

Non-Identifiability of Simultaneous Spatial Autoregressive Model and Singularity of Fisher Information Matrix

Yuuki Rikimaru^{1,2} & Ritei Shibata³

¹ School of International Liberal Studies, Waseda University, Tokyo, Japan

² School of Fundamental Science and Technology, Keio University, Kanagawa, Japan

³ Department of Mathematics, Keio University, Kanagawa, Japan

Correspondence: Yuuki Rikimaru, School of International Liberal Studies, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo, Japan. E-mail: yuuki@datascience.jp

Received: May 4, 2017 Accepted: May 31, 2017 Online Published: June 15, 2017

doi:10.5539/ijsp.v6n4p31 URL: <https://doi.org/10.5539/ijsp.v6n4p31>

Abstract

Simultaneous spatial autoregressive model is widely used for spatial data analysis, observed at a set of grid points in a space. However a problem, not so well known, is that there exists no unique model unlike time series AR model for given autocovariances or spectral density. We show that such a non-identifiability of the model implies existence of multiple maximum likelihood estimates under Gaussianity and causes non-estimability of parameters and the singularity of Fisher information matrix. Several types of necessary and sufficient conditions for the singularity are given.

Keywords: Fisher information matrix, maximum likelihood estimation, non-identifiability, simultaneous spatial autoregressive model, singularity.

1. Introduction

A simultaneous spatial autoregressive model for a weakly stationary random field $\{X_{\mathbf{v}}; \mathbf{v} \in \mathbb{Z}^n\}$ with the mean 0 and the autocovariance function $\gamma_{\mathbf{h}} = E(X_{\mathbf{v}}X_{\mathbf{v}+\mathbf{h}})$, $\mathbf{h} \in \mathbb{Z}^n$ is the model which satisfies the equation

$$P(T_1, \dots, T_n)X_{\mathbf{v}} = \varepsilon_{\mathbf{v}}, \quad \mathbf{v} \in \mathbb{Z}^n, \quad (1)$$

where $\{\varepsilon_{\mathbf{v}}; \mathbf{v} \in \mathbb{Z}^n\}$ is a set of uncorrelated random variables with the mean 0 and the variance σ^2 , $\sigma > 0$. Here the operator

$$P(T_1, \dots, T_n) = \sum_{\mathbf{k} \in \mathcal{K}} \beta_{\mathbf{k}} T_1^{k_1} \cdots T_n^{k_n}$$

is an n -dimensional transfer function with the real coefficients $\beta_{\mathbf{k}}$, $\mathbf{k} \in \mathcal{K}$ where $\mathbf{0} \in \mathcal{K}$ is a set of finite points $\mathbf{k} = (k_1, k_2, \dots, k_n)$ on \mathbb{Z}^n and $\beta_{\mathbf{0}} = 1$. We denote the number of elements of \mathcal{K} is m , so that the number of regression parameters as $m - 1$. The operators T_j , $j = 1, \dots, n$ are shift operators such as

$$T_j X_{\mathbf{v}} = X_{v_1, \dots, v_j+1, \dots, v_n}.$$

We assume the following for the weak stationary of the simultaneous spatial autoregressive model throughout this paper.

Assumption 1 $P(z_1, \dots, z_n) = \sum_{\mathbf{k} \in \mathcal{K}} \beta_{\mathbf{k}} z_1^{k_1} \cdots z_n^{k_n}$ is non zero on the domain $D = \{(z_1, \dots, z_n) ; |z_1| = \cdots = |z_n| = 1\}$ in \mathbb{C}^n .

The spectral density of the model (1) is then

$$f(\omega) = \frac{\sigma^2}{|P(z_1, \dots, z_n)|^2},$$

where $z_j = \exp(i\omega_j)$, $j = 1, \dots, n$ and $\omega = (\omega_1, \omega_2, \dots, \omega_n)$.

2. Non-identifiability of Simultaneous Spatial Autoregressive Model

We first note that any polynomial $P(z_1, \dots, z_n)$ is decomposable into a product of prime factors $h_k(z_1, \dots, z_n)$, $k = 1, \dots, p$ as

$$P(z_1, \dots, z_n) = \prod_{k=1}^p h_k(z_1, \dots, z_n).$$

Therefore, there exist 2^p choices in selecting $h_k(z_1, \dots, z_n)$ or $\overline{h_k(z_1, \dots, z_n)}$ for $k = 1, \dots, p$ to have a transfer function $P(z_1, \dots, z_n)$ which leads us to the spectral density

$$f(\omega) = \frac{\sigma^2}{\prod_{k=1}^p h_k(z_1, \dots, z_n) \overline{h_k(z_1, \dots, z_n)}}. \tag{2}$$

There is also freedom to add a factor of the form $cz_1^{\ell_1} z_2^{\ell_2} \dots z_n^{\ell_n}$ to the transfer function $P(z_1, \dots, z_n)$ for any constant c and integers $\ell_1, \ell_2, \dots, \ell_n$, since the constant c can be absorbed into the parameter σ^2 .

EXAMPLE 1. Let us consider a simple one-dimensional autoregressive model,

$$X_v + \beta_1 X_{v+1} + \beta_{-1} X_{v-1} = \varepsilon_v. \tag{3}$$

Then there exist $2^2 = 4$ different choices of transfer function for the spectral density

$$f(\omega) = \frac{\sigma_0^2}{(z - \alpha_1)(z - \alpha_2)(z^{-1} - \bar{\alpha}_1)(z^{-1} - \bar{\alpha}_2)}, \tag{4}$$

where $z = e^{i\omega}$ and $\alpha_1, \alpha_2 \in \mathbb{C}$ are the roots of the polynomial $P(z) = z + \beta_1 z^2 + \beta_{-1}$. In fact, there exist the following four different transfer functions for the spectral density (4).

$$P_1(z) = -\frac{1}{\alpha_1 + \alpha_2} z^{-1}(z - \alpha_1)(z - \alpha_2) = 1 - \frac{1}{\alpha_1 + \alpha_2} z - \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} z^{-1},$$

$$P_2(z) = -\frac{1}{\bar{\alpha}_1 + \bar{\alpha}_2} z(z^{-1} - \bar{\alpha}_1)(z^{-1} - \bar{\alpha}_2) = 1 - \frac{\bar{\alpha}_1 \bar{\alpha}_2}{\bar{\alpha}_1 + \bar{\alpha}_2} z - \frac{1}{\bar{\alpha}_1 + \bar{\alpha}_2} z^{-1},$$

$$P_3(z) = \frac{1}{1 + \alpha_1 \bar{\alpha}_2} (z - \alpha_1)(z^{-1} - \bar{\alpha}_2) = 1 - \frac{\bar{\alpha}_2}{1 + \alpha_1 \bar{\alpha}_2} z - \frac{\alpha_1}{1 + \alpha_1 \bar{\alpha}_2} z^{-1}$$

and

$$P_4(z) = \frac{1}{1 + \bar{\alpha}_1 \alpha_2} (z^{-1} - \bar{\alpha}_1)(z - \alpha_2) = 1 - \frac{\bar{\alpha}_1}{1 + \bar{\alpha}_1 \alpha_2} z - \frac{\alpha_2}{1 + \bar{\alpha}_1 \alpha_2} z^{-1}.$$

It is easy to show that each transfer function has real coefficients, providing us a model (3) with different coefficients. The variance parameter σ^2 varies from transfer function to transfer function, $\sigma^2 = \sigma_0^2/|\alpha_1 + \alpha_2|^2$ for $P_1(z)$ and $P_2(z)$, and $\sigma^2 = \sigma_0^2/|1 + \alpha_1 \bar{\alpha}_2|^2$ for $P_3(z)$ and $P_4(z)$. It is easy to see that $P_1(z)$ and $P_2(z)$ become identical if and only if α_1 and α_2 are real and $\alpha_1 \alpha_2 = 1$, and the $P_3(z)$ and $P_4(z)$ become identical if and only if α_1 and α_2 are real and $\alpha_1 = \alpha_2$. By noting Assumption 1, we see that such conditions are summarized as $\beta_1 = \beta_{-1}$ with $\beta_1^2 < 1/4$, that is, time reversible simultaneous spatial autoregressive model. However, it does not mean unique transfer function for the spectral density of time reversible model. The conditions $\alpha_1 = \alpha_2$ and $\alpha_1 \alpha_2 = 1$ are not compatible because of Assumption 1. Only two of the four transfer functions become identical and two others are not time reversible. We now see that there is no unique model for the given spectral density (4).

3. Maximum Likelihood Estimate

It is well known that the exact likelihood of simultaneous spatial autoregressive model has no closed form in terms of parameters even if the Gaussianity is assumed. Historically, a lot of approximations of the log-likelihood have been proposed. One of such approximations is that based on a modified periodogram, proposed by Guyon (1982). However, the estimation procedure is not only expensive in computation but also inaccurate because it requires multiple integration of the spectral density for each parameter value. In this respect, the approximation recently proposed by Rikimaru & Shibata (2016) is stronger and more straightforward, and closed in time domain. They also proved that the parameter estimate which maximises the approximation L_A in the following is asymptotically efficient.

Let us assume that the observations $\{x_v, \mathbf{v} \in \mathcal{N}\}$ are on a rectangular lattice $\mathcal{N} = \{(v_1, v_2, \dots, v_n); 1 \leq v_j \leq n_j, j = 1, \dots, n\}$. The $N = n_1 n_2 \dots n_n$ observations are arranged to make a vector \mathbf{x} in lexicographic order. By combining the $m - 1$ dimensional regression parameter vector $\boldsymbol{\beta}$ whose elements are arranged in lexicographic order of $\mathbf{k} \neq \mathbf{0} \in \mathcal{K}$ with σ , we have the whole parameter vector $\boldsymbol{\theta}$. An approximation of the log-likelihood of $\boldsymbol{\theta}$ proposed by Rikimaru & Shibata (2016) is then

$$L_A = \frac{1}{2} \log \det(A) - \frac{N}{2} \log 2\pi - \frac{1}{2} \mathbf{x}^T \tilde{A} \mathbf{x},$$

where

$$A = \frac{1}{\sigma^2} \sum_{k,k' \in \mathcal{K}} \beta_k \beta_{k'} W_{n_1}^{-k_1+k'_1} \otimes \dots \otimes W_{n_n}^{-k_n+k'_n}$$

and

$$\tilde{A} = \frac{1}{\sigma^2} \sum_{k,k' \in \mathcal{K}} \beta_k \beta_{k'} \alpha_{n_1}^{|k_1-k'_1|} \dots \alpha_{n_n}^{|k_n-k'_n|} W_{n_1}^{-k_1+k'_1} \otimes \dots \otimes W_{n_n}^{-k_n+k'_n}.$$

Here the symbol \otimes is Kronecker product and $\alpha_{n_j} = 1 + 1/n_j, j = 1, \dots, n$ are shrinkage factors to retain \sqrt{N} consistency.

The matrix W_n is an $n \times n$ circulant matrix such that the off-diagonal elements $(W_n)_{j,j+1}$ are all 1 for $j = 1, \dots, n - 1$, $(W_n)_{n,1} = 1$ and the other elements are all 0. It is clear that $W_n^n = W_n^0 = I_n$ and $W_n^{-1} = W_n^T$ where I_n is the $n \times n$ identity matrix.

The asymptotic efficiency proved is that the covariance matrix of the parameter estimate converges to the lower bound given by the inverse of the Fisher information matrix $I(\theta)$, whose elements are given by

$$I_{pq}(\theta) = \frac{1}{2 \cdot (2\pi)^n} \int \dots \int \frac{\partial \log f(\omega; \theta)}{\partial \theta_p} \frac{\partial \log f(\omega; \theta)}{\partial \theta_q} d\omega, \quad p, q = 1, 2, \dots, m, \tag{5}$$

(Whittle, 1954; Guyon, 1982; Robinson & Vidal Sanz, 2006), provided that $I(\theta)$ is non-singular which is a key assumption for the proof. It is rather unusual that the Fisher information matrix is singular in ordinary theory of statistics, but it often happens in case of simultaneous spatial autoregressive model. Before investigating when and why it happens, we will see other problems caused by non-identifiability of the model in maximum likelihood estimation by the following example.

EXAMPLE 2. Consider the same model (3) as in Example 1. Assume that $\{X_v\}$ has the spectral density (4) with $\alpha_1 = -0.85$, $\alpha_2 = -9.15$ and $\sigma_0 = 0.1$. Then four possible transfer functions for the spectral density are :

Transfer function	β_1	β_{-1}	σ
$P_1(z)$	0.10000	0.77775	0.01000
$P_2(z)$	0.77775	0.10000	0.01000
$P_3(z)$	1.04244	0.09683	0.01139
$P_4(z)$	0.09683	1.04244	0.01139

This suggests that the Gaussian likelihood function has the same value for such four sets of parameter values, since they share the same covariance structure. Therefore the likelihood function always has four maximum points on parameter space unless some of four transfer functions are identical. In fact, the following result of numerical experiment demonstrates this. In the experiment, $N = 1000$ random numbers are generated for $\{X_v\}$ by using the transfer function $P_1(z)$ with

$$\beta_1 = -\frac{1}{\alpha_1 + \alpha_2} = 0.1, \quad \beta_{-1} = -\frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} = 0.77775$$

and

$$\sigma = -\frac{\sigma_0}{\alpha_1 + \alpha_2} = 0.01.$$

Then, the following four maximum likelihood estimates are obtained by maximising L_A in this experiment.

	$\hat{\beta}_1$	$\hat{\beta}_{-1}$	$\hat{\sigma}$	$\max L_A$
Estimate1	0.10146	0.77066	0.00973	3124.091
Estimate2	0.77066	0.10146	0.00973	3124.091
Estimate3	1.04858	0.09804	0.01116	3124.091
Estimate4	0.09804	1.04858	0.01116	3124.091

Therefore, although the maximum likelihood estimate is consistent and asymptotically efficient as is proved, there is no global unique solution. This implies that we always have several different estimates of parameters, which may depend on the initial values of parameters for optimisation algorithm. There would be no good way to avoid such a problem in practice, because the problem is not over-parametrisation but non-identifiability of transfer function for given spectral density or covariances. Only a possible remedy would be to restrict our attention into a specific region of parameter space, which is meaningful for the underlying problem and effective for restricting the transfer function into a unique one. We might have to search for all possible solutions anyway since it would not be so easy to restrict the region beforehand.

4. Singularity of the Fisher Information Matrix $I(\theta)$

We have seen that several different parameters, $\theta_1, \theta_2, \dots, \theta_{2p}$ are mapped from a given simultaneous spatial autoregressive spectral density. The problem of simultaneous spatial autoregressive model is not only on such a non-identifiability but also on the singularity of the Fisher information matrix which is closely related to the non-identifiability. We will concentrate our attention into the singularity of Fisher information matrix $I(\theta)$ in (5), which is also the limit of

$$-\frac{1}{N} \frac{\partial^2 \log f(\mathbf{x}, \theta)}{\partial \theta \partial \theta^T} \tag{6}$$

The following theorem states that the Fisher information matrix becomes singular if some of the parameters are duplicated.

Theorem 1 *Fisher information matrix $I(\theta_0)$ becomes singular when some of the parameters are duplicated for the spectral density identified by θ_0 .*

Proof. Let $f_0(\omega)$ be the spectral density identified by θ_0 . Choose another $f(\omega)$ in the neighbourhood of $f_0(\omega)$, which leads us different θ_1 and θ_2 in the neighbourhood of θ_0 . We have then two corresponding maximum likelihood estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, which are zeros of the gradients of the log-likelihood

$$g_k(\theta) = \frac{\partial \log f(\mathbf{x}, \theta)}{\partial \theta_k}, k = 1, 2, \dots, m.$$

By Taylor expansion, we see that

$$0 = (\hat{\theta}_1 - \hat{\theta}_2)^T \frac{\partial}{\partial \theta} g_k(\theta_k^*), k = 1, 2, \dots, m \tag{7}$$

holds true for θ_k^* between $\hat{\theta}_1$ and $\hat{\theta}_2$. Rewrite (7) as

$$0 = N^{1/2-\epsilon} \left((\hat{\theta}_1 - \theta_1) - (\hat{\theta}_2 - \theta_2) + (\theta_1 - \theta_2) \right)^T \frac{1}{N} \frac{\partial}{\partial \theta} g_k(\theta_k^*), k = 1, 2, \dots, m.$$

Then by setting $\theta_1 = \theta_0 + N^{-1/2+\epsilon} \delta$ and $\theta_2 = \theta_0 - N^{-1/2+\epsilon} \delta$, we have

$$0 = \delta^T I(\theta_0).$$

The following example illustrates what happens if the Fisher information matrix is singular. It would be clear if we note that the Hessian matrix of the log-likelihood (6) is likely to be singular if it happened.

EXAMPLE 3. Let us consider the same model as in Example 1. As is already seen, if $\beta_1 = \beta_{-1}$ and $\beta_1^2 < 1/4$, then the transfer functions P_1 and P_2 or P_3 and P_4 are identical and the Fisher information matrix becomes singular as

$$I(\theta) = \begin{pmatrix} a & a & b \\ a & a & b \\ b & b & c \end{pmatrix},$$

where

$$a = \frac{-1 + 8\beta_1^2 + (\sqrt{1 - 4\beta_1^2})^3}{\beta_1^2(\sqrt{1 - 4\beta_1^2})^3}, b = \frac{-2(1 - \sqrt{1 - 4\beta_1^2})}{\sigma\beta_1\sqrt{1 - 4\beta_1^2}} \text{ and } c = \frac{4}{\sigma^2}.$$

From the maximum likelihood equation,

$$\mathbf{0} = \frac{1}{\sqrt{N}} \frac{\partial \log f(\mathbf{x}, \theta)}{\partial \theta} \Big|_{\hat{\theta}} \approx \frac{1}{\sqrt{N}} \frac{\partial \log f(\mathbf{x}, \theta)}{\partial \theta} + \frac{1}{N} \frac{\partial^2 \log f(\mathbf{x}, \theta)}{\partial \theta \partial \theta^T} \cdot \sqrt{N}(\hat{\theta} - \theta),$$

we see that

$$\begin{pmatrix} a\sqrt{N}\{(\hat{\beta}_1 + \hat{\beta}_{-1}) - (\beta_1 + \beta_{-1})\} + b\sqrt{N}(\hat{\sigma} - \sigma) \\ a\sqrt{N}\{(\hat{\beta}_1 + \hat{\beta}_{-1}) - (\beta_1 + \beta_{-1})\} + b\sqrt{N}(\hat{\sigma} - \sigma) \\ b\sqrt{N}\{(\hat{\beta}_1 + \hat{\beta}_{-1}) - (\beta_1 + \beta_{-1})\} + c\sqrt{N}(\hat{\sigma} - \sigma) \end{pmatrix}$$

is asymptotically normally distributed, so that we can only estimate $\beta_1 + \beta_{-1}$ and σ but not individual β_1 or β_{-1} .

4.1 Conditions for Non-Singularity of $I(\theta)$

As is seen from Example 3, singularity of the Fisher information matrix $I(\theta)$ causes more serious problem, non-estimability of individual parameters. It would be worthy of investigating what kind of conditions is necessary for the singularity of $I(\theta)$ because the Fisher information matrix is a complicated function of parameters and it is not feasible to check it as it is. We first derive a simple necessary and sufficient condition directly derived from the quadratic form of $I(\theta)$,

$$Q = \mathbf{y}^T I(\theta) \mathbf{y} = \frac{1}{2 \cdot (2\pi)^n} \int \cdots \int \left(\sum_{j=1}^m y_j \frac{\partial \log f(\omega)}{\partial \theta_j} \right)^2 d\omega. \tag{8}$$

Clearly a necessary and sufficient condition for the non-singularity is that the vector $\mathbf{y} = (y_1, y_2, \dots, y_m)$ is zero whenever $Q = 0$.

Theorem 2 A necessary and sufficient condition for $I(\theta)$ to be non-singular is

$$\frac{\partial \log f(\omega)}{\partial \theta_j}, \quad j = 1, \dots, m \text{ are linearly independent.}$$

Corollary 1 A sufficient condition for non-singularity of $I(\theta)$ is that

$$\frac{\partial \gamma_{\mathbf{k}}}{\partial \theta_j}, \quad j = 1, 2, \dots, m \text{ are linearly independent for a } \mathbf{k}.$$

Proof. We see that

$$\begin{aligned} Q &= \sum_{p,q=1}^m y_p y_q I_{pq}(\theta) = \lim_{n_1, \dots, n_n \rightarrow \infty} \frac{1}{2N} \sum_{p,q=1}^m y_p y_q \text{tr} \left(\Sigma^2 \frac{\partial \Sigma^{-1}}{\partial \theta_p} \frac{\partial \Sigma^{-1}}{\partial \theta_q} \right) \\ &= \lim_{n_1, \dots, n_n \rightarrow \infty} \frac{1}{2N} \text{tr} \left\{ \left(\Sigma^{-1} \sum_{j=1}^m y_j \frac{\partial \Sigma}{\partial \theta_j} \right)^2 \right\} \end{aligned}$$

from Proposition B.1 in Rikimaru & Shibata (2016) by noting that $\partial \Sigma^{-1} / \partial \theta_p = \Sigma^{-1} (\partial \Sigma / \partial \theta_p) \Sigma^{-1}$. Since the eigenvalues of Σ^{-1} are bounded away from 0, we have

$$\text{tr} \left\{ \left(\Sigma^{-1} \sum_{j=1}^m y_j \frac{\partial \Sigma}{\partial \theta_j} \right)^2 \right\} \geq C \text{tr} \left\{ \left(\sum_{j=1}^m y_j \frac{\partial \Sigma}{\partial \theta_j} \right)^2 \right\}$$

for a constant $C > 0$. It is enough to note that at most N elements of the matrix $\sum_{j=1}^m y_j \partial \Sigma / \partial \theta_j$ are $\sum_{j=1}^m y_j \partial \gamma_{\mathbf{k}} / \partial \theta_j$.

Corollary 2 If $\beta_{\mathbf{k}} = \beta_{-\mathbf{k}}$ for any $\mathbf{k} \in \mathcal{K}$, then $I(\theta)$ is singular.

Proof. If $\beta_{\mathbf{k}} = \beta_{-\mathbf{k}}$ for all $\mathbf{k} \in \mathcal{K}$, then

$$P(z_1, \dots, z_n) = P(z_1^{-1}, \dots, z_n^{-1}).$$

It is enough to note that

$$\frac{\partial \log f(\omega)}{\partial \beta_{\mathbf{k}}} = \frac{\partial \log f(\omega)}{\partial \beta_{-\mathbf{k}}}$$

holds true since

$$\frac{\partial \log f(\omega)}{\partial \beta_{\mathbf{k}}} = -\frac{z_1^{k_1} \dots z_n^{k_n}}{P(z_1, \dots, z_n)} - \frac{z_1^{-k_1} \dots z_n^{-k_n}}{P(z_1^{-1}, \dots, z_n^{-1})}$$

on the domain D .

EXAMPLE 4. Consider a 2-dimensional model,

$$X_{v_1, v_2} + \beta_{10} X_{v_1+1, v_2} + \beta_{-10} X_{v_1-1, v_2} + \beta_{01} X_{v_1, v_2+1} + \beta_{0-1} X_{v_1, v_2-1} = \varepsilon_{v_1, v_2},$$

then the spectral density is

$$f(\omega_1, \omega_2) = \frac{\sigma^2}{|1 + \beta_{10} z_1 + \beta_{-10} z_1^{-1} + \beta_{01} z_2 + \beta_{0-1} z_2^{-1}|^2}.$$

There exists only two transfer functions $P(z_1, z_2) = 1 + \beta_{10} z_1 + \beta_{-10} z_1^{-1} + \beta_{01} z_2 + \beta_{0-1} z_2^{-1}$ and $P(z_1^{-1}, z_2^{-1})$ for this spectral density. This is because $P(z_1, z_2)$ is prime factor. In fact, there exist no polynomials $Q_1(z_1, z_2)$ and $Q_2(z_1, z_2)$ of at most order 2 with respect to z_1 and z_2 , such that

$$z_1^{-1} z_2^{-1} P(z_1, z_2) = \beta_{10} z_1^2 z_2 + \beta_{01} z_1 z_2^2 + z_1 z_2 + \beta_{0-1} z_1 + \beta_{-10} z_2 = Q_1(z_1, z_2) Q_2(z_1, z_2).$$

Therefore, $P(z_1, z_2)$ is not decomposable into a product of transfer functions which accommodate with the underlying model. It is clear that $P(z_1, z_2)$ and $P(z_1^{-1}, z_2^{-1})$ are identical if and only if $\beta_{10} = \beta_{-10}$ and $\beta_{01} = \beta_{0-1}$. The singularity of $I(\theta)$ follows from Theorem 1 as well as from Corollary 2 in this case.

A practical procedure to check if the Fisher information matrix is singular would be through the matrix,

$$B = (\beta_{\ell_1} \quad \beta_{\ell_2} \quad \dots \quad \beta_{\ell_L})$$

where

$$\beta_{\ell_j} = \begin{pmatrix} \beta_{\mathbf{k}_1 + \ell_j} + \beta_{\mathbf{k}_1 - \ell_j} \\ \beta_{\mathbf{k}_2 + \ell_j} + \beta_{\mathbf{k}_2 - \ell_j} \\ \vdots \\ \beta_{\mathbf{k}_m + \ell_j} + \beta_{\mathbf{k}_m - \ell_j} \end{pmatrix}, \quad j = 1, \dots, L.$$

Here $\ell_j, j = 1, 2, \dots, L$ are ordered indices in $\mathcal{L} = \{\ell \mid \mathbf{k} - \ell$ or $\mathbf{k} + \ell \in \mathcal{K}$ for a $\mathbf{k} \in \mathcal{K}\}$. We should choose either ℓ_j or $-\ell_j$ in \mathcal{L} since $\beta_{\ell_j} = \beta_{-\ell_j}$. Let us $\mathbf{k}_i, i = 1, \dots, m$ are indices arranged in lexicographic order in \mathcal{K} and $\beta_{\mathbf{k}} = 0$ for $\mathbf{k} \notin \mathcal{K}$ as a convention. Note that it is always true that $L > m - 1$.

Theorem 3 A necessary and sufficient condition for the non-singularity of $I(\theta)$ is that B is of full rank.

Proof. We may restrict our attention into the non-singularity of the first $(m - 1) \times (m - 1)$ submatrix of $I(\theta)$, since the last row and column off-diagonal elements are all 0 and the diagonal element is $(2/\sigma)^2$. By setting $y_m = 0$ in (8) and introducing $Y(z_1, \dots, z_n) = \sum_{\mathbf{k}_j \neq \mathbf{0} \in \mathcal{K}} y_j z_1^{k_1} \dots z_n^{k_n}$, we have

$$Q = \frac{2}{(2\pi)^n} \int \left\{ \text{Re} \left(\frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right) \right\}^2 d\omega.$$

Thus, $Q = 0$ implies

$$Y(z_1, \dots, z_n) \bar{P}(z_1, \dots, z_n) + \bar{Y}(z_1, \dots, z_n) P(z_1, \dots, z_n) = 0.$$

A necessary and sufficient condition for the non-singularity of $I(\theta)$ is now that

$$\sum_{\mathbf{k}_j \neq \mathbf{0} \in \mathcal{K}} y_j (\beta_{\mathbf{k}_j - \ell} + \beta_{\mathbf{k}_j + \ell}) = 0 \text{ for any } \ell \in \mathcal{L}$$

implies $y_j = 0, j = 1, 2, \dots, m - 1$. This completes the proof.

EXAMPLE 5. The matrix B for the model in Example 1 is derived from

$$\beta_0 = \begin{pmatrix} \beta_{-1+0} + \beta_{-1-0} \\ \beta_{0+0} + \beta_{0-0} \\ \beta_{1+0} + \beta_{1-0} \end{pmatrix} = 2 \begin{pmatrix} \beta_{-1} \\ 1 \\ \beta_1 \end{pmatrix}, \beta_1 = \begin{pmatrix} \beta_{-1+1} + \beta_{-1-1} \\ \beta_{0+1} + \beta_{0-1} \\ \beta_{1+1} + \beta_{1-1} \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_1 + \beta_{-1} \\ 1 \end{pmatrix} \text{ and } \beta_2 = \begin{pmatrix} \beta_{-1+2} + \beta_{-1-2} \\ \beta_{0+2} + \beta_{0-2} \\ \beta_{1+2} + \beta_{1-2} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ 0 \\ \beta_{-1} \end{pmatrix},$$

as

$$B = \begin{pmatrix} 2\beta_{-1} & 1 & \beta_1 \\ 2 & \beta_1 + \beta_{-1} & 0 \\ 2\beta_1 & 1 & \beta_{-1} \end{pmatrix}.$$

The determinant

$$\det(B) = 2(\beta_1 - \beta_{-1})(1 + \beta_1 + \beta_{-1})(1 - \beta_1 - \beta_{-1})$$

is zero if and only if $\beta_1 = \beta_{-1}$ since $1 + \beta_1 + \beta_{-1} \neq 0$ and $1 - \beta_1 - \beta_{-1} \neq 0$ from Assumption 1. Thus, we see that the condition $\beta_1 = \beta_{-1}$ with $\beta_1^2 < 1/4$ is not only necessary and sufficient condition for some of transfer functions being identical, but also for the singularity of the Fisher information matrix in this example.

4.2 Unilateral Simultaneous Spatial Autoregressive Model

It is taken it for granted that unilateral simultaneous spatial autoregressive model including AR model in time series,

$$P(T_1, \dots, T_n) = \sum_{k_1, \dots, k_n \geq 0, \mathbf{k} \in \mathcal{K}} \beta_{\mathbf{k}} T_1^{k_1} \dots T_n^{k_n}$$

is always identifiable and the Fisher information matrix $I(\theta)$ is non-singular. However, it would be worthy of proving in the frame work of simultaneous spatial autoregressive model. Then, it becomes clearer that the problems we have discussed are due to the lack of unilaterality of general simultaneous spatial autoregressive model.

Theorem 4 *Unilateral simultaneous spatial autoregressive model is always identifiable and the Fisher information matrix $I(\theta)$ is always non-singular.*

Proof. It is only possible to choose $h_k(z_1, z_2, \dots, z_n), k = 1, 2, \dots, p$ to find out transfer function $P(z_1, z_2, \dots, z_n)$ for the spectral density (2). Any other choice contradicts with the unilaterality of the model. Therefore unilateral simultaneous spatial autoregressive model is always unique for given spectral density. On the other hand the quadratic form (8) is then rewritten as

$$\begin{aligned} Q &= \frac{2}{(2\pi)^n} \int \left\{ \operatorname{Re} \left(\frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right) - \frac{y_m}{\sigma} \right\}^2 d\omega \\ &= \frac{2}{(2\pi)^n} \int \left[\left\{ \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right\}^2 + 2 \left| \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right|^2 + \left\{ \frac{Y(z_1^{-1}, \dots, z_n^{-1})}{P(z_1^{-1}, \dots, z_n^{-1})} \right\}^2 \right. \\ &\quad \left. - \frac{y_m}{\sigma} \left\{ \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} + \frac{Y(z_1^{-1}, \dots, z_n^{-1})}{P(z_1^{-1}, \dots, z_n^{-1})} \right\} + \frac{y_m^2}{\sigma^2} \right] d\omega, \end{aligned}$$

where $Y(z_1, \dots, z_n) = \sum_{k_1, \dots, k_n \geq 0, \mathbf{k} \neq \mathbf{0} \in \mathcal{K}} y_{\mathbf{k}} z_1^{k_1} \dots z_n^{k_n}$. Since $P(z_1, \dots, z_n)$ and $Y(z_1, \dots, z_n)$ are both analytic non zero function on the domain D , we have

$$\int \left\{ \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right\}^k d\omega = - \int_D \left\{ \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right\}^k \frac{1}{z_1 \dots z_n} dz_1 \dots dz_n = 0$$

and

$$\int \left\{ \frac{Y(z_1^{-1}, \dots, z_n^{-1})}{P(z_1^{-1}, \dots, z_n^{-1})} \right\}^k d\omega = - \int_D \left\{ \frac{Y(z_1^{-1}, \dots, z_n^{-1})}{P(z_1^{-1}, \dots, z_n^{-1})} \right\}^k \frac{1}{z_1 \dots z_n} dz_1 \dots dz_n = 0$$

for any integer k . Therefore

$$Q = \frac{2}{(2\pi)^n} \int \left\{ 2 \left| \frac{Y(z_1, \dots, z_n)}{P(z_1, \dots, z_n)} \right|^2 + \frac{y_m^2}{\sigma^2} \right\} d\omega.$$

is zero if and only if $Y(z_1, \dots, z_n) = 0$ and $y_m = 0$. This proves the non-singularity of $I(\theta)$.

5. Concluding Remarks

We have shown that simultaneous spatial autoregressive model is non-identifiable from the covariance structure or the spectral density. Several different regression parameters with different standard deviation of the disturbance are mapped from a spectral density. Therefore, we have to be careful about estimation of parameters based on the second moments, for example, estimation by Gaussian maximum likelihood principle. There could be many other estimates even if an estimate had been obtained by giving an initial value to an optimisation algorithm. A practical procedure would be to find out all estimates and pick up one which is most meaningful for the underlying phenomena. This non-identifiability of the model has been already mentioned in the context of two sided moving average model (Rosenblatt, 1980). A cure he proposed is to employ bispectrum, which can be applied for the model, too. But we leave it for future investigation, together with an investigation of the type of parameter mapping from the spectral density.

Another problem we have investigated in this paper is possible singularity of the Fisher information matrix, where not all parameters are estimable. Theorem 1 demonstrates that it happens when some of parameters mapped from a spectral density are duplicated. Non-identifiability of simultaneous spatial autoregressive model leads us not only to multiple estimates of parameters but also non-estimable parameters. We need to check such a singularity before estimation. Otherwise, we may face unconvergence of optimisation algorithm or instability of the estimate. Several types of conditions given in Section 4 would be useful for the check. There are a lot of open problems left, for example, converse of Theorem 1 or any other type of necessary and sufficient condition for the non-singularity than that given in Theorem 3.

References

- Guyon, X. (1982). Parameter Estimation for a Stationary Process on a d-Dimensional Lattice, *Biometrika*, 69, 95-105. <https://doi.org/10.2307/2335857>
- Rikimaru, Y., & Shibata, R. (2016). A good approximation of the Gaussian likelihood of simultaneous autoregressive model which yields us an asymptotically efficient estimate of parameters. *Journal of Statistical Planning and Inference*, 173, 31-46. <https://doi.org/10.1016/j.jspi.2016.01.003>
- Robinson, P. M., & J. Vidal Sanz (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis*, 97, 1090-1120. <https://doi.org/10.1016/j.jmva.2005.05.013>
- Rosenblatt, M. (1980). Linear processes and bispectra. *Journal of Applied Probability*, 17, 265-270. <https://doi.org/10.2307/3212945>
- Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika*, 41, 434-449. <https://doi.org/10.2307/2332724>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).