Asymptotic efficiency of estimating function estimators for nonlinear time series models

TOMOYUKI AMANO

Abstract

The conditional least squares (CLS) estimator proposed by Tjϕstheim (1986) is convenient and important for nonlinear time series models. However this convenient estimator is not generally asymptotically efficient. Hence Chandra and Taniguchi (2001) proposed a G estimator based on Godambe's asymptotically optimal estimating function. For important nonlinear time series models, e.g., RCA, GARCH, nonlinear AR models, we show the asymptotic variance of the G estimator is smaller than that of the CLS estimator and the G estimator is asymptotically efficient if the innovation is Gaussian. Numerical studies for the comparison of the asymptotic variance of the G estimator, that of the CLS estimator and the Fisher information are also given. They elucidate some interesting features of the G estimator.

Some Key words: asymptotic efficiency; conditional least squares estimator; estimating function; GARCH model; local asymptotic normality; nonlinear AR model; RCA model.

1. Introduction

In finance, biology and natural science, there are many cases which linear models can not fully fit data and wide classes of nonlinear time series models have been proposed. Random coefficient autoregressive (RCA) model is one of such models and introduced to describe occasional sharp spikes, exhibited in many fields such as the engineering, econometrics and biology. This important model is represented as autoregressive coefficients of AR model are random and discussed by many authors (see Nicholls and Quinn (1982) and Feigin and Tweedie (1985), etc).

Autoregressive conditional heteroskedastic (ARCH) model is the most popular time series model in econometrics, which is proposed by Engle (1982). In econometrics, there are many cases a one-period forecast variance is not constant and traditional time series models could not overcome this difficulty. To overcome this, ARCH model was introduced. Bollerslev (1986) generalized ARCH model, which is defined as generalized autoregressive conditional heteroskedastic (GARCH) model. This model is extremely used in econometric time series.

One of the most important and fundamental estimators for nonlinear time series models is the conditional least squares (CLS) estimator which is introduced by Tjϕstheim (1986). This estimator has an advantage, which may be represented as a simple linear form. However it is not generally asymptotically efficient. Amano and Taniguchi

(2008) applied this convenient estimator to ARCH model and derived a condition for its asymptotic efficiency. However this condition is strict. On the other hand Chandra and Taniguchi (2001) constructed G estimators for RCA and ARCH models based on Godambe's asymptotically optimal estimating function. They showed these asymptotic normality and the G estimators are better than the CLS estimators by simulations. In this paper, we applied this estimator to some important nonlinear time series models, e.g., RCA, GARCH, nonlinear AR models and show the G estimators are better than the CLS estimators in the sense of the magnitude of the asymptotic variances. Furthermore we show the G estimators are asymptotically efficient, if the innovations are Gaussian. Numerical studies give interesting features of the G estimators.

This paper is organized as follows. Section 2 summarizes the definitions of G and CLS estimators. In Section 3, we give the asymptotic results of these estimators for RCA model and show the asymptotic variance of the G estimator is smaller than that of CLS estimator and if the innovation is Gaussian the G estimator becomes asymptotically efficient. In Section 4, when the model is GARCH model, we prove G estimator is better than the CLS estimator in the sense of the magnitude of the asymptotic variances and under a condition that the innovation is Gaussian the G estimator is asymptotically efficient. In Section 5, when the model follows nonlinear AR model, it is shown that the CLS estimator coincides with the G estimator and if the innovation is Gaussian, G and CLS estimators are asymptotically efficient. Section 6 provides numerical studies, which show how the G estimator is good. Proofs of theorems are relegated to Section 7.

2. CLS and G estimators for nonlinear time series models

One of the most fundamental estimators for the parameters of nonlinear time series model $\{X_t\}$ is the conditional least squares (CLS) estimator $\hat{\theta}_n^{(CL)}$ introduced by Tjϕstheim (1986). It is obtained by minimizing the penalty function

$$Q_n(\theta) \equiv \sum_{t=k+1}^{n} [X_t - E[X_t|\mathbf{F}_t(k)]]^2 \tag{2·1}$$

where $\mathbf{F}_t(k)$ is the σ-algebra generated by $\{X_s : t-k \leq s \leq t-1\}$ and k is an appropriate positive integer (e.g. if $\{X_t\}$ follows a m-th order nonlinear autoregressive model, we can take $k = m$). The CLS estimator has the simple expression generally. However, it is not asymptotically efficient in general. Hence Chandra and Taniguchi (2001) constructed an estimator $\hat{\theta}_n^{(G)}$ based on Godambe's asymptotically optimal estimating function for nonlinear time series models. For the definition of $\hat{\theta}_n^{(G)}$, we prepare the following estimating function $G(\theta)$. Let $\{X_t\}$ be a stochastic process which is depending on the k-dimensional parameter θ_0, then $G(\theta)$ is

$$G(\theta) = \sum_{t=1}^{n} \mathbf{a}_{t-1} h_t \tag{2·2}$$

where $\mathbf{a}_{t-1}$ is a k-dimensional vector depending on $X_1, \cdots, X_{t-1}$ and θ, $h_t = X_t - E[X_t|\mathbf{F}_{t-1}]$ and $\mathbf{F}_{t-1}$ is the σ-field generated by $\{X_s, s \leq t-1\}$. An estimating function estimator $\hat{\theta}_n^{(E)}$

for the parameter θ_0 is defined as $G(\hat{\theta}_n^{(E)}) = 0$. Chandra and Taniguchi (2001) derived the asymptotic variance of $\sqrt{n}(\hat{\theta}_n^{(E)} - \theta_0)$ is

$$\left[\frac{1}{n}E\frac{\partial}{\partial\theta}G(\theta_0)\right]^{-1}\frac{E\{G(\theta_0)G'(\theta_0)\}}{n}\left(\left[\frac{1}{n}E\frac{\partial}{\partial\theta}G(\theta_0)\right]^{-1}\right)' \tag{2·3}$$

and gived the following lemma by extending the result of Godambe (1985).

Lemma 2.1. *The asymptotic variance (2·3) is minimized if* $G(\theta) = G^*(\theta)$ *where*

$$G^*(\theta) = \sum_{t=1}^{n} \mathbf{a}_{t-1}^* h_t, \tag{2·4}$$

$$\mathbf{a}_{t-1}^* = E\left[\left.\frac{\partial h_t}{\partial\theta}\right|\mathbf{F}_{t-1}\right]E\left[h_t^2|\mathbf{F}_{t-1}\right]^{-1}. \tag{2·5}$$

3. Efficiency of the estimators for RCA model

In this section, we apply CLS and G estimators to Random coefficient autoregressive (RCA) model and discuss the efficiency of CLS and G estimators for the parameter θ_0 of RCA model. Then we prove the G estimator is more efficient than the CLS estimator and under $\{\epsilon_t\}$ and $\{\mathbf{z}_t\}$ are Gaussian, it is asymptotically efficient based on the local asymptotic normality (LAN).

RCA model of order k is

$$X_t = \sum_{j=1}^{k}(\theta_{0,j} + z_t(j))X_{t-j} + \epsilon_t \tag{3·1}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d random variables with mean 0, variance σ^2, spectral density $f(\cdot)$ and $\mathbf{z}_t = (z_t(1), \cdots, z_t(k))'$ is a sequence of i.i.d random vectors with mean vector 0 and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_k^2 \end{pmatrix} \tag{3·2}$$

and we assume $\{\epsilon_t\}, \{z_t(1)\}, \cdots, \{z_t(k)\}$ are independent.

If we apply the CLS estimator to RCA model, the penalty function is

$$Q_n(\theta) = \sum_{t=k+1}^{n}[X_t - \theta'\mathbf{X}_{t-1}]^2 \tag{3·3}$$

where $\theta \equiv (\theta_1, \cdots, \theta_k)'$, $\mathbf{X}_{t-1} = (X_{t-1}, \cdots, X_{t-k})'$ and due to the usual linear regression theory, this estimator has the following representaion

$$\hat{\theta}_n^{(CL)} = \left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right)^{-1}\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}X_t\right). \tag{3·4}$$

We impose the following assumptions to describe the asymptotics of $\hat{\theta}_n^{(CL)}$.

Assumption 3.1. *(i) Let $\lambda(\mathbf{A})$ be the maximum eigenvalue of a matrix $\mathbf{A}$ in modulus, then we suppose the parameter $\theta_0 = (\theta_{0,1}, \cdots, \theta_{0,k})'$ and the covariance matrix $\mathbf{\Sigma}$ satisfy*

$$\lambda(\theta_0\theta_0' + \mathbf{\Sigma}) < 1 \tag{3·5}$$

(ii)

$$E[X_t^2] < \infty \tag{3·6}$$

Assumption (i) implies $\{X_t\}$ to be strict stationary and ergotic (see Feigin and Tweedie (1985)). From Tjϕstheim (1986), the following lemma holds.

Lemma 3.1. *Under Assumption 3.1,*

$$\sqrt{n}(\hat{\theta}_n^{(CL)} - \theta_0) \xrightarrow{d} N(0, \mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}) \tag{3·7}$$

where

$$\mathbf{U} = E\left[\mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right] \tag{3·8}$$

$$\mathbf{W} = E\left[(X_t - \theta_0'\mathbf{X}_{t-1})^2\mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right] = E\left[var(X_t|\mathbf{X}_{t-1})\mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right]. \tag{3·9}$$

The CLS estimator has a simple and explicit form (3·4) and not require the knowledge of the distribution of $\{\mathbf{z}_t\}$ and $\{\epsilon_t\}$.

Based on the estimating function $G^*(\theta)$ in Lemma 2.1 and the observations $\{X_1, \cdots, X_n\}$, Chandra and Taniguchi (2001) derived a Godambe's asymptotically optimal estimating function estimator $\hat{\theta}_n^{(G)}$ for the parameter of θ_0 of RCA model

$$\hat{\theta}_n^{(G)} = \left(\sum_{t=k+1}^{n} \frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'}{\phi_t}\right)^{-1} \left(\sum_{t=k+1}^{n} \frac{\mathbf{X}_{t-1}X_t}{\phi_t}\right) \tag{3·10}$$

where $\phi_t = E[h_t^2|\mathbf{F}_{t-1}] = \sigma^2 + \mathbf{X}_{t-1}'\mathbf{\Sigma}\mathbf{X}_{t-1}$ and gived the following lemma.

Lemma 3.2. *Under Assumption 3.1, if $\mathbf{V} \equiv E\left[\frac{\mathbf{X}_{t-1}\mathbf{X}_{t-1}'}{var(X_t|\mathbf{X}_{t-1})}\right]$ is a positive definite matrix with bounded elements, then $\hat{\theta}_n^{(G)}$ for the parameter θ_0 of RCA model satisfies*

$$\sqrt{n}(\hat{\theta}_n^{(G)} - \theta_0) \xrightarrow{d} N(0, \mathbf{V}^{-1}). \tag{3·11}$$

Next, we state the main theorem to compare efficiencies of the G estimator and the CLS estimator, whose proof is given in section **??**

Theorem 3.1. *Suppose that Assumption 3.1 holds. Then the following statements hold true.*

(i) The asymptotic variances of $\hat{\theta}_n^{(CL)}$ and $\hat{\theta}_n^{(G)}$ satisfy the following inequality

$$\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1} \geq \mathbf{V}^{-1}. \tag{3·12}$$

(ii) The equality holds if and only if

$$var(X_t|\mathbf{X}_{t-1}) = c \quad a.s \tag{3·13}$$

for some constant c.

From this theorem it is implied that the G estimator is more efficient than the CLS estimator and a condition for the efficiency of $\hat{\theta}_n^{(CL)}$ to equal that of $\hat{\theta}_n^{(G)}$ is strict. Next we give a condition $\hat{\theta}_n^{(G)}$ is asymptotically efficient based on LAN. The following lemma is given by Hwang and Basawa (1993).

Lemma 3.3. *Under Assumption 3.1, RCA model has LAN with the central sequence*

$$\Delta_n \equiv \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \dot{\phi}_t(\theta_0) \tag{3·14}$$

and the Fisher information matrix

$$\mathbf{\Gamma} = \mathbf{\Gamma}(\theta_0) = 4E\left[\dot{\phi}_t(\theta_0)\dot{\phi}_t(\theta_0)'\right] \tag{3·15}$$

where $g_\theta(X_t|\mathbf{X}_{t-1})$ is the conditional density of X_t given $\mathbf{X}_{t-1}$ under the parameter θ, $\phi_t(\theta^*, \theta_0) = \left[\frac{g_{\theta^*}(X_t|\mathbf{X}_{t-1})}{g_{\theta_0}(X_t|\mathbf{X}_{t-1})}\right]^{\frac{1}{2}}$ and $\dot{\phi}_t(\theta_0) = \frac{\partial}{\partial\theta^*}\phi_t(\theta^*, \theta_0)\Big|_{\theta^*=\theta_0}$.

Let $\{Y_t\}$ be a stochastic process which depends on the p-dimensional parameter γ and define $P_{\gamma,n}$ as the distribution of $(Y_1, Y_2, \cdots, Y_n)$. We define an estimator $\{\mathbf{T}_n\}$ for the parameter γ of $\{Y_t\}$ is regular if the following condition holds.

$$\sqrt{n}(\mathbf{T}_n - \gamma_n) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}^{-1}(\gamma)) \qquad \text{under } P_{\gamma_n,n} \tag{3·16}$$

where $\gamma_n = \gamma + \frac{\mathbf{h}}{\sqrt{n}}$, $\mathbf{h}$ is an any p dimensional constant vector and a matrix $\mathbf{\Sigma}^{-1}(\gamma)$ depends on γ. If $\{P_{\gamma,n}\}$ has LAN and its Fisher information matrix is $\mathbf{\Gamma}(\gamma)$, for any regular estimator, $\mathbf{\Sigma}^{-1}(\gamma)$ satisfies (see Hall and Mathiason (1990)).

$$\mathbf{\Sigma}^{-1}(\gamma) \geq \mathbf{\Gamma}^{-1}(\gamma). \tag{3·17}$$

Hence if $\mathbf{\Sigma}^{-1}(\gamma)$ corresponds with $\mathbf{\Gamma}^{-1}(\gamma)$, we say that the regular estimator is asymptotically efficient. The following theorem gives the condition that $\hat{\theta}_n^{(G)}$ is asymptotically efficient, that is $\mathbf{V}^{-1} = \mathbf{\Gamma}^{-1}(\theta_0)$.

Theorem 3.2. *Under Assumption 3.1, the following statements are satisfied.*

(i) The asymptotic variance of $\hat{\theta}_n^{(G)}$ satisfy

$$\mathbf{V}^{-1} \geq \mathbf{\Gamma}^{-1}. \tag{3·18}$$

(ii) If $\{\epsilon_t\}$ and $\{\mathbf{z}_t\}$ are Gaussian, then $\hat{\theta}_n^{(G)}$ is asymptotically efficient, that is

$$\mathbf{V}^{-1} = \mathbf{\Gamma}^{-1}. \tag{3·19}$$

Theorem 3.1 implies, the condition for the asymptotic variance of $\hat{\theta}_n^{(CL)}$ to equal that of $\hat{\theta}_n^{(G)}$ is severe. Hence we consider the efficiency of $\hat{\theta}_n^{(CL)}$ relative to $\hat{\theta}_n^{(G)}$. For convenience we assume $\theta_0 = \mathbf{0}$. If we define $R(0) = var(X_t)$, $\mathbf{U} = R(0)\mathbf{I}_k$ where $\mathbf{I}_k$ is the identity matrix of order k. Hence the asymptotic variance of $\hat{\theta}_n^{(CL)}$ is

$$\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1} = \frac{1}{R^2(0)}E[\phi_t \mathbf{X}_{t-1}\mathbf{X}'_{t-1}] \tag{3·20}$$

$$\sim \frac{1}{(n-k)}\frac{1}{R^2(0)}\sum_{t=k+1}^{n}\phi_t \mathbf{X}_{t-1}\mathbf{X}'_{t-1} \tag{3·21}$$

$$= \frac{1}{R(0)}\mathbf{X}'\mathbf{\Phi}\mathbf{X}, \tag{3·22}$$

where

$$\mathbf{X} = \frac{1}{\sqrt{n-k}\sqrt{R(0)}}\begin{pmatrix} X_{n-1} & \cdots & X_{n-k} \\ \vdots & \ddots & \vdots \\ X_k & \cdots & X_1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Phi} = \begin{pmatrix} \phi_n & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \phi_{k+1} \end{pmatrix}. \tag{3·23}$$

Similarly, the asymptotic variance of $\hat{\theta}_n^{(G)}$ is approximated by

$$\mathbf{V}^{-1} \sim \frac{1}{R(0)}\left(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}\right)^{-1}. \tag{3·24}$$

If we define a measure of efficiency of $\hat{\theta}_n^{(CL)}$ relative to $\hat{\theta}_n^{(G)}$ as $\frac{|\mathbf{V}^{-1}|}{|\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}|}$, then from the above discussion this efficiency is approximated by

$$\frac{|\mathbf{V}^{-1}|}{|\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}|} \sim (|\mathbf{X}'\mathbf{\Phi}\mathbf{X}||\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|)^{-1} \tag{3·25}$$

where $|\cdot|$ is the determinant of the matrix. Since $\mathbf{X}'\mathbf{X} \sim \mathbf{I}_k$, the right hand side of (3·25) has the lower bound (Bloomfield and Watson (1975))

$$(|\mathbf{X}'\mathbf{\Phi}\mathbf{X}||\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|)^{-1} \geq \Pi_{l=1}^{k}\frac{4\tilde{\phi}_l\tilde{\phi}_{n-k-l+1}}{(\tilde{\phi}_l + \tilde{\phi}_{n-k-l+1})^2}. \tag{3·26}$$

where $(\tilde{\phi}_1, \cdots, \tilde{\phi}_{n-k})$ is the ordered sequence of $(\phi_{k+1}, \cdots, \phi_n)$.

4. Asymptotics and efficiency of GARCH models

In this section, CLS and G estimators are applied to GARCH model which is defined as,

$$\begin{cases} X_t = \epsilon_t\sqrt{U_t} \\ U_t = a_0 + \sum_{j=1}^{q} a_j X_{t-j}^2 + \sum_{i=1}^{p} b_i U_{t-i} \end{cases} \tag{4·1}$$

where $a_0 > 0$, $a_j \geq 0$, $j = 1, \cdots, q$, $b_i \geq 0$, $i = 1, \cdots, p$, and $\{\epsilon_t\}$ is a sequence of i.i.d random variables with mean 0, variance 1, fourth-order cumulant κ_4 and denstity $g(\cdot)$. We impose the following assumption for this model.

Assumption 4.1.

$$\sum_{j=1}^{q} a_j + \sum_{i=1}^{p} b_i < 1 \tag{4·2}$$

This assumption implies GARCH(p,q) model is strict stationary, ergotic and has second order moments (see Ling and Li (1997)).

We estimate the parameter $\theta_0 = (a_0, a_1, \cdots, a_q, b_1, \cdots, b_p)'$. For construction of the estimators, we let $Y_t \equiv X_t^2$ and $m = \max(p, q)$. Then the CLS estimator for the parameter of the squared strech of GARCH(p,q) model is obtained by minimizing

$$Q_n(\theta) = \sum_{t=m+1}^{n} [Y_t - E[Y_t|\mathbf{F}_t(m)]]^2 \tag{4·3}$$

$$= \sum_{t=m+1}^{n} [Y_t - U_t]^2 \tag{4·4}$$

and it may be written as the followng explicit representation

$$\hat{\theta}_n^{(CL)} = \left(\sum_{t=m+1}^{n} \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' \right)^{-1} \left(\sum_{t=m+1}^{n} \mathbf{Y}_{t-1} Y_t \right) \tag{4·5}$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, \cdots, Y_{t-q}, U_{t-1}, \cdots, U_{t-p})'$. For the asymptotic normality of this estimator, the following assumption is imposed.

Assumption 4.2.

$$E[X_t^4] < \infty \tag{4·6}$$

Due to Tjϕstheim (1986), the following lemma holds.

Lemma 4.1. *Under Assumptions 4.1 and 4.2,*

$$\sqrt{n}(\hat{\theta}_n^{(CL)} - \theta_0) \xrightarrow{d} N(0, \mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}) \tag{4·7}$$

where

$$\mathbf{U} = E[\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'], \tag{4·8}$$

$$\mathbf{W} = (\kappa_4 + 2)E\left[U_t^2 \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\right], \tag{4·9}$$

$$\mathbf{Y}_{t-1} = (1, Y_{t-1}, \cdots, Y_{t-q}, U_{t-1}, \cdots, U_{t-p})'. \tag{4·10}$$

Next, we define the G estimator for the parameter of GARCH(p,q) model. The martingale difference h_t of Y_t becomes

$$h_t = Y_t - E[Y_t|\mathbf{F}_{t-1}] = Y_t - \theta'\mathbf{Y}_{t-1} = U_t(\epsilon_t^2 - 1) \tag{4·11}$$

and from this representation, we can obtain

$$E[h_t^2|\mathbf{F}_{t-1}] = (\kappa_4 + 2)U_t^2. \tag{4·12}$$

If we differentiate h_t with respect to θ, it is seen that

$$\frac{\partial h_t}{\partial \theta} = -\mathbf{Y}_{t-1} \tag{4·13}$$

and its conditional expectation under $\mathbf{F}_{t-1}$ is

$$E\left[\frac{\partial h_t}{\partial \theta}\middle| \mathbf{F}_{t-1}\right] = -\mathbf{Y}_{t-1}. \tag{4·14}$$

Hence from Lemma 2.1, the G estimator for the parameter θ_0 is

$$\hat{\theta}_n^{(G)} = \left(\sum_{t=m+1}^{n} \frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'}{U_t^2}\right)^{-1}\left(\sum_{t=m+1}^{n} \frac{\mathbf{Y}_{t-1}Y_t}{U_t^2}\right). \tag{4·15}$$

To evaluate the efficiency of CLS and G estimators, the following lemma is given.

Lemma 4.2. *Under Assumptions 4.1 and 4.2, if* $\mathbf{V} \equiv \frac{1}{(\kappa_4+2)}E\left[\frac{\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'}{U_t^2}\right]$ *is a positive definite matrix with bounded elements, then*

$$\hat{\theta}_n^{(G)} \xrightarrow{a.s} \theta_0 \tag{4·16}$$

and

$$\sqrt{n}(\hat{\theta}_n^{(G)} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{-1}). \tag{4·17}$$

The proof of this lemma is relegated to Section 7. Now we state the main theorem for the comparison of the efficiencies of CLS and G estimators.

Theorem 4.1. *Under Assumptions 4.1 and 4.2, it holds that*

$$\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1} \geq \mathbf{V}^{-1} \tag{4·18}$$

and the necessary and sufficient condition for the equality is there exists some constant c such that

$$a_0 + \sum_{j=1}^{q} a_j X_{t-j}^2 + \sum_{i=1}^{p} b_i U_{t-i} = c \quad a.s. \tag{4·19}$$

The condition that the asymptotic variances of CLS and G estimators coincide with is severe. Thus we discuss the asymptotic efficiency of the G estimator based on LAN. We impose the following assumptions.

Assumption 4.3. *The polynomials* $(1 - \sum_{i=1}^{p} b_i z^i)$ *and* $\sum_{j=1}^{q} a_j z^{j-1}$ *have no common roots.*

Assumption 4.4. *The innovation density $g(\cdot)$ is symmetric, twice continuously differentiable, and satisfies*

(i)

$$0 < \int \left\{ \frac{\dot{g}(u)}{g(u)} \right\}^2 g(u)du < \infty \tag{4·20}$$

$$\int \left\{ \frac{\dot{g}(u)}{g(u)} \right\}^4 g(u)du < \infty \tag{4·21}$$

(ii)

$$\lim_{|u|\to\infty} ug(u) = 0 \qquad \lim_{|u|\to\infty} u^2\dot{g}(u) = 0 \tag{4·22}$$

Lee and Taniguchi (2005) showed the following results.

Lemma 4.3. *Under Assumptions 4.1, 4.2, 4.3 and 4.4, GARCH(p,q) model has LAN with the central sequence*

$$\Delta_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ -\frac{1}{2U_t} \left(\frac{\dot{g}(u_t)}{g(u_t)} u_t + 1 \right) \mathbf{Y}_{t-1} \right\} \tag{4·23}$$

and the Fisher information matrix

$$\mathbf{\Gamma} = \mathbf{\Gamma}(\theta_0) = \frac{1}{4} E \left[\left(\frac{\dot{g}(u_t)}{g(u_t)} u_t + 1 \right)^2 \frac{\mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}}{U_t^2} \right]. \tag{4·24}$$

From the Lemma 4.3, the asymptotic variance of the G estimator $\mathbf{V}^{-1}$ satisfies

$$\mathbf{V}^{-1} \geq \mathbf{\Gamma}^{-1} \tag{4·25}$$

and for this inequality we obtain the following theorem.

Theorem 4.2. *The G estimator is asymptotically efficient, that is*

$$\mathbf{V}^{-1} = \mathbf{\Gamma}^{-1} \tag{4·26}$$

if and only if $\{u_t\}$ is Gaussian.

5. Nonlinear AR models

Nonlinear AR model is very important model, which is motivated by dynamical systems directly. In this section, we apply CLS and G estimators to this model which has the representation

$$X_t = F_{\theta_0}(X_{t-1}, \cdots, X_{t-k}) + u_t \tag{5·1}$$

where θ_0 is a p-dimensional parameter, $F_{\theta_0} : R^k \to R$ is a measurable function which depends on θ_0, $\{u_t\}$ is a sequence of i.i.d random variables with mean 0, variance 1 and density f. This model includes SETAR model

$$X_t = \sum_{j=1}^{q}(a_{j0} + \sum_{i=1}^{k} a_{ji}X_{t-i})\chi(X_{t-d} \in \mathbf{I}_j) + u_t \qquad (5\cdot2)$$

where $\mathbf{I}_1 = (-\infty, r_1)$, $\mathbf{I}_2 = [r_1, r_2), \cdots, \mathbf{I}_j = [r_{j-1}, \infty)$, d is a positive integer and $\chi(\cdot)$ is the indicator function. Another model included in nonlinear AR model is EXPAR model

$$X_t = \{a_1 + b_1 \exp(-cX_{t-d}^2)\}X_{t-1} + \cdots + \{a_k + b_k \exp(-cX_{t-d}^2)\}X_{t-k} + u_t \qquad (5\cdot3)$$

where $c \geq 0$, $d \in \mathbf{N}$, a_j, b_j are real constants. To construct the CLS estimator for the parameter θ_0, let $\mathbf{X}_t = (X_{t-1}, \cdots, X_{t-k})'$. Then the conditional expectation of X_t under $\mathbf{F}_t(k)$ and the parameter θ is

$$E[X_t|\mathbf{F}_t(k)] = F_\theta(\mathbf{X}_t') \qquad (5\cdot4)$$

and the penalty function is

$$Q_n(\theta) = \sum_{t=k+1}^{n} [X_t - F_\theta(\mathbf{X}_t')]^2. \qquad (5\cdot5)$$

The CLS estimator $\hat{\theta}_n^{(CL)}$ is given by minimizing the penalty function $Q_n(\theta)$, that is which satisfies

$$\frac{\partial Q_n(\theta)}{\partial \theta} = 0. \qquad (5\cdot6)$$

From (5·6), it can be seen that $\hat{\theta}_n^{(CL)}$ is obtained by solving the followin equality

$$\sum_{t=k+1}^{n} X_t \frac{\partial F_\theta(\mathbf{X}_t')}{\partial \theta} = \sum_{t=k+1}^{n} F_\theta(\mathbf{X}_t') \frac{\partial F_\theta(\mathbf{X}_t')}{\partial \theta}. \qquad (5\cdot7)$$

Next we construct the G estimator $\hat{\theta}_n^{(G)}$. For nonlinear AR model, the martingale difference h_t is given by

$$h_t = X_t - E[X_t|\mathbf{F}_{t-1}] = X_t - F_\theta(\mathbf{X}_t') \qquad (5\cdot8)$$

and

$$E[h_t^2|\mathbf{F}_{t-1}] = E[u_t^2] = 1. \qquad (5\cdot9)$$

Differential h_t with respect to θ becomes

$$\frac{\partial h_t}{\partial \theta} = -\frac{\partial F_\theta(\mathbf{X}_t')}{\partial \theta} \qquad (5\cdot10)$$

and its conditional expectation is

$$E\left[\frac{\partial h_t}{\partial \theta}\bigg|\mathbf{F}_{t-1}\right] = -\frac{\partial F_\theta(\mathbf{X}'_t)}{\partial \theta}. \tag{5·11}$$

Hence from Lemma 2.1, the G estimator satisfies the following equation

$$\sum_{t=k+1}^{n} \frac{\partial F_\theta(\mathbf{X}'_t)}{\partial \theta}(X_t - F_\theta(\mathbf{X}'_t)) = 0. \tag{5·12}$$

Thus $\hat{\theta}_n^{(G)}$ corresponds with $\hat{\theta}_n^{(CL)}$. In order to obtain the asymptotics of G and CLS estimators, we impose the following assumption.

Assumption 5.1. *(i) There exist a positive number $\lambda < 1$ and a constant c such that*

$$|F_{\theta_0}(x_1, \cdots, x_k)| \le \lambda \max(|x_1|, \cdots, |x_k|) + c \tag{5·13}$$

(ii)

$$E_\theta |F_\theta(X_{t-1}, \cdots, X_{t-k})|^2 < \infty \tag{5·14}$$

(iii) $F_\theta(\mathbf{X}'_t)$ is almost surely three times continuously differentiable.

(iv) For $j, k = 1, \cdots, p$

$$E\left[\left|\frac{\partial}{\partial \theta_j} F_{\theta_0}(\mathbf{X}'_t)\right|^2\right] < \infty \quad \textit{and} \quad E\left[\left|\frac{\partial^2}{\partial \theta_j \partial \theta_k} F_{\theta_0}(\mathbf{X}'_t)\right|^2\right] < \infty \tag{5·15}$$

(v) If $c_1, \cdots, c_p$ are arbitrary real numbers such that

$$E\left[\left|\sum_{j=1}^{p} c_j \frac{\partial}{\partial \theta_j} F_{\theta_0}(\mathbf{X}'_t)\right|^2\right] = 0, \tag{5·16}$$

then $c_1 = c_2 = \cdots = c_p = 0$.

(vi) There exist functions $G_{t-1}^{ijk}(\mathbf{X}_t)$ and $H_t^{ijk}(X_t, \cdots, X_1)$ such that for $i, j, k = 1, \cdots, p$

$$\left|\frac{\partial}{\partial \theta_i} F_\theta(\mathbf{X}'_t) \frac{\partial^2}{\partial \theta_j \partial \theta_k} F_\theta(\mathbf{X}'_t)\right| \le G_{t-1}^{ijk}, \quad E[G_{t-1}^{ijk}] < \infty \tag{5·17}$$

and (5·18)

$$\left|\{X_t - F_\theta(\mathbf{X}'_t)\} \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} F_\theta(\mathbf{X}'_t)\right| \le H_t^{ijk}, \quad E[H_t^{ijk}] < \infty \tag{5·19}$$

(vii)

$$E\left[\frac{\partial}{\partial \theta} F_\theta(\mathbf{X}'_t)' \frac{\partial}{\partial \theta} F_\theta(\mathbf{X}'_t)\right] < \infty \tag{5·20}$$

The condition (i) implies nonlinear AR model $\{X_t\}$ is strict stationary and ergodic. Due to Tjϕstheim (1986), we have the following lemma.

Lemma 5.1. *Under Assumption 5.1,*

$$\sqrt{n}(\hat{\theta}_n^{(CL)} - \theta_0) \xrightarrow{d} N(0, \mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}) \tag{5·21}$$

where

$$\mathbf{U} = E\left[\frac{\partial F_{\theta_0}(\mathbf{X}_t')}{\partial \theta}\frac{\partial F_{\theta_0}(\mathbf{X}_t')}{\partial \theta'}\right] \tag{5·22}$$

$$\mathbf{W} = E\left[\frac{\partial F_{\theta_0}(\mathbf{X}_t')}{\partial \theta}\frac{\partial F_{\theta_0}(\mathbf{X}_t')}{\partial \theta'}u_t^2\right]. \tag{5·23}$$

Note that u_t is independent of $\frac{\partial F_{\theta_0}(\mathbf{X}_t')}{\partial \theta}$ and $E[u_t^2] = 1$, hence $\mathbf{W} = \mathbf{U}$ and the asymptotic variance of CLS and G estimator is

$$\mathbf{U}^{-1}. \tag{5·24}$$

Since the efficiency of CLS equals that of the G estimator, we are going to discuss its asymptotic efficiency. We set down the following assumption.

Assumption 5.2. *(i)*

$$\lim_{|x|\to\infty} |x| f(x) = 0 \tag{5·25}$$

(ii) The continuous derivative $\dot{f}$ *of* $f(\cdot)$ *exists and*

$$\int |f^{-1}\dot{f}|^4 f(x)dx < \infty \qquad \int |x^2||f^{-1}\dot{f}|^2 f(x)dx < \infty \tag{5·26}$$

The following lemma is due to Kato, Taniguchi and Honda (2006).

Lemma 5.2. *Under Assumptions 5.1 and 5.2, nonlinear AR model has LAN and its Fisher information matrix* Γ *is*

$$\mathbf{\Gamma} = E\left[\left(\frac{\dot{f}(u_t)}{f(u_t)}\right)^2\right]\mathbf{U}. \tag{5·27}$$

This lemma implies

$$\mathbf{U}^{-1} \geq \mathbf{\Gamma}^{-1}. \tag{5·28}$$

If the equality holds, CLS and G estimator is asymptotically efficient. We give the folloing necessary and sufficient condition for this asymptotic efficiency.

Theorem 5.1. *A necessary and sufficient condition that the folloing equality holds is* $\{u_t\}$ *is Gaussian.*

$$\mathbf{U}^{-1} = \mathbf{\Gamma}^{-1} \tag{5·29}$$

6. Numerical Study

In this section, by use of some measures, we compare the asymptotic variances $\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}$, $\mathbf{V}^{-1}$ and the Fisher bound $\mathbf{\Gamma}^{-1}$ numerically. When the model satisfies RCA model, the asymptotic variance of the G estimator is compared with that of CLS in Example 6.1 and with the Fisher bound in Example 6.2. Under GARCH model, we compare the asymptotic variance of the G estimator with that of CLS in Example 6.3 and with the Fisher bound in Example 6.4. Then we see how the G estimator is better than CLS and close to the efficient estimator. Some interesting features are also obtained.

Example 6.1. *Let us consider the following RCA(k) model.*

$$X_t = \sum_{j=1}^{k} z_t(j)X_{t-j} + \epsilon_t \tag{6·1}$$

Here $\{\mathbf{z}_t = (z_t(1), \cdots, z_t(k))\}$ *is i.i.d.* $N(\mathbf{0}, \sigma^2\mathbf{I})$ *and* $\{\epsilon_t\}$ *is i.i.d.* $N(0,1)$. *For (6·1), the lower bound (3·26) of* $\frac{|\mathbf{V}^{-1}|}{|\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}|}$ *is calculated with the length of observations* $n = 100$. *Based on* 1000 *times simulation, we give the sample mean of this lower bound. In Figure 1, we set* $k = 5$ *and these lower bounds in the case of* $0.1 \leq \sigma \leq 0.9$ *are plotted.*

Figure 1 is about here.

From Figure 1, we see that the lower bounds of the efficiency of CLS relative to the G estimator decreases as σ is large. This implies the conditional variance of X_t under the information up to t becomes small, when the variaions of random coefficients are large.

Example 6.2. *We consider the following RCA(1) models.*

$$X_t = (\theta_0 + z_t)X_{t-1} + \epsilon_t \tag{6·2}$$

where θ_0 *is a constant,* $\{\epsilon_t\}$ *is a sequence of i.i.d. random variables with mean* 0 *and variance* 1 *and* $\{z_t\}$ *is a sequence of i.i.d. random variables with mean* 0 *and variance* σ^2. $\{\epsilon_t\}$ *and* $\{z_t\}$ *are assumed to be independent. For (6·2), based on* 1000 *times simulation we calculate the asymptotic variance* V^{-1} *of the G estimator and the Fisher bound* Γ^{-1}. *In Figure 2, we set* $z_t \sim BEx(\frac{\sigma}{\sqrt{2}})$, $\epsilon_t \sim BEx(\frac{1}{\sqrt{2}})$ *and* $\theta_0 = 0.5$. *Then* V^{-1} *and* Γ^{-1} *are plotted in the case of* $0 \leq \sigma \leq 0.5$.

Figure 2 is about here.

From Figure 2, the difference of V^{-1} and Γ^{-1} is constant even if σ changes.

Example 6.3. *Let* $\{X_t\}$ *satisfy the following ARCH(1) models.*

$$X_t = \epsilon_t \sqrt{a_0 + a_1 X_{t-1}^2} \tag{6·3}$$

where $a_0 > 0$, $a_1 \geq 0$ *and* $\{\epsilon_t\}$ *is i.i.d.* $N(0,1)$. *For (6·3), based on* 1000 *times simulation the asymptotic variances* $\mathbf{V}^{-1}$ *of the G estimator and* $\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}$ *of the CLS estimator are calculated. In Figure 3, we set* $a_0 = 0.5$ *and the efficiency of CLS relative to the G estimator* $\frac{|\mathbf{V}^{-1}|}{|\mathbf{U}^{-1}\mathbf{W}\mathbf{U}^{-1}|}$ *are plotted in the case of* $0.1 \leq a_1 \leq 0.9$.

Figure 3 is about here.

From Figure 3, the efficiency of CLS relative to the G estimator is small when the variance of the innovation ϵ_t is lage. Due to this, if the variation of the conditonal variance $a_0 + a_1 X_{t-1}^2$ is larege, the efficiency becomes small.

Example 6.4. *The following GARCH(p,q) models are considered.*

$$\begin{cases} X_t = \epsilon_t \sqrt{U_t} \\ U_t = a_0 + \sum_{j=1}^{q} a_j X_{t-j}^2 + \sum_{i=1}^{p} b_i U_{t-i} \end{cases} \tag{6·4}$$

where $a_0 > 0$, $a_j \geq 0$, $j = 1, \cdots, q$, $b_i \geq 0$, $i = 1, \cdots, p$, *and* $\{\epsilon_t\}$ *is a sequence of i.i.d random variables with mean* 0*, variance* 1 *and the density* $g(\cdot)$*. Based on* 1000 *times simulation, we calculate the efficiency of the G estimator* $\frac{|\Gamma^{-1}|}{|V^{-1}|}$*. Since* $\frac{|\Gamma^{-1}|}{|V^{-1}|} = \frac{4}{(E[\epsilon_t^4]-1)E[\frac{\dot{g}(\epsilon_t)}{g(\epsilon_t)}\epsilon_t+1]^2}$*, the efficiency doesn't depend on the parameter* a_0 *and* a_1*. Hence in Figure 4, we set* $\{\epsilon_t\}$ *are i.i.d. t-distribution with degrees of freedom* ν *and we move the parameter* ν*.*

Figure 4 is about here.

From Figure 4, the efficiency of the G estimator becomes small as the degrees of freedom decreases. This means the efficiency decreases as the spread of the distribution is more than that of Gaussian.

Acknowledgements

The author would like to express his deepest gratitude to Professor Akihiko Kitada of Waseda University for his warm and comfort support and advice. The author has no words with which to thank Professor Masanobu Taniguchi of Waseda University for his valuable comments and advice. The author would like to thank for the reviewer for their comments, which improved the original version of this paper.

REFERENCES

[1] Amano, T. & Taniguchi, M. (2008) Asymptotic efficiency of conditional least squares estimators for ARCH models. Statist. Probab. Lett. 78, 179-185.

[2] Bloomfield, P. & Watson, G. S. (1975) The inefficiency of least squares. Biometrika 62, 121-128.

[3] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307-327.

[4] Chandra, S. A. & Taniguchi, M. (2001) Estimating functions for nonlinear time series models. Nonlinear non-Gaussian models and related filtering methods. Ann. Inst. Statist. Math. 53, 125-141.

[5] Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

[6] Feigin, P. D. & Tweedie, R. L. (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. J. Time Ser. Anal. 6, 1-14.

[7] Godambe, V. P. (1985) The foundations of finite sample estimation in stochastic processes. Biometrika 72, 419-428.

[8] Hall, W. J. & Mathiason, D. J. (1990) On large-sample estimation and testing in parametric models. Internat. Statist. Rev. 58, 77-97.

[9] Hwang, S. Y. & Basawa, I. V. (1993) Asymptotic optimal inference for a class of nonlinear time series models. Stochastic Process. Appl. 46, 91-113.

[10] Ibragimov, I. A. (1963) A Central Limit Theorem for a Class of Dependent Random Variables. Theory Probab. Appl. 8, 83-89.

[11] Kato, H., Taniguchi, M. & Honda, M. (2006) Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. IEEE Trans. Signal Processing. 54, 3414-3425.

[12] Kholevo, A. S. (1969) On estimates of regression coefficients. Theor. Prob. Appl. 14, 79-104.

[13] Lee, S. & Taniguchi, M. (2005) Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. Statist. Sinica 15, 215-234.

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

[15] Nicholls, D. F. & Quinn, B. G. (1982) *Random coefficient autoregressive models: an introduction.* Lecture Notes in Statistics, 11. New York: Springer-Verlag.

[16] Tjϕstheim, D. (1986) Estimation in nonlinear time series models. Stochastic Process. Appl. 21, 251-273.

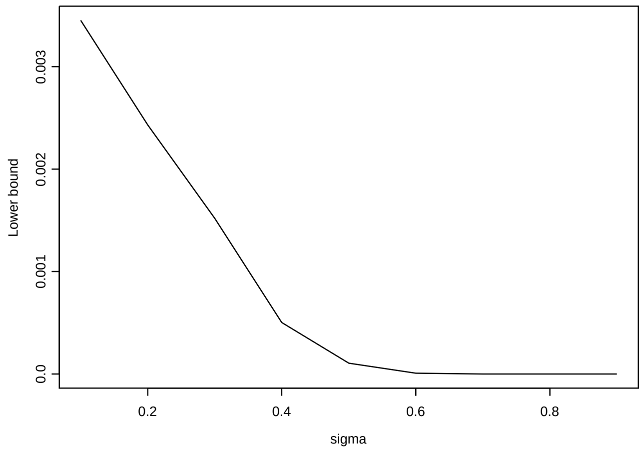

Figure 1: The lower bound of the efficiency of CLS relative to the G estimator for RCA(5) model.

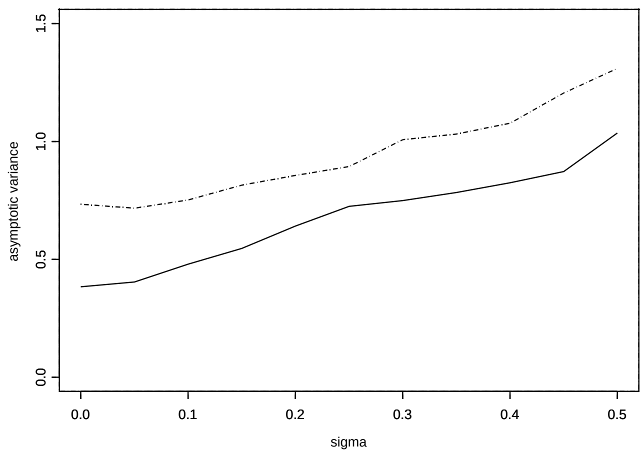

Figure 2: $\mathbf{V}^{-1}$ (dashed line) and $\boldsymbol{\Gamma}^{-1}$ (solid line) for RCA(1) models ($X_t = (\theta_0+z_t)X_{t-1}+\epsilon_t$) with $z_t \sim BEx(\frac{\sigma}{\sqrt{2}})$, $\epsilon_t \sim BEx(\frac{1}{\sqrt{2}})$ and $\theta_0 = 0.5$, $0 \leq \sigma \leq 0.5$.

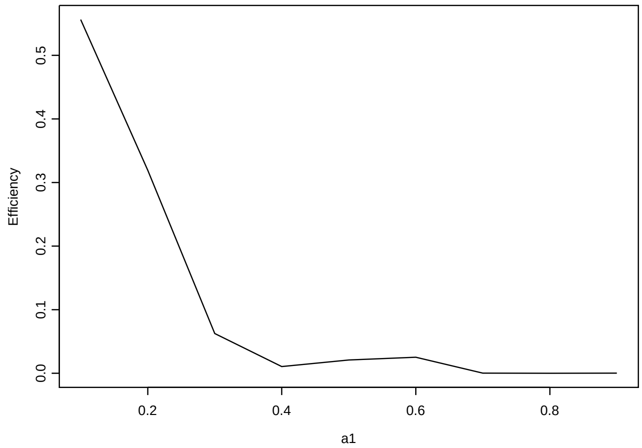

Figure 3: The efficiency of CLS relative to the G estimator for ARCH(1) model with $0.1 \leq a_1 \leq 0.9$.

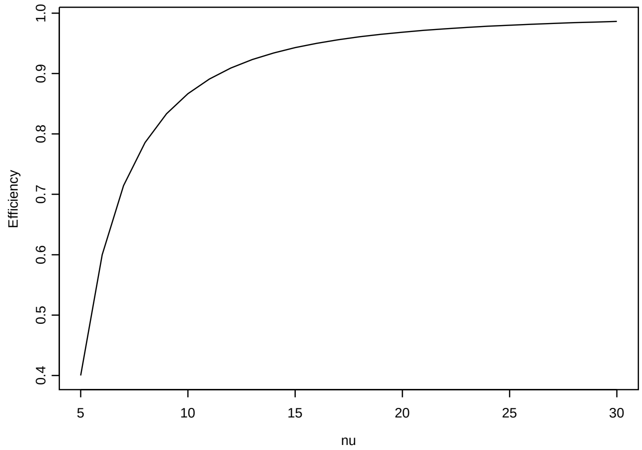

Figure 4: The efficiency of the G estimator for GARCH(k) model where the innovation is t-distribution with the degrees of freedom $5 \leq \nu \leq 30$.