97-F-31

On Estimation of the Simultaneous Switching Autoregressive Models

Naoto Kunitomo
University of Tokyo

October 1997

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

On Estimation of the Simultaneous Switching Autoregressive Models *

by Naoto Kunitomo†

September 1997

Abstract

The simultaneous switching autoregressive (SSAR) model is a non-linear Markovian time series model, which was originally proposed by Kunitomo and Sato (1996) in *Structural Change and Economic Dynamics* and some of its statistical properties have been investigated by Sato and Kunitomo (1996) in *Journal of Time Series Analysis.* Since these papers have omitted some derivations of their theoretical results, this note gives more detailed expositions on them with additional technical remarks. We discuss some sufficient conditions for the geometric ergodicity of the SSAR model and the existence of moments. Also we give some sufficient conditions for the consistency and asymptotic normality of the maximum likelihood estimator for the unknown parameters in the SSAR models. Some corrections in the previous papers are given.

Key Words

Asymmetry, Non-linear Time Series, Simultaneous Switching Autoregressive (SSAR) Model, Geometric Ergodicity, Maximum Likelihood Estimation, Asymptotic Properties.

*This note is a revised and expanded version of a technical memorandum, which was originally prepared and written in August 1996 as a supplement to Sato and Kunitomo (1996) published in *Journal of Time Series Analysis.*

†Faculty of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113, JAPAN.

1. Introduction

Recently Kunitomo and Sato (1996) have introduced a simple stationary simultaneous switching autoregressive (SSAR) time series model. Let $\{y_t\}$ be a sequence of scalar time series satisfying

$$y_t = \begin{cases} Ay_{t-1} + \sigma_1 u_t & \text{if} \quad y_t \geq y_{t-1} \\ By_{t-1} + \sigma_2 u_t & \text{if} \quad y_t < y_{t-1} \end{cases} , \tag{1.1}$$

where A, B, σ_i $(\sigma_i > 0, i = 1, 2)$ are unknown parameters, and $\{u_t\}$ are a sequence of independently and identically distributed (i.i.d.) random variables with $E(u_t) = 0$ and $E(u_t^2) = 1$. By imposing the condition given by

$$\frac{1-A}{\sigma_1} = \frac{1-B}{\sigma_2} = r , \tag{1.2}$$

this time series model has the Markovian representation

$$y_t = y_{t-1} + [\sigma_1 I(u_t \geq ry_{t-1}) + \sigma_2 I(u_t < ry_{t-1})][-ry_{t-1} + u_t] , \tag{1.3}$$

where r is an unknown parameter and $I(\cdot)$ is the indicator function. When $\sigma_1 = \sigma_2 = \sigma$, then the SSAR model becomes the standard $AR(1)$ model by re-parametrizing $A = B = 1 - \sigma r$. As we have shown (Kunitomo and Sato (1996)), even this simplest univariate SSAR model (called SSAR(1)) gives us some explanations and descriptions on an important aspect of the asymmetrical movement of time series in two different (up-ward and down-ward) phases. The simple SSAR model has been introduced from an econometric application and there are some intuitive reasons why the SSAR models are useful for econometric applications. Also it should be noted that the SSAR time series models are different from the threshold autoregressive (TAR) models, which have been extensively discussed in the non-linear time series analysis. See Tong (1990) for the details of the TAR models.

More generally, let $\boldsymbol{y}_t$ be an $m \times 1$ vector of the endogenous variables. The SSAR model we consider in this note is represented by

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\mu}_1 + \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{D}_1\boldsymbol{u}_t & \text{if} \quad \boldsymbol{e}_1'\boldsymbol{y}_t \geq \boldsymbol{e}_1'\boldsymbol{y}_{t-1} \\ \boldsymbol{\mu}_2 + \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{D}_2\boldsymbol{u}_t & \text{if} \quad \boldsymbol{e}_1'\boldsymbol{y}_t < \boldsymbol{e}_1'\boldsymbol{y}_{t-1} \end{cases} , \tag{1.4}$$

where $\boldsymbol{e}_1' = (1, 0, \cdots, 0)$ and $\boldsymbol{\mu}_i'$ $(i = 1, 2)$ are $1 \times m$ vectors of constants, and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{D}_i$ $(i = 1, 2)$ are $m \times m$ matrices. We note that the condition $\boldsymbol{e}_1'\boldsymbol{y}_t \geq \boldsymbol{e}_1'\boldsymbol{y}_{t-1}$ in (1.4) has been used instead of the condition $\boldsymbol{e}_m'\boldsymbol{y}_t \geq \boldsymbol{e}_m'\boldsymbol{y}_{t-1}$ in Kunitomo and Sato (1996), for instance. This change in our formulation does not harm any essential argument below. The disturbance terms in (1.4) satisfy $E(\boldsymbol{u}_t) = 0$ and the variance-covariance matrix of $\boldsymbol{D}_i\boldsymbol{u}_t$ is denoted by $\boldsymbol{\Sigma}_i$ $(= \boldsymbol{D}_i\boldsymbol{D}_i', i = 1, 2)$.

We assume either

(a) $\{\boldsymbol{u}_t\}$ are absolutely continuous (mutually) independent random variables with the density function $g(\boldsymbol{u})$ which is continuous and everywhere positive in $\boldsymbol{R}^m$,

or
(b) $\boldsymbol{D}_i\boldsymbol{u}_t = \sigma_i\boldsymbol{e}_1 u_t$ and $\{u_t\}$ are absolutely continuous (mutually) independent random variables with the density function $g(u)$ which is continuous and everywhere positive in $\boldsymbol{R}$.
In the first case the disturbance terms $\{\boldsymbol{u}_t\}$ are distributed with $E(\boldsymbol{u}_t\boldsymbol{u}_t') = I_m$ and we assume that $\boldsymbol{\Sigma}_i$ $(i = 1,2)$ are positive definite matrices. In the second case the disturbance terms $\{u_t\}$ are distributed with $E(u_t^2) = 1$ and it corresponds to the Markovian representation of the univariate SSAR(p) model given by

$$(1.5) \qquad y_t = \begin{cases} a_0 + \sum_{j=1}^{p} a_j y_{t-j} + \sigma_1 u_t & \text{if} \quad y_t \geq y_{t-1} \\ \\ b_0 + \sum_{j=1}^{p} b_j y_{t-j} + \sigma_2 u_t & \text{if} \quad y_t < y_{t-1} \end{cases} ,$$

where $\{a_j\}$ and $\{b_j\}$ $(j = 0, \cdots, p)$ are unknown coefficients. This is because if we define $p \times 1$ vectors $\boldsymbol{y}_t$ and $\boldsymbol{\mu}_i$ $(i = 1,2)$ by

$$(1.6) \qquad \boldsymbol{y}_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} , \quad \boldsymbol{\mu}_1 = \begin{pmatrix} a_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} , \quad \boldsymbol{\mu}_2 = \begin{pmatrix} b_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

and $p \times p$ matrices

$$(1.7) \qquad \boldsymbol{A} = \begin{pmatrix} a_1 & \cdots & \cdots & a_p \\ 1 & & \boldsymbol{0} & \\ & \ddots & & \\ \boldsymbol{0} & & 1 & 0 \end{pmatrix} , \quad \boldsymbol{B} = \begin{pmatrix} b_1 & \cdots & \cdots & b_p \\ 1 & & \boldsymbol{0} & \\ & \ddots & & \\ \boldsymbol{0} & & 1 & 0 \end{pmatrix} ,$$

then the resulting model can be regarded as a special case of (1.4) if we set $\boldsymbol{D}_i\boldsymbol{u}_t = \sigma_i\boldsymbol{e}_1 u_t$ and $m = p$.

We note that in (1.4) 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 do not cause any logical inconsistency as a statistical model. This problem has been called the coherency problem and the condition for the logical consistency has been called the coherency condition. The conditions of $\boldsymbol{e}_1'\boldsymbol{y}_t \geq \boldsymbol{e}_1'\boldsymbol{y}_{t-1}$ and $\boldsymbol{e}_1'\boldsymbol{y}_t < \boldsymbol{e}_1'\boldsymbol{y}_{t-1}$ can be rewritten as

$$(1.8) \qquad \boldsymbol{e}_1'\boldsymbol{D}_1\boldsymbol{u}_t \geq \boldsymbol{e}_1'(I_m - \boldsymbol{A})\boldsymbol{y}_{t-1} - \boldsymbol{e}_1'\boldsymbol{\mu}_1 ,$$

and

$$(1.9) \qquad \boldsymbol{e}_1'\boldsymbol{D}_2\boldsymbol{u}_t < \boldsymbol{e}_1'(I_m - \boldsymbol{B})\boldsymbol{y}_{t-1} - \boldsymbol{e}_1'\boldsymbol{\mu}_2 ,$$

respectively. A set of the coherency conditions for (1.4) can be summarized by a $1 \times (m+1)$ vector of unknown parameters

$$(1.10) \qquad \frac{1}{\sigma_1}[\boldsymbol{e}_1'(I_m - \boldsymbol{A}),\ -\boldsymbol{e}_1'\boldsymbol{\mu}_1] = \frac{1}{\sigma_2}[\boldsymbol{e}_1'(I_m - \boldsymbol{B}),\ -\boldsymbol{e}_1'\boldsymbol{\mu}_2] = [\boldsymbol{r}', r_0] ,$$

where $\boldsymbol{r}'$ is a $1 \times m$ vector, r_0 is a scalar, and the scale parameters σ_j $(j = 1, 2)$ satisfy $\sigma_j^2 = \boldsymbol{e}_1'\boldsymbol{\Sigma}_j\boldsymbol{e}_1 = \boldsymbol{e}_1'\boldsymbol{D}_j\boldsymbol{D}_j'\boldsymbol{e}_1$. For the normalizaion of the scale parameters, we may use a $1 \times m$ vector

$$\frac{1}{\sigma_1}\boldsymbol{e}_1'\boldsymbol{D}_1 = \frac{1}{\sigma_2}\boldsymbol{e}_1'\boldsymbol{D}_2 = \boldsymbol{d}' \tag{1.11}$$

where we take $\boldsymbol{d}'\boldsymbol{d} = 1$. It is apparent from our formulation that the condition given by (1.11) is automatically satisfied for the $p-$th order univariate SSAR model.

In this note we shall give some derivations omitted in Kunitomo and Sato (1996), and Sato and Kunitomo (1996). We shall re-state some of their results and discuss some related theoretical results on the SSAR model. Some useful lemmas will be given in the Appendix.

2. Asymptotic Properties of the Maximum Likelihood Estimator

Kunitomo and Sato (1996) have proposed to use the maximum likelihood (ML) estimation for the SSAR models. Given the initial condition $\boldsymbol{y}_0$, the ML estimator is defined by maximizing the log-likelihood equation

$$\begin{aligned} \log L_T(\boldsymbol{\theta}) &= -\frac{mT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2} I_t^{(i)}\log(|\boldsymbol{\Sigma}_i|) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2}(\boldsymbol{y}_t - \boldsymbol{\mu}_i - \boldsymbol{A}_i\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_i - \boldsymbol{A}_i\boldsymbol{y}_{t-1})I_t^{(i)}, \end{aligned} \tag{2.1}$$

where $I_t^{(1)} = I(\boldsymbol{e}_1'\boldsymbol{y}_t \geq \boldsymbol{e}_1'\boldsymbol{y}_{t-1})$ and $I_t^{(2)} = I(\boldsymbol{e}_1'\boldsymbol{y}_t < \boldsymbol{e}_1'\boldsymbol{y}_{t-1})$ for the indicator function $I(\cdot)$. We note that the above maximization should be done by using the coherency condition given by (1.10). Sato and Kunitomo (1996) have stated that the ML estimator is consistent and asymptotically normally distributed. We restate their Theorem 2 in a slightly different way.

Theorem 1 : *For the SSAR(1) model given by (1.4), suppose (i) the sufficient conditions (1.10) for the coherency hold, (ii) a set of sufficient conditions for the geometric ergodicity hold as stated in one of the following lemmas (Lemma 1, 3, 4, or 6), (iii) the moments of initial conditions* $\boldsymbol{y}_0$ *exist up to 3, and (iv) either (a) the disturbances terms* $\{\boldsymbol{u}_t\}$ *are independent normal random variables* $N(0, I_m)$ *with* $|\boldsymbol{\Sigma}_i| \neq 0$ $(i = 1, 2)$*, or (b) the disturbances terms* $\{u_t\}$ *are independent normal random variables* $N(0, 1)$*. Also assume (v) the true parameter vector* $\boldsymbol{\theta}_0$ *is an interior point of the parameter space* $\boldsymbol{\Theta}$*.*
Then the ML estimator $\hat{\boldsymbol{\theta}}_{ML}$ *of unknown parameter* $\boldsymbol{\theta}$ *is consistent and asymptotically normally distributed as*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left[0,\ I(\boldsymbol{\theta}_0)^{-1}\right], \tag{2.2}$$

where

$$I(\boldsymbol{\theta}_0) = \lim_{T\to\infty}\frac{1}{T}\left[-\frac{\partial^2\log L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]. \tag{2.3}$$

Since we have omitted many detailed derivations in Kunitomo and Sato (1996), and Sato and Kunitomo (1996), this note covers the most essential issues omitted and could help understanding our results already reported. The results reported in our earlier papers hold with minor modifications.

Let us begin our discussions, first, by giving some technical comments including some corrections on the Appendix of Sato and Kunitomo (1996). The proof of our main theorem is a direct consequence of Lemmas stated in Section 3. For instance, the convergence in L_1 in Line 5 of Page 303 should be read as convergence a.s. because of the Ergodic Theorem for the Markov chain with the general state space. (See Chapter 17 of Meyn and Tweedie (1993), for instance.) Assuming the boundedness of second order moments on $\boldsymbol{y}_t = (y_{ti})$ $(i = 1, \cdots, m)$ when $m \geq 1$, we have (A.7) by applying the following Lemma 9 and the Ergodic Theorem. For instance, if we take $X_{t1} = v_t I(v_t \geq r^{(0)} y_{t-1})$ when $m = 1$, then $\mu_{t1} = \phi(r^{(0)} y_{t-1})$ ($r^{(0)}$ is the true value of r and $\phi(\cdot)$ is the density function of the standard normal random variable) and $\mathcal{F}_t$ is the $\sigma-$field generated by $\{v_1, \cdots, v_t, y_0, y_1, \cdots, y_t\}$ in Lemma 9 below. This is possibly the simplest case (and actually Lemma 9 is not even necessary), but the rest of our arguments for other terms appeared in (A.6) are essentially the same. The sufficient conditions for the existence of moments for the SSAR model in the general case can be the same as those for the geometric ergodicity; some of them shall be given as the following Lemma 1, Lemma 3, Lemma 4, and Lemma 7. (See Kunitomo and Sato (1996) for the related discussions.) The existence of higher order moments and their boundedness can be fully examined when $m = p = 1$. (See Lemma 1 in the following.)

Secondly, the parameter space $\boldsymbol{\Theta}$ should be restricted into a (sufficiently large) compact subset $\boldsymbol{\Theta}_1 \subset \boldsymbol{\Theta}$. This can be easily done by taking small $c_1^{(i)}$ $(i = 1, 2)$ and large $c_2^{(i)}$ $(i = 1, 2)$ such that $0 < c_1^{(i)} \leq \sigma_i \leq c_2^{(i)}$ $(i = 1, 2)$, for instance. Then we can apply Theorem 4.1.1 of Amemiya. Since both $Q_T(\boldsymbol{\theta})$ and $Q(\boldsymbol{\theta})$ are concave functions with respect to $\boldsymbol{\theta}$ and the existence of second moments under the assumptions, we can take $\boldsymbol{\Theta}_1$ such that

$$P(Q_T(\boldsymbol{\theta}_0) \leq \sup_{\boldsymbol{\theta} \notin \boldsymbol{\Theta}_1} Q_T(\boldsymbol{\theta}))$$

is arbitrary small. (See the discussions on Condition D in Section 4 of Amemiya (1985), for instance.) Hence we have the consistency of the maximum likelihood estimator if we can show the positive definiteness of the information matrix.

Thirdly, for the asymptotic normality of the ML estimator we need an additional condition that

$$\sup_{t \geq 1} E\left[|y_{tj} y_{tk} y_{tl}|\right] < +\infty \tag{2.4}$$

for $\boldsymbol{y}_t = (y_{ti})$ when $m > 1$. Again this condition can be easily checked under the necessary and sufficient conditions for the ergodicity when $m = 1$ and under a set of sufficient conditions when $m > 1$ because of the normal disturbances. Under this condition the asymptotic normality of $\hat{\boldsymbol{\theta}}_{ML}$ can be easily established under the present situation. We can expand the likelihood equation around the true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and

$$\frac{1}{\sqrt{T}} \frac{\partial \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{ML}} \tag{2.5}$$

$$= \frac{1}{\sqrt{T}} \frac{\partial \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{T} \frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \times \sqrt{T}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)$$
$$= 0 \ ,$$

where $\boldsymbol{\theta}^* = (\theta_k^*) = \boldsymbol{\theta}_0 + \delta_1(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0) \ \ (0 < |\delta_1| \leq 1)$.
We note that

$$\frac{1}{T} \frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \frac{1}{T} \frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{T} \sum_{k=1}^{K} \frac{\partial^3 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \partial \theta_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{**}} \times (\theta_k^* - \theta_{0k}) \tag{2.6}$$

where $\boldsymbol{\theta}^{**} = (\theta_k^{**}) = \boldsymbol{\theta}_0 + \delta_2(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)$ $(0 < |\delta_2| \leq 1)$, $\boldsymbol{\theta}_0 = (\theta_{0k})$, and K is the number of components of $\boldsymbol{\theta}$.
The expectation of third order derivatives

$$E\left[\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l}\bigg|_{\boldsymbol{\theta}^{**}} \right]$$

are bounded for $j, k, l = 1, \cdots, K$ because of (i) the functional form of their components in the integrands, and (ii) the boundedness of third order moments of (y_{ti}), where $l_t(\boldsymbol{\theta}) = \log L_t(\boldsymbol{\theta}) - \log L_{t-1}(\boldsymbol{\theta})$. By using the Ergodic theorem for the Markov chain and $\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 \xrightarrow{p} 0$, we have

$$-\frac{1}{T} \frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \xrightarrow{p} \boldsymbol{I}(\boldsymbol{\theta}_0) \tag{2.7}$$

as $T \longrightarrow +\infty$.

The rest of the proof is a standard routine. (See Basawa, Feigin, and Heyde (1976), or Chapter 4 of Amemiya (1985), for instance.) The positive definiteness of the information matrix can be proven for (a) the multivariate SSAR(1) model as stated in Lemma 5 and for (b) the p-th order univariate SSAR model as stated in Lemma 8.

3. Technical Details

3.1 Useful Results

When $m = p = 1$, we have the necessary and sufficient condition on the geometric erogodicity for the SSAR model. (See Theorem 5.2 in Kunitomo and Sato (1996).) It is also a sufficient condition on the existence of moments if we assume the existence of moments for the disturbance terms. For the sake of completeness and an illustration, we state this result and its proof in a formal way, whose method will be repeatedly used in more general cases.

Lemma 1 *In the SSAR model (1.4) when $m = p = 1$, assume (i) the coherency conditions, (ii) the necessary and sufficient conditions for the geometric ergodicity; that is*

$$A < 1, B < 1, AB < 1 \,, \tag{3.1}$$

(iii) $\sup_{t\geq 1} E[|u_t|^k] < +\infty$, *for any positive integer* $k \geq 1$, *and (iv)* $E[|y_0|^k] < +\infty$ *for any positive integer* $k \geq 1$. *Then*

$$\sup_{t\geq 1} E[|y_t|^k] < +\infty \,. \tag{3.2}$$

Proof : When $m = 1$, we can take the criterion function

$$g(x) = \begin{cases} k_1^k x^k + c_1 & x > 0 \\ k_2^k |x|^k + c_1 & x \leq 0 \end{cases} , \tag{3.3}$$

where k is any positive integer, and k_1, k_2, and c_1 are positive constants. Without loss of generality, we assume (3.2) for $k = 1$ and try to show (3.2) for $k = 2$. We notice that $E[|u_t y_t|] \leq c_2 E[u_t^2 + |u_t||y_{t-1}|]$ for some constant c_2. Then from our assumption (iii) with $k = 1$ and (3.2) with $k = 1$, we have

$$\sup_{t\geq 1} E[|u_t y_t|] < +\infty \,. \tag{3.4}$$

We first consider the case when $y_{t-1} = x > M > 0$. Then

$$\begin{aligned} E[G(y_t)|y_{t-1} = x] &\leq c_3 + c_4 x + k_1^2 A^2 x^2 P\{v_t \geq rx\} \\ &+ k_1^2 B^2 x^2 P\{(r - \frac{1}{\sigma_2})x < v_t < rx\} \\ &+ k_2^2 B^2 x^2 P\{v_t \leq (r - \frac{1}{\sigma_2})x\} \,, \end{aligned} \tag{3.5}$$

where $c_i (i = 3, 4)$ are positive constants. Because $A < 1, B < 1$, and $AB < 1$, we can take $k_1 > 0$ and $k_2 > 0$ such that $1 > A > -k_2/k_1$ and $1 > B > -k_1/k_2$ and then $k_2^k > (-A)^k k_1^k$ for $A \leq 0$ and $k_1^k > (-B)^k k_2^k$ for $B \leq 0$. We note that the conditions $B < 0$ and $0 \leq B < 1$ correspond to the cases when $1/\sigma_2 < r < 1/\sigma_2 + 1/\sigma_1$ and $0 < r \leq 1/\sigma_2$, respectively. When $0 < r \leq 1/\sigma_2$ $(0 \leq B < 1)$, the coefficients of third and fifth terms on the right-hand side of (3.5) can be small. Then by taking a sufficiently large M, we have

$$E[G(y_t)|y_{t-1} = x] \leq c_5(M) + \delta_1 k_1^2 x^2 \,, \tag{3.6}$$

where $0 < \delta_1 < 1$ and $c_5(M)$ is a positive constant depending M. When $1/\sigma_2 < r \leq 1/\sigma_2 + 1/\sigma_1$ $(B < 0)$, the coefficients of the third and fourth terms on the right-hand side of (3.5) can be small. Because $k_2^2 B^2 < k_1^2$ in this case and we can take a sufficiently large M, we also have $E[G(y_t)|y_{t-1} = x] \leq c_6(M) + \delta_2 k_1^2 x^2$, where $0 < \delta_2 < 1$ and $c_6(M)$ depending on M is a positive constant. By taking $\max\{\delta_1, \delta_2\} < \delta_3 < 1$, we have $E[G(y_t)|y_{t-1} = x] \leq c_7(M) + \delta_3 G(x)$, where $c_7(M)$ is a positive constant. We can also use the similar arguments for the case when $y_{t-1} = x < -M < 0$. Then we can take positive constants $0 < \delta < 1$ and $c_8(M)$ depending on M for any $y_{t-1} = x$ such that

$$E[G(y_t)|y_{t-1} = x] \leq c_8(M) + \delta G(x) \,. \tag{3.7}$$

Because

$$E[G(y_t)|y_0 = x] = E\{E[G(y_t)|y_{t-1}]|y_0\}$$
$$\leq c_8(M)[1 + \delta + \cdots + \delta^{t-1}] + \delta^t G(y_0) ,$$

(3.8)

is bounded, we have the desired result for $k = 2$. We can use the induction for any positive integer k. (Q.E.D.)

Lemma 2 *Under the assumptions we have made in Theorem 1,*

(3.9) $$C_{yy}^{(i)} = E\left[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}' I_t^{(i)} \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\right] \quad (i = 1, 2)$$

are positive definite matrices.

Proof : Take the case of $i = 1$. Suppose this is not the case. Then there exists an $m \times 1$ non-zero vector $\boldsymbol{\lambda}$ such that

(3.10) $$\boldsymbol{\lambda}' E\left[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}' I_t^{(1)}\right] \boldsymbol{\lambda} = E\left[\boldsymbol{\lambda}'\boldsymbol{y}_{t-1} I_t^{(1)}\right]^2 = 0 \ .$$

Then $(\boldsymbol{\lambda}'\boldsymbol{y}_{t-1})^2 I_t^{(1)} = 0$ (a.s.). By taking the conditional expectation w.r.t. $\mathcal{F}_{t-1}$ and using that the density function of $f(\boldsymbol{u}_t)$ is positive everywhere, we have $\boldsymbol{\lambda}'\boldsymbol{y}_{t-1} = 0$ (a.s.). Also by multiplying $I_t^{(i)}$ $(i = 1, 2)$,

(3.11) $$(\boldsymbol{\lambda}'\boldsymbol{y}_t)^2 = \sum_{i=1}^{2}(\boldsymbol{\lambda}'\boldsymbol{A}_i\boldsymbol{y}_{t-1} + \boldsymbol{\lambda}'\boldsymbol{D}_i\boldsymbol{u}_t)^2 I_t^{(i)} = 0 ,$$

and thus we have $\boldsymbol{\lambda}'\boldsymbol{A}_i\boldsymbol{y}_{t-1} + \boldsymbol{\lambda}'\boldsymbol{D}_i\boldsymbol{u}_t = 0$ $(a.s.)$, where $\boldsymbol{A}_1 = \boldsymbol{A}$ and $\boldsymbol{A}_2 = \boldsymbol{B}$. This condition contradicts that $\{\boldsymbol{u}_t\}$ are i.i.d. $N(\mathbf{o}, \boldsymbol{I}_m)$ random variables unless $\boldsymbol{\lambda} = 0$. For the Markov representation for the univariate $p-$th order SSAR model, the above condition also leads to a contradiction unless $\boldsymbol{\lambda} = 0$. (Q.E.D.)

We give some sufficient conditions for the existence of higher order moments and their boundedness when $m \geq 1$. All of them are sufficient, but they are often too strong and we do not necessarily need those conditions. (See the conditions in Lemma 1.) The proofs of the following lemmas are straightforward and so brief.

Lemma 3 *In the SSAR model when $m \geq 1$, assume (i) the coherency conditions, (ii) a sufficient condition for the geometric ergodicity* $0 \leq \rho_1 < 1$, *where*

$$\rho_1 = \max\{\lambda_{\max}(\boldsymbol{A}'\boldsymbol{A}), \lambda_{\max}(\boldsymbol{B}'\boldsymbol{B})\},$$

and $\lambda_{\max}(\boldsymbol{C})$ is the maximum characteristic root of $\boldsymbol{C}$ in its absolute value,
(iii) $\sup_{t\geq 1} E[\|\boldsymbol{u}_t\|^k] < +\infty$ for any positive integer $k \geq 1$, and (iv) $E[\|\boldsymbol{y}_0\|^k] < +\infty$ for any positive integer $k \geq 1$. Then

(3.12) $$\sup_{t\geq 1} E[\|\boldsymbol{y}_t\|^k] < +\infty \ .$$

Proof : When $m \geq 1$, we can take the criterion function

$$G(\boldsymbol{x}) = \|\boldsymbol{x}\|^k , \tag{3.13}$$

where $\boldsymbol{x} = (x_i)$. Without loss of generality, we only show (13.12) for $k = 1$. When $k = 1$, we have

$$\begin{aligned} E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1} = \boldsymbol{x}] &\leq E[\|\boldsymbol{A}(t)\|]\|\boldsymbol{x}\| + E[\|\boldsymbol{D}(t)\boldsymbol{u}_t\|] \\ &\leq c + \sqrt{\rho_1}G(\boldsymbol{x}) , \end{aligned} \tag{3.14}$$

where $\boldsymbol{A}(t) = \boldsymbol{A}I_t^{(1)} + \boldsymbol{B}I_t^{(2)}$ and c is a positive constant. The rest of arguments and those for $k \geq 2$ are essentially the same as the proof of Lemma 1. (Q.E.D.)

Lemma 4 *In the SSAR model when $m \geq 1$, assume (i) the coherency conditions, (ii) a sufficient condition for the geometric ergodicity $0 \leq \rho_2 < 1$ or $0 \leq \rho_3 < 1$, where*

$$\rho_2 = \max_{1 \leq j \leq m} \{\sum_{i=1}^{m} |a_{ij}|, \sum_{i=1}^{m} |b_{ij}|\},$$

and

$$\rho_3 = \max_{1 \leq i \leq m} \{\sum_{j=1}^{m} |a_{ij}|, \sum_{j=1}^{m} |b_{ij}|\},$$

(iii) $\sup_{t \geq 1} E[\|\boldsymbol{u}_t\|^k] < +\infty$, for any positive integer $k \geq 1$, and (iv) $E[\|\boldsymbol{y}_0\|^k] < +\infty$ for any positive integer $k \geq 1$. Then

$$\sup_{t \geq 1} E[\|\boldsymbol{y}_t\|^k] < +\infty . \tag{3.15}$$

Proof : For $\boldsymbol{x} = (x_i)$, we take the criterion function

$$G(\boldsymbol{x}) = (\sum_{i=1}^{m} |x_i|)^k \tag{3.16}$$

for the first condition in (ii) and

$$G(\boldsymbol{x}) = (\max_{i=1,\cdots,m} |x_i|)^k \tag{3.17}$$

for the second condition in (ii), respectively. Then we use the same arguments as the proofs of Lemma 1 and Lemma 3. (Q.E.D.)

For the proof of consistency and asymptotic normality of the ML estimator, we need the condition that the information matrix evaluated at the true parameter values is non-singular. We first give the result for the multivariate SSAR model.

Lemma 5 *Suppose the assumptions we made in Theorem 1 hold for the SSAR model given by (1.4) with $\boldsymbol{\Sigma}_i$ ($i = 1, 2$) being positive definite. Then the information matrix $I(\boldsymbol{\theta}_0)$ in (2.3) is non-singular.*

Proof : We define a function

$$
\begin{aligned}
(3.18)\quad Q(\boldsymbol{\theta}) &= -\frac{1}{2}\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\sum_{i=1}^{2} I_t^{(i)}\log(|\boldsymbol{\Sigma}_i|) \\
&\quad -\frac{1}{2}\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\sum_{i=1}^{2}(\boldsymbol{y}_t-\boldsymbol{\mu}_i-\boldsymbol{A}_i\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_i-\boldsymbol{A}_i\boldsymbol{y}_{t-1})I_t^{(i)} \\
&= -\frac{1}{2}\sum_{i=1}^{2}E[I_t^{(i)}]\log(|\boldsymbol{\Sigma}_i|) \\
&\quad -\frac{1}{2}\sum_{i=1}^{2}E[(\boldsymbol{y}_t-\boldsymbol{\mu}_i-\boldsymbol{A}_i\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_i-\boldsymbol{A}_i\boldsymbol{y}_{t-1})I_t^{(i)}]\,,
\end{aligned}
$$

where the expectations are taken at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. We note that we have the above expression due to the Ergodic theorem for the Markov chain and the existence of moments under the assumptions of Theorem 1. Then the information matrix for $\boldsymbol{\theta}_0$ can be partitioned as

$$
(3.19)\qquad I(\boldsymbol{\theta}_0)=\left(\begin{array}{cc} -\dfrac{\partial^2 Q}{\partial\boldsymbol{\theta}^{(1)}\partial\boldsymbol{\theta}^{(1)'}} & -\dfrac{\partial^2 Q}{\partial\boldsymbol{\theta}^{(1)}\partial\boldsymbol{\theta}^{(2)'}} \\ -\dfrac{\partial^2 Q}{\partial\boldsymbol{\theta}^{(2)}\partial\boldsymbol{\theta}^{(1)'}} & -\dfrac{\partial^2 Q}{\partial\boldsymbol{\theta}^{(2)}\partial\boldsymbol{\theta}^{(2)'}} \end{array}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},
$$

where

$$
\boldsymbol{\theta}^{(1)} = vec\left(\begin{array}{cc} \boldsymbol{\mu}_{11} & \boldsymbol{A}_{11} \\ \boldsymbol{\mu}_{12} & \boldsymbol{A}_{12} \\ r_\mu & \boldsymbol{r}' \end{array}\right),
$$

with $\boldsymbol{\mu}_{11}$ and $\boldsymbol{\mu}_{12}$ being $(m-1)\times 1$ lower sub-vectors of $\boldsymbol{\mu}_i$ $(i=1,2)$, $\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{12}$ being $(m-1)\times m$ lower sub-matrices of $\boldsymbol{A}_{1i}$ $(i=1,2)$, and $\boldsymbol{\theta}^{(2)}=(vech(\boldsymbol{\Sigma}_1),vech(\boldsymbol{\Sigma}_2))'$.

Let us re-parametrize the variance-covariance matrices $\boldsymbol{\Sigma}_i$ $(i=1,2)$ such that

$$
\boldsymbol{\Sigma}_i=\left(\begin{array}{cc} \sigma_i & O \\ O & \boldsymbol{I}_{m-1}\end{array}\right)\boldsymbol{\Omega}_i^{-1}\left(\begin{array}{cc} \sigma_i & O \\ O & \boldsymbol{I}_{m-1}\end{array}\right).
$$

Then we have

$$
\boldsymbol{\Sigma}_i^{-1}=\left(\begin{array}{cc} \sigma_i^{-1} & O \\ O & \boldsymbol{I}_{m-1}\end{array}\right)\boldsymbol{\Omega}_i\left(\begin{array}{cc} \sigma_i^{-1} & O \\ O & \boldsymbol{I}_{m-1}\end{array}\right).
$$

By a result of straightforward calculations, we have a representation of the upper-left corner of the information matrix as

$$
(3.20)\qquad I(\boldsymbol{\theta}^{(1)})=E[\boldsymbol{C}_t\otimes\left(\begin{array}{c}1\\ \Delta\boldsymbol{y}_{t-1}\end{array}\right)\left(\begin{array}{c}1\\ \Delta\boldsymbol{y}_{t-1}\end{array}\right)'],
$$

where

$$
(3.21)\qquad \boldsymbol{C}_t=\left(\begin{array}{ccc} (\boldsymbol{\Omega}_1)_{22}I_t^{(1)} & O & -(\boldsymbol{\Omega}_1)_{21}I_t^{(1)} \\ O & (\boldsymbol{\Omega}_2)_{22}I_t^{(2)} & -(\boldsymbol{\Omega}_2)_{21}I_t^{(2)} \\ -(\boldsymbol{\Omega}_1)_{12}I_t^{(1)} & -(\boldsymbol{\Omega}_2)_{12}I_t^{(2)} & (\boldsymbol{\Omega}_1)_{11}I_t^{(1)}+(\boldsymbol{\Omega}_2)_{11}I_t^{(2)} \end{array}\right)
$$

and we have partitioned $m \times m$ matrices $\boldsymbol{\Omega}_i (i = 1,2)$ into $[1 + (m-1)] \times [1 + (m-1)]$ submatrices

$$\boldsymbol{\Omega}_i = \begin{pmatrix} (\boldsymbol{\Omega}_i)_{11} & (\boldsymbol{\Omega}_i)_{12} \\ (\boldsymbol{\Omega}_i)_{21} & (\boldsymbol{\Omega}_i)_{22} \end{pmatrix} .$$

By taking the conditional expectation of the first term of (3.20) given $\mathcal{F}_{t-1}$, its determinant can be written as

$$|E_{t-1}[\boldsymbol{C}_t]| = |(\boldsymbol{\Omega}_1)_{22.1}||(\boldsymbol{\Omega}_2)_{22.1}||E_{t-1}[I_t^{(1)}]||E_{t-1}[I_t^{(2)}]| \times |\sum_{i=1}^{2}(\boldsymbol{\Omega}_i)_{11} E_{t-1}[I_t^{(i)}]| , \tag{3.22}$$

where $E_{t-1}(\cdot)$ is the conditional expectation given $\mathcal{F}_{t-1}$ and

$$(\boldsymbol{\Omega}_i)_{22.1} = (\boldsymbol{\Omega}_i)_{22} - (\boldsymbol{\Omega}_i)_{21}(\boldsymbol{\Omega}_i)_{11}^{-1}(\boldsymbol{\Omega}_i)_{12}$$

for $i = 1,2$. Since (3.22) is positive a.s., the matrix $\boldsymbol{C}_t$ is positive definite a.s.. Hence (3.20) is positive definite because the matrix

$$E[E[\boldsymbol{C}_t|\mathcal{F}_{t-1}] \otimes \begin{pmatrix} 1 \\ \Delta \boldsymbol{y}_{t-1} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta \boldsymbol{y}_{t-1} \end{pmatrix}^{'}]$$

is positive definite. (If it were not the case, then we can lead to a contradition because there exists a non-zero vector $\boldsymbol{\lambda}$ such that $(1, \Delta \boldsymbol{y}_{t-1}^{'})\boldsymbol{\lambda} = 0 \; a.s.$.) Then by using Lemma 10 and by applying Lemma 11 to $\sum_{i=1}^{2} \log|\boldsymbol{\Omega}_i|\sigma_i^{-2} I_t^{(i)}$ for $\boldsymbol{\theta}^{'} = (\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'})$, the informatiom matrix $I(\boldsymbol{\theta}_0)$ is positive definite. (Q.E.D.)

3.2 The SSAR(p) Model

We give some sufficient conditions for the geometric ergodicity and existence of higher order moments for the $p-$th order SSAR model. We need some special consideration because the disturbance term in the Markovian representation is degenerate in a sense. The conditions we give are sufficient, but often too strong and we do not necessarily need those conditions.

Lemma 6 *In the $p-$th order SSAR model when $m = 1$*

$$y_t = \begin{cases} a_0 + \sum_{j=1}^{p} a_j y_{t-j} + \sigma_1 u_t & y_t \geq y_{t-1} \\ b_0 + \sum_{j=1}^{p} b_j y_{t-j} + \sigma_2 u_t & y_t < y_{t-1} \end{cases} , \tag{3.23}$$

assume (i) the coherency conditions

$$\begin{aligned} r_0 &= -\frac{a_0}{\sigma_1} = -\frac{b_0}{\sigma_2} , \\ r_1 &= \frac{1 - a_1}{\sigma_1} = \frac{1 - b_1}{\sigma_2} , \\ r_j &= -\frac{a_j}{\sigma_1} = -\frac{b_j}{\sigma_2} \; (j = 2, \cdots, p) , \end{aligned} \tag{3.24}$$

(ii) a sufficient condition $0 \le \rho_4 < 1$, *where*

$$\rho_4 = \max\{\sum_{j=1}^{p} |a_j|, \sum_{j=1}^{p} |b_j|\},$$

(iii) u_t *has an absolutely continuous distribution with respect to the Lebesgue measure on* $\boldsymbol{R}$*, and its density function* $f(u)$ *is continuous and strictly positive almost everywhere, and (iv)* $\sup_{t\ge1} E[|u_t|] < +\infty$. *Then* $\{y_t\}$ *is geometrically ergodic.*

Proof :
(**Step** 1) We apply the method used in Chan and Tong (1985) for the threshold autoregressive models with minor modifications. Under the assumptions ($0 \le \rho_4 < 1$) we can take $\xi_1 > \xi_2 > \cdots > \xi_m > 0$ ($m = p$) and θ such that

$$1 > \theta > \max\{\sum_{j=1}^{p} |a_j|\frac{\xi_1}{\xi_j}, \sum_{j=1}^{p} |b_j|\frac{\xi_1}{\xi_j}\} \tag{3.25}$$

and $\theta > \xi_{j+1}/\xi_j$ ($j = 1, \cdots, p-1$). We take the criterion function

$$G(\boldsymbol{x}) = 1 + \max_{1\le j\le p} |x_j|\xi_j \,. \tag{3.26}$$

Let a vector process $\boldsymbol{y}_t' = (y_t, y_{t-1}, \cdots, y_{t-p+1})$ and consider a Markovian representation for $\{\boldsymbol{y}_t\}$. Then it is straightforward to show

$$\begin{aligned} &E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1} = \boldsymbol{x}] \\ \le\;& c_1 + E[\max\{\sum_{j=1}^{p} |A_j(t)||y_{t-j}|\xi_1, |x_1|\xi_2, \cdots, |x_{p-1}|\xi_p\}|\boldsymbol{y}_{t-1} = \boldsymbol{x}] \\ \le\;& c_2 + E[\max\{\sum_{j=1}^{p} |A_j(t)||y_{t-j}|\xi_1, \theta|x_1|\xi_1, \cdots, \theta|x_{p-1}|\xi_{p-1}\}|\boldsymbol{y}_{t-1} = \boldsymbol{x}] \\ \le\;& c_3 + \max\{[\max\{\sum_{j=1}^{p} |a_j|\frac{\xi_1}{\xi_j}, \sum_{j=1}^{p} |b_j|\frac{\xi_1}{\xi_j}\}] \\ &\times[\max\{|x_1|\xi_1, \cdots, |x_p|\xi_p\}], \theta|x_1|\xi_1, \cdots, \theta|x_{p-1}|\xi_{p-1}\} \\ \le\;& c_4 + \theta G(\boldsymbol{x}), \end{aligned} \tag{3.27}$$

where $A_j(t) = a_j I_t^{(1)} + b_j I_t^{(2)}$ and $c_i (i = 1, \cdots, 4)$ are some positive constants.

(**Step** 2) Without loss of generality, we only consider $p = m = 2$ and take a set $\boldsymbol{A} = (a_1, b_1) \times (a_2, b_2) \in \boldsymbol{R}^2$ with $a_i < b_i (i = 1, 2)$. Define a function

$$\sigma(x - y) = (x - y)[\frac{1}{\sigma_1} 1_{\{x-y\ge0\}} + \frac{1}{\sigma_2} 1_{\{x-y<0\}}] \tag{3.28}$$

for $x - y$. Then for $\boldsymbol{x}' = (x_1, x_2)$

$$P(\boldsymbol{x}, \boldsymbol{A}) = \int_{\sigma(a_1-x_1)}^{\sigma(b_1-x_1)} f[u - (r_0, r_1, r_2)\begin{pmatrix}1\\ x_1\\ x_2\end{pmatrix}]du \,. \tag{3.29}$$

Also we have

$$(3.30) \qquad P^2(\boldsymbol{x},\boldsymbol{A}) = \int_{\sigma(a_2-x_1)}^{\sigma(b_2-x_1)}\{\int_{\sigma(a_1-y_2)}^{\sigma(b_1-y_2)} f[y_2-(r_0,r_1,r_2)\begin{pmatrix}1\\ x_1\\ x_2\end{pmatrix}] \times f[y_1-(r_0,r_1,r_2)\begin{pmatrix}1\\ y_2\\ x_1\end{pmatrix}]dy_1\}dy_2 .$$

By using the assumptions in (iii), we have

$$(3.31) \qquad \inf_{\boldsymbol{x}\in\boldsymbol{K}} P^m(\boldsymbol{x},\boldsymbol{A}) > 0$$

for $m=2$ and a compact set $\boldsymbol{K}$ in $\boldsymbol{R}^2$.

(Step 3) From Step 1, we can take $d>0, B>0$, and $0\le\delta<1$ such that
(i) $E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1}=\boldsymbol{x}] < B < \infty$ for $\|\boldsymbol{x}\|\le d$
and
(ii) $E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1}=\boldsymbol{x}] < \delta G(\boldsymbol{x})$ for $\|\boldsymbol{x}\| > d$.
Because the Markov chain for $\{\boldsymbol{y}_t\}$ is aperiodic and $\phi-$irreducible, we apply Theorem 4 in Tweedie (1983) and we can establish that $\{y_t\}$ is geometrically ergodic. (Q.E.D.)

Lemma 7 *In the p−th order univariate SSAR model (3.23) assume (i) the coherency conditions (2.24), (ii) a sufficient condition for the geometric ergodicity $\rho_4<1$, (iii) $\sup_{t\ge1}E[|u_t|^k]<+\infty$ for any positive integer $k\ge1$, and (iv) $E[|y_0|^k]<+\infty$ for any positive integer $k\ge1$. Then*

$$(3.32) \qquad \sup_{t\ge1} E[|y_t|^k] < +\infty .$$

Proof : The method of proof is similar to the first part of the proof of Lemma 6. We take the criterion function

$$(3.33) \qquad G(\boldsymbol{x}) = 1 + (\max_{1\le j\le p}|x_j|\xi_j)^k$$

for $\boldsymbol{x}=(x_i)$, where ξ_j $(j=1,\cdots,p)$ are defined as in the proof of Lemma 6. Then we consider the Markovian representation for $\boldsymbol{y}_t^{'}=(y_t,y_{t-1},\cdots,y_{t-p+1})$. For $k\ge1$, we have

$$(3.34) \qquad E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1}=\boldsymbol{x}] \le c+\theta G(\boldsymbol{x}),$$

for some positive c and $0<\theta<1$. The rest of our arguments is the same as the proof of Lemma 1. (Q.E.D.)

We should mention again that the above conditions given in this note are quite strong and sufficient, but they are not necessary and could be improved. Some of the results can be extended to more general cases easily. For an illustration, we will show the existence of moments for the SSAR(p) model with the MA error.

Let $\{v_t\}$ be the i.i.d disturbance terms satisfying the condition (iii) with $k=1$ in Lemma 7 and

$$(3.35) \qquad u_t = \sum_{j=0}^{q} c_j v_{t-j} ,$$

where $\{c_j\}$ are constants with $c_0 = 1$ for normalization. If we use a vector process $\boldsymbol{y}_t^{'} = (y_t, y_{t-1}, \cdots, y_{t-p+1}, v_t, v_{t-1}, \cdots, v_{t-q+1})$, then we have a Markovian representation for $\{\boldsymbol{y}_t\}$. By taking the criterion function

$$G(\boldsymbol{x}) = 1 + \max_{1 \le j \le p+q} |x_j| \xi_j \ , \tag{3.36}$$

where ξ_j $(j = 1, \cdots, m)$ are defined as in the proof of Lemma 6 and $m = p + q$. Then we have an inequality

$$E[G(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1} = \boldsymbol{x}] \le \ c[1 + \sum_{j=1}^{q} |v_{t-j}|] + \theta G(\boldsymbol{x}), \tag{3.37}$$

where $0 < \theta < 1$ and c is some constant.
By repeating the above procedure and taking the conditional expectations, we have

$$E[G(\boldsymbol{y}_t)|\boldsymbol{y}_0 = \boldsymbol{x}] \le \ c \sum_{k=0}^{t-1} \theta^k E[1 + \sum_{j=1}^{q} |v_{t-k-j}||\boldsymbol{y}_0 = \boldsymbol{x}] + \theta^t G(\boldsymbol{x}) \ . \tag{3.38}$$

Then by taking the expectation with respect to the initial distribution, we finally have

$$\sup_{t \ge 1} E[|y_t|] < +\infty \ , \tag{3.39}$$

provided that we assume the condition (iv) with $k = 1$ in Lemma 7 and the condition $E[|v_s|] < \infty$ for $s \le 0$. We can use the similar arguments to obtain

$$\sup_{t \ge 1} E[|y_t|^k] < +\infty \tag{3.40}$$

for an arbitrary integer $k \ge 1$.

Finally, we shall give the result that the information matrix evaluated at the true parameter values is non-singular for the p–th order univariate SSAR model. The following lemma is equivalent to Lemma 5 when $m = p = 1$. We can confirm this by noticing $(\boldsymbol{\Omega}_1)_{11} = (\boldsymbol{\Omega}_2)_{11} = 1$ in (3.21) for this particular case.

Lemma 8 *In the p–th order univariate SSAR model (3.23), assume the coherency condition (3.24), the sufficient condition for the geometric ergodicity in Lemma 6, and the normal distributions for mutually independent disturbances. Then the information matrix*

$$\boldsymbol{I}(\boldsymbol{\theta}_0) = \lim_{T \to \infty} \frac{1}{T} [-\frac{\partial^2 \log L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}] |_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \tag{3.41}$$

is positive definite for the parameter vector $\boldsymbol{\theta}' = (\sigma_1, \sigma_2, r_0, r_1, \cdots, r_p)$.

Proof : Given the initial conditions $\boldsymbol{y}_0' = (y_{-p+1}, \cdots, y_0)$, the log-likelihood function for $\{y_t\}$ is proportional to

$$\begin{aligned} \log L_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) \\ &= -\sum_{t=1}^{T} \sum_{i=1}^{2} I_t^{(i)} \log \sigma_i - \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{2} I_t^{(i)} [\frac{1}{\sigma_i} \Delta y_t + r_0 + \boldsymbol{r}' \boldsymbol{y}_{t-1}]^2, \end{aligned} \tag{3.42}$$

where we defined the vectors by $\boldsymbol{r}' = (r_1, \cdots, r_p)$ and $\boldsymbol{y}'_{t-1} = (y_{t-1}, \cdots, y_{t-p})$. Then by direct calculations, we have $\partial^2 l_t(\boldsymbol{\theta})/\partial\sigma_1\partial\sigma_2 = 0$,

$$(3.43)\qquad I_t\left(\begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}, \begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}'\right) = E\left[-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}\partial\begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}'}\right] = E\left[\begin{pmatrix} 1 \\ \boldsymbol{y}_{t-1}\end{pmatrix}\begin{pmatrix} 1 \\ \boldsymbol{y}_{t-1}\end{pmatrix}'\right],$$

$$(3.44)\qquad I_t\left(\begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}, \sigma_i\right) = E\left[-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\begin{pmatrix} r_0 \\ \boldsymbol{r}\end{pmatrix}\partial\sigma_i}\right] = (-1)\frac{1}{\sigma_i^2}E\left[\begin{pmatrix} 1 \\ \boldsymbol{y}_{t-1}\end{pmatrix}\Delta y_t I_t^{(i)}\right],$$

and

$$(3.45)\qquad I_t(\sigma_i, \sigma_i) = E\left[-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\sigma_i^2}\right] = \frac{1}{\sigma_i^4}E\{[3(\Delta y_t)^2 + 2c_t\Delta y_t\sigma_i - \sigma_i^2]I_t^{(i)}\}\ \ (i = 1, 2),$$

where $c_t = r_0 + \boldsymbol{r}'\boldsymbol{y}_{t-1}$.
Because we assume the normal distribution for mutually independent disturbances $\{u_t\}$, we can utilize the relation

$$(3.46)\qquad E[(u_t^2 - 1 - u_t c_t)I_t^{(i)}|\mathcal{F}_{t-1}] = 0.$$

Then by applying (3.46) and using the identity $\Delta y_t I_t^{(i)} = [u_t - c_t]I_t^{(i)}\ \ (i = 1, 2)$, we can re-write (3.45) as

$$(3.47)\qquad I_t(\sigma_i, \sigma_i) = \frac{1}{\sigma_i^4}E[(\Delta y_t)^2 I_t^{(i)}] + \frac{1}{\sigma_i^2}E[I_t^{(i)}].$$

Hence we can write

$$(3.48)\qquad I_t(\boldsymbol{\theta}) = E\left[\begin{pmatrix} \frac{1}{\sigma_1^2}I_t^{(1)}\Delta y_t \\ \frac{1}{\sigma_2^2}I_t^{(2)}\Delta y_t \\ -1 \\ -\boldsymbol{y}_{t-1}\end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_1^2}I_t^{(1)}\Delta y_t \\ \frac{1}{\sigma_2^2}I_t^{(2)}\Delta y_t \\ -1 \\ -\boldsymbol{y}_{t-1}\end{pmatrix}'\right] + E\left[\begin{pmatrix} \frac{1}{\sigma_1}I_t^{(1)} \\ \frac{1}{\sigma_2}I_t^{(2)} \\ 0 \\ O\end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_1}I_t^{(1)} \\ \frac{1}{\sigma_2}I_t^{(2)} \\ 0 \\ O\end{pmatrix}'\right].$$

The positive definiteness of the information matrix $I(\boldsymbol{\theta}_0)$ can be established by using Lemma 10 and the Ergodic theorem for the Markov chain. (Q.E.D.)

4. Appendix

In this Appendix, we give some lemmas useful for the derivations of our results in Section 3. The first lemma (Lemma 9) is useful to justify the procedure of (A.7) and

(A.9) in Sato and Kunitomo (1996) as stated in Section 2, which was omitted in the eariler papers mainly because of the space limitation. Since the last lemma (Lemma 11) has been known (Theorem 7.6.7 in Horn and Johnson(1985), for instance), we have omitted its proof.

Lemma 9 *Let $\mathcal{F}_t$ be the $\sigma-$algebra generated by a sequence of random vectors $(\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_t)$ for $t = 0, 1, \cdots$ and $\boldsymbol{X}_t = (X_{ti})$. Let $\boldsymbol{\mu}_t = E[\boldsymbol{X}_t | \mathcal{F}_{t-1}]$ for $t = 1, 2, \cdots$ and $\boldsymbol{\mu}_t = (\mu_{ti})$. Assume that (i)*

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\mu}_t \rightarrow \boldsymbol{\mu} \ a.s., \tag{4.1}$$

and (ii)

$$\sum_{t=1}^{\infty} t^{-2} E[\|\boldsymbol{X}_t - \boldsymbol{\mu}_t\|^2] < +\infty \ . \tag{4.2}$$

Then $lim_{T\rightarrow+\infty}(1/T)\sum_{t=1}^{T}\boldsymbol{X}_t = \boldsymbol{\mu}$ with probability one.

Proof : We write

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\mu}_t + \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{X}_t - \boldsymbol{\mu}_t) \ . \tag{4.3}$$

The sum $\boldsymbol{z}_T = \sum_{t=1}^{T} t^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}_t)$ is a martingale because $E[t^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}_t)|\mathcal{F}_{t-1}] = 0$ with probability one. The variance of $\boldsymbol{z}_T$ is given by

$$E[\|\boldsymbol{z}_T\|^2] = \sum_{t=1}^{T} t^{-2} E[\|\boldsymbol{X}_t - \boldsymbol{\mu}_t\|^2] \ , \tag{4.4}$$

which is bounded by Assumption (ii). Hence the second term in (4.3) converges to zero with probability one by the martingale convergence theorem and Kronecker's lemma. Then we have the desired result by using Assumption (i). (Q.E.D.)

Lemma 10 *Let an $(m+n)\times(m+n)$ non-negative matrix $\boldsymbol{A}$ be partitioned as*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} ,$$

where $\boldsymbol{A}_{22}$ is a positive definite matrix. Then for any $m \times m$ positive definite matrix $\boldsymbol{B}_{11}$,

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}_{11} + \boldsymbol{B}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \tag{4.5}$$

is positive definite.

Proof : Because $\boldsymbol{A}_{22}$ is a positive definite matrix, then we have

$$|\boldsymbol{C}| = |\boldsymbol{A}_{22}||\boldsymbol{A}_{11} + \boldsymbol{B}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}| \ . \tag{4.6}$$

Then because $\boldsymbol{A}_{11.2} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ is non-negative definite and $\boldsymbol{B}_{11}$ is positive definite, $|\boldsymbol{C}| \neq 0$ and we have the result. (Q.E.D.)

Lemma 11 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times m$ positive definite matrices. Then for any $0 < \alpha < 1$,*

$$\log |\alpha \boldsymbol{A} + (1-\alpha)\boldsymbol{B}| \geq \alpha \log |\boldsymbol{A}| + (1-\alpha) \log |\boldsymbol{B}| \ , \tag{4.7}$$

with the equality holds if and only if $\boldsymbol{A} = \boldsymbol{B}$.

5. Concluding Remarks

In this note we have given some derivations omitted in Kunitomo and Sato (1996), and Sato and Kunitomo (1996). Also we have given the related sufficient conditions on the geometric ergodicity and the existence of moments for the simultaneous switching autoregressive (SSAR) models. Many of our derivations and discussions in this note are rather straightforward and may be redundant for some well-trained econometricians as well as statisticians. However, we hope that this note could help convincing some readers that the results stated in our eariler papers are technically valid under mild additional conditions.

Finally, we should mention that the standard SSAR models discussed in this note have been recently extended to a class of non-stationary SSAR models by Kunitomo and Sato (1997a,b). They are useful for some econometric applications including the analyses of financial time series data.

References

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

[2] Basawa, I.V., Feigin, P.D., and Heyde, C.C. (1976), "Asymptotic Properties of Maximum Likelihood Estimators for Stochastic Processes," *Sankya*, **36**, 259-270

[3] Chan, K.S. and H. Tong (1985), "On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equation," *Adv. Appl. Prob.*, **17**, 666-678.

[4] Horn, R. and C. Johnson (1985), *Matrix Analysis*, Cambridge University Press.

[5] Kunitomo, N. and S. Sato (1996), "Asymmetry in Economic Time Series and Simultaneous Switching Autoregressive Model," *Structural Change and Economic Dynamics*, **7**, 1-34.

[6] Kunitomo, N. and S. Sato (1997a), "Stationary and Non-stationary Simultaneous Switching Autoregressive Models with an Application to Financial Time Series," Unpublished manuscript.

[7] Kunitomo, N. and S. Sato (1997b), "Estimation of Asymmetrical Volatility for Asset Prices : The Simultaneous Switching Autoregressive ARIMA Approach," Unpublished manuscript.

[8] Meyn, S.P. and R.L. Tweedie (1993), *Markov Chains and Stochastic Stability*, Springer.

[9] Sato, S. and N. Kunitomo (1996), "Some Properties of the Maximum Likelihood Estimator in the Simultaneous Switching Autoregressive Model," *Journal of Time Series Analysis*, **17**, 287-307.

[10] Tong, H. (1990), *Non-linear Time Series*, Oxfor Universty Press, Oxford.

[11] Tweedie, R. L. (1983), "Criteria for Rates of Convergence of Markov Chains with Application to Queueing and Storage Theory," In *Papers in Probability, Statistics, and Analysis*, ed. J.F.C.Kingman and G.E.H.Reuter, Cambridge University Press, 260-276.