Statistica Sinica **15**(2005), 215-234

ASYMPTOTIC THEORY FOR ARCH-SM MODELS: LAN AND RESIDUAL EMPIRICAL PROCESSES

Sangyeol Lee and Masanobu Taniguchi

Seoul National University and Waseda University

Abstract: In this paper, we have two asymptotic objectives: the LAN and the residual empirical process for a class of ARCH(∞)-SM (stochastic mean) models, which covers finite-order ARCH and GARCH models. First, we establish the LAN for the ARCH(∞)-SM model and, based on it, construct an asymptotically optimal test when the parameter vector contains a nuisance parameter. Also, we discuss asymptotically efficient estimators for unknown parameters when the innovation density is known and when it is unknown. For the residual empirical process, we investigate its asymptotic behavior in ARCH(q)-SM models. We show that, unlike the usual autoregressive model, the limiting distribution in this case depends upon the estimator of the regression parameter as well as those of the ARCH parameters.

Key words and phrases: ARCH model, ARCH(∞)-SM model, asymptotically efficient estimator, asymptotically optimal test, GARCH model, Gaussian process, LAN, residual empirical process, weak convergence.

1. Introduction

Traditional time series models such as ARMA models assume a constant one-period forecast variance. However, in actual practice, this assumption is often violated, especially in economic time series. In order to circumvent this difficulty, Engle (1982) and Bollerslev (1986) introduced, respectively, the autoregressive conditional heteroscedastic (ARCH) model and the generalized ARCH (GARCH) model. Since then, a great number of theoretical and empirical studies have been conducted for them (cf. Engle (1995), Linton (1993) and Chandra and Taniguchi (2001)). Recently, Giraitis, Kokoszaka and Leipus (2000) discussed a class of ARCH(∞) models, which includes the ARCH and GARCH models as special cases, and established sufficient conditions for the existence of a stationary solution and its explicit representation. In this paper, we deal with the ARCH(∞) model with stochastic mean (ARCH(∞)-SM model) rather than ARCH(∞) models themselves. In particular, we focus on local asymptotic normality (LAN) and the residual empirical process (REP) of the ARCH-SM model. Both LAN and REP are well known to be meaningful in the estimation and the testing of hypotheses for statistical models. For a review of the related history and background, we refer to LeCam (1986), LeCam and Yang (1990) and Shorack and Wellner (1986).

In this paper, based on the LeCam (1986) and Swensen (1985) approach, we establish the LAN theorem for a class of ARCH(∞)-SM models (there is another approach to prove the LAN for time series, see Drost, Klassen and Werker (1997)). Since the central sequence is not measurable with respect to given data, we provide a version of the LAN theorem described by a data-measurable central sequence. Based on the result, we consider the problem of testing for ARCH(∞)-SM models, and construct a locally asymptotically optimal test in terms of the data-measurable central sequence. Also, we discuss asymptotically efficient estimators for the unknown parameter of the ARCH(∞)-SM model when the innovation density $g(\cdot)$ is known, and when it is unknown. The details are presented in Sections 2 and 3.

On the REP side, we investigate the asymptotic behavior of the REP from an ARCH(q)-SM model. In fact, the asymptotic properties for the REP have been derived by many authors in autoregressive models. For instance, Boldin (1982), Kreiss (1991), Koul (1992), Ling (1998) and Lee and Wei (1999) derived its limiting distribution in stationary and unstable autoregressive models. However, those are directed toward handling the REP from the autoregressive model with i.i.d. errors with constant variance.

The REP from the ARCH model is intrinsically of great interest since, in general, the scale parameter makes the analysis a lot more complicated (cf. Lee and Wei (1999, Section 3.1)) and furthermore, the conditional variance of ARCH models varies with time. Recently, Boldin (2000) considered the REP for ARCH(1) models. However, his result does not cover the ARCH-SM model and cannot be extended in a straightforward manner. Also, Koul (2002) obtained a general result for a class of ARCH models, but his model does not include the ARCH-SM model either. Our analysis shows that the REP is asymptotically the same as the sum of the true empirical process and a stochastic process induced from the parameter estimators. Moreover, it is revealed that, unlike the usual autoregressive model, the estimator of the regression parameter severely affects the asymptotic behavior of the REP. The details are addressed in Section 4. Although the result itself is worthwhile, in actual practice it is more important to figure out the covariance structure of the limiting Gaussian process of the REP since goodness of fit tests, such as the chi-square type test, can be easily constructed provided the covariance structure is known. Therefore, at the end of Section 4, we discuss the limiting distribution of the REP from an ARCH(1)-SM model when the Gaussian MLE of the ARCH parameters are substituted into the REP. The proofs of the theorems presented in Sections 2-4 are in Section 5.

2. ARCH(∞) Model with Stochastic Regression

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space, and $\{\mathcal{F}_t; t \in \mathbb{Z}\}$ is a sequence of sub-σ-fields of $\mathcal{F}$ satisfying $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, $t \in \mathbb{Z}$. We consider the ARCH(∞)-SM model $\{Y_t; t \in \mathbb{Z}\}$ defined by

$$\begin{cases} Y_t - \beta'\mathbf{z}_t = \sigma_t u_t, \quad t \in \mathbb{Z}, \\ \qquad \sigma_t^2 = a + \sum_{j=1}^{\infty} b_j (Y_{t-j} - \beta'\mathbf{z}_{t-j})^2, \end{cases} \tag{2.1}$$

where $a > 0$, $b_j \geq 0$, $j = 1, 2, \ldots, \{u_t; t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with density $g(\cdot)$, and u_t is $\mathcal{F}_t$-measurable and independent of $\mathcal{F}_{t-1}$. Here $\beta = (\beta_1, \ldots, \beta_p)'$ is an unknown vector and the $\mathbf{z}_t = (z_{t1}, \ldots, z_{tp})'$ are observable $p \times 1$ random vectors which are $\mathcal{F}_{t-1}$-measurable. If $\beta = 0$, this ARCH(∞)-SM model becomes the ARCH(∞) model proposed by Giraitis et al. (2000). Note that the class of ARCH(∞)-SM models is larger than the class of GARCH(r, s)-SM models defined by

$$\begin{cases} Y_t - \beta'\mathbf{z}_t = \sigma_t u_t, \quad t \in \mathbb{Z}, \\ \qquad \sigma_t^2 = a + \sum_{j=1}^{r} a_j \sigma_{t-j}^2 + \sum_{j=1}^{s} b_j (Y_{t-j} - \beta'\mathbf{z}_{t-j})^2, \end{cases} \tag{2.2}$$

where the associated polynomials of the second equation of (2.2) satisfy the invertible condition. If we take $\mathbf{z}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$, then (2.2) becomes the AR(p)-GARCH (r, s) model, which indicates that the class of ARCH(∞)-SM models is sufficiently broad. Recently, there was an attempt to study an adaptive estimation for ARMA-GARCH models (cf. Ling and McAleer (2003)).

In order to develop the asymptotic theory we impose the following assumptions.

Assumption 1.
(i) $Eu_t = 0$, $Var(u_t) = 1$ and $Eu_t^4 < \infty$.
(ii) a and b_j's are functions of an unknown parameter $\eta = (\eta_1, \ldots, \eta_q)'$, i.e., $a = a(\eta)$ and $b_j = b_j(\eta)$ for $j \geq 1$ and $\eta \in \mathcal{H}$, where $\mathcal{H}$ is an open subset of $\mathbb{R}^q$. The functions $a(\eta)$ and $b_j(\eta)$ are twice continuously differentiable with respect to η.
(iii) There exist $\tilde{a} > 0$ and $\tilde{b_j} \geq 0$ satisfying $\sum_{j=1}^{\infty} \tilde{b_j} < 1$, such that $a(\eta) \geq \tilde{a}$ and $b_j(\eta) \leq \tilde{b_j}$ for all $j \geq 1$ and $\eta \in \mathcal{H}$, which entails

$$\sum_{j=1}^{\infty} b_j(\eta) < 1 \quad \text{for all } \eta \in \mathcal{H}. \tag{2.3}$$

(iv) There exist $\tilde{a}^{(1)}$, $\tilde{a}^{(2)} > 0$ and $\tilde{b}_j^{(1)}$, $\tilde{b}_j^{(2)} \geq 0$ satisfying $\sum_{j=1}^{\infty} \tilde{b}_j^{(i)} < \infty$, $i = 1, 2$, such that $\|\partial a(\eta)/\partial \eta\| \leq \tilde{a}^{(1)}$, $\|(\partial^2/\partial\eta\partial\eta')a(\eta)\| \leq \tilde{a}^{(2)}$ for all $\eta \in \mathcal{H}$, and

$||(\partial/\partial\eta)b_j(\eta)\| \leq \tilde{b_j}^{(1)}$ and $\|(\partial^2/(\partial\eta\partial\eta'))b_j(\eta)\| \leq \tilde{b}_j(\eta)^{(2)}$ for all $j \geq 1$ and $\eta \in \mathcal{H}$, where $\|a\|$ denotes the Euclidean norm of a vector or matrix defined by $\sqrt{tr(a'a)}$.

The condition (2.3) guarantees the existence of a strictly stationary solution for $\{Y_t - \beta'\mathbf{z}_t\}$ in (2.1) (cf. Giraitis et al. (2000)). For a class of GARCH models, Bougerol and Picard (1992) gave a necessary and sufficient condition for the strict stationarity which, however, is different from that for usual ARCH models.

Assumption 2. $\{E(u_t^4)\}^{1/2}\sum_{j=1}^{\infty} b_j < 1$ and $E||\mathbf{z}_t||^4 < \infty$.

The condition $\{E(u_t^4)\}^{1/2}\sum_{j=1}^{\infty} b_j < 1$ implies $E(Y_t^4) < \infty$ (see Giratis et al. (2000)). Note that in a special case of GARCH models in (2.2), a necessary and sufficient condition for the existence of the fourth moment of $Y_t - \beta'\mathbf{z}_t$ has been established by Ling and Li (1997), Chen and An (1998) and Ling and McAleer (2002).

Assumption 3. The innovation density $g(\cdot)$ is symmetric, twice continuously differentiable, and satisfies (i) $0 < I(g) := \int\{g'(u)/g(u)\}^2 g(u)du < \infty$, $\int\{g'(u)/g(u)\}^4 g(u)du < \infty$, and (ii) $\lim_{|u|\to\infty} ug(u) = 0$, $\lim_{|u|\to\infty} u^2 g'(u) = 0$.

From Assumption 3, it follows that $E\{g'(u_t)/g(u_t)\} = 0$, $E\{u_t(g'(u_t)/g(u_t))\} = -1$, $E\{(g'(u_t)/g(u_t))'\} = -I(g)$, $E\{u_t(g'(u_t)/g(u_t))'\} = 0$ and $E\{u_t^2(g'(u_t)/g(u_t))'\} = -J(g) + 2$, where $J(g) := E\{u_t^2(g'(u_t)/g(u_t))^2\}$. Symmetry of $g(\cdot)$ makes the asymptotics related to $g(\cdot)$ simple. Also it is known that the symmetry is essential for adaptive estimation in ARCH (see Linton (1993)). The variables $\{\mathbf{z}_t\}$ can be both non-random and random, and satisfy the following condition.

Assumption 4. The matrix $n^{-1}\sum_{t=1}^{n}\mathbf{z}_t\mathbf{z}_t'/\sigma_t^2$ converges to a finite limit $M(0)$ in L_2-sense, where $M(0)$ is positive definite.

We write $\theta = (\beta', \eta')' \in \Theta$ and $\dim\Theta = r = p + q$, where Θ is an open subset of $\mathbb{R}^r$. Then the σ_t's are functions of θ and Y_{t-j}, $j \geq 1$. In order to stress the dependence of θ, we sometimes write $\sigma_t = \sigma_t(\theta)$. Let $P_\theta^{(n)}$ be the distribution of $(u_s, s \leq 0, Y_1, \ldots, Y_n)$. For two hypothetical values θ, $\theta' \in \Theta$, the log-likelihood ratio is written as

$$\Lambda_n(\theta, \theta') := \log\frac{dP_{\theta'}^{(n)}}{dP_\theta^{(n)}} = 2\sum_{t=1}^{n}\log\Phi_t^{(n)}(\theta, \theta'), \tag{2.4}$$

where $\Phi_t^{(n)}(\theta, \theta') = [\{g\{\phi_t(\theta')\}\sigma_t(\theta)\}/\{g\{\phi_t(\theta)\}\sigma_t(\theta')\}]^{1/2}$ with $\phi_t(\theta) = (Y_t - \beta'\mathbf{z}_t)/\sigma_t(\theta)$. From Giraitis et al. (2000), $(Y_t - \beta'\mathbf{z}_t)$ can be expressed as a nonlinear function of $\{u_t, u_{t-1}, \ldots\}$, say, $f(u_t, u_{t-1}, \ldots)$. In (2.4) we understand that all the $(Y_t - \beta'\mathbf{z}_t)'$s are replaced by $f(u_t, u_{t-1}, \ldots)$ if $t \leq 0$. We denote by $H(g;\theta)$

the hypothesis under which the underlying parameter is $\theta \in \Theta$ and the density of u_t is $g = g(\cdot)$. We define $\theta_n = \theta + (\sqrt{n})^{-1}\xi$, $\xi = (\kappa', h')' \in \mathcal{S} \subset \mathbb{R}^r$, where $\kappa = (\kappa_1, \ldots, \kappa_p)'$, $h = (h_1, \ldots, h_q)'$, and $\mathcal{S}$ is an open subset of $\mathbb{R}^r$. We write $\theta_n = (\beta_n', \eta_n')'$, and take n so that $\theta_n \in \Theta$. Henceforth we denote by $\mathbb{R}^{\mathbb{Z}}$ the product space $\cdots \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \cdots$, whose component spaces correspond to the coordinate spaces of $(\ldots, u_{-1}, u_0, Y_1, Y_2, \ldots)$, and write its Borel σ-field by $\mathcal{B}^{\mathbb{Z}}$.

Now we state the LAN theorem for the ARCH(∞)-SM model (2.1).

Theorem 2.1. *Suppose that Assumptions* $1-4$ *hold. Then the sequence of experiments* $E_n = \{\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \{P_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^r\}\}$, $n \in N$, *is locally asymptotically normal and equicontinuous on compact subset* C *of* $\mathcal{S}$. *That is,*

(i) *For all* $\theta \in \Theta$, *the log-likelihood ratio* $\Lambda_n(\theta, \theta_n)$ *admits the following asymptotic representation under* $H(g; \theta)$*:*

$$\Lambda_n(\theta, \theta_n) = (\kappa', h')\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\Delta_{1,t}, \Delta_{2,t})' - \frac{1}{2}\, \xi' F \xi + o_P(1), \tag{2.5}$$

where $\Delta_{1,t} = -(\mathbf{z}_t/\sigma_t)(g'(\phi_t)/g(\phi_t)) - (2\sigma_t^2)^{-1}(\partial\sigma_t^2/\partial\beta)\{1 + \phi_t(g'(\phi_t)/g(\phi_t))\}$, $\Delta_{2,t} = -(2\sigma_t^2)^{-1}\{1 + \phi_t(g'(\phi_t)/g(\phi_t))\}(\partial/\partial\eta)\sigma_t^2$, *and*

$$F = \begin{pmatrix} F_{11}, F_{12} \\ F_{21}, F_{22} \end{pmatrix}$$

with $F_{11} = I(g) \cdot M(0) + \{J(g) - 1\}E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\beta)(\partial\sigma_t^2/\partial\beta')]$, $F_{12} = \{J(g) - 1\}\ E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\beta)(\partial\sigma_t^2/\partial\eta')]$, $F_{22} = \{J(g) - 1\}E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\eta)(\partial\sigma_t^2/\partial\eta')]$.

(ii) *Under* $H(g; \theta)$, $\Delta_n \xrightarrow{d} N(0, F)$, *where* $\Delta_n = n^{-1/2}\sum_{t=1}^{n}(\Delta_{1,t}, \Delta_{2,t})'$.

(iii) *For all* $n \in N$ *and* $\xi \in \mathcal{S}$, *the mapping* $\xi \to P_{\theta_n}^n$ *is continuous with respect to the variational distance* $\|P - Q\| = \sup\{|P(A) - Q(A)|; A \in \mathcal{B}^{\mathbb{Z}}\}$.

The term Δ_n, called the central sequence, is measurable with respect to u_s, $s \le 0$, Y_j, $\mathbf{z}_j$, $1 \le j \le n$, but it is not so with respect to the observable sequence Y_j, $\mathbf{z}_j$, $1 \le j \le n$. Therefore, we will make a $(Y_1, \ldots, Y_n, \mathbf{z}_1, \ldots, \mathbf{z}_n)$-measurable version $\tilde{\Delta}_n$ of Δ_n. For this task, we introduce the truncated versions of σ_t^2 and ϕ_t: $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\beta, \eta) := a + \sum_{j=1}^{t-1} b_j(Y_{t-j} - \beta'\mathbf{z}_{t-j})^2$, and $\tilde{\phi}_t := (Y_t - \beta'\mathbf{z}_t)/\tilde{\sigma}_t$, and assume the following.

Assumption 5.

(i) For some $r \in [0, 1)$, the coefficients $b_j = b_j(\eta)$ satisfy $b_j(\eta) = O(r^j)$ for all $j \ge 1$ and $\eta \in \mathcal{H}$.

(ii) For some $v_1, \ldots, v_q \in [0, 1)$, the derivatives $(\partial b_j(\eta)/\partial\eta_k)$, $k = 1, \ldots, q$, satisfy $|(\partial b_j(\eta)/\partial\eta_k)| = O(v_k^j)$ for all $j \ge 1$ and $\eta \in \mathcal{H}$.

The following theorem shows that the starting values $\{u_s; s \le 0\}$ have no influence on the LAN form in (2.6) below.

Theorem 2.2. *Under Assumptions* $1-5$, *the log-likelihood ratio* $\Lambda_n(\theta, \theta_n)$ *admits, under* $H(g;\theta)$, *as* $n \to \infty$, *the asymptotic representation*

$$\Lambda_n(\theta, \theta_n) \ = \ (\kappa', h')\tilde{\Delta}_n - \frac{1}{2}\ \xi'\ F\ \xi + o_P(1), \tag{2.6}$$

where $\tilde{\Delta}_n = n^{-1/2}\sum_{t=1}^{n}(\tilde{\Delta}_{1,t}, \tilde{\Delta}_{2,t}')$, $\tilde{\Delta}_{1,t} = -(\mathbf{z}_t\tilde{\sigma}_t)(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t)) - (2\tilde{\sigma}_t^2)^{-1}(\partial\tilde{\sigma}_t^2/\partial\beta)\{1+\tilde{\phi}_t(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t))\}$, $\tilde{\Delta}_{2,t} = -(2\tilde{\sigma}_t^2)^{-1}\{1+\tilde{\phi}_t(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t))\}(\partial/\partial\eta)\tilde{\sigma}_t^2$. *Here,* $\tilde{\Delta}_n \xrightarrow{d} N(0, F)$ *under* $H(g;\theta)$.

3. Applications to Estimation and Testing Problems

First, we discuss the estimation theory for θ. In what follows the distribution law of a random vector Y_n under $P_\theta^{(n)}$ is denoted by $\mathcal{L}(Y_n|P_\theta^{(n)})$, and the weak convergence to Z is denoted by $\mathcal{L}\{Y_n|P_\theta^{(n)}\} \xrightarrow{d} Z$. We define the class $\mathcal{A}$ of sequences of estimators of θ, $\{S_n\}$, as

$$\mathcal{A} \ = \ [\{S_n\}; \mathcal{L}\{\sqrt{n}(S_n - \theta_n)|P_{\theta_n}^{(n)}\} \xrightarrow{d} Z_\theta,\ \text{a probability distribution}\ ],$$

where Z_θ, in general, depends on $\{S_n\}$. Let L be the class of all loss functions $l: \mathbb{R}^r \to [0,\infty)$ of the form $l(x) = \tau(|x|)$ which satisfies $\tau(0) = 0$ and $\tau(a) \le \tau(b)$ if $a \le b$. Typical examples are $l(x) \ = \ I(|x| > a)$ and $l(x) \ = \ |x|^p$, $p \ge 1$, where $I(\cdot)$ denotes the indicator.

Assume that the LAN property (2.5) holds. Then, a sequence $\{\hat{\theta}_n\}$ of estimators of θ is said to be a sequence of asymptotically centering estimators if $\sqrt{n}(\hat{\theta}_n - \theta) - F^{-1}\Delta_n \ = \ o_P(1)$ in $P_\theta^{(n)}$. The following proposition can be verified by following the arguments in Strasser (1985, Section 83), Jeganathan (1995), and Taniguchi and Kakizawa (2000, p.69).

Proposition 3.1. *Assume the LAN property* (2.5) *for the ARCH*(∞)*-SM model. Let* $\{S_n\}$ *be a sequence of estimators of* θ *that belongs to* $\mathcal{A}$. *Let* Δ *be a random vector distributed as* $N(0_r, F)$. *Then the following statements hold.*

(i) *For any* $l \in L$ *with* $El(\Delta) < \infty$, $\liminf_{n\to\infty} E[l\{\sqrt{n}(S_n - \theta)\}|P_\theta^{(n)}] \ge E\{l(F^{-1}\Delta)\}$.
(ii) *If* $\limsup_{n\to\infty} E[l\{\sqrt{n}(S_n - \theta)\}|P_\theta^{(n)}] \le E\{l(F^{-1}\Delta)\}$ *for a nonconstant* $l \in L$ *with* $El(\Delta) < \infty$, *then* S_n *is a sequence of asymptotically centering estimators.*

In view of the above result, a sequence of estimators $\{\hat{\theta}_n\} \in \mathcal{A}$ of θ is asymptotically efficient if it is a sequence of asymptotically centering estimators. Then we can construct an asymptotically efficient estimator. For any sequence of estimators $\tilde{\theta}_n$, the discretized estimator $\bar{\theta}_n$ of $\tilde{\theta}_n$ is defined by the nearest vertex of $\{\theta; \theta = n^{-1/2}(i_1, \ldots, i_r)', i_j \text{ integers}\}$. First we assume that the innovation density $g(\cdot)$ is known. We denote the Fisher information matrix F and the central sequence $\tilde{\Delta}_n$ by $F(\theta, g)$ and $\tilde{\Delta}_n(\theta, g)$, respectively. Let $\hat{\theta}_n = \bar{\theta}_n + n^{-1/2}F(\bar{\theta}_n, g)^{-1}\tilde{\Delta}_n(\bar{\theta}_n, g)$, where $\bar{\theta}_n$ is a discrete and $\sqrt{n}$-consistent estimator of θ (for technical justification for our use of discrete estimators, see Kreiss (1987, p.120). In our model, we can use the least squares estimator (LSE) $\hat{\beta}_{LS}$ for β. Then η is estimated by the conditional LSE (see Tjøstheim (1986)),

$$\begin{aligned}\hat{\eta}(\hat{\beta}_{LS}) &= \text{argmin}_\eta \sum_{t=2}^{n} [(Y_t - \hat{\beta}'_{LS}\mathbf{z}_t)^2 - a(\eta) - \sum_{j=1}^{t-1} b_j(\eta)(Y_{t-j} - \hat{\beta}'_{LS}\mathbf{z}_{t-j})^2]^2 \\ &= \text{argmin}_\eta\, S(\hat{\beta}_{LS}, \eta), \quad \text{say.}\end{aligned}$$

Strictly speaking, Tjϕstheim's conditional LSE is $\hat{\eta}(\beta)$. However, in view of Assumptions 1, 4 and 5, we can show that $\text{Var}\,(\hat{\beta}_{LS}) = O(n^{-1})$, $n^{-1/2}[(\partial/\partial\eta) S(\hat{\beta}_{LS}, \eta) - (\partial/\partial\eta)S(\beta, \eta)] = o_p(1)$ and $n^{-1}[(\partial^2/\partial\eta\partial\eta')S(\hat{\beta}_{LS}, \eta) - (\partial^2/\partial\eta\partial\eta') S(\beta, \eta)] = o_p(1)$. Thus $\hat{\eta}(\hat{\beta}_{LS})$ has the same asymptotics as $\hat{\eta}(\beta)$. Hence the estimator $(\hat{\beta}'_{LS}, \hat{\eta}'(\hat{\beta}_{LS}))'$ becomes a candidate of $\tilde{\theta}_n$. Here, similarly to LeCam (1986) and Linton (1993), it can be shown that $\hat{\theta}_n$ is asymptotically efficient. If $g(\cdot)$ is unknown, substituting an appropriate nonparametric density estimator $\hat{g}_n(\cdot)$ for $g(\cdot)$, we can suggest $\hat{\hat{\theta}}_n = \bar{\theta}_n + n^{-1/2}F(\bar{\theta}_n, \hat{g}_n)^{-1}\tilde{\Delta}_n(\bar{\theta}_n, \hat{g}_n)$, but we cannot guarantee its asymptotic efficiency. For the ARCH(p) case, Linton (1993) reparameterized the model, and gave an asymptotic efficient estimator of the coefficients of ARCH model. Our case is not straightforward because infinitely many b_j's depend on unknown parameter η.

Next, we discuss the problem of testing. Let $\mathcal{M}(B)$ be the linear space spanned by the columns of a matrix B. The problem consists of testing the null hypothesis H, under which $\sqrt{n}(\theta - \theta_0) \in \mathcal{M}(B)$ for some given $r \times (r-l)$ matrix B of full rank and given vector $\theta_0 \in \mathbb{R}^r$. Then, similarly to Strasser (1985, Section 8.2) and Taniguchi and Kakizawa (2000, p.78), we can see that the test based on the quadratic form $\|[\mathbf{I}_r - F^{1/2}B(B'FB)^{-1}B'F^{1/2}]F^{-1/2}\tilde{\Delta}_n\|^2_{\theta=\bar{\theta}_n}$, which has the χ^2_l null distribution asymptotically, is locally asymptotically optimal. Here $\mathbf{I}_r$ denotes the $r \times r$ identity matrix.

4. Residual Empirical Process

In this section we study the residual empirical process from ARCH(q)-SM models. Consider the model

$$\begin{cases} Y_t = \beta'\mathbf{z}_t + \sigma_t u_t, \\ \sigma_t^2 = a + \sum_{j=1}^{q} b_j(Y_{t-j} - \beta'\mathbf{z}_{t-j})^2, \end{cases} \tag{4.1}$$

where β is a p-dimensional unknown vector, $a > 0$, $b_j \geq 0$, and $0 \leq \sum_{j=1}^q b_j < 1$, $\{u_t\}$ is a sequence of i.i.d. r.v.'s, such that u_t is independent of $\mathbf{z}_s$, $s \leq t$, and $\{\mathbf{z}_t\}$ is a strictly stationary and ergodic process.

Assumption 6.

(i) The density $g(\cdot)$ of $\{u_t\}$ is positive and differentiable; it is 'increasing and convex' on $(-\infty, -M]$ and 'decreasing and convex' on $[M, \infty)$ for some positive real number M; $|x|g(x) \to 0$ as $|x| \to \infty$ and $\sup_x x^2|g'(x)| < \infty$.

(ii) $Eu_t = 0$, $Eu_t^2 = 1$, $Eu_t^4 < \infty$ and $E||\mathbf{z}_t||^4 < \infty$.

Note that the normal density satisfies Assumption 6 (i). Let $\hat{a}$, $\hat{b}_j$ and $\hat{\beta}$ be the estimators of a, b_j and β with

$$n^{1/2}(\hat{a} - a) = O_P(1),\ n^{1/2}(\hat{b}_j - b_j) = O_P(1),\ \text{and}\ n^{1/2}(\hat{\beta} - \beta) = O_P(1). \quad (4.2)$$

Then the residuals are given by $\hat{u}_t = (Y_t - \hat{\beta}'\mathbf{z}_t)/\hat{\sigma}_t$, where $\hat{\sigma}_t^2 = \hat{a} + \sum_{j=1}^q \hat{b}_j(Y_{t-j} - \hat{\beta}'\mathbf{z}_{t-j})^2$.

Our goal here is to investigate the asymptotic behavior of $\hat{\mathcal{E}}_n$, where

$$\hat{\mathcal{E}}_n(x) = n^{-1/2} \sum_{t=1}^n \{I(\hat{u}_t \leq x) - G(x)\}, \quad -\infty < x < \infty, \quad (4.3)$$

where $G(x) = \int_{-\infty}^x g(v)dv$. We split $\hat{\mathcal{E}}_n(x)$ into $\mathcal{E}_n(x) + I_n + II_n$, where

$$\begin{aligned}
\mathcal{E}_n(x) &= n^{-1/2} \sum_{t=1}^n \{I(u_t \leq x) - G(x)\} \\
I_n &= n^{-1/2} \sum_{t=1}^n \{G(\hat{\sigma}_t x/\sigma_t + (\hat{\beta} - \beta)'\mathbf{z}_t/\sigma_t) - G(x)\} \\
II_n &= n^{-1/2} \sum_{t=1}^n \{I(\hat{u}_t \leq x) - G(\hat{\sigma}_t x/\sigma_t + (\hat{\beta} - \beta)'\mathbf{z}_t/\sigma_t) + G(x) - I(u_t \leq x)\}.
\end{aligned}$$

Since

$$\sup_x |II_n| = o_P(1), \quad (4.4)$$

for which the proof is provided in Section 5, only $\mathcal{E}_n$ and I_n determine the limiting distribution of $\hat{\mathcal{E}}_n$. In particular, I_n involves the estimators $\hat{\beta}$, $\hat{a}$ and $\hat{b}_j$, so that the asymptotic behavior of the residual empirical process is affected by a choice of estimators. Below we analyze I_n and see how the estimators affect the limiting distribution of $\hat{\mathcal{E}}_n$.

Using Taylor's series expansion, we can write

$$I_n = n^{-1/2}\sum_{t=1}^{n}\{(\hat{\sigma}_t/\sigma_t - 1)x + (\hat{\beta}-\beta)'\mathbf{z}_t/\sigma_t\}g(x) \tag{4.5}$$

$$+n^{-1/2}\sum_{t=1}^{n}\{(\hat{\sigma}_t/\sigma_t - 1)x + (\hat{\beta}-\beta)'\mathbf{z}_t/\sigma_t\}^2 g'(\zeta_{tx})/2,$$

$$= I_{1n} + I_{2n}, \quad \text{say},$$

where ζ_{tx} is a real number between x and $x\hat{\sigma}_t/\sigma_t + (\hat{\beta}-\beta)z_t/\sigma_t$.

First, we deal with I_{1n}. Note that

$$\hat{\sigma_t^2} = \hat{a} + \sum_{j=1}^{q}\hat{b}_j\xi_{t-j} - 2\sum_{j=1}^{q} b_j(\hat{\beta}-\beta)'\nu_{t-j} + \theta_{nt}, \tag{4.6}$$

where $\xi_t = \sigma_t^2 u_t^2$, $\nu_t = \sigma_t u_t \mathbf{z}_t$, and $\theta_{nt} = -2\sum_{j=1}^{q}(\hat{b}_j - b_j)(\hat{\beta}-\beta)'\nu_{t-j} + \sum_{j=1}^{q}\hat{b}_j[(\hat{\beta}-\beta)'\mathbf{z}_{t-1}]^2$. Utilizing Assumption 6 (ii) and (4.2), we can readily check that

$$\max_{1\le t\le n}|\theta_{nt}| = o_P(n^{-1/2}), \tag{4.7}$$

and $n^{-1/2}\sum_{t=1}^{n}(\hat{\sigma}_t^2 - \sigma_t^2)^2 = o_P(1)$. Then, by simple algebra using this and (4.7), we have that

$$I_{1n} = n^{1/2}(\hat{\beta}-\beta)'E(\mathbf{z}_t/\sigma_t)g(x) \tag{4.8}$$

$$+\{n^{1/2}(\hat{a}-a)E(1/2\sigma_t^2) + n^{1/2}\sum_{j=1}^{q}(\hat{b}_j - b_j)E(u_{t-j}^2\sigma_{t-j}^2/2\sigma_t^2)\}xg(x) + \eta_n(x)$$

with $\sup_x|\eta_n(x)| = o_P(1)$.

Now we deal with I_{2n}. Let $K > 0$ be any real number. In view of (4.2), (4.6) and (4.7), we can write

$$\sup_x|x^2 g'(\zeta_{tx})| \le K^2\sup_x|g'(x)| + \sup_{|x|\ge K}(\frac{x}{\zeta_{tx}})^2\sup_x|x^2 g'(x)|$$

$$\le O(1) + \sup_{|x|\ge K, |\theta|\le\delta_n, |\tilde{\theta}|\le\tilde{\delta_n}}|\frac{x}{(1+\theta)x+\tilde{\theta}}|^2 = O_P(1), \tag{4.9}$$

where $\{\delta_n\}$ and $\{\tilde{\delta_n}\}$ are some sequences of r.v.'s with $\delta_n = o_P(1)$ and $\tilde{\delta_n} = o_P(1)$. Also,

$$n^{-1/2}\{\sum_{t=1}^{n}(\hat{\sigma}_t/\sigma_{t-1} - 1)^2 + ((\hat{\beta}-\beta)'\sum_{t=1}^{n}(\mathbf{z}_t/\sigma_t))^2\} = o_P(1).$$

Hence, $\sup_x |I_{2n}| = o_P(1)$ which, together with (4.8), yields

$$I_n = n^{1/2}(\hat{\beta} - \beta)' E(\mathbf{z}_t/\sigma_t) g(x) + \{n^{1/2}(\hat{a} - a)E(1/2\sigma_t^2) + n^{1/2} \sum_{j=1}^{q} (\hat{b}_j - b_j) E(u_{t-j}^2 \sigma_{t-j}^2 / 2\sigma_t^2)\} x g(x) + \tilde{\eta}_n(x),$$

with $\sup_x |\tilde{\eta}_n(x)| = o_P(1)$. This leads to the following theorem.

Theorem 4.1. *Under Assumption* 6 *and the conditions in* (4.5)*, we have*

$$\hat{\mathcal{E}}_n(x) = \mathcal{E}_n(x) + \mathcal{N}_n(x) + \eta_n^*(x), \tag{4.10}$$

where $\mathcal{N}_n(x) = n^{1/2}(\hat{\beta} - \beta)' E(\mathbf{z}_t/\sigma_t) g(x) + \{n^{1/2}(\hat{a} - a)E(1/2\sigma_t^2) + n^{1/2} \sum_{j=1}^{q} (\hat{b}_j - b_j) E(u_{t-j}^2 \sigma_{t-j}^2 / 2\sigma_t^2)\} x g(x)$, *and* $\sup_x |\eta_n^*(x)| = o_P(1)$.

Notice that $\mathcal{N}_n$ is not independent of the true empirical process $\mathcal{E}_n$. This phenomenon is not a surprise and even occurs in the empirical process with parameter estimators in i.i.d. samples (cf. Durbin (1973)). Although we do not pursue this matter in detail here, one can guess that the REP in the ARCH(∞)-SM process would have the same representation with $q = \infty$ as in (4.10) in view of the result of Theorem 4.1 and the fact that $\sum_{j=1}^{\infty} b_j < 1$.

The following is a direct result of Theorem 4.1.

Theorem 4.2. *Under Assumption* 6 *and the conditions in* (4.5)*, we have*

$$\hat{\mathcal{E}}_n(G^{-1}(s)) = \mathcal{E}_n(G^{-1}(s)) + \mathcal{N}_n(G^{-1}(s)) + \eta_n^{**}(s), \quad 0 \le s \le 1, \tag{4.11}$$

where $\sup_s |\eta_n^{**}(s)| = o_P(1)$.

From the above theorem it transpires that the REP depends upon the estimators of parameters and therefore its limiting distribution is not a standard Brownian bridge. A fundamental difference between the REP of the ARCH-SM model and the usual AR model is that the limiting distribution of the former depends upon the estimator of β as well as those of ARCH parameters, while it is not true for the latter (cf. Boldin (1982)).

As we mentioned in the Introduction, Boldin already showed that the REP of the ARCH(1) model depends upon the estimators of the ARCH parameters. However, Theorem 4.2 indicates that special attention should be paid in applying Boldin's result to the ARCH-SM model. For example, the Kolmogorov-Smirnov (KS) test in Boldin (2000), aimed at performing a Gaussian test, is no longer applicable in the presence of $\hat{\beta}$. Indeed, $\hat{\beta}$ makes the situation somewhat complicated when applying Theorem 4.2 to a goodness of fit test. Considering this complexity, one might be able to use the chi-square type test statistic as proposed in Lee (1996). For instance, if $\hat{\mathcal{E}}_n(G^{-1}(\cdot))$ converges weakly to a Gaussian process

$\mathcal{Z}(\cdot)$, we reject the null hypothesis $H_0 : \{u_t\} \sim G$ when $T_n := \mathbf{l}_k' \hat{\Sigma}_k \mathbf{l}_k$ is large, where $\mathbf{l}_k = (\hat{\mathcal{E}_n}(G^{-1}(s_1)), \ldots, \hat{\mathcal{E}_n}(G^{-1}(s_k))'$ for some $s_j \in (0,1)$, $j = 1, \ldots, k$, and $\hat{\Sigma}_k$ is a consistent estimator of Σ whose (j,l)th entry is $Cov(\mathcal{Z}(s_j), \mathcal{Z}(s_l))$. Obviously, T_n asymptotically follows a chi-square distribution with k degrees of freedom.

As mentioned above, in order to apply Theorem 4.2 to a goodness of test, the asymptotic expansion form of the parameter estimators should be addressed. Usually, the estimators of ARCH parameters after normalization appear to be the sum of martingale differences. This guarantees that the REP converges weakly to a Gaussian process. Here we give an example of an ARCH(1)-SM model in which we employ the least squares estimator for the regression parameter and the Gaussian MLE for the ARCH parameters.

Consider the ARCH(1)-SM model, viz., the model (4.1) with $p = q = 1$:

$$\begin{cases} Y_t = \beta z_t + \sigma_t u_t, \\ \sigma_t^2 = a + b(Y_{t-1} - \beta z_{t-1})^2. \end{cases}$$

Assuming $y_0 = z_0 = 0$, the Gaussian MLE is obtained to maximize the Gaussian log-likelihood function with the estimator $\hat{\beta}$ substituted:

$$\begin{aligned} l(a,b) = c - (1/2) \sum_{t=2}^{n} \log(a + b\hat{\epsilon}_{t-1}^2) \\ -(1/2) \sum_{t=2}^{n} \hat{\epsilon}_t^2/(a + b\hat{\epsilon}_{t-1}^2) - \log a/2 - \hat{\epsilon}_1^2/2a, \end{aligned}$$

where c is a constant and $\hat{\epsilon}_t = Y_t - \hat{\beta} z_t$. Then it can be seen that the following expressions hold:

$$n^{1/2}(\hat{a} - a) = n^{-1/2} \sum_{t=1}^{n} \{(u_t^2 - 1)(\tau_{11} + \tau_{12} u_{t-1}^2 \sigma_{t-1}^2)/\sigma_t^2\} + o_P(1), \quad (4.12)$$

$$n^{1/2}(\hat{b} - b) = n^{-1/2} \sum_{t=1}^{n} \{(u_t^2 - 1)(\tau_{12} + \tau_{22} u_{t-1}^2 \sigma_{t-1}^2)/\sigma_t^2\} + o_P(1), \quad (4.13)$$

where τ_{ij} is the (i,j)-th entry of the 2×2 matrix D^{-1} and $D = (D_{ij})_{i,j=1}^2$ is the matrix with $D_{11} = -E(1/\sigma_t^4)$, $D_{12} = D_{21} = -E(\sigma_{t-1}^2 u_{t-1}^2/\sigma_t^4)$, $D_{22} = -E(u_{t-1}^4 \sigma_{t-1}^4/\sigma_t^4)$.

Meanwhile, if $\hat{\beta}$ is the least squares estimator, we have $n^{1/2}(\hat{\beta} - \beta) = n^{-1/2}(Ez_t^2)^{-1} \sum_{t=1}^n z_t \sigma_t u_t + o_P(1)$. In view of this, (4.12), (4.13), and Theorem 4.2, utilizing the Martingale Central Limit theorem one can see that the REP converges weakly to a Gaussian process. In fact, simple algebra shows that $\hat{\mathcal{E}_n}(G^{-1}(\cdot))$ converges weakly to a Gaussian process $\mathcal{E}^*$ with mean 0 and the covariance structure such that, for all $0 \le s,\, t \le 1$,

$$\begin{aligned} Cov(\mathcal{E}^*(s), \mathcal{E}^*(t)) = s \wedge t - st & \\ & +E\lambda_{11}(Eu_1^2 I(u_1 \le G^{-1}(s)) - s)G^{-1}(t)g(G^{-1}(t)) \\ & +E\lambda_{21}Eu_1 I(u_1 \le G^{-1}(s))g(G^{-1}(t)) \\ & +E\lambda_{11}(Eu_1^2 I(u_1 \le G^{-1}(t)) - t)G^{-1}(s)g(G^{-1}(s)) \\ & +E\lambda_{21}Eu_1 I(u_1 \le G^{-1}(t))g(G^{-1}(s)) \\ & +E\lambda_{11}\lambda_{21}Eu_1^3(G^{-1}(s) + G^{-1}(t))g(G^{-1}(s))g(G^{-1}(t)) \\ & +E\lambda_{11}^2 E(u_1^2 - 1)^2 G^{-1}(s)G^{-1}(t)g(G^{-1}(s))g(G^{-1}(t)) \\ & +E\lambda_{21}^2 g(G^{-1}(s))g(G^{-1}(t)), \end{aligned}$$

where

$$\lambda_{11} = \{E1/2\sigma_1^2\}(\tau_{11} + \tau_{12}u_0^2\sigma_0^2)/\sigma_1^2 + \{Eu_0^2\sigma_0^2/2\sigma_1^2\}(\tau_{12} + \tau_{22}u_0^2\sigma_0^2)/\sigma_1^2$$

and $\lambda_{21} = \{Ez_1^2\}^{-1}E(z_1/\sigma_1)z_1\sigma_1$. Although we do not pursue details, the above can be naturally extended to a more general case as long as the asymptotic behavior of the estimators is known.

5. Proofs

In this section, we give the proofs of the theorems presented in the previous sections. In order to prove Theorem 2.1, we need the following. Recall that $g(\cdot)$ is the innovation density satisfying Assumption 3. For given $a > 0$ and p-dimensional vector b, we introduce $G(x;\rho,\delta) := g^{1/2}[\{1 + a\rho\}(x - b'\delta)]\{1/a + \rho\}^{1/2}\{a/g(x)\}^{1/2} - 1$, so

$$DG(x) := \begin{pmatrix} (\frac{1}{2}g(x)^{-1}g'(x)ax + \frac{a}{2} \\ -\frac{1}{2}g(x)^{-1}g'(x)b \end{pmatrix}$$

which denotes the derivative of $G(x;\rho,\delta)$ with respect to ρ and δ evaluated at $\rho = 0$ and $\delta = 0$. In the same way as in Garel and Hallin (1995), we can prove the following lemma.

Lemma 5.1. *Suppose that Assumption 3 holds. Let $v = (\rho, \delta')'$ and $\bar{G}(x;v) := G(x;\rho,\delta)g(x)^{1/2}$, then the following statements hold.*

(i) *For all $v \in \mathbb{R}^{p+1}$, $\int[G(x;\rho,\delta) - v'DG(x)]^2 g(x)dx = O(a^2\rho^2) + O\{(\delta' b)^2\}$.*

(ii) *For all $v \to 0$ $(0 \ne v \in \mathbb{R}^{p+1})$, $(v'v)^{-1}\int[\bar{G}(x;v) - v'DG(x)g(x)^{1/2}]^2 dx \to 0$.*

(iii) *For any $v_n \to 0$ $(v_n \in \mathbb{R}^{p+1})$ and $c > 0$,*

$$\begin{aligned} \lim_{n\to\infty} \sup_{\|u\|\le c} \int \|n^{-\frac{1}{2}}u + v_n\|^{-2}\{\bar{G}(x; n^{-\frac{1}{2}}u + v_n) \\ -(n^{-\frac{1}{2}}u + v_n)'DG(x)g(x)^{\frac{1}{2}}\}^2 dx = 0. \end{aligned}$$

Now we write $U_t^{(n)} = \Phi_t^{(n)}(\theta, \theta_n) - 1$ *and* $W_t^{(n)} = (2\sqrt{n})^{-1}\{\kappa'\Delta_{1,t} + h'\Delta_{2,t}\}$, *where* $\Phi_t^{(n)}$, *is given in* (2.4), *and* $\Delta_{i,t}$, $i = 1, 2$, *are defined in Theorem* 2.1.

Proof of Theorem 2.1. In order to prove the theorem, we check the LAN conditions (S1)-(S6) below (cf. Swensen (1985)). Let $\mathcal{F}_t$ be the σ-algebra generated by $\{u_s, s \le 0, Y_1, \ldots, Y_t\}$.

(S1) $E\{W_t^{(n)}|\mathcal{F}_{t-1}\} = 0$ *a.e.* is easily checked.

(S2) $\lim_{n\to\infty} E[\sum_{t=1}^n \{U_t^{(n)} - W_t^{(n)}\}^2] = 0$. This condition is checked by showing (a) $\lim_{n\to\infty} E[\sum_{t=1}^n \{U_t^{(n)} - W_t^{*(n)}\}^2] = 0$, and (b) $\lim_{n\to\infty} E[\sum_{t=1}^n \{W_t^{*(n)} - W_t^{(n)}\}^2] = 0$, where $W_t^{*(n)} = (\rho, \delta')DG(x)|_{x=\phi_t, \rho=(\sigma_t(\theta_n))^{-1}-(\sigma_t(\theta))^{-1}, \delta=n^{-1/2}\kappa} = v'DG(\phi_t)$, say. For every $c_1 > 0$, we have

$$
\begin{aligned}
E[\sum_{t=1}^n \{U_t^{(n)} - W_t^{*(n)}\}^2] &= \sum_{t=1}^n E[I(\sqrt{n}\|v\| \le c_1)\{U_t^{(n)} - W_t^{*(n)}\}^2] \\
&\quad + \sum_{t=1}^n E[I(\sqrt{n}\|v\| > c_1)\{U_t^{(n)} - W_t^{*(n)}\}^2] \\
&= E_1 + E_2, \quad \text{say.}
\end{aligned}
$$

From Lemma 5.1 (iii), it follows that

$$
E_1 \ = \ \sum_{t=1}^n E[I(\sqrt{n}\|v\| \le c_1) E\{(U_t^{(n)} - W_t^{*(n)})^2|\mathcal{F}_{t-1}\}] \ \le \ [\sum_{t=1}^n E\|v\|^2] o_{c_1}(1),
$$

where $\lim_{n\to\infty} o_{c_1}(1) = 0$ for any given $c_1 > 0$. Here, $E\|v\|^2 \le 2E|\rho|^2 + O(n^{-1})$. From Assumption 1 (iii), we have $\sigma_t^2(\theta) \ge \tilde{a}$ and $\sigma_t^2(\theta_n) \ge \tilde{a}$, and thus $E\|v\|^2 \le O[E|\sigma_t(\theta_n) - \sigma_t(\theta)|^2] + O(n^{-1})$. By the Mean Value Theorem, we can see that $\sigma_t(\theta_n) - \sigma_t(\theta)$ is equal to

$$
(\theta_n - \theta)' \begin{pmatrix} -\frac{1}{\sigma_t(\theta^*)} \sum_{j=1}^{t-1} b_j(\eta^*)(Y_{t-j} - \beta^{*'} z_{t-j}) z_{t-j} \\ \frac{1}{2\sigma_t(\theta^*)} \{\frac{\partial a}{\partial \eta}(\eta^*) + \sum_{j=1}^{t-1} \frac{\partial b_j(\eta^*)}{\partial \eta}(Y_{t-j} - \beta^{*'} z_{t-j})^2 + \sum_{j=t}^{\infty} \frac{\partial b_j(\eta*)}{\partial \eta} u_{t-j}^2\} \end{pmatrix}, \tag{5.1}
$$

where θ^*, η^* and β^* are on the segments $\theta \mapsto \theta_n$, $\eta \mapsto \eta_n$ and $\beta \mapsto \beta_n$, respectively. Thus, from Assumption 1 (iii) and (iv) and Assumption 2, it follows that $E\|v\|^2 = O(n^{-1})$, which implies $\lim_{n\to\infty} E_1 = 0$ for any given $c_1 > 0$. We next evaluate E_2. Using Lemma 5.1 (i), one can see that

$$
E_2 \le \sum_{t=1}^n E[I(\sqrt{n}\|v\| \ge c_1)\{O(\sigma_t^2(\theta) \cdot \rho^2) + O(\delta' \frac{\mathbf{z}_t \mathbf{z}_t'}{\sigma_t^2(\theta)} \delta)\}]. \tag{5.2}
$$

From Assumptions 1 and 2, we can see that (5.2) is dominated by

$$\frac{1}{n}\sum_{t=1}^{n} E[I(\sqrt{n}\|v\| \geq c_1)\cdot O\{\|\sqrt{n}v\|^2\}]. \tag{5.3}$$

In view of (5.1), it is not difficult to show that $\sqrt{n}v$ converges to a random vector v^0 in L_2-sense. Hence, $\sqrt{n}v$ is uniformly integrable (e.g., Ash (1972, p.297)), which implies that (5.3) converges to zero as $c_1 \to \infty$. Therefore, $E_2 \to 0$, and (a) is proven. The assertion (b) follows from the definition of $W_t^{*(n)}$ and $W_t^{(n)}$, (5.1), and Assumptions 1, 2 and 3. Hence, (S2) is established.

(S3) $\sup_n E\{\sum_{t=1}^n {W_t^{(n)}}^2\} < \infty$. Recall the definition of $W_t^{(n)}$:

$$\begin{aligned} W_t^{(n)} = &-\frac{\kappa'\mathbf{z}_t}{2\sigma_t^2(\theta)\sqrt{n}}\frac{g'(\phi_t)}{g(\phi_t)} - \frac{1}{4\sigma_t^2(\theta)}\frac{\kappa'}{\sqrt{n}}\frac{\partial\sigma^2(\theta)}{\partial\beta}\{1+\phi_t\frac{g'(\phi_t)}{g(\phi_t)}\} \\ &-\frac{1}{4\sigma_t^2(\theta)\sqrt{n}}h'\frac{\partial}{\partial\eta}\sigma_t^2(\theta)\{1+\phi_t\frac{g'(\phi_t)}{g(\phi_t)}\}. \end{aligned}$$

From the stationarity of Y_t and Assumption 4, we have that

$$\begin{aligned} E\{\sum_{t=1}^n {W_t^{(n)}}^2\} &= \frac{1}{n}\sum_{t=1}^n E\{n{W_t^{(n)}}^2\} \\ &= \kappa' M(0)I(g)\kappa + o(1) \\ &\quad +\{J(g)-1\}[\kappa' E\{\frac{1}{16\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\beta}\frac{\partial\sigma_t^2(\theta)}{\partial\beta'}\}\kappa \\ &\quad +h'E\{\frac{1}{16\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\eta}\frac{\partial\sigma_t^2(\theta)}{\partial\eta'}\}h \\ &\quad +\kappa' E\{\frac{1}{16\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\beta}\frac{\partial\sigma_t^2(\theta)}{\partial\eta'}\}h] \;=\; \frac{\tau^2}{4}+o(1), \text{ say,} \end{aligned}$$

which implies (S3).

The remaining (S4), (S5) and (S6) can be shown similarly as in Taniguchi and Kakizawa (2000, pp.44-45), and their proofs are omitted for brevity.

(S4) $\sum_{t=1}^n {W_t^{(n)}}^2 \xrightarrow{P} \tau^2/4$,

(S5) $\max_{1\leq t\leq n}|W_t^{(n)}| \xrightarrow{P} 0$,

(S6) $\sum_{t=1}^n E[{W_t^{(n)}}^2 I(|W_t^{(n)}|>\delta)|\mathcal{F}_{t-1}] \xrightarrow{P} 0$ for some $\delta > 0$.

Finally, part (iii) of the theorem follows from Scheffé's theorem (cf. Bhattacharya and Rao (1976)) and the continuity of $g(\cdot)$.

Proof of Theorem 2.2. By the Mean Value Theorem, there exists $\sigma_t^* \in (\tilde{\sigma}_t, \sigma_t)$ such that

$$\tilde{\phi}_t - \phi_t \;=\; \{\frac{\sigma_t - \tilde{\sigma}_t}{\tilde{\sigma}_t}\}\phi_t = \frac{1}{2}\frac{\sum_{j=t}^{\infty} b_j (Y_{t-j} - \beta' z_{t-j})^2}{\tilde{\sigma}_t \sigma_t^*}\phi_t.$$

From Assumption 1 (i) and (ii) and Assumption 5 (i), we have $\tilde{\phi}_t - \phi_t = r^t O_P(1)$. Similarly, it is shown that

$$\frac{g'(\tilde{\phi}_t)}{g(\tilde{\phi}_t)} \;=\; \frac{g'(\phi_t)}{g(\phi_t)} + (\tilde{\phi}_t - \phi_t)(\frac{g'}{g})'_{\phi_t = \phi_t^*} \;=\; \frac{g'(\phi_t)}{g(\phi_t)} + r^t O_P(1), \tag{5.4}$$

where $\tilde{\phi}_t \le \phi_t^* \le \phi_t$ or $\phi_t \le \phi_t^* \le \tilde{\phi}_t$. Then, from Assumption 5 (i) and (ii), it follows that

$$\frac{\partial}{\partial \eta}\tilde{\sigma}_t^2 = \frac{\partial}{\partial \eta}\sigma_t^2 + \sum_{k=1}^{q} O(v_k^t) \cdot O_P(1), \tag{5.5}$$
$$\frac{\partial}{\partial \beta}\tilde{\sigma}_t^2 = \frac{\partial}{\partial \beta}\sigma_t^2 + r^t O_P(1).$$

In view of (5.4)-(5.5), we can see that $\Delta_n - \tilde{\Delta}_n \xrightarrow{P} 0$ under $H(g;\theta)$, which yields $\tilde{\Delta}_n \xrightarrow{d} N(0, F)$ under $H(g;\theta)$.

The following two lemmas are useful for proving (4.4).

Lemma 5.2. *Let G be a strictly increasing distribution function and let $g = G'$. Suppose that $h : R \to R$ is continuous with $\lim_{|x|\to\infty} |h(x)| = 0$. Then for any sequence of positive real numbers $\{\delta_n\}$ decaying to 0, $\sup_{|G(x)-G(y)|\le\delta_n} |h(x) - h(y)| \to 0$ as $n \to \infty$.*

Proof. Suppose that the lemma does not hold. Then we can find a positive real number c and a subsequence n' of positive integers, such that (a) $|h(x_{n'}) - h(y_{n'})| \ge c > 0$ for all n', and (b) $|G(x_{n'}) - G(y_{n'})| \le \delta_{n'}$. If there is a subsequence $\{x_{n''}\}$ of $\{x_{n'}\}$ with $x_{n''} \to \pm\infty$, then $y_{n''}$ should diverge to $\pm\infty$ by (5.6). But this is in contradiction to (a). Thus $(x_{n'}, y_{n'})$ must be in a compact subset of $\mathbf{R}^2$ and has a limit point (x_0, y_0). Then, by the continuity of h and (a), we have $|h(x_0) - h(y_0)| \ge c > 0$. Furthermore, from (b) we have $G(x_0) = G(y_0)$ and so $x_0 = y_0$, which leads to a contradiction. This completes the proof.

Lemma 5.3. *Suppose that a density function g is positive and differentiable. Furthermore, assume that g is decreasing and convex on $[M, \infty)$ and increasing and convex on $(-\infty, -M]$ for some $M > 0$. Let x_i denote the real numbers such that $-\infty = x_0 < \cdots < x_n = \infty$ and $G(x_i) = i/n$, where G is the distribution function corresponding to g. Then, $\sup_{1\le i\le n} \sup_{j=i,i-1} (x_i - x_{i-1}) g(x_j) \to 0$.*

Proof. Put $S_n^+ = \{i; M \le x_{i-1}\}$, $S_n^- = \{i; x_i \le -M\}$, and $S_n^0 = \{1, \ldots, n\} - S_n^+ - S_n^-$. If $i \in S_n^+$,

$$(x_i - x_{i-1})g(x_i) \le G(x_i) - G(x_{i-1}) = 1/n. \tag{5.6}$$

Now we write $(x_i - x_{i-1})g(x_{i-1}) = I + II$, where

$$I = \{g^{'}(x_i)(x_{i-1} - x_i) + 2g(x_i)\}(x_i - x_{i-1})/2,$$
$$II = \{2g(x_{i-1}) - g^{'}(x_i)(x_{i-1} - x_i) - 2g(x_i)\}(x_i - x_{i-1})/2.$$

Note that $I - II \ge g(x_i)(x_i - x_{i-1}) \ge 0$ and $I \le G(x_i) - G(x_{i-1}) = 1/n$. Hence, we have $(x_i - x_{i-1})g(x_{i-1}) \le 2I \le 2/n$. This together with (5.6) yields $\sup_{i \in S_n^+} \sup_{j=i,i-1} (x_i - x_{i-1})\, g(x_j) \le 2/n \to 0$. Similarly, we have $\sup_{i \in S_n^-} \sup_{j=i,i-1}(x_i - x_{i-1})g(x_j) \le 2/n \to 0$. Since the following holds manifestly, $\sup_{i \in S_n^0} \sup_{j=i,i-1}(x_i - x_{i-1})g(x_j) \to 0$, the lemma is established.

Proof of (4.4). Let x_i be real numbers such that $-\infty = x_0 < x_1 < \cdots < x_n = \infty$ and $G(x_i) = i/n$. For $\mathbf{s}_i \in \mathbf{R}^p$, $i = 1, 4$, $\mathbf{s}_3 \in \mathbf{R}^q$ and s_2, $s_5 \in \mathbf{R}$, we define $\Gamma_t(x, \mathbf{s}_1^{'}, s_2, \mathbf{s}_3^{'}, \mathbf{s}_4^{'}, s_5) = n^{-1/2}\mathbf{s}_1^{'}\mathbf{z}_t/\sigma_t + \Gamma_t^*(x, s_2, \mathbf{s}_3^{'}, \mathbf{s}_4^{'}, s_5)$, where $\Gamma_t^*(x, s_2, \mathbf{s}_3^{'}, \mathbf{s}_4^{'}, s_5) = (\Lambda_t(s_2, \mathbf{s}_3, \mathbf{s}_4)x)/(\sigma_t(2\sigma_t + s_5/n^{1/2}))$, and

$$\Lambda_t(s_2, \mathbf{s}_3, \mathbf{s}_4) = s_2/n^{1/2} + \sum_{j=1}^{q} s_{3j}\xi_{t-j}/n^{1/2} + \mathbf{s}_4^{'} \sum_{j=1}^{q} b_j \nu_{t-j}/n^{1/2}$$
$$+\mathbf{s}_4^{'} \sum_{j=1}^{q} s_{3j}\nu_{t-j}/n + \sum_{j=1}^{q} [(n^{-1/2}s_{3j} + b_j)\mathbf{s}_4^{'}\mathbf{z}_{t-j}]^2/n,$$

where s_{ij} denotes the j-th entry of $\mathbf{s}_i$. We set $\mathbf{S} = (\mathbf{s}_1^{'}, s_2, \mathbf{s}_3^{'}, \mathbf{s}_4^{'}, s_5) = (S_1, \ldots, S_r)$ with $r = 2q + p + 2$. In view of (4.2), (4.6), (4.7) and the identity $\hat{\sigma}_t/\sigma_t - 1 = (\hat{\sigma}_t^2 - \sigma_t^2)/\sigma_t(2\sigma_t + (\hat{\sigma}_t - \sigma_t))$, it suffices to prove

$$\sup_{\mathcal{T}} |n^{-1/2} \sum_{i=1}^{n} \{I(u_t \le x + \Gamma_t(x, \mathbf{S})) - G(x + \Gamma_t(x, \mathbf{S}))$$
$$+G(x) - I(u_t \le x)\}| = o_P(1)$$

for any $K > 0$ and $0 < \lambda < a$, where a is the real number in (4.1), and

$$\mathcal{T} = \{(x, \mathbf{S}); -\infty < x < \infty, |S_i| \le K, i = 1, \ldots, r-1, |S_r| \le \lambda\}.$$

For $i = 1, \ldots, r-1$ and $j = 1, \ldots, n^2$, we put $S_{ij} = -K + 2Kj/n^2$ and let $S_{rj} = -\lambda + 2\lambda j/n^2$. We denote by $\{\mathcal{C}_j\}$, $j = 1, \ldots, N_n = n^{2r}$, the subrectangles generated by the vertices S_{ij}'s. We define $\Gamma_t^+(i, j) = \sup\{\Gamma_t(x, \mathbf{S}); x_{i-1} < x \le$

$x_i,\ \mathbf{S} \in \mathcal{C}_j\}$ and $\Gamma_t^-(i,j) = \inf\{\Gamma_t(x,\mathbf{S}); x_{i-1} < x \leq x_i,\ \mathbf{S} \in \mathcal{C}_j\}$. Then for $x \in (x_{i-1}, x_i]$ and $\mathbf{S} \in \mathcal{C}_j$, using Taylor's series expansion we have

$$\begin{aligned} & n^{-1/2}\sum_{t=1}^{n}\{G(x_i+\Gamma_t^+(i,j)) - G(x+\Gamma_t(x,\mathbf{S}))\} \\ &= n^{-1/2}\sum_{t=1}^{n}\{G(x_i)-G(x)\} \\ &\quad +n^{-1/2}\sum_{t=1}^{n}\{\Gamma_t^+(i,j))g(x_i) - \Gamma_t(x,\mathbf{S})g(x)\} \\ &\quad +n^{-1/2}\sum_{t=1}^{n}\{(\Gamma_t^+(i,j))^2 g'(\zeta_{ti}) - (\Gamma_t(x,\mathbf{S}))^2 g'(\tilde{\zeta_{ti}})\}/2 \\ &= I_1 + I_2 + I_3 \ \text{ say,} \end{aligned}$$

where ζ_{ti} is a real number between x_i and $x_i + \Gamma_t^+(i,j)$, and $\tilde{\zeta}_{ti}$ is a real number between x and $x + \Gamma_t(x,\mathbf{S})$.

Obviously, $\sup_x |I_1| = o_P(1)$. Since the sup is taken over a bounded set, we can write

$$\Gamma_t^+(i,j) = (\mathbf{s}_1^*)'\mathbf{z}_t/\sigma_t n^{1/2} + \frac{\Lambda(s_2^*, \mathbf{s}_3^*, \mathbf{s}_4^*)y_i}{\sigma_t(2\sigma_t + s_5^*/n^{1/2})}$$

for some $\mathbf{S}^* = \mathbf{S}_t^* \in \bar{\mathcal{C}}_j$, where y_i is either x_i or x_{i-1} according to whether $\Lambda(s_2^*, \mathbf{s}_3^*, \mathbf{s}_4^*)$ is positive or negative. Then, using Lemmas 5.2 and 5.3, we have that

$$\begin{aligned} & |n^{-1/2}\sum_{t=1}^{n}\{\Gamma_t^+(i,j)g(x_i) - \Gamma_t(x,\mathbf{S})g(x)\}| \\ &= n^{-1/2}|\sum_{t=1}^{n}\{(\mathbf{s}_1^*-\mathbf{s}_1)'\mathbf{z}_t g(x_i)/\sigma_t n^{1/2}| \\ &\quad +n^{-1/2}|\sum_{t=1}^{n}\{(\mathbf{s}_1^*-\mathbf{s}_1)'\mathbf{z}_t(g(x_i)-g(x))/\sigma_t n^{1/2}\}| \\ &\quad +n^{-1/2}|\sum_{t=1}^{n}\{\frac{\Lambda(s_2^*,\mathbf{s}_3^*,\mathbf{s}_4^*)}{\sigma_t(2\sigma_t+s_5^*/n^{1/2})} - \frac{\Lambda(s_2,\mathbf{s}_3,\mathbf{s}_4)}{\sigma_t(2\sigma_t+s_5/n^{1/2}))}\}y_i g(x_i)| \\ &\quad +n^{-1/2}\ |\sum_{t=1}^{n}\{\frac{\Lambda(s_2^*,\mathbf{s}_3^*,\mathbf{s}_4^*)}{\sigma_t(2\sigma_t+s_5^*/n^{1/2})} - \frac{\Lambda(s_2,\mathbf{s}_3,\mathbf{s}_4)}{\sigma_t(2\sigma_t+s_5/n^{1/2})}\}(y_i g(x_i) - x g(x))| = o_P(1), \end{aligned}$$

which implies $I_2 = o_P(1)$.

For I_3, notice that due to our assumption, $\max_{j=i,i-1}\sup_{|\theta|\leq a_n, |\tilde{\theta}|\leq b_n} |x_j^2 g'(x_i +\theta + \tilde{\theta}x_j)| = O_P(1)$ for any sequences of r.v.'s $a_n = o_P(1)$ and $b_n = o_P(1)$, and similarly, $\sup_{|\theta|\leq a_n, |\tilde{\theta}|\leq b_n} |x^2 g'(x+\theta+\tilde{\theta}x)| = O_P(1)$ (cf. (4.9)). Hence, in view of

(4.2), (4.6) and (4.7), we have $I_3 = o_P(1)$. As a consequence,

$$\sup_{\mathcal{T}_{ij}} |n^{-1/2} \sum_{t=1}^{n} \{G(x_i + \Gamma_t^+(i,j)) - G(x + \Gamma_t(x, \mathbf{S})\}| = o_P(1), \tag{5.7}$$

where $\mathcal{T}_{ij} = \{(x, \mathbf{S}); x_{i-1} < x \le x_i, \mathbf{S} \in \mathcal{C}_j\}$. Similarly, we have

$$\sup_{\mathcal{T}_{ij}} |n^{-1/2} \sum_{t=1}^{n} \{G(x_{i-1} + \Gamma_t^-(i,j)) - G(x + \Gamma_t(x, \mathbf{S})\}| = o_P(1). \tag{5.8}$$

In view of (5.7), (5.8), the fact that for $\mathbf{S} \in C_j$, one has $I(u_t \le x_{i-1} + \Gamma_t^-(i,j)) \le I(u_t \le x + \Gamma_t(x, \mathbf{S})) \le I(u_t \le x_i + \Gamma_t^+(i,j))$, and the fact that $\sup_{|G(x)-G(y)| \le \tau_n} n^{-1/2} |\sum_{t=1}^n \{I(u_t \le x) - G(x) + G(y) - I(u_t \le y)\}| = o_P(1)$ for any sequence of positive real numbers $\{\tau_n\}$ decaying to 0 (cf. Billingsley (1968, p.106)), we can see that $\sup_x |II_n| = o_P(1)$ if

$$\sup_{1\le i\le n} \sup_{1\le j\le N_n} |n^{-1/2} \sum_{t=1}^{n} \{I(u_t \le x_i + \Gamma_t^+(i,j)) - G(x_i + \Gamma_t^+(i,j)) + G(x_i) - I(u_t \le x_i)\}| = o_P(1), \tag{5.9}$$

$$\sup_{1\le i\le n} \sup_{1\le j\le N_n} |n^{-1/2} \sum_{t=1}^{n} \{I(u_t \le x_{i-1} + \Gamma_t^-(i,j)) - G(x_{i-1} + \Gamma_t^-(i,j)) + G(x_{i-1}) - I(u_t \le x_{i-1})\}| = o_P(1). \tag{5.10}$$

Here, we only prove (5.9) since (5.10) can be proven in a similar fashion.

To this end, we set $d_t = I(u_t \le x_i + \Gamma_t^+(i,j)) - G(x_i + \Gamma_t^+(i,j)) + G(x_i) - I(u_t \le x_i)$ and $\mathcal{D}_L^t = (\sum_{i=1}^t |\xi_{i-j}| \le Ln,\ \sum_{i=1}^t ||\nu_{i-j}|| \le Ln,\ \sum_{i=1}^t ||\mathbf{z}_{i-j}||^2 \le Ln,\ j = 1, \ldots, q)$. Since $P((\mathcal{D}_L^n)^c)$ can be made arbitrarily small by taking a large L due to (4.8), it suffices to prove that

$$P(\sup_{1\le i\le n} \sup_{1\le j\le N_n} n^{-1/2} |\sum_{t=1}^{n} d_t| > \delta, \mathcal{D}_L^n) \to 0 \quad \text{for all } L, \delta > 0. \tag{5.11}$$

For each t, we define $\tilde{d}_t = d_t I(\mathcal{D}_L^t)$. Then (5.11) holds if $\sum_{i=1}^n \sum_{j=1}^{N_n} P(n^{-1/2} |\sum_{t=1}^n \tilde{d}_t| > \delta) \to 0$ for all L, $\delta > 0$. Since $P(\delta_t \ne \tilde{d}_t$ for some $t = 1, \ldots, n, \mathcal{D}_L^n) = P(\emptyset) = 0$. Note that $|\tilde{d}_t| \le 1$, $E\tilde{d}_t = 0$ and

$$\sum_{t=1}^{n} E|\tilde{d}_t|^2 \le E \sum_{t=1}^{n} |G(x_i + \Gamma_t^+(i,j)) - G(x_i)| \le \tilde{K} n^{1/2} \quad \text{for some } \tilde{K} > 0.$$

Applying Bernstein's inequality for a sequence of martingale differences, as in Lee and Wei (1999), we obtain $P(n^{-1/2} |\sum_{t=1}^n \tilde{d}_t| > \delta) \le e^{-\eta n^{1/2}}$ for some $\eta > 0$, which immediately implies $\sum_{i=1}^{n}\sum_{j=1}^{N_n} P(n^{-1/2}|\sum_{t=1}^{n}\tilde{d}_t| > \delta) \leq nN_n e^{-\eta n^{1/2}} \to 0$. Hence, the theorem is established.

Acknowledgement

We are grateful to the Editor, an associate editor and the two referees for their valuable comments which improved the quality of the paper. The first version of this paper was published as Preprint Series No.(S-51, 2000, Sept.) by the Department of Mathematical Science, Osaka University, Japan. In the reviewing process, we were informed by the Editor that the result of Theorem 4.1 has been obtained for a class of ARCH models independently by others (cf. Chap. 8 of Koul (2002)). The first author wishes to acknowledge that this research was supported (in part) by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University.

References

Ash, R. B. (1972). *Real Analysis and Probability.* Academic Press, New York.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Boldin, M. V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27**, 886-871.

Boldin, M. V. (2000). On empirical processes in heteroscedastic time series and their use for hypothesis testing and estimation. *Math. Methods Statist.* **9**, 65-89.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647-671.

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

Bollerslev, T.(1986). General autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327.

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

Chandra, S. A. and Taniguchi, M. (2001). Estimating functions for nonlinear time series models. *Ann. Inst. Statist. Math.* **53**, 125-141.

Chen, M. and An, H. Z. (1998). A note on the stationarity and the existence of moments of the GARCH models. *Ststist. Sinica* **8**, 505-510.

Drost, F. C., Klaassen, C. A. J. and Werker, B. J. M. (1997). Adaptive estimation in time series models. *Ann. Statist.* **25**, 786-817.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279-290.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1006.

Engle, R. (1995). *ARCH Selected Readings.* Oxford University Press, New York.

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.* **47**, 551-579.

Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* **16**, 3-22.

Hallin, M., Taniguchi, M., Serroukh, A. and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27**, 2054-2080.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11**, 818-887.

Koul, H. L. (1992). *Weighted Empiricals and Linear Models*. IMS Lecture Notes-Monograph Series **21**. Hayward, Calif.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. 2nd edition. Lecture Notes in Statistics 166. Springer, New York.

Kreiss, J. P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**, 112-133.

Kreiss, J. P. (1991). Estimation of the distribution function of noise in stationary processes. *Metrika* **38**, 285-297.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

LeCam, L. and Yang, G. L. (1990). *Asymptotics in Statistics; Some Basic Concepts*. Springer-Verlag, New York.

Lee, S. (1996). The asymptotic maxmin property of chi-squared type tests based on the empirical process. *Statist. Probab. Letters* **29**, 285-292.

Lee, S. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with application to time series. *Ann. Statist.* **27**, 237-261.

Ling, S. (1998). Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.* **56**, 741-754.

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

Ling, S. and McAleer, M. (2002). Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18**, 722-729.

Ling, S. and McAleer, M. (2003). On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.* **31**, 642-674.

Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* **9**, 539-569.

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

Strasser, H. (1985). *Mathematical Theory of Statistics*. Walter de Gruyter, New York.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16**, 54-70.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York.

Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Process. Appl.* **21**, 251-273.

Department of Statistics, Seoul National University, Seoul, 171-742, Korea.

E-mail: sylee@stat.snu.ac.kr

Department of Mathematical Sciences, School of Science and Engineering, Waseda University, 3-4-1, Okubo, Shinju-ku, Tokyo, 169-8555, Japan.

E-mail: taniguchi@waseda.jp

(Received August 2002; accepted March 2004)