Empirical process of the squared residuals of an ARCH sequence*

Lajos Horváth
University of Utah

Piotr Kokoszka
University of Liverpool

Gilles Teyssière
The European Commission and GREQAM

We derive the asymptotic distribution of the sequential empirical process of the squared residuals of an ARCH(p) sequence. Unlike the residuals of an ARMA process, these residuals do not behave in this context like asymptotically independent random variables, and the asymptotic distribution involves a term depending on the parameters of the model. We show that in certain applications, including the detection of changes in the distribution of the unobservable innovations, our result leads to asymptotically distribution free statistics.

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

Running head: Empirical process for ARCH

AMS 2000 subject classification: Primary 62G30, Secondary 62G20

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 complex 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 (2001) 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.

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

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

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

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

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

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

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

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

We assume that the parameter vector $\mathbf{b}$ is estimated from a sample $y_1, y_2, \dots, y_n$ by an estimator $\hat{\mathbf{b}}_n = (\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)$ which admits the representation

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

The functions l_i and f_i are regular in the sense that

$$El_i(\varepsilon_0^2) = 0, \quad 0 \le i \le p, \tag{1.4}$$

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

and

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

We show in Section 2 that commonly used estimators, see e.g. Chapter 4 of Gouriéroux (1997), admit representation (1.3).

The squared residuals are defined as

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

where

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

In this paper we study the weak convergence of the sequential (or two–time parameter) empirical process of the squared residuals

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

where

$$\hat{F}_n(t,s) = \frac{1}{ns} \sum_{p<k\le ns} \mathbf{I}\left\{\hat{\varepsilon}_k^2 \le t\right\}, \qquad p/n < s \le 1$$

and $\hat{F}_n(t,s) = 0$ if $0 \le s \le p/n$. We note that $\hat{e}_n(t,1)$ is the empirical process of the $\hat{\varepsilon}_k^2$, $p < k \le n$.

Following Giraitis *et al.* (2000) 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 (1999). It is also well–known that ARCH(p) and more general sequences from the GARCH family are not only ergodic but also mixing with geometric rate, we refer to Lu and Cheng (1997), Remark 4.2, for further 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:

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

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

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

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

$$\begin{aligned} r(t,t',s,s') &= (s \wedge s')\left(F(t \wedge t') - F(t)F(t')\right) \\ &+ tf(t)ss' \sum_{0 \le i \le p} \beta_i g_i(t') + t'f(t')ss' \sum_{0 \le i \le p} \beta_i g_i(t) \\ &+ tf(t)t'f(t')ss' \sum_{0 \le i,j \le p} \beta_i \gamma_{ij} \beta_j, \end{aligned}$$

where $s \wedge t = \min(s,t)$.

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

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

where the convergence is in the Skorokhod space $\mathcal{D}([0,\infty] \times [0,1])$ *and* Γ *is a Gaussian proces with*

$$E\Gamma(t,s) = 0 \quad \text{and} \quad E\left[\Gamma(t,s)\Gamma(t',s')\right] = r(t,t',s,s').$$

The convergence in Theorem 1.1 is equivalent to the convergence $\hat{e}_n(F^{-1}(x), s) \to \Gamma(F^{-1}(x), s)$, $0 \le x, s \le 1$, in $\mathcal{D}([0,1] \times [0,1])$. ($F^{-1}(0) \ge -\infty$ and $F^{-1}(1) \le \infty$ denote the end points of the support of F.)

Before presenting detailed proofs, we outline the argument. Set

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

and observe that

$$\sup_{0 \le t < \infty} \sup_{0 \le s \le 1} |\hat{e}_n(t,s) - (\hat{e}_{n,1}(t,s) + \hat{e}_{n,2}(t,s))| = O_P(n^{-1/2}),$$

where

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

and

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

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

$$e_n(t,s) = n^{-1/2} \sum_{p < i \le ns} \left(\mathbf{I}\left\{ \varepsilon_i^2 \le t \right\} - F(t) \right) \tag{1.9}$$

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

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

Observe that

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

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

$$= ns \left\{ (\hat{b}_0 - b_0)\beta_0 + (\hat{b}_1 - b_1)\beta_1 + \ldots + (\hat{b}_p - b_p)\beta_p + o_P(n^{-1/2}) \right\}.$$

Hence

$$\sup_{0 \le t < \infty} \sup_{0 \le s \le 1} \left| \hat{e}_n(t,s) - \left(e_n(t,s) + tf(t)s \sum_{0 \le i \le p} n^{1/2}(\hat{b}_i - b_i)\beta_i \right) \right| = o_P(1) \tag{1.11}$$

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

Relation (1.11) explains the structure of $\hat{e}_n(t,s)$ and also the formula for the covariance function of the limiting Gaussian process. It is interesting to compare (1.11) with Theorem 1 of Bai (1994) which shows that for the residuals in an ARMA model $\hat{e}_n(t,s) - e_n(t,s)$ is uniformly $o_P(1)$. Boldin (1998) showed that an additional term, analogous to $tf(t)s\sum_{0 \le i \le p} n^{1/2}(\hat{b}_i - b_i)\beta_i$, makes an asymptotic contribution to the empirical process of residuals (not squared residuals) of an ARCH(1) model.

The covariance function $r(\cdot)$ depends on several unknown parameters and functions. In Section 2 we obtain explicit formulas for l_i, f_i, $0 \le i \le p$, in the case of most commonly used estimators. Observe that $\alpha_i, g_i(t)$ and β_i are expected values and can be consistently estimated by the corresponding averages. Note also that $\{\gamma_{ij},\ 0 \le i, j \le p\}$ is the asymptotic covariance matrix of $n^{1/2}\left(\hat{\mathbf{b}}_n - \mathbf{b}\right)$ and its estimation is discussed in Gouriéroux (1997). In some applications, see Section 3, it is not necessary to estimate the parameters in $r(\cdot)$.

We will show in the proof of Theorem 1.1 that

$$\left\{ e_n(t,s),\ n^{1/2} \sum_{0 \le i \le p} \left(\hat{b}_i - b_i\right) \beta_i,\ 0 \le t < \infty, 0 \le s \le 1 \right\}$$

converges weakly to a Gaussian vector valued process $\{K(F(t), s), \xi,\ 0 \le t < \infty, 0 \le s \le 1\}$ with

$$E\left[K(F(t),s)\right] = 0, \quad E\xi = 0, \tag{1.12}$$

$$E\left[K(F(t),s)K(F(t'),s')\right] = (s \wedge s')\left(F(t \wedge t') - F(t)F(t')\right), \tag{1.13}$$

$$E\xi^2 = \sum_{0 \le i,j \le p} \beta_i \gamma_{ij} \beta_j \tag{1.14}$$

$$E\left[K(F(t),s)\xi\right] = s \sum_{0 \le i \le p} \beta_i g_i(t). \tag{1.15}$$

Observe that $\{K(x,s), 0 \le x, s \le 1\}$ is a Kiefer process. Also, the limit process $\Gamma(t,s)$ in Theorem 1.1 can be written as

$$\{\Gamma(t,s),\ 0 \le t < \infty, 0 \le s \le 1\} \stackrel{d}{=} \{K(F(t),s) + tf(t)s\xi,\ 0 \le t < \infty, 0 \le s \le 1\}. \tag{1.16}$$

In particular, the process $\Gamma(\cdot, 1)$ admits the representation

$$\{\Gamma(t,1),\ 0 \le t < \infty\} \stackrel{d}{=} \{B(F(t)) + tf(t)\xi,\ 0 \le t < \infty\}, \tag{1.17}$$

where $\{B(x),\ 0 \le x \le 1\}$ is a Brownian bridge.

The results in (1.11), (1.16) and (1.17) are similar to main theorems on parameter estimated processes. Durbin (1973a, 1973b) was the first who considered the weak convergence of the 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 (1.17), but in the iid case, ξ is a stochastic integral of a deterministic function with respect to $B(F(\cdot))$. Our case is somewhat different.

In the next section we consider three examples when condition (1.3) is satisfied. Section 3 discusses some applications of Theorem 1.1 whose proof is postponed until Section 4.

2 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.

In the remainder of this section we use the notation $\mathbf{s} = (s_0, s_1, \ldots, s_p) \in [0, \infty)^{p+1}$ and

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

2.1 Pseudo maximum likelihood estimation

Let

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

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

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

As pointed out in Section 4.1 of Gouriéroux (1997) $\hat{\mathbf{b}}_n$ 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

$$\left|\left|\hat{\mathbf{b}}_n - \mathbf{b}\right|\right| = O_P(n^{-1/2}), \tag{2.1}$$

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

$$\left|\frac{1}{n}\mathcal{L}_n''(\mathbf{b}) - \mathcal{J}(\mathbf{b})\right| \xrightarrow{P} 0, \tag{2.2}$$

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

$$\mathcal{L}_n''(\boldsymbol{\beta})\left(\hat{\mathbf{b}}_n - \mathbf{b}\right) = -\mathcal{L}_n'(\mathbf{b}), \tag{2.3}$$

where $\boldsymbol{\beta}$ is a point between $\hat{\mathbf{b}}_n$ and $\mathbf{b}$. The matrix $\mathcal{J}(\mathbf{b})$ is invertible, see e.g. Gouriéroux (1997) pp. 50–51, so (2.3) yields

$$\hat{\mathbf{b}}_n - \mathbf{b} = -\frac{1}{n}\mathcal{J}^{-1}(\mathbf{b})\mathcal{L}_n'(\mathbf{b}) - \mathcal{J}^{-1}(\mathbf{b})\left(\frac{1}{n}\mathcal{L}_n''(\boldsymbol{\beta}) - \mathcal{J}(\mathbf{b})\right)(\hat{\mathbf{b}}_n - \mathbf{b}). \tag{2.4}$$

To see that the second term on the right–hand side of (2.4) is $o_P(n^{-1/2})$ use (2.1) and the decomposition

$$\frac{1}{n}\mathcal{L}_n''(\boldsymbol{\beta}) - \mathcal{J}(\mathbf{b}) = \frac{1}{n}\left(\mathcal{L}_n''(\boldsymbol{\beta}) - \mathcal{L}_n''(\mathbf{b})\right) + \left(\frac{1}{n}\mathcal{L}_n''(\mathbf{b}) - \mathcal{J}(\mathbf{b})\right). \tag{2.5}$$

The first term on the right-hand sideof (2.5) is $o_P(1/n)$ because $\boldsymbol{\beta} \xrightarrow{P} \mathbf{b}$ whereas the second is $o_P(1)$ by (2.2). Representation (1.3) now follows since

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

2.2 Conditional least squares

By definition$E\left(y_k^2|\mathcal{F}_{k-1}\right) = \sigma_i^2(\mathbf{s})$, so the conditional sum of squares is

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

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

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

In order to establish (1.3) we proceed similarly as in Subsection 2.1. Direct verification shows that $Q_n''(\mathbf{s}) (= Q_n''(\mathbf{b}))$ does not depend on $\mathbf{s}$ and by the ergodic theorem

$$n^{-1} Q_n''(\mathbf{b}) \xrightarrow{P} \Lambda(\mathbf{b}). \tag{2.6}$$

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

$$-Q_n'(\mathbf{b}) = Q_n''(\mathbf{b}) \left(\hat{\mathbf{b}}_n - \mathbf{b} \right). \tag{2.7}$$

Relations (2.7) and (2.6) in conjunction with a Central Limit Theorem for the squares of an ARCH process imply that $n^{1/2} \left(\hat{\mathbf{b}}_n - \mathbf{b} \right)$ is asymptotically normal. Consequently, arguing as in Subsection 2.1,

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

Representation (1.3) now follows on observing that

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

and for $1 \le k \le p$

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

2.3 Conditional likelihood

Let h denote the density function of ε_0. The conditional density of y_i given $\mathcal{F}_{i-1}$ is $h(t/\sigma_i)/\sigma_i$. Hence the conditional log likelihood function is

$$\mathcal{L}_n^*(\mathbf{s}) = -\frac{1}{2} \sum_{1\le i\le n} \log \sigma_i^2(\mathbf{s}) + \sum_{1\le i\le n} \log h(y_i/\sigma_i(\mathbf{s})).$$

If ε_0 is normal, then $\mathcal{L}_n^*$ coincides with $\mathcal{L}_n$ in Subsection 2.1. The conditional likelihood estimator $\hat{\mathbf{b}}_n^*$ is the solution to $\mathcal{L}_n^{*\prime}(\hat{\mathbf{b}}_n^*) = \mathbf{0}$, where

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

Assuming that $n^{-1} E \mathcal{L}_n^{*\prime\prime}(\mathbf{s}) \to \mathcal{L}^{*\prime\prime}(\mathbf{s})$ and $\mathcal{L}^{*\prime\prime}(\mathbf{b})$ is a positive definite matrix we obtain that $\mathcal{L}^*(\mathbf{b}) > \mathcal{L}^*(\mathbf{s})$ in a neighbourhood of $\mathbf{b}$ where $\mathcal{L}^*(\mathbf{s}) = \lim_{n\to\infty} \mathcal{L}_n^*(\mathbf{s})/n$. This implies that

$$\left\| \hat{\mathbf{b}}_n^* - \mathbf{b} \right\| = o_P(1), \tag{2.8}$$

c.f. Amemiya (1985). Standard arguments show that (2.8) implies that (2.1) also holds for $\hat{\mathbf{b}}_n^*$ and therefore the arguments used in Subsection 2.1 give

$$\hat{\mathbf{b}}_n^* - \mathbf{b} = -\left[\mathcal{L}^{*\prime\prime}(\mathbf{b}) \right]^{-1} \frac{1}{n} \mathcal{L}_n^{*\prime}(\mathbf{b}) + o_P(n^{-1/2}).$$

Since

$$\mathcal{L}_n^{*\prime}(\mathbf{b}) = -\frac{1}{2} \sum_{1 \le i \le n} \left\{ \varepsilon_i \frac{h'(\varepsilon_i)}{h(\varepsilon_i)} + 1 \right\} \frac{1}{\sigma_i^2(\mathbf{b})} \left(1, y_{i-1}^2, \ldots, y_{i-p}^2\right),$$

we showed that (1.3) holds.

3 Some applications

Similarly to the parameter estimated process, the limit of $\hat{e}_n(t,s)$ depends on the unknown parameter $\mathbf{b}$ and the correlation between $\mathbf{I}\{\varepsilon_i^2 \le t\}$ and $\mathbf{b}$. The martingale approach of Khamaladze (1981) was used to transform the parameter estimated process so that the limit is a Brownian motion and therefore distribution free. The transformation is based on a martingale representation of the empirical process and in case of parameter estimation the transformation must be constructed from the data. Khamaladze's martingale approach was extended by Koul and Stute (1999) to regression and autoregression. These authors use an m-dependent version of (1.3), corresponding to $f_i(\cdot) \equiv 1$, $l_i = l_i(\varepsilon_j^2, \ldots, \varepsilon_{j-m}^2)$, $1 \le i \le p$. It is not immediately clear how the martingale approach can be adapted to ARCH sequences. Csörgő (1983) and Horváth (1985) suggested kernel–transforms of the parameter estimated empirical process to get a distribution (parameter) free limit. Their approach uses the covariance function (or its estimator), so, in principle, it might be used to transform $\hat{e}_n(t,s)$. However, the required calculation of the eigenfunctions and eigenvalues of the estimated covariance function is computationally unreasonable.

Next we consider some simple and immediate consequences of Theorem 1.1.

Example 3.1 The detection of possible changes in the structure of observation has been extensively studied in various settings, see e.g Csörgő and Horváth (1997). Chu (1995) and Mikosch and Stărică (1999) investigated changes in GARCH models. Theorem 1.1 can be used to test for a possible change in the distribution of the innovations in an ARCH model. Following the method described in detail in Section 2.6 of Csörgő and Horváth (1997), the process

$$w_n(t,s) = \begin{cases} 0, & 0 \le s \le p/n \\ \frac{[ns](n-[ns])}{n^{3/2}} \left(\hat{F}_n(t,s) - \hat{F}_n^*(t,s)\right), & p/n < s \le (n-1)/n \\ 0 & (n-1)/n < s \le 1, \end{cases}$$

with

$$\hat{F}_n^*(t,s) = \frac{1}{n - ns} \sum_{ns < i \le n} \mathbf{I}\left\{\hat{\varepsilon}_i^2 \le t\right\},$$

compares the empirical distribution function of $\hat{\varepsilon}_{p+1}^2, \ldots, \hat{\varepsilon}_{[ns]}^2$ to that of $\hat{\varepsilon}_{[ns]+1}^2, \ldots, \hat{\varepsilon}_n^2$. Observe that $w_n(t,s)$ has the same limit in $\mathcal{D}([0,\infty] \times [0,1])$ as $\hat{e}_n(t,s) - s\hat{e}_n(t,1)$ which by (1.11) has the same limit as $e_n(t,s) - se_n(t,1)$ (the contributions from the function h_n defined in (1.10) cancel). Thus Theorem 1.1 implies that $w_n(t,s)$ converges in $\mathcal{D}([0,\infty] \times [0,1])$ to $K^*(F(t),s)$, where $K^*(u,s)$ is a "tied-down" Kiefer process, i.e. a Gaussian process with $E[K^*(u,s)] = 0$ and

$$E[K^*(u,s)K^*(u',s')] = (u \wedge u' - uu')(s \wedge s' - ss')$$

Since F is continuous,

$$\int_0^1 \int_0^\infty w_n^2(t,s) d\hat{F}_n(t,1) ds \xrightarrow{d} \int_0^1 \int_0^1 \left[K^*(u,s)\right]^2 du ds$$

and

$$\sup_{0\le t<\infty,\ 0\le s\le 1} |w_n(t,s)| \xrightarrow{d} \sup_{0\le u,s\le 1} |K^*(u,s)|\,.$$

Blum *et al.* (1961), Cotteril and Csörgő (1985) and Martynov (1992) studied the distribution function of $\int_0^1\int_0^1 [K^*(t,s)]^2 dtds$. They computed its characteristic function and inverting the characteristic function one can compute the distribution function at points of interest. Tables for the distribution function of $\int_0^1\int_0^1 [K^*(t,s)]^2 dtds$ can be found in Blum *et al.* (1961) and Cotteril and Csörgő (1985). The distribution of $\sup_{0\le u,s\le 1}|K^*(u,s)|$ is tabulated in Picard (1985).

EXAMPLE 3.2 In this example we consider sums of functions of the squared residuals. Let $\psi(\cdot)$ be a function on $[0,\infty)$ with finite variation on $[0,\infty)$. Theorem 1.1 and integration by parts give

$$\int_0^\infty \psi(t)d\hat{e}_n(t,1) \xrightarrow{d} \int_0^\infty \psi(t)d\Gamma(t,1).$$

If

$$\int_0^\infty \psi(t)d(tf(t)) = 0, \tag{3.1}$$

then by (1.17) we have

$$\frac{n^{1/2}\left(\frac{1}{n}\sum_{p<i\le n}\psi(\hat{\varepsilon}_i^2) - \int_0^\infty \psi(t)dF(t)\right)}{\left(\mathrm{Var}\psi(\varepsilon_0^2)\right)^{1/2}} \xrightarrow{d} N(0,1), \tag{3.2}$$

where $N(0,1)$ stands for the standard normal random variable. We note that neither condition (3.1) nor the norming constants in (3.2) depend on the unknown $\mathbf{b}$.

EXAMPLE 3.3 In this example a very simple χ^2-test is suggested. It can be used to check if the distribution function of $\hat{\varepsilon}_1^2,\ldots,\hat{\varepsilon}_{[ns]}^2$ is F, for any $0<s\le 1$. Let $M\ge 1$ be an integer and $0<s<1$. Define $0\le t_0<t_1<t_2<\ldots<t_M\le\infty$. By Theorem 1.1, the vector $\mathbf{Z}_M=(Z_{1,M},\ldots,Z_{M,M})$,

$$\begin{aligned} Z_{i,M} &= n^{1/2}\left[\frac{1}{ns}\sum_{p\le j\le ns}\left(\mathbf{I}\left\{t_{i-1}\le\hat{\varepsilon}_j^2<t_i\right\}-(F(t_i)-F(t_{i-1}))\right)\right.\\ &\left.-\ \frac{1}{n(1-s)}\sum_{ns<j\le n}\left(\mathbf{I}\left\{t_{i-1}\le\hat{\varepsilon}_j^2<t_i\right\}-(F(t_i)-F(t_{i-1}))\right)\right], \end{aligned}$$

is asymptotically normal with zero mean and covariance matrix $[s(1-s)]^{-1}\boldsymbol{D}$, $\boldsymbol{D}=\{d_{ij},\ 1\le i,j\le M\}$, where

$$d_{ij}=\begin{cases} (F(t_i)-F(t_{i-1}))-(F(t_i)-F(t_{i-1}))^2, & i=j\\ -(F(t_i)-F(t_{i-1}))(F(t_j)-F(t_{j-1})), & i\ne j.\end{cases}$$

It is well–known (just observe that $\boldsymbol{D}$ is the covariance matrix of $e_n(t_{i-1},1)-e_n(t_i,1),\ 1\le i\le M$) that the rank of $\boldsymbol{D}$ is $M-1$. If $\boldsymbol{D}^{-1}$ denotes a generalized inverse of $\boldsymbol{D}$ (c.f. Seber (1977), p. 76), then $s(1-s)\mathbf{Z}_M\boldsymbol{D}^{-1}\mathbf{Z}_M^T$ is asymptotically $\chi^2(M-1)$ (c.f. Theorem 2.8 in Seber (1977)).

Since the asymptotic distribution of $\hat{e}_n(t,1)$ depends on unknown parameters, classical goodness-of-fit tests, like the Kolmogorov-Smirnov and Cramer-von Mises tests, are not directly applicable. However, Theorem 1.1 shows that appropriate functionals of $\hat{e}_n(t,1)$ have an asymptotic distribution, and it may be hoped that bootstrap goodness-of-fit tests may be developed. In Horváth *et al.* (2000) we proposed and examined by means of a simulation study such tests for ARCH(p) and more general GARCH models. Bootstrap tests based on $\hat{e}_n(t,1)$ can, in principle, be expected to detect any departure from the null hypothesis of independent, identically distributed squared innovations ε_i^2 with specified distribution function F. In practice, however, certain alternatives, like, for example, the change-point alternative discussed in Example 3.1, are not reliably detected and it may be expected that the asymptotic test described in Example 3.1 will have higher power. Other alternatives, like, for example, distribution function F even slightly different from the postulated one, are detected with probability around 0.8 even for series of length 200 (financial time series based on intra-daily trading have often lengths of several thousand). The tests have close to perfect size even for series of length 50. We refer to Horváth *et al.* (2000) for further details.

It appears to be often the case that for GARCH and related models many important statistics are not asymptotically pivotal and alternative approaches like bootstrap and and response surface analysis must be used in such situations, see Frances and van Dijk (2000) for an analysis of a related problem of outlier detection.

4 Proof of Theorem 1.1

Set

$$\zeta_n = n^{-1/2} \sum_{0\le k\le p} \beta_k \sum_{1\le j\le n} l_k(\varepsilon_j^2) f_k(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) \tag{4.1}$$

and

$$u_n(t,s) = tf(t)s\zeta_n. \tag{4.2}$$

In light of (1.16), Theorem 1.1 follows from Lemmas 4.6 and 4.7 below. The proof of Lemma 4.7 is fairly standard. In order to establish Lemma 4.6 we need a number of auxiliary lemmas.

In the following, $||\cdot||$ stands for the maximum norm of a vector.

Lemma 4.1 *If conditions* (1.3) – (1.8) *hold, then*

$$\sqrt{n}\left|\left|\hat{\mathbf{b}}_n - \mathbf{b}\right|\right| = O_P(1). \tag{4.3}$$

Proof. Define $X_{j,i} = l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)$ and observe that for any $0 \le i \le p$, the variables $X_{j,i}$ are zero mean, uncorrelated and form a stationary sequence. Therefore $\mathrm{Var}[\sum_{1\le j\le n} X_{j,i}] = O(n)$ and so (4.3) follows from (1.3) and the Chebyshev inequality. ∎

Recall now the definition of the function $h_n(t,s)$ given in (1.10).

Lemma 4.2 *If* (1.1) – (1.8) *hold, then*

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

PROOF. By the mean value theorem

$$F(t\delta_i^2) - F(t) = f(\vartheta_i)t(\delta_i^2 - 1) = tf(t)(\delta_i^2 - 1) + t(f(\vartheta_i) - f(t))(\delta_i^2 - 1), \tag{4.4}$$

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

$$\begin{aligned}
\hat{e}_{n,2}(t,s) - h_n(t,s) &= n^{-1/2} \sum_{p<i\le ns} t\,(f(\vartheta_i) - f(t))\left(\delta_i^2 - 1\right) \\
&= n^{-1/2}\left(\hat{b}_0 - b_0\right) \sum_{p<i\le ns} \frac{1}{\sigma_i^2} t\,(f(t) - f(\vartheta_i)) \\
&+ n^{-1/2} \sum_{1\le k\le p} (\hat{b}_k - b_k) \sum_{p<i\le ns} \frac{y_{i-k}^2}{\sigma_i^2} t\,(f(t) - f(\vartheta_i)).
\end{aligned}$$

We will verify below that

$$\max_{p<i\le n} \sup_{0\le t<\infty} t|f(t) - f(\vartheta_i)| = o_P(1), \tag{4.5}$$

which together with Lemma 4.1 will imply that

$$\sup_{0\le s\le 1} \sup_{0\le t<\infty} |\hat{e}_{n,2}(t,s) - h_n(t,s)| = o_P(1)\frac{1}{n}\sum_{p<i\le n}\frac{1}{\sigma_i^2}\left[1 + \sum_{1\le k\le p} y_{i-k}^2\right],$$

and so the claim will follow from the ergodic theorem.

In order to verify (4.5), we first show that

$$\max_{p<i\le n} \sup_{0\le t<\infty} \frac{|\vartheta_i - t|}{t} = o_P(1). \tag{4.6}$$

Observe that

$$|\vartheta_i - t| \le \frac{t}{b_0}\left\{|\hat{b}_0 - b_0| + |\hat{b}_1 - b_1|y_{i-1}^2 + \ldots + |\hat{b}_p - b_p|y_{i-p}^2\right\}. \tag{4.7}$$

Since $\{y_i^2\}$ is a stationary sequence with $Ey_i^4 < \infty$,

$$\max_{1\le i\le n} y_i^2 = o_P(n^{1/2}). \tag{4.8}$$

Putting together (4.7), Lemma 4.1 and (4.8) we obtain (4.6).

Now to verify (4.5) fix $\varepsilon > 0$ and $0 < T < \infty$ so that $\sup_{T/2\le t<\infty} tf(t) \le \varepsilon$, c.f. (1.2). By continuity of f i t is not difficult to see that (4.6) together with (1.1) and the first relation in (1.2) yield

$$\max_{p<i\le n} \sup_{0\le t\le T} t|f(t) - f(\vartheta_i)| = o_P(1).$$

Let

$$\tilde{\Omega}_n = \left\{\omega : \frac{\vartheta_i}{t} \ge \frac{1}{2} \text{ for all } 0 \le t < \infty, p < i \le n\right\}.$$

By (4.6), we have $\lim_{n\to\infty} P\{\tilde{\Omega}_n\} = 1$. On the event $\tilde{\Omega}_n$ we have

$$\max_{p<i\le n} \sup_{T\le t<\infty} t|f(t) - f(\vartheta_i)| \le \varepsilon + \max_{p<i\le n} \sup_{T\le t<\infty} \frac{t}{\vartheta_i} \max_{p<i\le n} \sup_{T\le t<\infty} \vartheta_i f(\vartheta_i)$$

$$\leq \varepsilon + 2 \sup_{T/2\leq t<\infty} tf(t) \leq 3\varepsilon,$$

and so (4.5) is proved. ∎

For any $\mathbf{a} = (a_0, a_1, \ldots, a_p) \in R^{p+1}$ define

$$\gamma_i(\mathbf{a}) = a_0 \frac{1}{\sigma_i^2} + a_1 \frac{y_{i-1}^2}{\sigma_i^2} + \ldots + a_p \frac{y_{i-p}^2}{\sigma_i^2},$$

so that $\delta_i^2 = 1 + \gamma_i(\hat{\mathbf{b}}_n - \mathbf{b})$. Let

$$e_n^*(t, s, \mathbf{a}) = n^{-1/2} \sum_{p<i\leq ns} \left(\mathbf{I}\left\{\varepsilon_i^2 \leq t + tn^{-1/2}\gamma_i(\mathbf{a})\right\} - F(t + tn^{-1/2}\gamma_i(\mathbf{a})) \right).$$

Observe that

(4.9) $$e_n^*(t, s, n^{1/2}(\hat{\mathbf{b}}_n - \mathbf{b})) = \hat{e}_{n,1}(t, s).$$

LEMMA 4.3 *If conditions* (1.1) – (1.8) *hold, then for any* $x > 0$

$$P\left\{ \sup_{0\leq s\leq 1} |e_n^*(t, s, \mathbf{a}) - e_n(t, s)| \geq x \right\} \leq \frac{1}{n} \frac{C}{x^4} \left(||\mathbf{a}||^2 + 1 \right),$$

if $n \geq 16\, ||\mathbf{a}||^2 / b_0^2$, *with some constant* C *(which depends on* f, p, b_0, Ey_0^4 *but not on* t*).*

PROOF. Let

$$\xi_i(t, \mathbf{a}) = \mathbf{I}\left\{\varepsilon_i^2 \leq t + tn^{-1/2}\gamma_i(\mathbf{a})\right\} - \mathbf{I}\left\{\varepsilon_i^2 \leq t\right\} - \left(F(t + tn^{-1/2}\gamma_i(\mathbf{a})) - F(t) \right),$$

so that

$$e_n^*(t, s, \mathbf{a}) - e_n(t, s) = n^{-1/2} \sum_{p<i\leq ns} \xi_i(t, \mathbf{a}).$$

Since t and $\mathbf{a}$ are fixed, we write in the following ξ_i and γ_i instead of $\xi_i(t, \mathbf{a})$ and $\gamma_i(\mathbf{a})$. The σ-algebra generated by $\{\varepsilon_j,\ -\infty < j \leq i\}$ will be denoted by $\mathcal{F}_i$. It is not difficult to verify that $E\{\xi_i^2 | \mathcal{F}_{i-1}\} \leq \left| F(t + tn^{-1/2}\gamma_i) - F(t) \right|$ and so we have

(4.10) $$\sum_{1\leq i\leq n} E\left\{\xi_i^2 | \mathcal{F}_{i-1}\right\} \leq \sum_{1\leq i\leq n} |F(t + tn^{-1/2}\gamma_i) - F(t)|.$$

There is ξ_i^* between t and $t + tn^{-1/2}\gamma_i$ such that $F(t + tn^{-1/2}\gamma_i) - F(t) = f(\xi_i^*)tn^{-1/2}\gamma_i$ and so, by conditions (1.1) and (1.2),

$$\left| F(t + tn^{-1/2}\gamma_i) - F(t) \right| = \xi_i^* f(\xi_i^*)(t/\xi_i^*) n^{-1/2} |\gamma_i| \leq c_1 (t/\xi_i^*) n^{-1/2} |\gamma_i|,$$

where $c_1 = \sup_{0\leq u\leq\infty} uf(u)$. If $n^{-1/2}\gamma_i \geq -1/2$, then $t/2 \leq \min(t, t + n^{-1/2}\gamma_i)$, so $t/\xi_i^* \leq 2$ and therefore we have

(4.11) $$\left| F(t + tn^{-1/2}\gamma_i) - F(t) \right| \leq 2c_1 n^{-1/2} |\gamma_i| \mathbf{I}\left\{n^{-1/2}\gamma_i \geq -1/2\right\} + \mathbf{I}\left\{n^{-1/2}\gamma_i < -1/2\right\}.$$

By (4.10) and (4.11),

$$\left[\sum_{1\le i\le n} E\left\{\xi_i^2|\mathcal{F}_{i-1}\right\}\right]^2 \le \frac{4c_1^2}{n}\sum_{1\le i,j\le n}|\gamma_i||\gamma_j| + 4c_1 n^{-1/2}\sum_{1\le i,j\le n}|\gamma_i|\mathbf{I}\left\{n^{-1/2}\gamma_j<-1/2\right\} + \sum_{1\le i,j\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i<-1/2\right\}\mathbf{I}\left\{n^{-1/2}\gamma_j<-1/2\right\}. \tag{4.12}$$

Note that (recall the definition of σ_i^2)

$$|\gamma_i| \le ||\mathbf{a}||\, b_0^{-1}(1+y_{i-1}^2+\ldots+y_{i-p}^2) \tag{4.13}$$

and therefore $E\gamma_i^2 \le c_2\,||\mathbf{a}||^2$, where c_2 depends on p, b_0 and Ey_0^4. By the Cauchy-Schwarz inequality we also have that $E[|\gamma_i||\gamma_j|]\le c_2\,||\mathbf{a}||^2$. Using again (4.13) and the existence of Ey_0^4 we get that

$$P\{\gamma_i<-n^{1/2}/2\}\le P\{|\gamma_i|\ge n^{1/2}/2\} \tag{4.14}$$

$$\le P\left\{1+y_{i-1}^2+\ldots+y_{i-p}^2\ge\frac{n^{1/2}b_0}{2\,||\mathbf{a}||}\right\}\le P\left\{y_{i-1}^2+\ldots+y_{i-p}^2\ge\frac{n^{1/2}b_0}{4\,||\mathbf{a}||}\right\}$$

$$\le\frac{16p^2\,||\mathbf{a}||^2\,Ey_0^4}{nb_0^2},$$

provided $n\ge 16\,||\mathbf{a}||^2/b_0^2$. Hence

$$E\left[|\gamma_i|\mathbf{I}\left\{n^{-1/2}\gamma_i<-1/2\right\}\right]\le\left(E\gamma_i^2\right)^{1/2}\left(P\left\{n^{-1/2}\gamma_i<-1/2\right\}\right)^{1/2}\le c_3\,||\mathbf{a}||^2\,n^{-1/2}$$

and

$$E\left[\mathbf{I}\left\{n^{-1/2}\gamma_i<-1/2\right\}\mathbf{I}\left\{n^{-1/2}\gamma_j<-1/2\right\}\right]$$

$$\le\left(P\left\{n^{-1/2}\gamma_i<-1/2\right\}P\left\{n^{-1/2}\gamma_j<-1/2\right\}\right)^{1/2}\le\frac{16p^2\,||\mathbf{a}||^2\,Ey_0^4}{nb_0^2}.$$

Therefore, by (4.12),

$$E\left[\sum_{1\le i\le n}E\left\{\xi_i^2|\mathcal{F}_{i-1}\right\}\right]^2\le c_4\,||\mathbf{a}||^2\,n.$$

Since $|\xi_i|\le 2$, the last inequality and Theorem 2.11 on p. 23 of Hall and Heyde (1980) give

$$E\left[\max_{p<k\le n}\left|\sum_{p<i\le k}\xi_i\right|^4\right]\le c_5\left[c_4\,||\mathbf{a}||^2\,n+2^4\right]\le c_6\left(||\mathbf{a}||^2+1\right)n$$

and therefore by the Markov inequality

$$P\left\{n^{-1/2}\max_{p<k\le n}\left|\sum_{p<i\le k}\xi_i\right|\ge x\right\}=P\left\{\max_{p<k\le n}\left|\sum_{p<i\le k}\xi_i\right|^4\ge x^4n^2\right\}\le\frac{c_6\left(||\mathbf{a}||^2+1\right)}{x^4n}.$$

This completes the proof of Lemma 4.3. ∎

Lemma 4.4 *If conditions* (1.1)–(1.8) *hold, then for any* $\mathbf{a}\in R^{p+1}$ *we have*

$$\sup_{0\le s\le 1}\sup_{0\le t<\infty}|e_n^*(t,s,\mathbf{a})-e_n(t,s)|=o_P(1).$$

PROOF. Following the notation introduced at the beginning of the proof of Lemma 4.3, we write $\xi_i(t)$ and γ_i instead of $\xi_i(t,\mathbf{a})$ and $\gamma_i(\mathbf{a})$ ($\mathbf{a}$ is fixed). To prove the claim, we will verify that for any $x>0$,

$$\lim_{T\to 0}\limsup_{n\to\infty} P\left\{\sup_{0\le s\le 1}\sup_{0\le t\le T}|e_n^*(t,s,\mathbf{a})-e_n(t,s)|\ge x\right\}=0, \tag{4.15}$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left\{\sup_{0\le s\le 1}\sup_{T\le t\le \delta n^{1/2}}|e_n^*(t,s,\mathbf{a})-e_n(t,s)|\ge x\right\}=0 \tag{4.16}$$

for any $T>0$, and

$$\sup_{0\le s\le 1}\sup_{\delta n^{1/2}\le t<\infty}|e_n^*(t,s,\mathbf{a})-e_n(t,s)|=o_P(1) \tag{4.17}$$

for any $\delta>0$.

Verification of (4.15)*:* Fix $\epsilon>0$. By condition (1.2) there is $T^*\le 1$ such that $\sup_{0\le t\le T^*} tf(t)\le \epsilon$. Define $0=t_0<t_1<\ldots<t_N<T^*/2\le t_{N+1}$ satisfying

$$F(t_{i+1})-F(t_i)=\epsilon n^{-1/2},\quad 0\le i\le N.$$

It is clear that $N\le n^{1/2}\epsilon^{-1}+1$ and so by Lemma 4.3 we have

$$P\left\{n^{-1/2}\max_{1\le j\le N+1}\max_{p<k\le n}\left|\sum_{p<i\le k}\xi_i(t_j)\right|\ge x\right\}\le c_1 n^{-1/2}. \tag{4.18}$$

It thus remains to show that

$$\limsup_{n\to\infty} P\left\{n^{-1/2}\max_{0\le j\le N}\sup_{t_j\le t\le t_{j+1}}\max_{p<k\le n}\left|\sum_{p<i\le k}\xi_i(t)-\sum_{p<i\le k}\xi_i(t_j)\right|\ge x\right\}=0. \tag{4.19}$$

For any $t_j\le t\le t_{j+1}$ we have

$$\begin{aligned}
&\max_{p<k\le n}\sup_{t_j\le t\le t_{j+1}}\left|\sum_{p<i\le k}\Big[\Big(\mathbf{I}\left\{\varepsilon_i^2\le t+tn^{-1/2}\gamma_i\right\}-F(t+tn^{-1/2}\gamma_i)\Big)\right.\\
&-\left.\Big(\mathbf{I}\left\{\varepsilon_i^2\le t_j+t_jn^{-1/2}\gamma_i\right\}-F(t_j+t_jn^{-1/2}\gamma_i)\Big)\Big]\right|\\
\le&\max_{p<k\le n}\sup_{t_j\le t\le t_{j+1}}\left|\sum_{p<i\le k}\Big(\mathbf{I}\left\{\varepsilon_i^2\le t+tn^{-1/2}\gamma_i\right\}-\mathbf{I}\left\{\varepsilon_i^2\le t_j+t_jn^{-1/2}\gamma_i\right\}\Big)\right|\\
+&\max_{p<k\le n}\sup_{t_j\le t\le t_{j+1}}\left|\sum_{p<i\le k}\Big(F(t+tn^{-1/2}\gamma_i)-F(t_j+t_jn^{-1/2}\gamma_i)\Big)\right|\\
\le&\max_{p<k\le n}\sum_{p<i\le k}\Big(\mathbf{I}\left\{\varepsilon_i^2\le t_{j+1}+t_{j+1}n^{-1/2}\gamma_i\right\}-\mathbf{I}\left\{\varepsilon_i^2\le t_j+t_jn^{-1/2}\gamma_i\right\}\Big)\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}\\
+&\max_{p<k\le n}\sum_{p<i\le k}\Big(F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i)-F(t_j+t_jn^{-1/2}\gamma_i)\Big)\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}\\
+&\ 2\sum_{1\le i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i<-1/2\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le \max_{p<k\le n}\Bigg|\sum_{p<i\le k}\Big[\Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_{j+1}+t_{j+1}n^{-1/2}\gamma_i\right\} - F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i)\Big)\\
&- \Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_j+t_jn^{-1/2}\gamma_i\right\} - F(t_j+t_jn^{-1/2}\gamma_i)\Big)\Big]\mathbf{I}\left\{n^{-1/2}\gamma_i \ge -1/2\right\}\Bigg|\\
&+ 2\sum_{p<i\le n}\Big(F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i) - F(t_j+t_jn^{-1/2}\gamma_i)\Big)\mathbf{I}\left\{n^{-1/2}\gamma_i \ge -1/2\right\}\\
&+ 2\sum_{1\le i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i < -1/2\right\}\\
&\le \max_{p<k\le n}\Bigg|\sum_{p<i\le k}\Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_{j+1}+t_{j+1}n^{-1/2}\gamma_i\right\} - F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i)\Big)\\
&- \sum_{p<i\le k}\Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_j+t_jn^{-1/2}\gamma_i\right\} - F(t_j+t_jn^{-1/2}\gamma_i)\Big)\Bigg|\\
&+ 2\sum_{p<i\le n}\Big(F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i) - F(t_j+t_jn^{-1/2}\gamma_i)\Big)\mathbf{I}\left\{n^{-1/2}\gamma_i \ge -1/2\right\}\\
&+ 4\sum_{1\le i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i < -1/2\right\}\\
&\le \max_{p<k\le n}\left[\left|\sum_{p<i\le k}\xi_i(t_j)\right| + \left|\sum_{p<i\le k}\xi_i(t_{j+1})\right|\right]\\
&+ \max_{p<k\le n}|V_k(j)| + 2W(j) + 4\sum_{1\le i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i < -1/2\right\},
\end{aligned}
$$

where

$$
V_k(j) = \sum_{p<i\le k}\Big[\Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_{j=1}\right\} - F(t_{j+1})\Big) - \Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_j\right\} - F(t_j)\Big)\Big]
$$

and

$$
W(j) = \sum_{p<i\le n}\Big(F(t_{j+1}+t_{j+1}n^{-1/2}\gamma_i) - F(t_j+t_jn^{-1/2}\gamma_i)\Big)\mathbf{I}\left\{n^{-1/2}\gamma_i \ge -1/2\right\}.
$$

A similar, in fact simpler, argument shows that

$$
\begin{aligned}
&\max_{p<k\le n}\sup_{t_j\le t\le t_{j+1}}\left|\sum_{p<i\le k}\Big[\Big(\mathbf{I}\left\{\varepsilon_i^2 \le t\right\} - F(t)\Big) - \Big(\mathbf{I}\left\{\varepsilon_i^2 \le t_j\right\} - F(t_j)\Big)\Big]\right|\\
&\qquad \le \max_{p<k\le n}|V_k(j)| + 2n(F(t_{j+1}) - F(t_j)).
\end{aligned}
$$

Thus we get

$$
\begin{aligned}
&\max_{p<k\le n}\sup_{t_j\le t\le t_{j+1}}\left|\sum_{p<i\le k}\xi_i(t) - \sum_{p<i\le k}\xi_i(t_j)\right|\\
&\le \max_{p<k\le n}\left[\left|\sum_{p<i\le k}\xi_i(t_j)\right| + \left|\sum_{p<i\le k}\xi_i(t_{j+1})\right|\right] + 2\max_{p<k\le n}|V_k(j)| + 2W(j)\\
&+ 4\sum_{1\le i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i < -1/2\right] + 2n\left(F(t_{j+1}) - F(t_j)\right).
\end{aligned}
\tag{4.20}
$$

Now we establish appropriate upper bounds for the terms on the right-hand side of (4.20) which (when devided by $n^{1/2}$) will imply (4.19). The first two terms are asymptotically negligible by (4.18). Focusing on the $V_k(j)$, we see that by the modulus of continuity of the empirical process (and (4.22) below)

$$\max_{0\le j\le N}\sup_{0\le s\le 1}|e_n(t_{j+1},s)-e_n(t_j,s)|=o_P(1), \tag{4.21}$$

c.f. e.g. Shorack and Wellner (1986) p. 542. By the definition of the t_j, $0\le j\le N+1$, we have

$$\max_{0\le j\le N} n\left(F(t_{j+1})-F(t_j)\right)\le \epsilon n^{1/2}. \tag{4.22}$$

If $n^{-1/2}\gamma_i\ge -1/2$, then we have $t/(t\wedge(t+tn^{-1/2}\gamma_i))\le 2$. Hence by the mean-value theorem and the choice of T^* we have

$$\begin{aligned}
&\max_{0\le j\le N} W(j)\\
\le\ & 2\max_{0\le j\le N+1}\sum_{p<i\le n}\left|F(t_j+t_jn^{-1/2}\gamma_i)-F(t_j)\right|\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}\\
+\ & \max_{0\le j\le N} n\left(F(t_{j+1})-F(t_j)\right)\\
\le\ & 2\max_{1\le j\le N+1}\sum_{p<i\le n} f(\xi_{i,j})t_jn^{-1/2}|\gamma_i|\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}+\epsilon n^{1/2}\\
\le\ & 2\max_{0\le j\le N+1}\sum_{p<i\le n} \xi_{i,j}f(\xi_{i,j})(t_j/\xi_{i,j})n^{-1/2}|\gamma_i|\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}+\epsilon n^{1/2}\\
\le\ & 4\epsilon n^{-1/2}\sum_{p<i\le n}|\gamma_i|+\epsilon n^{1/2}
\end{aligned} \tag{4.23}$$

where $\xi_{i,j}$ is between t_j and $t_j+t_jn^{-1/2}\gamma_i$. Using the definitions of γ_i and σ_i and the ergodic theorem we have

$$\frac{1}{n}\sum_{p<i\le n}|\gamma_i|\le\frac{||\mathbf{a}||+1}{b_0}\frac{p+1}{n}\sum_{p<i\le n}y_i^2=O_P(1). \tag{4.24}$$

Applying (4.14) we conclude that

$$E\sum_{p<i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i<-1/2\right\}=O(1). \tag{4.25}$$

Verification of (4.16)*:* We will use again Lemma 4.3 and (4.20). Let $t_0=T$ and $t_i=T+i\delta^{1/6}n^{-1/2}$, $1\le i\le M-1$, $t_{M-1}\le\delta n^{1/2}\le t_M$. We note that $M\le n\delta^{5/6}+1$. By Lemma 4.3 we have

$$P\left\{\max_{0\le j\le M}\max_{p<k\le n} n^{-1/2}\left|\sum_{p<i\le k}\xi_i(t_j)\right|\ge x\right\}\le\frac{c_2}{x^4}\frac{M+1}{n}\le\frac{c_3}{x^4}\left(\delta^{5/6}+\frac{1}{n}\right). \tag{4.26}$$

Using again the modulus of continuity of $e_n(t,s)$, we get

$$\max_{0\le j\le M}\max_{p<k\le n}|V_k(j)|=o_P(n^{1/2}). \tag{4.27}$$

Since $f(t)$ is bounded on $[T,\infty)$, say, by c_4, we obtain

$$\max_{0\le j\le M} n\left(F(t_{j+1})-F(t_j)\right)\le c_4\delta^{1/6}n^{1/2}, \tag{4.28}$$

and by the mean-value theorem

$$\max_{0\le j\le M} W(j) \le c_4 \max_{0\le j\le M} \sum_{p<i\le n} |t_{j+1}-t_j|(1+n^{-1/2}|\gamma_i|)\mathbf{I}\left\{n^{-1/2}\gamma_i \ge -1/2\right\}$$
$$\le c_4\delta^{1/6}n^{-1/2}\left(n+n^{-1/2}\sum_{p<i\le n}|\gamma_i|\right) \le \delta^{1/6}n^{1/2}O_P(1) \tag{4.29}$$

on account of (4.24). Relation (4.16) follows from (4.20), (4.26)–(4.29) and (4.25).

Verification of (4.17)*:* We use Lemma 4.3 and the following bound:

$$\max_{p<k\le n}\ \sup_{\delta n^{1/2}\le t<\infty}\left|\sum_{p<i\le k}\xi_i(t)\right| \tag{4.30}$$
$$\le \max_{p<k\le n}\ \sup_{\delta n^{1/2}\le t<\infty}\sum_{p<i\le k}\left(1-\mathbf{I}\left\{\varepsilon_i^2\le t+tn^{-1/2}\gamma_i\right\}\right)$$
$$+ \max_{p<k\le n}\ \sup_{\delta n^{1/2}\le t<\infty}\sum_{p<i\le k}\left(1-F(t+tn^{-1/2}\gamma_i)\right)$$
$$+ \sum_{1\le i\le n}\left(1-\mathbf{I}\left\{\varepsilon_i^2\le \delta n^{1/2}\right\}\right)+n\left(1-F(\delta n^{1/2})\right)$$
$$\le \sum_{p<i\le n}\left(1-\mathbf{I}\left\{\varepsilon_i^2\le \delta n^{1/2}+\delta\gamma_i\right\}\right)\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}$$
$$+ \sum_{p<i\le n}\left(1-F(\varepsilon_i^2\le \delta n^{1/2}+\delta\gamma_i)\right)\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}$$
$$+ 2\sum_{p<i\le n}\mathbf{I}\left\{n^{1/2}\gamma_i< -1/2\right\}+\left|\sum_{1\le i\le n}\left(\mathbf{I}\left\{\varepsilon_i^2\le \delta n^{1/2}\right\}-F(\delta n^{1/2})\right)\right|$$
$$+ 2n\left(1-F(\delta n^{1/2})\right)$$
$$\le \left|\sum_{p<i\le n}\xi_i(\delta n^{1/2})\right|+2\left|\sum_{1\le i\le n}\left(\mathbf{I}\left\{\varepsilon_i^2\le \delta n^{1/2}\right\}-F(\delta n^{1/2})\right)\right|$$
$$+ 2\sum_{p<i\le n}\left(1-F(\delta n^{1/2}+\delta\gamma_i)\right)\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}$$
$$+ 4\sum_{p<i\le n}\mathbf{I}\left\{n^{-1/2}\gamma_i< -1/2\right\}+4n\left(1-F(\delta n^{1/2})\right).$$

By Lemma 4.3,

$$\left|\sum_{p<i\le n}\xi_i(\delta n^{1/2})\right| = o_P(n^{1/2}). \tag{4.31}$$

Observing that $E(\sum_{1\le i\le n}(I\{\epsilon_i^2\le \delta n^{1/2}\}-F(\delta n^{1/2})))^2 = nF(\delta n^{1/2})(1-F(\delta n^{1/2})) = o(n)$, by the Chebyshev Inequality we have

$$\left|\sum_{1\le i\le n}\left(\mathbf{I}\left\{\varepsilon_i^2\le \delta n^{1/2}\right\}-F(\delta n^{1/2})\right)\right| = o_P(n^{1/2}). \tag{4.32}$$

Also

(4.33)
$$\sum_{p<i\le n}\Big(1-F(\delta n^{1/2}+\delta\gamma_i)\Big)\,\mathbf{I}\left\{n^{-1/2}\gamma_i\ge -1/2\right\}\le n\left(1-F(\delta n^{1/2}/2)\right)=o(1),$$

by assumption (1.7) which implies that $nP(\varepsilon_0^4>n)\to 0$. Relations (4.30)–(4.33) and (4.25) imply (4.17). ∎

For any $A>0$ define $\mathbf{A}=[-A,A]^{p+1}$.

LEMMA 4.5 *If conditions* (1.1)–(1.8) *hold, then for any* $A>0$

$$\sup_{\mathbf{a}\in\mathbf{A}}\sup_{0\le s\le 1}\sup_{0\le t<\infty}|e_n^*(t,s,\mathbf{a})-e_n(t,s)|=o_P(1).$$

PROOF. Denote

$$X_n(\mathbf{a})=\sup_{0\le s\le 1}\sup_{0\le t<\infty}\left|n^{-1/2}\sum_{p<i\le ns}\xi_i(t,\mathbf{a})\right|.$$

In Lemma 4.4. we showed that $X_n(\mathbf{a})\xrightarrow{P}0$ for every $\mathbf{a}$. Here we show that $\sup_{\mathbf{a}\in\mathbf{A}}X_n(\mathbf{a})\xrightarrow{P}0$.

Consider $\epsilon>0$ such that $2A/\epsilon$ is an integer and which will be specified at the end of the proof. Define $a(k)=-A+k\epsilon,\ 1\le k\le K\equiv 2A/\epsilon$. Set $\mathbf{k}=(k_0,k_1,\dots,k_p),\ 1\le k_0,k_1,\dots,k_p\le K$ and consider the grid of $(p+1)K$ points $\mathbf{a}(\mathbf{k})=(a(k_0),a(k_1),\dots,a(k_p))$. Consider also the corresponding $(p+1)K$ cells

$$\mathbf{A}(\mathbf{k})=\{(a_0,a_1,\dots,a_p)\in\mathbf{A}:a(k_i)-\epsilon\le a_i\le a(k_i)\}$$

and the points $\mathbf{a}^*(\mathbf{k})=(a(k_0)-\epsilon,a(k_1)-\epsilon,\dots,a(k_p)-\epsilon)$. Observe that

$$\begin{aligned}&F\left(t+tn^{-1/2}\gamma_i(a_0,\dots,a_{j-1},a(k_j),a_{j+1},\dots,a_p)\right)\\ -&F\left(t+tn^{-1/2}\gamma_i(a_0,\dots,a_{j-1},a(k_j)-\epsilon,a_{j+1},\dots,a_p)\right)\\ &=\begin{cases}f(\xi_{i0}^*)tn^{-1/2}\epsilon/\sigma_i^2 & \text{if}\quad j=0\\ f(\xi_{ij}^*)tn^{-1/2}\epsilon y_{i-j}^2/\sigma_i^2 & \text{if}\quad 1\le j\le p,\end{cases}\end{aligned}$$

where ξ_{ij}^* is a point between the arguments of F above. Consider the set

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

and notice that by (4.8) $\lim_{n\to\infty}P\{\Omega_n^*\}=0$. In the remainder of the proof we work on the complement of Ω_n^*. Since $t/\xi_{ij}^*\le 2$,

(4.34)
$$\begin{aligned}&F\left(t+tn^{-1/2}\gamma_i(a_0,\dots,a_{j-1},a(k_j),a_{j+1},\dots,a_p)\right)\\ -&F\left(t+tn^{-1/2}\gamma_i(a_0,\dots,a_{j-1},a(k_j)-\epsilon,a_{j+1},\dots,a_p)\right)\\ &\le\begin{cases}2c_1tn^{-1/2}\epsilon/\sigma_i^2 & \text{if}\quad j=0\\ 2c_1tn^{-1/2}\epsilon y_{i-j}^2/\sigma_i^2 & \text{if}\quad 1\le j\le p,\end{cases}\end{aligned}$$

where $c_1 = \sup_{0\le t<\infty} tf(t)$. Applying (4.34) consecutively to each coordinate, we obtain for any $t > 0$ and any $\mathbf{k}$

$$\sup_{\mathbf{a}\in\mathbf{A}(\mathbf{k})} \left| \sum_{p<i\le k} \left[F\left(t + tn^{-1/2}\gamma_i(\mathbf{a})\right) - F\left(t + tn^{-1/2}\gamma_i\left(\mathbf{a}(\mathbf{k})\right)\right)\right] \right| \le 2c_1\epsilon n^{1/2} b_0^{-1} \left[1 + n^{-1} \sum_{p<i\le n} \sum_{1\le j\le p} y_{i-j}^2 \right], \tag{4.35}$$

and so

$$\begin{aligned} &\sup_{\mathbf{a}\in\mathbf{A}(\mathbf{k})} \left| \sum_{p<i\le k} \left[\xi_i(t,\mathbf{a}) - \xi_i(t,\mathbf{a}(\mathbf{k}))\right] \right| \\ \le\ & 2c_1\epsilon n^{1/2} b_0^{-1} \left[1 + n^{-1} \sum_{p<i\le n} \sum_{1\le j\le p} y_{i-j}^2 \right] \\ +\ & 2 \sum_{p<i\le k} \left[\mathbf{I}\left\{\varepsilon_i^2 \le t + tn^{-1/2}\gamma_i(\mathbf{a}(\mathbf{k}))\right\} - \mathbf{I}\left\{\varepsilon_i^2 \le t + tn^{-1/2}\gamma_i(\mathbf{a}^*(\mathbf{k}))\right\} \right] \\ \le\ & 4c_1\epsilon n^{1/2} b_0^{-1} \left[1 + n^{-1} \sum_{p<i\le n} \sum_{1\le j\le p} y_{i-j}^2 \right] \\ +\ & \left| \sum_{p<i\le k} \xi_i(t,\mathbf{a}(\mathbf{k})) \right| + \left| \sum_{p<i\le k} \xi_i(t,\mathbf{a}^*(\mathbf{k})) \right|. \end{aligned} \tag{4.36}$$

Observe that

$$\sup_{\mathbf{a}\in\mathbf{A}} X_n(\mathbf{a}) \le \sup_{\mathbf{k}} \sup_{\mathbf{a}\in\mathbf{A}(\mathbf{k})} \sup_{0\le s\le 1} \sup_{0\le t<\infty} \left| n^{-1/2} \sum_{p<i\le k} \left[\xi_i(t,\mathbf{a}) - \xi_i(t,\mathbf{a}(\mathbf{k}))\right] \right| + \sup_{\mathbf{k}} X_n(\mathbf{a}(\mathbf{k})).$$

Hence, by (4.36)

$$\begin{aligned} \sup_{\mathbf{a}\in\mathbf{A}} X_n(\mathbf{a}) &\le c_2\epsilon \left[1 + n^{-1} \sum_{p<i\le n} \sum_{1\le j\le p} y_{i-j}^2 \right] \\ &+ 2\sup_{\mathbf{k}} X_n(\mathbf{a}(\mathbf{k})) + \sup_{\mathbf{k}} X_n(\mathbf{a}^*(\mathbf{k})) =: \epsilon Y_n + Z_n. \end{aligned}$$

Fix $r > 0$. We must show that $\lim_{n\to\infty} P\left\{\sup_{\mathbf{a}\in\mathbf{A}} X_n(\mathbf{a}) > r\right\} = 0$. By the ergodic theorem $Z_n := c_2\left[1 + n^{-1}\sum_{p<i\le n}\sum_{1\le j\le p} y_{i-j}^2\right]$ tends in probability to a constant, say, c_3. We now fix ϵ so small that $\epsilon c_3 < r/2$. With ϵ fixed, $Y_n := 2\sup_{\mathbf{k}} X_n(\mathbf{a}(\mathbf{k})) + \sup_{\mathbf{k}} X_n(\mathbf{a}^*(\mathbf{k}))$ is a maximum over a finite number of points and so $P\{Z_n > r/2\} \to 0$ by Lemma 4.4 and $P\{\epsilon Y_n > r/2\} \to 0$ by the choice of ϵ. ∎

Lemma 4.6 *If* (1.1) – (1.8) *hold, then*

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

where $u_n(t,s)$ *is defined by* (4.2) *and* (4.1).

PROOF. We use the inequality

$$
\begin{aligned}
|\hat{e}_n(t,s) - (e_n(t,s) + u_n(t,s))| &\le |\hat{e}_{n,1}(t,s) - e_n(t,s)| \\
&+ |\hat{e}_{n,2}(t,s) - h_n(t,s)| + |h_n(t,s) - u_n(t,s)|,
\end{aligned}
$$

where h_n is given by (1.10).

Lemmas 4.1 and 4.5 yield (c.f. (4.9))

$$
\sup_{0\le s\le 1}\sup_{0\le t<\infty} |\hat{e}_{n,1}(t,s) - e_n(t,s)| = o_P(1). \tag{4.37}
$$

Indeed, denoting, as in the proof of Lemma 4.5, $X_n(\mathbf{a}) = \sup_{0\le s\le 1}\sup_{0\le t<\infty} |e_n^*(t,s,\mathbf{a}) - e_n(t,s)|$ and $\hat{\mathbf{a}}_n = n^{1/2}(\hat{\mathbf{b}}_n - \mathbf{b})$, we have

$$
P\{X_n(\hat{\mathbf{a}}_n) > r\} \le P\left\{\sup_{\mathbf{a}\in\mathbf{A}} X_n(\mathbf{a}) > r\right\} + P\{\hat{\mathbf{a}}_n \notin \mathbf{A}\}.
$$

By Lemma 4.5 this implies $\limsup_{n\to\infty} P\{X_n(\hat{\mathbf{a}}_n) > r\} \le P\{\hat{\mathbf{a}}_n \notin \mathbf{A}\}$ and so (4.37) follows because by Lemma 4.1, letting $A \to \infty$, $P\{\hat{\mathbf{a}}_n \notin \mathbf{A}\}$ can be made arbitrarily small.

By Lemma 4.2

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

To verify that

$$
\sup_{0\le s\le 1}\sup_{0\le t<\infty} |u_n(t,s) - h_n(t,s)| = o_P(1) \tag{4.38}
$$

observe that by (1.3)

$$
\begin{aligned}
& h_n(t,s) - u_n(t,s) \\
= {} & \left[\frac{1}{n}\sum_{p<i\le ns}\frac{1}{\sigma_i^2} - s\beta_0\right] tf(t)n^{-1/2}\sum_{1\le j\le n} l_0(\varepsilon_j^2) f_0(\varepsilon_{j-1},\varepsilon_{j-2},\ldots) \\
+ {} & \sum_{1\le k\le p}\left[\frac{1}{n}\sum_{p<i\le ns}\frac{y_{i-k}^2}{\sigma_i^2} - s\beta_k\right] tf(t)n^{-1/2}\sum_{1\le j\le n} l_k(\varepsilon_j^2) f_k(\varepsilon_{j-1},\varepsilon_{j-2},\ldots) \\
+ {} & o_P(1)tf(t)\frac{1}{n}\sum_{p<i\le ns}\frac{1}{\sigma_i^2}\left(1 + \sum_{1\le k\le p} y_{i-k}^2\right).
\end{aligned}
$$

In view of (1.1)–(1.2) and (1.4)–(1.6) in remains to verify that for each $1 \le k \le p$

$$
\sup_{0\le s\le 1}\left[\frac{1}{n}\sum_{p<i\le ns}\frac{y_{i-k}^2}{\sigma_i^2} - s\beta_k\right] = o_P(1), \tag{4.39}
$$

with an obvious modification for $k = 0$. Since for each fixed k, the random variables $y_{i-k}^2/\sigma_i^2 - \beta_k$ form a stationary geometrically mixing sequence, see Guegan and Diebolt (1994), with zero mean and finite variance, their normalized partial sum process converges weakly to a Brownian motion, and so the left-hand side of (4.39) is in fact $O_P(n^{-1/2})$. ■

LEMMA 4.7 *Let* $e_n(t,s)$ *and* ζ_n *be defined by* (1.9) *and* (4.1) *respectively. If* (1.1)–(1.8) *hold, then* $\{(e_n(t,s),\zeta_n),\ 0 \le t < \infty, 0 \le s \le 1\}$ *converges weakly in* $\mathcal{D}([0,\infty]\times[0,1])\times R$ *to the Gaussian process* $\{(K(F(t),s),\zeta),\ 0 \le t < \infty, 0 \le s \le 1\}$ *defined by* (1.12)–(1.15).

PROOF. We will verify tightness and the convergence of the finite dimensional distributions. Since $e_n(t,s)$ is a sequential empirical process of independent, identically distributed random variables, its tightness is well-known, see, for example, Csörgő and Révész (1981). If we show that ζ_n converges in distribution to a normal random variable, the tightness will be proved because the vector valued process $(e_n(t,s),\zeta_n)$ is tight if the coordinates are tight.

Fix an integer M, real numbers $\lambda_1,\lambda_2\ldots,\lambda_M,\lambda_{M+1},\ 0\le t_1,t_2,\ldots,t_M<\infty,\ 0=s_0\le s_1\le s_2\le\ldots\le s_M\le 1$ and define

$$Z(n)=\sum_{1\le m\le M}\lambda_m e_n(t_m,s_m)+\lambda_{M+1}\zeta_n.$$

It is easy to see that

$$n^{1/2}Z(n)=\sum_{1\le j\le n}\tau_j,$$

where, setting $f_k^-(j)=f_k(\varepsilon_{j-1},\varepsilon_{j-2},\ldots)$,

$$\tau_j=\sum_{m(j)\le m\le M}\lambda_m\left(\mathbf{I}\left\{\varepsilon_j^2\le t_m\right\}-F(t_m)\right)+\lambda_{M+1}\sum_{0\le k\le p}\beta_k l_k(\varepsilon_j^2)f_k^-(j), \tag{4.40}$$

where $m(j)=i$ if $(ns_{i-1})\vee p<j\le ns_i, 1\le i\le M$ and

$$\tau_j=\lambda_{M+1}\sum_{0\le k\le p}\beta_k l_k(\varepsilon_j^2)f_k^-(j),\quad \text{if}\ \ 1\le j\le p\ \ \text{or}\ \ ns_M<j\le n.$$

Recall that $\mathcal{F}_k$ is the σ-algebra generated by $\varepsilon_k,\varepsilon_{k-1},\ldots$. We will use Theorem 3.5 of Hall and Heyde (1980) to show that $Z(n)$ is asymptotically normal. To do so, we will verify that

$$\max_{1\le j\le n}E\left\{\tau_j^2|\mathcal{F}_{j-1}\right\}=o_P(n^{1/2}), \tag{4.41}$$

$$E\left|\sum_{1\le i\le n}\left(E\left\{\tau_j^2|\mathcal{F}_{j-1}\right\}-E\tau_j^2\right)\right|=o(n) \tag{4.42}$$

and

$$E\left|\sum_{1\le i\le n}\left(\tau_j^2-E\tau_j^2\right)\right|=o(n). \tag{4.43}$$

Focusing on the more complex case (4.40), we write

$$\begin{aligned}\tau_j^2=&\left[\sum_{m(j)\le m\le M}\lambda_m\left(\mathbf{I}\left\{\varepsilon_j^2\le t_m\right\}-F(t_m)\right)\right]^2\\&+\lambda_{M+1}^2\sum_{0\le k,k'\le p}\beta_k\beta_{k'}l_k(\varepsilon_j^2)l_{k'}(\varepsilon_j^2)f_k^-(j)f_{k'}^-(j)\\&+2\lambda_{M+1}\sum_{0\le k\le p}\beta_k f_k^-(j)l_k(\varepsilon_j^2)\sum_{m(j)\le m\le M}\lambda_m\left(\mathbf{I}\left\{\varepsilon_j^2\le t_m\right\}-F(t_m)\right)\end{aligned} \tag{4.44}$$

and therefore

$$E\left\{\tau_j^2|\mathcal{F}_{j-1}\right\}=\sum_{m(j)\le m\le M}\lambda_m^2\left[F(t_m)-F^2(t_m)\right] \tag{4.45}$$

$$+\lambda_{M+1}^2 \sum_{0\le k,k'\le p} \beta_k\beta_{k'} E[l_k(\varepsilon_j^2) l_{k'}(\varepsilon_j^2)] f_k^-(j) f_{k'}^-(j)$$

$$+2\lambda_{M+1} \sum_{0\le k\le p} \beta_k f_k^-(j) \sum_{m(j)\le m\le M} \lambda_m \int_0^{t_m} l_k(x) f(x)dx$$

and

$$E\tau_j^2 = \sum_{m(j)\le m\le M} \lambda_m^2 \Big[F(t_m) - F^2(t_m)\Big] + \lambda_{M+1}^2 \sum_{0\le k,k'\le p} \beta_k \gamma_{kk'} \beta_{k'} \tag{4.46}$$

$$+2\lambda_{M+1} \sum_{0\le k\le p} \beta_k \sum_{m(j)\le m\le M} \lambda_m g_k(t_m).$$

Since for each k, $f_k^-(j) \equiv f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots)$, $1 \le j < \infty$ is a stationary sequence satisfying (1.6) , $\max_{1\le j\le n} |f_i^-(j)| = o_P(n^{1/2})$, and so (4.41) follows from (4.45). Using expressions (4.44)–(4.46), we see that (4.42) and (4.43) follow from the mean ergodic theorem.

By computing $EZ^2(n)$, we obtain the desired covariance structure. ∎

Acknowledgement. We would like to thank both referees for their careful reading of the paper and their useful comments.

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, MA.

Bai, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics*, **22,** 2051–2061.

Blum, J. R., Kiefer, J. and Rossenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32,** 485–498.

Boldin, M. V. (1998). On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen*, **235,** 49–66.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* 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.

Chu, C.-S. J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14,** 241–266.

Cotteril, D. S. and Csörgő, M. (1985). On the limiting distribution of and the critical values for the Hoeffding, Blum, Kiefer and Rosenblatt independence criterion. *Statistics & Decisions*, **3,** 1–48.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

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

Csörgő, S. (1983). Kernel–transformed empirical processes. *Journal of Multivariate Analysis*, **13,** 511–533.

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.

Frances, P. H. and van Dijk, D. (2000). Outlier detection in GARCH models. Technical Report. Erasmus University. Econometric Institute Research Report, EI-9926/rv.

Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Testing for long memory in the presence of a general trend. preprint.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer-Verlag, New York.

Guegan, D. and Diebolt, J. (1994). Probabilistic properties of β-arch model. *Statistica Sinica*, **4,** 71–87.

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

He, C. and Teräsvirta, T. (1999). Fourth moment structure of the GARCH(p, q) process. *Econometric Theory*, **15,** 824–846.

Horváth, L. (1985). Empirical kernel transforms of parameter–estimated empirical processes. *Acta Scientiarum Mathematicarum*, **48,** 201–213.

Horváth, L. and Kokoszka, P. (2001). Large sample distribution of ARCH(p) squared residual correlations. forthcoming in *Econometric Theory.*

Horváth, L., Kokoszka, P. and Teyssiére, G. (2000). Bootstrap specification tests for ARCH. preprint.

Khamaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.*, **26,** 240–257.

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *The Annals of Statistics*, **27,** 204–236.

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.

Martynov, G. V. (1992). Statistical tests based on empirical processes and related questions. *J. Sov. Math.*, **61,** 2195–2273.

Mikosch, T. and Stărică, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. preprint available at `http://www.cs.nl/~eke/iwi/preprints`.

Picard, D. (1985). Testing and estimating change–point in time series. *Adv. in Appl. Probab.*, **17,** 841–867.

Seber, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

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.