Ann. Inst. Statist. Math.
Vol. 57, No. 4, 747–765 (2005)

THE EMPIRICAL DISTRIBUTION FUNCTION AND PARTIAL SUM PROCESS OF RESIDUALS FROM A STATIONARY ARCH WITH DRIFT PROCESS

JANUSZ KAWCZAK[1], REG KULPERGER[2] AND HAO YU[2]

[1]*Department of Mathematics and Statistics, University of North Carolina, Charlotte, NC 28223, U.S.A.*
[2]*Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada N6A 5B7*

(Received October 7, 2002; revised August 30, 2004)

Abstract. The weak convergence of the empirical process and partial sum process of the residuals from a stationary ARCH-M model is studied. It is obtained for any $\sqrt{n}$ consistent estimate of the ARCH-M parameters. We find that the limiting Gaussian processes are no longer distribution free and hence residuals cannot be treated as i.i.d. In fact the limiting Gaussian process for the empirical process is a standard Brownian bridge plus an additional term, while the one for partial sum process is a standard Brownian motion plus an additional term. In the special case of a standard ARCH process, that is an ARCH process with no drift, the additional term disappears. We also study a sub-sampling technique which yields the limiting Gaussian processes for the empirical process and partial sum process as a standard Brownian bridge and a standard Brownian motion respectively.

Key words and phrases: Weak convergence, residuals, ARCH, drift, empirical distribution.

1. Introduction

In nonlinear time series, and in particular econometric and discrete time financial modeling, Engle's (1982) ARCH model plays a fundamental role; see Campbell *et al.* (1997), Gourieroux (1997) or the volume Rossi (1996) which contains several papers by Nelson. The simplest of these is of the form

$$X_t = \sigma_t \epsilon_t, \tag{1.1}$$

where $\{\epsilon_t, t \geq 1\}$ is a sequence of iid random variables (r.v.'s) with mean zero and finite variance. Throughout this paper we make the additional assumption that the variance term $E(\epsilon_1^2) = 1$ so that σ_t^2 is the conditional variance of X_t given $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(X_s : s \leq t)$ is the sigma field generated by the data up to time t, that is $\{X_s : s \leq t\}$. The conditional variance term σ_t^2 is F_t adapted. For an ARCH(1) model the conditional variance is of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2, \qquad \alpha_0 > 0, \qquad \alpha_1 > 0,$$

so that it is a known form parametric function of the most recent observation. Other forms of σ_t^2 are also used to capture various properties, such as non symmetric conditional variances, or higher order lag dependencies as functions of $X_{t-1}, X_{t-2}, \dots, X_{t-p}$. For example an ARCH(p) model has

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \qquad \alpha_0 > 0, \qquad \alpha_j \geq 0, \qquad j = 1, \dots, p.$$

Often an assumption that the innovations ϵ_t are $N(0, \sigma^2)$ is also made. More generally ϵ_t are iid with distribution function F which is assumed throughout this paper.

To address some weaknesses of ARCH models, Engle *et al.* (1982) introduced the ARCH-M model which extends the ARCH model to allow the conditional variance to affect the mean. The ARCH model (1.1) then becomes

$$X_t = \mu + \delta m(\sigma_t) + \sigma_t \epsilon_t, \tag{1.2}$$

where μ and δ are additional parameters with the deterministic function m usually chosen as $m(x) = x, \sqrt{x}$ or $\exp(x)$. In this paper we consider only the case with $\delta = 0$, that is an ARCH-M model with non-zero mean or drift parameter μ. The ARCH-M(1) is given by (1.3) and (1.4) below.

Horváth *et al.* (2001) investigate the empirical process of the squared residuals arising from fitting an ARCH type model with mean $\mu = 0$. They obtained a distribution free limiting process for a specially transformed empirical process.

This article proposes to fill in the gap by establishing the limiting process of the residuals from fitting and ARCH-M model. The limiting Gaussian process is not distribution free and depends on the distribution of the innovations. This does not create a drawback for applications since quite powerful nonparametric methods for the density estimation are readily available. Gourieroux (1997) discusses estimation of parameters for GARCH models with a non-zero unknown mean or drift parameter μ. Koul (2002) presents some ideas on the estimation of the parameters in the ARCH type modelling.

In this paper we consider a special case of ARCH-M model (1.2) with $\delta = 0$. First we study the ARCH-M(1) process

$$X_t = \mu + \sigma_t \epsilon_t \tag{1.3}$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 (X_{t-1} - \mu)^2, \qquad \alpha_0 > 0, \qquad \alpha_1 > 0. \tag{1.4}$$

Later we show how to extend our results to the ARCH-M(p).

Consider the ARCH-M(1) process (1.3) with observed data X_t, $t = 0, \dots, n$. Consider any $\sqrt{n}$ consistent estimators of the parameters (see for example Gourieroux (1997)) and the residuals (2.2) obtained from this fit. From these residuals one constructs the empirical distribution function (EDF) $\hat{F}_n$ and the partial sum $\hat{S}_n$ defined below by (2.3) and (2.4), respectively. We study the asymptotic properties of $\hat{F}_n$ and $\hat{S}_n$. In particular we study the empirical process

$$E_n(x) = \sqrt{n}(\hat{F}_n(x) - F(x)), \qquad -\infty < x < \infty \tag{1.5}$$

and the partial sum process

$$B_n(u) = \frac{1}{\sqrt{n}} \hat{S}_{[nu]}, \qquad 0 \leq u \leq 1, \tag{1.6}$$

where $[x]$ denotes the integer part of x. The results are given in Section 2.

An assumption that the ARCH-M process is stationary and ergodic is made throughout this paper. See for example An *et al.* (1997) for relevant conditions on the ARCH parameters. Section 2 defines the ARCH-M(1) residuals EDF process and partial sum process and states the main theorems. Extension of ARCH-M(1) to ARCH-M(p) will be discussed in the end of Section 2. Section 3 gives the proofs.

2. ARCH-M residuals and results

In this section we first consider the EDF process and partial sum process for ARCH-M(1) residuals. At the end of the section the changes required for the residual processes from an ARCH-M(p) process are given.

Let $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\mu})$ be an estimator of $\theta = (\alpha_0, \alpha_1, \mu)$ based on the sample of size n. Also suppose that the estimator is $\sqrt{n}$ consistent. Such estimators are obtained in Engle *et al.* (1982) and are discussed in the monograph by Gourieroux (1997). The conditional variance $\sigma_t^2 = h(X_{t-1}, \theta)$ of (1.4) is estimated by

$$\hat{\sigma}_t^2 = h(X_{t-1}, \hat{\theta}) = \hat{\alpha}_0 + \hat{\alpha}_1 (X_{t-1} - \hat{\mu})^2, \tag{2.1}$$

where $h : \mathbb{R}^4 \to \mathbb{R}^+$ is a deterministic function. Thus the residual at time $t \in \{1, 2, \ldots, n\}$ is

$$\hat{\epsilon}_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}_t} = \frac{\mu - \hat{\mu} + \sigma_t \epsilon_t}{\hat{\sigma}_t} = \frac{\sqrt{n}(\mu - \hat{\mu})}{\sqrt{n h(X_{t-1}, \hat{\theta})}} + \epsilon_t \sqrt{\frac{h(X_{t-1}, \theta)}{h(X_{t-1}, \hat{\theta})}} \tag{2.2}$$

by (1.3), (1.4) and (2.1).

The EDF of the residuals is defined as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{I}(\hat{\epsilon}_t \le x), \qquad -\infty < x < \infty \tag{2.3}$$

and the partial sums of the residuals as

$$\hat{S}_0 = 0, \qquad \hat{S}_k = \sum_{t=1}^{k} \hat{\epsilon}_t, \qquad k = 1, 2, \ldots, n. \tag{2.4}$$

We now introduce some notation that is necessary in our study of the EDF and partial sum processes. Let $s = (s_0, s_1, s_2) \in \mathbb{R}^3$ and define the function g_n as

$$g_n(x, s) = \frac{\sqrt{n}(h(x, \theta + n^{-1/2} s) - h(x, \theta))}{h(x, \theta)}.$$

Define also

$$\hat{F}_n(x, s) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{I}\left(\epsilon_t \le \left(x + \frac{s_2}{\sqrt{n h(X_{t-1}, \theta + n^{-1/2} s)}} \right) \sqrt{1 + \frac{g_n(X_{t-1}, s)}{\sqrt{n}}} \right) \tag{2.5}$$

and

$$\hat{\epsilon}_t(s) = \frac{\epsilon_t}{\sqrt{1 + \frac{g_n(X_{t-1}, s)}{\sqrt{n}}}}. \tag{2.6}$$

From (2.2), (2.3) and (2.5) we obtain

$$\hat{F}_n(x) = \hat{F}_n(x, \sqrt{n}(\hat{\theta} - \theta)) \tag{2.7}$$

and from (2.2), (2.4) and (2.6) we obtain

$$\hat{S}_k = -\sum_{t=1}^{k} \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{nh(X_{t-1}, \hat{\theta})}} + \sum_{t=1}^{k} \hat{\epsilon}_t(\sqrt{n}(\hat{\theta} - \theta)), \qquad k = 1, \ldots, n. \tag{2.8}$$

Note that for a given s,

$$\begin{aligned} g_n(x, s) &= \frac{s_0 + s_1(x-\mu)^2 - 2\alpha_1(x-\mu)s_2}{\alpha_0 + \alpha_1(x-\mu)^2} \\ &\quad + \frac{1}{\sqrt{n}} \frac{\alpha_1 s_2^2 - 2(x-\mu)s_1 s_2}{\alpha_0 + \alpha_1(x-\mu)^2} + \frac{1}{n} \frac{s_1 s_2^2}{\alpha_0 + \alpha_1(x-\mu)^2} \end{aligned} \tag{2.9}$$

which leads us easily to the following conclusion

$$\sup_{x \in \mathbb{R}} |g_n(x, s)| \le \sum_{i=1}^{3} \frac{C_i(\theta) \|s\|_\infty^i}{n^{(i-1)/2}} \tag{2.10}$$

where $\|s\|_\infty = \max\{|s_0|, |s_1|, |s_2|\}$ is the sup norm and $C_i(\theta)$, $i = 1, 2, 3$, are finite positive constants depending only on θ. As to the function h, it is easy to see that for $n > \max(\|s\|_\infty^2/\alpha_1^2, 4\|s\|_\infty^2/\alpha_0^2)$

$$\inf_{x \in \mathbb{R}} |h(x, \theta + n^{-1/2}s)| \ge \alpha_0 - \|s\|_\infty/\sqrt{n} > \alpha_0/2. \tag{2.11}$$

It is clear from (2.5), (2.7), (2.10) and (2.11), that

$$\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{I}(\epsilon_t \le (x + O_p(1/\sqrt{n}))\sqrt{1 + O_p(1/\sqrt{n})}).$$

Hence the EDF $\hat{F}_n$ of $\hat{\epsilon}$ will be consistent for F, although the uniformity of O_p in x will be shown later.

Define the processes

$$\begin{aligned} E_n(x, s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Bigg\{ & \mathrm{I}\left(\epsilon_t \le \left(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}}\right) \cdot \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)}\right) \\ & - F\left(\left(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}}\right) \cdot \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)}\right) \Bigg\} \end{aligned} \tag{2.12}$$

and

$$B_n(u,s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \hat{\epsilon}_t(s), \qquad 0 \le u \le 1. \tag{2.13}$$

Note that $E_n(x,0)$ and $B_n(x,0)$ are the usual EDF process and partial sum process of the iid sequence $\{\epsilon_t, t \ge 1\}$, respectively, and hence converge to a standard Brownian bridge and a standard Brownian motion, respectively. Also straightforward algebra applied to (1.5) and (2.7) yields

$$\begin{aligned} E_n(x) &= E_n(x, \sqrt{n}(\hat{\theta}-\theta)) \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ F\left(\left\{ x + \frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{nh(X_{t-1},\hat{\theta})}} \right\} \right.\right. \\ &\left.\left. \cdot \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, \sqrt{n}(\hat{\theta}-\theta))} \right) - F(x) \right\}. \end{aligned} \tag{2.14}$$

Similarly straightforward algebra applied to (1.6) and (2.8) yields

$$B_n(u) = B_n(u, \sqrt{n}(\hat{\theta}-\theta)) - \frac{1}{n} \sum_{t=1}^{[nu]} \frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{h(X_{t-1},\hat{\theta})}}. \tag{2.15}$$

PROPOSITION 2.1. *Suppose that the process $\{X_t, t \ge 0\}$ is stationary and ergodic, for example it satisfies the conditions of An et al. (1997). Suppose also that F has continuous density f that is positive on the open support of F, and $\lim_{x\to\pm\infty} |x|f(x) = 0$. Then for any $b > 0$*

$$\sup_{\|s\|_\infty \le b} \sup_{x\in\mathbb{R}} |E_n(x,s) - E_n(x,0)| \to 0$$

in probability as $n \to \infty$.

PROPOSITION 2.2. *Suppose that the process $\{X_t, t \ge 0\}$ is stationary and ergodic, for example it satisfies the conditions of An et al. (1997). Suppose also that F has continuous density f, and the iid sequence $\{\epsilon_t, t \ge 1\}$ has mean zero and variance 1. Then for any $b > 0$*

$$\sup_{\|s\|_\infty \le b} \sup_{0\le u\le 1} |B_n(u,s) - B_n(u,0)| \to 0$$

in probability as $n \to \infty$.

Define $g : \mathbb{R} \to \mathbb{R}^3$ and $k : \mathbb{R} \to \mathbb{R}^3$ to be

$$g(x) = \frac{(1, (x-\mu)^2, -2\alpha_1(x-\mu))}{\alpha_0 + \alpha_1(x-\mu)^2}, \qquad k(x) = \frac{(0,0,1)}{\sqrt{\alpha_0 + \alpha_1(x-\mu)^2}}.$$

It is important to notice that g and k are bounded functions. This fact is used in the proof of Propositions 2.3 and 2.4.

Using ergodicity of the process $\{X_t, t \geq 0\}$, we define

$$\Psi(\theta) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} g(X_{t-1}) \quad \text{a.s.} \quad \text{and} \quad \Phi(\theta) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} k(X_{t-1}) \quad \text{a.s.}$$

PROPOSITION 2.3. *Under the conditions of Proposition* 2.1,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{F\left(\left(x+\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{nh(X_{t-1},\hat{\theta})}}\right)\sqrt{1+\frac{1}{\sqrt{n}}g_n(X_{t-1},\sqrt{n}(\hat{\theta}-\theta))}\right)-F(x)\right\}$$
$$= \left\langle (\Phi(\theta)+\frac{1}{2}x\Psi(\theta))f(x), \sqrt{n}(\hat{\theta}-\theta)\right\rangle + o_P(1),$$

where the $o_P(1)$ is uniform in x and $\langle a, b\rangle$ is the inner product of vectors a and b.

PROPOSITION 2.4. *Let*

$$\phi(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{h(X_{t-1},\theta)}} \qquad a.s.$$

Then, under the conditions of Proposition 2.2,

$$\frac{1}{n}\sum_{t=1}^{[nu]}\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{h(X_{t-1},\hat{\theta})}} = u\phi(\theta)\sqrt{n}(\hat{\mu}-\mu)+o_P(1),$$

where the $o_P(1)$ is uniform in u.

Combining Propositions 2.1 and 2.3 yields the following theorem.

THEOREM 2.1. *Suppose that the process $\{X_t, t \geq 0\}$ satisfies the conditions of Propositions* 2.1 *(and hence the conditions of Proposition* 2.3*). Assume that $\hat{\theta}$ is a $\sqrt{n}$-consistent estimate of θ. Then we have*

$$\sup_{x\in\mathbb{R}}\left|E_n(x)-\left\{E_n(x,0)+\left\langle(\Phi(\theta)+\frac{1}{2}x\Psi(\theta))f(x),\sqrt{n}(\hat{\theta}-\theta)\right\rangle\right\}\right| \to 0$$

in probability as $n \to \infty$.

COROLLARY 2.1. *Suppose that $\{E_n(x,0), \sqrt{n}(\hat{\theta}-\theta) : x \in \mathbb{R}\}$ converges weakly to a Gaussian process $\{E(F(x)), Z : x \in \mathbb{R}\}$ on $D(-\infty,+\infty)\times\mathbb{R}^3$, where E is a standard Brownian bridge. Then under the conditions of Theorem* 2.1, *$E_n(x)$ converges weakly to the Gaussian process*

$$E(F(x)) + \left\langle\left(\Phi(\theta)+\frac{1}{2}x\Psi(\theta)\right)f(x), Z\right\rangle.$$

Remark 1. For ARCH model (1.1), that is (1.3) with known $\mu = 0$, the term $\Phi(\theta)f(x)$ will disappear in the limiting Gaussian process after assuming $\sqrt{n}(\hat{\theta} - \theta) = (\sqrt{n}(\hat{\alpha}_0 - \alpha_0), \sqrt{n}(\hat{\alpha}_1 - \alpha_1))^T \Rightarrow Z = (Z_0, Z_1)^T$ in distribution as $n \to \infty$.

Remark 2. Theorem 2.1 (or Corollary 2.1) shows that we cannot treat residuals of a ARCH-M process as if they were the iid innovations $\{\epsilon_t, t \geq 1\}$. A preliminary Monte Carlo simulation has shown that the distribution of functional statistics such as goodness-of-fit tests based on a Kolmogorov-Smirnov does differ from one based on iid sample. In fact, the simulation shows that Kolmogorov-Smirnov test using the usual asymptotic critical values based on iid samples gives a much smaller nominal size under the null hypothesis of an ARCH-M(1) model. The only exception is the case of ARCH(1) [$\mu = 0$ known] with F as standard normal, where this additional term $\langle \frac{1}{2}x\Psi(\theta)f(x), Z\rangle$ plays little role in tests based on a Kolmogorov-Smirnov or Cramer Von Mises statistic based on the form in D'Agostino and Stephens (1986).

We consider a simulation of normal ARCH-M(1) models and the Kolmogorov-Smirnov statistic. The size and critical value of near 1.31, in the case $\mu = 0$ known is close to the asymptotic Kolmogorov-Smirnov size and critical value of 1.358. Table 1 gives the size , in the case of μ is estimated by the MLE, for various ARCH effect parameters α_1 and various sample sizes n when the asymptotic 0.05 critical value of 1.358 is used. Table 2 gives the empirical quantile in this case. When μ is estimated the critical value is near 1.09 and the size is very far from .05 when the standard critical value is used.

Remark 3. For ARCH(1) model, a similar result of Theorem 2.1 (or Corollary 2.1) was obtained by Boldin (1998). However, we believe that his conditions on the distribution function of iid innovations are stronger than ours. In particular, he requires that $F^{-1}(u)f(F^{-1}(u))$, $0 < u < 1$ to be uniformly continuous which may fail for heavy tail distributions. For example, let $f(x) = \exp(-|x|)/2$. Then it is easy to check that $F^{-1}(u)f(F^{-1}(u))$ is not uniformly continuous when u is near 0 or 1. Thus Boldin's result

Table 1. Empirical size of KS statistic for ARCH(1) with μ unknown.

	$n = 100$	$n = 500$	$n = 1000$
$\alpha_1 = 0.85$	0.0048	0.0048	0.0046
$\alpha_1 = 0.90$	0.0061	0.0053	0.0051
$\alpha_1 = 0.95$	0.0060	0.0064	0.0079
$\alpha_1 = 0.98$	0.0055	0.0062	0.0091

Table 2. 95% empirical quantile of KS statistic for ARCH(1) with μ unknown.

	$n = 100$	$n = 500$	$n = 1000$
$\alpha_1 = 0.85$	1.062093	1.087335	1.088622
$\alpha_1 = 0.90$	1.071449	1.098601	1.096711
$\alpha_1 = 0.95$	1.079876	1.107994	1.114918
$\alpha_1 = 0.98$	1.098522	1.112108	1.108957

could not be applied while ours is valid for the given $f(x)$. Our method of proof is different allowing us to avoid conditions such as his uniform continuity of $F^{-1}(u)f(F^{-1}(u))$, $0 < u < 1$.

In the following, we consider a technique called sub-sampling to deal with the problem of distribution dependent limiting Gaussian process in Corollary 2.1. We construct the empirical process $\tilde{E}_m(x)$ based on the first m ($< n$) residuals as

$$\tilde{E}_m(x) = \sqrt{m}(\hat{F}_m(x) - F(x)) = \frac{1}{\sqrt{m}} \sum_{t=1}^{m} (\mathrm{I}(\hat{\epsilon}_t \le x) - F(x)), \qquad -\infty < x < \infty$$

and the process $\tilde{E}_m(x, s)$ as

$$\begin{aligned}\text{(2.16)} \qquad \tilde{E}_m(x,s) = \frac{1}{\sqrt{m}} \sum_{t=1}^{m} \Bigg\{ \mathrm{I}\Bigg(\epsilon_t \le \Bigg(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}}\Bigg) \\ \cdot \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)}\Bigg) \\ -F\Bigg(\Bigg(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}}\Bigg) \\ \cdot \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)}\Bigg)\Bigg\}.\end{aligned}$$

THEOREM 2.2. *Assume that $m = m(n) \to \infty$ and $m = o(n)$ as $n \to \infty$. Then under the conditions of Theorem 2.1, we have*

$$\sup_{x \in \mathbb{R}} |\tilde{E}_m(x) - \tilde{E}_m(x, 0)| \to 0$$

in probability as $n \to \infty$. Hence $\tilde{E}_m(x)$ converges weakly to $E(F(x))$.

Theorem 2.2 enables us to construct several asymptotic test statistics for statistical inference about the unknown distribution $F(x)$. A well known statistic will be the Kolmogorov-Smirnov goodness-of-test. Theorem 2.2 implies that as $n \to \infty$

$$P\left(\sup_{x \in \mathbb{R}} |\tilde{E}_m(x)| \ge K_\delta\right) \to P\left(\sup_{0 \le u \le 1} |E(u)| \ge K_\delta\right) = \delta,$$

where $K_\delta = 1.3581$ for $\delta = 0.05$ and $K_\delta = 1.6276$ for $\delta = 0.01$. These are the usual critical values based on a limit obtained from iid r.v.'s.

Remark 4. Even though the empirical process $\tilde{E}_m(x)$ is based on only part of residuals, residuals themselves are constructed from the whole data, that is, the estimation $\hat{\theta}$ is based on the whole data. For many discrete time financial data, the sample sizes are often in the order of several hundred or several thousand. Hence this sub-sampling is feasible to implement. For example, we can choose $m = [n/\log n]$ and the empirical process $\tilde{E}_m(x)$ can be constructed from any consecutive data block of residuals in size m. This sub-sampling can safeguard against some impacts that may come from the additional term in Corollary 2.1 although Monte Carlo simulation has shown that it plays little role in statistical inference, at least in a normal ARCH case (cf. Remark 2).

Combining Propositions 2.2 and 2.4 yields the following theorem.

THEOREM 2.3. *Suppose that the process $\{X_t, t \geq 0\}$ obeys the conditions of Proposition 2.2 and hence of Proposition 2.4. Assume that $\hat{\theta}$ is a $\sqrt{n}$-consistent estimate of θ. Then we have*

$$\sup_{0 \leq u \leq 1} |B_n(x) - \{B_n(u,0) - u\phi(\theta)\sqrt{n}(\hat{\mu} - \mu)\}| \to 0$$

in probability as $n \to \infty$.

COROLLARY 2.2. *Suppose that $\{B_n(u,0), \sqrt{n}(\hat{\mu} - \mu) : 0 \leq u \leq 1\}$ converges weakly to a Gaussian process $\{B(u), Z_2 : 0 \leq u \leq 1\}$ on $D[0,\ 1] \times \mathbb{R}$, where B is a standard Brownian motion. Then under the conditions of Theorem 2.3, $B_n(u)$ converges weakly to the Gaussian process*

$$B(u) - u\phi(\theta)Z_2.$$

Remark 5. For ARCH model (1.1), that is (1.3) with known $\mu = 0$, the term $u\phi(\theta)Z_2$ will disappear in the limiting Gaussian process. Hence the limit process of $B_n(u)$ is just a standard Brownian motion.

Remark 6. The $\Phi(\theta)$ in Proposition 2.3 and $\phi(\theta)$ in Proposition 2.4 are essentially the same. The vector version $\Phi(\theta)$ is used for the sake of a simple representation in Proposition 2.3 and hence the limiting Gaussing process in Corollary 2.1.

Based on the same sub-sampling technique, we redefine the partial sum process $\tilde{B}_m(u)$ as

$$\tilde{B}_m(u) = \frac{1}{\sqrt{m}} \hat{S}_{[mu]}, \qquad 0 \leq u \leq 1$$

and the process $\tilde{B}_m(u,s)$ as

$$\tilde{B}_m(u,s) = \frac{1}{\sqrt{m}} \sum_{t=1}^{[mu]} \hat{\epsilon}_t(s), \qquad 0 \leq u \leq 1. \tag{2.17}$$

THEOREM 2.4. *Assume that $m = m(n) \to \infty$ and $m = o(n)$ as $n \to \infty$. Then under the conditions of Theorem 2.3, we have*

$$\sup_{0 \leq u \leq 1} |\tilde{B}_m(u) - \tilde{B}_m(u,0)| \to 0$$

in probability as $n \to \infty$. Hence $\tilde{B}_m(u)$ converges weakly to standard Brownian motion $B(u)$.

For an ARCH-M(p) process the following changes must be made. With these changes, Propositions 2.1 to 2.4 and Theorems 2.1 to 2.4 are valid. In addition,

$$g_n(x_1,\dots,x_p,s) = \frac{\sqrt{n}(h(x_1,\dots,x_p,\theta+n^{-1/2}s) - h(x_1,\dots,x_p,\theta))}{h(x_1,\dots,x_p,\theta)}$$

is needed to obtain analogue of the proofs which are otherwise identical to the case ARCH-M(1).

(i) $\theta = (\alpha_0,\alpha_1,\dots,\alpha_p,\mu)$ is $\mathbb{R}^{p+2}$ valued and $\{X_t, t\geq 0\}$ is stationary and ergodic. To be specific, suppose it satisfies the conditions of An *et al.* (1997)

(ii) $h : \mathbb{R}^{2p+2} \mapsto \mathbb{R}^+$

$$h(x_1,\dots,x_p,\theta) = \alpha_0 + \sum_{i=1}^{p} \alpha_i (x_i-\mu)^2.$$

(iii) $g : \mathbb{R}^p \mapsto \mathbb{R}^{p+2}$

$$g(x_1,\dots,x_p) = \frac{1}{h(x_1,\dots,x_p,\theta)}\left(1,(x_1-\mu)^2,\dots,(x_p-\mu)^2,-2\sum_{i=1}^{p}\alpha_i(x_i-\mu)\right).$$

(iv) $k : \mathbb{R}^p \mapsto \mathbb{R}^{p+2}$

$$k(x_1,\dots,x_p) = \frac{(0,\dots,0,1)}{\sqrt{h(x_1,\dots,x_p,\theta)}}.$$

(v) Ψ is replaced by

$$\Psi(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=p}^{n} g(X_{t-1},\dots,X_{t-p}) \qquad \text{a.s.}$$

(vi) Φ is replaced by

$$\Phi(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=p}^{n} k(X_{t-1},\dots,X_{t-p}) \qquad \text{a.s.}$$

(vii) ϕ is replaced by

$$\phi(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=p}^{n} \frac{1}{\sqrt{h(X_{t-1},\dots,X_{t-p},\theta)}} \qquad \text{a.s.}$$

3. Proofs

This section contains proofs for Propositions 2.1 to 2.4, and the Theorems 2.1 to 2.4. Throughout this section it is assumed that $\hat{\theta}$ is a $\sqrt{n}$-consistent estimate of θ. The following lemma which will be used throughout this section.

LEMMA 3.1. *Let $b > 1$ be any fixed constant. Then for any sequence $\{X_t, t\geq 0\}$, we have that for any $n \geq 1$*

$$\sup_{\|s\|_\infty\leq b}\max_{1\leq t\leq n} |g_n(X_{t-1},s)| \leq C(\theta)b^3, \tag{3.1}$$

and for $n > \max(b^2/\alpha_1^2, 4b^2/\alpha_0^2)$

$$\inf_{\|s\|_\infty \le b} \min_{1\le t\le n} h(X_{t-1}, \theta + n^{-1/2}s) \ge \alpha_0 - b/\sqrt{n} > \alpha_0/2, \tag{3.2}$$

$$\sup_{\|s\|_\infty \le b} \max_{1\le t\le n} \left| \frac{1}{\sqrt{h(X_{t-1}, \theta + n^{-1/2}s)}} - \frac{1}{\sqrt{h(X_{t-1}, \theta)}} \right| \le \frac{\sqrt{2}C(\theta)b^3}{\sqrt{\alpha_0 n}}, \tag{3.3}$$

where $C(\theta)$ *is a positive constant depending on* θ *only.*

PROOF. (3.1) and (3.2) are followed easily by using (2.10) and (2.11), respectively, which in terms imply (3.3) by the definition of g_n and straightforward calculation. This completes the proof of Lemma 3.1.

3.1 *Proof of Proposition* 2.3

Applying a first order Taylor approximation with remainder to the second term of (2.14) gives

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} f(x) & \left(\left(x + \frac{\sqrt{n}(\hat\mu - \mu)}{\sqrt{nh(X_{t-1}, \hat\theta)}} \right) \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, \sqrt{n}(\hat\theta - \theta))} - x \right) \\ & + R_n(x, \sqrt{n}(\hat\theta - \theta)), \end{aligned} \tag{3.4}$$

where $R_n(x, \sqrt{n}(\hat\theta - \theta))$ is the remainder term.

LEMMA 3.2. *Under the conditions of Proposition* 2.3, *we have*

$$\sup_{x\in\mathbb{R}} |R_n(x, \sqrt{n}(\hat\theta - \theta))| = o_P(1).$$

PROOF. Let $b > 1$ be a fixed constant. Since $|\sqrt{1+v} - 1| \le |v|$ for $|v| \le 1$, by (3.1) and (3.2), we have that for any $x \in \mathbb{R}$ and large n

$$\sup_{\|s\|_\infty \le b} \max_{1\le t\le n} \left| \left(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}} \right) \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)} - x \right| \le \frac{D(|x|+1)}{\sqrt{n}},$$

where D is a constant depending on θ and b only.

Let $M > 1$ be a constant, chosen later to be suitably large. Using the conditions in Proposition 2.3 and the bound above, one can obtain a simple bound on the remainder R_n of (3.4) as

$$\begin{aligned} & \sup_{\|s\|_\infty \le b; x\in\mathbb{R}} |R_n(x, s)| \\ & \le \frac{1}{\sqrt{n}} \sup_{\|s\|_\infty \le b; x\in\mathbb{R}} \sup_{|y-x|\le D(|x|+1)/\sqrt{n}} |f(y) - f(x)| \\ & \quad \cdot \sum_{t=1}^{n} \left| \left(x + \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}} \right) \sqrt{1 + \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)} - x \right| \\ & \le D \sup_{|x|>M} (|x|+1) \sup_{|y-x|\le D(|x|+1)/\sqrt{n}} |f(y) - f(x)| \\ & \quad + D \sup_{|x|\le M} (|x|+1) \sup_{|y-x|\le D(|x|+1)/\sqrt{n}} |f(y) - f(x)|. \end{aligned} \tag{3.5}$$

The second last term in (3.5) is bounded by (except the constant D)

$$\begin{aligned}
&\sup_{|x|>M} 2|x|f(x) + \sup_{|x|>M} 2|x| \sup_{|y-x|\le 2D|x|/\sqrt{n}} f(y) \\
&\le \sup_{|x|>M} 2|x|f(x) + \sup_{|y|>M(1-2D/\sqrt{n})} 2|y|f(y) \sup_{|y-x|\le 2D|x|/\sqrt{n}} \left|\frac{x}{y}\right| \\
&= o(1) + o(1)O(1) = o(1)
\end{aligned}$$

for all n sufficiently large, as $M \to \infty$. The last term in (3.5) is $o(1)$ as $n \to \infty$ on $|x| \le M$ by the continuity of f.

Thus on the set $\{\sqrt{n}\|\hat{\theta} - \theta\|_\infty \le b\}$, R_n converges to zero uniformly in x, and hence Lemma 3.2 is proved by letting $b \to \infty$, after taking $n \to \infty$, in the following inequality

$$\begin{aligned}
P\left(\sup_{x\in\mathbb{R}} |R_n(x, \sqrt{n}(\hat{\theta} - \theta))| \ge \epsilon\right) &\le P\left(\sup_{\|s\|_\infty \le b; x\in\mathbb{R}} |R_n(x, s)| \ge \epsilon\right) \\
&\quad + P(\sqrt{n}\|\hat{\theta} - \theta\|_\infty > b).
\end{aligned}$$

To prove Proposition 2.3, we use the same technique that is used in proving Lemma 3.2, that is to work with the first sum in (3.4) on the set $\{\sqrt{n}\|\hat{\theta} - \theta\|_\infty \le b\}$ only.

It is easy to verify that $|\sqrt{1+v} - 1 - v/2| \le v^2$ for $|v| \le 1/2$. Hence, by Lemma 3.1 and $\sup_{x\in\mathbb{R}} |x|f(x) < \infty$, the first sum in (3.4) becomes

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{h(X_{t-1}, \theta)}} + \frac{1}{2}x g_n(X_{t-1}, \sqrt{n}(\hat{\theta} - \theta))\right) f(x) + o_P(1),$$

where the $o_P(1)$ is uniform in x. Finally Proposition 2.3 can be proved by (2.9) and the definitions of $\Psi(\theta)$ and $\Phi(\theta)$.

3.2 *Proof of Proposition* 2.1

The following Lemma will be used repeatedly for finding upper bound probabilities for the increments of the process (2.12). It appears in Levental (1989) and it was independently obtained by Hitczenko (1990).

LEMMA 3.3. *Let $\{d_i, \mathcal{F}_i\}$ be a martingale difference sequence with $E(d_j \mid \mathcal{F}_{j-1}) = 0$, $|d_j| \le c$, for $0 < c < \infty$, $E(d_j^2 \mid \mathcal{F}_{j-1}) = \sigma_j^2$ and $V_n^2 = \sum_{i=1}^n \sigma_i^2$. Then, for all $x, y > 0$,*

$$P\left(\left|\sum_{i=1}^{n} d_i\right| \ge x, V_n^2 \le y \text{ for some } n\right) \le 2\exp\left\{-\frac{x}{2c}\operatorname{arcsinh}\left(\frac{cx}{2y}\right)\right\}. \tag{3.6}$$

PROOF OF PROPOSITION 2.1. The technique employed in the proof resembles the approach undertaken by Koul (1992) and Kawczak (1998) when dealing with the additive structure of the perturbations under the indicator function. Since we are considering an empirical process with the multiplicative terms under the indicator function, a new scheme for the proof had to be developed. The changes are explained where the Proposition 2.1 is proven.

The proof is split into two parts. First, we establish that for each fixed s

$$\sup_{x\in\mathbb{R}} |E_n(x,s) - E_n(x,0)| = o_p(1). \tag{3.7}$$

Using this result we then prove that

$$\sup_{\|s\|_\infty \le b} \sup_{x\in\mathbb{R}} |E_n(x,s) - E_n(x,0)| = o_p(1). \tag{3.8}$$

On $\mathbb{R}$ define the pseudo-metric $\rho(x,y) := |F(x) - F(y)|^{(1/2)}$ for all $x, y \in \mathbb{R}$. Since the limiting process is a Gaussian process it is natural to introduce the pseudo-metric ρ and study the ρ-equicontinuity of the paths. The choice of the metric is dictated by the form of the covariance function of limiting process of $E_n(x,s)$. The square root ensures that the variance of the process is less than ρ^2. The metric ρ makes $\mathbb{R}$ totally bounded. Hence the closure of the index space becomes a compact set with the metric ρ. Therefore, to prove (3.7), it suffices to prove (A) and (B) below

(A) $\forall\, x \in \mathbb{R}$, $|E_n(x,s) - E_n(x,0)| = o_p(1)$

(B) $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{\rho(x,y)\le\delta} |E_n(x,s) - E_n(y,s)| > \varepsilon\right) = c(\delta),$$

where $c(\delta) \to 0$ as $\delta \to 0^+$.

PROOF OF (A). Let

$$Y_{1,t} := \frac{s_2}{\sqrt{nh(X_{t-1}, \theta + n^{-1/2}s)}}$$

$$Y_{2,t} := \frac{1}{\sqrt{n}} g_n(X_{t-1}, s)$$

and define

$$D_t^{[1]}(x) := \mathrm{I}(\epsilon_t \le (x + Y_{1,t})\sqrt{1 + Y_{2,t}}) - F((x + Y_{1,t})\sqrt{1 + Y_{2,t}})$$

and

$$D_t^{[2]}(x) := \mathrm{I}(\epsilon_t \le x) - F(x).$$

Then for every s, $\{D_t^{[1]} - D_t^{[2]}, \mathcal{F}_t\}$, where $\mathcal{F}_t = \sigma\{X_0, \epsilon_1, \epsilon_2, \ldots, \epsilon_{t-1}\}$, is a martingale difference sequence. Thus, by taking $d_t = (D_t^{[1]} - D_t^{[2]})/\sqrt{n}$, with $c = 1/\sqrt{n}$, and using Lemma 3.3 we get that $\forall\, \varepsilon > 0$ and $y > 0$

$$P(|E_n(x,s) - E_n(x,0)| > \varepsilon) \le 2\exp\left\{-\frac{\varepsilon\sqrt{n}}{2} \operatorname{arcsinh}\left(\frac{\varepsilon}{2\sqrt{ny}}\right)\right\} + P(V_n^2 > y). \tag{3.9}$$

where V_n^2 is the variance of the conditionally centered bounded random variables in $E_n(x,s) - E_n(x,0)$. Since s is fixed and $\sup_{x\in\mathbb{R}} |x| f(x) < \infty$, by a calculation similar to the proof of Proposition 2.3 (setting $\sqrt{n}(\hat{\theta} - \theta) = s$; see Subsection 3.1), it can be shown that

$$V_n^2 \le \frac{1}{n}\sum_{t=1}^{n} |F((x + Y_{1,t})\sqrt{1 + Y_{2,t}}) - F(x)| \le \frac{C}{\sqrt{n}},$$

where $C > 0$ is a constant independent of x. Thus, by taking $y = C/\sqrt{n}$, we have that $P(V_n^2 > y) = 0$.

This proves (A).

PROOF OF (B). The proof relies on the restricted chaining argument as presented in Pollard (1984), and a repeated application of the exponential inequality in Lemma 3.3.

For a fixed $\varepsilon > 0$ define the mesh grid

$$\mathcal{G}_n := \left\{x_k; x_k = F^{-1}\left(\frac{k\varepsilon}{\sqrt{n}}\right), k = 0, 1, \ldots, k_n = \left[\frac{\sqrt{n}}{\varepsilon}\right]\right\}, \qquad n \geq 1,$$

where $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 \leq u < 1$ is the usual quantile function. Since we assume that f is positive on the open support of F, F^{-1} is continuous on $(0, 1)$. Thus, for any large $M > 0$,

$$\lim_{n\to\infty} \max_{k: -M \leq x_k \leq x_{k+1} \leq M} |x_{k+1} - x_k| = 0. \tag{3.10}$$

As it is explained in Pollard ((1984), p. 160), the chaining can continue in an ordinary way until the "little" links come to start contributing to the cumulative sum of the increments. Special care has to be taken regarding those small links. For $u > 0$, consider a pair of points in $\mathcal{G}_n$ that are at least $u\varepsilon/\sqrt{n}$ apart in the ρ^2 metric (that is $\rho^2(x, y) \geq u\varepsilon/\sqrt{n}$). Then, by applying Lemma 3.3 to $\{D_t^{[1]}(x) - D_t^{[1]}(y), \mathcal{F}_t\}$, we get

$$P(|E_n(x, s) - E_n(y, s)| > u) \leq 2 \exp\left\{-\frac{u\sqrt{n}}{2} \operatorname{arcsinh}\left(\frac{\varepsilon}{2\sqrt{n}v}\right)\right\} + P(V_n^2 > v)$$

for all $u > 0$ and $v > 0$. With the same argument for V_n^2 in the proof of (A), one obtains that

$$V_n^2 \leq \frac{1}{n}\sum_{t=1}^{n} |F((x + Y_{1,t})\sqrt{1 + Y_{2,t}}) - F((y + Y_{1,t})\sqrt{1 + Y_{2,t}})| \leq \frac{2C}{\sqrt{n}} + \rho^2(x, y).$$

Hence letting $v = (\frac{2C}{u\varepsilon} + 1)\rho^2(x, y)$ and using the fact that $\operatorname{arcsinh}(x) \geq x/2$ for small $x > 0$, we obtain that

$$P(|E_n(x, s) - E_n(y, s)| > u) \leq 2 \exp\left\{-\frac{(u\varepsilon)^2}{8(2C + u\varepsilon)\rho^2(x, y)}\right\}.$$

At this point, the only thing left is to connect the points from $\mathbb{R}$ with a point in $\mathcal{G}_n$ and the chaining will go as in Pollard.

For each $x \in \mathbb{R}$, define $x_{k(x)} \in \mathcal{G}_n$ such that it is closest to x in the ρ-metric amongst all points in $\mathcal{G}_n$ and the relationship is satisfied: $x_{k(x)} \leq x$. The following needs to be proved in order to establish (B):

$$\forall\, \varepsilon > 0 \quad \limsup_{n\to\infty} P\left(\sup_{x\in\mathbb{R}} |E_n(x, s) - E_n(x_{k(x)}, s)| > 3\varepsilon\right) = 0. \tag{3.11}$$

PROOF OF (3.11). Notice that by Lemma 3.1, for large n, $\max_{1\le t\le n}|Y_{i,t}| = O(1/\sqrt{n})$, $i = 1, 2$. Using the properties of the indicator function and the monotonicity on F we get

$$\begin{aligned}\sup_{x\in\mathbb{R}} &|E_n(x,s) - E_n(x_{k(x)}, s)| \\ &\le \max_{0\le k\le k_n-1} |E_n(x_{k+1}, s) - E_n(x_k, s)| \\ &\quad + \max_{0\le k\le k_n-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F((x_{k+1} + Y_{1,t})\sqrt{1+Y_{2,t}}) \\ &\qquad\qquad - F((x_k + Y_{1,t})\sqrt{1+Y_{2,t}})).\end{aligned} \tag{3.12}$$

Taking $s = \sqrt{n}(\hat{\theta} - \theta)$, we apply Proposition 2.3 to the last part of the above inequality and obtain

$$\begin{aligned}&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F((x_{k+1} + Y_{1,t})\sqrt{1+Y_{2,t}}) - F((x_k + Y_{1,t})\sqrt{1+Y_{2,t}})) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F((x_{k+1} + Y_{1,t})\sqrt{1+Y_{2,t}}) - F(x_{k+1})) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F((x_k + Y_{1,t})\sqrt{1+Y_{2,t}}) - F(x_k)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F(x_{k+1}) - F(x_k)) \\ &= \left\langle \left(\Phi(\theta) + \frac{1}{2}x_{k+1}\Psi(\theta)\right) f(x_{k+1}), s \right\rangle - \left\langle \left(\Phi(\theta) + \frac{1}{2}x_k\Psi(\theta)\right) f(x_k), s \right\rangle \\ &\quad + \eta(x_{k+1}) + \eta(x_k) + \varepsilon,\end{aligned}$$

where $\eta(x_{k+1})$ and $\eta(x_k)$ are the uniform $o_p(1)$ terms from Proposition 2.3. Thus

$$\begin{aligned}\max_{0\le k\le k_n-1} &\frac{1}{\sqrt{n}} \left| \sum_{t=1}^{n} (F((x_{k+1} + Y_{1,t})\sqrt{1+Y_{2,t}}) - F((x_k + Y_{1,t})\sqrt{1+Y_{2,t}})) \right| \\ &\le \max_{0\le k\le k_n-1} |f(x_{k+1}) - f(x_k)| |\langle \Phi(\theta), s\rangle| \\ &\quad + \max_{0\le k\le k_n-1} |x_{k+1}f(x_{k+1}) - x_k f(x_k)| \left| \left\langle \frac{1}{2}\Psi(\theta), s \right\rangle \right| \\ &\quad + \max_{0\le k\le k_n-1} |\eta(x_{k+1}) + \eta(x_k)| + \varepsilon.\end{aligned}$$

Since f is continuous and $\lim_{x\to\infty} |x|f(x) = 0$, similar to the proof of Lemma 3.2, by (3.10), the first two parts tend to zero as $n \to \infty$. This proves

$$\begin{aligned}\lim_{n\to\infty} P \Bigg(\max_{0\le k\le k_n-1} \frac{1}{\sqrt{n}} \Bigg| &\sum_{t=1}^{n} (F((x_{k+1} + Y_{1,t})\sqrt{1+Y_{2,t}}) \\ &- F((x_k + Y_{1,t})\sqrt{1+Y_{2,t}})) \Bigg| > 2\varepsilon \Bigg) = 0.\end{aligned} \tag{3.13}$$

Next, we consider the first part of the RHS of equation (3.12).

$$P\left(\max_{0\leq k\leq k_n-1}|E_n(x_{k+1},s)-E_n(x_k,s)|>\varepsilon\right) \leq \sum_{k=0}^{k_n-1} P\left(\left|\sum_{t=1}^{n}(D_t^{[1]}(x_{k+1})-D_t^{[1]}(x_k))\right|>\varepsilon\sqrt{n}\right). \tag{3.14}$$

Yet another application of the Lemma 3.3, together with (3.13), gives us upper bound on the probability in equation (3.14)

$$\sum_{k=0}^{k_n-1} P\left(\left|\sum_{t=1}^{n}(D_t^{[1]}(x_{k+1})-D_t^{[1]}(x_k))\right|>\varepsilon\sqrt{n}\right) \leq \frac{3\sqrt{n}}{\varepsilon}\exp\{-O(\sqrt{n})\}+o(1)\to 0.$$

Now the proof of (B) is complete.

Next, we prove the uniform closeness of the processes expressed by equation (3.8).

Proof of (3.8). The proof relies on repeated application of the result from equation (3.7) and the compactness property of the set $\mathcal{S}(b) := \{s : \|s\|_\infty \leq b\}$.

The proof of equation (3.7) needs to be modified to accommodate the multiplicative nature of the perturbations under the indicator functions in (2.12). It is easily recognized that when $x \in [-M, M]$ for $M < \infty$ the proof of equation (3.8) will follow from equation (3.7), the compactness of $\mathcal{S}(b)$ and the proof of Theorem 3.2.9 in Kawczak (1998). Therefore, we should concentrate on the case when x is outside of $[-M, M]$. First, let us look at the case when $x > M$. Notice that the technique used so far cannot be utilized in the direct fashion because of the complication arising from the multiplicity rather than the additivity of the terms under the indicator function. The following provides the remedy to the problem.

Expanding $\sqrt{1+x} = 1+\frac{1}{2}x-\frac{1}{8}x^2\cdots$ yields

$$\left(x+\frac{s_2}{\sqrt{nh(X_{t-1},\theta+n^{-1/2}s)}}\right)\sqrt{\left(1+\frac{1}{\sqrt{n}}g_n(X_{t-1},s)\right)} = x\left(1+\frac{1}{2\sqrt{n}}g_n(X_{t-1},s)\right)+\frac{s_2}{\sqrt{nh(X_{t-1},\theta+n^{-1/2}s)}}+O_p(n^{-1}). \tag{3.15}$$

Thus it is sufficient to consider the following empirical process

$$\mathcal{G}_n(x,\nu,\psi) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\mathrm{I}\left(\epsilon_t\leq x\left(1+\frac{\nu\xi_{1t}}{\sqrt{n}}\right)+\frac{\psi\xi_{2t}}{\sqrt{n}}\right) - F\left(x\left(1+\frac{\nu\xi_{1t}}{\sqrt{n}}\right)+\frac{\psi\xi_{2t}}{\sqrt{n}}\right)\right\} \tag{3.16}$$

for bounded random variables, ξ_{1t} and ξ_{2t} that are $\mathcal{F}_{t-1}$-measurable and independent of ϵ_t. ν and ψ constants in $\mathbb{R}$.

The empirical partial sum processes with an additive term like $\frac{\psi\xi_{2t}}{\sqrt{n}}$ have been treated by Boldin (1982), Koul (1992) and Kawczak (1998), among others. However, by mimicking the proof of (3.7) the conditions required are as stated, thus we do not need verify any additional conditions such as higher moments as used by Boldin (1982) and Koul (1992).

It follows that further simplification to the process is possible. Hence, the process of interest becomes

$$\mathcal{H}_n(x,\nu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \mathrm{I}\left(\varepsilon_t \le x\left(1+\frac{\nu\xi_{1t}}{\sqrt{n}}\right)\right) - F\left(x\left(1+\frac{\nu\xi_{1t}}{\sqrt{n}}\right)\right)\right\}. \tag{3.17}$$

Therefore, the equation (3.8) will be implied by

$$\sup_{|\nu|\le b}\sup_{x>M} \|H_n(x,\nu) - H_n(x,0)\| = o_p(1). \tag{3.18}$$

The case of $x < -M$ can be treated in an analogous manner. Because of multiplicative form of the fluctuations we need to devise the partition on x in a suitable way. One possibility is to take

$$K_{n,k} := \left\{ x_k : x_k = \ln F^{-1}\left(\frac{k\varepsilon}{\sqrt{n}}\right), k(M) \le k \le \left[\frac{\sqrt{n}}{\varepsilon}\right]\right\}, \tag{3.19}$$

where $k(M) = [\frac{F(e^M)\sqrt{n}}{\varepsilon}]$ for the given $\varepsilon > 0$.

For $x_k \le x < x_{k+1}$ we have

$$\begin{aligned}
&\|\mathcal{H}_n(x,\nu) - \mathcal{H}_n(x,0)\| \\
&\quad \le \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{\mathrm{I}(x_k \le \varepsilon_t \le x_{k+1}) - F(x_{k+1}) + F(x_k)\}\right| \\
&\quad + \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ F(x_{k+1}) - F(x_k) - F\left(x\left(1+\frac{\nu\xi_{1t}}{\sqrt{n}}\right)\right) + F(x)\right\}\right|.
\end{aligned}$$

The form of the RHS of the above display allows us to adopt the technique of the proof of Theorem 3.2.9 in Kawczak (1998). Hence, most of the details are omitted. Thus, we get the equation (3.18) for x outside the region $[-M, M]$. Now, combining it with the remaining part for the $x \in [-M, M]$ we get

$$\sup_{|\nu|\le b}\sup_{x\in\mathbb{R}} \|H_n(x,\nu) - H_n(x,0)\| = o_p(1). \tag{3.20}$$

This concludes the proof of Proposition 2.1.

3.3 *Proof of Proposition* 2.2

Using the inequality $|1/\sqrt{1+x} - 1 + x/2| \le x^2$ for $|x| \le 1/2$, (2.6), (2.13), and the local boundedness of g_n property (3.1), one obtains

$$\begin{aligned}
&\sup_{\|s\|_\infty \le b}\sup_{0\le u\le 1} |B_n(u,s) - B_n(u,0)| \\
&\quad \le \sup_{\|s\|_\infty \le b}\sup_{0\le u\le 1} \left| \frac{1}{2n}\sum_{t=1}^{[nu]} \epsilon_t g_n(X_{t-1},s)\right| + \frac{1}{n^{3/2}} C^2(\theta) b^6 \sum_{t=1}^{n} |\epsilon_t|.
\end{aligned}$$

The second term on the right hand side of the above inequity tends to zero in probability, since ϵ_t has a finite first moment. As to the first term on the right hand side, by (2.9), we obtain

$$\sup_{\|s\|_\infty \le b} \sup_{0\le u\le 1} \left| \frac{1}{2n} \sum_{t=1}^{[nu]} \epsilon_t g_n(X_{t-1}, s) \right| \le \sum_{i=1}^{3} \frac{b^i}{n^{(i-1)/2}} \max_{1\le k\le n} \left| \frac{1}{n} \sum_{t=1}^{k} \epsilon_t g_{i,n}(X_{t-1}) \right|,$$

where $g_{i,n}(x)$, $i = 1, 2, 3$, are defined by those terms in (2.9) after all s's are removed. The $g_{i,n}(x)$, $i = 1, 2, 3$, are bounded functions of x and independent of s. Thus Proposition 2.2 can be proved if we can show that

$$\max_{1\le k\le n} \left| \frac{1}{n} \sum_{t=1}^{k} \epsilon_t g_{i,n}(X_{t-1}) \right| \to 0$$

in probability as $n \to \infty$. This can be proved immediately by the fact that $\{\epsilon_t g_{i,n}(X_{t-1}), \mathcal{F}_t\}$ is a martingale difference sequence and by Kolmogorov's maximum inequality with the assumption $E\epsilon_1^2 < \infty$.

3.4 *Proof of Proposition* 2.4

By (3.3) in Lemma 3.1, we have for any $b > 1$

$$\sup_{\|s\|_\infty \le b} \sup_{0\le u\le 1} \frac{1}{n} \left| \sum_{t=1}^{[nu]} \frac{s_2}{\sqrt{h(X_{t-1}, \theta + n^{-1/2}s)}} - \sum_{t=1}^{[nu]} \frac{s_2}{\sqrt{h(X_{t-1}, \theta)}} \right| \le \frac{\sqrt{2}C(\theta)b^4}{\sqrt{\alpha_0 n}}.$$

Hence, to complete the proof of Proposition 2.4, we need to verify on the set $\{\sqrt{n}\|\hat{\theta}-\theta\| \le b\}$ that

$$\sup_{0\le u\le 1} \left| \frac{1}{n} \sum_{t=1}^{[nu]} \frac{1}{\sqrt{h(X_{t-1}, \theta)}} - u\phi(\theta) \right| = o_p(1). \tag{3.21}$$

To this end, by (3.2), we have for any $0 < \eta < 1$ and large n,

$$\begin{aligned} \sup_{0\le u\le 1} \left| \frac{1}{n} \sum_{t=1}^{[nu]} \frac{1}{\sqrt{h(X_{t-1}, \theta)}} - u\phi(\theta) \right| &\le O(\eta) + \sup_{\eta\le u\le 1} \left| \frac{1}{n} \sum_{t=1}^{[nu]} \frac{1}{\sqrt{h(X_{t-1}, \theta)}} - u\phi(\theta) \right| \\ &\le O(\eta) + \sup_{k\ge [n\eta]} \left| \frac{1}{k} \sum_{t=1}^{k} \frac{1}{\sqrt{h(X_{t-1}, \theta)}} - \phi(\theta) \right|. \end{aligned}$$

The last term in the above inequality converges to 0 in probability by the definition of convergence almost surely, i.e., as $n \to \infty$,

$$\sup_{k\ge n} \left| \frac{1}{k} \sum_{t=1}^{k} \frac{1}{\sqrt{h(X_{t-1}, \theta)}} - \phi(\theta) \right| = o_P(1).$$

This proves (3.21).

3.5 *Proofs of Theorems* 2.1 *to* 2.4

The proofs of Theorems 2.1 and 2.3 rely on the same technique that is used in proving Propositions 2.1 to 2.4, that is one just needs to work on the set $\{\sqrt{n}\|\hat{\theta} - \theta\|_\infty \le b\}$ for any fixed $b > 0$. Thus Theorem 2.1 follows easily from (2.14) and Propositions 2.1 and 2.3. Similarly, Theorem 2.3 follows from (2.15) and Propositions 2.2 and 2.4.

To prove Theorems 2.2 and 2.4, we need to slightly modify the proofs of Propositions 2.3 and 2.3 respectively. Specifically, there are only m terms in the sum instead of n terms. However, because $\hat{\theta}$ is still assumed to be $\sqrt{n}$ consistent, an extra factor $\sqrt{m}/\sqrt{n}$ appears in the right side of representations in Propositions 2.3 and 2.4. Hence as $n \to \infty$, they will converge in a sup norm in probability to 0. Thus Theorems 2.2 and 2.4 follow easily. The details are omitted.

Acknowledgements

The authors would like to thank anonymous referees for their valuable comments that improved overall presentation and for pointing us to most recent articles that are relevant to our paper.

References

An, H.-Z., Chen, M. and Huang, F. (1997). The geometric ergodicity and existence of moments for a class of non-linear time series model, *Statistics and Probability Letters*, **31**, 213–224.

Boldin, M. V. (1982). Estimation of the distribution of the noise in an autoregression scheme, *Theory of Probability and Its Applications* (Translation of *Teorija Verojatnostei i ee Primenenija*), **27**, 866–871.

Boldin, M. V. (1998). On residual empirical distribution functions in ARCH models with applications to testing and estimation, *Mitteilungen aus dem Mathematischen Seminar Giessen*, **235**, 49–66.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

D'Agostino, R. B. and Stephens, M. A. (eds.) (1986). *Goodness of Fit Techniques*, Marcel Dekker, New York.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, **50**, 987–1008.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1982). Estimating time varying risk premia in the term structure: ARCH-M model, *Econometrica*, **55**, 391–407.

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

Hitczenko, P. (1990). Best constants in martingale version of Rosenthal's inequality, *The Annals of Probability*, **18**, 1656–1668.

Horváth, L., Kokoszka, P. and Teyssière, G. (2001). Empirical processes of the squared residuals of an ARCH sequence, *The Annals of Statistics*, **29**(2), 445–469.

Kawczak, J. (1998). Weak convergence of a certain class of residual empirical processes, Ph.D. Thesis, The University of Western Ontario.

Koul, H. L. (1992). *Weighted Empirical and Linear Models*, Lecture Notes-Monograph Series, Vol. 21, Institute of Mathematical Statistics, Hayward, California.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed., Springer-Verlag, New York.

Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale differences, *Journal of Theoretical Probability*, **2**, 271–287.

Pollard, D. (1984).*Convergence of Stochastic Processes*, Springer-Verlag, New York.

Rossi, P. (1996). *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time*, Academic Press, Toronto.