Journal of the Korean Statistical Society (2003), **32**: *4, pp 385–399*

ROBUST ESTIMATION USING QUASI-SCORE ESTIMATING FUNCTIONS FOR NONLINEAR TIME SERIES MODELS†

KYUNGYUP CHA[1], SAHMYEONG KIM[2] AND SUNGDUCK LEE[3]

ABSTRACT

We first introduce the quasi-score estimating function and applied the quasi-score estimating function to nonlinear time series models. We proposed the M quasi-score estimating functions bounded functions for the quasi-score estimating functions. Also, we investigated the asymptotic properties of quasi-likelihood estimators and M quasi-likelihood estimators. Simulation results show that the M quasi-likelihood estimators work better than the least squares estimators under the heavy-tailed distributions

AMS 2000 subject classifications. Primary 62M10; Secondary 62F35.
Keywords. Estimating functions, nonlinear time series, quasi-likelihood estimation, robust estimation, asymptotic normality.

1. INTRODUCTION

The most common approach to the study of point estimation is to suppose that alternative estimators are given as functions of the data and is to compare the properties of estimators in some sense. Uniformly minimum variance unbiased estimators (UMVUEs) are one of the typical targets in point estimation. The other approach is to study estimating functions based on the data and parameters. Godambe (1960) proposed the theory of estimating function which established on

Received May 2002; accepted July 2003.
†This work was supported by grant No. R05-2000-000-00023-0 from the Basic Research Program of the Korea Science & Engineering Foundation.
[1]Corporate risk management department, Korea Credit Guarantee Fund, Seoul 121-744, Korea (e-mail : kycha68@empal.com)
[2]Department of Statistics, ChungAng University, Seoul 156-756, Korea (e-mail : sahm@cau.ac.kr)
[3]Department of Statistics, Chungbuk National University, Cheongju 361-763, Korea (e-mail : sdlee@cbucc.chungbuk.ac.kr)

optimal property of maximum likelihood estimator for identically and independently distributed observations. And quasi-likelihood estimating function which is an optimal estimating function in terms of Godambe's criteria has been very useful tool for estimating the parameters of interest when the underlying distribution is not fully specified except for the mean and the variance structures. The original idea of quasi-likelihood estimation was proposed by Wedderburn (1974) and this has been widely used in generalized linear models. The theory of estimating functions and quasi-likelihood estimation developed independently. Godambe (1985) showed that the quasi-score function of Wedderburn (1974) satisfied the optimal properties of estimatings for discrete time stochastic processes with martingale structure. Huber (1964) proposed robust estimation under the heavy-tailed distributions of the errors for regression models. And recently robust estimation methods have been applied to linear time series models. Denby and Martin (1979) applied Huber's M estimator to AR (Auto-Regressive) models and compared the efficiency with least squares estimators. Bustos and Yohai (1986) proposed the RA (Residual Autocovariance) estimation and investigate asymptotic properties of RA estimator for ARMA (Auto-Regressive Moving Average) models. Also, Martin and Yohai (1985) summarized the concepts and method of robustness in linear time series model. But not much research for robust estimation is in process for nonlinear time series models. In the paper, we applied the quasi-score estimating function and the proposed M quasi-score estimating function comprising bounded functions for nonlinear time series models. Also we investigated the asymptotic properties of estimators. We compare the unbiasedness and efficiency for the (conditional) least squares estimator and the proposed estimators in RCA (Random Coefficient Auto-Regressive) models and ARCH (Auto-Regressive Conditional Heteroscadastic) models with several error distributions by simulation.

2. Quasi-score Estimating Functions for Nonlinear Time Series Models

Various authors have considered estimators which are solutions of the estimating equations and investigated their asymptotic properties under the assumption that the model is correct. Consider a nonlinear time series model satisfying a recursive equation, *i.e.*,

$$X_t = H(\mathbf{X}_{t-1}, \mathbf{Z}_t; \boldsymbol{\theta}) + \epsilon_t, \;\; t \geq 1, \tag{2.1}$$

where $\{\epsilon_t\}$ is an unobserved sequence of *iid* random variables with mean zero and variance σ^2. Also, H is a known function, $\mathbf{X}_{t-1} = (X_{t-1}, \ldots, X_{t-p})$, $p \geq 1$, $\boldsymbol{\theta}$ is an unknown $p \times 1$ vector parameter, and $\{\mathbf{Z}_t\}$ is a sequence of unobserved *iid* random vectors independent of $\{\epsilon_t\}$. It is assumed that $\{X_t\}$ is stationary and ergodic. Let $S_n(\boldsymbol{\theta})$ be a quasi-score estimating function of the form

$$S_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \{X_t - \mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})\} \frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\boldsymbol{\theta}} \tag{2.2}$$

where $\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta}) = E(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_{t-1}$ is a σ-field generated by $X_{t-1}, \ldots, X_{t-p}$, $t \geq 1$. An estimator $\widehat{\boldsymbol{\theta}}_n$ is obtained by solving the estimating equation $S_n(\boldsymbol{\theta}) = \mathbf{0}$. Under regularity conditions, existence of a $\sqrt{n}$-consistent choice of such a $\widehat{\boldsymbol{\theta}}_n$ can be obtained. On the other hand, a solution of $S_n(\boldsymbol{\theta}) = \mathbf{0}$ is the same as the conditional least squares (CLS) estimator of $\boldsymbol{\theta}$ which minimizes $\sum_{t=1}^{n} U_t^2(\boldsymbol{\theta})$ where

$$U_t(\boldsymbol{\theta}) = X_t - \mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta}). \tag{2.3}$$

Hwang *et al.* (1994) established the consistency and the asymptotic normality of the CLS estimators for a class of nonlinear time series models. Consider the following regularity conditions:

1. For each $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$, $\{X_t,\, t \geq 1\}$ is stationary and ergodic.

2. The conditional expectation $\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})$ is a function $\mathbf{X}_{t-1}$ and $\boldsymbol{\theta}$ only not depending on any other nuisance parameters.

3. The $p \times p$ matrix $F(\boldsymbol{\theta})$ defined by

$$F(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[\left\{ \frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\} \left\{ \frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right\}^T \right]$$

   exists and F is positive definite for each $\boldsymbol{\theta} \in \Theta$, where $d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})/d\boldsymbol{\theta}$ denotes the $p \times 1$ vector of first derivatives of μ_t with respect to $\boldsymbol{\theta}$.

4. For all sequences $\boldsymbol{\theta}^*$ such that $\boldsymbol{\theta}^* \xrightarrow{p} \boldsymbol{\theta}$,

$$\frac{1}{n} \sum_{t=1}^{n} \left\{ g_t(\boldsymbol{\theta}^*) - g_t(\boldsymbol{\theta}) \right\} \xrightarrow{p} 0,$$

   where

$$g_t(\boldsymbol{\theta}) = \frac{d}{d\boldsymbol{\theta}} \left[\{X_t - \mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})\} \cdot \frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right].$$

REMARK 2.1. Hwang *et al.* (1994) pointed out that it is possible to relax condition 2 and allow $\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})$ to depend on unknown nuisance parameters. Sufficient conditions for condition 1 are given in Tong (1990). Hwang *et al.* (1994) and Klimko and Nelson (1978) assumed that the third derivative of $\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})$ is bounded in a neighborhood of $\boldsymbol{\theta}$.

THEOREM 2.1. *Let $\widehat{\boldsymbol{\theta}}_n$ be a consistent solution of $S_n(\boldsymbol{\theta}) = \mathbf{0}$. Then, under the conditions 1 to 4, we have*

(i) $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$, *as* $n \to \infty$,

$$\text{(ii)}\ \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, F^{-1}(\boldsymbol{\theta})), \tag{2.4}$$

where

$$F(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\left\{\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right\}\left\{\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right\}^T\right].$$

PROOF. (i) By ergodic theorem, $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$.
(ii) Since $S_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$, we have, by Taylor's expansion, for a $\boldsymbol{\theta}^*$ lying on the line segment jointing $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$,

$$\begin{aligned} -\frac{1}{\sqrt{n}}S_n(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\left.\frac{dS_n(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \\ &= \left\{\frac{1}{n}\sum_{t=1}^{n} g_t(\boldsymbol{\theta}^*)\right\}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}). \end{aligned} \tag{2.5}$$

Note that

$$\frac{1}{n}\sum_{t=1}^{n} g_t(\boldsymbol{\theta}^*) - \frac{1}{n}\sum_{t=1}^{n} g_t(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{0} \tag{2.6}$$

by condition 4. We have

$$\begin{aligned} g_t(\boldsymbol{\theta}) &= \frac{d}{d\boldsymbol{\theta}}\left[\{X_t - \mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})\}\cdot\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right] \\ &= -\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\left\{\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right\}^T + \{X_t - \mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})\}\frac{d^2\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}. \end{aligned}$$

Note that

$$E_{\boldsymbol{\theta}}\left[\{X_t - \mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})\}\frac{d^2\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\right]$$
$$= E_{\boldsymbol{\theta}}\left[\frac{d^2\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}E_{\boldsymbol{\theta}}\{(X_t - \mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta}))\big|\,\mathcal{F}_{t-1}\}\right]$$
$$= \mathbf{0}.$$

Therefore, by using the ergodic theorem, we have

$$-\frac{1}{n}\sum_{t=1}^{n} g_t(\boldsymbol{\theta}) \xrightarrow{a.s.} E_{\boldsymbol{\theta}}\left[\left\{\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right\}\left\{\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right\}^T\right] = F(\boldsymbol{\theta}). \quad (2.7)$$

It is also seen that $\{S_n(\boldsymbol{\theta}), \mathcal{F}_n\}$, $n \geq 1$, is a zero mean martingale with covariance matrix $F(\boldsymbol{\theta})$. By CLT for martingales (Hall and Heyde, 1980)

$$\frac{1}{\sqrt{n}}S_n(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, F(\boldsymbol{\theta})). \quad (2.8)$$

Therefore, by (2.6), (2.7) and (2.8),

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, F^{-1}(\boldsymbol{\theta})). \quad (2.9)$$

□

To illustrate this result, we consider simple examples of the nonlinear time series models.

EXAMPLE 2.1 (*RCA(1) model*). Let

$$X_t = (\theta + Z_t)X_{t-1} + \epsilon_t,\ t = 1, \ldots, n$$

where $\{\epsilon_t\}$ are *iid* random variables with zero mean and finite variance σ^2 and $\{Z_t\}$ *iid* random variables with zero mean and variance σ_z^2. Note that $\{Z_t\}$ are independent of $\{\epsilon_t\}$, $t = 1, \ldots, n$. Then the estimating function is

$$S_n(\theta) = \sum_{t=1}^{n}(X_t - \theta X_{t-1})X_{t-1}. \quad (2.10)$$

The solution of $S_n(\theta) = 0$ is

$$\widehat{\theta} = \frac{\sum_{t=1}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_{t-1}^2} \quad (2.11)$$

and

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, F^{-1}(\theta)), \tag{2.12}$$

where

$$F^{-1}(\theta) = \frac{\sigma_z^2 E(X_{t-1}^4) + \sigma^2 E(X_{t-1}^2)}{\{E(X_{t-1}^2)\}^2}.$$

EXAMPLE 2.2 (*ARCH(1) model*). Let $X_1, X_2, \ldots, X_n$ be observations from the ARCH(1) model defined as follow.

$$X_t = \theta X_{t-1} + \epsilon_t,$$

where $|\theta| < 1$ and

$$E(\epsilon_t | \mathcal{F}_{t-1}) = 0,$$

$$\mathrm{Var}(\epsilon_t | \mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 \epsilon_{t-1}^2, \ \alpha_0 > 0, \ 0 < \alpha_1 < 1.$$

We have

$$\mu_t(X_{t-1}; \theta) = E(X_t | X_{t-1}) = \theta X_{t-1}$$

and

$$v_t(\theta) = \mathrm{Var}(\epsilon_t | X_{t-1}) = \alpha_0 + \alpha_1 \epsilon_{t-1}^2.$$

The quasi-likelihood estimating function is

$$\begin{aligned} S_n(\theta) &= \sum_{t=1}^{n} \{X_t - \mu_t(X_{t-1}; \theta)\} \cdot {v_t}^{-1}(\theta) \cdot \frac{d\mu_t(X_{t-1}; \theta)}{d\theta} \\ &= \sum_{t=1}^{n} \left(\frac{\epsilon_t X_{t-1}}{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right). \end{aligned} \tag{2.13}$$

If $\widehat{\theta}$ is a consistent solution of $S_n(\theta) = 0$, we have

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, F^{-1}(\theta)), \tag{2.14}$$

where

$$F(\theta) = E\left(\frac{X_{t-1}^2}{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right).$$

3. M QUASI-SCORE ESTIMATING FUNCTIONS FOR NONLINEAR TIME SERIES MODELS

In this section, we will investigate the asymptotic properties of robust estimators of parameters in nonlinear time series models by using maximum likelihood type estimators (M-estimators). Under the general model in (2.1), consider the estimating function

$$\begin{aligned} S_n^*(\boldsymbol{\theta}) &= \sum_{t=1}^{n} \psi_1\left(\epsilon_t(\boldsymbol{\theta})\right) \psi_2\left(\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right) \\ &= \sum_{t=1}^{n} \begin{pmatrix} g_1(X_t;\boldsymbol{\theta}) \\ \vdots \\ g_p(X_t;\boldsymbol{\theta}) \end{pmatrix} \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} \mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta}) &= E(X_t|\mathcal{F}_{t-1}), \\ \mathcal{F}_{t-1} &= \sigma(X_{t-1},\dots,X_{t-p}), \\ \epsilon_t(\boldsymbol{\theta}) &= X_t - H(\mathbf{X}_{t-1}, Z_t;\boldsymbol{\theta}). \end{aligned}$$

Note that

$$\psi_2\left(\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right) = \begin{pmatrix} \psi_2\left(\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\theta_1}\right) \\ \vdots \\ \psi_2\left(\frac{d\mu_t(\mathbf{X}_{t-1};\boldsymbol{\theta})}{d\theta_p}\right) \end{pmatrix}.$$

We will call $S_n^*(\boldsymbol{\theta})$ as the M quasi-score estimating function. Next, we will establish the asymptotic properties of the estimator satisfying the estimating equation $S_n^*(\boldsymbol{\theta}) = \mathbf{0}$ in (3.1). Consider the regularity conditions:

1. $\{X_t, t \geq 1\}$ is stationary and ergodic;
2. ψ_1 and ψ_2 are bounded and differentiable in $\boldsymbol{\theta}$;
3. $E\{\psi_1(\epsilon_1)\} = 0.$;
4. For all $\boldsymbol{\theta}^*$ such that $\boldsymbol{\theta}^* \xrightarrow{p} \boldsymbol{\theta}$,

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{d}{d\boldsymbol{\theta}}g(X_t;\boldsymbol{\theta}) - \frac{d}{d\boldsymbol{\theta}}g(X_t,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \xrightarrow{p} \mathbf{0};$$

5. The $p \times p$ matrix $\mathbf{A} = E_{\boldsymbol{\theta}}(dg_i/d\theta_j)$ exists for all $i, j = 1, \ldots, p$ and $\mathbf{A}$ is a non-singular matrix;

6. The $p \times p$ matrix $\mathbf{B} = E_{\boldsymbol{\theta}}(dg_i/d\theta_j)$ exists for all $i, j = 1, \ldots, p$ and $\mathbf{B}$ is a positive definite matrix.

THEOREM 3.1. *Under the conditions 1, 2, 3 and 6, we have the following result*

$$\frac{1}{\sqrt{n}} S_n^*(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}). \tag{3.2}$$

PROOF. Let $\mathbf{C} = (C_1, \ldots, C_p)$ be an non-zero vector of constants. Define

$$\begin{aligned} T_n(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\, \mathbf{C}^T S_n^*(\boldsymbol{\theta}) \\ &= \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \sum_{j=1}^{P} C_j\, \psi_1(\epsilon_t(\boldsymbol{\theta}))\, \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_j} \right) \\ &= \sum_{t=1}^{n} Z_t(\boldsymbol{\theta}) \end{aligned}$$

where

$$Z_t(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{p} C_j\, \psi_1(\epsilon_t(\boldsymbol{\theta}))\, \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_j} \right).$$

Note that $T_n(\boldsymbol{\theta})$ is a zero mean martingale. Since ψ_1 and ψ_2 are bounded, using Corollary 3.1 of Hall and Heyde (1980), it is enough to show that

$$\sum_{t=1}^{n} E_{\boldsymbol{\theta}} \left\{ Z_t^2(\boldsymbol{\theta})\ \mathcal{F}_{t-1} \right\} \xrightarrow{a.s.} \mathbf{C}^T \mathbf{B} \mathbf{C}. \tag{3.3}$$

Note that

$$\begin{aligned} &\sum_{t=1}^{n} E_{\boldsymbol{\theta}} \left(Z_t^2(\boldsymbol{\theta}) | \mathcal{F}_{t-1} \right) \\ &= \frac{1}{n} \sum_{t=1}^{n} \sum_{i,j=1}^{p} C_i\, C_j\, \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_i} \right) \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_j} \right) E \left\{ \psi_1^2(\epsilon_t(\boldsymbol{\theta})) | \mathcal{F}_{t-1} \right\} \\ &\xrightarrow{a.s.} \sum_{i,j=1}^{p} C_i\, C_j\, E \left\{ \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_i} \right) \psi_2 \left(\frac{d\mu_t(\mathbf{X}_{t-1}; \boldsymbol{\theta})}{d\theta_j} \right) \right\} E \left\{ \psi_1^2(\epsilon_1(\boldsymbol{\theta})) \right\} \\ &\quad = \mathbf{C}^T \mathbf{B} \mathbf{C}. \end{aligned}$$

Thus

$$\frac{1}{\sqrt{n}}\mathbf{C}^T S_n^*(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{C}^T\mathbf{B}\mathbf{C}) \tag{3.4}$$

Hence by the Cramer-Wold device,

$$\frac{1}{\sqrt{n}} S_n^*(\boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}) \tag{3.5}$$

completing the proof of Theorem 3.1. □

The following theorem establishes the asymptotic normality of the estimator which satisfies the estimating equation $S_n^*(\boldsymbol{\theta}) = \mathbf{0}$ in (3.1).

THEOREM 3.2. *Suppose that* $\widehat{\boldsymbol{\theta}}_n$ *is a* $\sqrt{n}$*-consistent solution of* $S_n^*(\boldsymbol{\theta}) = \mathbf{0}$ *in (3.1). Then under the conditions 1-6, we have*

(i) $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$, *as* $n \to \infty$,

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$, (3.6)

where

$$\mathbf{V} = (\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A})^{-1}.$$

PROOF. (i) It is easily seen that $\widehat{\boldsymbol{\theta}}_n$ is a consisitent estimator by the ergodic theorem.
(ii) Since $S_n^*(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$, by a first order Taylor's expansion, we have

$$\begin{aligned} -\frac{1}{\sqrt{n}} S_n^*(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}} \frac{dS_n^*(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \\ &= \frac{1}{n}\sum_{t=1}^{n} \frac{dg(X_t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}), \end{aligned}$$

for a $\boldsymbol{\theta}^*$ lying between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n$. Note that

$$\frac{1}{n}\sum_{t=1}^{n} \frac{dg(X_t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \frac{1}{n}\sum_{t=1}^{n} \frac{dg(X_t;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \xrightarrow{p} \mathbf{0}, \tag{3.7}$$

by condition 4 and that $\widehat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent for $\boldsymbol{\theta}$. By ergodicity, we have

$$-\frac{1}{n}\sum_{t=1}^{n} \frac{dg(X_t;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \xrightarrow{a.s.} \mathbf{A}, \tag{3.8}$$

where $\mathbf{A} = E_{\boldsymbol{\theta}}(dg_i/d\theta_j)$, $i,j = 1,\ldots,p$. Therefore, we have the final result by using (3.4) and (3.8)

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{d} N(\mathbf{0},\mathbf{V}), \tag{3.9}$$

where

$$\mathbf{V} = (\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A})^{-1}.$$

□

REMARK. The existence of a consistent solution of $S_n^*(\boldsymbol{\theta}) = \mathbf{0}$ can be established as in the context of Huber (1981).

EXAMPLE 3.1 (*RCA(1) model*). The M quasi-score estimating function for the RCA(1) Model is

$$S_n^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} \psi_1(\epsilon_t(\boldsymbol{\theta}))\,\psi_2(X_{t-1}). \tag{3.10}$$

Suppose that $\widehat{\boldsymbol{\theta}}_{M1}$ is a consistent solution of $S_n^*(\boldsymbol{\theta}) = \mathbf{0}$ in (3.10). Then, we have the limiting distribution for a solution of the estimating equation in (3.9). Thus,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{M1} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V_1(\boldsymbol{\theta})), \tag{3.11}$$

where

$$V_1(\boldsymbol{\theta}) = E\Big\{\psi_1(\epsilon_t(\boldsymbol{\theta}))\,\psi_2(X_{t-1})\Big\}^2 E\Big\{\frac{d\psi_1(\epsilon_t(\boldsymbol{\theta}))}{d\boldsymbol{\theta}}\cdot\psi_2(X_{t-1})\Big\}^{-2}.$$

Note that if $\psi_1(x) = \psi_2(x) = x$, then

$$V_1(\boldsymbol{\theta}) = \{\sigma^2 E(X_{t-1}^2) + \sigma_Z^2 E(X_{t-1}^4)\}\,E(X_{t-1}^2)^{-2} \tag{3.12}$$

where $\sigma^2 = \mathrm{Var}(\epsilon_t)$ and $\sigma_Z^2 = \mathrm{Var}(Z_t)$.

EXAMPLE 3.2 (*ARCH(1) model*). The M quasi-score estimating function for the ARCH(1) Model is

$$S_n^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} \psi_1(\epsilon_t(\boldsymbol{\theta}))\,\psi_2(X_{t-1}). \tag{3.13}$$

Suppose that $\widehat{\boldsymbol{\theta}}_{M2}$ is a consistent solution of $S_n^*(\boldsymbol{\theta})$ in (3.13). Then, we have the limiting distribution such as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{M2} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V_2(\boldsymbol{\theta})), \tag{3.14}$$

where

$$V_2(\boldsymbol{\theta}) = E\Big\{\psi_1(\epsilon_t(\boldsymbol{\theta}))\,\psi_2\,(X_{t-1})\Big\}^2 E\Big\{\frac{d\psi_1(\epsilon_t(\boldsymbol{\theta}))}{d\boldsymbol{\theta}}\,\psi_2\,(X_{t-1})\Big\}^{-2}.$$

Note that if $\psi_1(x) = \psi_2(x) = x$, then

$$V_2(\boldsymbol{\theta}) = E\left(\frac{X_{t-1}^2}{\alpha_0 + \alpha_1 \epsilon_{t-1}^2}\right)^{-1}.$$

4. Simulation Results

A simulation study is carried out to compare the bias and the limiting variances for conditional least squares estimators, quasi-likelihood estimators. and M quasi-likelihood estimators in RCA(1) and ARCH(1) model. We consider 3 different error distributions, *i.e.*, standard normal distribution, *i.e.*, $\gamma N(0,1) + (1-\gamma)N(0,\sigma_c^2)$ where $\sigma_c^2 = 10$ and $\gamma = 0.05$ and the standard double exponential distribution. The sample size of n=200, 400 are used. We consider the various bounded functions, *i.e.* Huber function (Huber, 1964), Tukey function (Beaton and Tukey, 1974), Andrews function (Andrews *et al.*, 1972), and Hampel function (Andrews *et al.*, 1972).

1. Huber functions(Hu) : $\psi(x) = \begin{cases} x, & |x| \le k, \\ k\,\mathrm{sgn}(x), & |x| > k, \end{cases}$

2. Tukey function(Tu) : $\psi(x) = \begin{cases} x\left(1 - \dfrac{x^2}{k^2}\right)^2, & |x| < k, \\ 0, & |x| > k, \end{cases}$

3. Andrews function(An) : $\psi(x) = \begin{cases} \sin\left(\dfrac{x}{k}\right), & |x| < k\pi, \\ 0, & |x| > k\pi, \end{cases}$

4. Hampel function (Ha) : $\psi(x) = \begin{cases} x, & |x| \le k, \\ k\,\mathrm{sgn}(x), & k < |x| \le l, \\ \dfrac{k\,\mathrm{sgn}(x)\,(m - |x|)}{m - l}, & l < |x| \le m, \\ 0, & |x| > m. \end{cases}$

TABLE 4.1 *Bias and ARE in RCA(1) model with Standard Normal error*

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.0583	-	0.0474	-	0.0547	-	0.0377	-
QL	0.0573	0.9046	0.0417	0.9034	0.0515	0.8817	0.0383	0.8815
MQL_Hu	0.0592	0.9424	0.0412	0.9276	0.0555	0.9155	0.0382	0.9623
MQL_Tu	0.0613	0.9712	0.0433	0.9684	0.0570	0.9915	0.0405	0.9959
MQL_An	0.0594	0.9433	0.0422	0.9316	0.0548	0.9136	0.0380	0.8996
MQL_Ha	0.0664	1.1122	0.0443	0.9674	0.0549	0.9237	0.0383	0.8734

TABLE 4.2 *Bias and ARE in RCA(1) model with Double Exponential error*

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.0872	-	0.0693	-	0.0843	-	0.0651	-
QL	0.0764	0.7431	0.0533	0.7549	0.0631	0.7988	0.0547	0.7443
MQL_Hu	0.0637	0.9214	0.0463	0.9346	0.0593	0.9025	0.0414	0.9176
MQL_Tu	0.0683	0.9026	0.0470	0.9635	0.0601	0.9217	0.0476	0.9395
MQL_An	0.0732	0.9416	0.0564	0.9477	0.0673	0.9436	0.0499	0.9518
MQL_Ha	0.0712	0.9327	0.0474	0.9286	0.0597	0.9388	0.0498	0.9614

TABLE 4.3 *Bias and ARE in RCA(1) model with Contaminated Normal error*
(γ = 0.05 , σ_c^2 =10)

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.0891	-	0.0619	-	0.0826	-	0.0594	-
QL	0.0823	0.7484	0.0595	0.7175	0.0717	0.7192	0.0536	0.6857
MQL_Hu	0.0598	0.9100	0.0433	0.8925	0.0646	0.8713	0.0394	0.8498
MQL_Tu	0.0687	0.9488	0.0492	0.9364	0.0679	0.9012	0.0393	0.8221
MQL_An	0.0673	0.9237	0.0473	0.9118	0.0634	0.8834	0.0393	0.8551
MQL_Ha	0.0682	0.9637	0.0473	0.9413	0.0641	0.8439	0.0436	0.9326

TABLE 4.4 *Bias and ARE in ARCH(1) model with Standard Normal error*

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.0631	-	0.0450	-	0.0580	-	0.0416	-
QL	0.0599	1.0074	0.0424	1.0036	0.0538	1.0045	0.0392	0.9917
MQL_Hu	0.0586	1.0518	0.0445	1.0759	0.0549	1.0056	0.0390	1.0606
MQL_Tu	0.0577	1.2138	0.0463	1.2329	0.0579	1.1960	0.0411	1.2329
MQL_An	0.0549	1.0746	0.0442	1.0789	0.0553	1.0534	0.0388	1.0789
MQL_Ha	0.0573	1.0647	0.0452	1.0684	0.0566	1.1024	0.0391	1.0684

TABLE 4.5 *Bias and ARE in ARCH(1) model with Double Exponential error*

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.0897	-	0.0832	-	0.0728	-	0.0708	-
QL	0.0599	0.7614	0.0563	0.7298	0.0639	0.7413	0.0577	0.7129
MQL_Hu	0.0603	0.9893	0.0473	0.9907	0.0579	0.9832	0.0483	0.9876
MQL_Tu	0.0647	1.0725	0.0482	1.0011	0.0606	1.0789	0.0488	1.0736
MQL_An	0.0638	1.0636	0.0493	0.9999	0.0603	1.0334	0.0479	1.0024
MQL_Ha	0.0625	1.0412	0.0461	0.9982	0.0598	1.0477	0.0474	1.0981

TABLE 4.6 *Bias and ARE in ARCH(1) model with Contaminated Normal error* (γ = 0.05 , σ_c^2 =10)

Estimator	$\theta = 0.3$				$\theta = 0.5$			
	Sample size				Sample size			
	200		400		200		400	
.	Bias	ARE	Bias	ARE	Bias	ARE	Bias	ARE
CLS	0.1015	-	0.0726	-	0.0894	-	0.0607	-
QL	0.0839	0.6578	0.0607	0.6357	0.0723	0.6461	0.0529	0.6344
MQL_Hu	0.0678	0.9543	0.0470	0.9401	0.0606	0.9378	0.0452	0.9634
MQL_Tu	0.0689	1.1967	0.0473	1.1879	0.0602	1.1833	0.0463	1.0634
MQL_An	0.0684	0.9877	0.0510	0.9889	0.0610	0.9893	0.0467	0.9338
MQL_Ha	0.0670	0.9725	0.0513	1.0634	0.5995	0.9623	0.0464	0.9684

Let $\widehat{\boldsymbol{\theta}}_{CLS}$ be the conditional least squares estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{QL}$ be the quasi-likelihood estimator, and $\widehat{\boldsymbol{\theta}}_{MQL}$ be the M quasi-likelihood estimator. Define asymptotic relative efficiency (ARE) of the conditional least squares estimator with respect to the quasi-likelihood estimator and M quasi-likelihood estimator as follows:

$$ARE_{QL} = \frac{\mathrm{Var}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_{QL} - \boldsymbol{\theta})\}}{\mathrm{Var}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta})\}}, \tag{4.1}$$

$$ARE_{MQL} = \frac{\mathrm{Var}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_{MQL} - \boldsymbol{\theta})\}}{\mathrm{Var}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta})\}}. \tag{4.2}$$

We repeated the whole procedure 500 times and we computed the average of biases and AREs. For RCA(1) model with standard normal error, quasi-likelihood estimators, and M quasi-likelihood estimators using Huber function have the smaller bias than those of the other estimators, And when error distributions are double exponential and contaminated normal distribution, M quasi-likelihood estimators using Huber function have smaller biases than those of the least squares and quasi-likelihood estimators.

For ARCH(1) model with standard normal error, least squares estimators have the smaller variance than those of the quasi-likelihood estimators and M quasi-likelihood estimators. But under the contaminated normal and double exponential distributions, quasi-likelihood estimators have the small variance than those of the other estimators. Also, the M quasi-likelihood estimators except the Huber function have bigger variances than the least square estimators. And M quasi-likelihood estimators using Huber function have smaller bias than the other estimators.

5. Conclusion

Robustness concepts have been a main role to estimate parameters of interest under non-normal situations such as heavy-tailed distributions of the data structure to be investigated. We have introduced a new type of robust estimating functions to investigate the asymptotic properties of the solutions of the estimating equations for nonlinear time seires models. Simulation results show that when error distributions are double exponential and contaminated normal distribution, M quasi-likelihood estimators using Huber function have smaller biases than those of the least squares and quasi-likelihood estimators for RCA(1) model.

But for ARCH(1) model, under the contaminated normal and double exponential distributions, quasi-likelihood estimators have the small variance than those of the other estimators.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. AND TUKEY, J. W. (1972). *Robust Estimates of Location : Survey and Advance*, Princeton University Press, Princeton.

BEATON, A. E. AND TUKEY, J. W. (1974). "The fitting of power series, meaning polynomials, illustrated on Band-Spectroscopic Data", *Technometrics*, **16**, 147–185.

BILLINGSLEY, P. (1961). "The Lindeberg-Levy theorem for martingales", *Proceedings of the American Mathematical Society*, **12**, 788–792.

BUSTOS, O. H. AND YOHAI, V. J. (1986). "Robust estimates for ARMA models", *Journal of the American Statistical Association*, **81**, 155–168.

DENBY, L. AND MARTIN, R. D. (1979). "Robust estimation of the first-order autoregressive parameter", *Journal of the American Statistical Association*, **74**, 140–146.

GODAMBE, V. P. (1960). "An optimum property of regular maximum likelihood estimation" *The Annals of Mathematical Statistics*, **31**, 1208–1211.

GODAMBE, V. P. (1985). "The foundations of finite sample estimation in stochastic processes", *Biometrika*, **72**, 419–428.

HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press, New York.

HUBER, P. J. (1964). "Robust estimation of a location parameter", *The Annals of Mathematical Statistics*, **35**, 73–101.

HUBER, P. J. (1981). *Robust Statistics*, Wiley, New York.

HWANG, S. Y., BASAWA, I. V. AND REEVES, J. (1994). "The asymptotic distributions of residual autocorrelations and related tests of fit for a class of nonlinear time series models", *Statistica Sinica*, **4**, 107–125.

KIM, S. (1998). *Inference for Nonlinear Time Series Models via Estimating Functions*, Ph.D. Dissertations, University of Georgia, Athens, Georgia.

KLIMKO, L. A. AND NELSON, P. I. (1978). "On conditional least squares estimation for stochastic processes", *The Annals of Statistics*, **6**, 629–642.

MARTIN, R. D. AND YOHAI, V. J. (1985). "Robustness in time series and estimating ARMA models", In *Handbook of Statistics V*, North-Holland, Amsterdam, 119–155.

NICHOLLS, D. F. AND QUINN, B. G. (1982). *Random Coefficient Autoregressive Models : An Introduction*, No. 11 of Lecture Note in Statistics, Springer-Verlag, New York.

TONG, H. (1990). *Nonlinear Time Series : A Dynamical System Approach*, Oxford University Press, Oxford.

WEDDERBURN, R. W. M. (1974). "Quasi-likelihood functions, generalized linear models and the Gauss-Newton method", *Biometrika*, **61**, 439–447.