# Parameter Estimation of 

## Simultaneous Spatial Autoregressive Model

August 2017

Yuuki RIKIMARU

A Thesis for the Degree of Ph.D. in Science Parameter Estimation of Simultaneous Spatial Autoregressive Model

August 2017

Graduate School of Science and Technology
Keio University

## Preface

The first manifestations of statistics for spatial data appear to have arisen in the form of data maps. For example, Halley (1686) superimposed, onto a map of land forms, directions of trade winds and monsoons between and near the tropics, and attempted to assign them a physical cause (Cressie, 1993).

Importance of spatial data analysis was first advocated by Fisher(1935) since spatial correlation might remain significant in agricultural experiment even if the sampling design were well organised. Whittle (1954) is the first person who proposed a spatial model to describe such a spatial correlation. The model is an autoregressive type model which now called simultaneous spatial autoregressive model. It is attractive as a spatial model because of its simplicity but the estimation procedure proposed there turned out not $\sqrt{N}$-consistent for the number $N$ of observations and the computation is not straightforward. Various articles have been devoted since then for the improvement of the behaviour of the estimator but it is still not popular yet.

In this thesis, an approximation on space domain of the log-likelihood of simultaneous spatial autoregressive model is proposed. It is proved that the estimation procedure so as to maximise the approximation provides us a consistent and asymptotically efficient estimator under the assumption that the underlying process is weakly stationary. Simultaneous spatial autoregressive model is however not always identifiable. The non-identifiability causes not only problem in finding a global solution of the maximum likelihood equation but also non-estimable problem in the estimation. It is shown that several types of sub-models of the simultaneous spatial autoregressive model are effective to avoid such a non-identifiability problem.

Chapter 1 is a brief introduction to spatial data and its model. Various
models including simultaneous spatial autoregressive model are introduced together with the parameter estimation procedure ever proposed. A handy weak stationarity condition for spatial autoregressive model is also given with the proof.

In Chapter 2, an approximation on space domain to the log-likelihood of the model is proposed (Rikimaru and Shibata, 2016). It looks a mimic to that for time series autoregressive model but several new ideas will be introduced to accommodate simultaneous spatial autoregressive model. One is to modify the translation matrix of observations to errors to be a circulant matrix. The modification is applicable for any simultaneous spatial autoregressive model although only 2 -dimensional model is investigated in this thesis and also enables us to develop transparent mathematical theory. The other is an introduction of shrinkage factor to the quadratic form of observations to retain the asymptotic efficiency of the estimator which maximises the approximation. It is in fact proved that the estimator based on this new approximation is consistent and asymptotically efficient. The result of random number experiments supports that the estimation procedure provides us estimates which have less bias and variance than the other procedure even if the number of observations is small. An effective random number generation algorithm is also developed.

In Chapter 3, it is shown that simultaneous spatial autoregressive model is not always identifiable for the given 2 nd moments or the spectral density (Rikimaru and Shibata, 2017). This implies that there could be multiple global solutions of the maximum likelihood equation. Therefore different estimates of parameters may come out depending on the initial value given to the optimisation algorithm used. It is also shown that Fisher information matrix could become singular if some of models were overlapped which share the same 2nd moments. The singularity of Fisher information matrix does not only destroy the asymptotic efficiency of the estimator but also result in non-estimability of some of parameters or the instability of the solution of the maximum likelihood equation. This suggests that we need to check if the Fisher information matrix is singular or not when the parameters are estimated. Several types of necessary and sufficient conditions are given for the check. Also it is shown that unilateral or symmetric
sub-model of simultaneous spatial autoregressive model is free from such a non-identifiability problem.

Chapter 4 is concluding remarks of this thesis. Summarising the results in this thesis, several problems left for future works are raised, including extension of the theorems for the case of general $d$ dimensional space together with modelling the expectation by multiple regression, compatibility of the orthogonality and the innovation property of error variables, application of the results to spatial lag model, investigation of identifiable sub-models of simultaneous spatial autoregressive model, and application of the results to conditional autoregressive model.

## Contents

Preface ..... i
1 Introduction ..... 1
1.1 Spatial Data ..... 1
1.2 Simultaneous Spatial Autoregressive Model ..... 3
1.2.1 Spatial Autoregressive Model ..... 3
1.2.2 Weak Stationarity of Spatial Autoregressive Model ..... 5
1.2.3 Spatial Lag Model ..... 9
1.3 Parameter Estimation of SAR model ..... 10
1.3.1 Aprroximations of the Likelihood of SAR model ..... 10
1.3.2 A Space Domain Approximation ..... 14
1.4 Non-identifiability of SAR Model ..... 15
2 A Simple Statistically Efficient Approximation of the Gaussian Likelihood of SAR Model ..... 17
2.1 Introduction ..... 17
2.2 SAR Model ..... 18
2.3 An Approximation of the Log-likelihood ..... 19
2.4 Goodness of the Approximation ..... 22
2.5 Consistency ..... 24
2.6 Asymptotic Efficiency ..... 26
2.7 Numerical Experiments ..... 33
2.7.1 Random Number Generation ..... 34
2.7.2 Parameter Estimation ..... 35
2.8 Remarks ..... 38
2.9 Appendix ..... 39
2.9.1 Fundamental Lemmas and Propositions ..... 39
2.9.2 Lemma and Proposition Used for the Proof of Theorem 2.3 ..... 46
2.9.3 Representation of Gradient and Hessian ..... 49
3 Non-identifiability of SAR Model ..... 51
3.1 Introduction ..... 51
3.2 Non-identifiability of SAR Model ..... 52
3.3 Maximum Likelihood Estimate ..... 54
3.4 Singularity of the Fisher Information Matrix $I(\boldsymbol{\theta})$ ..... 57
3.4.1 Conditions for Non-Singularity of $I(\boldsymbol{\theta})$ ..... 59
3.4.2 Unilateral SAR Model ..... 63
3.4.3 Symmetric SAR Model ..... 64
3.5 Remarks ..... 65
4 Concluding Remarks ..... 67
Acknowledgements ..... 71

## Chapter 1

## Introduction

Spatial data analysis plays an important role in various fields of science, for example, geostatistics, economics, environmental science, ecology, agricultural science, image modeling, and so on. In this thesis, we concentrate our attention into "simultaneous spatial autoregressive model" which is one of standard autoregessive models in spatial data analysis and closely related spatial lag model.

### 1.1 Spatial Data

Spatial data observed on a space is roughly classified into two categories, one is "point pattern data" and the other is "values on a space".

Point Pattern Data Point pattern data is a record of locations or sites where an event occurred, for example, a record of tree locations on an area. One of main concerns of the analysis would be the type of occurrence patterns, essentially regular, completely random or aggregated, where essentially regular pattern appears when events occur repulsively, completely random pattern appears when events do not interact with each other and aggregated pattern appears when events occur attractively. A basic stochastic model is a point process with the intensity function which describes the type of the pattern. For works on the analysis of this type of
data, see for example, Pielou (1959), Getis and Boots (1978), Marquiss et al. (1978), Ripley (1981), Diggle (1983) and Upton and Fingleton (1985) (Cressie, 1993).

Values on a Space The data is obtained as a set of measurements at fixed sites. An example would be the amount of rainfall at each site in an area. It is, for example, the case when geographical distribution of the amount of rainfall is main concern of the analysis. However, if the main concern were the pattern of rainfall, the point pattern data, rainy or not is an appropriate data, which would be newly collected or aggregated from the amount of rainfall data. Therefore point pattern data and values on a space data are not exclusive. It changes by the aim of analysis. Stochastic model for such values on a space is random field. When the observation sites are designed on a mesh, for example, a subset of $\mathbb{Z}^{d}$, the observations are called lattice data. Image data is an example of such a lattice data since pixels are arranged on a mesh. One of main concerns of the analysis is to quantify spatial correlations and to understand mechanism of phenomena behind the data, as well as prediction or interpolation of the value. Related works are summarised, for example, in Cressie (1993) and Gaetan and Guyon (2010).

Point pattern data and values on a space data are totally different type of data, reflecting the aim of data collection or the analysis. Mathematical models for such data are also different. A typical model for the point pattern is a point process which describes the probability of occurrence at each site, whereas the model for the values on a space is a random field which describes the distribution and the correlations of the values over sites. Both are interesting but we will focus our interest on the values on a space in this thesis because of limited time and resource.

### 1.2 Simultaneous Spatial Autoregressive Model

### 1.2.1 Spatial Autoregressive Model

Spatial autoregressive model is a model by which each value on a site is explained by a linear combination of several values on the neighbourhood. For a random field $\left\{X_{\boldsymbol{v}}, \boldsymbol{v}=\left(v_{1}, v_{2}, \ldots, v_{d}\right)^{T} \in \mathcal{N}\right\}$ on a lattice $\mathcal{N} \subset \mathbb{Z}^{d}$ with mean zero, spatial autoregressive model is defined by the equations,

$$
\begin{equation*}
\sum_{k \in \mathcal{K}} \beta_{\boldsymbol{k}} X_{\boldsymbol{v}+\boldsymbol{k}}=\varepsilon_{\boldsymbol{v}}, \boldsymbol{v} \in \mathcal{N} \tag{1.1}
\end{equation*}
$$

where $\beta_{\boldsymbol{k}}$ 's are regression coefficients. The error process $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ is weakly stationary with mean zero and variance $\sigma^{2}$. We also assume that $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ has a positive spectral density. The set $\mathcal{K}$ is an index set for the neighbouring sites, including $\mathbf{0}$. For simplicity, we assume that any $X_{v}$ has mean zero. By this model, each $X_{v}$ is represented by the neighbouring values $\left\{X_{\boldsymbol{v}+\boldsymbol{k}}, \boldsymbol{k}=\left(k_{1}, k_{2}, \ldots, k_{d}\right)^{T} \neq \mathbf{0} \in \mathcal{K}\right\}$ with the error $\varepsilon_{\boldsymbol{v}}$. The equation (1.1) can be rewritten as

$$
P\left(T_{1}, T_{2}, \ldots, T_{d}\right) X_{v}=\varepsilon_{\boldsymbol{v}}
$$

by using a transfer function

$$
P\left(T_{1}, T_{2}, \ldots, T_{d}\right)=\sum_{k \in \mathcal{K}} \beta_{\boldsymbol{k}} T_{1}^{k_{1}} T_{2}^{k_{2}} \cdots T_{d}^{k_{d}}
$$

with forward shift operators

$$
T_{i} X_{v}=X_{v_{1}, \ldots, v_{i}+1, \ldots, v_{d}}, i=1,2, \ldots, d
$$

Spatial autoregressive model is roughly classified into two categories. One is unilateral spatial autoregressive model where the regression specified by the index set $\mathcal{K}$ only extends over a direction and the other is multilateral spatial autoregessive model where it extends over all directions.

## Unilateral Spatial Autoregressive Model

Unilateral spatial autoregressive model is, for example, defined as

$$
\sum_{k_{1}, k_{2}, \ldots, k_{d} \leq 0} \beta_{\boldsymbol{k}} X_{\boldsymbol{v}+\boldsymbol{k}}=\varepsilon_{\boldsymbol{v}}
$$

where $\beta_{\boldsymbol{k}}$ 's are regression coefficients and $\varepsilon_{\boldsymbol{v}}$ 's are error variables. This unilateral model is often called quarter-plane order autoregressive model because a value at a site is represented by values on the third quadrant when the origin is at the site on a two-dimensional plane. In other words, an order of sites is introduced in the model such that $\boldsymbol{u}>\boldsymbol{v}$ as $u_{i} \geq v_{i}$ for all $i=1, \ldots, d$ unless $\boldsymbol{u}=\boldsymbol{v}$ and the index set $\mathcal{K}$ for the neighbours is taken to be $\mathcal{K}_{q} \cup\{\mathbf{0}\}$ where $\mathcal{K}_{q}=\{\boldsymbol{k} ; \boldsymbol{k}<\mathbf{0}\}$ (Tjøstheim, 1978).

There is another popular unilateral model on a two-dimensional plane, which is referred as half-plane model in case of two-dimensional plane. In the model, the index set $\mathcal{K}$ is similar to the case of quarter-plane order and $\mathcal{K}_{-} \cup\{\mathbf{0}\}$ where $\mathcal{K}_{-}=\{\boldsymbol{k} ; \boldsymbol{k}<\mathbf{0}\}$ but the order of sites is different, which is defined as $\boldsymbol{u}>\boldsymbol{v}$ when either if $u_{1}>v_{1}$ or if $u_{1}=v_{1}$ and $u_{2}>v_{2}$. Then the values on two-dimensional plane can be rearranged on a line since the order is lexicographic (Whittle, 1954). The model is also extended for the case of general $d$-dimensional space (Dimitriou-Fakalou, 2009).

When $d=1$ and $\mathcal{K}$ is a finite set, unilateral spatial autoregressive model becomes well known time series autoregressive model,

$$
\sum_{k=-p}^{0} \beta_{k} X_{v+k}=\varepsilon_{v} \text { with } \beta_{0}=1
$$

## Multilateral Spatial Autoregressive Model

When the spatial dependence extends over all directions, the model is called multilateral spatial autoregressive model. We hereafter use the word "multilateral spatial autoregressive model" in a strict sense, excluding unilateral spatial autoregressive model. When $d=1$, we call multilateral spatial autoregressive model in particular bilateral spatial autoregressive
model since the spatial dependence exists only for the "past" and "future". In case of multilateral spatial autoregressive model, the orthogonality of $\varepsilon_{\boldsymbol{v}}$ 's and the orthogonality of $\varepsilon_{\boldsymbol{v}}$ to $X_{\boldsymbol{u}}$ for $\boldsymbol{u} \neq \boldsymbol{v}$ are not compatible. This is because the spatial dependence extends over all directions. Therefore, multilateral spatial autoregressive model is classified into two categories, simultaneous spatial autoregressive model and conditional spatial autoregressive model, by the choice of the orthogonality. We will discuss into detail multilateral spatial autoregressive model under the assumption of weak stationarity of $\left\{X_{v}\right\}$ in the next section.

### 1.2.2 Weak Stationarity of Spatial Autoregressive Model

Weak stationarity of a process or the model is one of keys for effective statistical inference. It is hard to derive any effective statistical inference without such an assumption, as far as a single set of observations are only available on a lattice. The stochastic process $\left\{X_{\boldsymbol{v}}, \boldsymbol{v} \in \mathcal{N}\right\}$ is weakly stationary if and only if any autocovariance between $X_{v}$ and $X_{v+\boldsymbol{k}}$ depends only on the lag $\boldsymbol{k}$. A necessary and sufficient condition for a spatial autoregressive model to be weakly stationary is given in the following condition. A primitive idea of this condition is in Whittle(1954) but not well known, even in the context of time series analysis. Therefore, we will give this fact as Proposition 1.1 with the proof.

## Condition 1.1.

The transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\sum_{k \in \mathcal{K}} \beta_{\boldsymbol{k}} z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}$ has no zeros on a domain $D=\left\{\left(z_{1}, z_{2}, \ldots, z_{d}\right) ;\left|z_{1}\right|=\left|z_{2}\right|=\cdots=\left|z_{d}\right|=1\right\}$ in $\mathbb{C}^{d}$.

Condition 1.1 is a bit mathematically sophisticated but the check is not so hard in practice. For example, it is easily seen that the condition is equivalent to $1+\beta_{1}+\beta_{-1} \neq 0,1-\beta_{1}-\beta_{-1} \neq 0$ and $\beta_{1} \neq \beta_{-1}$ or $\left|\beta_{1}\right|>1 / 2$
for $X_{v}+\beta_{1} X_{v+1}+\beta_{-1} X_{v-1}=\varepsilon_{v}$. Similar conditions can be found for any other case, even if they are not necessarily necessary condition. A numerical validation procedure of the condition would be to find out all roots of the transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ and check if they are on $D$ or not.

Two types of weakly stationary spatial autoregressive models are well known. One is so called Simultaneous Spatial Autoregressive (SAR) model (Whittle, 1954) and the other is Conditional Spatial Autoregessive (CAR) model (Besag, 1974). The difference is on the property of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$. An orthogonal process $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ is assumed for SAR model and a stationary process $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ such that any $X_{\boldsymbol{v}}$ is orthogonal to $\left\{\varepsilon_{\boldsymbol{v}+\boldsymbol{k}}, \boldsymbol{k} \neq \mathbf{0}\right\}$ is assumed for CAR model.

Therefore the role of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ is different from SAR model to CAR model. It is thought to be a disturbance in SAR model and an error or residual in CAR model. In fact, the error variable $\left\{\varepsilon_{v}\right\}$ in CAR model becomes $X_{v}-$ $E\left(X_{v} \mid X_{\boldsymbol{u}}, \boldsymbol{u} \neq \boldsymbol{v}\right)$ as far as $\left\{X_{v}\right\}$ is Gaussian. It is worthy of noting that such two properties of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ are not always compatible in case of spatial models. For other differences between SAR and CAR models, see Besag(1974), Cliff and $\operatorname{Ord}(1975)$ or $\operatorname{Haining}(1979)$.

In this thesis, we focus on the parameter estimation of SAR model, which is one of the oldest spatial autoregressive model but still there are many problems unsolved, for example, space domain analysis, efficiency of parameter estimator, random number generation or identifiability of the model.

Proposition 1.1. Condition 1.1 is a necessary and sufficient condition for spatial autoregressive model to be weakly stationary.

## Proof of Proposition 1.1

We first prove the sufficiency. If $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ has no zeros on $D$, then the

Laurent expansion

$$
P\left(z_{1}, z_{2}, \ldots, z_{d}\right)^{-1}=\sum_{j_{1}, j_{2}, \ldots, j_{d}=-\infty}^{\infty} \alpha_{\boldsymbol{j}} z_{1}^{j_{1}} z_{2}^{j_{2}} \cdots z_{d}^{j_{d}}
$$

is valid on a product of circular annuli $\left\{\left(z_{1}, z_{2}, \ldots, z_{d}\right) ; 1-\delta<\left|z_{i}\right|<1+\delta, i=\right.$ $1,2, . ., d\}$ for a $\delta>0$ (Shabat, 1992), where $\alpha_{j}$ is the coefficient of $z_{1}^{j_{1}} z_{2}^{j_{2}} \cdots z_{d}^{j_{d}}$ in the Laurent expansion of $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)^{-1}$. Then,

$$
\left|\alpha_{\boldsymbol{j}}\right|<K\left(1 \pm \frac{\delta}{2}\right)^{-j_{1}}\left(1 \pm \frac{\delta}{2}\right)^{-j_{2}} \cdots\left(1 \pm \frac{\delta}{2}\right)^{-j_{d}}
$$

for a constant $K$ where the + is for $j_{k}>0$ and the - is for $j_{k}<0$ on the right hand side of the inequality. We then have $\sum_{\boldsymbol{j}}\left|\alpha_{\boldsymbol{j}}\right|<\infty$ so that

$$
X_{\boldsymbol{v}}=P\left(T_{1}, T_{2}, \ldots, T_{d}\right)^{-1} \varepsilon_{\boldsymbol{v}}=\sum_{j_{1}, j_{2}, \ldots, j_{d}=-\infty}^{\infty} \alpha_{\boldsymbol{j}} \varepsilon_{\boldsymbol{v}+\boldsymbol{j}}
$$

is well defined. The weak stationarity of $\left\{X_{\boldsymbol{v}}\right\}$ is clear from that of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$.
We now prove the necessity. If $\left\{X_{\boldsymbol{v}}\right\}$ is weakly stationary, then Karhunen's general representation theorem of weakly stationary process (Grenander, 1981) yields us the representation,

$$
X_{\boldsymbol{v}}=\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i \boldsymbol{\omega}^{T} \boldsymbol{v}} \mathrm{~d} W_{\boldsymbol{\omega}}
$$

where $\left\{W_{\boldsymbol{\omega}}\right\}$ is an orthogonal increment process such that $E\left(\Delta W_{\boldsymbol{\omega}} \overline{\Delta W_{\boldsymbol{\omega}^{\prime}}}\right)=0$ for increment $\Delta W_{\boldsymbol{\omega}}=W_{\boldsymbol{\omega}+\Delta \boldsymbol{\omega}}-W_{\boldsymbol{\omega}}$ as far as $\boldsymbol{\omega} \neq \boldsymbol{\omega}^{\prime}$. Since $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ is also weakly stationary, it has a spectral representation,

$$
\varepsilon_{\boldsymbol{v}}=\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i \boldsymbol{\omega}^{T} \boldsymbol{v}} \mathrm{~d} W_{\boldsymbol{\omega}}^{\varepsilon}
$$

where $\left\{W_{\boldsymbol{\omega}}^{\varepsilon}\right\}$ is the other orthogonal increment process. We have then

$$
\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} P\left(e^{i \omega_{1}}, e^{i \omega_{2}}, \ldots, e^{i \omega_{d}}\right) e^{i \boldsymbol{\omega}^{T}} \boldsymbol{v} \mathrm{~d} W_{\boldsymbol{\omega}}=\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i \boldsymbol{\omega}^{T} \boldsymbol{v}} \mathrm{~d} W_{\boldsymbol{\omega}}^{\varepsilon}
$$

From Plancherel's theorem, we now have that

$$
\begin{equation*}
\left|P\left(e^{i \omega_{1}}, e^{i \omega_{2}} \ldots, e^{i \omega_{d}}\right)\right|^{2}\left\|\mathrm{~d} W_{\boldsymbol{\omega}}\right\|^{2}=\left\|\mathrm{d} W_{\boldsymbol{\omega}}^{\varepsilon}\right\|^{2} \tag{1.2}
\end{equation*}
$$

for the norm $\|X\|^{2}=E|X|^{2}$. This implies that Condition 1.1 is satisfied because the spectral desnity $f_{\varepsilon}(\boldsymbol{\omega})=\left\|\mathrm{d} W_{\boldsymbol{\omega}}^{\varepsilon}\right\|^{2} / \mathrm{d} \boldsymbol{\omega}$ of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ being positive.

From the proof, we see that the spectral density of SAR model is given by

$$
f(\boldsymbol{\omega})=\frac{\sigma^{2}}{\left|P\left(z_{1}, z_{2}, \ldots, z_{d}\right)\right|^{2}},
$$

since $\left\|\mathrm{d} W_{\omega}^{\varepsilon}\right\|^{2}=\sigma^{2} \mathrm{~d} \boldsymbol{\omega}$.
In case of CAR model, there is a requirement that $\varepsilon_{v}$ is always orthogonral to $X_{\boldsymbol{u}}$ for $\boldsymbol{u} \neq \boldsymbol{v}$. Let us write

$$
\varepsilon_{\boldsymbol{v}}=Q\left(T_{1}, T_{2}, . ., T_{d}\right) \xi_{v}
$$

with a transfer function $Q\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ and an orthogonal process $\left\{\xi_{v}\right\}$. Then $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\left|Q\left(z_{1}, z_{2}, \ldots, z_{d}\right)\right|^{2}$ is equivalent to the orthogonality requirement. This implies that the transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ should take only non-negative values. A common practice is to assume that symmetric regression coefficients, $\beta_{\boldsymbol{k}}=\beta_{-\boldsymbol{k}}$ for all $\boldsymbol{k} \in \mathcal{K}$ and $1+$ $2 \sum_{\boldsymbol{k} \in \mathcal{K}^{+}} \beta_{\boldsymbol{k}} \cos \left(\boldsymbol{k}^{T} \boldsymbol{\omega}\right)>0$, where $\mathcal{K}^{+}$is the positive half space of $\mathcal{K}$ (Besag, 1974; Besag and Moran, 1975; Besag, 1977). The assumption also leads us to a positive spectral density $\left\|\mathrm{d} W_{\boldsymbol{\omega}}^{\varepsilon}\right\|^{2} / \mathrm{d} \boldsymbol{\omega}=\sigma^{2} P\left(e^{i \omega_{1}}, e^{i \omega_{2}}, \ldots, e^{i \omega_{d}}\right)$ of $\left\{\varepsilon_{\boldsymbol{v}}\right\}$, which accommodates with the assumption on $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ in Section 1.2.1. Note that the variance of $\xi_{v}$ is equal to that of $\varepsilon_{\boldsymbol{v}}$. Now we can cancel one $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ out from the both sides of (1.2) and we have

$$
P\left(e^{i \omega_{1}}, e^{i \omega_{2}}, \ldots, e^{i \omega_{d}}\right)\left\|\mathrm{d} W_{\boldsymbol{\omega}}\right\|^{2}=\sigma^{2} \mathrm{~d} \boldsymbol{\omega}
$$

so that the spectral density of CAR model is given by

$$
f(\boldsymbol{\omega})=\sigma^{2}\left(1+2 \sum_{\boldsymbol{k} \in \mathcal{K}^{+}} \beta_{\boldsymbol{k}} \cos \left(\boldsymbol{k}^{T} \boldsymbol{\omega}\right)\right)^{-1} .
$$

We here note that the autocovariance $\gamma_{\boldsymbol{h}}=E\left(X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{k}}\right)$ of a weakly stationary process $\left\{X_{v}\right\}$ also has a spectral representation,

$$
\gamma_{\boldsymbol{h}}=\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i \boldsymbol{\omega}^{T} h} f(\boldsymbol{\omega}) \mathrm{d} \boldsymbol{\omega} .
$$

### 1.2.3 Spatial Lag Model

Spatial lag model frequently used in spatial econometrics is also a type of spatial autoregressive model. However, the model is often mixed up with the SAR (simultaneous spatial autoregressive) model already introduced and abbreviated as SAR model (LeSage and Pace, 2009), too. In fact the spatial lag model looks similar to SAR model but not the same. We will clarify the similarity and the difference in this section.

Suppose that $Y$ is a real-valued $N \times p$ matrix of $N$ sets of the values of $p$ explanatory variables. Spatial lag model is a model

$$
\boldsymbol{x}=\rho W \boldsymbol{x}+Y \boldsymbol{\beta}+\boldsymbol{\varepsilon}
$$

for $N$-dimensional vector $\boldsymbol{x}$ of observations with unknown scalar parameter $\rho$ and $p$-dimensional parameter vector $\boldsymbol{\beta}$. The vector $\boldsymbol{\varepsilon}$ is assumed to be a vector of realisations of independent and identically distributed variables with mean zero and variance $\sigma^{2}$ (Cliff and Ord, 1973; Cliff and Ord, 1981; Ord, 1975). Spatial correlation is described by the $N \times N$ spatial-weighting matrix $W$. Typical choices of $W$ are the following (Tiefelsdorf et al., 1999).

1. Adjacency matrix $W=\left(w_{i, j}\right)$ such that

$$
w_{i, j}= \begin{cases}1 & \text { if } i \text { and } j \text { are adjacent }, \\ 0 & \text { otherwise }\end{cases}
$$

2. Inverse-distance matrix $W=\left(w_{i, j}\right)$ such that $w_{i, j}=1 / d_{i, j}$ for the distance $d_{i, j}$ between $i$ and $j$ (Tobler, 1970).

As is easily seen, spatial lag model is a simple extension of multiple regression model. The observation vector $\boldsymbol{x}$ can be thought of a vector of
realisations of the SAR model if $\boldsymbol{\beta}$ is $\mathbf{0}$. For example, in case of adjacency matrix $W$, realisations of the SAR model can be written in the form of spatial lag model except observations on the edge of the lattice, by taking $\mathcal{K}=\{(-1,-1),(-1,0),(-1,1),(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1)\}$ and setting $\beta_{\boldsymbol{k}}=-\rho$ for any $\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}$. However, no such stochastic process is assumed behind spatial lag model, so that no attempt has been done to evaluate the behaviour of parameter estimator, for example, consistency or asymptotic efficiency of the maximum likelihood estimator, which will be proved in Chapter 2 of this thesis.

### 1.3 Parameter Estimation of SAR model

A typical estimation procedure of unknown parameters $\left\{\beta_{\boldsymbol{k}}\right\}$ and $\sigma$ is based on the likelihood under the assumptions of Gaussianity and weak stationarity.

### 1.3.1 Aprroximations of the Likelihood of SAR model

Let assume that $N=n_{1} n_{2} \cdots n_{d}$ observations are available on a lattice $\mathcal{N}=$ $\left\{\boldsymbol{v}=\left(v_{1}, v_{2}, \ldots, v_{d}\right)^{T} ; 1 \leq v_{j} \leq n_{j}, j=1, \ldots, d\right\}$. The exact log-likelihood of the observational vector $\boldsymbol{x}$ under Gaussianity is then

$$
L=-\frac{1}{2} \log \operatorname{det}(\Sigma)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} \Sigma^{-1} \boldsymbol{x}
$$

where $\Sigma$ is the covariance matrix of $\boldsymbol{x}$. However, there remain several problems in direct evaluation of $L$, particularly in the evaluation of $\operatorname{det}(\Sigma)$, since it is a complicated function of the parameters $\left\{\beta_{k}\right\}$ and $\sigma$ of SAR model, although there exist exceptional cases. Therefore a variety of approximations have been proposed since then. The need of a good approximation is not only for the easiness of computation but also for the simplicity of the evaluation of the behaviour of parameter estimator.

## Whittle's Approximation

Whittle (1954) proposed a maximum likelihood estimator of parameters
based on an approximation of the log-likelihood when he first introduced SAR model. He approximated the log-likelihood as

$$
\begin{equation*}
\frac{1}{4 \pi^{2}} \iint\left[\log f\left(\omega_{1}, \omega_{2}\right)+\frac{F\left(\omega_{1}, \omega_{2}\right)}{f\left(\omega_{1}, \omega_{2}\right)}\right] \mathrm{d} \omega_{1} \mathrm{~d} \omega_{2} \tag{1.3}
\end{equation*}
$$

where $f\left(\omega_{1}, \omega_{2}\right)$ is the spectral density function and $F\left(\omega_{1}, \omega_{2}\right)$ is the periodogram,

$$
F\left(\omega_{1}, \omega_{2}\right)=\sum_{j_{1}=-n_{1}}^{n_{1}} \sum_{j_{2}=-n_{2}}^{n_{2}} \hat{\gamma}_{j_{1}, j_{2}} e^{i\left(j_{1} \omega_{1}+j_{2} \omega_{2}\right)}
$$

and $\hat{\gamma}_{j_{1}, j_{2}}$ is the empirical covariance,

$$
\hat{\gamma}_{j_{1}, j_{2}}=\frac{1}{n_{1} n_{2}} \sum_{v_{1}=1}^{n_{1}-j_{1}} \sum_{v_{2}=1}^{n_{2}-j_{2}} X_{v_{1}, v_{2}} X_{v_{1}+j_{1}, v_{2}+j_{2}} .
$$

However, it turned out true that the estimator based on Whittle's approximation is not $\sqrt{N}$-consistent. To overcome this problem, various improvements have been proposed since then.

## Guyon's Approximation

Guyon (1982) proposed a replacement of $\hat{\gamma}_{j}$ by

$$
\tilde{\gamma}_{j}=\frac{1}{\left(n_{1}-\left|j_{1}\right|\right)\left(n_{2}-\left|j_{2}\right|\right) \cdots\left(n_{d}-\left|j_{d}\right|\right)} \sum_{v_{1}=1}^{n_{1}-j_{1}} \sum_{v_{2}=1}^{n_{2}-j_{2}} \cdots \sum_{v_{d}=1}^{n_{d}-j_{d}} X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{j}}
$$

to retain $\sqrt{N}$-consistency. However, the estimator based on his approximation is not asymptotically efficient because $\tilde{\gamma}_{j}$ has large variance for large lags as is shown in Mardia and Marshall (1984).

## Dahlhaus and Künsch's Approximation

Dahlhaus and Künsch (1987) introduced a tapering into $\tilde{\gamma}_{j}$ to remove the large variance for large lags. Their tapering function $h(u), u \in[0,1]$ with smoothness parameter $\rho$ is

$$
h(u)= \begin{cases}w(2 u / \rho) & \left(0 \leq u<\frac{1}{2} \rho\right) \\ 1 & \left(\frac{1}{2} \rho \leq u \leq \frac{1}{2}\right) \\ h(1-u) & \left(\frac{1}{2}<u \leq 1\right)\end{cases}
$$

for a continuous increasing function $w$ with $w(0)=0$ and $w(1)=1$. The tapered empirical covariance is then

$$
\begin{aligned}
\gamma_{\boldsymbol{j}}^{D H}= & \left\{\prod_{i=1}^{d} \sum_{s=1}^{n} h\left\{\left(s-\frac{1}{2}\right) / n_{i}\right\}^{2}\right\}^{-1} \sum_{v_{1}=1}^{n_{1}-j_{1}} \sum_{v_{2}=1}^{n_{2}-j_{2}} \cdots \sum_{v_{d}=1}^{n_{d}-j_{d}} X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{j}} \\
& \times\left[\prod_{i=1}^{d} h\left\{\left(t_{i}-\frac{1}{2}\right) / n_{i}\right\} h\left\{\left(t_{i}+k_{i}-\frac{1}{2}\right) / n_{i}\right\}\right] .
\end{aligned}
$$

## Robinson and Vidal Sanz's Approximation

Robinson and Vidal Sanz (2006) proposed a discrete approximation of the integrals in (1.3). The estimator is obtained by a recursion. They also proved asymptotic efficiency by introduction of a truncated version of the periodogram,

$$
F_{R}(\boldsymbol{\omega})=\sum_{\left|j_{1}\right| \leq g\left(n_{1}\right)} \sum_{\left|j_{2}\right| \leq g\left(n_{2}\right)} \cdots \sum_{\left|j_{d}\right| \leq g\left(n_{d}\right)} \tilde{\gamma}_{k} e^{i\left(\boldsymbol{\omega}^{T} j\right)}
$$

where $g(x)$ is a positive, integer-valued, monotonically increasing function such that $g(x)$ goes to infinity as $x$ tends to infinity and for all sufficiently large positive $x$, some $C>1$ exists such that $g(x) \leq C x$.

## Kent and Mardia's Approximation

A different approach from Whittle is by Kent and Mardia (1996). They proposed an approximation of the log-likelihood,

$$
-\frac{1}{2} \log \operatorname{det}(C)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} C^{-1} \boldsymbol{x}
$$

by using a circulant matrix $C$. An advantage of employing such a circulant matrix is that $\operatorname{det}(C)$ or $C^{-1}$ is a simple function of the autocovariances $\gamma_{\boldsymbol{h}}$ 's. They proved that their approximation gives us a good approximation of the log-likelihood up to the second derivatives but no estimation procedure nor evaluation of the behaviour are given. A practical problem would be how to evaluate the values of $\gamma_{j}$ 's or its derivatives although they suggested the use
of spectral density.

All approximations above are common in frequency domain analysis. In other words, they are all based on the underlying spectral density. Another type of approximation is on space domain.

## Cressie's Approximation

Cressie (1993) suggested an approximation by writing the observation vector $\boldsymbol{x}$ as

$$
B \boldsymbol{x}=\varepsilon
$$

where $\varepsilon$ is a vector arranged $\varepsilon_{\boldsymbol{v}}$ for $\boldsymbol{v} \in \mathcal{N}$ and $B$ is the spatial dependence matrix. This is the same idea as in spatial lag model but already an approximation to the SAR model where the values on the edge of $\mathcal{N}$ are not represented well. Then, an approximate maximum likelihood estimator can be obtained, in principle, by maximising

$$
\begin{equation*}
\frac{1}{\left(2 \pi \sigma^{2}\right)^{\frac{N}{2}}}|B| \exp \left(-\frac{\boldsymbol{x}^{T} B^{T} B \boldsymbol{x}}{2 \sigma^{2}}\right) \tag{1.4}
\end{equation*}
$$

If we restrict our attention into a special case $B=I-\rho W$ as in spatial lag model, diagonally dominant sparce matrix approximation can be used (Gaetan and Guyon, 2010; Banerjee et al., 2015, Pace and Barry, 1997a,b), but no general efficient algorithm known for the evaluation of the value of (1.4) as far as we have investigated.

The approximation proposed in this thesis is an approximation on space domain, where a circulant matrix approximation is used for the transformation from observation to error vector. The approximation is simple and applicable for any type of SAR model without restriction. The consistency and asymptotic efficiency of the resulting parameter estimator is proved together with the result of small sample size numerical experiment in

Chapter 2. We thus see that the approximation error caused by the circulant matrix approximation is quite marginal.

### 1.3.2 A Space Domain Approximation

Space domain approximation proposed here is

$$
L_{A}=\frac{1}{2} \log \operatorname{det}(A)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} \tilde{A} \boldsymbol{x}
$$

where

$$
A=\frac{1}{\sigma^{2}} \sum_{k, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes \cdots \otimes W_{n_{d}}^{-k_{d}+k_{d}^{\prime}}
$$

and

$$
\tilde{A}=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \cdots \alpha_{n_{d}}^{\left|k_{d}-k_{d}^{\prime}\right|} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes \cdots \otimes W_{n_{d}}^{-k_{d}+k_{d}^{\prime}}
$$

with Kronecker product $\otimes$ and $\alpha_{n_{i}}=1+1 / n_{i}$ for $i=1,2, \ldots, d$. The matrix $W_{n}$ is an $n \times n$ circulant matrix such that the off-diagonal elements $\left(W_{n}\right)_{j, j+1}$ are all 1 for $j=1,2, \ldots, n-1,\left(W_{n}\right)_{n, 1}=1$ 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 idea behind this approximation is similar to that for time series AR model. By transforming the observation to error vector, the log-likelihood is simplified. However, it is not enough in the case of SAR model. This is because the transformation matrix is triangular in case of time series AR model but not so in case of SAR model. Therefore, we approximated the transformation matrix by a circulant matrix and the approximation of the inverse of the covariance matrix by $A$ is obtained as a result. The shrinkage factor $\alpha_{n}$ introduced in $\tilde{A}$ is to retain the asymptotic efficiency of the parameter estimator which maximises $L_{A}$. The reason why the factor $\alpha_{n}$ is called "shrinkage factor" here is that it plays a role to shrink $\tilde{A}^{-1}$.

Numerical experiments in Section 2.7 show that our estimate has less bias and variance than Guyon's estimate. In the numerical experiments, we propose a random number generation although it is limited to the case where the transfer function is decomposed as the product of a polynomial of $z_{1}$ and a polynomial of $z_{2}$. The idea is transforming from the error variables to the observation variables by using the inverse of transfer function. The precision is verified by the comparison between the periodograom and the spectral density.

### 1.4 Non-identifiability of SAR Model

Space domain approximation we have proposed in Chapter 2 provides us an optimal estimation procedure. However, a problem behind SAR model is that the model is not always identifiable for given 2nd moments or the spectral density. In other words several different SAR models share the same spectral density. Therefore multiple solutions of the maximum likelihood equation may exist since the Gaussian likelihood is completely determined by the 2nd moments or the spectral density. In fact, many different estimators are available depending on the initial values for optimisation algorithm used. A remedy would be to introduce some restriction on parameter space. Such a restriction may come from the meaning of the SAR model. Otherwise introduction of unilateral or symmetric SAR model would be worthwhile. The non-identifiability of SAR model causes another problem, singularity of Fisher information matrix. The Fisher information matrix becomes singular if some of SAR models which share the same spectral density are not isolated. Good properties of the maximum likelihood estimator will be lost if it happens. Some of parameters then become un-estimable. It would be indispensable to check if the Fisher information is non-singular in practice since the check of non-isolation of some of models which share the same spectral density is not so easy. Several types of necessary and
sufficient conditions for the singularity of Fisher information matrix are given in Section 3.4.

## Chapter 2

## A Simple Statistically Efficient Approximation of the Gaussian Likelihood of SAR Model

A good approximation of the Gaussian likelihood of SAR model is proposed. The approximation yields us an asymptotically efficient estimator of the parameters. It is a straightforward approximation on space domain so that the estimation procedure becomes transparent and applicable for any SAR model without restriction on the parameter space except the stationarity. Numerical experiments show that our estimator has less bias and variance than the other estimator, although our comparison is limited to the case where random number generation and the calculation of the other estimate are both feasible at this stage.

### 2.1 Introduction

It is well known that the exact likelihood of SAR model has no closed form in the parameters even when the Gaussianity is assumed, as the covariance function is not a simple function of the parameters. Historically a lot of approximations of the log-likelihood have been proposed. Whittle (1954) is a pioneer. His approximation however does not seem feasible for practical use, since the estimator is not only $\sqrt{N}$-consistent but also the estimation
procedure requires multiple integrations on a high-dimensional unit sphere.
There are various proposals for the improvements of the method. The use of a modified periodogram is by Guyon(1982) to retain the $\sqrt{N}$-consistency and a tapered use of the periodogram is by Dahlhaus and Künsch (1987) to improve efficiency. As far as we know, Robinson (2006) only proposed a discrete approximation of integrations needed for the calculation.

The aim of this chapter is to derive a simple approximation on space domain, that is, the approximation without using spectral density, which yields us an asymptotically efficient estimator. The order of the approximation is evaluated in Section 2.4. The consistency of the estimator $\hat{\boldsymbol{\theta}}$ is shown in Theorem 2.2 in Section 2.5. Theorem 2.3 in Section 2.6 shows that the estimator $\hat{\boldsymbol{\theta}}$ is asymptotically efficient as far as the Fisher information matrix is non-singular for large enough $N$. The results of numerical experiments are given in Section 2.7, where our estimator is compared with that by Guyon(1982).

### 2.2 SAR Model

We have already introduced SAR model in Chapter 1 but we define it again to clarify the framework of the following mathematical analysis. The SAR model considered here is for a real valued weakly stationary random field $\left\{X_{\boldsymbol{v}} ; \boldsymbol{v}=\left(v_{1}, v_{2}\right)^{T} \in \mathbb{Z}^{2}\right\}$ on a two dimensional lattice with the mean zero and the autocovariance function $\gamma_{\boldsymbol{h}}=E\left(X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{h}}\right), \boldsymbol{h} \in \mathbb{Z}^{2}$. The model is that $\left\{X_{v}\right\}$ satisfies the equation

$$
\begin{equation*}
P\left(T_{1}, T_{2}\right) X_{v}=\varepsilon_{\boldsymbol{v}} \tag{2.1}
\end{equation*}
$$

where $\left\{\varepsilon_{\boldsymbol{v}} ; \boldsymbol{v} \in \mathbb{Z}^{2}\right\}$ is a set of orthogonal random variables with the mean zero and the variance $\sigma^{2}, \sigma>0$. The operator

$$
P\left(T_{1}, T_{2}\right)=\sum_{k \in \mathcal{K}} \beta_{\boldsymbol{k}} T_{1}^{k_{1}} T_{2}^{k_{2}}
$$

is a two dimensional transfer function with $\beta_{\mathbf{0}}=1$. The index set $\mathcal{K}=\{\boldsymbol{k}=$ $\left.\left(k_{1}, k_{2}\right)^{T}\right\}$ is a finite set of lattice points on $\mathbb{Z}^{2}$ and the operators $T_{1}$ and $T_{2}$ are forward shift operators such as

$$
T_{1} X_{\boldsymbol{v}}=X_{v_{1}+1, v_{2}} \text { and } T_{2} X_{\boldsymbol{v}}=X_{v_{1}, v_{2}+1}
$$

The following stationarity assumption, which is already explained in Section 1.2.2 in a general manner, is assumed throughout this chapter for SAR model since the underlying random field $\left\{X_{v}\right\}$ is weakly stationary.

Assumption 2.1. $P\left(z_{1}, z_{2}\right)=\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} z_{1}^{k_{1}} z_{2}^{k_{2}}$ has no zeros on a domain $\left\{\left(z_{1}, z_{2}\right) ;\left|z_{1}\right|=\left|z_{2}\right|=1\right\}$ in $\mathbb{C}^{2}$.

### 2.3 An Approximation of the Log-likelihood

Let us assume that the observations $\left\{x_{\boldsymbol{v}}, \boldsymbol{v} \in \mathcal{N}\right\}$ are available on a rectangular lattice $\mathcal{N}=\left\{\boldsymbol{v}=\left(v_{1}, v_{2}\right)^{T} ; 1 \leq v_{1} \leq n_{1}, 1 \leq v_{2} \leq\right.$ $\left.n_{2}\right\}$. The $N=n_{1} n_{2}$ observations are arranged as a vector $\boldsymbol{x}=$ $\left(x_{1,1}, x_{1,2}, \ldots, x_{1, n_{2}}, x_{2,1}, \ldots, x_{n_{1}, n_{2}}\right)^{T}$ in lexicographic order. To simplify the notation, we put together the parameters $\beta_{\boldsymbol{k}}, \boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}$ and $\sigma$ into an $m$ dimensional vector $\boldsymbol{\theta} \in \Theta$, where $\mathcal{K}$ is an index set which specifies neighbouring sites in relative to each site $\boldsymbol{v}$. We hereafter assume that the parameter space for each $\beta_{\boldsymbol{k}}$ is a bounded open set for any $\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}$ and that for $\sigma$ is $\left\{\sigma ; \sigma>\sigma_{0}\right\}$ for a $\sigma_{0}>0$ and the whole parameter space $\Theta$ is a product of them.

The log-likelihood $L$ of the observation $\boldsymbol{x}$ under Gaussianity is written as

$$
L=-\frac{1}{2} \log \operatorname{det}(\Sigma)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} \Sigma^{-1} \boldsymbol{x}
$$

where $\Sigma$ is the $N \times N$ covariance matrix of $\boldsymbol{x}$. As is shown later, in case of SAR model, $\operatorname{det}(\Sigma)$ or $\Sigma^{-1}$ does not take any simple form of the parameters. It would not be a good idea to directly calculate them. The approximation
we propose here

$$
\begin{equation*}
L_{A}=\frac{1}{2} \log \operatorname{det}(A)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} \tilde{A} \boldsymbol{x} \tag{2.2}
\end{equation*}
$$

where

$$
\begin{equation*}
A=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}} \tag{2.3}
\end{equation*}
$$

and

$$
\tilde{A}=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}
$$

with Kronecker product $\otimes$. We introduced such shrinkage factors $\alpha_{n_{1}}=$ $1+1 / n_{1}$ and $\alpha_{n_{2}}=1+1 / n_{2}$ to retain the asymptotic efficiency. The matrix $W_{n}$ is an $n \times n$ circulant matrix such that the off-diagonal elements $\left(W_{n}\right)_{j, j+1}$ are all 1 for $j=1,2, \ldots, n-1,\left(W_{n}\right)_{n, 1}=1$ 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.

Key of the approximation $L_{A}$ is that the circulant matrix $A$ well approximates $\Sigma^{-1}$. The idea is almost the same as that for the approximation of the likelihood of time series autoregressive model, which has the conditional likelihood

$$
\frac{\left\{\operatorname{det}\left(B^{T} B\right)\right\}^{1 / 2}}{\left(\sqrt{2 \pi \sigma^{2}}\right)^{n}} \exp \left(-\frac{1}{2 \sigma^{2}} \boldsymbol{x}^{T} B^{T} B \boldsymbol{x}\right)
$$

of $\boldsymbol{x}=\left(x_{1}, \ldots, x_{n}\right)^{T}$ given $x_{-p+1}=x_{-p+2}=\cdots=x_{0}=0$ where

$$
B=\left(\begin{array}{cccccccc}
1 & & & & & & & \\
\beta_{1} & \ddots & & & & & & \\
\vdots & \ddots & \ddots & & & & & \\
\beta_{p} & & \ddots & \ddots & & & 0 & \\
& \ddots & & \ddots & \ddots & & & \\
& & \ddots & & \ddots & \ddots & & \\
& 0 & & \ddots & & \ddots & \ddots & \\
& & & & \beta_{p} & \cdots & \beta_{1} & 1
\end{array}\right) .
$$

However, things are much more complicated in SAR model. The $n \times n$ transformation matrix $B$ is not simple as in time series. To explain our approximation, let us consider one dimensional bilateral model with the transfer function $P(z)=\sum_{k=-p_{1}}^{p_{2}} \beta_{k} z^{k}$ as an example. The following circulant matrix $\mathcal{B}$ plays a role of $B$ in time series autoregressive model,

$$
\mathcal{B}=\sum_{k=-p_{1}}^{p_{2}} \beta_{k} W_{n}^{k}=\left(\begin{array}{ccccccccc}
1 & \beta_{1} & \cdots & \beta_{p_{2}} & & & \beta_{-p_{1}} & \cdots & \beta_{-1} \\
\beta_{-1} & \ddots & \ddots & & \ddots & & & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & & \ddots & & 0 & \\
\beta_{-p_{1}} & & \ddots & \ddots & \ddots & & \ddots & & \\
& \ddots & & \ddots & \ddots & \ddots & & \ddots & \\
& & \ddots & & \ddots & \ddots & \ddots & & \beta_{p_{1}} \\
\beta_{p_{2}} & & 0 & & \ddots & & \ddots & \ddots & \ddots \\
\vdots & \ddots & & & \ddots & & \ddots & \ddots & \beta_{1} \\
\beta_{1} & \cdots & \beta_{p_{2}} & & & \beta_{-p_{1}} & \cdots & \beta_{-1} & 1
\end{array}\right)
$$

given $x_{-p_{1}+1}=x_{-p_{1}+2}=\cdots=x_{0}=0$ and $x_{n+1}=x_{n+2}=\cdots=x_{n+p_{2}}=0$. The extra triangular submatrices on the lower left and the upper right corner of $\mathcal{B}$ are introduced only for the simplicity of the expression of $\mathcal{B}$. Then, the product

$$
\mathcal{B}^{T} \mathcal{B}=\sum_{k=-p_{1}}^{p_{2}} \sum_{k^{\prime}=-p_{1}}^{p_{2}} \beta_{k} \beta_{k^{\prime}} W_{n}^{-k+k^{\prime}}
$$

is $\sigma^{2} A$ in (2.3) in case of $n_{1}=n$ and $n_{2}=1$.
An advantage of using such a circulant matrix is the easiness of the calculation of the eigenvalues or of the determinant. In fact, the eigenvalues of $A$ are easily obtained by a simple Fourier transform of $\beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}}$ as is shown in (2.20) in Section 2.9.1, and

$$
\log \operatorname{det}(A)=2 \sum_{\boldsymbol{j} \in \mathcal{N}} \log \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \exp \left(-i \boldsymbol{\omega}_{n, \boldsymbol{j}}^{T} \boldsymbol{k}\right)-N \log \sigma^{2}
$$

gives us a good approximation of $-\log \operatorname{det}(\Sigma)$. As is seen in Section 2.9.3, the gradient and Hessian of $L_{A}$ have the explicit representations, so that
the implementation of our algorithm is straightforward. Moreover, the simple representation of the eigenvalues helps us to prove goodness of the approximation, consistency and asymptotic efficiency of the estimator, which are shown in Section 2.4, 2.5 and 2.6.

The reason why the matrix $\tilde{A}$ is used in place of $A$ for the quadratic form of $\boldsymbol{x}$ in (2.2) is to ensure the asymptotic efficiency of the maximum likelihood estimator of parameters based on $L_{A}$. The approximation $L_{A}$ without a shrinkage factor like $\alpha_{n}=1+1 / n$ does not yield us asymptotically efficient estimator since there is a significant edge effect of observations on a lattice.

A practical estimation procedure would be to find out the parameter $\boldsymbol{\theta}$ which maximises $L_{A}$, for example, by using a function nlminb in R for optimisation with boundaries. The calculation of $A$ or $\tilde{A}$ in $L_{A}$ is straightforward. The gradient and Hessian of $L_{A}$ given in Section 2.9.3 will be helpful for accelerating the convergence of the algorithm.

### 2.4 Goodness of the Approximation

We first note that the spectral density and its derivatives are all bounded and bounded away from zero up to the second order. The spectral density of SAR model is given by

$$
f(\boldsymbol{\omega})=\frac{\sigma^{2}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}}=\sum_{\boldsymbol{h} \in \mathbb{Z}^{2}} \gamma_{\boldsymbol{h}} z_{1}^{h_{1}} z_{2}^{h_{2}}
$$

where $\boldsymbol{\omega}=\left(\omega_{1}, \omega_{2}\right)^{T}$, and $z_{1}=\exp \left(i \omega_{1}\right)$ and $z_{2}=\exp \left(i \omega_{2}\right)$. The boundedness and positivity of $f(\boldsymbol{\omega})$ is clear from the continuity. The boundedness of the derivatives,
$0<\left|\partial_{\theta_{p}} f(\boldsymbol{\omega})\right|<\mathrm{c}_{1}$ and $0<\left|\partial_{\theta_{p} \theta_{q}}^{2} f(\boldsymbol{\omega})\right|<\mathrm{c}_{2}$ for a positive constant $c_{1}, c_{2}$
follow from the formulas,
$\partial_{\beta_{s}} f(\boldsymbol{\omega})=-2 f(\boldsymbol{\omega}) \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right), \quad \partial_{\sigma} f(\boldsymbol{\omega})=\frac{2 f(\boldsymbol{\omega})}{\sigma}$,
$\partial_{\beta_{s} \beta_{s^{\prime}}}^{2} f(\boldsymbol{\omega})=2 f(\boldsymbol{\omega})\left\{2 \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right) \operatorname{Re}\left(\frac{z_{1}^{s_{1}^{\prime}} z_{2}^{s_{2}^{\prime}}}{P\left(z_{1}, z_{2}\right)}\right)+\operatorname{Re}\left(\frac{z_{1}^{s_{1}+s_{2}^{\prime}} z_{2}^{s_{2}+s_{2}^{\prime}}}{P\left(z_{1}, z_{2}\right)^{2}}\right)\right\}$,
$\partial_{\beta_{s} \sigma}^{2} f(\boldsymbol{\omega})=\frac{2 \partial_{\beta_{s}} f(\boldsymbol{\omega})}{\sigma} \quad$ and $\quad \partial_{\sigma \sigma}^{2} f(\boldsymbol{\omega})=\frac{2 f(\boldsymbol{\omega})}{\sigma^{2}}$.
Here, $\partial_{\theta_{p}}$ and $\partial_{\theta_{p} \theta_{q}}^{2}$ are simplified notations for differential operators $\partial / \partial \theta_{p}$ and $\partial^{2} / \partial \theta_{p} \partial \theta_{q}$ respectively. We will also use two types of matrix norms. One is the operator norm,

$$
\|M\|=\max _{j}\left(\mu_{j}\right)
$$

and the other is the trace class norm,

$$
\|M\|_{1}=\sum_{j=1}^{m} \mu_{j}
$$

where $\mu_{j} \geq 0, j=1,2, \cdots, m$ are singular values of $m \times m$ matrix $M$. Then, the order of the approximation $L_{A}$ parallels the result of Kent and Mardia (1996).

Theorem 2.1. The log-likelihood $L_{A}$ is asymptotically equivalent to the exact log-likelihood $L$ in the sense that,
(i) $\frac{1}{N}\left(L-L_{A}\right)=\mathrm{O}_{\mathrm{P}}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)$,
(ii) $\frac{1}{N}\left(\partial_{\theta_{p}} L-\partial_{\theta_{p}} L_{A}\right)=\mathrm{O}_{\mathrm{P}}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)$,
(iii) $\frac{1}{N}\left(\partial_{\theta_{p} \theta_{q}}^{2} L-\partial_{\theta_{p} \theta_{q}}^{2} L_{A}\right)=\mathrm{O}_{\mathrm{P}}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)$
for any $\theta_{p}$ and $\theta_{q}$ as $\min \left(n_{1}, n_{2}\right)$ tends to infinity.

Proof. We first show the (i). Since the first term on the right hand side of

$$
\frac{2}{N}\left(L-L_{A}\right)=\frac{1}{N}\left(\log \operatorname{det}\left(A^{-1}\right)-\log \operatorname{det}(\Sigma)\right)+\frac{1}{N} \boldsymbol{x}^{T}\left(\tilde{A}-\Sigma^{-1}\right) \boldsymbol{x}
$$

is of the order of $1 / n_{1}+1 / n_{2}$ as is shown in Proposition A2 of Section 2.9.1, it is now enough to note that

$$
\begin{equation*}
\left\|\tilde{A}-\Sigma^{-1}\right\|_{1}=\mathrm{O}\left(n_{1}+n_{2}\right) \tag{2.4}
\end{equation*}
$$

in view of Lemma 4.2 of Kent and Mardia (1996). The evaluation on (2.4) is shown in Proposition A3 in Section 2.9.1.

For the proofs of (ii) and (iii), it is enough to apply similar techniques used for the proof of (i) to

$$
\begin{gathered}
\frac{2}{N}\left(\partial_{\theta_{p}} L-\partial_{\theta_{p}} L_{A}\right)=\frac{1}{N}\left(\partial_{\theta_{p}} \log \operatorname{det}\left(A^{-1}\right)-\partial_{\theta_{p}} \log \operatorname{det}(\Sigma)\right) \\
+\frac{1}{N} \boldsymbol{x}^{T}\left(\partial_{\theta_{p}} \tilde{A}-\partial_{\theta_{p}}\left(\Sigma^{-1}\right)\right) \boldsymbol{x}
\end{gathered}
$$

and

$$
\begin{aligned}
\frac{2}{N}\left(\partial_{\theta_{p} \theta_{q}}^{2} L-\partial_{\theta_{p} \theta_{q}}^{2} L_{A}\right)= & \frac{1}{N}\left(\partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}\left(A^{-1}\right)-\partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(\Sigma)\right) \\
& +\frac{1}{N} \boldsymbol{x}^{T}\left(\partial_{\theta_{p} \theta_{q}}^{2} \tilde{A}-\partial_{\theta_{p} \theta_{q}}^{2}\left(\Sigma^{-1}\right)\right) \boldsymbol{x}
\end{aligned}
$$

Before showing the asymptotic efficiency of $\hat{\boldsymbol{\theta}}$ which maximise $L_{A}$, we establish the consistency. It is worthy of noting that the approximations given in Kent and Mardia (1996) or that in Whittle (1954) does not necessarily yield us asymptotically efficient estimator (Kent and Mardia, 1996) although consistent.

### 2.5 Consistency

We need the following assumption for the consistency.

Assumption 2.2. The limit of the Fisher information matrix $I(\boldsymbol{\theta})$ is non-singular, whose $(p, q)$ element is

$$
I_{p q}(\boldsymbol{\theta})=\frac{1}{8 \pi^{2}} \iint \partial_{\theta_{p}} \log f(\boldsymbol{\omega}) \partial_{\theta_{q}} \log f(\boldsymbol{\omega}) \mathrm{d} \boldsymbol{\omega}
$$

The following theorem holds true when the observation $\boldsymbol{x}$ follows the SAR model with a parameter $\boldsymbol{\theta}_{0}$ in $\Theta$.

Theorem 2.2. Under Assumption 2.1 and 2.2, the estimator $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_{0}$ in probability as $n_{1}$ and $n_{2}$ tend to infinity.

Proof. It is enough to show that

$$
\begin{equation*}
\frac{1}{N} L_{A}-\frac{1}{N} E_{\boldsymbol{\theta}_{0}} L_{A}=\frac{1}{N}\left(L_{A}-L\right)+\frac{1}{N}\left(L-E_{\boldsymbol{\theta}_{0}} L\right)+\frac{1}{N}\left(E_{\boldsymbol{\theta}_{0}} L-E_{\boldsymbol{\theta}_{0}} L_{A}\right) \tag{2.5}
\end{equation*}
$$

converges to zero in probability for any $\boldsymbol{\theta}$. The convergence to zero in probability of the first term on the right hand side of (2.5) follows from Theorem 2.1. The convergence of the second term is clear from the law of large numbers. Rewrite the third term as

$$
\begin{equation*}
\frac{1}{2 N}\left(\log \operatorname{det}\left(A^{-1}\right)-\log \operatorname{det}(\Sigma)\right)+\frac{1}{2 N} \mathrm{E}_{\boldsymbol{\theta}_{0}}\left(\boldsymbol{x}^{T}\left(\tilde{A}-\Sigma^{-1}\right) \boldsymbol{x}\right) . \tag{2.6}
\end{equation*}
$$

Then the convergence to zero of the first term on (2.6) is clear from (i) of Proposition A2 in Section 2.9.1. The second term on (2.6) is rewritten as

$$
\frac{1}{2 N} \operatorname{tr}\left(\Sigma_{0}\left(\tilde{A}-\Sigma^{-1}\right)\right)
$$

which is bounded by

$$
\left\|\Sigma_{0}\right\|\left\|\tilde{A}-\Sigma^{-1}\right\|_{1}
$$

where $\Sigma_{0}$ is the covariance matrix of the observation when $\boldsymbol{\theta}=\boldsymbol{\theta}_{0}$. Now we have the desired result from Proposition A3 in Section 2.9.1.

### 2.6 Asymptotic Efficiency

We further need the following assumption for the asymptotic efficiency of $\hat{\boldsymbol{\theta}}$.
Assumption 2.3. max $\left(n_{1} / n_{2}^{3}, n_{2} / n_{1}^{3}\right)$ goes to zero as $n_{1}$ and $n_{2}$ tend to infinity.

In order to prove Theorem 2.3, we need the following two propositions. The notation $L_{A}(\boldsymbol{\theta})$ is used in place of $L_{A}$ to clarify the dependence of $L_{A}$ on $\boldsymbol{\theta}$.

Proposition 2.1. Under Assumption 2.1, 2.2 and 2.3,

$$
\frac{\partial_{\boldsymbol{\theta}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}} \rightarrow \mathrm{~N}(\mathbf{0}, I(\boldsymbol{\theta})) \text { as } n_{1}, n_{2} \rightarrow \infty
$$

Proposition 2.2. Under Assumption 2.1, 2.2 and 2.3,

$$
-\frac{\partial_{\theta_{p} \theta_{q}}^{2} L_{A}(\boldsymbol{\theta})}{N} \rightarrow I_{p q}(\boldsymbol{\theta}) \text { in probability as } n_{1}, n_{2} \rightarrow \infty
$$

for any $p$ and $q=1,2, \ldots, m$.
The following Theorem 2.3 shows the asymptotic efficiency of $\hat{\boldsymbol{\theta}}$ when the observation $\boldsymbol{x}$ follows the SAR model with a parameter $\boldsymbol{\theta} \in \Theta$.

Theorem 2.3. Under Assumption 2.1, 2.2 and 2.3, $\hat{\boldsymbol{\theta}}$ is asymptotically efficient, that is,

$$
\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \rightarrow \mathrm{N}\left(\mathbf{0}, I(\boldsymbol{\theta})^{-1}\right) \text { as } n_{1}, n_{2} \rightarrow \infty
$$

Proof. The desired result follows from Proposition 2.1 and 2.2. Note the Taylor expansion of the equation $\partial_{\theta_{p}} L_{A}(\hat{\boldsymbol{\theta}}) / \sqrt{N}=0$,

$$
\frac{\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}}+\sum_{q=1}^{m}\left(\frac{\partial_{\theta_{p} \theta_{q}}^{2} L_{A}(\tilde{\boldsymbol{\theta}})}{N}\right) \sqrt{N}\left(\hat{\theta}_{q}-\theta_{q}\right)=0
$$

where $\tilde{\boldsymbol{\theta}}$ is a mid-value between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$.

We give the proofs of Proposition 2.1 and 2.2 at the rest of this section. The following lemmas are necessary for the proofs.

Lemma 2.1. Under Assumption 2.1, 2.2 and 2.3,

$$
\mathrm{E}\left(\frac{\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}}\right)=\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}^{3}}, \frac{n_{2}}{n_{1}^{3}}\right)^{\frac{1}{2}}\right)
$$

for any $\theta_{p}$.

Proof. The proof is divided into the two parts.
(a) When $\theta_{p}=\beta_{s}$

Note that

$$
\begin{align*}
\mathrm{E}\left(\partial_{\beta_{\boldsymbol{s}}} L_{A}(\boldsymbol{\theta})\right)= & \sum_{\boldsymbol{j} \in \mathcal{N}} \frac{\exp \left(i \boldsymbol{\omega}_{n, j}^{T} \boldsymbol{s}\right)}{\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \exp \left(i \boldsymbol{\omega}_{n, j}^{T} \boldsymbol{k}\right)} \\
& -\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|} \mathrm{E}\left\{\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x}\right\} \tag{2.7}
\end{align*}
$$

because

$$
\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x}=\boldsymbol{x}^{T}\left(W_{n_{1}}^{k_{1}-s_{1}} \otimes W_{n_{2}}^{k_{2}-s_{2}}\right) \boldsymbol{x}
$$

It is easily seen that the first term on the right hand side of (2.7) is evaluated as

$$
\frac{N}{4 \pi^{2}} \int \frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)} \mathrm{d} \boldsymbol{\omega}+\mathrm{O}(1)
$$

and the second term is evaluated as

$$
\begin{equation*}
-\frac{N}{4 \pi^{2}} \int \frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)} \mathrm{d} \boldsymbol{\omega}+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right) . \tag{2.8}
\end{equation*}
$$

In fact, by noting

$$
\begin{align*}
& \mathrm{E}\left\{\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x}\right\} \\
& =\left(n_{1}-\left|k_{1}-s_{1}\right|\right)\left(n_{2}-\left|k_{2}-s_{2}\right|\right) \gamma_{-k_{1}+s_{1},-k_{2}+s_{2}} \\
& +\left(n_{1}-\left|k_{1}-s_{1}\right|\right)\left|k_{2}-s_{2}\right| \gamma_{-k_{1}+s_{1},-n_{2}-k_{2}+s_{2}} \\
& +\left|k_{1}-s_{1}\right|\left(n_{2}-\left|k_{2}-s_{2}\right|\right) \gamma_{-n_{1}-k_{1}+s_{1},-k_{2}+s_{2}} \\
& +\left|k_{1}-s_{1}\right|\left|k_{2}-s_{2}\right| \gamma_{-n_{1}-k_{1}+s_{1},-n_{2}-k_{2}+s_{2}} \tag{2.9}
\end{align*}
$$

we have

$$
\begin{array}{r}
\alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|} \mathrm{E}\left\{\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x}\right\} \\
=N \gamma_{-\boldsymbol{k}+\boldsymbol{s}}+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right) \tag{2.10}
\end{array}
$$

since

$$
\alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|}\left(n_{1}-\left|k_{1}-s_{1}\right|\right)\left(n_{2}-\left|k_{2}-s_{2}\right|\right)=N+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right)
$$

and the other terms on the right hand side of (2.9) are negligible even if they are multiplied by $\alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|}$. We finally have (2.8) from

$$
-\frac{N}{\sigma^{2}} \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \gamma_{-\boldsymbol{k}+\boldsymbol{s}}=-\frac{N}{4 \pi^{2}} \int \frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)} \mathrm{d} \boldsymbol{\omega}
$$

(b) When $\theta_{p}=\sigma$

Note that

$$
\begin{aligned}
& \mathrm{E}\left(\partial_{\sigma} L_{A}(\boldsymbol{\theta})\right) \\
& =-\frac{N}{\sigma}+\frac{1}{\sigma^{3}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}}\left[\alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} \mathrm{E}\left\{\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right) \boldsymbol{x}\right\}\right]
\end{aligned}
$$

The second term on the right hand side is evaluated as

$$
\frac{N}{\sigma}+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right)
$$

because the evaluation

$$
\begin{array}{r}
\alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} \mathrm{E}\left(\boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right) \boldsymbol{x}\right) \\
=N \gamma_{-\boldsymbol{k}+\boldsymbol{k}^{\prime}}+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right)
\end{array}
$$

holds true by similar reason as for (2.10).

We hereafter use the notation $\lambda_{\boldsymbol{n}, \boldsymbol{j}}(M), \boldsymbol{n}=\left(n_{1}, n_{2}\right)^{T}, \boldsymbol{j}=\left(j_{1}, j_{2}\right)^{T}, j_{1}=$ $1,2, \ldots, n_{1}, j_{2}=1,2, \ldots, n_{2}$ for the eigenvalues of $N \times N$ matrix $M$.

Lemma 2.2. Under Assumption 2.1, 2.2 and 2.3,

$$
\mathrm{E}\left(\frac{\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}} \frac{\partial_{\theta_{q}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}}\right) \rightarrow I_{p q}(\boldsymbol{\theta}) \text { as } n_{1}, n_{2} \rightarrow \infty
$$

for any $p$ and $q=1,2, \ldots, m$.

Proof. We prove the proposition through the evaluation of the moment generating function of $\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta}) / \sqrt{N}$. The moment generating function is written as

$$
\begin{align*}
& M(\boldsymbol{t})=\mathrm{E}\left(\exp \left(\frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p}\left(\partial_{\theta_{p}} \log \operatorname{det}(A)-\boldsymbol{x}^{T} \partial_{\theta_{p}} \tilde{A} \boldsymbol{x}\right)\right)\right) \\
& =\exp \left(\frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p} \partial_{\theta_{p}} \log \operatorname{det}(A)\right) \operatorname{det}\left\{I+\frac{1}{\sqrt{N}} \sum_{p=1}^{m} t_{p}\left(\Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \Sigma^{\frac{1}{2}}\right)\right\}^{-\frac{1}{2}} \tag{2.11}
\end{align*}
$$

of $\boldsymbol{t}=\left(t_{1}, t_{2}, \ldots, t_{m}\right)^{T}$ since

$$
\begin{aligned}
& E\left(\exp \left(-\frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p} \boldsymbol{x}^{T} \partial_{\theta_{p}} \tilde{A} \boldsymbol{x}\right)\right) \\
& =\int e^{-\frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p} \boldsymbol{y}^{T} \Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \Sigma^{\frac{1}{2}} \boldsymbol{y}} \cdot \frac{1}{(\sqrt{2 \pi})^{N}} e^{-\frac{1}{2} \boldsymbol{y}^{T} \boldsymbol{y}} \mathrm{~d} \boldsymbol{y} \\
& =\frac{1}{(\sqrt{2 \pi})^{N}} \int e^{-\frac{1}{2} \boldsymbol{y}^{T}\left(I+\frac{1}{\sqrt{N}} \sum_{p=1}^{m} t_{p} \Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \Sigma^{\frac{1}{2}}\right) \boldsymbol{y}} \mathrm{d} \boldsymbol{y} \\
& =\operatorname{det}\left(I+\frac{1}{\sqrt{N}} \sum_{p=1}^{m} t_{p} \Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \Sigma^{\frac{1}{2}}\right)^{-\frac{1}{2}}
\end{aligned}
$$

where $\boldsymbol{y}=\Sigma^{-\frac{1}{2}} \boldsymbol{x}$.
We first see that

$$
\begin{aligned}
& \left.\partial_{t_{p} t_{q}}^{2} M(\boldsymbol{t})\right|_{\boldsymbol{t}=\mathbf{0}} \\
& =\frac{1}{2 N} \sum_{j \in \mathcal{N}} \lambda_{n, j}\left(\Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \Sigma^{\frac{1}{2}}\right) \lambda_{n, j}\left(\Sigma^{\frac{1}{2}} \partial_{\theta_{q}} \tilde{A} \Sigma^{\frac{1}{2}}\right)+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right) \\
& =\frac{1}{2 N} \operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \tilde{A} \partial_{\theta_{q}} \tilde{A}\right)+\mathrm{O}\left(\max \left(\frac{n_{1}}{n_{2}}, \frac{n_{2}}{n_{1}}\right)\right)
\end{aligned}
$$

for any $p$ and $q=1, \ldots, m$ from Lemma 2.1. From Lemma 3.1 (b) of Kent and Mardia (1996),

$$
\begin{aligned}
& \left|\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \tilde{A} \partial_{\theta_{q}} \tilde{A}\right)-\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right)\right| \\
& =\left|\operatorname{tr}\left(\Sigma^{2}\left(\partial_{\theta_{p}} \tilde{A}-\partial_{\theta_{p}} \Sigma^{-1}\right) \partial_{\theta_{q}} \tilde{A}\right)+\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1}\left(\partial_{\theta_{q}} \tilde{A}-\partial_{\theta_{q}} \Sigma^{-1}\right)\right)\right| \\
& \leq\left\|\Sigma \left|\left\|^{2}| | \partial_{\theta_{p}} \tilde{A}-\partial_{\theta_{p}} \Sigma^{-1} \mid\right\|_{1}\left\|\partial_{\theta_{q}} \tilde{A}\right\|+\|\Sigma\|^{2}\left\|\partial_{\theta_{p}} \Sigma^{-1}\right\|\left\|\partial_{\theta_{q}} \tilde{A}-\partial_{\theta_{q}} \Sigma^{-1}\right\|_{1}\right.\right.
\end{aligned}
$$

holds true. We have now

$$
\begin{equation*}
\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \tilde{A} \partial_{\theta_{q}} \tilde{A}^{T}\right)=\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right)+\mathrm{O}\left(n_{1}+n_{2}\right) \tag{2.12}
\end{equation*}
$$

from Lemma A1, A2 and Proposition A3 in Section 2.9.1. It is enough to note that

$$
I_{p q}(\boldsymbol{\theta})=\lim _{n_{1}, n_{2} \rightarrow \infty} \frac{1}{2 N} \operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right)
$$

which is proved in Proposition B1 in Section 2.9.2.

Proof of Proposition 2.1. It is enough to show the asymptotic normality of $\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta}) / \sqrt{N}$ for any $p=1,2, \ldots, m$ from Lemma 2.1 and 2.2. We prove that $\log M(\boldsymbol{t})$ is asymptotically quadratic as $N$ tends to infinity for proof of asymptotic normality. From (2.11), we have that

$$
\begin{align*}
\log M(\boldsymbol{t})= & \frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p} \partial_{\theta_{p}} \log \operatorname{det}(A)-\frac{1}{2} \sum_{j=1}^{N} \log \left(1+\frac{1}{\sqrt{N}} \sum_{p=1}^{m} t_{p} \lambda_{j}\left(\Delta_{p}\right)\right) \\
= & \frac{1}{2 \sqrt{N}} \sum_{p=1}^{m} t_{p} \partial_{\theta_{p}} \log \operatorname{det}(A)-\frac{1}{2} \sum_{j=1}^{N}\left\{\frac{1}{\sqrt{N}} \sum_{p=1}^{m} t_{p} \lambda_{j}\left(\Delta_{p}\right)\right. \\
& \left.\quad-\frac{1}{2 N}\left(\sum_{p=1}^{m} t_{p} \lambda_{j}\left(\Delta_{p}\right)\right)^{2}+\frac{1}{3 N \sqrt{N}}\left(\sum_{p=1}^{m} s_{p} \lambda_{j}\left(\Delta_{p}\right)\right)^{3}\right\} \\
= & \sum_{p=1}^{m} t_{p} E\left(\frac{\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}}\right)+\frac{1}{2} \sum_{p, q=1}^{m} t_{p} t_{q} E\left(\frac{\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}} \frac{\partial_{\theta_{q}} L_{A}(\boldsymbol{\theta})}{\sqrt{N}}\right) \\
& -\frac{1}{6 N \sqrt{N}} \sum_{j \in \mathcal{N}}\left(\sum_{p=1}^{m} s_{p} \lambda_{j}\left(\Delta_{p}\right)\right)^{3} \tag{2.13}
\end{align*}
$$

where $\Delta_{p}=\Sigma^{\frac{1}{2}} \partial_{\theta_{p}} \tilde{A} \sum^{\frac{1}{2}}$ and $s_{p}$ is a mid value between 0 and $t_{p}$ for $p=$ $1,2, \ldots, m$.

The third term of (2.13) tends to zero as $N$ tends to infinity since

$$
\begin{aligned}
\sum_{j=1}^{N} \lambda_{j}\left(\Delta_{p}\right) \lambda_{j}\left(\Delta_{q}\right) \lambda_{j}\left(\Delta_{r}\right) & =\operatorname{tr}\left(\Sigma^{3} \partial_{\theta_{p}} \tilde{A} \partial_{\theta_{q}} \tilde{A} \partial_{\theta_{r}} \tilde{A}\right) \\
& \leq N\|\Sigma\|^{3}\left\|\partial_{\theta_{p}} \tilde{A} \partial_{\theta_{q}} \tilde{A}\right\|\left\|\partial_{\theta_{r}} \tilde{A}\right\|=\mathrm{O}(N)
\end{aligned}
$$

from Lemma A1 and A2 in Section 2.9.1.

Proof of Proposition 2.2. It is enough to show that the moment generating function $M(t)$ of each $-\partial_{\theta_{p} \theta_{q}}^{2} L_{A}(\boldsymbol{\theta}) / \sqrt{N}-I_{p q}(\boldsymbol{\theta})$ converges to 1. It is equivalent to show

$$
\begin{array}{r}
\lim _{n_{1}, n_{2} \rightarrow \infty} \exp \left(-\frac{t}{2 N} \partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(A)\right) E\left(\exp \left(\frac{t}{2 N} \boldsymbol{x}^{T} \partial_{\theta_{p} \theta_{q}}^{2} \tilde{A} \boldsymbol{x}\right)\right) \\
=\exp \left(t I_{p q}(\boldsymbol{\theta})\right) \tag{2.14}
\end{array}
$$

for any $p$ and $q=1,2, \ldots, m$. By taking the logarithm of the both sides of (2.14), we see that it is equivalent to prove the convergence of

$$
\begin{equation*}
-\frac{t}{2 N} \partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(A)-\frac{1}{2} \log \operatorname{det}\left(I-\frac{t}{N} \Sigma^{\frac{1}{2}} \partial_{\theta_{p} \theta_{q}}^{2} \tilde{A} \Sigma^{\frac{1}{2}}\right) \tag{2.15}
\end{equation*}
$$

to $t I_{p q}(\boldsymbol{\theta})$ as $n_{1}$ and $n_{2}$ tend to infinity. The first term on (2.15) is evaluated as

$$
\frac{t}{2 N} \partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(\Sigma)+\mathrm{O}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)
$$

from the (iii) of Proposition A2 in Section 2.9.1, and the second term is evaluated as

$$
\begin{aligned}
& -\frac{1}{2} \sum_{j \in \mathcal{N}} \log \left(1-\frac{t}{N} \lambda_{n, j}\left(\Sigma^{\frac{1}{2}} \partial_{\theta_{p} \theta_{q}}^{2} \tilde{A} \Sigma^{\frac{1}{2}}\right)\right) \\
& =\frac{t}{2 N} \sum_{j \in \mathcal{N}} \lambda_{n, j}\left(\Sigma^{\frac{1}{2}} \partial_{\theta_{p} \theta_{q}}^{2} \tilde{A} \Sigma^{\frac{1}{2}}\right)+\mathrm{O}\left(\frac{1}{N}\right) \\
& =\frac{t}{2 N} \operatorname{tr}\left(\Sigma \partial_{\theta_{p} \theta_{q}}^{2} \tilde{A}\right)+\mathrm{O}\left(\frac{1}{N}\right) \\
& =\frac{t}{2 N} \operatorname{tr}\left(\Sigma \partial_{\theta_{p} \theta_{q}}^{2} \Sigma^{-1}\right)+\mathrm{O}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)
\end{aligned}
$$

where the last equality can be proved in a similarly way as in (2.12). Combining those result, we see that (2.15) is evaluated as

$$
\frac{t}{2 N} \operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right)+\mathrm{O}\left(\frac{1}{n_{1}}+\frac{1}{n_{2}}\right)
$$

Note here that

$$
\partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(\Sigma)=-\operatorname{tr}\left(\Sigma^{-2} \partial_{\theta_{p}} \Sigma \partial_{\theta_{q}} \Sigma\right)+\operatorname{tr}\left(\Sigma^{-1} \partial_{\theta_{p} \theta_{q}}^{2} \Sigma\right)
$$

and

$$
\partial_{\theta_{p} \theta_{q}}^{2} \Sigma^{-1}=\Sigma^{-1}\left(\partial_{\theta_{p}} \Sigma \Sigma^{-1} \partial_{\theta_{q}} \Sigma+\partial_{\theta_{q}} \Sigma \Sigma^{-1} \partial_{\theta_{p}} \Sigma-\partial_{\theta_{p} \theta_{q}}^{2} \Sigma\right) \Sigma^{-1} .
$$

We now obtain the desired result from Proposition B2 in Section 2.9.2.

### 2.7 Numerical Experiments

The aim of this section is to numerically check the behaviour of the estimator $\hat{\boldsymbol{\theta}}$ and compare it with other estimator. A naive estimation procedure would be to find out the maximum of the exact likelihood. The autocovariances which appear in the covariance matrix $\Sigma$ are not simple functions of the regression parameters but the integral functions as

$$
\gamma_{\boldsymbol{h}}=\frac{\sigma^{2}}{4 \pi^{2}} \iint \frac{z_{1}^{h_{1}} z_{2}^{h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}} \mathrm{~d} \boldsymbol{\omega}
$$

This integration is a burden in practice for the calculation of the exact likelihood, since no residue theorem is known for such a multivariable complex function. This is the reason why Whittle (1954) proposed an approximation of the likelihood on frequency domain, which is later improved by Guyon (1982). We therefore compare our estimator with the Guyon type estimator $\hat{\boldsymbol{\theta}}_{G}$ which maximises

$$
\begin{equation*}
L_{G}=-\frac{1}{4 \pi^{2}} \iint \log \left|P\left(z_{1}, z_{2}\right)\right|^{2} \mathrm{~d} \boldsymbol{\omega}+\sum_{k_{1}, k_{2}} c_{k_{1}, k_{2}} \tilde{\gamma}_{k_{1}, k_{2}}, \tag{2.16}
\end{equation*}
$$

where $c_{k_{1}, k_{2}}$ are the Fourier coefficients of $f(\boldsymbol{\omega})^{-1}=\left|P\left(z_{1}, z_{2}\right)\right|^{2} / \sigma^{2}$ and $\tilde{\gamma}_{k_{1}, k_{2}}$ are the empirical covariances,

$$
\tilde{\gamma}_{k_{1}, k_{2}}=\frac{1}{\left(n_{1}-\left|k_{1}\right|\right)\left(n_{2}-\left|k_{2}\right|\right)} \sum_{\boldsymbol{v}, \boldsymbol{v}+\boldsymbol{k} \in \mathcal{N}} X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{k}}
$$

### 2.7.1 Random Number Generation

An algorithm has been developed for generation of random numbers which follow SAR model for numerical experiments, as the direct use of the equation (2.1) as for time series autoregressive model does not work well. This is probably because $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ is not innovation of $\left\{X_{\boldsymbol{v}}\right\}$ in any sense. The following algorithm instead works well, although it is applicable only for the case when the transfer function $P\left(z_{1}, z_{2}\right)$ is decomposed as the product $P_{1}\left(z_{1}\right) P_{2}\left(z_{2}\right)$ of

$$
P_{1}\left(z_{1}\right)=\sum_{k=-p_{1}}^{p_{2}} b_{k} z_{1}^{k} \text { and } P_{2}\left(z_{2}\right)=\sum_{k=-p_{1}^{\prime}}^{p_{2}^{\prime}} b_{k}^{\prime} z_{2}^{k}
$$

We here assume that $b_{0}=b_{0}^{\prime}=1$ for the unique of the decomposition. From Assumption 2.1, $P_{1}\left(z_{1}\right)$ can be factorized as

$$
P_{1}\left(z_{1}\right)=c \prod_{j=1}^{p_{1}} \prod_{k=-p_{2}}^{-1}\left(1-\alpha_{j} z_{1}\right)\left(1-\alpha_{k} z_{1}^{-1}\right)
$$

on $\left|z_{1}\right|=1$ for a constant $c \neq 0$. The inverse function

$$
P_{1}\left(z_{1}\right)^{-1}=c^{-1} \prod_{j=1}^{p_{1}} \prod_{k=-p_{2}}^{-1}\left(1-\alpha_{j} z_{1}\right)^{-1}\left(1-\alpha_{k} z_{1}^{-1}\right)^{-1}
$$

can be expanded on $\left|z_{1}\right|=1$. In fact, if $\left|\alpha_{j}\right|<1$,

$$
\begin{equation*}
\left(1-\alpha_{j} z_{1}\right)^{-1}=\sum_{\ell=0}^{\infty}\left(\alpha_{j} z_{1}\right)^{\ell} \tag{2.17}
\end{equation*}
$$

and else if $\left|\alpha_{\boldsymbol{j}}\right|>1$,

$$
\begin{equation*}
\left(1-\alpha_{j} z_{1}\right)^{-1}=-\left(\alpha_{j} z_{1}\right)^{-1} \sum_{\ell=0}^{\infty}\left(\alpha_{j} z_{1}\right)^{-\ell} \tag{2.18}
\end{equation*}
$$

Note that there is no case where $\left|\alpha_{\boldsymbol{j}}\right|=1$ from the stationarity assumption Assumption 2.1. Similar expansion is possible for $\left(1-\alpha_{k} z_{1}^{-1}\right)^{-1}$ by replacing $z_{1}$ by $z_{1}^{-1}$ in the above. As a result, we have a random number generation algorithm from $\left\{\varepsilon_{\boldsymbol{v}}\right\}$ through the formula,

$$
x_{\boldsymbol{v}}=P_{1}\left(T_{1}\right)^{-1} P_{2}\left(T_{2}\right)^{-1} \varepsilon_{\boldsymbol{v}}
$$

where the expansions (2.17) or (2.18) are truncated into finite sums.
Figure 2.1 is an example of random number experiments. The vertical line shows the periodogram and the spectral density and the horizontal line shows the frequency $\boldsymbol{\omega}$. The periodogram
$\hat{f}(\boldsymbol{\omega})=\frac{1}{N} \sum_{k_{1}=-n_{1}+1}^{n_{1}-1} \sum_{k_{2}=-n_{2}+1}^{n_{2}-1} w\left(k_{1}\right) w\left(k_{2}\right) x_{v_{1}, v_{2}} x_{v_{1}+k_{1}, v_{2}+k_{2}} \exp \left(i\left(\omega_{1} k_{1}+\omega_{2} k_{2}\right)\right)$
is shown by the solid lines in the nine panels for $0 \leq \omega_{1}<\pi$, where the first row panels are for $\omega_{2}=\pi / 9,2 \pi / 9,3 \pi / 9$, the second row panels are for $\omega_{2}=4 \pi / 9,5 \pi / 9,6 \pi / 9$ and the third panels are for $\omega_{2}=7 \pi / 9,8 \pi / 9,9 \pi / 9$. The dotted line is the true density $f(\boldsymbol{\omega})$ for the transfer function $P\left(z_{1}, z_{2}\right)=$ $\left(1+\beta_{1,0} z_{1}+\beta_{-1,0} z_{1}^{-1}\right)\left(1+\beta_{0,1} z_{2}+\beta_{0,-1} z_{2}^{-1}\right)$ with $\beta_{1,0}=0.6, \beta_{-1,0}=0.3, \beta_{0,1}=$ $0.4, \beta_{0,-1}=0.5$ and $\sigma=0.01$. The number of the observations are $n_{1}=60$ and $n_{2}=60$, and the window used here is $w(k)=0.2-0.25 \cos \left(\pi k /\left(n_{i}-\right.\right.$ 1) $+\pi)+0.3 \cos \left(2 \pi k /\left(n_{i}-1\right)+2 \pi\right)-0.25 \cos \left(3 \pi k /\left(n_{i}-1\right)+3 \pi\right)$ for $i=1,2$.

### 2.7.2 Parameter Estimation

In view of the easiness of random number generation as is demonstrated in the previous section, we restrict our attention into the case when the transfer function is a product two single variate functions as $P\left(z_{1}, z_{2}\right)=P_{1}\left(z_{1}\right) P_{2}\left(z_{2}\right)$, particularly when $P\left(z_{1}, z_{2}\right)=\left(1+\beta_{1,0} z_{1}+\beta_{-1,0} z_{1}^{-1}\right)\left(1+\beta_{0,1} z_{2}+\beta_{0,-1} z_{2}^{-1}\right)$. In the following, the estimate $\hat{\boldsymbol{\theta}}$ or the estimate $\hat{\boldsymbol{\theta}}_{G}$ is obtained by a non-linear optimisation function nlminb in R.

For the $\hat{\boldsymbol{\theta}}$, the explicit formulas of the gradient $\partial_{\theta_{p}} L_{A}(\boldsymbol{\theta})$ and Hessian $\partial_{\theta_{p} \theta_{q}}^{2} L_{A}(\boldsymbol{\theta})$ given in Section 2.9.3 are used for the maximisation of $L_{A}(\boldsymbol{\theta})$ in (2.2) by the R function nlminb.

Since the parameter $\sigma$ is obtained from the formula

$$
\sigma^{2}=\frac{1}{N} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right) \boldsymbol{x}
$$



Figure 2.1: Estimation of spectral density from generated random numbers
as the solution of $\partial_{\sigma} L_{A}(\boldsymbol{\theta})=0$, the parameter $\sigma$ is excluded from the parameters for the maximisation of $L_{A}$.

For the $\hat{\boldsymbol{\theta}}_{G}$, the hardest part is the approximation of the first term on the right hand side of (2.16). In this experiment, we make use of the series expansion suggested in Whittle(1954),

$$
\begin{align*}
& \iint \log \left|P\left(z_{1}, z_{2}\right)\right|^{2} \mathrm{~d} \boldsymbol{\omega} \\
& =\int \log \left|1+\beta_{1,0} z_{1}+\beta_{-1,0} z_{1}^{-1}\right|^{2} \mathrm{~d} \omega_{1} \int \log \left|1+\beta_{0,1} z_{2}+\beta_{0,-1} z_{2}^{-1}\right|^{2} \mathrm{~d} \omega_{2} \\
& =\sum_{j=1}^{\infty} \frac{(2 j)!}{j!j!j}\left\{\left(\beta_{1,0} \beta_{-1,0}\right)^{j}+\left(\beta_{0,1} \beta_{0,-1}\right)^{j}\right\} \tag{2.19}
\end{align*}
$$

The summation is truncated at $j=40$ in this experiment.
Table 2.1 and Table 2.2 are the summaries of 60 times random number
experiments for the case when $n_{1}=n_{2}=30$ and $n_{1}=n_{2}=40$ respectively. Computational time for both estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{G}$ are almost the same, 0.6 second and 0.9 second respectively. The machine used is a Windows 7 Home Premium PC with a Intel Core i7-3667U 2.00 GHz dual-core processor with 8 GB of memory. However the accuracies of the estimates are not the same. The $\hat{\boldsymbol{\theta}}$ has less bias than the $\hat{\boldsymbol{\theta}}_{G}$ for $\beta_{1,0}$ and $\beta_{-1,0}$ with no notable difference in the bias for the other parameters and less standard deviation for whole parameters in Table 2.1. Almost the same observation follows from Table 2.2, that is, the $\hat{\boldsymbol{\theta}}$ has less bias than the $\hat{\boldsymbol{\theta}}_{G}$ for $\beta_{1,0}$ with no notable difference in the bias for other parameters and less standard deviation for whole parameters.

Also it is worthy of noting that the Guyon type estimation procedure may involve instability. Even for such a simple case, 32 times in the 60 experiments nlminb results in false convergence with the code 8 for both cases. The code says that the gradient may be computed incorrectly, the other stopping tolerances may be too tight, or either the function of the gradient may be discontinuous near the current iterate. Most probable reason would be that the computation of (2.19) or its gradient is not accurate enough. This suggests that there would be cases where the estimation procedure for $\hat{\boldsymbol{\theta}}_{G}$ becomes unstable for more complex SAR model.

Table 2.1: Average and standard deviation of the estimates when $n_{1}=n_{2}=$ 30.

|  |  | $\beta_{1,0}$ | $\beta_{-1,0}$ | $\beta_{0,1}$ | $\beta_{0,-1}$ | $\sigma$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0.1 | 0.8 | 0.2 | 0.7 | 0.01 |
| Average | $\hat{\boldsymbol{\theta}}$ | 0.09855 | 0.7962 | 0.2066 | 0.6926 | 0.01000 |
|  | $\hat{\boldsymbol{\theta}}_{G}$ | 0.08698 | 0.8133 | 0.2007 | 0.7015 | 0.01006 |
| Standard <br> deviation | $\hat{\boldsymbol{\theta}}$ | 0.05916 | 0.05971 | 0.05855 | 0.05726 | 0.000984 |
|  | $\hat{\boldsymbol{\theta}}_{G}$ | 0.06322 | 0.06732 | 0.06084 | 0.06143 | 0.001020 |

Table 2.2: Average and standard deviation of the estimates when $n_{1}=n_{2}=$ 40.

|  |  | $\beta_{1,0}$ | $\beta_{-1,0}$ | $\beta_{0,1}$ | $\beta_{0,-1}$ | $\sigma$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0.1 | 0.8 | 0.2 | 0.7 | 0.01 |
| Average | $\hat{\boldsymbol{\theta}}$ | 0.10037 | 0.7978 | 0.1999 | 0.6970 | 0.01005 |
|  | $\hat{\boldsymbol{\theta}}_{G}$ | 0.09560 | 0.8046 | 0.1978 | 0.7001 | 0.01007 |
| Standard <br> deviation | $\hat{\boldsymbol{\theta}}^{2}$ | 0.04614 | 0.04810 | 0.03570 | 0.03562 | 0.0006973 |
|  | $\boldsymbol{\theta}_{G}$ | 0.04753 | 0.05059 | 0.03635 | 0.03662 | 0.0007050 |

### 2.8 Remarks

We have demonstrated the goodness of the estimator which maximises the approximate $\log$-likelihood $L_{A}$. Applications of the SAR model are wide spread from astronomy to ecology. We hope that our estimation procedure will help people in such various fields of applications for analysing spatial data, together with the random number generation algorithm developed in Section 2.7.1. Also a remaining work would be to prove the theorems without Gaussianity of the underlying random field.

Optimisation algorithm used here for numerical experiments is an implementation of Newton-Raphson type algorithm with constraint nlminb in R. However, a problem of such a Newton-Raphson type algorithm is that the convergence of the algorithm becomes slow or even does not converge as an increase of the number of parameters. A possible alternative is to employ Bayesian Markov chain Monte Carlo (MCMC). The algorithm is advantageous over Newton-Raphson type algorithm because calculation of the likelihood for given parameters values is only needed, and no search path is constructed. They are already implemented as an R package, CARBayes for CAR model and HSAR for Hierarchical spatial lag model, based on the approximation of the likelihood proposed by Cressie (1993) and Banerjee et al. (2003). It is part of the Comprehensive R Archive Network (CRAN) and is freely available
at https://cran.r-project.org/web/packages/CARBayes/index.html or https://cran.r-project.org/web/packages/HSAR/index.html. Unfortunately, there is no implementation of MCMC algorithm for SAR model probably because there was no good approximation of the likelihood. We hope that an implementation of MCMC algorithm for SAR model soon appears based on our approximation proposed.

### 2.9 Appendix

### 2.9.1 Fundamental Lemmas and Propositions

Several fundamental lemmas and propositions which are used in this chapter are given in this section.

Lemma A1. (Kent and Mardia,1996, p.386) $\|\Sigma\|,\left\|\partial_{\theta_{p}} \Sigma\right\|$ and $\left\|\partial_{\theta_{p} \theta_{q}}^{2} \Sigma\right\|$ are all uniformly bounded and bounded away from zero with respect to $n_{1}$ and $n_{2}$.

Lemma A2. $\|A\|,\left\|\partial_{\theta_{p}} A\right\|$ and $\left\|\partial_{\theta_{p} \theta_{q}}^{2} A\right\|$ are all uniformly bounded and bounded away from zero with respect to $n_{1}$ and $n_{2}$.

We only show the outline of the proof of Lemma A2. It is enough to note that the eigenvalues of $A, \partial_{\theta_{p}} A$ and $\partial_{\theta_{p} \theta_{q}}^{2} A$ are written as the following finite sums,

$$
\begin{gather*}
\lambda_{n, \boldsymbol{j}}(A)=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k} \in \mathcal{K}} \sum_{\boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \exp \left(i \boldsymbol{\omega}_{\boldsymbol{n}, \boldsymbol{j}}^{T}\left(-\boldsymbol{k}+\boldsymbol{k}^{\prime}\right)\right),  \tag{2.20}\\
\lambda_{\boldsymbol{n}, \boldsymbol{j}}\left(\partial_{\theta_{p}} A\right)=\frac{2}{\sigma^{2}} \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \cos \left(\boldsymbol{\omega}_{\boldsymbol{n}, \boldsymbol{j}}^{T}(-\boldsymbol{k}+\boldsymbol{s})\right), \\
\lambda_{\boldsymbol{n}, \boldsymbol{j}}\left(\partial_{\theta_{p} \theta_{q}}^{2} A\right)=\frac{2}{\sigma^{2}} \cos \left(\boldsymbol{\omega}_{\boldsymbol{n}, \boldsymbol{j}}^{T}\left(-\boldsymbol{s}^{\prime}+\boldsymbol{s}\right)\right)
\end{gather*}
$$

when $\theta_{p}=\beta_{\boldsymbol{s}}$ and $\theta_{q}=\beta_{\boldsymbol{s}^{\prime}}$ for $\boldsymbol{\omega}_{n, j}=\left(\omega_{n_{1}, j_{1}}, \omega_{n_{2}, j_{2}}\right)^{T}$ with $\omega_{n_{1}, j_{1}}=2 \pi\left(j_{1}-\right.$ 1)/ $n_{1}$ and $\omega_{n_{2}, j_{2}}=2 \pi\left(j_{2}-1\right) / n_{2}$ (Martin, 1986). The convergence is uniform since only $\boldsymbol{\omega}_{n, j}$ depends on $n_{1}$ and $n_{2}$. Note that this lemma holds true even if $A$ is replaced by $\tilde{A}$.

Proposition A1. $\left\|A^{-1}-\Sigma\right\|_{1},\left\|\partial_{\theta_{p}} A^{-1}-\partial_{\theta_{p}} \Sigma\right\|_{1}$ and $\left\|\partial_{\theta_{p} \theta_{q}}^{2} A^{-1}-\partial_{\theta_{p} \theta_{q}}^{2} \Sigma\right\|_{1}$ are all of the order of $n_{1}+n_{2}$ for any $\theta_{p}$ and $\theta_{q}$.

Proof. Let us introduce the singular auxiliary matrix $U_{n}$ such that $\left(U_{n}\right)_{j, k}=$ $\delta_{j+1, k}$ for $j, k=1,2, \ldots, n$, and denote $U_{n}^{T}$ as $U_{n}^{-1}$ although $U_{n}^{T}$ is not the inverse matrix but a generalised inverse of $U_{n}$. We can then decompose $W_{n}^{g}$ into $W_{n}^{g}=U_{n}^{g}+\left(U_{n}^{T}\right)^{n-g}$. Since the matrix $A$ is circular, the inverse matrix is nicely written as

$$
A^{-1}=\sum_{h_{1}=-\infty}^{\infty} \sum_{h_{2}=-\infty}^{\infty} \gamma_{\boldsymbol{h}} W_{n_{1}}^{h_{1}} \otimes W_{n_{2}}^{h_{2}} .
$$

This is due to the fact that the eigenvectors of $A$ are also circular, whose elements are of the form $\exp \left(i \boldsymbol{\omega}_{n, j}\right)$, and the eigenvalues of $A^{-1}$ are given by

$$
\left\{\lambda_{n, \boldsymbol{j}}(A)\right\}^{-1}=\frac{\sigma^{2}}{\left|P\left(\exp \left(i \omega_{n_{1}, j_{1}}\right), \exp \left(i \omega_{n_{2}, j_{2}}\right)\right)\right|^{2}}=\sum_{\boldsymbol{h} \in \mathbb{Z}^{2}} \gamma_{\boldsymbol{h}} \exp \left(i \boldsymbol{\omega}_{\boldsymbol{n}, \boldsymbol{j}}\right)
$$

which is shown in (2.20) of Section 2.9.1.
From the representation,

$$
\Sigma=\sum_{h_{1}=-n_{1}+1}^{n_{1}-1} \sum_{h_{2}=-n_{2}+1}^{n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{h_{2}}
$$

we see that

$$
\begin{align*}
& A^{-1}-\Sigma \\
&= \sum_{h_{1}=1}^{2 n_{1}-1} \sum_{h_{2}}^{n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{-n_{1}+h_{1}} \otimes U_{n_{2}}^{h_{2}}+\sum_{h_{1}=-2 n_{1}+1}^{-1} \sum_{h_{2}=-n_{2}+1}^{n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{n_{1}+h_{1}} \otimes U_{n_{2}}^{h_{2}} \\
&+\sum_{h_{1}=-n_{1}+1}^{n_{1}-1} \sum_{h_{2}=1}^{2 n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{-n_{2}+h_{2}}+\sum_{h_{1}=-n_{1}+1}^{n_{1}-1} \sum_{h_{2}=-2 n_{2}+1}^{-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{n_{2}+h_{2}} \\
&+\sum_{h_{1}=1}^{2 n_{1}-1} \sum_{h_{2}=1}^{2 n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{-n_{1}+h_{1}} \otimes U_{n_{2}}^{-n_{2}+h_{2}}+\sum_{h_{2}=-2 n_{2}+1}^{2 n_{1}-1} \sum_{\boldsymbol{h}}^{-1} U_{n_{1}}^{-n_{1}+h_{1}} \otimes U_{n_{2}}^{n_{2}+h_{2}} \\
&+ \sum_{h_{1}}^{-1} \sum_{h_{2}=-2 n_{1}+1}^{2 n_{2}-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{n_{1}+h_{1}} \otimes U_{n_{2}}^{-n_{2}+h_{2}}+\sum_{h_{1}=-2 n_{1}+1}^{-1} \sum_{h_{2}=-2 n_{2}+1}^{-1} \gamma_{\boldsymbol{h}} U_{n_{1}}^{n_{1}+h_{1}} \otimes U_{n_{2}}^{n_{2}+h_{2}} . \tag{2.21}
\end{align*}
$$

We can prove that the $\|\cdot\|_{1}$ norms of the first two terms on the right hand side of $(2.21)$ are of the order $n_{2}$, the next two terms are of the order $n_{1}$ from the following Lemma A3 and Lemma A4 and the others are of the order of constant. For example, the norm of the first term is bounded by

$$
\sum_{h_{1}=1}^{2 n_{1}-1} \sum_{h_{2}=-n_{2}+1}^{n_{2}-1}\left|\gamma_{\boldsymbol{h}}\right|\left(n_{1}-\left|n_{1}-h_{1}\right|\right)\left(n_{2}-\left|h_{2}\right|\right)
$$

which is further bounded by a constant times of

$$
\begin{equation*}
\left(\sum_{h_{1}=1}^{n_{1}-1} \eta^{h_{1}} h_{1}+\sum_{h_{1}=n_{1}}^{2 n_{1}-1} \eta^{h_{1}}\left(2 n_{1}-h_{1}\right)\right)\left(\sum_{h_{2}=1}^{n_{2}-1} \eta^{h_{2}}\left(n_{2}-h_{2}\right)+\sum_{h_{2}=-n_{2}+1}^{0} \eta^{-h_{2}}\left(n_{2}+h_{2}\right)\right) \tag{2.22}
\end{equation*}
$$

for $0<\eta<1$. We see that the terms in each brackets on (2.22) are of the order $n_{1} \eta^{n_{1}}$ and $n_{2}$ respectively since

$$
\sum_{h=1}^{n} h \eta^{h}=\frac{\left\{1-(n+1) \eta^{n}+n \eta^{n+1}\right\} \eta}{(1-\eta)^{2}}=\frac{\eta}{(1-\eta)^{2}}+\mathrm{O}\left(n \eta^{n}\right)
$$

Then the first term on the right hand side of $(2.21)$ is of the order $n_{2}$. The proofs for the other terms are similar.

Lemma A3. For any $0<\eta<1$, there exists a positive finite constant $C_{1}$ such that

1. $\left|\gamma_{s}\right| \leq C_{1} \eta^{\left|s_{1}\right|+\left|s_{2}\right|}$,
2. $\left|\partial_{\theta_{p}} \gamma_{s}\right| \leq C_{1} \eta^{\left|s_{1}\right|+\left|s_{2}\right|}$,
3. $\left|\partial_{\theta_{p} \theta_{q}}^{2} \gamma_{s}\right| \leq C_{1} \eta^{\left|s_{1}\right|+\left|s_{2}\right|}$
for any $\theta_{p}$ and $\theta_{q}$.

Proof. We only give the proof of (i) since the other proofs are similar. As same as in time series, $1 / P\left(z_{1}, z_{2}\right)$ exists and is regular on a domain $\mathcal{R}=$ $\left\{\left(z_{1}, z_{2}\right) ; 1-\delta<\left|z_{1}\right|,\left|z_{2}\right|<1+\delta, \delta>0\right\}$. In the Laurent expansion

$$
\frac{1}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}}=\sum_{s}\left(\sum_{k} \alpha_{k+s} \alpha_{k}\right) z_{1}^{s_{1}} z_{2}^{s_{2}}
$$

on the domain $\mathcal{R}$, the coefficients of the expansion are bounded as

$$
\left|\sum_{k} \alpha_{k+s} \alpha_{k}\right|<C \eta^{\left|s_{1}\right|+\left|s_{2}\right|}
$$

for an $\eta=\max (1 /(1+\delta), 1-\delta)$. Note that $\left|z_{j}^{s_{j}}\right|<(1+\delta)^{s_{j}}$ if $s_{j}>0$ and $\left|z_{j}^{s_{j}}\right|<(1-\delta)^{s_{j}}$ else for $j=1,2$. Therefore we have

$$
\left|\gamma_{s}\right|=\left|\sum_{k} \sum_{k^{\prime}} \alpha_{\boldsymbol{k}} \alpha_{\boldsymbol{k}^{\prime}} E\left(\varepsilon_{\boldsymbol{v}+\boldsymbol{k}} \varepsilon_{\boldsymbol{v}+\boldsymbol{k}^{\prime}+\boldsymbol{s}}\right)\right| \leq \sigma^{2} C \eta^{\left|s_{1}\right|+\left|s_{2}\right|} .
$$

## Lemma A4.

$$
\left\|U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{h_{2}}\right\|_{1}=\left(n_{1}-\left|h_{1}\right|\right)\left(n_{2}-\left|h_{2}\right|\right)
$$

for any $0 \leq\left|h_{1}\right| \leq n_{1}-1$ and $0 \leq\left|h_{2}\right| \leq n_{2}-1$.

Proof. Note that

$$
U=\left(U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{h_{2}}\right)\left(U_{n_{1}}^{h_{1}} \otimes U_{n_{2}}^{h_{2}}\right)^{T}=U_{n_{1}}^{h_{1}}\left(U_{n_{1}}^{h_{1}}\right)^{T} \otimes U_{n_{2}}^{h_{2}}\left(U_{n_{2}}^{h_{2}}\right)^{T} .
$$

Since

$$
U_{n}^{h}\left(U_{n}^{h}\right)^{T}=\left(\begin{array}{cc}
I_{n-h} & 0 \\
0 & 0
\end{array}\right) \text { for } h>0, \quad U_{n}^{h}\left(U_{n}^{h}\right)^{T}=\left(\begin{array}{cc}
0 & 0 \\
0 & I_{n-|h|}
\end{array}\right) \quad \text { for } h<0
$$

and $U_{n}^{h}\left(U_{n}^{h}\right)^{T}=I_{n}$ for $h=0$, the matrix $U$ has only 0 or 1 eigenvalues where the number of 1's is $\left(n_{1}-\left|h_{1}\right|\right)\left(n_{2}-\left|h_{2}\right|\right)$. This leads us the desired result.

We need the following propositions for the proofs of Theorem 2.1, 2.2 and 2.3.

## Proposition A2.

(i) $\log \operatorname{det}\left(A^{-1}\right)-\log \operatorname{det}(\Sigma)=\mathrm{O}\left(n_{1}+n_{2}\right)$,
(ii) $\partial_{\theta_{p}} \log \operatorname{det}\left(A^{-1}\right)-\partial_{\theta_{p}} \log \operatorname{det}(\Sigma)=\mathrm{O}\left(n_{1}+n_{2}\right)$,
(iii) $\partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}\left(A^{-1}\right)-\partial_{\theta_{p} \theta_{q}}^{2} \log \operatorname{det}(\Sigma)=\mathrm{O}\left(n_{1}+n_{2}\right)$
for any $\theta_{p}$ and $\theta_{q}$.

Proof of (i). By noting that

$$
\begin{aligned}
\log \operatorname{det}\left(A^{-1}\right)-\log \operatorname{det}(\Sigma) & =\log \operatorname{det}\left(\Sigma^{-1} A^{-1}\right) \\
& =\log \operatorname{det}\left\{I+\Sigma^{-1}\left(A^{-1}-\Sigma\right)\right\} \\
& =\log \prod_{j_{1}=1}^{n_{1}} \prod_{j_{2}=1}^{n_{2}}\left(1+\lambda_{n, j}\left(\Sigma^{-1}\left(A^{-1}-\Sigma\right)\right)\right) \\
& \leq \sum_{j_{1}=1}^{n_{1}} \sum_{j_{2}=1}^{n_{2}} \lambda_{n, j}\left(\Sigma^{-1}\left(A^{-1}-\Sigma\right)\right) \\
& \leq\left\|\Sigma^{-1}\right\|\left\|A^{-1}-\Sigma\right\|_{1}
\end{aligned}
$$

the desired result follows from Lemma A1 and Proposition A1. Reversing the roles of $\Sigma$ and $A$ leads us to the lower bound of the same order.

Proof of (ii). It is enough to show that

$$
\left|\operatorname{tr}\left(A \partial_{\theta_{p}} A^{-1}\right)-\operatorname{tr}\left(\Sigma^{-1} \partial_{\theta_{p}} \Sigma\right)\right|
$$

is of the order $n_{1}+n_{2}$, which can be proved in a same way as in the proof of (2.12).

Proof of (iii). It is enough to show

$$
\left|\operatorname{tr}\left(\partial_{\theta_{p}} A^{-1} \partial_{\theta_{q}} A\right)-\operatorname{tr}\left(\partial_{\theta_{p}} \Sigma \partial_{\theta_{q}} \Sigma^{-1}\right)\right|+\left|\operatorname{tr}\left(A \partial_{\theta_{p} \theta_{q}}^{2} A^{-1}\right)-\operatorname{tr}\left(\Sigma^{-1} \partial_{\theta_{p} \theta_{q}}^{2} \Sigma\right)\right|,
$$

is the order of $n_{1}+n_{2}$. The proofs are similar to the proof of (2.12).
Proposition A3. $\left\|\tilde{A}-\Sigma^{-1}\right\|_{1},\left\|\partial_{\theta_{p}} \tilde{A}-\partial_{\theta_{p}} \Sigma^{-1}\right\|_{1}$ and $\left\|\partial_{\theta_{p} \theta_{q}}^{2} \tilde{A}-\partial_{\theta_{p} \theta_{q}}^{2} \Sigma^{-1}\right\|_{1}$ are all of the order of $n_{1}+n_{2}$ for any $\theta_{p}$ and $\theta_{q}$.

Proof. We only give the proof for $\left\|\tilde{A}-\Sigma^{-1}\right\|_{1}$. Because of the boundedness of $\|\tilde{A}\|$ and $\left\|\Sigma^{-1}\right\|$, we have

$$
\begin{aligned}
\left\|\tilde{A}-\Sigma^{-1}\right\|_{1} & \leq\|\tilde{A}\|\left\|\Sigma^{-1}\right\|\left\|\Sigma-\tilde{A}^{-1}\right\|_{1} \\
& \leq C\left(\left\|\Sigma-A^{-1}\right\|_{1}+\left\|A^{-1}-\tilde{A}^{-1}\right\|_{1}\right)
\end{aligned}
$$

for a constant $C>0$. The fact that $\left\|\Sigma-A^{-1}\right\|_{1}=\mathrm{O}\left(n_{1}+n_{2}\right)$ is already proved in Proposition A1 and the fact that $\left\|A^{-1}-\tilde{A}^{-1}\right\|_{1}=\mathrm{O}\left(n_{1}+n_{2}\right)$ is proved in the following lemma.

Lemma A5. $\left\|A^{-1}-\tilde{A}^{-1}\right\|_{1},\left\|\partial_{\theta_{p}} A^{-1}-\partial_{\theta_{p}} \tilde{A}^{-1}\right\|_{1}$ and $\left\|\partial_{\theta_{p} \theta_{q}}^{2} A^{-1}-\partial_{\theta_{p} \theta_{q}}^{2} \tilde{A}^{-1}\right\|_{1}$ are all of the order of $n_{1}+n_{2}$ for any $\theta_{p}$ and $\theta_{q}$.

Proof. We first note that

$$
1-\alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|}=\mathrm{O}\left(\max \left(\frac{1}{n_{1}}, \frac{1}{n_{2}}\right)\right)
$$

and

$$
\left\|W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right\|_{1}=\mathrm{O}(N) .
$$

Because of the boundedness of $\left\|A^{-1}\right\|$ and $\left\|\tilde{A}^{-1}\right\|$,

$$
\begin{aligned}
& \left\|A^{-1}-\tilde{A}^{-1}\right\|_{1} \leq\left\|A^{-1}\right\|\left\|\tilde{A}^{-1}\right\|\|A-\tilde{A}\|_{1} \\
& \quad \leq C\left|\sum_{\boldsymbol{k} \in \mathcal{K}} \sum_{\boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}}\left(1-\alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|}\right)\right|\left\|W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right\|_{1}
\end{aligned}
$$

follows for a constant $C>0$ and the desired result follows.
Similarly, we can prove $\left\|\partial_{\theta_{p}} A^{-1}-\partial_{\theta_{p}} \tilde{A}^{-1}\right\|_{1}$ is of the order of $n_{1}+n_{2}$. It is enough to note that

$$
\left\|\partial_{\boldsymbol{\beta}_{s}} A-\partial_{\boldsymbol{\beta}_{s^{\prime}}} \tilde{A}\right\|_{1}=\left|\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}}\left(1-\alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|}\right)\right|\left\|W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right\|_{1}
$$

because $\left\|W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right\|_{1}=\left\|W_{n_{1}}^{k_{1}-s_{1}} \otimes W_{n_{2}}^{k_{2}-s_{2}}\right\|_{1}$, and

$$
\left\|\partial_{\sigma} A-\partial_{\sigma} \tilde{A}\right\|_{1}=\frac{2}{\sigma^{3}}\|A-\tilde{A}\|_{1} .
$$

In order to prove that $\left\|\partial_{\theta_{p} \theta_{q}}^{2} A^{-1}-\partial_{\theta_{p} \theta_{q}}^{2} \tilde{A}^{-1}\right\|$ is of the order $n_{1}+n_{2}$, it is enough to note that

$$
\begin{gathered}
\left\|\partial_{\beta_{s} \beta_{s^{\prime}}}^{2} A-\partial_{\beta_{s} \beta_{s^{\prime}}}^{2} \tilde{A}\right\|_{1}=\frac{2}{\sigma^{2}}\left|1-\alpha_{n_{1}}^{\left|s_{1}-s_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|s_{2}-s_{2}^{\prime}\right|}\right|\left\|W_{n_{1}}^{-s_{1}+s_{1}^{\prime}} \otimes W_{n_{2}}^{-s_{2}+s_{2}^{\prime}}\right\|_{1}, \\
\left\|\partial_{\beta_{s} \sigma}^{2} A-\partial_{\beta_{s} \sigma}^{2} \tilde{A}\right\|_{1}=\frac{2}{\sigma}\left\|\partial_{\beta_{s}} A-\partial_{\beta_{s}} \tilde{A}\right\|_{1},
\end{gathered}
$$

and

$$
\left\|\partial_{\sigma \sigma}^{2} A-\partial_{\sigma \sigma}^{2} \tilde{A}\right\|_{1}=\frac{6}{\sigma^{2}}\|A-\tilde{A}\|_{1} .
$$

### 2.9.2 Lemma and Proposition Used for the Proof of Theorem 2.3

Lemma B1. Under Assumption 2.1 and 2.2,

$$
\begin{equation*}
I_{p q}(\boldsymbol{\theta})=\frac{1}{2 \sigma^{4}} \sum_{\boldsymbol{h} \in \mathbb{Z}^{2}}\left(\sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \partial_{\theta_{p}} \gamma_{-\boldsymbol{k}+\boldsymbol{k}^{\prime}-\boldsymbol{h}}\right)\left(\sum_{\boldsymbol{\ell}, \ell^{\prime} \in \mathcal{K}} \beta_{\ell} \beta_{\ell^{\prime}} \partial_{\theta_{q}} \gamma_{-\ell+\ell^{\prime}+\boldsymbol{h}}\right), \tag{2.23}
\end{equation*}
$$

for any $p$ and $q=1,2, \ldots, m$.

Proof. We first note that

$$
\partial_{\beta_{\boldsymbol{s}}} \log f(\boldsymbol{\omega})=-2 \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right),
$$

$$
\partial_{\sigma} \log f(\boldsymbol{\omega})=\frac{2}{\sigma},
$$

$$
\partial_{\beta_{s}} \gamma_{\boldsymbol{h}}=-\frac{\sigma^{2}}{2 \pi^{2}} \int \frac{z_{1}^{h_{1}} z_{2}^{h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{4}} \operatorname{Re}\left(z_{1}^{-s_{1}} z_{2}^{-s_{2}} P\left(z_{1}, z_{2}\right)\right) \mathrm{d} \boldsymbol{\omega}
$$

and

$$
\partial_{\sigma} \gamma_{\boldsymbol{h}}=\frac{\sigma}{2 \pi^{2}} \int \frac{z_{1}^{h_{1}} z_{2}^{h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}} \mathrm{~d} \boldsymbol{\omega} .
$$

(a) When $\theta_{p}=\beta_{\boldsymbol{s}}$ and $\theta_{q}=\beta_{\boldsymbol{s}^{\prime}}$

$$
\begin{aligned}
& \frac{1}{2 \sigma^{4}} \sum_{\boldsymbol{h} \in \mathbb{Z}^{2}} \partial_{\beta_{\boldsymbol{s}}} \gamma_{\boldsymbol{h}}\left(\sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime}, \ell, \ell^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \beta_{\boldsymbol{\ell}} \beta_{\boldsymbol{\ell}^{\prime}} \partial_{\beta_{\boldsymbol{s}^{\prime}}} \gamma_{-\boldsymbol{k}+\boldsymbol{k}^{\prime}-\boldsymbol{\ell}+\boldsymbol{\ell}^{\prime}-\boldsymbol{h}}\right) \\
& =-\frac{1}{4 \pi^{2} \sigma^{2}} \sum_{\boldsymbol{h} \in \mathbb{Z}^{2}} \partial_{\beta_{\boldsymbol{s}}} \gamma_{\boldsymbol{h}} \\
& \quad \times \int \frac{\sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} z_{1}^{-k_{1}+k_{1}^{\prime}} z_{2}^{-k_{2}+k_{2}^{\prime}} \sum_{\ell, \ell^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{\ell}} \beta_{\boldsymbol{\ell}^{\prime}} z_{1}^{-\ell_{1}+\ell_{1}^{\prime}} z_{2}^{-\ell_{2}+\ell_{2}^{\prime}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{4}} \\
& \times z_{1}^{-h_{1}} z_{2}^{-h_{2}} \operatorname{Re}\left(z_{1}^{-s_{1}^{\prime}} z_{2}^{-s_{2}^{\prime}} P\left(z_{1}, z_{2}\right)\right) \mathrm{d} \boldsymbol{\omega} \\
& = \\
& =\frac{1}{2 \sigma^{2}} \frac{\sum_{\boldsymbol{h} \in \mathbb{Z}^{2}}\left(\partial_{\beta_{\boldsymbol{s}}} \gamma_{\boldsymbol{h}}\right)\left(\beta_{\boldsymbol{s}^{\prime}+\boldsymbol{h}}+\beta_{\boldsymbol{s}^{\prime}-\boldsymbol{h}}\right)}{4 \pi^{2}} \int \frac{\sum_{\boldsymbol{h} \in \mathbb{Z}^{2}}\left(\beta_{\boldsymbol{s}^{\prime}+\boldsymbol{h}}+\beta_{\boldsymbol{s}^{\prime}-\boldsymbol{h}}\right) z_{1}^{h_{1}} z_{2}^{h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}} \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right) \mathrm{d} \boldsymbol{\omega} \\
& =\frac{1}{2 \pi^{2}} \int \operatorname{Re}\left(\frac{z_{1}^{s_{1}}}{P\left(z_{1}, z_{2}\right)}\right) \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right) \mathrm{d} \boldsymbol{\omega}=I_{p q}(\boldsymbol{\theta}) .
\end{aligned}
$$

(b) When $\theta_{p}=\beta_{s}$ and $\theta_{q}=\sigma$

$$
\begin{aligned}
& -\frac{1}{8 \pi^{4} \sigma} \sum_{\boldsymbol{h} \in \mathbb{Z}^{2}} \int \frac{\sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} z_{1}^{-k_{1}+k_{1}^{\prime}-h_{1}} z_{2}^{-k_{2}+k_{2}^{\prime}-h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}} \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right) \mathrm{d} \boldsymbol{\omega} \\
& \times \int \frac{\sum_{\ell, \ell^{\prime} \in \mathcal{K}} \beta_{\ell} \beta_{\ell^{\prime}} z_{1}^{-\ell_{1}+\ell_{1}^{\prime}-h_{1}} z_{2}^{-\ell_{2}+\ell_{2}^{\prime}-h_{2}}}{\left|P\left(z_{1}, z_{2}\right)\right|^{2}} \mathrm{~d} \boldsymbol{\omega} \\
& =-\frac{1}{2 \pi^{2} \sigma} \int \operatorname{Re}\left(\frac{z_{1}^{s_{1}} z_{2}^{s_{2}}}{P\left(z_{1}, z_{2}\right)}\right) \mathrm{d} \boldsymbol{\omega}=I_{p q}(\boldsymbol{\theta}) .
\end{aligned}
$$

(c) When $\theta_{p}=\sigma$ and $\theta_{q}=\sigma$

Since

$$
\sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \partial_{\sigma} \gamma_{-\boldsymbol{k}+\boldsymbol{k}^{\prime}-\boldsymbol{h}}=2 \sigma \text { and } \sum_{\ell, \ell^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{\ell}} \beta_{\ell^{\prime}} \partial_{\sigma} \gamma_{-\boldsymbol{\ell}+\ell^{\prime}+\boldsymbol{h}}=2 \sigma
$$

the right hand side of $(2.23)$ is written as $2 / \sigma^{2}$.

Proposition B1. Under Assumption 2.1 and 2.2,

$$
I_{p q}(\boldsymbol{\theta})=\lim _{n_{1}, n_{2} \rightarrow \infty} \frac{1}{2 N} \operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right)
$$

for any $p$ and $q=1,2, \ldots, m$.

Proof. We can show

$$
\begin{aligned}
\operatorname{tr}\left(\Sigma^{2} \partial_{\theta_{p}} \Sigma^{-1} \partial_{\theta_{q}} \Sigma^{-1}\right) & =\operatorname{tr}\left(\Sigma^{-2} \partial_{\theta_{p}} \Sigma \partial_{\theta_{q}} \Sigma\right) \\
& =\operatorname{tr}\left(A^{2} \partial_{\theta_{p}} A^{-1} \partial_{\theta_{q}} A^{-1}\right)+\mathrm{O}\left(n_{1}+n_{2}\right)
\end{aligned}
$$

as similarly as in (2.12). Then

$$
\operatorname{tr}\left(A^{2} \partial_{\theta_{p}} A^{-1} \partial_{\theta_{q}} A^{-1}\right)=I_{p q}(\boldsymbol{\theta})
$$

follows from the fact that

$$
\operatorname{tr}\left(W_{n_{1}}^{\ell_{1}} \otimes W_{n_{2}}^{\ell_{2}}\right)=\left\{\begin{array}{cc}
N \quad \text { if } \ell_{1}=\ell_{2}=0 \\
0 & \text { otherwise }
\end{array}\right.
$$

### 2.9.3 Representation of Gradient and Hessian

The gradient and the Hessian of $L_{A}$ are represented as follows.

$$
\begin{aligned}
\partial_{\beta_{s}} L_{A}= & \sum_{\boldsymbol{j} \in \mathcal{N}} \frac{\exp \left(i \boldsymbol{\omega}_{n, j}^{T} \boldsymbol{s}\right)}{\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \exp \left(i \boldsymbol{\omega}_{n, j}^{T} \boldsymbol{k}\right)} \\
& -\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \alpha_{n_{1}}^{\left|k_{1}-s_{1}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x}, \\
\partial_{\sigma} L_{A}= & -\frac{N}{\sigma}+\frac{1}{\sigma^{3}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right) \boldsymbol{x}, \\
\partial_{\beta_{s} \beta_{s^{\prime}}}^{2} L_{A}= & -\sum_{\boldsymbol{j} \in \mathcal{N}} \frac{\exp \left(i \boldsymbol{\omega}_{n, j}^{T}\left(\boldsymbol{s}+\boldsymbol{s}^{\prime}\right)\right)}{\left\{\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \exp \left(i \boldsymbol{\omega}_{n, j}^{T} \boldsymbol{k}\right)\right\}^{2}} \\
& -\frac{1}{\sigma^{2}} \alpha_{n_{1}}^{\left|s_{1}-s_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|s_{2}-s_{2}^{\prime}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-s_{1}+s_{1}^{\prime}} \otimes W_{n_{2}}^{-s_{2}+s_{2}^{\prime}}\right) \boldsymbol{x}, \\
\partial_{\beta_{s} \sigma}^{2} L_{A}= & \frac{2}{\sigma^{3}} \sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} \alpha_{n_{1}}^{\left|k_{1}-s_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-s_{2}^{\prime}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+s_{1}} \otimes W_{n_{2}}^{-k_{2}+s_{2}}\right) \boldsymbol{x},
\end{aligned}
$$

and

$$
\partial_{\sigma \sigma}^{2} L_{A}=\frac{N}{\sigma^{2}}-\frac{3}{\sigma^{4}} \sum_{k, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \alpha_{n_{2}}^{\left|k_{2}-k_{2}^{\prime}\right|} \boldsymbol{x}^{T}\left(W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes W_{n_{2}}^{-k_{2}+k_{2}^{\prime}}\right) \boldsymbol{x} .
$$

## Chapter 3

## Non-identifiability of SAR Model

SAR 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.

### 3.1 Introduction

An SAR model for a random field $\left\{X_{v} ; \boldsymbol{v}=\left(v_{1}, v_{2}, \ldots, v_{d}\right)^{T} \in \mathbb{Z}^{d}\right\}$ with the mean zero and the autocovariance function $\gamma_{\boldsymbol{h}}=E\left(X_{\boldsymbol{v}} X_{\boldsymbol{v}+\boldsymbol{h}}\right), \boldsymbol{h} \in \mathbb{Z}^{d}$ is the model which satisfies the equation

$$
\begin{equation*}
P\left(T_{1}, T_{2}, \ldots, T_{d}\right) X_{v}=\varepsilon_{\boldsymbol{v}}, \quad \boldsymbol{v} \in \mathbb{Z}^{d} \tag{3.1}
\end{equation*}
$$

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

$$
P\left(T_{1}, T_{2}, \ldots, T_{d}\right)=\sum_{k \in \mathcal{K}} \beta_{\boldsymbol{k}} T_{1}^{k_{1}} T_{2}^{k_{2}} \cdots T_{d}^{k_{d}}
$$

is an $n$-dimensional transfer function with the real coefficients $\beta_{\boldsymbol{k}}, \boldsymbol{k} \in \mathcal{K}$ where $\mathcal{K}$ is a set of finite points $\boldsymbol{k}=\left(k_{1}, k_{2}, \ldots, k_{d}\right)$ on $\mathbb{Z}^{d}$, which always includes $\mathbf{0}$, 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 $T_{j}, j=1, \ldots, d$ are shift operators such as

$$
T_{j} X_{\boldsymbol{v}}=X_{v_{1}, \ldots, v_{j}+1, \ldots, v_{d}}
$$

We assume the following for the weakly stationary of the simultaneous spatial autoregressive model throughout this chapter.

Assumption 3.1. $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\sum_{\boldsymbol{k} \in \mathcal{K}} \beta_{\boldsymbol{k}} z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}$ has no zeros on the domain $D=\left\{\left(z_{1}, z_{2}, \ldots, z_{d}\right) ;\left|z_{1}\right|=\left|z_{2}\right|=\cdots=\left|z_{d}\right|=1\right\}$ in $\mathbb{C}^{d}$.

The spectral density of the SAR model (3.1) is then

$$
f(\boldsymbol{\omega})=\frac{\sigma^{2}}{\left|P\left(z_{1}, z_{2}, \ldots, z_{d}\right)\right|^{2}}
$$

where $z_{j}=\exp \left(i \omega_{j}\right), j=1, \ldots, n$ and $\boldsymbol{\omega}=\left(\omega_{1}, \omega_{2}, \ldots, \omega_{d}\right)$.

### 3.2 Non-identifiability of SAR Model

We first note that any polynomial $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ is decomposable into a product of prime factors $h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right), k=1,2, \ldots, p$ as

$$
P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\prod_{k=1}^{p} h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right)
$$

Therefore, there exist $2^{p}$ choices in selecting $h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ or $\overline{h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right)}$ for $k=1,2, \ldots, p$ to have a transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ which leads us to the spectral density

$$
\begin{equation*}
f(\boldsymbol{\omega})=\frac{\sigma^{2}}{\prod_{k=1}^{p} h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right) \overline{h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right)}} \tag{3.2}
\end{equation*}
$$

There is also freedom to add a factor of the form $c z_{1}^{\ell_{1}} z_{2}^{\ell_{2}} \cdots z_{d}^{\ell_{d}}$ to the transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ for any constant $c$ and integers $\ell_{1}, \ell_{2}, \ldots, \ell_{d}$, since the constant $c$ can be absorbed into the parameter $\sigma^{2}$.

Example 3.1. Let us consider a simple one-dimensional SAR model,

$$
\begin{equation*}
X_{v}+\beta_{1} X_{v+1}+\beta_{-1} X_{v-1}=\varepsilon_{v} \tag{3.3}
\end{equation*}
$$

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

$$
\begin{equation*}
f(\omega)=\frac{\sigma_{0}^{2}}{\left(z-\alpha_{1}\right)\left(z-\alpha_{2}\right)\left(z^{-1}-\bar{\alpha}_{1}\right)\left(z^{-1}-\bar{\alpha}_{2}\right)}, \tag{3.4}
\end{equation*}
$$

where $z=\mathrm{e}^{i \omega}$ 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 (3.4).

$$
\begin{aligned}
& P_{1}(z)=-\frac{1}{\alpha_{1}+\alpha_{2}} z^{-1}\left(z-\alpha_{1}\right)\left(z-\alpha_{2}\right)=1-\frac{1}{\alpha_{1}+\alpha_{2}} z-\frac{\alpha_{1} \alpha_{2}}{\alpha_{1}+\alpha_{2}} z^{-1} \\
& P_{2}(z)=-\frac{1}{\bar{\alpha}_{1}+\bar{\alpha}_{2}} z\left(z^{-1}-\bar{\alpha}_{1}\right)\left(z^{-1}-\bar{\alpha}_{2}\right)=1-\frac{\bar{\alpha}_{1} \bar{\alpha}_{2}}{\bar{\alpha}_{1}+\bar{\alpha}_{2}} z-\frac{1}{\bar{\alpha}_{1}+\bar{\alpha}_{2}} z^{-1} \\
& P_{3}(z)=\frac{1}{1+\alpha_{1} \bar{\alpha}_{2}}\left(z-\alpha_{1}\right)\left(z^{-1}-\bar{\alpha}_{2}\right)=1-\frac{\bar{\alpha}_{2}}{1+\alpha_{1} \bar{\alpha}_{2}} z-\frac{\alpha_{1}}{1+\alpha_{1} \bar{\alpha}_{2}} z^{-1}
\end{aligned}
$$

and

$$
P_{4}(z)=\frac{1}{1+\bar{\alpha}_{1} \alpha_{2}}\left(z^{-1}-\bar{\alpha}_{1}\right)\left(z-\alpha_{2}\right)=1-\frac{\bar{\alpha}_{1}}{1+\bar{\alpha}_{1} \alpha_{2}} z-\frac{\alpha_{2}}{1+\bar{\alpha}_{1} \alpha_{2}} z^{-1}
$$

It is easy to show that each transfer function has real coefficients, providing us an SAR model (3.3) with different coefficients. The variance parameter $\sigma^{2}$ varies from transfer function to transfer function, $\sigma^{2}=\sigma_{0}^{2} /\left|\alpha_{1}+\alpha_{2}\right|^{2}$ for $P_{1}(z)$ and $P_{2}(z)$, and $\sigma^{2}=\sigma_{0}^{2} /\left|1+\alpha_{1} \bar{\alpha}_{2}\right|^{2}$ 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 $\alpha_{1}$ and $\alpha_{2}$ are real and $\alpha_{1} \alpha_{2}=1$, and the $P_{3}(z)$ and $P_{4}(z)$ become identical if and only if $\alpha_{1}$ and $\alpha_{2}$
are real and $\alpha_{1}=\alpha_{2}$. By noting Assumption 3.1, we see that such conditions are summarized as $\beta_{1}=\beta_{-1}$ with $\beta_{1}^{2}<1 / 4$, that is, time reversible SAR model. However, it does not mean unique transfer function for the spectral density of time reversible SAR model. The conditions $\alpha_{1}=\alpha_{2}$ and $\alpha_{1} \alpha_{2}=1$ are not compatible because of Assumption 3.1. Only two of the four transfer functions become identical and two others are not time reversible. We now see that there is no unique SAR model for the given spectral density (3.4) .

### 3.3 Maximum Likelihood Estimate

It is well known that the exact likelihood of SAR 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 and Shibata (2016) is stronger and more straightforward, and closed on space domain. They also proved that the parameter estimator which maximises the approximation $L_{A}$ in the following is asymptotically efficient.

Let us assume that the observations $\left\{x_{\boldsymbol{v}}, \boldsymbol{v} \in \mathcal{N}\right\}$ are on a rectangular lattice $\mathcal{N}=\left\{\boldsymbol{v}=\left(v_{1}, v_{2}, \ldots, v_{d}\right)^{T} ; 1 \leq v_{j} \leq n_{j}, j=1,2, \ldots, d\right\}$. The $N=$ $n_{1} n_{2} \cdots n_{d}$ observations are arranged to make a vector $\boldsymbol{x}$ in lexicographic order. By combining the $m-1$ dimensional regression parameter vector $\boldsymbol{\beta}$ whose elements are arranged in lexicographic order of $\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}$ with $\sigma$, we have the whole parameter vector $\boldsymbol{\theta}$. An approximation of the log-likelihood of $\boldsymbol{\theta}$ proposed by Rikimaru and Shibata (2016) is then

$$
L_{A}=\frac{1}{2} \log \operatorname{det}(A)-\frac{N}{2} \log 2 \pi-\frac{1}{2} \boldsymbol{x}^{T} \tilde{A} \boldsymbol{x}
$$

where

$$
A=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes \cdots \otimes W_{n_{d}}^{-k_{d}+k_{d}^{\prime}}
$$

and

$$
\tilde{A}=\frac{1}{\sigma^{2}} \sum_{\boldsymbol{k}, \boldsymbol{k}^{\prime} \in \mathcal{K}} \beta_{\boldsymbol{k}} \beta_{\boldsymbol{k}^{\prime}} \alpha_{n_{1}}^{\left|k_{1}-k_{1}^{\prime}\right|} \cdots \alpha_{n_{d}}^{\left|k_{d}-k_{d}^{\prime}\right|} W_{n_{1}}^{-k_{1}+k_{1}^{\prime}} \otimes \cdots \otimes W_{n_{d}}^{-k_{d}+k_{d}^{\prime}} .
$$

Here the symbol $\otimes$ is Kronecker product and $\alpha_{n_{j}}=1+1 / n_{j}, j=1,2, \ldots, d$ are shrinkage factors to retain $\sqrt{N}$ consistency. The asymptotic efficiency proved that the covariance matrix of the parameter estimator converges to the lower bound given by the inverse of the Fisher information matrix $I(\boldsymbol{\theta})$, whose elements are given by

$$
\begin{equation*}
I_{p q}(\boldsymbol{\theta})=\frac{1}{2 \cdot(2 \pi)^{d}} \int \cdots \int \frac{\partial \log f(\boldsymbol{\omega} ; \boldsymbol{\theta})}{\partial \theta_{p}} \frac{\partial \log f(\boldsymbol{\omega} ; \boldsymbol{\theta})}{\partial \theta_{q}} \mathrm{~d} \boldsymbol{\omega}, p, q=1,2, \ldots, m \tag{3.5}
\end{equation*}
$$

(Whittle, 1954; Guyon, 1982; Robinson and Vidal Sanz, 2006), provided that $I(\boldsymbol{\theta})$ 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 the case of SAR model. Before investigating when and why it happens, we will see other problems caused by non-identifiability of SAR model in maximum likelihood estimation by the following example.

Example 3.2. Consider the same SAR model (3.3) as in Example 3.1. Assume that $\left\{X_{\mathbf{v}}\right\}$ has the spectral density (3.4) 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 <br> function | $\beta_{1}$ | $\beta_{-1}$ | $\sigma$ |
| :--- | :--- | :--- | :--- |
| $P_{1}(z)$ | 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 $\left\{X_{v}\right\}$ by using the transfer function $P_{1}(z)$ with

$$
\beta_{1}=-\frac{1}{\alpha_{1}+\alpha_{2}}=0.1 \quad, \quad \beta_{-1}=-\frac{\alpha_{1} \alpha_{2}}{\alpha_{1}+\alpha_{2}}=0.77775
$$

and

$$
\sigma=-\frac{\sigma_{0}}{\alpha_{1}+\alpha_{2}}=0.01
$$

Then, the following four maximum likelihood estimate 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 estimator 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 autocovariances. 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.

### 3.4 Singularity of the Fisher Information Matrix $I(\boldsymbol{\theta})$

We have seen that several different parameters, $\boldsymbol{\theta}_{1}, \boldsymbol{\theta}_{2}, \ldots, \boldsymbol{\theta}_{2^{p}}$ are mapped from a given SAR spectral density. The problem of SAR 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 $I(\boldsymbol{\theta})$ in (3.5), which is also the limit of

$$
\begin{equation*}
-\frac{1}{N} \frac{\partial^{2} \log f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{T}} \tag{3.6}
\end{equation*}
$$

Let $\boldsymbol{\theta}_{j}(f), j=1,2, \ldots, r$ be all possible mappings of a spectral density $f$ to the parameter $\boldsymbol{\theta}_{j}(f)$. We call a parameter $\boldsymbol{\theta}$ non-singular if the Fisher infomation matrix $I(\boldsymbol{\theta})$ is non-singular.

Definition Parameter $\boldsymbol{\theta}=\boldsymbol{\theta}_{j}(f)$ has multiplicity if there exists a $k \neq j$ and $\boldsymbol{\delta} \neq \mathbf{0}$ such that $\boldsymbol{\theta}_{j}\left(f_{n}\right)-\boldsymbol{\theta}_{k}\left(f_{n}\right)=a_{n} \boldsymbol{\delta}$ for a nonzero sequence $\left\{a_{n}\right\}$ which converges to zero and a sequence of spectral densities $\left\{f_{n}\right\}$ which converges to $f$, where $\boldsymbol{\theta}_{j}\left(f_{n}\right)$ and $\boldsymbol{\theta}_{k}\left(f_{n}\right)$ are all non-singular parameters.

Theorem 3.1. Fisher information matrix $I\left(\boldsymbol{\theta}_{0}\right)$ becomes singular if $\boldsymbol{\theta}_{0}$ has multiplicity.

Proof. If $\boldsymbol{\theta}_{0}=\boldsymbol{\theta}_{0}(f)$ has multiplicity then we can find a sequence $f_{N}$ of spectral density which converges to $f$. Then there exists a small enough $\epsilon>0$ and a sequence $\left\{n_{N}\right\}$ such that $\boldsymbol{\theta}_{1}\left(f_{n_{N}}\right)-\boldsymbol{\theta}_{2}\left(f_{n_{N}}\right)=N^{-\frac{1}{2}+\epsilon}(1+\mathrm{o}(1)) \boldsymbol{\delta}$ for a nonzero $\boldsymbol{\delta}$ from the multiplicity of $\boldsymbol{\theta}_{0}$. We have then two corresponding maximum likelihood estimators $\hat{\boldsymbol{\theta}}_{1}$ and $\hat{\boldsymbol{\theta}}_{2}$, which are zeros of the gradients of the log-likelihood

$$
g_{k}(\boldsymbol{\theta})=\frac{\partial \log f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_{k}}, k=1,2, \ldots, m
$$

By Taylor expansion, we see that

$$
\begin{equation*}
0=\left(\hat{\boldsymbol{\theta}}_{1}-\hat{\boldsymbol{\theta}}_{2}\right)^{T} \frac{\partial}{\partial \boldsymbol{\theta}} g_{k}\left(\boldsymbol{\theta}_{k}^{*}\right), k=1,2, \ldots, m \tag{3.7}
\end{equation*}
$$

holds true for $\boldsymbol{\theta}_{k}^{*}$ between $\hat{\boldsymbol{\theta}}_{1}$ and $\hat{\boldsymbol{\theta}}_{2}$. Rewrite (3.7) as
$0=N^{\frac{1}{2}-\epsilon}\left(\left(\hat{\boldsymbol{\theta}}_{1}-\boldsymbol{\theta}_{1}\right)-\left(\hat{\boldsymbol{\theta}}_{2}-\boldsymbol{\theta}_{2}\right)+\left(\boldsymbol{\theta}_{1}-\boldsymbol{\theta}_{2}\right)\right)^{T} \frac{1}{N} \frac{\partial}{\partial \boldsymbol{\theta}} g_{k}\left(\boldsymbol{\theta}_{k}^{*}\right), k=1,2, \ldots, m$.
Then, by setting $\boldsymbol{\theta}_{1}=\boldsymbol{\theta}_{0}+N^{-\frac{1}{2}+\epsilon} \boldsymbol{\delta}$ and $\boldsymbol{\theta}_{2}=\boldsymbol{\theta}_{0}-N^{-\frac{1}{2}+\epsilon} \boldsymbol{\delta}$, we have

$$
\mathbf{0}=\boldsymbol{\delta}^{T} I\left(\boldsymbol{\theta}_{0}\right)
$$

since $\boldsymbol{\theta}_{1}$ and $\boldsymbol{\theta}_{2}$ are non-singular parameters.
Example 3.3 Let $\boldsymbol{\theta}_{1}=\left(\beta_{1}, \beta_{-1}, \sigma\right)^{T}$ and $\boldsymbol{\theta}_{2}=\left(\beta_{-1}, \beta_{1}, \sigma\right)^{T}$ be parameter vector for $P_{1}(z)$ and $P_{2}(z)$ in Example 3.1. Suppose that there exists $\boldsymbol{\theta}_{1}$ and $\boldsymbol{\theta}_{2}$ in the neighbourhood of $\boldsymbol{\theta}_{0}$. When we set $\boldsymbol{\theta}_{1}=\boldsymbol{\theta}_{0}+a_{N} \boldsymbol{\delta}$ and $\boldsymbol{\theta}_{2}=\boldsymbol{\theta}_{0}-a_{N} \boldsymbol{\delta}$ where $a_{N}=N^{-\frac{1}{2}+\epsilon}$ and $\boldsymbol{\delta}=(\delta,-\delta, 0)^{T}, \boldsymbol{\theta}_{1}$ and $\boldsymbol{\theta}_{2}$ go to $\boldsymbol{\theta}_{0}$ respectively as $N$ tends infinity.

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 (3.6) is likely to be singular if it happens.

Example 3.4. Let us consider the same SAR model as in Example 3.1. 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

$$
I(\boldsymbol{\theta})=\left(\begin{array}{lll}
a & a & b \\
a & a & b \\
b & b & c
\end{array}\right)
$$

where

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

From the maximum likelihood equation,
$\mathbf{0}=\left.\frac{1}{\sqrt{N}} \frac{\partial \log f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \approx \frac{1}{\sqrt{N}} \frac{\partial \log f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}+\frac{1}{N} \frac{\partial^{2} \log f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{T}} \cdot \sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$, we see that

$$
\left(\begin{array}{l}
a \sqrt{N}\left\{\left(\hat{\beta}_{1}+\hat{\beta}_{-1}\right)-\left(\beta_{1}+\beta_{-1}\right)\right\}+b \sqrt{N}(\hat{\sigma}-\sigma) \\
a \sqrt{N}\left\{\left(\hat{\beta}_{1}+\hat{\beta}_{-1}\right)-\left(\beta_{1}+\beta_{-1}\right)\right\}+b \sqrt{N}(\hat{\sigma}-\sigma) \\
b \sqrt{N}\left\{\left(\hat{\beta}_{1}+\hat{\beta}_{-1}\right)-\left(\beta_{1}+\beta_{-1}\right)\right\}+c \sqrt{N}(\hat{\sigma}-\sigma)
\end{array}\right)
$$

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

### 3.4.1 Conditions for Non-Singularity of $I(\boldsymbol{\theta})$

As is seen from Example 3.4, singularity of the Fisher information matrix $I(\boldsymbol{\theta})$ 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 $I(\boldsymbol{\theta})$ 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 $I(\boldsymbol{\theta})$,

$$
\begin{equation*}
Q=\boldsymbol{y}^{T} I(\boldsymbol{\theta}) \boldsymbol{y}=\frac{1}{2 \cdot(2 \pi)^{d}} \int \cdots \int\left(\sum_{j=1}^{m} y_{j} \frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_{j}}\right)^{2} \mathrm{~d} \boldsymbol{\omega} . \tag{3.8}
\end{equation*}
$$

Clearly a necessary and sufficient condition for the non-singularity is that the vector $\boldsymbol{y}=\left(y_{1}, y_{2}, \ldots, y_{m}\right)$ is zero whenever $Q=0$. Thus, we have the following theorem.

Theorem 3.2. A necessary and sufficient condition for $I(\boldsymbol{\theta})$ to be non-singular is

$$
\frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_{j}}, j=1,2, \ldots, m \text { are linearly independent. }
$$

Corollary 3.1. A sufficient condition for non-singularity of $I(\boldsymbol{\theta})$ is that

$$
\frac{\partial \gamma_{k}}{\partial \theta_{j}}, j=1,2, \ldots, m \text { are linearly independent for a } \boldsymbol{k} .
$$

Proof. We see that

$$
\begin{aligned}
Q=\sum_{p, q=1}^{m} y_{p} y_{q} I_{p q}(\boldsymbol{\theta}) & =\lim _{n_{1}, \ldots, n_{d} \rightarrow \infty} \frac{1}{2 N} \sum_{p, q=1}^{m} y_{p} y_{q} \operatorname{tr}\left(\Sigma^{2} \frac{\partial \Sigma^{-1}}{\partial \theta_{p}} \frac{\partial \Sigma^{-1}}{\partial \theta_{q}}\right) \\
& =\lim _{n_{1}, \ldots, n_{d} \rightarrow \infty} \frac{1}{2 N} \operatorname{tr}\left\{\left(\Sigma^{-1} \sum_{j=1}^{m} y_{j} \frac{\partial \Sigma}{\partial \theta_{j}}\right)^{2}\right\}
\end{aligned}
$$

from Proposition B1 in Chapter 2 by noting that $\partial \Sigma^{-1} / \partial \theta_{p}=$ $\Sigma^{-1}\left(\partial \Sigma / \partial \theta_{p}\right) \Sigma^{-1}$. Since the eigenvalues of $\Sigma^{-1}$ are bounded away from 0 , we have

$$
\operatorname{tr}\left\{\left(\Sigma^{-1} \sum_{j=1}^{m} y_{j} \frac{\partial \Sigma}{\partial \theta_{j}}\right)^{2}\right\} \geq C \operatorname{tr}\left\{\left(\sum_{j=1}^{m} y_{j} \frac{\partial \Sigma}{\partial \theta_{j}}\right)^{2}\right\}
$$

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

Corollary 3.2. If $\beta_{\boldsymbol{k}}=\beta_{-\boldsymbol{k}}$ for any $\boldsymbol{k} \in \mathcal{K}$, then $I(\boldsymbol{\theta})$ is singular.

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

$$
P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=P\left(z_{1}^{-1}, z_{2}^{-1}, \ldots, z_{d}^{-1}\right) .
$$

It is enough to note that

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

holds true since

$$
\frac{\partial \log f(\boldsymbol{\omega})}{\partial \beta_{\boldsymbol{k}}}=-\frac{z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}}{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}-\frac{z_{1}^{-k_{1}} z_{2}^{-k_{2}} \cdots z_{d}^{-k_{d}}}{P\left(z_{1}^{-1}, z_{2}^{-1} \ldots, z_{d}^{-1}\right)}
$$

on the domain $D$.
Example 3.5. Consider a 2-dimensional SAR model,

$$
X_{v_{1}, v_{2}}+\beta_{1,0} X_{v_{1}+1, v_{2}}+\beta_{-1,0} X_{v_{1}-1, v_{2}}+\beta_{0,1} X_{v_{1}, v_{2}+1}+\beta_{0,-1} X_{v_{1}, v_{2}-1}=\varepsilon_{v_{1}, v_{2}}
$$

then the spectral density is

$$
f\left(\omega_{1}, \omega_{2}\right)=\frac{\sigma^{2}}{\left|1+\beta_{1,0} z_{1}+\beta_{-1,0} z_{1}^{-1}+\beta_{0,1} z_{2}+\beta_{0,-1} z_{2}^{-1}\right|^{2}} .
$$

There exists only two transfer functions $P\left(z_{1}, z_{2}\right)=1+\beta_{1,0} z_{1}+\beta_{-1,0} z_{1}^{-1}+$ $\beta_{0,1} z_{2}+\beta_{0,-1} z_{2}^{-1}$ and $P\left(z_{1}^{-1}, z_{2}^{-1}\right)$ for this spectral density. This is because $P\left(z_{1}, z_{2}\right)$ is prime polynomial. It is clear that $P\left(z_{1}, z_{2}\right)$ and $P\left(z_{1}^{-1}, z_{2}^{-1}\right)$ are identical if and only if $\beta_{1,0}=\beta_{-1,0}$ and $\beta_{0,1}=\beta_{0,-1}$. Then $I(\boldsymbol{\theta})$ is singular in this case from Theorem 3.1 as well as from Corollary 3.2.

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

$$
B=\left(\begin{array}{llll}
\boldsymbol{\beta}_{\ell_{1}} & \boldsymbol{\beta}_{\ell_{2}} & \cdots & \boldsymbol{\beta}_{\ell_{L}}
\end{array}\right)
$$

where

$$
\boldsymbol{\beta}_{\ell_{j}}=\left(\begin{array}{c}
\beta_{\boldsymbol{k}_{1}+\ell_{j}}+\beta_{\boldsymbol{k}_{1}-\ell_{j}} \\
\beta_{\boldsymbol{k}_{2}+\ell_{j}}+\beta_{\boldsymbol{k}_{2}-\ell_{j}} \\
\vdots \\
\beta_{\boldsymbol{k}_{m}+\ell_{j}}+\beta_{\boldsymbol{k}_{m}-\ell_{j}}
\end{array}\right), j=1, \ldots, L
$$

Here $\boldsymbol{\ell}_{j}, j=1,2, \ldots, L$ and $\boldsymbol{k}_{j}, j=1, \ldots, m$ are ordered indices in $\mathcal{L} \cup\{\mathbf{0}\}$ and $\mathcal{K}$ respectively, where $\mathcal{L}=\{\boldsymbol{\ell}>\mathbf{0} \mid \boldsymbol{k}-\boldsymbol{\ell}$ or $\boldsymbol{k}+\boldsymbol{\ell} \in \mathcal{K}$ for a $\boldsymbol{k} \in \mathcal{K}\}$ with the half-space order $>$ defined in Chapter 1 and the order of indices is lexicographic one.

Theorem 3.3. A necessary and sufficient condition for the non-singularity of $I(\boldsymbol{\theta})$ 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 $I(\boldsymbol{\theta})$, since the last row and column off-diagonal elements are all 0 and the diagonal element is $(2 / \sigma)^{2}$. By introducing a polynomial $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\sum_{\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}} y_{\boldsymbol{k}} z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}$, the $Q$ in (3.8) when $y_{m}=0$ can be rewritten as

$$
Q=\frac{2}{(2 \pi)^{d}} \int\left\{\operatorname{Re}\left(\frac{Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)}{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}\right)\right\}^{2} \mathrm{~d} \boldsymbol{\omega} .
$$

Thus, $Q=0$ implies

$$
Y\left(z_{1}, z_{2}, \ldots, z_{d}\right) \overline{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}+\overline{Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)} P\left(z_{1}, z_{2}, \ldots, z_{d}\right)=0
$$

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

$$
\sum_{\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}} y_{\boldsymbol{k}}\left(\beta_{\boldsymbol{k}-\boldsymbol{\ell}}+\beta_{\boldsymbol{k}+\ell}\right)=0 \text { for any } \boldsymbol{\ell} \in \mathcal{L} \cup\{\mathbf{0}\}
$$

implies $y_{\boldsymbol{k}}$ for $\boldsymbol{k} \neq \mathbf{0} \in \mathcal{K}$. This completes the proof.

Example 3.6. The matrix $B$ for the SAR model in Example 3.1 is derived from

$$
\begin{gathered}
\boldsymbol{\beta}_{0}=\left(\begin{array}{c}
\beta_{-1+0}+\beta_{-1-0} \\
\beta_{0+0}+\beta_{0-0} \\
\beta_{1+0}+\beta_{1-0}
\end{array}\right)=2\left(\begin{array}{c}
\beta_{-1} \\
1 \\
\beta_{1}
\end{array}\right), \\
\boldsymbol{\beta}_{1}=\left(\begin{array}{c}
\beta_{-1+1}+\beta_{-1-1} \\
\beta_{0+1}+\beta_{0-1} \\
\beta_{1+1}+\beta_{1-1}
\end{array}\right)=\left(\begin{array}{c}
1 \\
\beta_{1}+\beta_{-1} \\
1
\end{array}\right), \\
\text { and } \boldsymbol{\beta}_{2}=\left(\begin{array}{c}
\beta_{-1+2}+\beta_{-1-2} \\
\beta_{0+2}+\beta_{0-2} \\
\beta_{1+2}+\beta_{1-2}
\end{array}\right)=\left(\begin{array}{c}
\beta_{1} \\
0 \\
\beta_{-1}
\end{array}\right)
\end{gathered}
$$

as

$$
B=\left(\begin{array}{ccc}
2 \beta_{-1} & 1 & \beta_{1} \\
2 & \beta_{1}+\beta_{-1} & 0 \\
2 \beta_{1} & 1 & \beta_{-1}
\end{array}\right)
$$

The determinant

$$
\operatorname{det}(B)=2\left(\beta_{1}-\beta_{-1}\right)\left(1+\beta_{1}+\beta_{-1}\right)\left(1-\beta_{1}-\beta_{-1}\right)
$$

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 3.1. 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.

### 3.4.2 Unilateral SAR Model

It is taken it for granted that unilateral SAR model including AR model in time series,

$$
P\left(T_{1}, T_{2}, \cdots, T_{d}\right)=\sum_{\boldsymbol{k} \in \mathcal{\mathcal { K } _ { q } \cup \{ 0 \}}} \beta_{\boldsymbol{k}} T_{1}^{k_{1}} T_{2}^{k_{2}} \cdots T_{d}^{k_{d}}
$$

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

Theorem 3.4. Unilateral SAR model is always identifiable and the Fisher information matrix $I(\boldsymbol{\theta})$ is always non-singular.

Proof. It is only possible to choose $h_{k}\left(z_{1}, z_{2}, \ldots, z_{d}\right), k=1,2, \ldots, p$ to find out transfer function $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ for the spectral density (3.2). Any other choice contradicts with the unilaterality of the model. Therefore unilateral SAR model is always unique for given spectral density. On the other hand
the quadratic form (3.8) is then rewritten as

$$
\begin{aligned}
Q= & \frac{2}{(2 \pi)^{d}} \int\left\{\operatorname{Re}\left(\frac{Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)}{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}\right)-\frac{y_{m}}{\sigma}\right\}^{2} \mathrm{~d} \boldsymbol{\omega} \\
= & \frac{1}{2(2 \pi)^{d}} \int\left[\left\{\frac{Y\left(z_{1}, \ldots, z_{d}\right)}{P\left(z_{1}, \ldots, z_{d}\right)}\right\}^{2}+2\left|\frac{Y\left(z_{1}, \ldots, z_{d}\right)}{P\left(z_{1}, \ldots, z_{d}\right)}\right|^{2}+\left\{\frac{Y\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}{P\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}\right\}^{2}\right. \\
& \left.\quad-\frac{2 y_{m}}{\sigma}\left\{\frac{Y\left(z_{1}, \ldots, z_{d}\right)}{P\left(z_{1}, \ldots, z_{d}\right)}+\frac{Y\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}{P\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}\right\}+\frac{4 y_{m}^{2}}{\sigma^{2}}\right] \mathrm{d} \boldsymbol{\omega},
\end{aligned}
$$

where $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\sum_{\boldsymbol{k} \in \mathcal{K}_{q}} y_{\mathbf{k}} z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}$ is a polynomial of $z_{1} z_{2} \cdots z_{d}$. Since $P\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ and $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)$ are both analytic non zero function on the domain $D$, we have

$$
\int\left\{\frac{Y\left(z_{1}, \ldots, z_{d}\right)}{P\left(z_{1}, \ldots, z_{d}\right)}\right\}^{k} \mathrm{~d} \boldsymbol{\omega}=-\int_{D}\left\{\frac{Y\left(z_{1}, \ldots, z_{d}\right)}{P\left(z_{1}, \ldots, z_{d}\right)}\right\}^{k} \frac{1}{z_{1} \cdots z_{d}} \mathrm{~d} z_{1} \cdots \mathrm{~d} z_{d}=0
$$

and
$\int\left\{\frac{Y\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}{P\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}\right\}^{k} \mathrm{~d} \boldsymbol{\omega}=-\int_{D}\left\{\frac{Y\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}{P\left(z_{1}^{-1}, \ldots, z_{d}^{-1}\right)}\right\}^{k} \frac{1}{z_{1} \cdots z_{d}} \mathrm{~d} z_{1} \cdots \mathrm{~d} z_{d}=0$
for any integer $k$. Therefore

$$
Q=\frac{2}{(2 \pi)^{d}} \int\left\{2 \frac{\left|Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)\right|^{2}}{\left|P\left(z_{1}, z_{2}, \ldots, z_{d}\right)\right|^{2}}+\frac{y_{m}^{2}}{\sigma^{2}}\right\} \mathrm{d} \boldsymbol{\omega}
$$

is zero if and only if $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=0$ and $y_{m}=0$. This proves the non-singularity of $I(\boldsymbol{\theta})$.

### 3.4.3 Symmetric SAR Model

The following theorem shows that symmetric SAR model,

$$
P\left(T_{1}, T_{2}, \ldots, T_{d}\right)=P\left(T_{1}^{-1}, T_{2}^{-1}, \ldots, T_{d}^{-1}\right)
$$

is always identifiable and the Fisher information matrix $I(\boldsymbol{\theta})$ is non-singular.

Theorem 3.5. Symmetric SAR model is always identifiable and the Fisher information matrix $I(\boldsymbol{\theta})$ is always non-singular.

Proof. Since the transfer function is real valued, $P\left(z_{1}, z_{2}, \cdots, z_{d}\right)=$ $\overline{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}$, the spectral density is written as

$$
f(\boldsymbol{\omega})=\frac{\sigma^{2}}{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)^{2}}
$$

Therefore, the transfer function is uniquely determined from the spectral density $f(\boldsymbol{\omega})$ as $P\left(e^{i \omega_{1}}, e^{i \omega_{2}}, \ldots, e^{i \omega_{d}}\right)=\sigma f(\boldsymbol{\omega})^{-1 / 2}$, when the negative sign is not available because the constant term of $P\left(e^{i \omega_{1}}, e^{i \omega_{2}}, \ldots, e^{i \omega_{d}}\right)$ is 1 but the constant term of Laurent expansion of $f(\boldsymbol{\omega})^{-1 / 2}$ is given by $\int f(\boldsymbol{\omega})^{-1 / 2} \mathrm{~d} \boldsymbol{\omega}$.

On the other hand, the quadratic form (3.8) is written as

$$
Q=\frac{2}{(2 \pi)^{d}} \int\left\{\frac{Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)}{P\left(z_{1}, z_{2}, \ldots, z_{d}\right)}-\frac{y_{m}}{\sigma}\right\}^{2} \mathrm{~d} \boldsymbol{\omega}
$$

where $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\sum_{\boldsymbol{k} \in \mathcal{K}_{-}} y_{\mathbf{k}}\left(z_{1}^{k_{1}} z_{2}^{k_{2}} \cdots z_{d}^{k_{d}}+z_{1}^{-k_{1}} z_{2}^{-k_{2}} \cdots z_{d}^{-k_{d}}\right)$. Therefore, $Q=0$ is equivalent to

$$
Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=\frac{y_{m}}{\sigma} P\left(z_{1}, z_{2}, \ldots, z_{d}\right)
$$

but the left hand side has no constant term but the right hand side has the constant term $y_{m} / \sigma$. This implies that $y_{m}=0$ if $Q=0$ so that $Y\left(z_{1}, z_{2}, \ldots, z_{d}\right)=0$. This proves the non-singularity of $I(\boldsymbol{\theta})$.

### 3.5 Remarks

We have shown that SAR model can be non-identifiable from the covariance structure or the spectral density. This is because several different regression parameters with different standard deviation of the disturbance are mapped from a spectral density. Therefore, we have to carefully estimate 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. Except for cases when non-identifiability causes no problem, for example, in case of prediction of unknown value of $X_{\boldsymbol{v}}$, 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 SAR model has been already mentioned by Whittle (1954) and also in the context of two sided moving average model by Rosenblatt (1980). A cure would be to employ bispectrum, which is also applicable for SAR model. But we leave it for future investigation, together with an investigation of possible mappings from a spectral density to parameters. Other strategy would be to put some constraints on parameter space like symmetry of transfer function as described in Section 3.4.3.

Another problem we have investigated in this chapter is possible singularity of the Fisher information matrix, where not all parameters are estimable. Theorem 3.1 demonstrates that it happens when some of parameters mapped from a spectral density are duplicated. Non-identifiability of SAR 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 non-convergence of optimisation algorithm or instability of the estimate. Several types of conditions given in Section 3.4 would be useful for the check. There are a lot of open problems left, for example, converse of Theorem 3.1 or any other type of necessary and sufficient condition for the non-singularity than that given in Theorem 3.3.

## Chapter 4

## Concluding Remarks

We have derived a space domain approximation of the Gaussian likelihood of SAR model. An advantage of this approximation is on the simplicity. The computation only requires simple matrix operations, not only for values of the likelihood but also for the derivatives up to the order 2 . The idea behind our approximation comes from the approximation of the likelihood of time series autoregressive model, where the observations are transformed into error variables which are independent and identically distributed. However, the transformation matrix is not only triangular but much more complicated in case of SAR model because of high dimensionality of the space. We have introduced a circular matrix approximation to the transformation matrix to solve this problem. It is then written in a simple form of circulant matrices $W_{n}$ 's whose elements are 0 or 1 by using Kronecker products.

The space domain approximation is not only advantageous to the approximation of the Gaussian likelihood but also to the proof of the consistency and asymptotic efficiency of the estimator. The simplicity of the approximation enabled us to do an advanced evaluation of the approximation error. The proof of the asymptotic efficiency is quite fragile and would not be possible without such an explicit formula of the approximation because of correlations spread over high dimensional space. An example is on the introduction of shrinkage factor $\alpha_{n}$ in $\tilde{A}$ to decrease the edge effect of
observations on a lattice. The shrinkage factor $\alpha_{n}=1+1 / n$ could not be found without such a careful evaluation of the expectation of the quadratic form of observations. No other form of the shrinkage factor $\alpha_{n}=1+c / n$ is possible without losing the asymptotic efficiency.

Random number experiments are conducted to compare the estimation procedure numerically with other procedure such as Guyon's approximation. The result shows that the estimation procedure is less bias and variance even when the observation size is relatively small like as $30 \times 30$ or $40 \times 40$, although the experiment is not exhaustive. One of reasons why we did not do any exhaustive experiment is that we had to develop random number generation algorithm from scratch in case of SAR model. We could find no article which suggests us a good random number generation algorithm. We tried every intuitive generation algorithm, but none of them was successful. It is probably because random number generation as a sequence following the regression formula does not fit well to the generation of random numbers on a space. Since initial values needed for the generation is are not a set of several values but a set of values on the edge of the lattice, the effect of initial values does not decay even after large number of random number generations. The random number generation algorithm introduced in Chapter 2 is based on an explicit representation of target variable by error variable, that is , $M A(\infty)$ representation. As is shown by estimated spectral density with the true spectral density, this algorithm worked well. It could be generalised to any transfer functions if multivariable Laurant expansion were employed, although a product of two one dimensional transfer functions case was focused there.

The non-identifiability of SAR model in Chapter 3 has been investigated when looking for the reason why singularity of Fisher information matrix arises, where parameters become non-estimable. The non-identifiability implies several different models for a given autocovariances or spectral density so that the Fisher information becomes singular if some of the models which
share the same spectral density are identical or duplicated.
What we have learned from this investigation is that casual use of SAR model may cause serious problems since the model can be non-identifiable or some of parameters are not estimable, although local optimality of the estimator has been established in Chapter 2 as far as parameters for which the Fisher information matrix is non-singular. A remedy would be to introduce some restriction on parameter space. A natural restriction may come from the meaning of the model for underlying phenomena. Otherwise, introduction of unilateral or symmetric SAR model is one of options. There would be many other options. We leave it for future investigation. The necessary and sufficient conditions for the singularity given in Chapter 3 are also helpful to check the reliability of estimated values.

There are many works have to be done. Generalisation of the results in Chapter 2 for $d$-dimensional case is one of interesting projects. Robustness of the estimation is also interesting when the underlying process is not Gaussian. Extension for the case when multivariate values would be straightforward but worthy of investigation. Application of the results in this thesis to the case of CAR model is most challenging because we expect that some of problems may disappear and some new problems appear although SAR models and CAR models look similar.

## Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Ritei Shibata, for his advice and continuous encouragement. He has given me careful and persistent advice for almost nine years since the research seminar held during spring break in my junior year in Keio University. I learned a lot from him, not only Mathematics and Statistics but also various principles which researchers should keep in mind. He is a real Data Scientist. I am extremely fortunate to have studied under his supervision. I am deeply grateful to Professor Mihoko Minami for her tremendous support for writing this thesis. Without her persistent support this thesis would not have been completed. I would like to thank Professor Hiroshi Shiraishi, Professor Kei Kobayashi and Professor Kenichi Hayashi for careful reading of this thesis and for giving me a number of useful suggestions. Special thanks go to Professor Masaaki Sibuya, Professor Kunio Shimizu and Professor Tomonari Sei for giving me valuable comments and suggestions on my research. They made my research life very productive. Thanks also go to the members of Data Science laboratory. I am especially grateful to Dr. Yuki Sugaya, Mr. Takayuki Okamoto, Mr. Akihiko Kumasaka and Mr. Mugen Setoguchi. My heartfelt appreciation goes to Professor Yu Hayakawa and Professor Hiroyuki Taniai at Waseda University, Professor Daisuke Yokouchi at Hitotsubashi University and the staff of Gakushuin Girl's Junior and Senior High School,
who gave me a lot of opportunities for learning various things while we were working together. I would like to thank Mrs. Haruko Shibata for her moral support. Finally, I wish to express my greatest appreciation to my husband, daughter, dogs, mother who passed away in the first year of my doctoral program and my father who has always watched over me.

To my beloved $K$ छ $K$
Yuuki Rikimaru

## Bibliography

[1] Banerjee, S., Carlin, B. and Gelfand, A. (2015). Hierarchical Modeling and Analysis for Spatial Dara, 2nd edn, Chapman \& Hall/CRC.
[2] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, Jouranl of the Royal Statistical Society. Series B (Mehodological) 36(2): 192-236.
[3] Besag, J. (1977). Efficiency of psuedo likelihood estimation for simple gaussian fields, Biometrika 64: 616-618.
[4] Besag, J. and Moran, P. (1975). On the estimation and testing of spatial interaction in gaussian lattice processes, Biometrika 62: 555-562.
[5] Cliff, A. and Ord, J. (1973). Spatial Autocorrelation, Pion.
[6] Cliff, A. and Ord, J. (1975). Model building and the analysis of spatial pattern in human geography, Journal of Royal Statistical Society. Series B (Methodological) 37(3): 297-348.
[7] Cliff, A. and Ord, J. (1981). Spatial Processes: Models and Applications., Pion.
[8] Cressie, N. (1993). Statistics for spatial data, revised edn, Wiley.
[9] Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields, Biometrika 74: 877-882.
[10] Diggle, P. (1983). Statistical Analysis of Spatial Point Patterns, Academic Press.
[11] Dimitiou-Fakalou, C. (2009). Modified gaussian likelihood estimators for arma models on $\mathbb{Z}^{d}$, Stochastic Processes and their Applications 119: 4149-4175.
[12] Fisher, R. (1935). The Design of Experiments, Oliver and Boyd.
[13] Gaetan, C. and Guyon, X. (2010). Spaitial Statistics and Modeling, Springer.
[14] Getis, A. and Boots, B. (1978). Models of Spatial Processes, Cambridge University Press.
[15] Grenander, U. (1981). Abstract inference, John Wiley \& Sons.
[16] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice, Biometrika 69: 95-105.
[17] Haining, R. (1979). Statistical tests and process generators for random field models, Geographical Analysis 11(1): 45-64.
[18] Halley, E. (1686). An historical account of the trade winds, and monsoons, observable in the seas between and near the tropicks, with an attempt to assign the phisical cause of the said winds, Philosophical Transactions 16: 153-168.
[19] Kent, J. and Mardia, K. (1996). Spectral and circulant approximations to the likelihood for stationary gaussian random fields, Journal of Statistical Planning and Inference 50: 379-394.
[20] Lesage, J. and Pace, R. (2009). Introduction to Spatial Economics, Champman \& Hall/CRC.
[21] Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71(1): 135-146.
[22] Marquiss, M., Newton, I. and Radcliffe, D. (1978). The decline of the raven, corvus corax, in relation to afforestation in southern scotland and northern england, Journal of Applied Ecology 15: 129-144.
[23] Martin, R. (1986). A note on the asymptotic eigenvalues and eigenvectors of the dispersion matrix of a second-order stationary process on a d-dimensional lattice, Journal of Applied Probability 23: 529-535.
[24] Ord, J. (1975). Estimation methods for spatial interaction, Journal of the American Statistical Association 70: 120-126.
[25] Pace, R. and Barry, R. (1997a). Sparce spatial autoregressions, Statistics \& Probability Letters 33: 291-297.
[26] Pace, R. and Barry, R. (1997b). Fast spatial estimation, Applied Economics Letters 4(5): 337-341.
[27] Pielou, E. (1959). The use of point-to-plant distances in the pattern of plant populations, Jounral of Ecology 47: 607-613.
[28] Rikimaru, Y. and Shibata, R. (2016). A good approximation of the gaussian likeihood of simultaneous autoregressive model which yields us an asymptotically efficient estimate of parameters, Journal of Statistical Planning and Inference 173: 31-46.
[29] Rikimaru, Y. and Shibata, R. (2017). Non-identifiability of simultaneous spatial autoregressive model and singulariy of fisher information matrix, International Journal of Statistics and Probability 6(4): 31-38.
[30] Ripley, B. (1981). Spatial Statistics, Wiley.
[31] Robinson, P. and Vidal Sanz, J. (2006). Modified whittle estimation of multilateral models on a lattice, Journal of Multivariate Analysis 97: 1090-1120.
[32] Rosenblatt, M. (1980). Linear processes and bispectra, Journal of Applied Probability 17: 265-270.
[33] Shabat, B. (1992). Introduction to Complex Analysis Part II. Functions of Several Variables, American Mathematical Society.
[34] Tiefelsdorf, M., Griffith, D. and Boots, B. (1999). A variance stabilizing scheme for spatial link matrices, Environment and planning A 31: 165-180.
[35] Tjøstheim, D. (1978). Statistical spatial series modeling, Advances in Applied Probability 10: 130-154.
[36] Tobler, W. (1970). A computer movie simulating urban growth in the detroit region, Economic Geography 46: 234-240.
[37] Upton, G. and Fingleton, B. (1985). Spatial Data Analysis by Example, Volume 1:Point Pattern and Quantitative Data, Wiley.
[38] Whittle, P. (1954). On stationary processes in the plane, Biometrika 41: 434-449.

