$$

The first term on the RHS of (3.5) is $o_P(1/n)$ because $\boldsymbol{\beta} \xrightarrow{P} \mathbf{b}$ whereas the second is $o_P(1)$ by (3.2). Representation (1.3) now follows since

$$\mathcal{L}'(\mathbf{b}) = \frac{1}{2}\sum_{1\le i\le n}\left(\varepsilon_i^2 - 1\right)\frac{1}{\sigma_i^2(\mathbf{b})}\left[\frac{\partial\sigma_i^2(\mathbf{s})}{\partial\mathbf{s}}\right]_{\mathbf{s}=\mathbf{b}}.$$

3.2 Conditional least squares

By definition, $E\,(y_k^2|\mathcal{F}_{k-1}) = \sigma_i^2(\mathbf{s})$, where $\sigma_i^2(\mathbf{s}) = s_0 + s_1 y_{i-1}^2 + \ldots s_p y_{i-p}^2$, so the conditional sum of squares is

$$Q(\mathbf{s}) = \sum_{p<i\leq n} \left(y_i^2 - \sigma_i^2(\mathbf{s})\right)^2,$$

The conditional least squares estimators of the b_k, $k = 0, 1, \ldots, p$, are the solutions of the equations

$$\frac{\partial Q}{\partial s_k} = -2 \sum_{p<i\leq n} \left(y_i^2 - \sigma_i^2(\mathbf{s})\right) \frac{\partial \sigma_i^2(\mathbf{s})}{\partial s_k} = 0, \quad k = 0, 1, \ldots, p.$$

In order to establish (1.3) we proceed similarly as in Subsection 3.1. Direct verification shows that $Q''(\mathbf{s})(= Q''(\mathbf{b}))$ does not depend on $\mathbf{s}$ and by the ergodic theorem

$$n^{-1} Q''(\mathbf{b}) \xrightarrow{P} \Lambda(\mathbf{b}). \tag{3.6}$$

Since $Q'(\hat{\mathbf{b}}) = \mathbf{0}$,

$$-Q'(\mathbf{b}) = Q''(\mathbf{b}) \left(\hat{\mathbf{b}} - \mathbf{b}\right). \tag{3.7}$$

Relations (3.7) and (3.6) in conjunction with a Central Limit Theorem for the squares of an ARCH process imply that $n^{1/2} \left(\hat{\mathbf{b}} - \mathbf{b}\right)$ is asymptotically normal. Consequently, arguing as in Subsection 3.1,

$$\hat{\mathbf{b}} - \mathbf{b} = -\frac{1}{n} \Lambda^{-1} Q'(\mathbf{b}) + o_P(n^{1/2}).$$

Representation (1.3) now follows on observing that

$$\left.\frac{\partial Q}{\partial s_0}\right|_{\mathbf{s}=\mathbf{b}} = -2 \sum_{p<i\leq n} \left(\varepsilon_i^2 - 1\right) \sigma_i^2(\mathbf{b})$$

and for $1 \leq k \leq p$

$$\left.\frac{\partial Q}{\partial s_0}\right|_{\mathbf{s}=\mathbf{b}} = -2 \sum_{p<i\leq n} \left(\varepsilon_i^2 - 1\right) \sigma_i^2(\mathbf{b}) y_{i-k}^2.$$

3.3 Conditional likelihood

The conditional distribution function of y_i^2/σ_i^2 given $\mathcal{F}_{i-1}$ is F. Thus the likelihood function is $\prod_{1\leq i\leq n} f(y_i^2/\sigma_i^2)$. Assuming that f is twice differentiable and proceeding along the lines of Subsection 3.1, one can show that (1.3) holds.

References

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

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theory and methods*, Second edn. Springer-Verlag, New York.

Burke, M. D., Csörgő, M., Csörgő, S. and Révész, P. (1979). Approximations of the empirical process when parameters are estimated. *The Annals of Probability*, **7,** 790–810.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Davis, R. A. and Mikosch, T. (1999). The sample autocorrelations of heavy–tailed processes with applications to ARCH. *The Annals of Statistics*, **26,** 2049–2080.

Durbin, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, **1,** 279–290.

Durbin, J. (1973b). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

Giraitis, L., Kokoszka, P. S. and Leipus, R. (1999). Stationary ARCH models: dependence structure and Central Limit Theorem. *Econometric Theory*, **00,** 000–000.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

He, C. and Teräsvirta, T. (1998). Fourth moment structure of the GARCH(p, q) process. forthcoming in Econometric Theory.

Horváth, L. and Kokoszka, P. (1999). Large sample theory for linear functionals of squared residual correlations of the ARCH(p) model. preprint.

Koul, H. L. (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference*, **29,** 291–308.

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.

Lu, Z. and Cheng, P. (1997). Distribution–free strong consistency for nonparametric kernel regression involving nonlinear time series. *Journal of Statistical Planning and Inference*, **65,** 67–86.

Lundbergh, S. and Teräsvirta, T. (1998). Evaluating GARCH models. Working paper No. 292, Stockholm School of Economics.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

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

Tjøstheim, D. (1999). Nonparametric specification procedures for time series. In *Asymptotics, Nonparametrics and Time Series, New York* (ed. S. Ghosh), pp. 149–200. Marcel Dekker.