94-F-22

Some Properties of the Maximum Likelihood Estimator in Simultaneous Switching Autoregressive Model

by

Seisho Sato
Tokyo Institute of Technology

Naoto Kunitomo
The University of Tokyo

July 1994

Some Properties of the Maximum Likelihood Estimator in Simultaneous Switching Autoregressive Model

Seisho Sato*
and
Naoto Kunitomo†

June 1994

Abstract

The simultaneous switching autoregressive (SSAR) model proposed by Kunitomo and Sato (1994a,b) is a Markovian nonlinear time series model. We investigate the finite sample as well as the asymptotic properties of the least squares estimator and the maximum likelihood (ML) estimator. Due to a specific simultaneity involved in the SSAR model, the least squares estimator is badly biased. However, the ML estimator under the assumption of Gaussian disturbances gives reasonable estimates.

Keywords : Simultaneous Switching Autoregressive Model, Nonlinear Time Series, Least Squares Estimator, Maximum Likelihood Estimator, Finite Sample Properties, Asymptotic Properties

1. Introduction

In the past decade, several non-linear time series models have been proposed by statisticians and econometricians. For instance, Granger and Andersen (1978) have introduced the bilinear time series models. Also Ozaki and Oda (1978), and Tong (1983) have proposed the exponential autoregressive (EXPAR) model and the threshold autoregressive (TAR) model, respectively, in the field of statistical time series analysis. In particular, a considerable attention has been paid on the TAR model by statisticians and econometricians and several related applications have been reported. The statistical details of many non-linear time series models have been discussed by Tong (1990).

*Graduate Student, Department of System Science, Tokyo Institute of Technology, Meguro-ku, Ôkayama 2-12-1, Tokyo 152, JAPAN.
†Professor, Faculty of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113, JAPAN.

In Kunitomo and Sato (1994a,b), we have introduced an alternative non-linear time series model, which is called the simultaneous switching autoregressive (SSAR) time series model. This model is a kind of Markovian switching time series model with a quite distinctive structure of simultaneity. The current state in the SSAR model is dependent upon not only the past values of states but also the current disturbances, which makes it different from the TAR models. We have proposed this statistical model because we have a conviction that the standard autoregressive moving-average (ARMA) time series model cannot describe one important aspect in many economic time series data, that is, the asymmetrical movements in the up-ward phase (or regime) and in the down-ward phase (or regime). There have been many intuitive observations on this aspect in economic time series among leading economists, but there has not been any useful statistical time series model as far as we know in the statistical time series analysis and the econometric literature. Kunitomo and Sato (1994a,b) have discussed the problems of coherency, ergodicity, the stationary distribution and its moments on a particular version of the multivariate SSAR model. Also they proposed the maximum likelihood estimation for estimating its unknown parameters. We have shown that the class of the SSAR models proposed gives us some explanations and descriptions to handle the very important aspect of asymmetrical movements in two different phases (or regimes). Though this characteristic of economic time series has been observed by a number of economists, enough attention has not been paid on the statistical modelling of this feature in economic time series.

The main purpose of this paper is to investigate the finite sample and asymptotic properties of the maximum likelihood estimator in a systematic way. Since the standard least squares method gives a severely biased estimator as we shall show in Section 3, we will give a fuller investigation on the finite sample distribution of the maximum likelihood (ML) estimator. For this purpose, we will present a fairly detailed tables of the distribution of the ML estimator in the simple univariate case. Because of the intractability of mathematical expressions of the distribution function of the estimators, we have utilized simulation procedures in this paper, but nevertheless the tables given have a high degree of accuracy, any error being in the third decimal place.

In Section 2, we define the SSAR model in this paper. In Section 3, we will discuss the serious problem in the least squares estimation for the SSAR model. In Section 4, we shall investigate the asymptotic properties of the ML estimator and show that it is consistent and asymptotically normal under a set of restrictive assumptions. Then in Section 5, we shall present tables of the distribution of the ML estimator in a simple case and give some comments on its implications. Finally, we give some concluding remarks. The proof of Theorem 2 is lengthy and given in Appendix.

2. Simultaneous Switching Autoregressive Model

In this paper we shall investigate the maximum likelihood estimation method for a version of the multivariate SSAR model. Let $\boldsymbol{y}_t$ be an $m \times 1$ vector of the endogenous variables. The model we consider in this section is represented by

$$(2.1)\qquad \boldsymbol{y}_t = \begin{cases} \boldsymbol{\mu}_1 + \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{D}_1\boldsymbol{u}_t & \text{if} \quad \boldsymbol{e}_m'\boldsymbol{y}_t \geq \boldsymbol{e}_m'\boldsymbol{y}_{t-1} \\ \boldsymbol{\mu}_2 + \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{D}_2\boldsymbol{u}_t & \text{if} \quad \boldsymbol{e}_m'\boldsymbol{y}_t < \boldsymbol{e}_m'\boldsymbol{y}_{t-1} \end{cases} ,$$

where $\boldsymbol{e}_m' = (0, \cdots, 0, 1)$ and $\boldsymbol{\mu}_i'$ $(i = 1, 2)$ are $1 \times m$ vector of constants, and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{D}_i$ $(i = 1, 2)$ are $m \times m$ matrices. The disturbance terms $\{\boldsymbol{u}_t\}$ are distributed with

$$(2.2)\qquad E(\boldsymbol{u}_t|\mathcal{F}_{t-1}) = 0 \ ,$$

and

$$(2.3)\qquad E(\boldsymbol{u}_t\boldsymbol{u}_t'|\mathcal{F}_{t-1}) = I_m \ ,$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\boldsymbol{y}_s,\ s \leq t-1\}$. Then the conditional covariance matrix of $\boldsymbol{D}_i\boldsymbol{u}_t$ given $\mathcal{F}_{t-1}$ is denoted by $\boldsymbol{\Sigma}_i (= \boldsymbol{D}_i\boldsymbol{D}_i', i = 1, 2)$. We assume that $\{\boldsymbol{u}_t\}$ are absolutely continuous random variables with the density function $g(\boldsymbol{u})$ which is everywhere positive in $\boldsymbol{R}^m$. We denote this model as $\mathrm{SSAR}_m(1)$ and also we simply denote $\mathrm{SSAR}_1(1)$ as SSAR(1). By using the standard Markovian representation the p-th order multivariate SSAR model can be reduced to the $\mathrm{SSAR}_m(1)$ model (See Section 5 of Kunitomo and Sato (1994b) for its details). Hence without loss of generality, we shall consider the $\mathrm{SSAR}_m(1)$ model given by (2.1) in this paper.

We note that in (2.1) there are two phases(or regimes) at time t given $\mathcal{F}_{t-1}$. Then there is a basic question that the simultaneity among two phases and the values of the endogenous variables does not cause a logical inconsistency as a statistical model. This problem has been called the coherency problem and the sufficient condition for the logical consistency has been called the coherency condition. The conditions of $\boldsymbol{e}_m'\boldsymbol{y}_t \geq \boldsymbol{e}_m'\boldsymbol{y}_{t-1}$ and $\boldsymbol{e}_m'\boldsymbol{y}_t < \boldsymbol{e}_m'\boldsymbol{y}_{t-1}$ can be rewritten as

$$(2.4)\qquad \boldsymbol{e}_m'\boldsymbol{D}_1\boldsymbol{u}_t \geq \boldsymbol{e}_m'(I_m - \boldsymbol{A})\boldsymbol{y}_{t-1} - \boldsymbol{e}_m'\boldsymbol{\mu}_1 \ ,$$

and

$$(2.5)\qquad \boldsymbol{e}_m'\boldsymbol{D}_2\boldsymbol{u} < \boldsymbol{e}_m'(I_m - \boldsymbol{B})\boldsymbol{y}_{t-1} - \boldsymbol{e}_m'\boldsymbol{\mu}_2 \ ,$$

respectively. When $\boldsymbol{D}_i$ $(i = 1, 2)$ are positive definite matrices, the necessary and sufficient conditions on the coherency problem for (2.1) can be summarized by a $1 \times (m+1)$ vector

$$(2.6)\qquad \begin{aligned} \frac{1}{\sigma_1}[\boldsymbol{e}_m'(I_m - \boldsymbol{A}),\ \boldsymbol{e}_m'\boldsymbol{\mu}_1] &= \frac{1}{\sigma_2}[\boldsymbol{e}_m'(I_m - \boldsymbol{B}),\ \boldsymbol{e}_m'\boldsymbol{\mu}_2] \\ &= \boldsymbol{r}' \ , \end{aligned}$$

where $\sigma_j^2 = \boldsymbol{e}_m'\boldsymbol{\Sigma}_j\boldsymbol{e}_m = \boldsymbol{e}_m'\boldsymbol{D}_j\boldsymbol{D}_j'\boldsymbol{e}_m$ $(j = 1, 2)$. (See Section 4 of Kunitomo and Sato (1994b).)

The SSAR model defined by (2.1) is a Markovian time series model, which is non-linear in the state variables. Thus, the problem of ergodicity is also not a trivial one contrary to the linear time series models including the autoregressive moving average models. Hence we assume the following strong conditions on (2.1).

Assumption I : *In (2.1) the coherency condition given by (2.6) is satisfied and the SSAR model is also ergodic.*

In Kunitomo and Sato (1994b) we have discussed the coherency problem in some details and also given a set of sufficient conditions for the geometric ergodicity for the SSAR model. In particular, they have shown the necessary and sufficient conditions on the ergodicity of the SSAR model when $m = 1$ are $A < 1$, $B < 1$, and $AB < 1$. We should note that the conditions $|A| < 1$ and $|B| < 1$ are too strong for the ergodicity of the SSAR model. This feature of the SSAR model leads to not only a new aspect in the statistical modelling of time series, but also some interesting economic interpretations.

3. Bias of the Least Squares Estimator

When $\boldsymbol{A} = \boldsymbol{B}$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ in (2.1), the SSAR model reduces to the multivariate autoregressive (AR) model. Since the least squares estimation method has been extensively used for the multivariate AR models, one may ask if we can use the least squares method to estimate the unknown parameters in the SSAR$_m(1)$ model given by (2.1). In this section, we shall show that the least squares method gives a seriously biased estimator.

For simplicity, we assume that $\boldsymbol{\mu}_i = 0$ $(i = 1, 2)$. Then the least squares estimation method based on the observed data $\{\boldsymbol{y}_t, 0 \leq t \leq T\}$ can be defined by minimizing the criterion function

$$S_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left\{\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{y}_{t-1}I_t^{(1)} - \boldsymbol{B}\boldsymbol{y}_{t-1}I_t^{(2)}\right\}' \times \left\{\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{y}_{t-1}I_t^{(1)} - \boldsymbol{B}\boldsymbol{y}_{t-1}I_t^{(2)}\right\} , \tag{3.1}$$

where $I_t^{(1)}$ and $I_t^{(2)} = 1 - I_t^{(1)}$ are the indicator functions defined by

$$I_t^{(1)} = \begin{cases} 1 & \text{if} \quad \boldsymbol{e}_m'\boldsymbol{y}_t \geq \boldsymbol{e}_m'\boldsymbol{y}_{t-1} \\ 0 & \text{if} \quad \boldsymbol{e}_m'\boldsymbol{y}_t < \boldsymbol{e}_m'\boldsymbol{y}_{t-1} \end{cases} . \tag{3.2}$$

and $\boldsymbol{\theta}$ is the vector of unknown parameters.
We rewrite (2.1) as

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1}I_t^{(1)} + \boldsymbol{B}\boldsymbol{y}_{t-1}I_t^{(2)} + \boldsymbol{w}_t , \tag{3.3}$$

where

$$\boldsymbol{w}_t = \boldsymbol{D}_1\boldsymbol{u}_tI_t^{(1)} + \boldsymbol{D}_2\boldsymbol{u}_tI_t^{(2)} . \tag{3.4}$$

Let also $\hat{\boldsymbol{A}}_{LS}$ and $\hat{\boldsymbol{B}}_{LS}$ be the least squares estimators of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Then we have

$$\begin{aligned} \hat{\boldsymbol{A}}_{LS} - \boldsymbol{A} &= \left(\sum_{t=1}^{T} \boldsymbol{w}_t\boldsymbol{y}_{t-1}'I_t^{(1)}\right)\left(\sum_{t=1}^{T} \boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'I_t^{(1)}\right)^{-1} \\ &= \boldsymbol{D}_1\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}_t\boldsymbol{y}_{t-1}'I_t^{(1)}\right)\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'I_t^{(1)}\right)^{-1} . \end{aligned} \tag{3.5}$$

When the sample size T increases, this quantity converges to

(3.6) $$\hat{\boldsymbol{A}}_{LS} - \boldsymbol{A} \xrightarrow{P} \boldsymbol{D}_1 \left[E(\boldsymbol{u}_t \boldsymbol{y}'_{t-1} I_t^{(1)})\right] \left[E(\boldsymbol{y}_{t-} \boldsymbol{y}'_{t-1} I_t^{(1)})\right]^{-1} ,$$

provided that the expectations on the right hand side of (3.6) exist. Similarly, we also have

(3.7) $$\hat{\boldsymbol{B}}_{LS} - \boldsymbol{B} \xrightarrow{P} \boldsymbol{D}_2 \left[E(\boldsymbol{u}_t \boldsymbol{y}'_{t-1} I_t^{(2)})\right] \left[E(\boldsymbol{y}_{t-} \boldsymbol{y}'_{t-1} I_t^{(2)})\right]^{-1} .$$

The important aspect in (3.6) and (3.7) is due to the fact that in general the random variables $\{\boldsymbol{u}_t\}$ and $\{\boldsymbol{y}_{t-1} I_t^{(i)}\}$ are contemporaneously correlated. In order to see the bias of the least squares estimator in more details, we assume the following condition.

Assumption II : *The random variables* $\{\boldsymbol{u}_t\}$ *are independently distributed as* $N_m(0, I_m)$.

Let

(3.8) $$\boldsymbol{v}_t^{(i)} = \frac{1}{\sigma_i} \boldsymbol{D}_i \boldsymbol{u}_t \ (i = 1, 2),$$

and

(3.9) $$v_{mt}^{(i)} = \frac{1}{\sigma_i} \boldsymbol{e}'_m \boldsymbol{D}_i \boldsymbol{u}_t \ (i = 1, 2) .$$

Then the condition $I_t^{(1)} = 1$ is equivalent to

(3.10) $$v_{mt}^{(1)} \geq \boldsymbol{r}' \boldsymbol{y}_{t-1} .$$

By the normality assumption on $\{\boldsymbol{u}_t\}$, we have

(3.11) $$E\left[\boldsymbol{v}_t^{(1)} | v_{mt}^{(1)}\right] = \frac{1}{\sigma_1^2} \boldsymbol{\Sigma}_1 \boldsymbol{e}_m v_{mt}^{(1)}$$

and

(3.12) $$\begin{aligned} & E\left[\boldsymbol{D}_1 \boldsymbol{u}_t I_t^{(1)} \mid \mathcal{F}_{t-1}\right] \\ = \ & \sigma_1 E\left\{E\left[\boldsymbol{v}_t^{(1)} \mid v_{mt}^{(1)}\right] I(v_{mt}^{(1)} \geq \boldsymbol{r}' \boldsymbol{y}_{t-1}) \mid \mathcal{F}_{t-1}\right\} \\ = \ & \frac{1}{\sigma_1} \boldsymbol{\Sigma}_1 \boldsymbol{e}_m E\left[v_{mt}^{(1)} I(v_{mt}^{(1)} \geq \boldsymbol{r}' \boldsymbol{y}_{t-1}) \mid \mathcal{F}_{t-1}\right] \\ = \ & \frac{1}{\sigma_1} \boldsymbol{\Sigma}_1 \boldsymbol{e}_m \phi(\boldsymbol{r}' \boldsymbol{y}_{t-1}) , \end{aligned}$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. Similarly, we have

(3.13) $$\begin{aligned} & E\left[\boldsymbol{y}_{t-1} \boldsymbol{y}'_{t-1} I_t^{(1)} \mid \mathcal{F}_{t-1}\right] \\ = \ & \boldsymbol{y}_{t-1} \boldsymbol{y}'_{t-1} \left(1 - \Phi(\boldsymbol{r}' \boldsymbol{y}_{t-1})\right) , \end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Hence the asymptotic biases of $\hat{\boldsymbol{A}}_{LS}$ and $\hat{\boldsymbol{B}}_{LS}$ can be further re-written as

$$ABIAS(\hat{\boldsymbol{A}}_{LS}) = \frac{1}{\sigma_1}\boldsymbol{\Sigma}_1\boldsymbol{e}_m\left\{E\left[\phi(\boldsymbol{r}'\boldsymbol{y}_{t-1})\boldsymbol{y}'_{t-1}\right]\right\} \times \left\{E\left[(1-\Phi(\boldsymbol{r}'\boldsymbol{y}_{t-1}))\boldsymbol{y}_{t-1}\boldsymbol{y}'_{t-1}\right]\right\}^{-1}, \tag{3.14}$$

and

$$ABIAS(\hat{\boldsymbol{B}}_{LS}) = -\frac{1}{\sigma_2}\boldsymbol{\Sigma}_2\boldsymbol{e}_m\left\{E\left[\phi(\boldsymbol{r}'\boldsymbol{y}_{t-1})\boldsymbol{y}'_{t-1}\right]\right\} \times \left\{E\left[\Phi(\boldsymbol{r}'\boldsymbol{y}_{t-1})\boldsymbol{y}_{t-1}\boldsymbol{y}'_{t-1}\right]\right\}^{-1}, \tag{3.15}$$

respectively.
We summarize the above result on the asymptotic bias of LS estimator in the following proposition.

Theorem 1 *Under Assumption I, the least squares estimators $\hat{\boldsymbol{A}}_{LS}$ and $\hat{\boldsymbol{B}}_{LS}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ are not consistent if*

$$E\left[\boldsymbol{u}_t\boldsymbol{y}'_{t-1}I_t^{(1)}\right] \neq 0 . \tag{3.16}$$

Also under Assumption II, the asymptotic biases of $\hat{\boldsymbol{A}}_{LS}$ and $\hat{\boldsymbol{B}}_{LS}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ are given by (3.14) and (3.15), respectively.

It should be important to note that the distribution of $\{\boldsymbol{y}_t\}$ is symmetrically distributed around zero when $\boldsymbol{A} = \boldsymbol{B}$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Then the asymptotic bias of the least squares estimator is zero. However, when $\boldsymbol{A} \neq \boldsymbol{B}$ and $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, the random variables $\{\boldsymbol{u}_t\}$ and $\{\boldsymbol{y}_{t-1}I_t^{(1)}\}$ are contemporaneously correlated and the condition (3.16) is satisfied.

In order to investigate the bias of the least squares estimator, we present some tables of its mean. We also present some figures of the distribution of the least squares estimator. Our simulation has been done by the following procedure:

1. First, generate standard normal random numbers for $\{u_t\}$.
2. Using the above sequense of $\{u_t\}$, obtain the simulated time series for the SSAR(1) model when $m = 1$.
3. Estimate A and B by $\hat{A}_{LS}$ and $\hat{B}_{LS}$.
4. Calculate the sample means for each estimator based on 5,000 replications.

We have chosen the values of parameters as $\{A, B\} = \{0.8, 0.5, 0.2, 0.0, -0.2, -0.5, -1.5\}$ and their sample size $\{T\} = \{50, 100, 500\}$, respectively. Table 1 shows the sample means of the least squares estimators. When $A = B$, the bias of the least squares estimators is negligible. This is expected because the least squares estimator is consistent and its asymptotic bias is zero when the sample size is infinity. However, when $A \neq B$,

Table 1: The mean of the least squares estimator [1]

(i) T = 50

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}_{LS}$	0.7985	0.7099	0.7437	0.7559	0.7653	0.7750	0.7875
		(0.202)	(0.106)	(0.069)	(0.056)	(0.046)	(0.036)	(0.022)
	$\hat{B}_{LS}$	0.8005	1.2277	1.4506	1.5676	1.6894	1.8551	2.3697
		(0.203)	(0.784)	(1.296)	(1.617)	(1.945)	(2.431)	(4.026)
A = 0.5	$\hat{A}_{LS}$	1.2182	0.4951	0.4315	0.4406	0.4500	0.4621	0.4835
		(0.777)	(0.197)	(0.131)	(0.106)	(0.091)	(0.076)	(0.048)
	$\hat{B}_{LS}$	0.7095	0.4940	0.4258	0.3852	0.3345	0.2226	-0.0788
		(0.107)	(0.201)	(0.409)	(0.593)	(0.767)	(1.015)	(1.903)
A = 0.2	$\hat{A}_{LS}$	1.4575	0.4254	0.1896	0.1680	0.1732	0.1785	0.1942
		(1.301)	(0.409)	(0.199)	(0.152)	(0.127)	(0.102)	(0.065)
	$\hat{B}_{LS}$	0.7430	0.4352	0.1919	0.0354	-0.1313	-0.3899	-1.3026
		(0.070)	(0.129)	(0.197)	(0.264)	(0.335)	(0.443)	(0.809)
A = 0.0	$\hat{A}_{LS}$	1.5771	0.3808	0.0317	-0.0064	-0.0127	-0.0056	0.0019
		(1.623)	(0.582)	(0.265)	(0.196)	(0.159)	(0.125)	(0.072)
	$\hat{B}_{LS}$	0.7575	0.4407	0.1768	-0.0027	-0.1858	-0.4920	-1.5332
		(0.055)	(0.107)	(0.155)	(0.200)	(0.246)	(0.311)	(0.558)
A = -0.2	$\hat{A}_{LS}$	1.6860	0.3112	-0.1327	-0.1907	-0.2028	-0.1977	-0.1877
		(1.940)	(0.751)	(0.329)	(0.250)	(0.195)	(0.150)	(0.078)
	$\hat{B}_{LS}$	0.7661	0.4475	0.1703	-0.0076	-0.2017	-0.5061	-1.5490
		(0.045)	(0.092)	(0.128)	(0.160)	(0.196)	(0.245)	(0.384)
A = -0.5	$\hat{A}_{LS}$	1.8508	0.2397	-0.3903	-0.4933	-0.5118	-0.4968	-0.4816
		(2.427)	(1.019)	(0.439)	(0.318)	(0.251)	(0.177)	(0.071)
	$\hat{B}_{LS}$	0.7748	0.4616	0.1808	-0.0049	-0.1971	-0.4952	-1.4733
		(0.036)	(0.077)	(0.102)	(0.123)	(0.148)	(0.180)	(0.188)
A = -1.5	$\hat{A}_{LS}$	2.3902	-0.0908	-1.3122	-1.5477	-1.5377	-1.4656	-1.5000
		(4.032)	(1.901)	(0.808)	(0.555)	(0.373)	(0.193)	(0.000)
	$\hat{B}_{LS}$	0.7877	0.4829	0.1941	0.0039	-0.1896	-0.4799	-1.5000
		(0.022)	(0.047)	(0.064)	(0.074)	(0.080)	(0.070)	(0.000)

[1] The value in the parentheses shows the root mean squared error.

Table 1: The mean of the least squares estimator (cont.)

(ii) T = 100

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}_{LS}$	0.8055	0.7152	0.7479	0.7601	0.7687	0.7767	0.7889
		(0.143)	(0.094)	(0.059)	(0.046)	(0.037)	(0.029)	(0.017)
	$\hat{B}_{LS}$	0.7978	1.2359	1.4515	1.5603	1.6583	1.8087	2.3009
		(0.141)	(0.759)	(1.268)	(1.577)	(1.877)	(2.331)	(3.843)
A = 0.5	$\hat{A}_{LS}$	1.2328	0.4972	0.4361	0.4405	0.4519	0.4636	0.4848
		(0.757)	(0.133)	(0.099)	(0.086)	(0.073)	(0.058)	(0.034)
	$\hat{B}_{LS}$	0.7160	0.4974	0.4297	0.3904	0.3317	0.2766	-0.0320
		(0.092)	(0.135)	(0.331)	(0.492)	(0.651)	(0.916)	(1.708)
A = 0.2	$\hat{A}_{LS}$	1.4507	0.4329	0.1966	0.1729	0.1720	0.1790	0.1931
		(1.267)	(0.331)	(0.133)	(0.107)	(0.092)	(0.073)	(0.045)
	$\hat{B}_{LS}$	0.7477	0.4360	0.1976	0.0353	-0.1269	-0.3824	-1.2837
		(0.059)	(0.099)	(0.131)	(0.181)	(0.236)	(0.321)	(0.574)
A = 0.0	$\hat{A}_{LS}$	1.5623	0.3874	0.0380	-0.0019	-0.0071	-0.0051	0.0028
		(1.580)	(0.495)	(0.180)	(0.134)	(0.108)	(0.086)	(0.050)
	$\hat{B}_{LS}$	0.7601	0.4424	0.1752	-0.0037	-0.1927	-0.4831	-1.5317
		(0.046)	(0.085)	(0.107)	(0.136)	(0.168)	(0.221)	(0.387)
A = -0.2	$\hat{A}_{LS}$	1.6645	0.3538	-0.1276	-0.1912	-0.2049	-0.1982	-0.1913
		(1.884)	(0.668)	(0.232)	(0.167)	(0.136)	(0.103)	(0.054)
	$\hat{B}_{LS}$	0.7689	0.4517	0.1744	-0.0074	-0.2017	-0.5009	-1.5537
		(0.037)	(0.072)	(0.088)	(0.110)	(0.135)	(0.173)	(0.267)
A = -0.5	$\hat{A}_{LS}$	1.8091	0.2625	-0.3910	-0.4814	-0.5020	-0.4972	-0.4891
		(2.332)	(0.896)	(0.316)	(0.218)	(0.170)	(0.123)	(0.046)
	$\hat{B}_{LS}$	0.7768	0.4644	0.1817	-0.0035	-0.1943	-0.4960	-1.4874
		(0.029)	(0.058)	(0.073)	(0.086)	(0.102)	(0.125)	(0.125)
A = -1.5	$\hat{A}_{LS}$	2.3123	-0.0295	-1.2830	-1.5303	-1.5517	-1.4913	-1.5000
		(3.856)	(1.697)	(0.577)	(0.388)	(0.268)	(0.125)	(0.000)
	$\hat{B}_{LS}$	0.7889	0.4846	0.1941	0.0042	-0.1910	-0.4887	-1.5000
		(0.017)	(0.034)	(0.045)	(0.053)	(0.055)	(0.047)	(0.000)

Table 1: The mean of the least squares estimator (cont.)

(iii) T = 500

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}_{LS}$	0.7984	0.7205	0.7511	0.7625	0.7704	0.7781	0.7899
		(0.064)	(0.081)	(0.050)	(0.039)	(0.031)	(0.023)	(0.012)
	$\hat{B}_{LS}$	0.8010	1.2453	1.4465	1.5496	1.6429	1.7806	2.2351
		(0.063)	(0.749)	(1.249)	(1.552)	(1.846)	(2.284)	(3.742)
A = 0.5	$\hat{A}_{LS}$	1.2476	0.4989	0.4387	0.4457	0.4554	0.4668	0.4856
		(0.751)	(0.059)	(0.070)	(0.060)	(0.051)	(0.039)	(0.020)
	$\hat{B}_{LS}$	0.7208	0.4968	0.4369	0.3991	0.3633	0.2968	0.0274
		(0.081)	(0.058)	(0.259)	(0.421)	(0.587)	(0.823)	(1.570)
A = 0.2	$\hat{A}_{LS}$	1.4484	0.4398	0.2007	0.1769	0.1774	0.1821	0.1935
		(1.251)	(0.261)	(0.058)	(0.051)	(0.044)	(0.036)	(0.021)
	$\hat{B}_{LS}$	0.7511	0.4387	0.1988	0.0404	-0.1199	-0.3785	-1.2688
		(0.050)	(0.069)	(0.058)	(0.088)	(0.127)	(0.179)	(0.331)
A = 0.0	$\hat{A}_{LS}$	1.5489	0.4044	0.0452	0.0007	-0.0094	-0.0054	0.0013
		(1.552)	(0.425)	(0.090)	(0.060)	(0.049)	(0.039)	(0.022)
	$\hat{B}_{LS}$	0.7627	0.4448	0.1779	0.0000	-0.1870	-0.4827	-1.5224
		(0.039)	(0.062)	(0.050)	(0.060)	(0.076)	(0.099)	(0.171)
A = -0.2	$\hat{A}_{LS}$	1.6439	0.3577	-0.1189	-0.1855	-0.1993	-0.1989	-0.1937
		(1.847)	(0.581)	(0.127)	(0.077)	(0.060)	(0.046)	(0.025)
	$\hat{B}_{LS}$	0.7706	0.4548	0.1772	-0.0077	-0.2020	-0.5013	-1.5503
		(0.031)	(0.051)	(0.045)	(0.050)	(0.059)	(0.076)	(0.124)
A = -0.5	$\hat{A}_{LS}$	1.7801	0.2921	-0.3759	-0.4810	-0.5020	-0.4989	-0.4965
		(2.284)	(0.819)	(0.180)	(0.098)	(0.076)	(0.055)	(0.020)
	$\hat{B}_{LS}$	0.7781	0.4666	0.1821	-0.0062	-0.1986	-0.4989	-1.5048
		(0.023)	(0.039)	(0.036)	(0.039)	(0.046)	(0.054)	(0.054)
A = -1.5	$\hat{A}_{LS}$	2.2297	0.0239	-1.2639	-1.5243	-1.5532	-1.5046	-1.5000
		(3.736)	(1.567)	(0.334)	(0.172)	(0.128)	(0.054)	(0.000)
	$\hat{B}_{LS}$	0.7899	0.4857	0.1936	0.0015	-0.1944	-0.4962	-1.5000
		(0.011)	(0.020)	(0.020)	(0.023)	(0.025)	(0.020)	(0.000)

the situation is dramatically changed and the bias of the least squares estimator is substantially large. The estimated values for cofficients often exceed the boundaries of the stationary region when the absolute values of A and B are large. Also it is interesting to see that the estimated signs of coefficients are not necessarily the same as the true signs. This causes a serious problem when we want to interpret the estimated coefficients. Our simulations indicate that the least squares method gives a badly biased estimate when $A \neq B$ and hence it is not adequate for estimating the SSAR(1) model. In Figure 1, we have shown the histograms of the least squares estimator for the case of $A = 0.2, B = 0.8$ and $A = 0.2, B = -0.2$. These figures vividly show the observations we have found on the distribution of the least squares estimator.

4. The ML Estimator and its Asymptotic Properties

In this section, we consider the SSAR$_m$(1) model given by (2.1) when $m \geq 1$. Kunitomo and Sato (1994a,b) have proposed the ML estimation method for estimating its unknown parameters. Since the SSAR$_m$(1) model is a Markovian process, the joint density function $p(\boldsymbol{y}_T,\ \boldsymbol{y}_{T-1},\ \cdots,\ \boldsymbol{y}_1)$ given $\boldsymbol{y}_0$ can be rewritten as

$$p(\boldsymbol{y}_T, \boldsymbol{y}_{T-1}, \cdots, \boldsymbol{y}_1 | \boldsymbol{y}_0) = p(\boldsymbol{y}_T | \boldsymbol{y}_{T-1}) p(\boldsymbol{y}_{T-1} | \boldsymbol{y}_{T-2}) \cdots p(\boldsymbol{y}_1 | \boldsymbol{y}_0) \ , \tag{4.1}$$

where $p(\boldsymbol{y}_t | \boldsymbol{y}_{t-1})$ is the conditional density function of $\boldsymbol{y}_t$ given $\boldsymbol{y}_{t-1}$. Then under Assumption II and $|\boldsymbol{\Sigma}_i| \neq 0$ $(i = 1, 2)$, the conditional log-likelihood function of $\{\boldsymbol{y}_t,\ 1 \leq t \leq T\}$ given $\boldsymbol{y}_0$ can be rewritten as

$$\begin{aligned} \log L_T(\boldsymbol{\theta}) &= -\frac{Tm}{2} \log 2\pi \\ &- \frac{1}{2} \log |\boldsymbol{\Sigma}_1| \sum_{t=1}^{T} I(\boldsymbol{e}_m' \boldsymbol{y}_t \geq \boldsymbol{e}_t' \boldsymbol{y}_{t-1}) \\ &- \frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{\mu}_1 - \boldsymbol{A} \boldsymbol{y}_{t-1})' \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu}_1 - \boldsymbol{A} \boldsymbol{y}_{t-1}) I(\boldsymbol{e}_t' \boldsymbol{y}_t \geq \boldsymbol{e}_m' \boldsymbol{y}_{t-1}) \\ &- \frac{1}{2} \log |\boldsymbol{\Sigma}_2| \sum_{t=1}^{T} I(\boldsymbol{e}_m' \boldsymbol{y}_t < \boldsymbol{e}_t' \boldsymbol{y}_{t-1}) \\ &- \frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{\mu}_2 - \boldsymbol{B} \boldsymbol{y}_{t-1})' \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu}_2 - \boldsymbol{B} \boldsymbol{y}_{t-1}) I(\boldsymbol{e}_t' \boldsymbol{y}_t < \boldsymbol{e}_m' \boldsymbol{y}_{t-1}) \ , \end{aligned} \tag{4.2}$$

where $\boldsymbol{\theta}' = \big(\boldsymbol{r}',\ (vech(\boldsymbol{\Sigma}_j))',\ \boldsymbol{e}_i'\boldsymbol{A},\ \boldsymbol{e}_i'\boldsymbol{B}\ \boldsymbol{e}_i'\boldsymbol{\mu}_j\ \ (i = 1, \cdots, m-1,\ j = 1, 2)\big)$ denotes the vector of unknown parameters in the SSAR$_m$(1) and the parameter space $\boldsymbol{\Theta}$ is defined correspondingly. Hence, the ML estimator we have proposed is the vector $\boldsymbol{\theta}$ which maximizes $\log L_T(\boldsymbol{\theta})$ under the restriction (2.6). Since it is difficult to obtain its analytical expression, we use the numerical maximization technique to get the ML estimate. In Kunitomo and Sato (1994b) we have claimed that the ML estimator $\hat{\boldsymbol{\theta}}_{ML}$ of $\boldsymbol{\theta}$ is consistent and asymptotically normal when the sample size increases. Since Kunitomo and Sato (1994b) did not give a complete proof of this result, we shall restate it and give its proof in Appendix.

Theorem 2 *For the $SSAR_m(1)$ model given by (2.1), suppose the sufficient conditions for the coherency and ergodicity hold and the disturbances terms $\{\boldsymbol{u}_t\}$ are independently distributed as $N(0, I_m)$ with $|\boldsymbol{\Sigma}_i| \neq 0$ $(i = 1, 2)$. Also suppose that the true parameter vector $\boldsymbol{\theta}$ is an interior point of the parameter space Θ. Then the ML estimator $\hat{\boldsymbol{\theta}}_{ML}$ of unknown parameter $\boldsymbol{\theta}$ is consistent and asymptotically normally distributed as*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left[0,\ I(\boldsymbol{\theta})^{-1}\right] \quad , \tag{4.3}$$

where

$$I(\boldsymbol{\theta}) = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\left[-\frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \quad . \tag{4.4}$$

We note that the assumptions in Theorem 2 suffice the existence of the second order moments for $\{\boldsymbol{y}_t\}$. In fact, in Section 5.3 of Kunitomo and Sato (1994b) we have shown that the existence of k-th order moments for disturbance terms implies the existence of k-th moments on $\{\boldsymbol{y}_t\}$. Hence the information matrix given by (4.4) is well-defined in the present situation.

5. Tables of the Distributions of the ML Estimator

In this section we consider the SSAR(1) model when $m = 1$. In this case (2.1) becomes

$$y_t - \mu = \begin{cases} A(y_{t-1} - \mu) + \sigma_1 u_t & \text{if} \quad y_t \geq y_{t-1} \\ B(y_{t-1} - \mu) + \sigma_2 u_t & \text{if} \quad y_t < y_{t-1} \end{cases} \quad , \tag{5.1}$$

where μ is the location parameter in this model.
In this simple case the coherency conditions are given by

$$r = \frac{1 - A}{\sigma_1} = \frac{1 - B}{\sigma_2} \tag{5.2}$$

and

$$\frac{\mu_1}{\sigma_1} = \frac{\mu_2}{\sigma_2} \quad . \tag{5.3}$$

By comparing (2.1) and (5.1), the second condition is also equal to $r\mu$. Since we can write another random variables $y_t' = y_t - \mu$, we can assume $\mu = 0$ from the beginning without loss of generality.

Under Assumption II and $\sigma_i > 0$ $(i = 1, 2)$, the conditional log-likelihood function of $\{y_t,\ 1 \leq t \leq T\}$ given y_0 can be rewritten as

$$\begin{aligned} \log L_T(\boldsymbol{\theta}) &= -\frac{T}{2}\log 2\pi \\ &- \frac{1}{2}\log\sigma_1^2 \sum_{t=1}^{T} I_t^{(1)} - \frac{1}{2\sigma_1^2}\sum_{t=1}^{T}(y_t - Ay_{t-1})^2 I_t^{(1)} \\ &- \frac{1}{2}\log\sigma_2^2 \sum_{t=1}^{T} I_t^{(2)} - \frac{1}{2\sigma_2^2}\sum_{t=1}^{T}(y_t - By_{t-1})^2 I_t^{(2)} \quad , \end{aligned} \tag{5.4}$$

where $\boldsymbol{\theta} = (A, B, \sigma_1, \sigma_2)'$ and

$$(5.5) \qquad I_t^{(1)} = \begin{cases} 1 & \text{if} \quad y_t \geq y_{t-1} \\ 0 & \text{if} \quad y_t < y_{t-1} \end{cases} ,$$

$$(5.6) \qquad I_t^{(2)} = 1 - I_t^{(1)} .$$

Since the analytical expression of the exact distribution of the ML estimator is not tractable, we have analyzed the properties of ML estimator for the unknown parameters by Monte Carlo simulations. We have generated the simulated time series $\{y_t\}$ for the SSAR(1) model by using the standard normal random numbers for the disturbances $\{u_t\}$. First, we obtained the tables of the sample mean of the ML estimators from 5,000 replications in each case in Table 2. Then we can compare these tables to those of the least squares estimator shown in Table 1. Table 2 shows that the biases of the ML estimator are vary small even when $A \neq B$ and they are negligible for paractical purposes. We also have shown some histograms of the ML estimator in Figure 2.

Next, we have shown the tables of the empirical distribution functions of the ML estimator for some cases. In order to make a comparison with the normal distribution possible, we calculated the values of the empirical distribution based upon the standardized estimator, that is,

$$(5.7) \qquad \hat{\theta}_i' = \frac{\hat{\theta}_i - m_\theta}{\sqrt{\dfrac{\sum_{j=1}^{N^{(sim)}} (\hat{\theta}_j - m_\theta)^2}{N^{(sim)} - 1}}} ,$$

where $\hat{\theta}_i$ is the estimate of a parameter in the i-th replication, $m_\theta = 1/N^{(sim)} \sum_{j=1}^{N^{(sim)}} \hat{\theta}_j$ and $N^{(sim)}$ is the number of replications. In our simulations it is 20,000 in all cases. In Table 3, we also have shown some percentiles of the empirical distributions. It seems that the distributions of the ML estimator are slightly asymmetrical when $T = 100$. However, when $T = 1,000$, the distribution function of the ML estimator is very close to the standard normal distribution function. The error of a sample distribution function is about 0.01 at 99 percent confidence level in this case by using the Kolomogorov-Smirnov statistic. From this observation we confirm that the asymptotic properties of the ML estimator discussed in Section 4 hold approximately when the sample size is 1,000. We also have plotted the distribution functions of the ML estimator obtained by simulations in Figure 3.

Finally, we should mention to the fact that the ML estimation of the SSAR model is based on the normality assumption for the disturbance terms. In order to see the robustness of the finite sample properties of the ML estimator, we did a small amount of Monte Carlo simulations. We calculated the pseudo-ML estimator when $\{u_t\}$ are not normally distributed random variables. The sample means of the empirical distributions have been calculated and given in Table 4 by assuming that $\{u_t\}$ are distributed as the Gaussian sum:

$$(5.8) \qquad u_t \sim \begin{cases} N(0, \frac{1}{3}) & \text{(Prob. 0.5)} \\ N(0, \frac{5}{3}) & \text{(Prob. 0.5)} \end{cases} .$$

Table 2: The mean of the ML estimator[1,2]

(i) T = 50

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7582	0.7833	0.7933	0.7950	0.7963	0.7970	0.7991
		(0.111)	(0.057)	(0.039)	(0.032)	(0.028)	(0.025)	(0.018)
	$\hat{B}$	0.7589	0.4379	0.1375	-0.0639	-0.2600	-0.5830	-1.6168
		(0.110)	(0.196)	(0.275)	(0.325)	(0.374)	(0.485)	(0.821)
A = 0.5	$\hat{A}$	0.4422	0.4650	0.4800	0.4870	0.4901	0.4941	0.4996
		(0.195)	(0.143)	(0.109)	(0.090)	(0.079)	(0.066)	(0.044)
	$\hat{B}$	0.7851	0.4665	0.1566	-0.0414	-0.2480	-0.5435	-1.5709
		(0.055)	(0.142)	(0.213)	(0.251)	(0.300)	(0.358)	(0.604)
A = 0.2	$\hat{A}$	0.1291	0.1570	0.1789	0.1842	0.1898	0.1929	0.2010
		(0.279)	(0.212)	(0.161)	(0.140)	(0.120)	(0.102)	(0.064)
	$\hat{B}$	0.7924	0.4819	0.1793	-0.0262	-0.2286	-0.5291	-1.5421
		(0.038)	(0.104)	(0.159)	(0.199)	(0.233)	(0.286)	(0.464)
A = 0.0	$\hat{A}$	-0.0736	-0.0446	-0.0248	-0.0178	-0.0097	-0.0035	0.0039
		(0.331)	(0.252)	(0.197)	(0.168)	(0.144)	(0.122)	(0.073)
	$\hat{B}$	0.7943	0.4864	0.1830	-0.0158	-0.2166	-0.5141	-1.5227
		(0.033)	(0.091)	(0.142)	(0.168)	(0.195)	(0.245)	(0.392)
A = -0.2	$\hat{A}$	-0.2674	-0.2491	-0.2261	-0.2235	-0.2109	-0.2007	-0.1935
		(0.384)	(0.293)	(0.229)	(0.203)	(0.169)	(0.136)	(0.077)
	$\hat{B}$	0.7959	0.4899	0.1891	-0.0077	-0.2128	-0.5054	-1.4928
		(0.029)	(0.077)	(0.121)	(0.148)	(0.172)	(0.206)	(0.302)
A = -0.5	$\hat{A}$	-0.5661	-0.5487	-0.5241	-0.5171	-0.5097	-0.4914	-0.4823
		(0.481)	(0.358)	(0.285)	(0.243)	(0.207)	(0.160)	(0.068)
	$\hat{B}$	0.7977	0.4929	0.1953	-0.0049	-0.1990	-0.4957	-1.4629
		(0.024)	(0.067)	(0.102)	(0.119)	(0.138)	(0.163)	(0.180)
A = -1.5	$\hat{A}$	-1.6170	-1.5670	-1.5452	-1.5200	-1.4974	-1.4652	NA
		(0.819)	(0.588)	(0.463)	(0.397)	(0.314)	(0.178)	(NA)
	$\hat{B}$	0.7991	0.4982	0.2000	0.0027	-0.1911	-0.4840	NA
		(0.018)	(0.045)	(0.066)	(0.073)	(0.077)	(0.068)	(NA)

1 The value in the parentheses shows the root mean squared error.

2 "NA" corresponds to the case when it is not ergodic. We did not have investigated the ML estimator in this case.

Table 2: The mean of the ML estimator (cont.)

(ii) T = 100

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7783	0.7930	0.7969	0.7964	0.7982	0.7986	0.7995
		(0.071)	(0.037)	(0.025)	(0.023)	(0.020)	(0.017)	(0.012)
	$\hat{B}$	0.7777	0.4726	0.1718	-0.0368	-0.2354	-0.5354	-1.5500
		(0.072)	(0.130)	(0.182)	(0.221)	(0.266)	(0.323)	(0.551)
A = 0.5	$\hat{A}$	0.4744	0.4829	0.4911	0.4941	0.4952	0.4972	0.4992
		(0.127)	(0.097)	(0.072)	(0.062)	(0.054)	(0.046)	(0.031)
	$\hat{B}$	0.7930	0.4813	0.1820	-0.0218	-0.2218	-0.5252	-1.5321
		(0.036)	(0.098)	(0.143)	(0.172)	(0.201)	(0.248)	(0.400)
A = 0.2	$\hat{A}$	0.1661	0.1807	0.1869	0.1936	0.1955	0.1979	0.2007
		(0.186)	(0.142)	(0.112)	(0.096)	(0.084)	(0.070)	(0.045)
	$\hat{B}$	0.7960	0.4915	0.1888	-0.0155	-0.2103	-0.5130	-1.5191
		(0.026)	(0.072)	(0.112)	(0.137)	(0.159)	(0.195)	(0.322)
A = 0.0	$\hat{A}$	-0.0347	-0.0187	-0.0128	-0.0082	-0.0064	-0.0029	0.0010
		(0.220)	(0.173)	(0.136)	(0.117)	(0.104)	(0.084)	(0.051)
	$\hat{B}$	0.7975	0.4944	0.1926	-0.0072	-0.2065	-0.5154	-1.5190
		(0.022)	(0.062)	(0.096)	(0.116)	(0.141)	(0.168)	(0.272)
A = -0.2	$\hat{A}$	-0.2322	-0.2275	-0.2135	-0.2093	-0.2036	-0.2003	-0.1965
		(0.263)	(0.205)	(0.159)	(0.135)	(0.119)	(0.096)	(0.053)
	$\hat{B}$	0.7983	0.4954	0.1959	-0.0054	-0.2034	-0.5073	-1.4995
		(0.020)	(0.054)	(0.083)	(0.102)	(0.119)	(0.145)	(0.215)
A = -0.5	$\hat{A}$	-0.5451	-0.5252	-0.5154	-0.5118	-0.5031	-0.4974	-0.4906
		(0.326)	(0.246)	(0.197)	(0.167)	(0.143)	(0.111)	(0.046)
	$\hat{B}$	0.7982	0.4972	0.1971	-0.0016	-0.1973	-0.4972	-1.4798
		(0.017)	(0.046)	(0.071)	(0.084)	(0.095)	(0.109)	(0.122)
A = -1.5	$\hat{A}$	-1.5685	-1.5450	-1.5251	-1.5163	-1.5044	-1.4788	NA
		(0.563)	(0.407)	(0.321)	(0.272)	(0.214)	(0.121)	(NA)
	$\hat{B}$	0.7993	0.4986	0.1995	0.0020	-0.1963	-0.4919	NA
		(0.012)	(0.031)	(0.045)	(0.051)	(0.053)	(0.045)	(NA)

Table 2: The mean of the ML estimator (cont.)

(iii) T = 500

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7951	0.7985	0.7997	0.7993	0.7998	0.7998	0.7999
		(0.028)	(0.015)	(0.011)	(0.009)	(0.008)	(0.007)	(0.005)
	$\hat{B}$	0.7949	0.4952	0.1950	-0.0094	-0.2056	-0.5064	-1.5076
		(0.029)	(0.054)	(0.078)	(0.096)	(0.112)	(0.137)	(0.238)
A = 0.5	$\hat{A}$	0.4932	0.4958	0.4974	0.4984	0.4988	0.4995	0.4997
		(0.055)	(0.041)	(0.032)	(0.027)	(0.024)	(0.020)	(0.014)
	$\hat{B}$	0.7983	0.4962	0.1946	-0.0034	-0.2050	-0.5044	-1.5074
		(0.015)	(0.042)	(0.063)	(0.074)	(0.089)	(0.108)	(0.178)
A = 0.2	$\hat{A}$	0.1940	0.1961	0.1974	0.1985	0.1995	0.1992	0.1998
		(0.078)	(0.062)	(0.049)	(0.042)	(0.038)	(0.032)	(0.020)
	$\hat{B}$	0.7992	0.4978	0.1970	-0.0032	-0.2022	-0.5035	-1.5038
		(0.011)	(0.031)	(0.049)	(0.060)	(0.071)	(0.087)	(0.143)
A = 0.0	$\hat{A}$	-0.0028	-0.0039	-0.0020	-0.0013	0.0001	0.0004	0.0002
		(0.093)	(0.077)	(0.060)	(0.052)	(0.044)	(0.037)	(0.022)
	$\hat{B}$	0.7997	0.4991	0.1979	-0.0020	-0.2012	-0.5004	-1.5028
		(0.009)	(0.027)	(0.042)	(0.052)	(0.061)	(0.074)	(0.118)
A = -0.2	$\hat{A}$	-0.2061	-0.2039	-0.2025	-0.2007	-0.2006	-0.2003	-0.1991
		(0.112)	(0.088)	(0.071)	(0.062)	(0.053)	(0.043)	(0.023)
	$\hat{B}$	0.7998	0.4988	0.1997	-0.0006	-0.2008	-0.5005	-1.5010
		(0.009)	(0.024)	(0.037)	(0.045)	(0.052)	(0.063)	(0.094)
A = -0.5	$\hat{A}$	-0.5043	-0.5031	-0.5026	-0.5013	-0.4992	-0.4984	-0.4979
		(0.139)	(0.106)	(0.085)	(0.075)	(0.064)	(0.049)	(0.019)
	$\hat{B}$	0.7997	0.5000	0.1991	-0.0006	-0.1990	-0.4985	-1.4953
		(0.007)	(0.020)	(0.031)	(0.038)	(0.043)	(0.048)	(0.051)
A = -1.5	$\hat{A}$	-1.5106	-1.5050	-1.5030	-1.5023	-1.5004	-1.4951	NA
		(0.244)	(0.176)	(0.140)	(0.118)	(0.093)	(0.052)	(NA)
	$\hat{B}$	0.7999	0.4997	0.2002	0.0006	-0.1995	-0.4984	NA
		(0.005)	(0.014)	(0.020)	(0.023)	(0.023)	(0.019)	(NA)

Table 3: The distribution function of the ML estimator

(i) The case $A = 0.4,\ B = 0.8,\ \sigma_1 = 0.6,\ \sigma_2 = 0.2$

	Normal		$T=$ 100				$T=$ 1000		
x		$\hat{A}$	$\hat{B}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{A}$	$\hat{B}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
-3.0	0.001	0.006	0.006	0.000	0.001	0.002	0.002	0.000	0.001
-2.5	0.006	0.015	0.015	0.002	0.004	0.008	0.009	0.004	0.005
-2.0	0.023	0.034	0.035	0.013	0.020	0.027	0.027	0.020	0.022
-1.4	0.081	0.088	0.089	0.070	0.076	0.084	0.085	0.079	0.080
-1.0	0.159	0.154	0.151	0.156	0.160	0.157	0.161	0.159	0.159
-0.8	0.212	0.198	0.197	0.218	0.216	0.207	0.211	0.213	0.212
-0.6	0.274	0.254	0.250	0.287	0.281	0.271	0.268	0.273	0.277
-0.4	0.345	0.318	0.313	0.362	0.351	0.337	0.331	0.348	0.346
-0.2	0.421	0.387	0.386	0.444	0.429	0.413	0.408	0.426	0.426
0.0	0.500	0.466	0.463	0.519	0.507	0.491	0.485	0.506	0.503
0.2	0.579	0.547	0.545	0.601	0.586	0.569	0.566	0.587	0.582
0.4	0.655	0.628	0.630	0.673	0.661	0.648	0.645	0.662	0.658
0.6	0.726	0.709	0.712	0.739	0.729	0.721	0.722	0.730	0.726
0.8	0.788	0.782	0.784	0.797	0.792	0.785	0.786	0.790	0.790
1.0	0.841	0.846	0.849	0.844	0.842	0.840	0.844	0.843	0.840
1.4	0.919	0.938	0.940	0.912	0.918	0.921	0.925	0.916	0.919
2.0	0.977	0.991	0.991	0.968	0.973	0.982	0.982	0.975	0.976
2.5	0.994	1.000	0.999	0.989	0.992	0.996	0.996	0.992	0.992
3.0	0.999	1.000	1.000	0.996	0.998	1.000	1.000	0.997	0.999
Q(2.5)	-1.960	-2.225	-2.217	-1.781	-1.905	-2.036	-2.043	-1.909	-1.951
Q(25)	-0.674	-0.612	-0.600	-0.700	-0.691	-0.661	-0.663	-0.680	-0.682
Q(50)	0.000	0.085	0.086	-0.050	-0.018	0.027	0.039	-0.014	-0.009
Q(75)	0.674	0.714	0.704	0.635	0.663	0.679	0.682	0.663	0.675
Q(97.5)	1.960	1.706	1.720	2.116	2.035	1.876	1.872	1.997	1.984
IQR	1.348	1.326	1.304	1.335	1.354	1.340	1.345	1.343	1.357
BIAS	—	-0.029	-0.006	0.000	-0.002	-0.003	-0.001	0.000	0.000
STD	—	0.142	0.031	0.077	0.018	0.043	0.009	0.024	0.006

1. $Q(p)$ is denoted as p percentile.
2. "IQR" shows the interquartile range of the distribution.
3. "BIAS" shows the mean bias from the true value.
4. "STD" shows the root mean squared error.

Table 3: The distribution function of the ML estimator (cont.)

(ii) The case $A = 0.2,\ B = -0.2,\ \sigma_1 = 0.8,\ \sigma_2 = 1.2$

	Normal	$T =$ 100				$T =$ 1000			
x		$\hat{A}$	$\hat{B}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{A}$	$\hat{B}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$
-3.0	0.001	0.003	0.004	0.001	0.001	0.002	0.002	0.002	0.001
-2.5	0.006	0.009	0.011	0.004	0.004	0.007	0.007	0.006	0.006
-2.0	0.023	0.028	0.030	0.020	0.018	0.023	0.025	0.021	0.022
-1.4	0.081	0.086	0.086	0.077	0.077	0.084	0.085	0.080	0.079
-1.0	0.159	0.158	0.159	0.158	0.159	0.156	0.158	0.159	0.158
-0.8	0.212	0.206	0.204	0.212	0.214	0.208	0.208	0.210	0.213
-0.6	0.274	0.266	0.262	0.280	0.280	0.268	0.268	0.274	0.273
-0.4	0.345	0.334	0.329	0.352	0.354	0.339	0.339	0.345	0.344
-0.2	0.421	0.406	0.402	0.430	0.430	0.417	0.413	0.424	0.422
0.0	0.500	0.486	0.481	0.509	0.509	0.501	0.496	0.504	0.501
0.2	0.579	0.568	0.562	0.588	0.590	0.579	0.574	0.581	0.580
0.4	0.655	0.647	0.643	0.666	0.664	0.653	0.651	0.658	0.656
0.6	0.726	0.720	0.715	0.732	0.733	0.726	0.723	0.728	0.729
0.8	0.788	0.787	0.784	0.791	0.791	0.788	0.788	0.791	0.793
1.0	0.841	0.841	0.842	0.841	0.841	0.841	0.842	0.841	0.846
1.4	0.919	0.925	0.928	0.918	0.917	0.920	0.922	0.920	0.919
2.0	0.977	0.983	0.985	0.972	0.972	0.979	0.979	0.975	0.975
2.5	0.994	0.997	0.997	0.991	0.991	0.994	0.995	0.993	0.993
3.0	0.999	0.999	1.000	0.998	0.998	0.999	0.999	0.998	0.998
Q(2.5)	-1.960	-2.046	-2.094	-1.896	-1.890	-1.966	-2.002	-1.936	-1.955
Q(25)	-0.674	-0.649	-0.640	-0.683	-0.689	-0.655	-0.658	-0.669	-0.672
Q(50)	0.000	0.032	0.049	-0.025	-0.023	-0.004	0.012	-0.008	-0.003
Q(75)	0.674	0.683	0.697	0.659	0.656	0.673	0.679	0.664	0.665
Q(97.5)	1.960	1.860	1.821	2.047	2.043	1.931	1.926	1.994	1.994
IQR	1.348	1.332	1.337	1.342	1.345	1.328	1.338	1.333	1.337
BIAS	—	-0.004	-0.013	-0.008	-0.008	0.000	-0.001	-0.001	-0.001
STD	—	0.083	0.159	0.071	0.116	0.026	0.049	0.022	0.037

Table 3: The distribution function of the ML estimator (cont.)

(iii) The case $A = 0.4,\ B = 0.8,\ \sigma_1 = 6.0,\ \sigma_2 = 2.0$

	Normal	$T=$ 100				$T=$ 1000			
x		$\hat{A}$	$\hat{B}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{A}$	$\hat{B}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$
-3.0	0.001	0.006	0.005	0.000	0.000	0.003	0.003	0.001	0.001
-2.5	0.006	0.014	0.013	0.002	0.003	0.009	0.009	0.005	0.006
-2.0	0.023	0.034	0.034	0.013	0.017	0.028	0.027	0.020	0.023
-1.4	0.081	0.088	0.088	0.070	0.079	0.085	0.086	0.077	0.079
-1.0	0.159	0.156	0.153	0.158	0.159	0.157	0.157	0.160	0.160
-0.8	0.212	0.201	0.201	0.217	0.214	0.208	0.204	0.213	0.217
-0.6	0.274	0.255	0.258	0.288	0.278	0.267	0.265	0.280	0.278
-0.4	0.345	0.318	0.321	0.361	0.351	0.334	0.331	0.351	0.346
-0.2	0.421	0.389	0.391	0.443	0.432	0.408	0.407	0.428	0.422
0.0	0.500	0.465	0.468	0.524	0.511	0.485	0.488	0.507	0.500
0.2	0.579	0.544	0.550	0.602	0.588	0.566	0.570	0.585	0.579
0.4	0.655	0.627	0.635	0.672	0.661	0.645	0.649	0.659	0.653
0.6	0.726	0.708	0.711	0.737	0.731	0.720	0.720	0.727	0.725
0.8	0.788	0.781	0.782	0.795	0.792	0.787	0.788	0.787	0.790
1.0	0.841	0.846	0.846	0.842	0.843	0.842	0.845	0.841	0.841
1.4	0.919	0.942	0.934	0.912	0.916	0.925	0.923	0.918	0.920
2.0	0.977	0.992	0.991	0.967	0.973	0.982	0.982	0.976	0.976
2.5	0.994	1.000	0.999	0.989	0.991	0.996	0.996	0.992	0.993
3.0	0.999	1.000	1.000	0.997	0.998	0.999	0.999	0.998	0.999
Q(2.5)	-1.960	-2.173	-2.176	-1.786	-1.887	-2.056	-2.048	-1.926	-1.960
Q(25)	-0.674	-0.617	-0.621	-0.701	-0.682	-0.654	-0.647	-0.686	-0.689
Q(50)	0.000	0.089	0.080	-0.057	-0.028	0.035	0.029	-0.018	0.001
Q(75)	0.674	0.715	0.705	0.647	0.656	0.688	0.681	0.673	0.673
Q(97.5)	1.960	1.692	1.738	2.109	2.044	1.884	1.885	1.987	1.967
IQR	1.348	1.331	1.325	1.348	1.339	1.342	1.328	1.359	1.362
BIAS	—	-0.027	-0.005	0.003	-0.018	-0.003	-0.001	-0.002	-0.002
STD	—	0.142	0.031	0.769	0.176	0.044	0.009	0.244	0.056

For this distribution the first four moments of u_t are

$$E(u_t) = 0, \;\; Var(u_t) = 1, \;\; Skewness(u_t) = 0 \text{ and } Kurtosis(u_t) = 4 + \frac{1}{3}.$$

Then we have made a comparison between the distributions of the ML estimator and the pseudo-ML estimator , which is defined as the ML estimator as if the disturbances were normally distributed. From Table 4 the pseudo-ML estimator based on the Gaussian likelihood does not have large biases in many cases when the distribution of u_t is not far from the normal distribution.

6. Concluding Remarks

In this paper, we have investigated the basic properties of two estimation methods in some details for the simultaneous switching autoregressive (SSSR) model,which is originally proposed by Kunitomo and Sato (1994a,b). In particular, we have investigated the finite sample as well as the asymptotic properties of the least squares estimator and the maximum likelihood estimator in the SSAR model.

First, the least squares estimator is asymptotically inconsistent. There is a simultaneity among the phases (or regimes) and the values of the states in the SSAR model, which makes the statistical estimation problem a non-trivial one. The current phase (or regime) is dependent upon not only the past values of states, but also the current unobservable disturbances in this model, which makes it different from the TAR models. From our limited number of simulations, the least squares estimator is fairly biased in many cases even when the sample size is not very large. Also the estimated signs of coefficients by the least squares method are often different from the true ones. These findings lead to the first conclusion that the least squares methods should not be used in our situation.

Second, we have shown that the ML estimator is asymptotically consistent and normally distributed. Also by our systematic simulations we have investigated the finite sample distributions of the ML estimator for the univariate SSAR(1) model. From the tables and figures of distributions and densities given in Section 5, we confirm that the ML estimator does not have any serious bias and its distribution can be well approximated by the normal distribution. We have found that this approximation is quite good especially when $|A| < 1$ and $|B| < 1$. Hence we have the second conclusion that the ML estimation method gives us reasonable estimation results for practical purposes. We expect that these properties of the ML estimator hold in more complicated SSAR models.

However, it should be noted that our results on the estimation of the SSAR model depends on the assumption for the distribution of disturbance terms. We have shown that the finite sample properties of the pseudo-ML estimator are not far from those of the ML estimator when the distribution of disturbances is a Gaussian sum in Section 5. However, the assumption on the distribution of disturbances could cause a serious consequence when the distributions of the disturbance terms are far from the nomal distribution. In this respect, we may need a further development in the statistical

Table 4: The mean of the pseudo-ML estimator

(i) T = 50

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7495	0.7829	0.7904	0.7926	0.7945	0.7957	0.7981
		(0.116)	(0.056)	(0.038)	(0.033)	(0.028)	(0.025)	(0.017)
	$\hat{B}$	0.7493	0.3909	0.0314	-0.2038	-0.4523	-0.8031	-2.0189
		(0.117)	(0.245)	(0.371)	(0.452)	(0.562)	(0.690)	(1.231)
A = 0.5	$\hat{A}$	0.3966	0.4583	0.4762	0.4834	0.4859	0.4899	0.4978
		(0.230)	(0.153)	(0.112)	(0.092)	(0.082)	(0.068)	(0.045)
	$\hat{B}$	0.7825	0.4579	0.1335	-0.0841	-0.3039	-0.6265	-1.7049
		(0.057)	(0.152)	(0.237)	(0.300)	(0.365)	(0.458)	(0.782)
A = 0.2	$\hat{A}$	0.0240	0.1310	0.1667	0.1755	0.1798	0.1883	0.1982
		(0.373)	(0.244)	(0.180)	(0.149)	(0.127)	(0.106)	(0.066)
	$\hat{B}$	0.7898	0.4792	0.1701	-0.0462	-0.2543	-0.5589	-1.5988
		(0.038)	(0.110)	(0.176)	(0.225)	(0.266)	(0.328)	(0.554)
A = 0.0	$\hat{A}$	-0.2010	-0.0833	-0.0474	-0.0313	-0.0188	-0.0049	0.0009
		(0.454)	(0.301)	(0.222)	(0.188)	(0.158)	(0.128)	(0.075)
	$\hat{B}$	0.7932	0.4845	0.1743	-0.0355	-0.2275	-0.5361	-1.5428
		(0.032)	(0.094)	(0.148)	(0.189)	(0.222)	(0.276)	(0.445)
A = -0.2	$\hat{A}$	-0.4407	-0.3005	-0.2515	-0.2313	-0.2141	-0.2075	-0.1949
		(0.553)	(0.368)	(0.264)	(0.221)	(0.184)	(0.145)	(0.078)
	$\hat{B}$	0.7940	0.4894	0.1842	-0.0151	-0.2154	-0.5155	-1.5052
		(0.028)	(0.081)	(0.127)	(0.158)	(0.186)	(0.227)	(0.342)
A = -0.5	$\hat{A}$	-0.8030	-0.6168	-0.5674	-0.5348	-0.5168	-0.5051	-0.4845
		(0.692)	(0.456)	(0.333)	(0.270)	(0.223)	(0.167)	(0.068)
	$\hat{B}$	0.7959	0.4924	0.1872	-0.0115	-0.2083	-0.5017	-1.4618
		(0.024)	(0.069)	(0.105)	(0.129)	(0.146)	(0.165)	(0.185)
A = -1.5	$\hat{A}$	-1.9950	-1.6677	-1.6052	-1.5496	-1.5015	-1.4629	NA
		(1.204)	(0.750)	(0.587)	(0.453)	(0.330)	(0.182)	(NA)
	$\hat{B}$	0.7978	0.4978	0.1984	0.0004	-0.1952	-0.4830	NA
		(0.017)	(0.046)	(0.066)	(0.075)	(0.077)	(0.069)	(NA)

Table 4: The mean of the pseudo-ML estimator (cont.)

(ii) T = 100

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7741	0.7910	0.7948	0.7953	0.7968	0.7979	0.7991
		(0.075)	(0.036)	(0.025)	(0.021)	(0.019)	(0.016)	(0.011)
	$\hat{B}$	0.7742	0.4266	0.0771	-0.1708	-0.3930	-0.7589	-1.9379
		(0.073)	(0.161)	(0.250)	(0.326)	(0.386)	(0.493)	(0.867)
A = 0.5	$\hat{A}$	0.4289	0.4779	0.4885	0.4938	0.4946	0.4963	0.4982
		(0.159)	(0.104)	(0.076)	(0.063)	(0.054)	(0.046)	(0.031)
	$\hat{B}$	0.7911	0.4775	0.1589	-0.0532	-0.2695	-0.5923	-1.6555
		(0.036)	(0.105)	(0.162)	(0.204)	(0.245)	(0.310)	(0.533)
A = 0.2	$\hat{A}$	0.0759	0.1547	0.1824	0.1879	0.1940	0.1942	0.2000
		(0.249)	(0.168)	(0.123)	(0.102)	(0.087)	(0.073)	(0.045)
	$\hat{B}$	0.7944	0.4896	0.1828	-0.0197	-0.2286	-0.5341	-1.5575
		(0.025)	(0.073)	(0.123)	(0.145)	(0.178)	(0.226)	(0.379)
A = 0.0	$\hat{A}$	-0.1578	-0.0574	-0.0270	-0.0155	-0.0082	-0.0059	0.0010
		(0.315)	(0.206)	(0.150)	(0.127)	(0.108)	(0.088)	(0.052)
	$\hat{B}$	0.7961	0.4931	0.1882	-0.0124	-0.2144	-0.5187	-1.5231
		(0.022)	(0.064)	(0.101)	(0.125)	(0.150)	(0.186)	(0.311)
A = -0.2	$\hat{A}$	-0.4031	-0.2746	-0.2260	-0.2142	-0.2091	-0.2044	-0.1992
		(0.389)	(0.246)	(0.179)	(0.149)	(0.126)	(0.100)	(0.054)
	$\hat{B}$	0.7972	0.4939	0.1951	-0.0090	-0.2081	-0.5068	-1.5034
		(0.018)	(0.054)	(0.088)	(0.106)	(0.127)	(0.152)	(0.230)
A = -0.5	$\hat{A}$	-0.7547	-0.5918	-0.5378	-0.5205	-0.5093	-0.5023	-0.4913
		(0.496)	(0.310)	(0.224)	(0.191)	(0.151)	(0.117)	(0.045)
	$\hat{B}$	0.7978	0.4973	0.1950	-0.0060	-0.2034	-0.4992	-1.4798
		(0.016)	(0.045)	(0.072)	(0.088)	(0.100)	(0.115)	(0.124)
A = -1.5	$\hat{A}$	-1.9115	-1.6634	-1.5655	-1.5222	-1.4908	-1.4794	NA
		(0.855)	(0.534)	(0.389)	(0.306)	(0.231)	(0.123)	(NA)
	$\hat{B}$	0.7987	0.4986	0.1993	0.0010	-0.1970	-0.4921	NA
		(0.012)	(0.031)	(0.046)	(0.051)	(0.054)	(0.044)	(NA)

Table 4: The mean of the pseudo-ML estimator (cont.)

(iii) T = 500

		B = 0.8	B = 0.5	B = 0.2	B = 0.0	B = -0.2	B = -0.5	B = -1.5
A = 0.8	$\hat{A}$	0.7945	0.7973	0.7982	0.7984	0.7985	0.7989	0.7994
		(0.029)	(0.015)	(0.011)	(0.009)	(0.008)	(0.007)	(0.005)
	$\hat{B}$	0.7946	0.4561	0.1073	-0.1271	-0.3593	-0.7136	-1.8784
		(0.029)	(0.075)	(0.133)	(0.174)	(0.214)	(0.282)	(0.500)
A = 0.5	$\hat{A}$	0.4545	0.4949	0.4995	0.4998	0.5002	0.4999	0.5000
		(0.076)	(0.043)	(0.032)	(0.027)	(0.024)	(0.020)	(0.013)
	$\hat{B}$	0.7969	0.4949	0.1815	-0.0320	-0.2431	-0.5573	-1.6170
		(0.015)	(0.043)	(0.072)	(0.091)	(0.112)	(0.140)	(0.247)
A = 0.2	$\hat{A}$	0.1055	0.1839	0.1960	0.1992	0.1994	0.2001	0.2005
		(0.134)	(0.070)	(0.053)	(0.044)	(0.038)	(0.031)	(0.020)
	$\hat{B}$	0.7978	0.4993	0.1952	-0.0061	-0.2110	-0.5203	-1.5310
		(0.011)	(0.032)	(0.052)	(0.066)	(0.078)	(0.099)	(0.163)
A = 0.0	$\hat{A}$	-0.1283	-0.0272	-0.0075	-0.0031	-0.0003	0.0000	0.0001
		(0.175)	(0.091)	(0.065)	(0.055)	(0.046)	(0.038)	(0.023)
	$\hat{B}$	0.7983	0.5005	0.1988	-0.0031	-0.2043	-0.5062	-1.5046
		(0.009)	(0.027)	(0.044)	(0.055)	(0.065)	(0.081)	(0.132)
A = -0.2	$\hat{A}$	-0.3578	-0.2413	-0.2115	-0.2048	-0.2035	-0.2007	-0.1999
		(0.215)	(0.110)	(0.078)	(0.065)	(0.055)	(0.044)	(0.024)
	$\hat{B}$	0.7986	0.5001	0.2001	-0.0006	-0.2023	-0.5016	-1.4919
		(0.008)	(0.024)	(0.038)	(0.047)	(0.055)	(0.068)	(0.103)
A = -0.5	$\hat{A}$	-0.7098	-0.5582	-0.5194	-0.5070	-0.5016	-0.5000	-0.4986
		(0.279)	(0.139)	(0.099)	(0.081)	(0.067)	(0.051)	(0.019)
	$\hat{B}$	0.7991	0.5004	0.1999	0.0000	-0.2004	-0.4997	-1.4937
		(0.007)	(0.020)	(0.032)	(0.038)	(0.044)	(0.051)	(0.053)
A = -1.5	$\hat{A}$	-1.8703	-1.6156	-1.5338	-1.5059	-1.4961	-1.4942	NA
		(0.494)	(0.249)	(0.166)	(0.135)	(0.103)	(0.053)	(NA)
	$\hat{B}$	0.7996	0.4999	0.2001	-0.0001	-0.1998	-0.4984	NA
		(0.005)	(0.013)	(0.020)	(0.023)	(0.024)	(0.019)	(NA)

estimation method of the SSAR model.

7. Appendix

For the proof of Theorem 2, the first preliminary result is taken from (5.19) of Kunitomo and Sato (1994b). Its proof is given in their Appendix.

Lemma 1 *The partial derivative of log-likelihood process* $\left\{\frac{\partial \log L_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$ *from (4.2) is a vector martingale process, that is,*

$$E\left[\frac{\partial \log L_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|\mathcal{F}_{t-1}\right] = \frac{\partial \log L_{t-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \ , \tag{A.1}$$

where $\mathcal{F}_{t-1}$ *is the* σ*-field generated by* $\{\boldsymbol{y}_s\,,\ s \leq t-1\}$.

The next two lemmas are the results of direct algebraic calculations by using the partitioned matrices. We omit their proofs.

Lemma 2 *Let a* $p \times p$ *positive definite matrix* $\boldsymbol{A}$ *be decomposed into* $(q+(p-q)) \times (q+(p-q))$ *submatrices* $\boldsymbol{A} = (\boldsymbol{A}_{ij})$. *For any* $q \times r$ *matrix* $\boldsymbol{B}$, $(p-q) \times r$ *matrix* $\boldsymbol{C}$ $(p-q>0)$, *and* $r \times r$ *matrix* D,

$$\min_{\boldsymbol{B}}\ tr\boldsymbol{A}\begin{pmatrix}\boldsymbol{B}\\ \boldsymbol{C}\end{pmatrix}D(\boldsymbol{B}',\boldsymbol{C}') = tr\ \boldsymbol{C}'(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})\boldsymbol{C}D \tag{A.2}$$

and the minimum occurs at $\boldsymbol{B} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{C}$.

Lemma 3 *Let a* $q \times q$ *positive definite matrix* $\boldsymbol{C}$ *be decomposed into* $(r+(q-r)) \times (r+(q-r))$ *submatrices* $\boldsymbol{C} = (\boldsymbol{C}_{ij})$. *For any* $p \times r$ *matrix* $\boldsymbol{A}$, $p \times (q-r)$ *matrix* $\boldsymbol{B}$ $(q-r>0)$,

$$\min_{\boldsymbol{B}}\ tr(\boldsymbol{A},\boldsymbol{B})\boldsymbol{C}\begin{pmatrix}\boldsymbol{A}'\\ \boldsymbol{B}'\end{pmatrix} = tr\ \boldsymbol{A}(\boldsymbol{C}_{11}-\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21})\boldsymbol{A}' \tag{A.3}$$

and the minimum occurs at $\boldsymbol{B} = -\boldsymbol{A}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}$.

Proof of Theorem 2 :

<u>*Step 1*</u> : We assume $\boldsymbol{\mu}_i = 0$ $(i=1,2)$ without loss of generality. Let $\boldsymbol{v}_t^{(i)} = (1/\sigma_i)\boldsymbol{D}_i\boldsymbol{u}_t$ $(i=1,2)$. Then the condition $\boldsymbol{e}_m'\boldsymbol{y}_t \geq \boldsymbol{e}_m'\boldsymbol{y}_{t-1}$ is equivalent to $v_{mt}^{(1)} \equiv (1/\sigma_1)\boldsymbol{e}_m'\boldsymbol{D}_1\boldsymbol{u}_t \geq \boldsymbol{r}'\boldsymbol{y}_{t-1}$. Let the indicator functions $I_t^{(i)}$ $(i=1,2)$ be defined by $I_t^{(1)} = I(v_{mt}^{(1)} \geq \boldsymbol{r}'\boldsymbol{y}_{t-1})$ and $I_t^{(2)} = I(v_{mt}^{(2)} < \boldsymbol{r}'\boldsymbol{y}_{t-1})$.

<u>*Step 2*</u> : We take the criterion function

$$Q_T(\boldsymbol{\theta}) = \sum_{i=1}^{2} Q_{iT}(\boldsymbol{\theta}) \tag{A.4}$$

where

$$\text{(A.5)} \qquad Q_{iT}(\boldsymbol{\theta}) = -\log|\boldsymbol{\Sigma}_i|\frac{1}{T}\sum_{t=1}^{T} I_t^{(i)} - \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{A}_i\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_t - \boldsymbol{A}_i\boldsymbol{y}_{t-1})I_t^{(i)} \ ,$$

and $\boldsymbol{A}_1 = \boldsymbol{A}$ and $\boldsymbol{A}_2 = \boldsymbol{B}$.
Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$ and we denote $\boldsymbol{\Sigma}_i^{(0)}$, $\boldsymbol{A}_i^{(0)}$, $\boldsymbol{r}^{(0)}$, $\sigma_i^{(0)}$ $(i = 1, 2)$ as the true value of the corresponding matrix or vector of parameters. Substituting (2.1) at the true parameter values for $\boldsymbol{y}_t$ in (A.5), we have

$$\text{(A.6)} \qquad Q_{iT}(\boldsymbol{\theta}) = -\log|\boldsymbol{\Sigma}_i|\frac{1}{T}\sum_{t=1}^{T} I_t^{(i)} - \frac{1}{T}\sum_{t=1}^{T}\left[(\boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i)\boldsymbol{y}_{t-1} + \sigma_i^{(0)}\boldsymbol{v}_t^{(i)}\right]'\boldsymbol{\Sigma}_i^{-1}\left[(\boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i)\boldsymbol{y}_{t-1} + \sigma_i^{(0)}\boldsymbol{v}_t^{(i)}\right] I_t^{(i)} \ .$$

Let also the moment matrices of $I_t^{(i)}$, $\boldsymbol{y}_{t-1}$ and $\boldsymbol{v}_t$ evaluated at the true value $\boldsymbol{\theta}_0$ be

$$\text{(A.7)} \qquad c^{(i)} = E\left[I_t^{(i)}\right] \ ,$$

$$\text{(A.8)} \qquad \boldsymbol{C}_{yy}^{(i)} = E\left[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}' I_t^{(i)}\right] \ ,$$

$$\text{(A.9)} \qquad \boldsymbol{C}_{vy}^{(i)} = E\left[\boldsymbol{v}_t\boldsymbol{y}_{t-1}' I_t^{(i)}\right] \ ,$$

and

$$\text{(A.10)} \qquad \boldsymbol{C}_{vv}^{(i)} = E\left[\boldsymbol{v}_t\boldsymbol{v}_t' I_t^{(i)}\right] \ .$$

Then $Q_{iT}(\boldsymbol{\theta})$ converges in L^1 to

$$\text{(A.11)} \qquad Q_i(\boldsymbol{\theta}) = - c^{(i)}\log|\boldsymbol{\Sigma}_i| - tr\left[(\boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i)\boldsymbol{C}_{yy}^{(i)}\right] - 2\sigma_i^{(0)} tr\left[(\boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i)'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{vy}^{(i)}\right] - {\sigma_i^{(0)}}^2 tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{vv}^{(i)}\right]$$

uniformly w.r.t. $\boldsymbol{\theta}$.
Now we want to show that $Q(\boldsymbol{\theta}) = \sum_{i=1}^{2} Q_i(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Since $Q_i(\boldsymbol{\theta})$ is a complicated function of $\boldsymbol{\theta}$, we shall do it in several steps.

First, we maximize $Q_i(\boldsymbol{\theta})$ with respect to $\boldsymbol{e}_j'\boldsymbol{A}_i$ $(j = 1, \ldots, m-1)$. We partition $\boldsymbol{A}_i^{(0)}$ and $\boldsymbol{A}_i$ into $((m-1)+1) \times m$ submatrices and

$$\text{(A.12)} \qquad \boldsymbol{A}_i^{(0)} - \boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{A}_{1i}^{(0)} - \boldsymbol{A}_{i1} \\ \sigma_i \boldsymbol{r}' - \sigma_i^{(0)}\boldsymbol{r}^{(0)\prime} \end{bmatrix} \ .$$

Using Lemma 2, we can maximize $Q_i(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{A}_{i1}$. Then we have

$$(A.13)\qquad Q_i(\boldsymbol{\theta}) \leq Q_i^*(\boldsymbol{\theta}) = -c^{(i)} \log|\boldsymbol{\Sigma}_i|$$
$$- \frac{1}{\sigma_i^2}\left[(\sigma_i \boldsymbol{r} - \sigma_i^{(0)}\boldsymbol{r}^{(0)})' - \sigma_i^{(0)}\boldsymbol{e}_m'\boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}\right]$$
$$\times \boldsymbol{C}_{yy}^{(i)}\left[(\sigma_i \boldsymbol{r} - \sigma_i^{(0)}\boldsymbol{r}^{(0)}) - \sigma_i^{(0)}\boldsymbol{C}_{yy}^{(i)^{-1}}\boldsymbol{C}_{yv}^{(i)}\boldsymbol{e}_m\right]$$
$$- \frac{1}{2}\left(\sigma_i^{(0)}\right)^2 tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{C}_{vv.y}^{(i)}\right] \quad ,$$

where $\boldsymbol{C}_{vv.y}^{(i)} = \boldsymbol{C}_{vv}^{(i)} - \boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}\boldsymbol{C}_{yv}^{(i)}$.
The equality in (A.13) holds when

$$(A.14)\qquad \boldsymbol{A}_{1i}^{(0)} - \boldsymbol{A}_{1i} - \sigma_i^{(0)} J_1' \boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}$$
$$= \frac{\sigma_{12}^{(i)}}{\sigma_i^2}\left[(\sigma_i \boldsymbol{r} - \sigma_i^{(0)}\boldsymbol{r}^{(0)})' - \sigma_i^{(0)}\boldsymbol{e}_m'\boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}\right] \quad ,$$

where $J_1' = (I_{m-1},\ 0)$.
Next, we partition an $m \times m$ matrix $\boldsymbol{\Sigma}_i^{-1}$ into $((m-1)+1)\times((m-1)+1)$ submatrices

$$(A.15)\qquad \boldsymbol{\Sigma}_i^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{(i)} & \sigma_{12}^{(i)} \\ \sigma_{21}^{(i)} & \sigma_i^2 \end{pmatrix}^{-1}$$
$$= \left(\frac{1}{\sigma_i^2}\right)\boldsymbol{e}_m\boldsymbol{e}_m' + \begin{pmatrix} I_{m-1}' \\ -\frac{\sigma_{21}^{(i)}}{\sigma_1^2} \end{pmatrix} \boldsymbol{\Sigma}_{11.2}^{(i)^{-1}} \left(I_{m-1}\ ,\ -\frac{\sigma_{21}^{(i)}}{\sigma_1^2}\right) \quad ,$$

where $\boldsymbol{\Sigma}_{11.2}^{(i)} = \boldsymbol{\Sigma}_{11}^{(i)} - \sigma_{12}^{(i)}\sigma_{21}^{(i)}/\sigma_i^2$.
We use $|\boldsymbol{\Sigma}_i| = |\boldsymbol{\Sigma}_{11.2}^{(i)}|\sigma_i^2$ and maximize $Q_i^*(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\Sigma}_{11.2}^{(i)}$ by using Lemma 2.6.1 of Anderson (1984). The resulting maximum occurs at

$$(A.16)\qquad \boldsymbol{\Sigma}_{11.2}^{(i)} = \frac{1}{c^{(i)}\sigma_i^{(0)2}}\left[I_{m-1},\ -\frac{\sigma_{12}^{(i)}}{\sigma_i^2}\right]\boldsymbol{C}_{vv.y}^{(i)}\begin{bmatrix} I_{m-1}' \\ -\frac{\sigma_{21}^{(i)}}{\sigma_i^2}\end{bmatrix} .$$

Also by using Lemma 3 we can maximize

$$(A.17)\qquad -\frac{c^{(i)}}{2}\log|\boldsymbol{\Sigma}_{11.2}^{(i)}|$$

w.r.t. $\sigma_{12}^{(i)}$. The resulting problem in this maximization is to investigate

$$(A.18)\qquad Q_i^{**}(\boldsymbol{\theta}) = -c^{(i)}\log\sigma_i^2$$

$$- \frac{1}{\sigma_i^2}\left[(\sigma_i \boldsymbol{r} - \sigma_i^{(0)}\boldsymbol{r}^{(0)})' - \sigma_i^{(0)}\boldsymbol{e}_m'\boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}\right]$$

$$\times \boldsymbol{C}_{yy}^{(i)}\left[(\sigma_i \boldsymbol{r} - \sigma_i^{(0)}\boldsymbol{r}^{(0)}) - \sigma_i^{(0)}\boldsymbol{C}_{yy}^{(i)^{-1}}\boldsymbol{C}_{yv}^{(i)}\boldsymbol{e}_m\right]$$

$$- \frac{\sigma_1^{(0)^2}}{\sigma_i^2}\boldsymbol{e}_m'\boldsymbol{C}_{vv.y}^{(i)}\boldsymbol{e}_m \quad .$$

Because $Q_i^{**}(\boldsymbol{\theta})$ is a concave function of σ_i, we differentiate it w.r.t. σ_i to obtain

(A.19)
$$\frac{\partial Q_i^{**}(\boldsymbol{\theta})}{\partial \sigma_i} = -2c^{(i)}\frac{1}{\sigma_i}$$

$$- \frac{2}{\sigma_i^2}\left[\sigma_i^{(0)}\boldsymbol{r}^{(0)'} + \sigma_i^{(0)}\boldsymbol{e}_m'\boldsymbol{C}_{vy}^{(i)}\boldsymbol{C}_{yy}^{(i)^{-1}}\right]\boldsymbol{C}_{yy}^{(i)}$$

$$\times \left[\left(\boldsymbol{r} - \frac{\sigma_i^{(0)}}{\sigma_i}\boldsymbol{r}^{(0)}\right) - \frac{\sigma_i^{(0)}}{\sigma_i}\boldsymbol{C}_{yy}^{(i)^{-1}}\boldsymbol{C}_{yv}^{(i)}\boldsymbol{e}_m\right]$$

$$+ 2\frac{\sigma_1^{(0)^2}}{\sigma_i^3}\boldsymbol{e}_m'\boldsymbol{C}_{vv.y}^{(i)}\boldsymbol{e}_m \quad .$$

By setting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have

(A.20)
$$\left.\frac{\partial Q_i^{**}(\boldsymbol{\theta})}{\partial \sigma_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \quad ,$$

because we can use the relation

(A.21)
$$\boldsymbol{e}_m'\boldsymbol{C}_{vv}^{(i)}\boldsymbol{e}_m - c^{(i)} - \boldsymbol{r}^{(0)'}\boldsymbol{C}_{yv}^{(i)}\boldsymbol{e}_m$$

$$= E\left\{\left[(v_{mt}^{(i)})^2 - 1 - (\boldsymbol{r}'\boldsymbol{y}_{t-1})v_{mt}^{(i)}\right] I_t^{(i)}\right\} = 0$$

by the assumption of normality on $\{\boldsymbol{u}_t\}$.
Finally, we note that

(A.22)
$$Q^{**}(\boldsymbol{\theta}) = \sum_{i=1}^{2} Q_i^{**}(\boldsymbol{\theta})$$

is a quadratic form of an $m \times 1$ vector $\boldsymbol{r}$. The maximum of (A.22) occurs at

(A.23)
$$\boldsymbol{r} - \frac{\sigma_i^{(0)}}{\sigma_i}\boldsymbol{r}^{(0)} = \sum_{i=1}^{2}\frac{\sigma_i^{(0)}}{\sigma_i}\boldsymbol{C}_{yv}^{(i)}\boldsymbol{e}_m$$

$$= \sum_{i=1}^{2}\frac{\sigma_i^{(0)}}{\sigma_i}E\left[\boldsymbol{y}_{t-1}\cdot v_{mt}^{(i)}I_t^{(i)}\right] \quad ,$$

which is zero at $\sigma_i = \sigma_i^{(0)}$.
This proves that $Q(\boldsymbol{\theta}_0)$ is the unique maximum of $Q(\boldsymbol{\theta})$. Then we apply Theorem 4.1.1

of Amemiya (1985) to show that $\hat{\boldsymbol{\theta}}_{ML}$ converges to $\boldsymbol{\theta}_0$ in probability as T goes to infinity.

Step 3 : Let θ_j be the j-th component of $\boldsymbol{\theta}$. From the log-likelihood function, we have

(A.24)
$$\frac{\partial \log L_T(\boldsymbol{\theta})}{\partial \theta_j} = \sum_{t=1}^{T}\sum_{i=1}^{2} \partial L_{it}(\theta_j) \quad ,$$

where

(A.25)
$$\begin{aligned}\partial L_{it}(\theta_j) \;=\; & -\frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma}_i|}{\partial \theta_j} I_t^{(i)} + \sigma_i \boldsymbol{y}_{t-1}' \left(\frac{\partial \boldsymbol{A}_i'}{\partial \theta_j}\right) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{v}_t I_t^{(i)} \\ & -\frac{\sigma_i^2}{2} tr\left\{\frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \theta_j} \boldsymbol{v}_t \boldsymbol{v}_t' I_t^{(i)}\right\} \quad .\end{aligned}$$

We also have

(A.26)
$$\frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} = \sum_{t=1}^{T}\sum_{i=1}^{2} \partial^2 L_{it}(\theta_j, \theta_k) \quad ,$$

where

(A.27)
$$\begin{aligned}\partial^2 L_{it}(\theta_j,\theta_k) = & -\frac{1}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}_i|}{\partial \theta_j \partial \theta_k} I_t^{(i)} - \boldsymbol{y}_{t-1}' \frac{\partial \boldsymbol{A}_i'}{\partial \theta_j} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{A}_i}{\partial \theta_k} \boldsymbol{y}_{t-1} I_t^{(i)} \\ + \; & \sigma_i \left[\boldsymbol{y}_{t-1}' \frac{\partial^2 \boldsymbol{A}_i'}{\partial \theta_j \partial \theta_k} \boldsymbol{\Sigma}_i^{-1} + \boldsymbol{y}_{t-1}' \frac{\partial \boldsymbol{A}_i'}{\partial \theta_j} \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \theta_k} + \boldsymbol{y}_{t-1}' \frac{\partial \boldsymbol{A}_i'}{\partial \theta_k} \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \theta_j}\right] \boldsymbol{v}_t I_t^{(i)} \\ - \; & \frac{\sigma_i^2}{2} \boldsymbol{v}_t' \frac{\partial^2 \boldsymbol{\Sigma}_i^{-1}}{\partial \theta_j \partial \theta_k} \boldsymbol{v}_t I_t^{(i)} \quad .\end{aligned}$$

Since $\{\boldsymbol{v}_t^{(i)}\}$ follows $N_m\left[0, (1/\sigma_i^2)\boldsymbol{\Sigma}_i\right]$, we have the following conditional expectation formulas

(A.28)
$$E\left[\boldsymbol{v}_t^{(i)} \;\middle|\; v_{mt}^{(i)}\right] = \frac{1}{\sigma_i^2} \boldsymbol{\Sigma}_i \boldsymbol{e}_m v_{mt}^{(i)} \quad ,$$

and

(A.29)
$$E\left[\boldsymbol{v}_t^{(i)} \boldsymbol{v}_t^{(i)\prime} \;\middle|\; v_{mt}^{(i)}\right] = \frac{1}{\sigma_i^2} \begin{pmatrix} \boldsymbol{\Sigma}_{11.2}^{(i)} & 0 \\ 0' & 0 \end{pmatrix} + \frac{1}{\sigma_i^4} \boldsymbol{\Sigma}_i \boldsymbol{e}_m \boldsymbol{e}_m' \boldsymbol{\Sigma}_i v_{mt}^{(i)2} \quad .$$

Then by using the repeated expectation formula, we have

(A.30)
$$\begin{aligned} & E\left[\partial^2 L_{it}(\theta_j,\theta_k) \;\middle|\; \mathcal{F}_{t-1}\right] \\ = \; & E\left\{E\left[\partial^2 L_{it}(\theta_j,\theta_k) \;\middle|\; v_{mt}^{(i)}\right] \;\middle|\; \mathcal{F}_{t-1}\right\} \\ = \; & \left[-\frac{1}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}_i|}{\partial \theta_j \partial \theta_k} - \boldsymbol{y}_{t-1}' \frac{\partial \boldsymbol{A}_i'}{\partial \theta_j} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{A}_i}{\partial \theta_k} \boldsymbol{y}_{t-1} - \frac{1}{2} tr\left(\frac{\partial^2 \boldsymbol{\Sigma}_i^{-1}}{\partial \theta_j \partial \theta_k} \boldsymbol{\Sigma}_i\right)\right] E\left[I_t^{(i)}\right]\end{aligned}$$

$$+\frac{1}{\sigma_i}\left[\boldsymbol{y}_{t-1}\frac{\partial^2 \boldsymbol{A}_i'}{\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{y}_{t-1}'\frac{\partial \boldsymbol{A}_i'}{\partial\theta_j}\frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial\theta_k}+\boldsymbol{y}_{t-1}'\frac{\partial \boldsymbol{A}_i'}{\partial\theta_k}\frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial\theta_j}\right]\boldsymbol{\Sigma}_i\boldsymbol{e}_m E\left[v_{mt}^{(i)}I_t^{(i)}\right]$$
$$-\frac{1}{2\sigma_i^2}tr\left[\frac{\partial^2\boldsymbol{\Sigma}_i^{-1}}{\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}_i\boldsymbol{e}_m\boldsymbol{e}_m'\boldsymbol{\Sigma}_i\right]E\left[(v_{mt}^{(i)2}-1)I_t^{(i)}\right].$$

Under the assumptions of Theorem 2, we need some tedious calculations to obtain the asymptotic covariance matrix. Then we have the $m \times m$ Fisher matrix

$$I(\boldsymbol{\theta}) = (I_{jk}) \quad , \tag{A.31}$$

where

$$\begin{aligned} I_{jk} &= E\left[-\sum_{i=1}^{2}\partial^2 L_{it}(\theta_j,\theta_k)\right] \\ &= E\left[\sum_{i=1}^{2}\partial L_{it}(\theta_j)\sum_{i=1}^{2}\partial L_{it}(\theta_k)\right] \quad . \end{aligned} \tag{A.32}$$

For illustration, we take $\theta_j = r_j$ and $\theta_k = \sigma_1$. Then in this case using $\frac{\partial^2 \log|\boldsymbol{\Sigma}_i|}{\partial\theta_j\partial\theta_k} = 0$, $\frac{\partial^2 \boldsymbol{A}_1}{\partial\theta_k\partial\theta_j} = -\boldsymbol{e}_m\boldsymbol{e}_j'$, $\frac{\partial^2\boldsymbol{\Sigma}_1^{-1}}{\partial\theta_j\partial\theta_k} = 0$, and $\frac{\partial\boldsymbol{\Sigma}_1^{-1}}{\partial\theta_k} = -2\sigma_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}_m\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}$,
we have

$$\begin{aligned} & E\left[\partial^2 L_{it}(\theta_j,\theta_k)\,\middle|\,\mathcal{F}_{t-1}\right] \\ &= -\left(\frac{1}{\sigma_1}\right)\left(\boldsymbol{e}_j'\boldsymbol{y}_{t-1}\right)\left(\boldsymbol{r}'\boldsymbol{y}_{t-1}\right)\sigma_1^2\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}_m E\left[I_t^{(1)}\right] \\ &\quad +\frac{1}{\sigma_1}\left(2\sigma_1^2\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}_m-1\right)E\left[\boldsymbol{v}_t^{(1)}I_t^{(1)}\right] \quad . \end{aligned} \tag{A.33}$$

On the other hand, we have

$$\begin{aligned} & E\left[\sum_{i=1}^{2}\partial L_{it}(\theta_j)\sum_{i=1}^{2}\partial L_{it}(\theta_k)\,\middle|\,\mathcal{F}_{t-1}\right] \\ &= E\left\{\left[-\boldsymbol{e}_j'\boldsymbol{y}_{t-1}\sum_{i=1}^{2}\sigma_i^2\boldsymbol{e}_m'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{v}_t^{(i)}I_t^{(i)}\right]\left[\partial L_{1t}(\theta_k)\right]\,\middle|\,\mathcal{F}_{t-1}\right\} \\ &= -\sigma_1^2\boldsymbol{e}_j'\boldsymbol{y}_{t-1}E\left\{\left(\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{v}_t\right)\left[\sigma_1^3\left(\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{v}_t^{(1)}\right)^2-\sigma_1\left(\boldsymbol{r}'\boldsymbol{y}_{t-1}\right)\left(\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{v}_t\right)\right.\right. \\ &\qquad \left.\left.-\sigma_1\boldsymbol{e}_m'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}_m\right]I_t^{(1)}|\mathcal{F}_{t-1}\right\} \\ &= -\frac{1}{\sigma_1}\boldsymbol{e}_j'\boldsymbol{y}_{t-1}E\left\{\left[w^3-(\boldsymbol{r}'\boldsymbol{y}_{t-1})w^2-dw\right]I_t^{(1)}\,\middle|\,\mathcal{F}_{t-1}\right\} \quad , \end{aligned} \tag{A.34}$$

where $w = \boldsymbol{e}'_m \boldsymbol{\Sigma}_1^{-1} \boldsymbol{v}_t^{(1)}$ and $d = \sigma_1^2 \boldsymbol{e}'_m \boldsymbol{\Sigma}_1^{-1} \boldsymbol{e}_m$.
Let $v = v_{mt}^{(1)}$. Then

$$(A.35) \qquad \begin{pmatrix} v \\ w \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & d \end{pmatrix}\right] ,$$

and we have $E\left[(w-v)^2 \mid v\right] = d-1$, $E\left[(w-v)^3 \mid v\right] = 0$. From these relations for the normal random variables, we can obtain the formulas

$$(A.36) \qquad E\left[w \mid v\right] = v ,$$

$$(A.37) \qquad E\left[w^2 \,\middle|\, v\right] = v^2 + d - 1 ,$$

$$(A.38) \qquad E\left[w^3 \,\middle|\, v\right] = 3v(d-1) + v^3 .$$

Hence the parenthesis part of (A.34) can be fuether re-written as

$$\begin{aligned} (A.39) \qquad & E\left\{E\left[w^3 - (\boldsymbol{r}'\boldsymbol{y}_{t-1})w^2 - dw \,\middle|\, v\right] I_t^{(1)}\right\} \\ = \; & E\left\{E\left[H_3(v) - (\boldsymbol{r}\boldsymbol{y}_{t-1})H_2(v) + 2dH_1(v) - d\boldsymbol{r}'\boldsymbol{y}_{t-1} \,\middle|\, v\right] I_t^{(1)}\right\} \\ = \; & E\left[(2d-1)\phi(\boldsymbol{r}'\boldsymbol{y}_{t-1}) - d\boldsymbol{r}'\boldsymbol{y}_{t-1} E\left(I_t^{(1)} | \mathcal{F}_{t-1}\right)\right] , \end{aligned}$$

where $H_j(v)$ are Hermite polynomials. ($H_3(v) = v^3 - 3v$, $H_2(v) = v^2 - 1$ and $H_1(v) = v$.) Hence we confirm that the minus of (A.33) and (A.34) are equivalent. Also we can show that each componemt of the minus of (A.33) and (A.34) are equivalent by tedious calculations. Under the assumptions of Theorem 2 we have

$$\begin{aligned} (A.40) \qquad I(\boldsymbol{\theta}) &= \lim_{T\to\infty} E\left[-\frac{1}{T}\frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \\ &= \lim_{T\to\infty} E\left[\frac{1}{T}\frac{\partial \log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right] . \end{aligned}$$

Finally, we shall apply the central limit theorem (CLT) to

$$(A.41) \qquad \frac{1}{\sqrt{T}}\frac{\partial \log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}_0} .$$

Since the partial derivatives of the log-likelihood function is a vector martingale process by Lemma 1 and the conditional Lindeberg condition is clearly satisfied under the assumptions we made in Theorem 2, we can use the martingale CLT developed by Dvoretzky (1972). By applying the standard arguments in the asymptotic theory (see Section 4 of Amemiya (1985), for instance), we have the desired result. (QED)

References

[1] Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell.

[2] Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis, 2nd. Edition*, John-Wiley.

[3] Dvoretzky, A. (1972), "Asymptotic Normality for Sums of Dependent Random Variables", *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **2**, University of California Press, 513-535.

[4] Granger, C. W. and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Vanderhoek and Ruprecht, Gottingen.

[5] Kunitomo, N. and Sato, S. (1994a), "A Non-linearity in Economic Time Series and Disequilibrium Econometric Models", in A. Takemura (ed.), *Theory and Application of Mathematical Statistics*, University of Tokyo Press, (In Japanese).

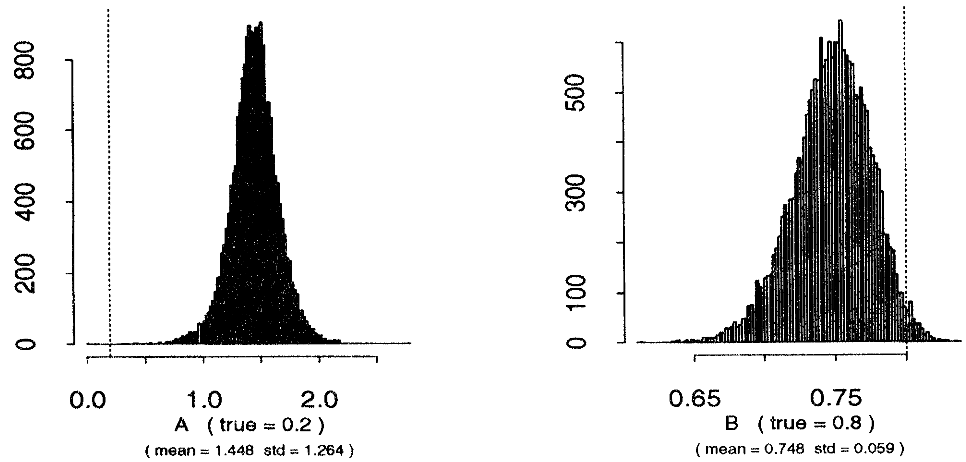
[6] Kunitomo, N. and Sato, S. (1994b), "Asymmetry in Economic Time Series and Simultaneous Switching Autoregressive Model", Discussion Paper No. 94-F-9, Faculty of Economics, University of Tokyo.

[7] Ozaki, T. and Oda, H. (1978), "Non-linear Time Series Model Identification by Akaike's Information Criterion", *Proceedings of IFAC Workshop on Information and Systems*, Campiegn, France.

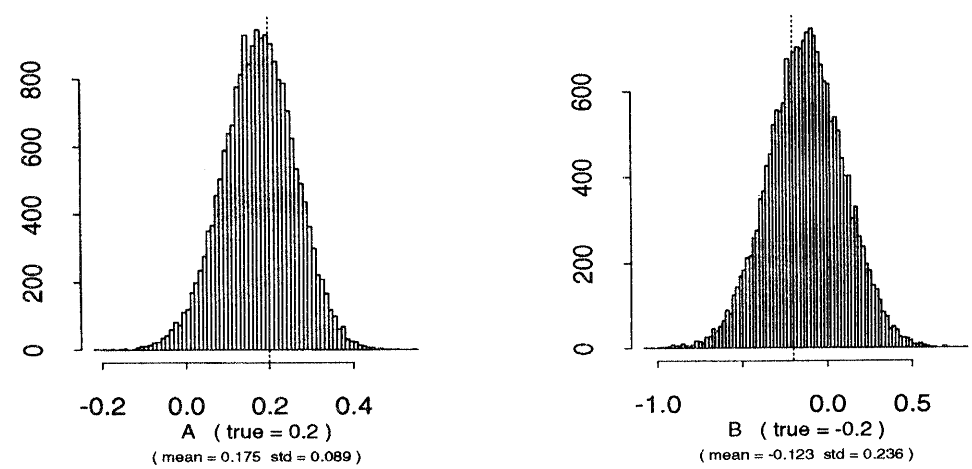
[8] Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*, Lecture Notes in Statistics No. 21, Springer, Heidelberg.

[9] Tong, H. (1990), *Non-linear Time Series*, Oxford.

Figure 1: The histogram of the least squares estimator



(i) $A = 0.2, \quad B = 0.8, \quad r = 1.0, \quad T = 100$



(ii) $A = 0.2, \quad B = -0.2, \quad r = 1.0, \quad T = 100$

Figure 2: The histogram of the ML estimator

$$A = 0.2, \quad B = 0.8, \quad r = 1.0, \quad T = 100$$

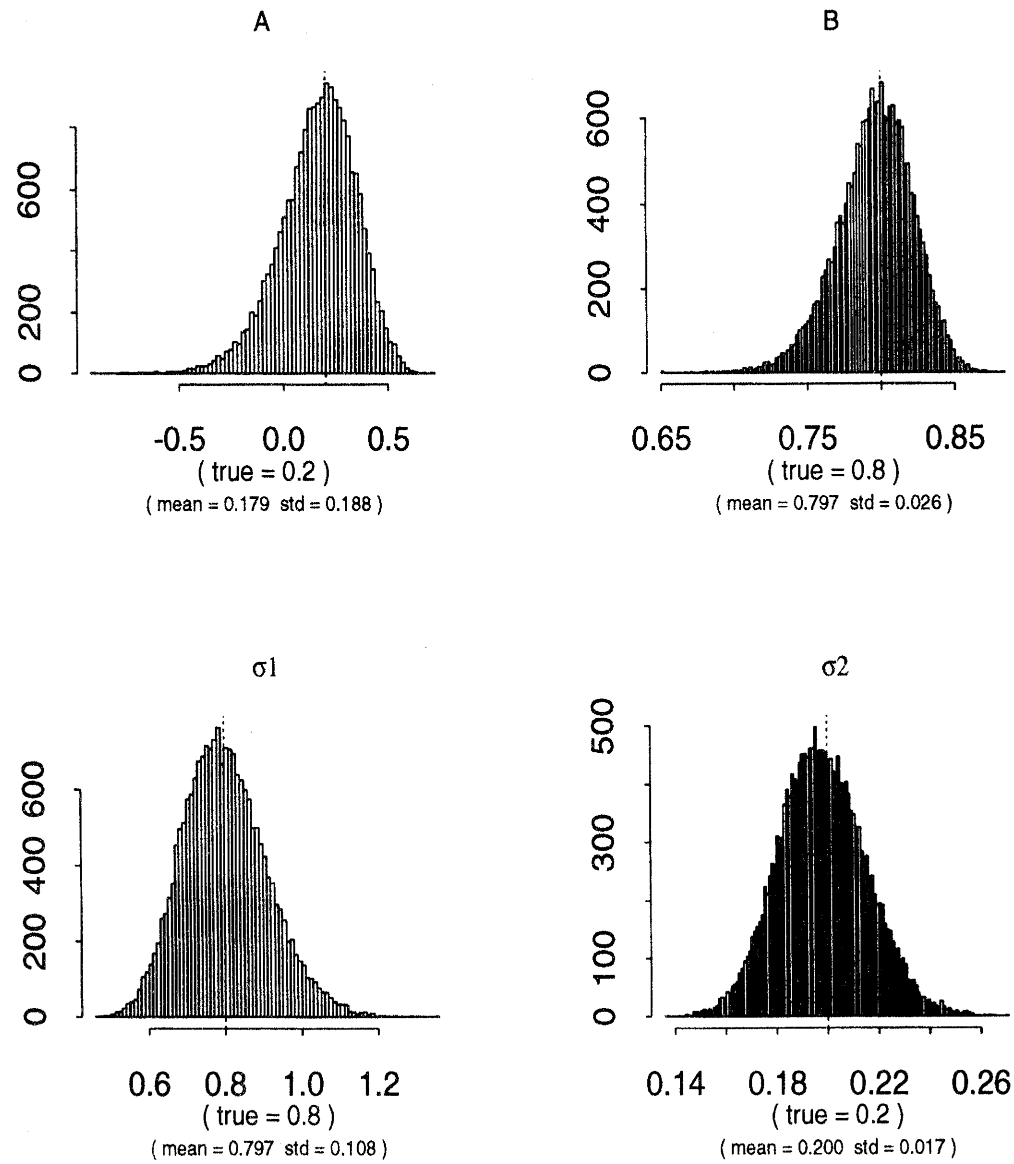


Figure 3: The distribution function of the ML estimator

$$A = 0.4, \quad B = 0.8, \quad \sigma_1 = 0.6, \quad \sigma_2 = 0.2$$