Journal of the Korean Statistical Society (2000), **29**: *1, pp 9–16*

Regularity of Maximum Likelihood Estimation for ARCH Regression Model with Lagged Dependent Variables [†]

Sun Y. Hwang[1]

ABSTRACT

This article addresses the problem of maximum likelihood estimation in ARCH regression with lagged dependent variables. Some topics in asymptotics of the model such as uniform expansion of likelihood funtion and construction of a class of MLE are discussed, and the regularity property of MLE is obtained. The error process here is possibly non-Gaussian.

Keywords: ARCH Regression ; Lagged dependent variables ; Local sequence of parameters ; Regularity of MLE

1. INTRODUCTION

In the traditional linear time series (or regression), the conditional variance of one-step ahead prediction does not depend on the past information. The class of models admitting additional information from the past to affect the conditional variance is typically two-fold : random coefficient autoregressions (RCA) and autoregressive conditional heteroscedastic processes(ARCH). Whereas RCA mainly studied by time series analysts assumes the conditional variance to vary with past observations, ARCH models postulate the conditional variance evolving with previous innovations (ϵ) and are usually investigated by econometricians.

In a seminal paper, Engle (1982) introduced ARCH concepts in a time series regression which have proven useful in modeling diverse econometric data including inflation rate and stock prices. See, for instance, Bougeral and Picard (1992) for a survey of numerous applications of ARCH models.

In this article we consider the following ARCH-regression model with lagged dependent variables first suggested in Section 5 of Engle (1982) :

$$y_t = x_t'\phi + \varepsilon_t \tag{1.1}$$

[†]This research was supported by a grant from KISTEP made in the program year of 1999.

[1]Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Korea.

where the innovations $\{\varepsilon_t\}$ are defined by

$$\varepsilon_t = h_t^{1/2} \cdot e_t \tag{1.2}$$

with

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \tag{1.3}$$

Here x_t is a vector of k - lagged dependent variables,

$$x_t = (y_{t-1}, \cdots, y_{t-k})' \tag{1.4}$$

It may be noted that ARCH(1) component in (1,3) can be replaced by higher order ARCH(m), viz.,

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 \tag{1.5}$$

Limiting results in this paper can be easily extended to ARCH(m) case and thus we retain ARCH(1) structure for simplicity of presentation.

For the parameter estimation, Engle(1982) assumed the conditional Gaussian model, equivalently, $\{e_t\}$ was assumed a sequence of iid N(0, 1) variables, and derived asymptotic distributions of " maximum liklihood estimator (MLE, for short) " of paramaters. Weiss (1986) discussed the non-normal $\{e_t\}$ case and provided limiting distributions of the least squares estimators as well as quasi-MLE obtained by maximizing quasi-likelihood function derived as though $\{e_t\}$ are, in fact, iid N(0, 1). Regarding the expansion of the likelihood function, Drost and Klassen(1997) investigated the local asymptotic normality for GARCH(1,1) processes.

In the present paper we are concerned with case when $\{e_t\}$ are possibly non-normal with *pdf* f($\cdot$), discussing some intriguing topics in parameter estimation via likelihood function for the above mentioned model. Specifically, under a *minimal* set of conditions, discussed are the *uniform* expansion of the log-likelihood function and the problem of existence and derivation of so called one-step MLE obtained by approximating score function. Moreover some desirable properties of MLE such as regularity and optimality are derived. In a general setting, in particular, covering pure ARCH models without regression component, discussions comparable to ours have been treated by Drost et al. (1997) under some broad assumptions. Those results established in this article have not yet been explicitly addressed in the literature.

2. PRELIMINARIES

Recall the model specified by (1.1) through (1.4) with possibly non-normal $\{e_t\}$. It will be assumed throughout that following conditions are satisfied.

(C.1) $\{e_t\}$ is a iid sequence of random variables with the marginal *pdf* f($\cdot$) of zero mean and variance unity. Also e_t is independent of ε_{t-s}, $s \geq 1$.

(C.2) $\alpha = (\alpha_0,\ \alpha_1)$ is such that $\alpha_0 > 0$, $0 \leq \alpha_1 < 1$.

(C.3) For the $k \times 1$ vector of parameters $\phi = (\phi_1, \cdots, \phi_k)$, all the roots of $1 - \sum_{j=1}^{k} \phi_j\ z^j = 0$ lie outside the unit circle of z.

Remarks : From (C.1) and (C.2), it follows that the innovation process $\{\varepsilon_t\}$ is strictly stationary, which in turn implies together with (C.3) that the observation process $\{y_t\}$ itself is also strictly stationary and ergodic.

Below we briefly mention the $\sqrt{n}$ - consistency of the least squares estimators of ϕ and α (which is due to Weiss (1986)) for the quick reference in discussing maximum likelihood estimation.

Denote the least squares estimators of $\phi = (\phi_1, \cdots, \phi_k)$ and $\alpha = (\alpha_0, \alpha_1)$ by $\hat{\phi}_{LS}$ and $\hat{\alpha}_{LS}$, respectively. Thus, $\hat{\phi}_{LS}$ and $\hat{\alpha}_{LS}$ are so obtained by minimizing the objective function $Q = \sum \varepsilon_t^2 / h_t$.

Result 1 *(Weiss (1986)) : Under (C.1) to (C.3), plus $Ey_t^4 < \infty$, $\hat{\phi}_{LS}$ and $\hat{\alpha}_{LS}$ are $\sqrt{n}$ consistent estimator of ϕ and α, respectively and are asymptotically normally distributed.*

Remarks : Regarding conditions for $Ey_t^4 < \infty$, it is well known for Gaussian $\{e_t\}$ (see, for instance, Engle(1982)) that the condition is equivalent to $\alpha_1^2 < 1/3$. When $\{e_t\}$ are not necessarily Gaussian, we refer to An et al. (1997) for the sufficient conditions for finite fourth (and higher) order moment of $\{y_t\}$.

3. MAIN RESULTS

Let $\{y_{-k+1}, \cdots, y_n\}$ be a sample and define the $(k+2) \times 1$ vector of paramters : $\beta = (\phi', \alpha')'$. Also use the notation $\varepsilon_t(\beta) = y_t - x_t'\phi$, $h_t(\beta) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2(\beta)$

and $e_t(\beta) = \varepsilon_t(\beta)/\sqrt{h_t(\beta)}$ in order to stress the fact that these are functions of parameters β. Denoting by $\ln(\beta)$ the log likelihood function conditional on the initial $y_{-k+1}, \cdots, y_0$, it follows that

$$\ln(\beta) = \sum_{t=1}^{n} \{ \log f[\varepsilon_t(\beta)/\sqrt{h_t(\beta)}] - \frac{1}{2} \log h_t(\beta) \} \tag{3.1}$$

Introduce $\Delta_n(\beta)$ and $\omega_n(\beta)$ for the score vector and the sample Fisher information matrix:

$$\Delta_n(\beta) = \partial \ln(\beta)/\partial\beta \ : \ (k+2) \times 1 \ vector \tag{3.2}$$

$$\omega_n(\beta) = -\partial^2 \ln(\beta)/\partial\beta^2 \ : \ (k+2) \times (k+2) \ matrix \tag{3.3}$$

Also, define the limiting average Fisher information by $F(\beta)$, viz.,

$$F(\beta) = plim[\ n^{-1}\omega_n(\beta)\ ] \tag{3.4}$$

where the existence of $F(\beta)$ is secured by the ergodic theorem.

Let us now consider the following Cramer-type conditions on the density $f(\cdot)$ of e_t.

(C.4) For the location-scale family of densities

$$\{\ g(e_t : a, b) = b^{-1} f[\frac{e_t - a}{b}],\ -\infty < a < \infty, b > 0\ \}$$

there exists a integrable $B(e_t)$ and a constant $c > 0$ such that for all a and b with $|a| < c$ and $|b-1| < c$

$$|\ \partial^3 \log g(e_t : a, b)/\partial a^i \partial b^j\ | \leq B(e_t)$$

where i and j are nonnegative integers summing to 3.

In what follows, fix β representing the true parameter value and introduce β_n : a sequence of parameters converging to the true value β (as the sample size tends to infinity) in such a way

$$\beta_n = \beta + \delta/\sqrt{n} \tag{3.5}$$

with δ being a finite $(k+2) \times 1$ vector of constants.

Theorem 1 : *Suppose that $Ey_t^2 < \infty$. Conditions (C.1) to (C.4) then imply as $n \to \infty$:*

For given $M > 0$,

(i) Uniform local expansion of log-likelihood ;

$$\sup_{|\delta| \le M} [\ \ln(\beta_n) - \ln(\beta) - \{\delta' \Delta_n(\beta)/\sqrt{n} - \frac{1}{2}\delta' F(\beta)\delta\ \}\ ] = o_p(1)$$

(ii) Asymptotic normality of $\Delta_n(\beta)$;

$$n^{-1/2}\Delta_n(\beta) \xrightarrow{d} N(0, F(\beta)) \tag{3.6}$$

(iii) Uniform local linearity of $\Delta_n(\beta)$;

$$\sup_{|\delta| \le M} [\ \Delta_n(\beta_n) - \Delta_n(\beta) + w_n(\beta)(\beta_n - \beta)\ ] = o_p(1)$$

Proof : (i) By a Taylor's expansion of $\ln(\beta)$ we obtain

$$\ln(\beta_n) - \ln(\beta) = \delta' \Delta_n(\beta)/\sqrt{n} - \frac{1}{2n}\delta' w_n(\beta)\delta + R_n(\beta^*, \delta) \tag{3.7}$$

where $R_n(\cdot, \cdot)$ is the remainder consisting of third order derivatives with respect to β and β^* lies between β and β_n. Denoting as in (C.4)

$$g(e_t : a, b) = b^{-1} f[\frac{e_t - a}{b}]$$

it can be written that

$$\ln(\beta_n) = \sum_{t=1}^{n} g[\ e_t\ ;\ n^{-1/2}x_t'\delta_1/h_t(\beta)\ ,\ (h_t(\beta)/h_t(\beta_n))^{1/2}\ ] \tag{3.8}$$

where $k \times 1$ vector of constants δ_1 is as defined in $\delta = (\delta_1', \delta_2')'$.
Also, it follows from $Ey_t^2 < \infty$ that $E[\sup_{|\delta| \le M} |x_t'\delta_1|]^2 < \infty$ and hence

$$\sup_{|\delta| \le M} \sup_{1 \le t \le n} |x_t'\delta_1| = o_p(\sqrt{n})$$

which in turn implies using $h_t(\beta) \ge \alpha_0 > 0$,

$$\sup_{|\delta| \le M} \sup_{1 \le t \le n} |n^{-1/2}x_t'\delta_1/h_t(\beta)| = o_p(1) \tag{3.9}$$

It can also be similarly verified that

$$\sup_{|\delta|\leq M}\sup_{1\leq t\leq n}|\frac{h_t(\beta)}{h_t(\beta_n)}-1|^{1/2}=o_p(1) \tag{3.10}$$

Consequently, combining (3.9) and (3.10) and using (C.4) it can be deduced that

$$\sup_{|\delta|\leq M}|R_n(\beta^*,\delta)|\leq n^{-3/2}M^3\cdot\sup_{1\leq t\leq n}B(e_t)/6 \tag{3.11}$$

where $B(\cdot)$ is defined in (C.4). The integrability of $B(\cdot)$ implies that

$$\sup_{1\leq t\leq n}B(e_t)=o_p(n)$$

which leads to the assertion (i) since $n^{-1}w_n(B)$ converges in probability to $F(\beta)$.

(ii) Note that $\Delta_n(\beta)$ is a sum of zero mean martingale differences. A little algebra reveals also that each martingale increments has a variance $F(\beta)$ defined in (3.4). This essentially yields (ii), by employing the martingale CLT.

(iii) The verification of (iii) follows on a similar lines as in the proof of (i), making crucial use of (C.4) and hence it does not bear repetition.

We now define C_{ML} : a class of maximum likelihood estimators of β, viz. ,

$$C_{ML}=\{\ \hat\beta_{ML}\,;\,\hat\beta_{ML}=t_n+w_n^{-1}(t_n)\Delta_n(t_n)\ \} \tag{3.12}$$

where t_n stands for a preliminary $\sqrt{n}$ - consistent estimator of β. It is worth indicating that C_{ML} is non-empty since one may substitute $\hat\beta_{LS}$: the least squares estimator of β for t_n. Furthermore, $\hat\beta_n$ being usually referred to as the MLE of β, obtained as a root of the ML-equation: $\Delta_n(\hat\beta_n)=0$, belongs to C_{ML}.

Theorem 2 : *Let $\hat\beta_{ML}$ be any member of C_{ML}. Under (C.1) to (C.4), $\hat\beta_{ML}$ is regular in the sense that along with β_n*

$$\sqrt{n}(\hat\beta_{ML}-\beta_n)\xrightarrow{d}N(0,F^{-1}(\beta)) \tag{3.13}$$

Remarks : (1) Under β_n, $\hat\beta_n-\beta_n$ represents estimation error converging in law to zero mean normal distribution, which is a desirable property that a "good" estimator of β may possess. (2) Setting $\delta=0$ in β_n, (3.13) reduces to the asymptotic normality of $\hat\beta_{ML}$ under the true paramer β, i. e.,

$$\sqrt{n}(\hat\beta_{ML}-\beta)\xrightarrow{d}N(0,F^{-1}(\beta))$$

Proof : It is deduced from (i) in Theorem 1 that two probability measures associated with β_n and β (call these P_{β_n} and P_β, respectively) are contiguous (cf. Roussas (1972), ch.2). Thus for any F_n-measurable random quantity, say S_n, $S_n = o_p(1)$ under P_{β_n} is equivalent to $S_n = o_p(1)$ under P_β. Moreover it can be verified from (i) and (ii) in Theorem 1 via the Le Cam's Third Lemma (e.g. Hall and Mathiason (1990)) that under P_{β_n}, $\ln(\beta_n) - \ln(\beta)$ converges to normal distribution with mean $\delta' F(\beta)\delta/2$ and variance $\delta' F(\beta)\delta$. Consequently it follows that along with β_n

$$n^{-1/2}\Delta_n(\beta) \xrightarrow{d} N(F(\beta)\delta, F(\beta)) \tag{3.14}$$

Write that for t_n such that $\sqrt{n}(t_n - \beta)$ is bounded in probability

$$\begin{aligned} \sqrt{n}\,(\,\hat{\beta}_{ML} - \beta\,) &= \sqrt{n}\,[\,t_n + w_n^{-1}(t_n)\Delta_n(t_n) - \beta\,] \\ &= \sqrt{n}\,[\,t_n + w_n^{-1}(\beta)\{\,\Delta_n(\beta) - w_n(\beta)(t_n - \beta)\}\ - \beta\,] + o_p(1) \end{aligned}$$

where the uniform local linearity of $\Delta_n(\beta)$ addressed in (iii) of Theorem 1 is used. After summarizing one can reach at

$$\sqrt{n}(\hat{\beta}_{ML} - \beta) = [\,n^{-1}w_n(\beta)\,]^{-1} n^{-1/2}\Delta_n(\beta) + o_p(1) \tag{3.15}$$

Also, $n^{-1}w_n(\beta)$ converges in probability to $F(\beta)$ under P_{β_n} and P_β as well, which in turn implies by exploiting (3.14) that

$$\sqrt{n}(\hat{\beta}_{ML} - \beta) \xrightarrow{d} N(\delta, F^{-1}(\beta)),\ \textit{under}\ \beta_n$$

Or, equivalently, along with β_n

$$\sqrt{n}(\hat{\beta}_{ML} - \beta_n) \xrightarrow{d} N(0, F^{-1}(\beta))$$

which concludes the theorem.

Indeed, it can be deduced further from Theorem 1 that $\hat{\beta}_{ML} \in C_{ML}$ enjoys certain asymptotic optimality properties under various criteria : asymptotic minimaxity and the minimum variance of the limiting distribution. Refer to Sweeting(1980) and Hall and Mathiason (1990) for relevant discussions.

ACKNOWLEDGEMENT

Thanks are due to two referees for a careful reading of the paper

REFERENCES

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

Bougeral, P. and Picard, N.(1992). "Stationarity of GARCH processes and some non-negative time series". *Journal of Econometrics*, 52, 115-127.

Drost, F.C. and Klaassen, C.A.J.(1997). "Efficient estimation in semiparametric GARCH models". *Journal of Econometrics*, 81, 193-221.

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

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

Hall, W.J. and Mathiason, D.J.(1990). "On large-sample estimation and testing in parametric models". *International Statistics Review*, 58, 77-97.

Roussas, G.G.(1972). *Contiguity of Probability Meausures: Some Applications in Statistics.* Cambridge Univ. Press, London.

Sweeting, T.J.(1980). "Uniform asymptotic normality of the maximum likelihood estimator". *Annals of Statistics*, 8, 1375-1381.

Weiss, A.A(1986). "Asymptotic theory for ARCH models: estimation and testing". *Econometric Theory*, 2, 107-131.