BOOTSTRAP METHODS TO EVALUATE THE EFFICIENCY OF THE ESTIMATORS OF THE SPATIAL UNILATERAL AR(1,1) MODEL

LATIFAH RAHAYU SIREGAR

UNIVERSITI SAINS MALAYSIA

2011

BOOTSTRAP METHODS TO EVALUATE THE EFFICIENCY OF THE ESTIMATORS OF THE SPATIAL UNILATERAL AR(1,1) MODEL

by

LATIFAH RAHAYU SIREGAR

Thesis submitted in fulfillment of the requirements for the degree of Master of Science

December 2011

ACKNOWLEDGEMENTS

Alhamdulillah, all praises to Allah SWT for the grace and the guidance until I have completed the research and finalize this thesis. Many people have helped and give support in the making process of this thesis. Firstly, I would like to give appreciation to my supervisor Dr. Norhashidah Awang, for her knowledge, guidance and time to help me from the beginning until the completion of this thesis.

Special thanks to Pemerintah Aceh who has provided the scholarship during my study in Universiti Sains Malaysia.

My thanks also to all the members of School of Mathematical Sciences, Universiti Sains Malaysia for their assistance during my study in USM.

To Dr. Hizir Sofyan, thank for your support. To all my friends especially Khairul Umam, Bintra Mailina, to my sister Sheilatina, I would like to thank them for their supports. For all those whose names are not mentioned here, thanks for moral, support and help given.

Thanks and the highest appreciation to my family, my mother Hj. Ratnawati and my father Ir. H. Hamarullah for their pray, love and support. To my husband, Yusmadie, thank you for giving spirit in my life. Lastly, this thesis is dedicated to my sibling Arib Alfaris (1987-2004) and Ilham Mirza (1993-2004), may Allah rests their soul.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	Autoregressive
CAR	Conditional Autoregressive
LS	Least Squares
ML	Maximum Likelihood
SAR	Simultaneous Autoregressive
YW	Yule-Walker

KAEDAH *BOOTSTRAP* UNTUK MENILAI EFISIENSI PENGANGGAR MODEL RERUANG SESISI AR(1,1)

ABSTRAK

Tiga kategori data reruang ialah data geostatistik, data kekisi dan data pola titik. Tesis ini memfokuskan aspek penganggaran model reruang sesisi autoregresi untuk data reruang kekisi pada grid biasa dua dimensi. Secara khusus, tesis ini menilai efisiensi penganggar model reruang sesisi autoregresi, AR(1,1), menggunakan kaedah *bootstrap*. Kajian perbandingan dilakukan untuk membandingkan prestasi antara kaedah yang tersedia untuk menganggar parameter model AR(1,1), iaitu kaedah Yule-Walker, kaedah Yule-Walker tak-pincang, kaedah kuasa dua terkecil dan kaedah kebolehjadian maksimum. Dua jenis kaedah *bootstrap* dipertimbangkan, iaitu kaedah *bootstrap* reja dan kaedah *bootstrap* blok yang biasa digunakan pada analisis siri masa. Ralat piawai anggaran digunakan sebagai kriteria untuk mengukur efisiensi penganggar-penganggar. Untuk menunjukkan kebolehpercayaan anggaran, dibina selang keyakinan piawai. Perbezaan prestasi antara dua jenis kaedah *bootstrap* juga diperiksa. Sebagai tambahan, contoh berangka juga diberikan untuk menjelaskan prosedur kaedah *bootstrap* dalam menilai efisiensi penganggar. Kajian tesis ini mendapati, secara umum, anggaran Yule-Walker adalah lebih efisien berbanding anggaran yang lain dan kaedah *bootstrap* reja lebih mudah dan lebih konsisten daripada kaedah *bootstrap* blok.

BOOTSTRAP METHODS TO EVALUATE THE EFFICIENCY OF THE ESTIMATORS OF THE SPATIAL UNILATERAL AR(1,1) MODEL

ABSTRACT

Three categories of the spatial data are geostatistical data, lattice data and point pattern data. This thesis focuses on the estimation aspect of the spatial unilateral autoregressive models for spatial lattice data on two-dimensional regular grid. Specifically, this thesis evaluate the efficiency of the estimators of the first order spatial unilateral autoregressive model, AR(1,1) using the *bootstrapping* methods. A comparative studies are done to compare the performance among the available methods for estimating the parameters of AR(1,1) model, namely the Yule-Walker, the unbiased Yule-Walker, the least squares and the maximum likelihood methods. Two types of bootstrap methods are considered, namely *bootstrapping the residual* and *block bootstrap*, which are commonly used in time series analysis. The standard error of the estimate is used as criterion to assess the efficiency of the estimators. To indicate the reliability of the estimate, the standard confidence intervals are constructed. The differences of the performance between two types of bootstrap methods are also being examined. In addition, the numerical examples are also given to illustrate the procedure of the bootstrapping methods to assess the efficiency of the estimators. The results of the thesis show that, in general, the Yule-Walker estimate is more efficient as compared to the other estimates and bootstrapping the residual method is easier and more consistent than the block bootstrap method.

CHAPTER 1

INTRODUCTION

Spatial statistics discuss about pattern, relationship and trends of space and time. The usefulness of spatial statistics has attracted many scientists to utilize it in their research. In many aspects, as in the field of economics, health, agricultural, public safety and environmental sciences, spatial statistics is used to get more geographical information of the data and to model the data.

Many models have been proposed to analyze spatial data and the procedures have been developed to estimate the parameters of the models. These include the Yule-Walker, the least squares and the maximum likelihood estimators. The asymptotic properties of the estimate have been established for some estimators. However, in practice, we always deal with small to moderate sample sizes. Therefore, it is of interest to study the behaviors of the estimates, such as accurateness and efficiency, for small to moderate sample sizes.

This introductory chapter provides background of spatial processes, the statement of problems, research objectives and the organization of thesis.

1.1 Introduction to Spatial Processes

Let $X \in \mathrm{R}^d$ be the data location in d-dimensional Euclidean space. Usually, d = 1 is used in time series, d = 2 is used in geographic area and d = 3 is used in spatiotemporal. The spatial with random processes can be given as $\{Y_X : X \in D\}$ where D is subset of R^d (see Cressie, 1993).

Three categories of spatial data were defined by Cressie (1993) namely, geostatistical data, lattice data and point patterns data. These categories of the spatial data will be discussed in details in the next paragraphs.

In geostatistical data, a spatial process is indexed over a continuous space. The index can be written as D, where D is fixed subset of R^d and Y_X is a random variable at location $X \in D$. Geostatistics is regarded as hybrid discipline of mining engineering, geology, mathematics and statistics. It recognizes spatial variability at both the large scale or spatial trend and the small scale or spatial correlation. Trend-surface methods take on large-scale variation and assuming independent error. An important process in geostatistics is kriging, which can predict the ore grade in a mining block from observed sample. Examples of geostatistical application include modeling soil, studies on ground water, rainfall, public health and many more.

Lattice data is spatial data indexed over a lattice in space. In this case, D is a fixed (regular or irregular) and graph in R^d and Y_X is a random variable at location $X \in D$. For spatial data on regularly spaced set of points, it is analogous to time series data. However, in time series, the observations are obtained over a regularly spaced set of points. An example of application of lattice data can be seen in remote sensing by means of satellites or aircraft, in which the data come in the form of small rectangular shaped regions called pixels.

The data are called point patterns when D is a point process in R^d or subset of R^d and Y_X is a random variable at location $X \in D$. In point patterns, the important variable to be analyzed are the location of events, and whether or not the pattern is exhibiting a complete spatial randomness. Examples of this case include the spread of infectious diseases and the long life pines in an old growth forest.

In this thesis, the discussion will focus on the spatial lattice data on two-dimensional regular grid.

1.2 Statement of Problems

In time series, there is the natural distinction of past and future, and the value of the observation depends only upon past values, whereas in spatial, the dependence extends in all directions (see Whittle, 1954). An example for spatial case is fertilizer which is applied at any point in a field will ultimately affect soil fertility in all directions.

Some models are used to analyze the spatial data and many procedures are used to estimate the parameter of the models. Since the dependence in spatial data extends in all directions, the process of estimating the parameter of spatial models is more complicated. Several studies have been conducted to remedy the estimation problem of spatial models. Martin (1979) discussed the separable models where this type of model has a product correlation structure which makes the estimation simpler. The estimation of separable processes is then equivalent to estimation for one-dimensional processes. Basu and Reinsel (1993) focused on the unilateral models where this type of model can be analyzed using extension of time series theory.

The model of spatial processes in lattice which receives much attention is the spatial unilateral autoregressive model. Several methods have been proposed to estimate the parameter of these models, namely the Yule-Walker method (Tjøstheim, 1978), the least squares method and the maximum likelihood method (Awang, 2005). Guyon (1982) found that the Yule-Walker method of estimation is asymptotically biased. Ha and Newton (1993) shows that in fact the Yule-Walker estimator is much more biased than the least squares estimator. Awang (2005) used the maximum likelihood estimates approach to estimate the parameters of the model with some modification at the border. Bustos *et al.* (2009) applied the Yule-Walker, the least squares and another version of the maximum likelihood method for estimating the parameters of the spatial autoregressive models used in image filtering based.

The asymptotic properties of Yule-Walker estimate have been discussed by Tjøstheim (1978), Guyon (1982) and Basu and Reinsel (1992). Guyon (1982) illustrated the implementation of the Yule-Walker and the least squares methods to estimate the parameter of the first-order spatial unilateral autoregressive model. Ha and Newton (1993) estimated the parameters of the first-order spatial autoregressive model by Yule-Walker, unbiased version Yule-Walker and least squares method and the methods are compared for small and moderate sample by developing simulation study. Ten sample sizes representing small to moderate samples, with 500 realizations each were considered in the study and the measure of biasness was used to compare the performance of the methods. The result showed that the Yule-Walker estimator was much more biased than the least squares and unbiased Yule-Walker estimators.

In this thesis, the efficiency of the estimators of the spatial unilateral AR(1,1) model for small to moderate samples are studied via *bootstrapping* and the criterion used to assess the efficiency of the estimates is the standard error. The standard confidence intervals are constructed to indicate the reliability of the parameter estimates. To our knowledge, no such study has been done.

1.3 Research Objectives

Specifically, the objectives of this research are as follows.

1. To evaluate the efficiency of the estimators of the spatial unilateral AR(1,1) model by bootstrapping methods. The criteria used are the standard error and the standard confidence intervals.
2. To compare the results obtained from bootstrapping the residual method and the block bootstrapping method.
3. To illustrate the bootstrapping procedures in assessing the efficiency of the estimates by fitting the AR(1,1) model to real data set using the methods stated in (1). Here, the data used is the yield of wheat grain by Mercer and Hall (1911).

1.4 Organization of Thesis

Chapter 2 reviews the spatial processes. The discussion begins with introduction of the simultaneous autoregressive (SAR) model and continued with the conditional autoregressive (CAR) model, the unilateral model, the separable model and the first-order spatial unilateral AR model. The estimation methods for the spatial unilateral AR (1,1) model, namely the Yule-Walker method, the least squares method and the maximum likelihood method are given too.

In Chapter 3, the methods to evaluate the efficiency of the estimators of the spatial unilateral AR(1,1) model are explained. The methods used are bootstrapping the residual and block bootstrap. Before presents the methods, the methodology of bootstrap method for time series models and method of block bootstrap in time series models are given. Then, the methodology of bootstrapping the residual and block bootstrap for the spatial unilateral AR(1,1) model are proposed and explained in the later section.

Chapter 4 evaluates the efficiency of the estimators of the spatial unilateral AR(1,1) model for small to moderate sample sizes using the bootstrapping the residual method. Here, the discussion is about the estimation of the standard error and the construction of standard confidence intervals for the estimate of the spatial unilateral AR(1,1) parameters. The estimators used are the Yule-Walker estimator, unbiased version Yule-Walker estimator, the least squares estimators and the maximum likelihood estimator.

Chapter 5 presents the results of the second method used to estimate the standard error and the standard confidence intervals of the parameter estimates of the spatial unilateral AR(1,1) model, namely the block bootstrap method.

Comparison studies of bootstrapping the residual and block bootstrap procedures are presented in Chapter 6. This chapter shows comparison of the performance between two procedures to estimate the standard error of the parameter estimates of the spatial unilateral AR(1,1) model.

Chapter 7 illustrated the procedure discussed in Chapter 4 and Chapter 5 by fitting the spatial unilateral AR(1,1) model to real data set. The data used are wheat yield data by Mercer and Hall (1911).

Finally, the summary and conclusion with recommendation for further research of the research are given in Chapter 8.

CHAPTER 2

LITERATURE REVIEW ON SPATIAL MODELS

This chapter reviews the available models for spatial lattice data. The discussion begins with the simultaneous autoregressive (SAR) model in Section 2.1 and then the conditional autoregressive (CAR) model in Section 2.2. Sections 2.3 through 2.5 defined the unilateral model, separable model and the first order spatial unilateral autoregressive model, respectively. The estimation methods for the spatial unilateral AR (1,1) model, namely the Yule-Walker method, the least squares method and the maximum likelihood method are discussed in Sections 2.5.1, 2.5.2 and 2.5.3, respectively.

2.1 The Simultaneous Autoregressive (SAR) model

The simultaneous autoregressive (SAR) model which first defined by Whittle (1954) is given as,

$$\Phi(B_1 B_2) Y_{ij} = \varepsilon_{ij}\,, \tag{2.1}$$

where $\Phi(B_1 B_2) = \sum_s \sum_t \alpha_{st} B_1^s B_2^t$, $(s,t = \ldots, -2, -1, 0, 1, 2, \ldots)$ with B_1 and B_2 are translation operators defined by $B_1^s Y_{ij} = Y_{i-s,j}$, $B_2^t Y_{ij} = Y_{i,j-t}$ and ε_{ij} are independent variables with $E(\varepsilon_{ij}) = 0$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$.

From equation (2.1), we have $Y_{ij} = \dfrac{\varepsilon_{ij}}{\Phi(B_1, B_2)}$, and the model of equation (2.1) can be written as $Y_{ij} = \sum_s \sum_t \theta_{st} \varepsilon_{i-s,j-t}$, if only if , $\Phi(z_1, z_2)$ is not zero for any z_1 and z_2 which simultaneously satisfy $|z_1| = 1, |z_2| = 1$.

Ord (1975) proposed a maximum likelihood method to estimate the parameter of SAR model defined as,

$$Y_{ij} = \rho\left(\sum_s \sum_t w_{st} B_1^s B_2^t\right) Y_{ij} + \varepsilon_{ij}, \tag{2.2}$$

for $\{Y_{ij}\}$ with zero-mean and $\varepsilon_{ij} \sim IN(0, \sigma^2)$ for i = 1, 2, …, m; j = 1, 2, …, n. $\{w_{st}\}$ are a set of nonnegative weights which represent the 'degree of possible interaction' between locations and $\Phi(B_1, B_2) = \sum_s \sum_t \alpha_{st} B_1^s B_2^t = 1 - \rho\left(\sum_s \sum_t w_{st} B_1^s B_2^t\right)$ with $\alpha_{00} = 1$ and therefore, $w_{00} = 0$. The term in (2.3) can be reformulated in matrix form as

$$\mathbf{Y} = \rho\, \mathbf{WY} + \boldsymbol{\varepsilon}, \tag{2.3}$$

where $\mathbf{W}$ is the (N×N) matrix of weights, $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are (N×1) vectors, $N = mn$, and ρ is the parameter to be estimated. From (2.2), we have $\boldsymbol{\varepsilon} = \mathbf{AY}$, where,

$$\mathbf{A} = \mathbf{I} - \rho\mathbf{W}. \tag{2.4}$$

The log-likelihood function for ρ, σ^2, given that $\mathbf{Y} = \mathbf{y}$, is,

$$L(\rho, \sigma^2) = -(N/2)\ln(2\pi\sigma^2) - (1/2\sigma^2)\mathbf{y}'\mathbf{A}'\mathbf{A}\mathbf{y} + \ln|\mathbf{A}|. \tag{2.5}$$

The maximum likelihood estimators are given by,

$$\hat{\sigma}^2 = (\mathbf{Ay})'(\mathbf{Ay})/N$$

and $\hat{\rho}$ as that value of ρ which maximizes as,

$$L(\rho,\hat{\sigma}^2) = \text{const} - (N/2)\ln\left(\hat{\sigma}^2 |\mathbf{A}|^{-2/N}\right), \qquad (2.6)$$

where $|\mathbf{A}|$ is an Nth order polynomial in ρ. When N is not small, the evaluation of ρ becomes time consuming. Whittle (1954) gave a large sample approach based on spectral method. Ord (1975) provided the alternative approach to solve the problem with computational procedure and defined

$$|\mathbf{A}| = \prod_{i=1}^{N}(1-\rho\lambda_i),$$

where λ_i, $i = 1, \ldots, N$ are the eigen values of **W**. The evaluation of $\{\lambda_i\}$ will usually be a computer job, and the time involved becomes large as N increases.

For two-dimensional autoregressive model, Whittle (1954) shows that the least squares estimator is inconsistent. Ord (1975) removes the lack of consistency and gives the solution to the least squares. Unfortunately, the efficiency of such estimators relative to the maximum likelihood estimators declines drastically as ρ increases.

Whittle (1954) established the unilateral representation of two-dimensional process and show that for a given set of autocorrelations of SAR processes, there is a unique process in which Y_{ij} can be expressed as an autoregression upon Y_{it} $(t < j)$ and Y_{st} $(s < i,\ t$ unrestricted). The useful of the unilateral representation is that it suggests a simplifying change of parameters. Details about the unilateral model will be discussed in section 2.3.

2.2 The Conditional Autoregressive (CAR) Model

Barlett (1971) defined that the conditional probability distribution of $Y_{ij,}$ given all other site values, should depend only upon the values at the four nearest sites to (i,j), namely $Y_{i-1,j}$, $Y_{i+1,j}$, $Y_{i,j-1}$ and $Y_{i,j+1}$. The model can be written as,

$$P\,(Y_{ij} \mid \text{all other site values}) = P\,(\,Y_{ij} \mid Y_{i-1,j}, Y_{i+1,j}, Y_{i,j-1}, Y_{i,j+1}). \qquad (2.7)$$

The conditional probability formulation may be said to have rather more intuitive appeal, but has disadvantages. Firstly, there is no obvious method of deducing the joint probability structure associated with conditional probability model. Secondly, the conditional probability structure itself is subject to some unobvious and highly restrictive consistency conditions, see Besag (1974).

2.3 The Unilateral Model

The real usefulness of the unilateral representation is that it suggests a simplifying change of parameters (Whittle, 1954). The model can be analysed using extensions of time series theory in some special cases and the model is useful in the field of digital filtering and systems theory.

Martin (1996) shows the stationary d-dimensional nearest-neighbour process NN (d) is defined as a unilateral quadrant autoregression with dependence on the d adjacent preceding neighbour, one in each direction. When d = 2, for AR(1,1), Pickard (1980) and Tory and Pickard (1992) considered the process (Pickard process) which includes the diagonally adjacent term

$$Y_{ij} = \alpha_{10} Y_{i-1,j} + \alpha_{01} Y_{i,j-1} + \alpha_{11} Y_{i-1,j-1} + \varepsilon_{i,j} . \qquad (2.8)$$

The pickard process is stationary if

$$|\alpha_{10} + \alpha_{01}| < 1 - \alpha_{11} , \qquad |\alpha_{10} - \alpha_{01}| < 1 + \alpha_{11} . \qquad (2.9)$$

Barlett (1978) gives that if in the equation (2.7) the $\alpha_{11} = 0$, it is called a nearest neighbour (NN) model. The process (2.7) can be referred to as the first quadrant autoregressive (QAR(1,1)) process (Tjøstheim, 1978). The QAR(1,1) process with $\alpha_{11} = -\alpha_{10}\alpha_{01}$ is called a doubly geometric process (Martin, 1979), that is one natural extension to the plane of the one dimensional Markov process. Detail about doubly geometric process will be discussed in section 2.4.

Genton and Koul (2008) state that the unilateral AR processes in equation (2.8) have two main important reasons. First, the processes are useful for practical modeling because the process include a fairly flexible range of spatial correlation structures (see Besag (1972) and Basu and Reinsel (1993)). Secondly, QAR processes are the building blocks for inference in SAR models because the process can be used as auxiliary models in an indirect inferential procedure (see Luna and Genton (2002)).

2.4 The Separable Model

Linear-by-linear processes are a subclass of the simultaneous schemes and are simple extensions to the plane of processes on the plane. This process is regarded as separable model. For AR(1,1), the separable model known as doubly geometric processes that were (Martin, 1979), defined by,

$$Y_{ij} = \alpha_{10} Y_{i-1,j} + \alpha_{01} Y_{i,j-1} + \alpha_{10}\alpha_{01} Y_{i-1,j-1} + \varepsilon_{i,j} . \qquad (2.10)$$

The correlation at lag (s,t) of the separable model have the reflection symmetry, $\rho_{st} = \rho_{-s,-t} = \rho_{-s,t} = \rho_{s,-t}$.

Scaccia and Martin (2004) provided separable (and axially symmetric) processes are used to simulate the null distribution of tests for separability (and axial symmetric), while non-separable (and non-axially symmetric) processes were used to simulate the distribution of the tests under alternative hypothesis.

2.5 The First-order Spatial Unilateral AR Model

For the unilateral models, the spatial unilateral autoregressive model received much attention. This model is denoted as AR $(p_1,1)$ and is defined by,

$$Y_{ij} = \alpha_{10} Y_{i-1,j} + \alpha_{01} Y_{i,j-1} + \alpha_{11} Y_{i-1,j-1} + \cdots + \alpha_{p_1 0} Y_{i-p_1,j} + \alpha_{p_1,1} Y_{i-p_1,j-1} + \varepsilon_{ij} . \qquad (2.11)$$

Many procedures have been developed to estimate the parameters of this model. Tjøstheim (1978) discussed about the extension of the Yule-Walker method in time series analysis to spatial series. However, Guyon (1982), Basu and Reinsel (1992) and Ha and Newton (1993) showed that the autocovariance function used in the Yule-Walker estimate is asymptotically biased and they proposed the unbiased version of the estimate. Whittle (1954) discussed about the correct equations for the least squares estimates in two dimensional case. Awang (2005) introduced an alternative method using the maximum likelihood method with certain modification at the border to obtain the estimate of the parameters of the AR$(p_1,1)$ model.

The first-order spatial unilateral AR model is defined as,

$$Y_{ij} = \alpha_{10} Y_{i-1,j} + \alpha_{01} Y_{i,j-1} + \alpha_{11} Y_{i-1,j-1} + \varepsilon_{ij}. \quad (2.12)$$

Basu and Reinsel (1993) showed that this model is stationary if it follows the below conditions

1. $|\alpha_{10}|, |\alpha_{01}|, |\alpha_{11}| < 1,$
2. $\left(1+\alpha_{10}^2 - \alpha_{01}^2 - \alpha_{11}^2\right)^2 - 4\left(\alpha_{10} + \alpha_{01}\alpha_{11}\right)^2 > 0,$
3. $1 - \alpha_{01}^2 > |\alpha_{10} + \alpha_{01}\alpha_{11}|.$

The convergent representation of this model is presented by

$$Y_{ij} = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} \alpha_{10}^{k}\alpha_{01}^{l}\alpha_{11}^{r} (\varepsilon_{i-k-r,j-l-r}). \quad (2.13)$$

Basu and Reinsel (1992) obtained the asymptotic distribution of the spatial Yule-Walker estimator and showed that the spatial Yule-Walker estimator is asymptotically biased. The Yule-Walker estimators from the first-order spatial AR model are compared with the exact maximum likelihood estimators. In the next subsections, the available methods to estimate the parameters of the first-order spatial unilateral AR are presented.

2.5.1 Yule-Walker Method

Tjøstheim (1978) considered Yule-Walker estimators to estimate the parameters of the spatial unilateral $\text{AR}(p_1, p_2)$ models. The biased sample autocovariance function at lag (*s*,*t*) and (*s*,-*t*) for $s \geq 0$ and $t \geq 0$ is defined by

$$\hat{R}(s,t) = \frac{1}{mn} \sum_{i=1}^{m-s}\sum_{j=1}^{n-t} Y_{ij} Y_{i+s,j+t} \quad (2.14a)$$

and

$$\hat{R}(s,-t) = \frac{1}{mn}\sum_{i=1}^{m-s}\sum_{j=t+1}^{n} Y_{ij}Y_{i+s,j-t.} \tag{2.14b}$$

For a spatial unilateral $\mathrm{AR}(p_1, p_2)$ model, the spatial analogue of the one-dimensional Yule-Walker equation as in time series is given as,

$$R(s,t) = \sum_{k=0}^{p_1}\sum_{l=0}^{p_2} \alpha_{kl} R(s-k, t-l) \quad (s \geq 0, t \geq 0). \tag{2.15}$$

If we define $\boldsymbol{\alpha} = (\alpha_{10} \ldots, \alpha_{p_1,0}, \alpha_{01}, \ldots, \alpha_{p_1,1}, \ldots, \alpha_{0,p_2}, \ldots, \alpha_{p_1,p_2})'$,

$$\mathbf{r} = \left(R(1,0), \ldots, R(p_1,0), R(0,1), \ldots, R(p_1,1), \ldots, R(0,p_2), \ldots, R(p_1,p_2)\right)',$$

and $$\mathbf{R} = \begin{bmatrix} R(0,0) & R(-1,0) & \cdots & R(1-p_1, p_2) \\ R(1,0) & R(0,0) & \cdots & R(2-p_1, -p_2) \\ \vdots & \vdots & \cdots & \vdots \\ R(p_1-1, p_2) & R(p_1-2, p_2) & \cdots & R(0,0) \end{bmatrix},$$

then, the spatial Yule-Walker estimator of $\boldsymbol{\alpha}$ is simplified by

$$\hat{\boldsymbol{\alpha}}' = \mathbf{R}^{-1}\mathbf{r}. \tag{2.16}$$

The Yule-Walker equations are then solved with the $\hat{R}$ s replacing the R's. The resulting estimators, denoted by $\hat{\boldsymbol{\alpha}}$, are called the Yule-Walker estimator. Guyon (1982) compared $\hat{R}(s,t)$ with the unbiased estimator of $\gamma(s,t) = E(Y_{ij}Y_{i+s,j+t})$, $R(s,t)$ $= mn\ \hat{R}(s,t)/(m-s)(n-t)$, and claimed that the unbiased estimator was preferred.

To illustrate the Yule-Walker method in AR models with $p_1 = 1$ and $p_2 = 1$, let choose a spatial data on regular grid of (5×5) grid size. The biased sample autocovariance function at lag (s, t) and $(s,-t)$ for $s \geq 0$ and $t \geq 0$ is defined by

$$\hat{R}(s,t) = \frac{1}{(5)(5)} \sum_{i=1}^{5-s} \sum_{j=1}^{5-t} Y_{ij} Y_{i+s,j+t}$$

and

$$\hat{R}(s,-t) = \frac{1}{(5)(5)} \sum_{i=1}^{5-s} \sum_{j=t+1}^{5} Y_{ij} Y_{i+s,j-t.}$$

The sample autocovariance model can be compute as,

$$\hat{R}(0,0) = \frac{1}{25} \sum_{i=1}^{5} \sum_{j=1}^{5} (y_{ij})^2$$

$$= \frac{1}{25}(y_{11}y_{11} + y_{12}y_{12} + y_{13}y_{13} + y_{14}y_{14} + y_{15}y_{15} + y_{21}y_{21} + y_{22}y_{22} + y_{23}y_{23}$$

$$+ y_{24}y_{24} + y_{25}y_{25} + y_{31}y_{31} + y_{32}y_{32} + y_{33}y_{33} + y_{34}y_{34} + y_{35}y_{35} + y_{41}y_{41}$$

$$+ y_{42}y_{42} + y_{43}y_{43} + y_{44}y_{44} + y_{45}y_{45} + y_{51}y_{51} + y_{52}y_{52} + y_{53}y_{53} + y_{54}y_{54}$$

$$+ y_{55}y_{55}).$$

$$\hat{R}(1,0) = \frac{1}{25} \sum_{i=1}^{4} \sum_{j=1}^{5} y_{ij} y_{i+s,j+t}$$

$$= \frac{1}{25}(y_{11}y_{21} + y_{21}y_{31} + y_{31}y_{41} + y_{12}y_{22} + y_{13}y_{23} + y_{14}y_{24} + y_{15}y_{25} + y_{22}y_{32}$$

$$+ y_{23}y_{33} + y_{24}y_{34} + y_{25}y_{35} + y_{32}y_{42} + y_{33}y_{43} + y_{34}y_{44} + y_{35}y_{45} + y_{42}y_{52}$$

$$+ y_{43}y_{53} + y_{44}y_{54} + y_{45}y_{55}).$$

$$\hat{R}(0,1) = \frac{1}{25}\sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}y_{i+s,j+t}$$

$$= \frac{1}{25}(y_{11}y_{12} + y_{21}y_{22} + y_{31}y_{32} + y_{41}y_{42} + y_{51}y_{52} + y_{12}y_{13} + y_{13}y_{14} + y_{14}y_{15}$$

$$+ y_{22}y_{23} + y_{23}y_{24} + y_{24}y_{25} + y_{32}y_{33} + y_{33}y_{34} + y_{34}y_{35} + y_{42}y_{43} + y_{43}y_{44}$$

$$+ y_{44}y_{45} + y_{52}y_{53} + y_{53}y_{54}).$$

$$\hat{R}(1,1) = \frac{1}{25}\sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij}y_{i+s,j+t}$$

$$= \frac{1}{25}(y_{11}y_{22} + y_{21}y_{32} + y_{31}y_{42} + y_{41}y_{52} + y_{12}y_{23} + y_{13}y_{24} + y_{14}y_{25} + y_{22}y_{33}$$

$$+ y_{23}y_{34} + y_{24}y_{35} + y_{32}y_{43} + y_{33}y_{44} + y_{34}y_{45} + y_{42}y_{53} + y_{43}y_{54} + y_{44}y_{55}).$$

$$\hat{R}(1,-1) = \frac{1}{25}\sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}y_{i+s,j-t}$$

$$= \frac{1}{25}(y_{12}y_{21} + y_{13}y_{22} + y_{14}y_{23} + y_{15}y_{24} + y_{22}y_{31} + y_{23}y_{32} + y_{24}y_{33} + y_{25}y_{34}$$

$$+ y_{32}y_{41} + y_{33}y_{42} + y_{34}y_{43} + y_{35}y_{44} + y_{42}y_{51} + y_{43}y_{52} + y_{44}y_{53} + y_{45}y_{54}).$$

The Yule-Walker equation for AR(1,1) model is given as,

$$R(s,t) = \sum_{k=0}^{1}\sum_{l=0}^{1} \alpha_{kl} R(s-k, t-l),$$

then

$$\hat{R}(1,0) = \alpha_{10}\hat{R}(0,0) + \alpha_{01}\hat{R}(1,-1) + \alpha_{11}\hat{R}(0,1)$$

$$\hat{R}(0,1) = \alpha_{10}\hat{R}(1,-1) + \alpha_{01}\hat{R}(0,0) + \alpha_{11}\hat{R}(1,0)$$

$$\hat{R}(1,1) = \alpha_{10}\hat{R}(0,1) + \alpha_{01}\hat{R}(0,1) + \alpha_{11}\hat{R}(0,0)$$

The matrix can be written as,

$$\begin{bmatrix} R(1,0) \\ R(0,1) \\ R(1,1) \end{bmatrix} = \begin{bmatrix} \hat{R}(0,0) & \hat{R}(1,-1) & \hat{R}(0,1) \\ \hat{R}(-1,1) & \hat{R}(0,0) & \hat{R}(1,0) \\ \hat{R}(0,1) & \hat{R}(1,0) & \hat{R}(0,0) \end{bmatrix} \begin{bmatrix} \alpha_{10} \\ \alpha_{01} \\ \alpha_{11} \end{bmatrix}.$$

If we define

$\boldsymbol{\alpha} = (\alpha_{10}, \alpha_{01}, \alpha_{11})', \mathbf{r} = (R(1,0), R(0,1), R(1,1))'$ and

$$\mathbf{R} = \begin{bmatrix} R(0,0) & R(1,-1) & R(0,1) \\ R(-1,1) & R(0,0) & R(1,0) \\ R(0,1) & R(1,0) & R(0,0) \end{bmatrix},$$

the spatial Yule-Walker estimator of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}' = \mathbf{R}^{-1}\mathbf{r}.$$

Ha and Newton (1993) provided the asymptotic distribution of Yule-Walker and least squares estimators for two-dimensional causal autoregressive processes observed on a rectangular part of a lattice. They showed that the Yule-Walker estimate is asymptotically biased. The unbiased Yule-Walker and the least squares estimators have the same asymptotic properties as the Yule-Walker estimator except that the asymptotic bias is zero. By simulation studies, they compare the performance of the Yule-Walker, the unbiased Yule-Walker and the least squares in small and moderate samples with simulation of 500 realizations for each of ten sample sizes (8×8), (20×20), (8×10), (16×20), (15×25), (6×10), (6×15), (12×30), (5×20), (10×40) for each of three AR(1,1) model, having coefficients $\boldsymbol{\alpha}' = (0.2, 0.3, 0.2)$, $\boldsymbol{\alpha}' = (0.1, 0.5, 0.1)$, $\boldsymbol{\alpha}' = (0.7, 0.8, -0.6)$. The numerical results are not given, but they conclude that the Yule-Walker estimators are much more biased than the least squares and unbiased Yule-Walker estimators and that the bias conforms very closely to the theoretical bias given in theorem. The performance of the least squares and unbiased Yule-Walker estimators are remarkably similar.

2.5.2 Least Squares Method

This section discusses two types of conditional least squares estimation procedure to estimate the parameters of the spatial unilateral AR model. In the first type estimation, for estimating the parameters of the AR(1,1) model, (we may then call it as Type 1), we assume that the unobserved border values are all zeros, that is

$$\mathbf{Y_b}' = \left(Y_{-1,0}, \ldots, Y_{-1,n}, Y_{00}, \ldots, Y_{0n}, Y_{10}, \ldots, Y_{m0}\right) = \mathbf{0}$$

$$\mathbf{Y} = \left(Y_{11}, Y_{12}, \ldots, Y_{1n}, Y_{21}, Y_{22}, \ldots, Y_{2n}, \ldots, Y_{m1}, Y_{m2}, \ldots, Y_{mn}\right)'$$

then

$$Y_{11} = \alpha_{10}Y_{0,1} + \alpha_{01}Y_{1,0} + \alpha_{11}Y_{0,0} + \varepsilon_{11} = \varepsilon_{11}$$

$$Y_{12} = \alpha_{10}Y_{0,2} + \alpha_{01}Y_{1,1} + \alpha_{11}Y_{0,1} + \varepsilon_{12} = \alpha_{01}Y_{1,1} + \varepsilon_{12}$$

$$\vdots$$

$$Y_{1n} = \alpha_{10}Y_{0,n} + \alpha_{01}Y_{1,n-1} + \alpha_{11}Y_{0,n-1} + \varepsilon_{1n} = \alpha_{01}Y_{1,n-1} + \varepsilon_{1n}$$

$$Y_{21} = \alpha_{10}Y_{1,1} + \alpha_{01}Y_{2,0} + \alpha_{11}Y_{1,0} + \varepsilon_{21} = \alpha_{10}Y_{1,1} + \varepsilon_{21}$$

$$Y_{22} = \alpha_{10}Y_{1,2} + \alpha_{01}Y_{2,1} + \alpha_{11}Y_{1,1} + \varepsilon_{22}$$

$$\vdots$$

$$Y_{2n} = \alpha_{10}Y_{1,n} + \alpha_{01}Y_{2,n-1} + \alpha_{11}Y_{1,n-1} + \varepsilon_{2n}$$

$$\vdots$$

$$Y_{m1} = \alpha_{10}Y_{m-1,1} + \alpha_{01}Y_{m,0} + \alpha_{11}Y_{m-1,0} + \varepsilon_{m1} = \alpha_{10}Y_{m-1,1} + \varepsilon_{m1}$$

$$Y_{m2} = \alpha_{10}Y_{m-1,2} + \alpha_{01}Y_{m,1} + \alpha_{11}Y_{m-1,1} + \varepsilon_{m2}$$

$$\vdots$$

$$Y_{mn} = \alpha_{10} Y_{m-1,n} + \alpha_{01} Y_{m,n-1} + \alpha_{11} Y_{m-1,n-1} + \varepsilon_{mn}$$

In matrix form

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ \vdots \\ Y_{m1} \\ Y_{m2} \\ \vdots \\ Y_{mn} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Y_{11} & 0 \\ \vdots & \vdots & \vdots \\ 0 & Y_{1,n-1} & 0 \\ Y_{11} & 0 & 0 \\ Y_{12} & Y_{21} & Y_{11} \\ \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2,n-1} & Y_{1,n-1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,1} & 0 & 0 \\ Y_{m-1,2} & Y_{m,1} & Y_{m-1,1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1[\ ]} \end{bmatrix} \begin{bmatrix} \alpha_{10} \\ \alpha_{01} \\ \alpha_{11} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \\ \vdots \\ \varepsilon_{m1} \\ \varepsilon_{m2} \\ \vdots \\ \varepsilon_{mn} \end{bmatrix}.$$

The least squares estimate of $\boldsymbol{\alpha}' = (\alpha_{10}, \alpha_{01}, \alpha_{11})$ is given as,

$$\hat{\boldsymbol{\alpha}}' = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}, \qquad (2.17)$$

where

$$\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{1n}, Y_{21}, Y_{22}, \ldots, Y_{2n}, \ldots, Y_{m1}, Y_{m2}, \ldots, Y_{mn})'. \qquad (2.18)$$

and **X** is a matrix of dimension (*mn*) ×3 given as,

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Y_{11} & 0 \\ \vdots & \vdots & \vdots \\ 0 & Y_{1,n-1} & 0 \\ Y_{11} & 0 & 0 \\ Y_{12} & Y_{21} & Y_{11} \\ \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2,n-1} & Y_{1,n-1} \\ Y_{21} & 0 & 0 \\ Y_{22} & Y_{31} & Y_{21} \\ \vdots & \vdots & \vdots \\ Y_{2n} & Y_{3,n-1} & Y_{2,n-1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,1} & 0 & 0 \\ Y_{m-1,2} & Y_{m,1} & Y_{m-1,1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1} \end{bmatrix}.$$

In the second type, **Type 2,** the conditional least squares estimate is obtained by conditioning on the given observed border, $\mathbf{Y}_{\mathbf{o}}^{'} = \left(Y_{11}, Y_{12}, \ldots, Y_{1n}, \ldots, Y_{m1}, Y_{m2}, \ldots, Y_{mn}\right)'$.

For AR(1,1) model,

$$Y_{22} = \alpha_{10} Y_{12} + \alpha_{01} Y_{21} + \alpha_{11} Y_{11} + \varepsilon_{22}$$

$$\vdots$$

$$Y_{2n} = \alpha_{10} Y_{1,n} + \alpha_{01} Y_{2,n-1} + \alpha_{11} Y_{1,n-1} + \varepsilon_{2n}$$

$$Y_{32} = \alpha_{10} Y_{22} + \alpha_{01} Y_{31} + \alpha_{11} Y_{21} + \varepsilon_{32}$$

$$\vdots$$

$$Y_{3n} = \alpha_{10} Y_{2n} + \alpha_{01} Y_{3,n-1} + \alpha_{11} Y_{2,n-1} + \varepsilon_{3n}$$

$$\vdots$$

$$Y_{m-1,2} = \alpha_{10} Y_{m-2,2} + \alpha_{01} Y_{m-1,1} + \alpha_{11} Y_{m-2,1} + \varepsilon_{m-1,2}$$

$$\vdots$$

$$Y_{mn} = \alpha_{10} Y_{m-1,n} + \alpha_{01} Y_{m,n-1} + \alpha_{11} Y_{m-1,n-1} + \varepsilon_{mn}$$

In matrix form

$$\begin{bmatrix} Y_{22} \\ \vdots \\ Y_{2n} \\ Y_{32} \\ \vdots \\ Y_{3n} \\ \vdots \\ Y_{m-1,2} \\ \vdots \\ Y_{mn} \end{bmatrix} = \begin{bmatrix} Y_{12} & Y_{21} & Y_{11} \\ \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2,n-1} & Y_{1,n-1} \\ Y_{22} & Y_{31} & Y_{21} \\ \vdots & \vdots & \vdots \\ Y_{2n} & Y_{3,n-1} & Y_{2,n-1} \\ \vdots & \vdots & \vdots \\ Y_{m-2,2} & Y_{m-1,1} & Y_{m-2,1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1} \end{bmatrix} \begin{bmatrix} \alpha_{10} \\ \alpha_{01} \\ \alpha_{11} \end{bmatrix} + \begin{bmatrix} \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \\ \varepsilon_{32} \\ \vdots \\ \varepsilon_{3n} \\ \vdots \\ \varepsilon_{m-1,2} \\ \vdots \\ \varepsilon_{mn} \end{bmatrix}.$$

The estimator is given as,

$$\hat{\boldsymbol{\alpha}}_{\mathbf{o}} = \left(\mathbf{X_o}'\mathbf{X_o}\right)^{-1}\mathbf{X_o}'\mathbf{Y}_{(1)}, \qquad (2.19)$$

where $\mathbf{Y}_{(1)}{}' = (Y_{22}, \ldots, Y_{2n}, Y_{32}, \ldots, Y_{3n}, \ldots, Y_{m2}, \ldots, Y_{mn})$ and $\mathbf{X_0}$ is a matrix of dimension $(m-1)(n-1) \times 3$ defined as,

$$\mathbf{X_0} = \begin{bmatrix} Y_{12} & Y_{21} & Y_{11} \\ \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2,n-1} & Y_{1,n-1} \\ Y_{22} & Y_{31} & Y_{21} \\ \vdots & \vdots & \vdots \\ Y_{2n} & Y_{3,n-1} & Y_{2,n-1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,2} & Y_{m1} & Y_{m-1,1} \\ \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1} \end{bmatrix}.$$

2.5.3 Maximum Likelihood Method

The parameters of the spatial autoregressive models can be estimated by the method of maximum likelihood (see Awang, 2005). For AR(p_1,1) model defined by (2.10), the unobserved values is given as,

$\mathbf{Y} = (Y_{11}, Y_{12}, \cdots, Y_{1n}, Y_{21}, Y_{22}, \cdots, Y_{2n}, \cdots, Y_{m1}, Y_{m2}, \cdots, Y_{mn})' = (\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y_m})'$,

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \cdots, Y_{in})'$, $i = 1, 2, \cdots, m$ and the error vector

$\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \cdots, \varepsilon_{1n}, \varepsilon_{21}, \varepsilon_{22}, \cdots . \varepsilon_{2n}, \cdots, \varepsilon_{m1}, \varepsilon_{m2}, \cdots, \varepsilon_{mn})' = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon_m})'$,

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in})'$, $i = 1, 2, \cdots, m$ the matrix form for equation (2.10) can be defined as,

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Phi}_1 & \mathbf{\Phi}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Phi}_2 & \mathbf{\Phi}_1 & \mathbf{\Phi}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \ddots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Phi}_{p_1} & \mathbf{\Phi}_{p_1-1} & \cdots & \mathbf{\Phi}_2 & \mathbf{\Phi}_1 & \mathbf{\Phi}_0 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}, \quad (2.20)$$

where $\mathbf{\Phi}_j$'s are $n \times n$ matrices defined as,

$$\mathbf{\Phi}_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{01} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{01} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{01} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{01} & 0 \end{pmatrix} \text{ and } \mathbf{\Phi}_j = \begin{pmatrix} \alpha_{j0} & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{j1} & \alpha_{j0} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{j1} & \alpha_{j0} & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{j1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \cdots & \alpha_{j1} & \alpha_{j0} \end{pmatrix}.$$

The compact form is,

$$\mathbf{Y} = \mathbf{\Phi Y} + \boldsymbol{\varepsilon} .$$

If the $\mathbf{\Phi}$ is decomposed into $2p_1 + 1$ matrices such that it isolate different parameters, $\mathbf{Y}$ can be written as,

$$\mathbf{Y} = (\alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \cdots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1})\mathbf{Y} + \boldsymbol{\varepsilon}, \quad (2.21)$$

where $\mathbf{\Phi} = \alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \cdots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1}$ and

$\mathbf{W}_{jk}$, $j = 1, 2, \ldots, p_1$; $k = 0,1$ are $N \times N$ lower triangular weight matrices with elements ones and zeroes.

To estimate the parameter of AR$(p_1,1)$ model, equation (2.20) can be written as,

$$\mathbf{Y} = \left(\mathbf{I}\left(\alpha_{10}\mathbf{W}_{10} + \alpha_{01}\mathbf{W}_{01} + \alpha_{11}\mathbf{W}_{11} + \cdots + \alpha_{p_1 0}\mathbf{W}_{p_1 0} + \alpha_{p_1 1}\mathbf{W}_{p_1 1}\right)\right)^{-1} + \boldsymbol{\varepsilon} \quad (2.22)$$

or

$$\mathbf{Y} = (\mathbf{I} - \mathbf{\Phi})^{-1}\boldsymbol{\varepsilon}, \quad (2.23)$$

where $\mathbf{I}$ is a $N \times N$ identity matrix.

The covariance matrix of $\mathbf{Y}$, $\mathbf{V}$ is defined as,

$$\mathbf{V} = \sigma^2 (\mathbf{I} - \mathbf{\Phi})^{-1}\left((\mathbf{I} - \mathbf{\Phi})^{-1}\right)' \quad (2.24)$$

and the determinant of equation (2.24) is defined as,

$$|\mathbf{V}|^{1/2} = \left(\sigma^2\right)^{N/2}\left|(\mathbf{I} - \mathbf{\Phi})^{-1}\right|. \quad (2.25)$$

Since $(\mathbf{I} - \mathbf{\Phi})$ is the lower triangular matrix with diagonal elements 1, $\left|(\mathbf{I} - \mathbf{\Phi})^{-1}\right| = 1$, equation (2.24) can be written as,

$$|\mathbf{V}|^{1/2} = \left(\sigma^2\right)^{N/2}. \quad (2.26)$$

By assuming the normality of ε, the likelihood function is defined as,

$$l = \frac{1}{(2\pi)^{N/2}|\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}\right\}$$

$$= (2\pi)^{N/2}(\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{Y}'\left[(\mathbf{I} - \mathbf{\Phi})^{-1}(\mathbf{I} - \mathbf{\Phi})^{-1}\right]^{-1}\mathbf{Y}\right\}$$

$$= (2\pi)^{N/2}(\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{Y}'(\mathbf{I} - \mathbf{\Phi})(\mathbf{I} - \mathbf{\Phi})\mathbf{Y}\right\},$$

and thus, the log likelihood function is formed as,