J. Appl. Math. & Informatics Vol. 27(2009), No. 1 - 2, pp. 315 - 326
Website: http://www.kcam.biz

EVALUATION OF PARAMETER ESTIMATION METHODS FOR NONLINEAR TIME SERIES REGRESSION MODELS

TAE SOO KIM* AND JUNG-HO AHN

ABSTRACT. The unknown parameters in regression models are usually estimated by using various existing methods. There are several existing methods, such as the least squares method, which is the most common one, the least absolute deviation method, the regression quantile method, and the asymmetric least squares method. For the nonlinear time series regression models, which do not satisfy the general conditions, we will compare them in two ways: 1) a theoretical comparison in the asymptotic sense and 2) an empirical comparison using Monte Carlo simulation for a small sample size.

AMS Mathematics Subject Classification : 62F12, 62E20, 62M10.
Key words and phrases : Regression estimators; consistency; asymptotic normality; Monte Carlo simulation

1. Introduction

Generally, the nonlinear regression model is $y_t = f(x_t, \theta_0) + \epsilon_t$, where $t = 1, 2, \cdots, T$, and $f(x_t, \theta_0)$ is a real valued nonlinear function defined on $R^{p_1+p_2}$, x_t is a $(1 \times p_2)$ observed vector; and the error terms ϵ_t are independent and identically distributed (i.i.d.) with finite variance. The parameter vector θ_0, which is interior point in a compact parameter space $\Theta \subset R^{p_1}$, is unknown and to be estimated. Given the observation y_t, any vector $\hat{\theta}_T$ in Θ minimizing the objective function $S_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}(y_t - f(x_t, \theta))^2$ is called the least squares estimator (LSE) of θ_0 based on $\{y_t\}_{t=1}^{T}$. Jennrich (1969) first rigorously proved the existence of

Received July 11, 2008. Accepted November 27, 2008. * Corresponding author.

the LSE and showed the strong consistency and asymptotic normality of the LSE with the several assumptions including the following: $F_T(\theta_1, \theta_2)$ converges uniformly to a continuous function $F(\theta_1, \theta_2)$ and $F_T(\theta_1, \theta_2) = 0$ if and only if $\theta_1 = \theta_2$, where $F_T(\theta_1, \theta_2) = \frac{1}{T}\sum(f_t(\theta_1) - f_t(\theta_2))^2$. Wu (1981) gave sufficient conditions under which the LSE converges to θ_0 strongly. When the condition of the above requirement of F_T is replaced by the following assumption : $f(\theta)$ are Lipschitz function on Θ and

$$\sup_{\theta_1 \neq \theta_2} \frac{|f_t(\theta_1) - f_t(\theta_2)|}{|\theta_1 - \theta_2|} \leq M \sup_{|\theta - \theta_0| \geq \delta} |f_t(\theta) - f_t(\theta_0)|$$

for some $\delta > 0$ and for all t, where M is independent of t and $|\theta_1 - \theta_2|$ is the Euclidean distance between θ_1 and θ_2. On the other hand, in spite of the theoretical and practical merits, certain criticisms of the procedures based on the least squares method in the past have been pointed to the robustness even with a single outlier or a slight departure from the normality assumption on the errors. When the error distribution is heavy-tailed such as Laplace or Cauchy distributed errors, the least squares method is deemed inadequate, and, automatically, the least absolute deviation estimator (LAD) which is defined as follows has attracted considerable attention in the part of the robust regression analysis. Given an observation y_t any vector $\breve{\theta}_T$ in Θ minimizing the following objective function

$$S_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} |y_t - f(x_t, \theta)|$$

is called the LAD estimator of θ_0 based on $\{y_t\}_{t=1}^{T}$.

The concept of the periodicity in time series is of fundamental interest, since it provides a mean for formalizing the notions of dependence or correlation between adjacent points. In this paper, we think about a sum of sinusoidal components :

$$y_t = \sum_{r=1}^{q} \{A_{r0}\cos(\omega_{r0}t) + B_{r0}\sin(\omega_{r0}t)\} + \epsilon_t\,, \tag{1.1}$$

where $\theta_0 = (A_{10}, B_{10}, \omega_{10}, \cdots, A_{q0}, B_{q0}, \omega_{q0})$, and for $q \geq 1$, A_{r0}, B_{r0} are some fixed unknown constants, ω_{r0} is unknown frequency lying between 0 to π $(1 \leq r \leq q)$ and in this case the observed value x_t means t.

But the above formula neither satisfies Jennrich (1969)'s assumptions nor Wu (1981)'s Lipschitz type condition-the methods which are proposed by Jennrich and Wu are not available. For this reason, Walker (1971) obtained the asymptotic properties of the approximate LSE. Hannan (1973) generalized the results of the Walker. Hannan (1973) considered the case when ϵ_t is generated by a strictly stationary random variable process. Kundu (1993) and Kundu and Mitra (1996) gave a direct proof of the strong consistency and asymptotic normality and observed that the approximate LSE and the LSE are asymptotically equal. And Oberhofer (1982) studied the weak consistency about the LAD estimators with the assumptions from B1 to B6 in his paper. But the assumption B5 in his paper is equivalent to the assumption of Jennrich (1969). And then using the different aforementioned methods, the asymptotic properties of LAD of this model is proved by T. S. Kim et al. (2000). But the LSE and LAD are inadequate for asymmetric model. In this case, the asymmetric model means that the error's distribution $G(\epsilon_t)$ satisfies $G(0) \neq \frac{1}{2}$. Accordingly, we need a substitute approach. Koenker, R and Bassett, G. (1978) introduced Regression Quantile Estimator (RQE) and Whitney K. Newey and James L. Powell (1987) studied the Asymmetric Least Squares (ALS) Estimator which is defined as follows. For the nonlinear regression model, consider the next objective function, when $\beta \neq \frac{1}{2}$ and $0 < \beta < 1$, $S_T(\theta;\beta) = \frac{1}{T}\sum_{t=1}^{T} \varphi_\beta(y_t - f(x_t, \theta))$, where $\varphi_\beta(\lambda)$is called a check function which is defined by:

$$\varphi_\beta(\lambda) = \begin{cases} \beta\lambda, & \lambda \geq 0, \\ (\beta - 1)\lambda, & \lambda < 0. \end{cases}$$

For the given observation y_t, any vector $\breve{\theta}_T(\beta)$ in Θ which minimizing the objective function $S_T(\theta;\beta)$ shall be called the RQE of θ_0 based on $\{y_t\}_{t=1}^{T}$. Lastly, we consider another new objective function such as:

$$S_T(\theta;\beta,\tau) = \frac{1}{T}\sum_{t=1}^{T} \phi_\tau(y_t - f(x_t, \theta)), \tag{1.2}$$

where τ is defined by β and $\phi_\tau(\lambda)$is called a check function which is defined by:

$$\phi_\tau(\lambda) = \begin{cases} \tau\lambda^2, & \lambda \geq 0, \\ (1 - \tau)\lambda^2, & \lambda < 0, \end{cases}$$

where $0 < \tau < 1$.

For the given observation y_t, any vector $\hat{\theta}_T(\beta)$ in Θ which minimizes the objective function $S_T(\theta;\beta,\tau)$ is called the ALS estimator of θ_0 based on $\{y_t\}_{t=1}^{T}$. Firstly, we studied the asymptotic properties of the four different estimators. But in the practical phenomenon, we deal with a finite data set. So, they are invalid and not adjusted in a small sample size. Then, using the Monte Carlo simulation, we check out the validation of the above different estimators under the various error distributions.

2. The asymptotic results

Theorem 2.1. *If $\hat{\theta}_T$ is the LSE of the non-linear time series regression model (1.1) with the assumptions :*

1) *ϵ_t is distributed independently and identically with $E\{\epsilon_t\} = 0$ and $E\{\epsilon_t^2\} = \sigma^2 < \infty$,*

2) $\lim_{T\to\infty} \min_{1\le r\ne s\le q} (T|\omega_{r0} - \omega_{s0}|) = \infty$,

then the LSE is a strongly consistent estimator of θ_0 and $(P_1(\hat{\theta}_{1T}), P_2(\hat{\theta}_{2T}), \cdots, P_q(\hat{\theta}_{qT}))$ converges in law $N\Big(0_{3q\times 1}, \sigma^2\Sigma^{-1}\Big)$, where $P_r(\hat{\theta}_{rT}) = \Big(\sqrt{T}(\hat{A}_{rT} - A_{r0}), \sqrt{T}(\hat{B}_{rT} - B_{r0}), \sqrt{T^3}(\hat{\omega}_{rT} - \omega_{r0})\Big)$, $\Sigma = (\Sigma_{rs})_{3q\times 3q}$, for $r, s = 1, 2, \cdots, q$, and

$$\Sigma_{rs} = \begin{cases} 0, & \text{if } r \ne s, \\ \begin{bmatrix} \frac{1}{2} & 0 & \frac{B_{r0}}{4} \\ 0 & \frac{1}{2} & \frac{-A_{r0}}{4} \\ \frac{B_{r0}}{4} & \frac{-A_{r0}}{4} & \frac{A_{r0}^2 + B_{r0}^2}{6} \end{bmatrix}, & \text{if } r = s. \end{cases}$$

Proof. For the detailed proof, see Walker (1971) and Kundu and Mitra (1996). □

Theorem 2.2. *If $\breve{\theta}_T$ is the LAD of the non-linear time series regression model (1.1) with the same assumptions of Theorem 2.1, then the LAD estimator is a strongly consistent estimator of θ_0 and $(P_1(\breve{\theta}_{1T}), P_2(\breve{\theta}_{2T}), \cdots, P_q(\breve{\theta}_{qT}))$ converges in law $N\Big(0_{3q\times 1}, \frac{1}{\{2g(0)\}^2}\Sigma^{-1}\Big)$, where $g(\epsilon_t)$ is a continuous probability density function of ϵ_t.*

Proof. For the detailed proof, see T. S. Kim et al.(2000). □

Theorem 2.3. *If $\breve{\theta}(\beta)$ is the RQE of the non-linear time series regression model (1.1) with the same assumptions in Theorem 2.1, except $E\{\epsilon_t\} = 0$ and with the additional condition $G(0) = \beta, (0 < \beta(\ne 0.5) < 1)$, where $G(\epsilon_t)$ is a distribution function of the error terms ϵ_t, then the RQE is a strongly consistent estimator of θ_0 and $\Big(P_1(\breve{\theta}_{1T}(\beta)), P_2(\breve{\theta}_{2T}(\beta)), \cdots, P_q(\breve{\theta}_{qT}(\beta))\Big)$ converges in law $N\Big(0_{3q\times 1}, \frac{\beta(1-\beta)}{\{g(0)\}^2}\Sigma^{-1}\Big)$.*

Proof. For the detailed proof, see T. S. Kim et al. (2002). □

First of all, for the case of $q = 1$, we will consider the asymptotic properties of the ALS $\hat{\theta}_T(\beta) = \hat{\theta}_{1T}(\beta) = (\hat{A}_{1T}, \hat{B}_{1T}, \hat{\omega}_{1T}) = (\hat{A}_T, \hat{B}_T, \hat{\omega}_T)$ for $\theta_0 = (A_{10}, B_{10}, \omega_{10}) = (A_0, B_0, \omega_0)$ in a time series with stationary independent residuals model (1.1).

Theorem 2.4. *If $\hat{\theta}_T(\beta)$ is the ALS estimator of the non-linear time series regression model (1.1) with the same assumptions of Theorem 2.3 and $q = 1$, then the ALS is a strongly consistent estimator of θ_0 and $P_1(\hat{\theta}_{1T}(\beta))$ converges in law $N\left(0_{3q\times 1}, \frac{(1-2\tau)d + \tau^2(\sigma^2+\mu^2)}{\{\beta+\tau(1-2\beta)\}^2}\Sigma^{-1}\right)$, where $\Sigma = \Sigma_{11}$, $\tau = \frac{b}{2b-\mu}$, $\mu = E(\epsilon_t)$, $b = \int_{-\infty}^{0} xg(x)dx$, $d = \int_{-\infty}^{0} x^2 g(x)dx$.*

Proof. Since $S_T(\theta_0;\beta)$ defined in (1.2) is independent of $\theta = \theta_1 = (A_1, B_1, \omega_1)$, the minimizer of $S_T(\theta;\beta)$ is equivalent to the minimizer of the new objective function : $D_T(\theta;\beta) = S_T(\theta;\beta) - S_T(\theta_0;\beta)$. Then, firstly, using the Kolmogorov's strong law of large numbers, we obtain $D_T(\theta;\beta)$ which uniformly converges to $\lim_{T\to\infty} E[D_T(\theta;\beta)]$. Under the direct calculations, we know

$$\lim_{T\to\infty} \frac{\partial^2 E[D_T(\theta;\beta)]}{\partial\theta'\partial\theta} = \lim_{T\to\infty} \nabla^2 E[D_T(\theta;\beta)]$$

is a positive matrix. It should be induced that $\lim_{T\to\infty} E[D_T(\theta;\beta)]$ has a unique minimizer θ_0 in Θ. But the above two facts are sufficient conditions for the strong consistency (see White (1980), Lemma 2.2). Since $S_T(\theta;\beta,\tau)$ is the minimum when $\theta = \hat{\theta}_T$, an application of the mean value theorem gives :

$$\begin{aligned}
(S_T)_{A_0} &= (S_T)_{\bar{A}\bar{A}}(A_0 - \hat{A}_T) + (S_T)_{\bar{A}\bar{B}}(B_0 - \hat{B}_T) + (S_T)_{\bar{A}\bar{\omega}}(\omega_0 - \hat{\omega}_T),\\
(S_T)_{B_0} &= (S_T)_{\bar{A}\bar{B}}(A_0 - \hat{A}_T) + (S_T)_{\bar{B}\bar{B}}(B_0 - \hat{B}_T) + (S_T)_{\bar{B}\bar{\omega}}(\omega_0 - \hat{\omega}_T),\\
(S_T)_{\omega_0} &= (S_T)_{\bar{\omega}\bar{A}}(A_0 - \hat{A}_T) + (S_T)_{\bar{\omega}\bar{B}}(B_0 - \hat{B}_T) + (S_T)_{\bar{\omega}\bar{\omega}}(\omega_0 - \hat{\omega}_T),
\end{aligned} \tag{2.1}$$

where $(S_T)_{A_0} = \frac{\partial S_T((A,B,\omega);\beta,\tau)}{\partial A}\Big|_{(A_0,B_0,\omega_0)}$, $(S_T)_{\bar{A}\bar{B}} = \frac{\partial^2 S_T((A,B,\omega);\beta,\tau)}{\partial A\partial B}\Big|_{(A_0,B_0,\omega_0)}$, etc., and we use generic notation $(\bar{A}_T, \bar{B}_T, \bar{\omega}_T)$ for a point on the line joining $\left(A_0, B_0, \omega_0\right)$ and $\left(\hat{A}_T, \hat{B}_T, \hat{\omega}_T\right)$, so that

$$(\bar{A}_T, \bar{B}_T, \bar{\omega}_T) = \gamma(A_0, B_0, \omega_0) + (1-\gamma)(\hat{A}_T, \hat{B}_T, \hat{\omega}_T)\ (0 < \gamma < 1).$$

The point $(\bar{A}_T, \bar{B}_T, \bar{\omega}_T)$ in (2.1) will, in general, not be the same, but to distinguish them would complicate the notation, and no ambiguity will arise by not doing so. But (2.1) is replaced by the following :

$$\begin{aligned}&\left(\sqrt{T}(S_T)_{A_0}, \sqrt{T}(S_T)_{B_0}, \frac{1}{\sqrt{T}}(S_T)_{\omega_0}\right)\\ &\quad = -\left(\sqrt{T}(\hat{A}_T - A_0), \sqrt{T}(\hat{B}_T - B_0), \sqrt{T^3}(\hat{\omega}_T - \omega_0)\right) \times W_T,\end{aligned}$$

where

$$W_T = \begin{bmatrix} (S_T)_{\bar{A}\bar{A}} & (S_T)_{\bar{A}\bar{B}} & T^{-1}(S_T)_{\bar{A}\bar{\omega}} \\ (S_T)_{\bar{B}\bar{A}} & (S_T)_{\bar{B}\bar{B}} & T^{-1}(S_T)_{\bar{B}\bar{\omega}} \\ T^{-1}(S_T)_{\bar{\omega}\bar{A}} & T^{-1}(S_T)_{\bar{\omega}\bar{B}} & T^{-2}(S_T)_{\bar{\omega}\bar{\omega}} \end{bmatrix} \tag{2.2}$$

On the other hand,

$$\sqrt{T}(S_T)_{A_0} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} 2|\tau - I_{\{\epsilon_t<0\}}|\epsilon_t \cos\omega_0 t, \tag{2.3}$$

$$\sqrt{T}(S_T)_{B_0} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} 2|\tau - I_{\{\epsilon_t<0\}}|\epsilon_t \sin\omega_0 t, \tag{2.4}$$

$$\frac{1}{\sqrt{T}}(S_T)_{\omega_0} = \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T} 2|\tau - I_{\{\epsilon_t<0\}}|\epsilon_t (A_0 t \sin\omega_0 t - B_0 t \cos\omega_0 t). \tag{2.5}$$

The sum in (2.3) are of the form $\sum_{t=1}^{T} U_t$, since $E(\epsilon_t) = \mu$ and $\tau = \frac{b}{2b-\mu}$, we have $E(U_t) = 0$ and $E(U_t^2) = \frac{4}{T}\Big[(1-2\tau)d + \tau^2(\sigma^2+\mu^2)\Big]\cos^2\omega_0 t$. Let $B_T^2 = \sum_{t=1}^{T} Var(U_t^2)$, then we have $B_T^2 = 2\Big[(1-2\tau)d + \tau^2(\sigma^2+\mu^2)\Big] + o(1) < \infty$.

With the same process, we have the similar following results. In (2.4), we have $B_T^2 = \sum_{t=1}^{T} Var(U_t^2) = 2\Big[(1-2\tau)d + \tau^2(\sigma^2+\mu^2)\Big] + o(1) < \infty$ and in (2.5),

$$B_T^2 = \sum_{t=1}^{T} Var(U_t^2) = 2[(1-2\tau)d + \tau^2(\sigma^2+\mu^2)]\frac{(A_0^2+B_0^2)}{3} + o(1) < \infty.$$

For all of the above cases, for any given $\epsilon > 0$, we get

$$\lim_{T\to\infty} \frac{1}{B_T^2}\sum_{t=1}^{T} E[U_t^2 \cdot I_{\{|U_t|\geq \epsilon B_T\}}] = 0.$$

Using the Lindberg theorem, we see that $\sqrt{T}(S_T)_{A_0}$ and $\sqrt{T}(S_T)_{B_0}$ converge in law to $N\Big(0, 2[(1-2\tau)d+\tau^2(\sigma^2+\mu^2)]\Big)$ respectively, and $\frac{1}{\sqrt{T}}(S_T)_{\omega_0}$ converges in law to $N\left(0, 2\Big[(1-2\tau)d+\tau^2(\sigma^2+\mu^2)\Big]\frac{(A_0^2+B_0^2)}{6}\right)$.

For the limiting joint distribution we consider the random variable :

$$V_T(\delta_1,\delta_2,\delta_3) = \delta_1\sqrt{T}(S_T)_{A_0} + \delta_2\sqrt{T}(S_T)_{B_0} + \delta_3\frac{1}{\sqrt{T}}(S_T)_{\omega_0},$$

where the $\delta_i (i = 1, 2, 3)$ are arbitrary real numbers. Now likewise (2.3)-(2.5), $V_T(\delta_1,\delta_2,\delta_3)$ is equal to :

$$\frac{1}{T}\sum_{t=1}^{T}\left[\delta_1\sqrt{T}(-\cos\omega_0 t) + \delta_2\sqrt{T}(-\sin\omega_0 t) + \delta_3\frac{1}{\sqrt{T}}(A_0 t\sin\omega_0 t - B_0 t\cos\omega_0 t)\right]$$
$$\times 2\left|1 - I_{\{\epsilon_t<0\}}\right|.$$

Now let $V_T(\delta_1,\delta_2,\delta_3) = \sum_{t=1}^{T} U_t$, and $B_T^2 = \sum_{t=1}^{T} Var(U_t)$, then we also have $E(U_t) = 0$, and

$$B_T^2 = 4\left[(1-2\tau)d+\tau^2(\sigma^2+\mu^2)\right]\cdot\left(\frac{\delta_1^2}{2}+\frac{\delta_2^2}{2}+\frac{A_0^2+B_0^2}{6}\delta_3^2+\frac{B_0}{2}\delta_1\delta_3-\frac{A_0}{2}\delta_2\delta_3\right)$$
$$+ o(1).$$

Hence, by the Lindberg theorem applied to the above sum, we see that $V_T(\delta_1,\delta_2,\delta_3)$ converges in law to a normal distribution with zero mean and variance

$$4\left[(1-2\tau)d+\tau^2(\sigma^2+\mu^2)\right]\cdot\left(\frac{\delta_1^2}{2}+\frac{\delta_2^2}{2}+\frac{A_0^2+B_0^2}{6}\delta_3^2+\frac{B_0}{2}\delta_1\delta_3-\frac{A_0}{2}\delta_2\delta_3\right).$$

Consequently, by the virtue of the Cramer-Wold device, we see that the joint distribution of $\left(\sqrt{T}(S_T)_{A_0}, \sqrt{T}(S_T)_{B_0}, \frac{1}{\sqrt{T}}(S_T)_{\omega_0}\right)$ converges in law to

$$N\Big((0,0,0), 4\Big[(1-2\tau)d+\tau^2(\sigma^2+\mu^2)\Big]\cdot\Sigma_{11}\Big),$$

where $\lim_{T\to\infty} W_T = 2\left[\beta+\tau(1-2\beta)\right]\cdot\Sigma_{11}$.

Therefore, we have $\left(\sqrt{T}(\hat{A}_T - A_0), \sqrt{T}(\hat{B}_T - B_0), \sqrt{T^3}(\hat{\omega}_T - \omega_0)\right)$ converges in law to :

$$N\left((0,0,0), \frac{(1-2\tau)d+\tau^2(\sigma^2+\mu^2)}{\{\beta+\tau(1-2\beta)\}^2}\cdot\Sigma_{11}^{-1}\right). \quad \square$$

Suppose now that the model (1.1) is generalized to $q > 1$. The function corresponding to (1.2) whose minimization yields estimators $\hat{\theta}_T = \Big(\hat{A}_{1T}, \hat{B}_{1T}, \hat{\omega}_{1T}, \cdots, \hat{A}_{qT}, \hat{B}_{qT}, \hat{\omega}_{qT}\Big)$ became (1.2), where $\theta = \Big(A_1, B_1, \omega_1, \cdots, A_q, B_q, \omega_q\Big)$.

Theorem 2.5. *If $\hat{\theta}_T(\beta)$ is the ALS estimator of the non-linear time series regression model (1.1) with the same assumptions of Theorem 2.3, then the ALS is a strongly consistent estimator of θ_0 and $\Big(P_1(\hat{\theta}_{1T}(\beta)), P_2(\hat{\theta}_{2T}(\beta)), \cdots, P_q(\hat{\theta}_{qT}(\beta)\Big)\Big)$ converges in law*

$$N\left(0_{3q\times 1}, \frac{(1-2\tau)d + \tau^2(\sigma^2+\mu^2)}{\{\beta+\tau(1-2\beta)\}^2}\cdot \Sigma^{-1}\right).$$

Proof. Likewise, for the one harmonic case, we could obtain the strong consistency which is proved in Theorem 2.4, and we have $\lim_{T\to\infty} \nabla^2 E[D_T(\theta;\beta)]$, which is a $3q \times 3q$ positive definite matrix. This fact indicates for $r = 1, 2, \cdots, q$, $\Big(\sqrt{T}(\hat{A}_{rT} - A_{r0}), \sqrt{T}(\hat{B}_{rT} - B_{r0}), \sqrt{T^3}(\hat{\omega}_{rT} - \omega_{r0})\Big) = o_p(1)$, where $o_p(1)$ denotes convergence in probability to zero, and we can also have the fact

$$\begin{aligned}&\left(\sqrt{T}(S_{rT})_{A_0}, \sqrt{T}(S_{rT})_{B_0}, \frac{1}{\sqrt{T}}(S_{rT})_{\omega_0}\right)\\ &= -\Big(\sqrt{T}(\hat{A}_{rT} - A_0), \sqrt{T}(\hat{B}_{rT} - B_0), \sqrt{T^3}(\hat{\omega}_{rT} - \omega_0)\Big) \times W_{rT},\end{aligned}$$

where W_{rT} in (2.2), and $r = 1, 2, \cdots, q$. □

Remark. Using the four results, we obtain the asymptotic efficiency of LSE relative to LAD estimator :

$$\lim_{T\to\infty} eff(\hat{\theta}_T|\breve{\theta}_T) = \lim_{T\to\infty}\frac{Var(\hat{\theta}_T)}{Var(\breve{\theta}_T)} = \{2g(0)\}^2\sigma^2$$

is smaller than one; and under the error distribution it is standard normal, and also larger than one under the heavy-tailed distribution likewise the Laplace distribution. As well, the asymptotic efficiency of ALS relative to RQE :

$$\lim_{T\to\infty} eff\Big(\hat{\theta}_T(\beta)|\breve{\theta}_T(\beta)\Big) = \frac{g^2(0)\Big[(1-2\tau)d + \tau^2(\sigma^2+\mu^2)\Big]}{\beta(1-\beta)\Big[\beta+\tau(1-2\beta)\Big]^2}$$

is also smaller than one under the normal distribution with mean one and variance one, larger than one under the Laplace distribution with parameter $\beta\,(0 < \beta(\neq 0.5) < 1)$.

3. Monte Carlo simulations

In this section, we consider the simplest sinusoidal model with one component, i.e., for $q = 1$ in the formula (1.1). We performed some Monte Carlo simulations to compare the four different estimators. Under the four different error distributions which are (a) the standard normal, (b) the Laplace with the parameter 3, (c) the skewed Laplace ($\beta = 0.4$) with the parameter 3 and lastly (d) the skewed normal distribution with mean one and variance one, we study the behavior of the four different methods of estimation for only small sample sizes, since they were already verified for the exact relations concerning the asymptotic sense. Numerical results are reported for $T = 10, 15$ and 25 and $\omega = 0.25\pi(\simeq 0.785398)$, $0.5\pi(\simeq 1.570796)$ and $0.75\pi(\simeq 2.356194)$.

For a particular T and ω, thousands of different sets of data were generated. The two linear parameters A and B are taken as 1.5, each. Under each given data set, we estimated the nonlinear parameter ω by the four methods. The rest in this simulation runs show the same results as shown in Table 1. In the Table 4, avg. EST means the average estimates using the asymmetric least squares methods ('ALS') and the regression quantile estimating method ('RQE'), avg. MSE means the average mean squared error, avg. C.I. is the average length of the 95% confidence intervals and C.P. means the coverage probability over 1,000 simulation runs.

Table 1. When error distributions are the standard normal.

True Value	#of Sample	$T = 10$ LSE	LAD	$T = 15$ LSE	LAD	$T = 25$ LSE	LAD
	avg. EST	0.7843	0.7810	0.7852	0.7852	0.7842	0.7833
$.25\pi$	avg. MSE	0.0052	0.0098	0.0016	0.0028	0.0004	0.006
	avg. C.I.	0.1352	0.1395	0.0699	0.0715	0.0361	0.0371
	C.P.	0.901	0.842	0.932	0.887	0.932	0.854
	avg. EST	1.5693	1.5718	1.5700	1.5690	1.5705	1.5701
$.50\pi$	avg. MSE	0.0052	0.0099	0.0016	0.0028	0.0003	0.0006
	avg. C.I.	0.1284	0.1335	0.0732	0.0751	0.0345	0.0350
	C.P.	0.904	0.863	0.928	0.883	0.941	0.882
	avg. EST	2.3555	2.3583	2.3566	2.3576	2.3563	2.3586
$.75\pi$	avg. MSE	0.0051	0.0096	0.0016	0.0028	0.0003	0.0006
	avg. C.I.	0.1213	0.1260	0.0767	0.0779	0.0339	0.0350
	C.P.	0.992	0.877	0.918	0.859	0.947	0.888

Table 2. When error distributions are the Laplace with the parameter 3.

True Value	#of Sample	$T=10$ LSE	LAD	$T=15$ LSE	LAD	$T=25$ LSE	LAD
	avg. EST	0.7535	0.76749	0.7576	0.7744	0.7741	0.7742
$.25\pi$	avg. MSE	0.1041	0.0550	0.0293	0.0146	0.0062	0.0032
	avg. C.I.	0.5535	0.5773	0.3060	0.3057	0.1468	0.1495
	C.P.	0.881	0.890	0.853	0.907	0.849	0.902
	avg. EST	1.5346	1.5732	1.5505	1.5593	1.5554	1.5563
$.50\pi$	avg. MSE	0.1018	0.0549	0.0289	0.0150	0.0063	0.0032
	avg. C.I.	0.5469	0.5714	0.3051	0.3115	0.1478	0.1495
	C.P.	0.848	0.925	0.878	0.912	0.859	0.883
	avg. EST	2.3507	2.3494	2.3322	2.3499	2.3483	2.3519
$.75\pi$	avg. MSE	0.1047	0.0522	0.0296	0.0146	0.0063	0.0031
	avg. C.I.	0.5507	0.5550	0.3080	0.3087	0.1465	0.1455
	C.P.	0.847	0.910	0.872	0.922	0.897	0.937

Table 3. The skewed Laplace ($\beta = 0.4$) with the parameter 3.

True Value	#of Sample	$T=10$ LSE	LAD	$T=15$ LSE	LAD	$T=25$ LSE	LAD
	avg. EST	0.7699	0.7792	0.7666	0.7718	0.7117	0.7801
$.25\pi$	avg. MSE	0.1163	0.1045	0.0327	0.0285	0.0073	0.0055
	avg. C.I.	0.5494	0.5619	0.3045	0.3097	0.1519	0.1513
	C.P.	0.860	0.893	0.784	0.916	0.566	0.915
	avg. EST	1.5748	1.5672	1.5580	1.5675	1.5315	1.5636
$.50\pi$	avg. MSE	0.1216	0.1071	0.0335	0.0286	0.0071	0.0055
	avg. C.I.	0.5637	0.5745	0.3115	0.3155	0.1499	0.1487
	C.P.	0.841	0.896	0.803	0.920	0.601	0.938
	avg. EST	2.3621	2.3557	2.3432	2.3531	2.3324	2.3557
$.75\pi$	avg. MSE	0.1190	0.1046	0.0329	0.0294	0.0071	0.0058
	avg. C.I.	0.5528	0.5706	0.3084	0.3151	0.1491	0.1483
	C.P.	0.812	0.892	0.836	0.928	0.633	0.934

4. Conclusions

The above tables show all four cases; as T increases the absolute difference between true parameter values and estimates, the average mean squared errors and the average length of confidence interval decrease and the coverage probabilities monotonically decrease to the asymptotic value in the case. And we have the results that when the error terms are the case (a), the LSE usually performs better than the other methods as far as the estimation of ω is concerned. Also, we know for the case (b), the LAD and for the (c), the RQE is superior to the other methods. From the asymptotic theory and Monte Carlo simulation runs, we conclude that to estimate the true parameters in the real phenomenon, we have to determine the error distribution using the general statistical analysis and then choose the suitable estimating method.

Table 4. The skewed normal with mean one and variance one.

True Value	#of Sample	$T=10$ LSE	$T=10$ LAD	$T=15$ LSE	$T=15$ LAD	$T=25$ LSE	$T=25$ LAD
	avg. EST	0.7844	0.7854	0.7854	0.7819	0.7857	0.7857
$.25\pi$	avg. MSE	0.0090	0.0135	0.0025	0.0036	0.0005	0.0007
	avg. C.I.	0.1583	0.1631	0.0790	0.0815	0.0411	0.0414
	C.P.	0.933	0.897	0.938	0.900	0.950	0.885
	avg. EST	1.5694	1.5662	1.5712	1.5703	1.5708	1.5704
$.50\pi$	avg. MSE	0.0091	0.0135	0.0025	0.0037	0.0005	0.0008
	avg. C.I.	0.1509	0.1549	0.0832	0.0848	0.0395	0.0400
	C.P.	0.948	0.905	0.939	0.893	0.951	0.892
	avg. EST	2.3553	2.3551	2.3566	2.3551	2.3562	2.3563
$.75\pi$	avg. MSE	0.0091	0.0135	0.0025	0.0036	0.0005	0.0008
	avg. C.I.	0.1447	0.1503	0.0857	0.0870	0.0384	0.0390
	C.P.	0.946	0.908	0.931	0.870	0.960	0.898

References

1. Debasis Kundu, *Asymptotic theory of least squares estimator of a particular nonlinear regression model*, Statistics & Probability Letters, 18(1993), 13-17.
2. Debasis Kundu, Amit Mitra, *Asymptotic theory of least squares estimator of a nonlinear time series regression model.* Commun. Statist.-Theory Math. , 23 (1)(1996), 133-141.
3. Hannan, E. J., *The estimation of Frequency.* Journal of Applied Probability 10(1973), 510-519.
4. Jennrich, R., *Asymptotic properties of nonlinear least squares estimations.* Annals of Mathematical Statistics, 40(1969), 633-643.
5. Koenker, R. and Bassett, G. , *Asymptotic theory of least absolute error regression.* Journal of the American Statistical Association 73(1978), 618-621.
6. Oberhofer, W., *The consistency of nonlinear regression minimizing the L1-norm.* The Annals of Statistics 10(1982), 316-319.
7. T. S. Kim, H. K. Kim and S. H. Choi, *Asymptotic Properties of LAD Estimators of a Nonlinear Time Series Regression Model.* Journal of the Korean Statistical Society, vol29, no2(2000), 187-199.
8. T. S. Kim, H. K. Kim and S. Hur, *Asymptotic properties of nonlinear regression quantile estimation.* Statistics & Probability Letters, 60(2002), 387-394.

9. Walkerr, A. M., *On the estimation of a harmonic component in a time series with stationary independent residuals.* Biometrika, 58(1971), 21-36.
10. Whitney K. Newey and James L. Powell, *Asymmetric least squares estimation and testing.* Econometrica, Vol. 53, No. 4(1987), 819-847.
11. Wu, C. F., *Asymptotic theory of nonlinear least squares estimation.* The Annals of Statistics, 9(1981), 501-513.

Tae Soo Kim received his Ph.D in mathematics at YonSei University. Since 2004 he has been at the Seoul National University of Technology(SNUT). His research interests center on the mathematical statistics and simulations.

School of Liberal Arts(Dept. Math), Seoul National University of Technology, Seoul, 139-743, South Korea
e-mail: tskim@snut.ac.kr

Jung Ho Ahn received his Ph.D in computer science department at Yonsei University. Since 2007 he has been at Kangnam University. His reserch interests are applied statistics, pattern analysis and computer vision.

School of Computer & Media Engineering, Kangnam University, Yongin, 446-702, South Korea.
email : jungho@kangnam.ac.kr