5-1-2014

Generalized method of moments approach for spatial-temporal binary data

Kimberly Ann Kaufeld

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

GENERALIZED METHOD OF MOMENTS APPROACH FOR SPATIAL-TEMPORAL BINARY DATA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Kimberly Ann Kaufeld

College of Education and Behavioral Science
Applied Statistics and Research Methods

May 2014

This Dissertation by: Kimberly A. Kaufeld

Entitled: *Generalized Method of Moments Approach for Spatial-Temporal Binary Data*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences, Program of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

_____

Trent Lalonde, Ph.D., Research Advisor

_____

Jay Schaffer, Ph.D., Co-Research Advisor

_____

Khalil Shafie, Ph.D., Committee Member

_____

Scott Franklin, Ph.D., Faculty Representative

Date of Dissertation Defense _____

Accepted by the Graduate School

_____

Linda L. Black, Ed.D.
Dean of the Graduate School and International Admissions

ABSTRACT

Kaufeld, Kimberly A. *Generalized Method of Moments Approach for Spatial-Temporal Binary Data*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2014.

Binary data that are correlated across space and time often occur in health and ecological studies. The centered spatial-temporal autologistic regression model (Wang & Zheng, 2013) accounts for the spatial and temporal dependence that can occur in binary data. Statistical inference for the autologistic model has been based upon pseudolikelihood, Monte Carlo maximum likelihood, Monte Carlo expectation maximization or Bayesian hierarchical models. However, these methods require the full conditional distribution to be defined and with the complexity of spatial and temporal dependence and interactions between observations can cause convergence problems and increase computation time. An alternative approach to likelihood based methods for spatial-temporal data is to use generalized method of moments, a method not currently used for spatial-temporal binary data. Two different generalized method of moments approaches based on a set of moment conditions is constructed with respect to spatial neighborhoods to account for the spatial and temporal dependence of the data. Comparisons of estimation methods using a small simulation and two data sets show that the generalized method of moments approaches perform well in specific data situations.

DEDICATION

For Kevin whose patience is beyond limits. You supported me in pursuing my dreams every step of the way.

ACKNOWLEDGMENTS

I am thankful for the support I had along the way at University of Northern Colorado. My achievements would not have been attainable without the encouragement that I was given there. Firstly, I want thank my advisor, Dr. Trent Lalonde, for his openness to take on a dissertation topic that paired our two areas of interests making this dissertation possible. I am grateful for the stimulating conversations, insight, patience, and guidance provided along the way.

Additionally, I am appreciative of my committee members. I am thankful to Jay Schaffer for the advice and support throughout my graduate experience. To Khalil Shafie, whose comments and discussions provided additional avenues to explore in my research. I am also grateful to Scott Franklin for additional insight on the ecological aspect of this work, it has strengthened my research and am thankful for the feedback.

I am immensely grateful for the support and encouragement from my family and particularly my husband, Kevin. They were always there to lend a ear or provide words of encouragement through countless conversations during the difficult times in graduate school. I am thankful for the advice from my parents and the backing they gave me through this process. I also feel fortunate to have met Jamie

Riggs whose drive and ambition prompted me to pursue opportunities outside the department, which enriched my graduate experience.

Thank you to the Colorado State Forest service that provided data for the Rocky Mountain Forest Service data. The trainings and assistance with GIS data offered from the Aerial Survey team was very valuable; meaningful interpretations could not be made without the additional help provided. Data was obtained from the PRISM dataset is publicly available at `http://www.prism.oregonstate.edu/`.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

SGLMM	Spatial Generalized Linear Mixed Models
SAR	Simultaneous Autoregressive Model
CAR	Conditional Autoregressive Model
MCMC	Markov Chain Monte Carlo
MCML	Monte Carlo Maximum Likelihood
PS	Perfect Simulation Sampling
PGS	Perfect Simulation with Gibbs Sampling
GMM	Generalized Method of Moments
GMM_2	Second Generalized Method of Moments Approach
GMM_3	Third Generalized Method of Moments Approach
MPL	Maximum Pseudolikelihood
EMPL	Expectation-Maximization Pseudolikelihood
MCEML	Monte Carlo Expectation-Maximization Likelihood

CHAPTER I

INTRODUCTION

Spatial-temporal models are progressively becoming an important method to assess public health and environmental issues. As data are collected for longer periods of time and over larger geographic regions efficient modeling and estimation techniques are needed. Spatial-temporal models are useful in explaining and understanding the underlying structure of data observations that occur across space and time while accounting for both spatial and temporal dependence among geographic locations. In particular, the binary spatial-temporal model helps to explain the data through applications such as modeling the spread of disease within a county, the presence of a species over a geographic region, or peer networks in health and educational settings.

Binary spatial-temporal data have values of either 0 or 1, which are measured repeatedly over time on a spatial lattice. Applications of binary data commonly appear in agriculture, ecology, climatology, geography and image analysis. In modeling binary areal data, there are two dominant approaches, the logistic spatial generalized linear mixed model (SGLMM) and the autologistic model. Both of these approaches are characterized by how they model spatial dependence. The

SGLMM indirectly accounts for spatial dependence using a latent Gaussian Markov random field applied over a lattice whereas the autologistic model directly accounts for the spatial dependence through the use of an autocovariate, a function of the observations themselves.

The increase in spatio-temporal modeling within multiple disciplines makes the development of methods to efficiently capture the nature of the data increasingly important. As there are different types of spatial data, it is important to be able to model the data effectively. One way to capture the spatial dependence of the data is to construct a stationary or non-stationary lattice, which will be used in this dissertation.

Binary Spatial Data Models

In spatial-temporal binary models the nature of the data is captured by the fact that observations that are closer in space and time are more similar than those that are further apart. A model that directly accounts for binary data that are correlated across space and time is the centered spatial-temporal autologistic regression model developed by Wang & Zheng (2013). It uses logistic regression to model a response variable on explanatory variables and autoregression on responses from spatial neighborhoods. The model is an extension of the atemporal version of the autologistic model developed by Besag (1972, 1974) and the non-centered spatial-temporal version by Zhu, Huang & Wu (2005). It uses logistic regression to model three components, (1) the response variable on explanatory variables (2) the au-

toregression on responses from spatial neighborhoods or locations due to the spatial dependence among sites, and (3) the autoregression of the temporal term due to different discrete times. It incorporates spatial correlation while modeling the relationship between the spatial binary response and the explanatory variables (Cressie, 1993; Zhu et al., 2005; Zheng & Zhu, 2008).

Autologistic regression is traditionally modeled by a conditional probability distribution. The model parameters are estimated using a joint conditional likelihood function that contains a normalizing constant, introducing difficulty in direct maximization of the function. To maximize the likelihood function, research has generally been based upon the maximum pseudolikelihood (MPL), Markov Chain Monte Carlo (MCMC) approximation of the likelihood, Bayesian hierarchical models and most recently Monte Carlo expectation-maximum likelihood estimation (MCEML).

Besag (1975) proposed maximizing pseudolikelihood functions as a method for estimation for the autologistic model, which has been widely used. Huffer & Wu (1998) later used MCMC to approximate the unknown normalizing constants with maximum likelihood estimation (MLE) for a spatial autologistic model. The pseudolikelihood estimates were generalized by Huang & Ogata (2002) as they proposed maximum generalized pseudo-likelihood estimates that connect the maximum pseudolikelihood estimates (MPLE) and MLE. The generalized method improved the efficiency and standard errors as compared to MPL and maximum likelihood estimators. Zheng & Zhu (2008) proposed a Bayesian hierarchical model

framework for the autologistic regression model and compared the performance of MPL, MCMC maximum likelihood to Bayesian inference. They discovered that the Bayesian and MCMC maximum likelihood performed similarly, however MPL was found to be statistically inefficient especially when spatial and/or temporal dependence is strong. The most recent estimation method proposed by Wang & Zheng (2013) is the expectation-maximization pseudolikelihood (EMPL) and Monte Carlo expectation-maximization (MCEML) for a centered spatial-temporal autologistic regression model.

The added complexity of spatial and temporal dependence and associations between observations, likelihood based methods may not be the most efficient estimation techniques. In complex spatial-temporal binary models the likelihood-based methods can have convergence issues as well as significant increases in computation time. This is due to the need for the full conditional probability distribution to be defined. An alternative approach to the autologistic regression model for analyzing binary spatial-temporal data was proposed using marginal models with quasi-likelihood estimating equations. The quasi-likelihood (QL) approach allows separate modeling of regression and spatial and temporal dependence of the response variables. Lin, Lee & Clayton (2009) developed quasi-likelihood estimating equations for non-separable spatial-temporal data and the efficiencies of the estimates of QL were compared to maximum pseudo-likelihood estimates. Despite these advances in estimation, the QL estimation equations, defined by Lin et al. (2009) for separable and non-separableLin (2010) spatial temporal binary data were con-

structed from the correlation structure of exponential models and distance measurements rather than the natural grid structure. A binary Markov random field is a nice alternative, which captures the binary data information through neighbor definitions. Generally, under these conditions, a joint conditional likelihood is defined and MCML, MPL, MCEML and Bayesian methods are performed to estimate the parameters.

Purpose

The purpose of this research is to develop an alternative to traditional likelihood methods using Generalized Method of Moments (GMM) for a centered spatial-temporal autologistic model with stationary and non-stationary data. The centered spatial-temporal autologistic model will be used for stationary data and a modified centered spatial-temporal model will be constructed to deal with non-stationary spatial frame i.e. social network framework. The alternative estimation method using GMM does not have the restriction that the full conditional distribution be defined, but rather can be specified by the first two moments. It will use a binary Markov random field to capture the natural structure of binary data. A set of estimating equations will be specified based upon a working correlation structure constructed to deal with the spatial and temporal dependence of the binary data. The following research questions are addressed:

Q1 How will generalized method of moments be developed for the spatial-temporal binary autologistic model?

Q2 Will the computation time of generalized method of moments be comparable to other estimation methods such as Monte Carlo Markov Chain maximum likelihood and Monte Carlo expectation-maximization estimation?

Q3 Do the estimation methods, generalized method of moments, Monte Carlo Markov Chain maximum likelihood and Monte Carlo expectation-maximization converge?

Q4 How does the bias of the generalized method of moments parameter estimates compare to other estimation methods such as Monte Carlo Markov Chain maximum likelihood and Monte Carlo expectation-maximization estimation?

Q5 How does the precision of generalized method of moments compare to other estimation methods such as Monte Carlo Markov Chain maximum likelihood and Monte Carlo expectation-maximization estimation?

To make comparisons a small simulation study constructed on a 5×5 and 10×10 spatial frame is conducted for two different levels of spatial dependence. In addition two real world datasets, the National Longitudinal Study of Adolescent Health (Add Health) data and Rocky Mountain Pine Beetle data are used to demonstrate the different estimation techniques in terms of convergence, computation time and parameter estimates.

The Rocky Mountain Pine Beetle data is a public dataset collected by the United States Forest Service through aerial survey methods across the Rocky Mountain Region from 2000-2010. The data are composed of categories of damage-causing agents, host tree species (trees in which the damage has taken place), number of dead trees, numbers of acres, surveyor identifications and numbers of dead trees per acre. A stationary grid of dimension 42×55 is constructed and at each site. A binary response is noted based upon presence or absence of the damage-causing

agent, bark beetle, at each spatial location. For each grid cell August mean maximum temperature in degrees Celsius, January mean minimum temperature in degrees Celsius, mean annual precipitation in inches, elevation in feet are calculated. Each of these components has been shown to have an impact on mountain pine beetle outbreaks in the western United States (see, for example, Waring & Pitman, 1985; Mitchell & Preisler, 1991; Negron & Popp, 2004). Weather variables were taken from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) dataset, which is publicly available at `http://www.prism.oregonstate.edu/`. The PRISM data estimates monthly weather data over a contiguous grid at a resolution of 0.0416 decimal degrees latitude and longitude ($\sim$4 km) cells (Daly, Gibson, Taylor, Johnson & Pasteris, 2002) and aligns with the resolution of the gridded mountain pine beetle data.

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States assessed over three different time periods. The survey combines data on social, economic, psychological and physical well being based upon social contexts such as friendships, family, and neighborhood. It contains detailed information about the respondents' characteristics and peer group networks over the course of several years. The Add Health data will be used to assess the effect of peer networks at multiple levels (friend and school) using a binary variable, adolescent drug use (yes or no) and alcohol use (yes or no).

The outline of the dissertation is as follows. Chapter 2 provides relevant background and literature review to develop the estimation technique for the centered spatial-temporal autologistic model. Chapter 3 develops the Generalized Method of Moments approach for spatial-temporal binary data on a stationary and non-stationary grid. In Chapter 4, GMM is compared to previous estimation techniques using a small simulation, National Longitudinal Study of Adolescent Health Data and Rocky Mountain Forest Service Data. In Chapter 5, conclusions for the Generalized Method of Moments methods are presented as well as future developments.

CHAPTER II

SPATIAL-TEMPORAL BINARY MODELS

This chapter reviews the literature of spatial data analysis and the development of spatial and spatial-temporal binary models used to provide the necessary background for the development of the model and estimation procedure that this dissertation serves. The introduction provides an overview of what spatial data is and how it is used. The spatial data section provides a definition for different types of binary spatial data. The spatial generalized linear mixed models (SGLMM) section describes the spatial generalized linear mixed model (SGLMM) commonly used for binary spatial data. The autologistic regression models section describes the autologistic model, a competing model to SGLMM, and its development to the centered spatial-temporal autologistic model. The statistical inference for binary spatial models section describes the different estimation methods used for SGLMM and autologistic models.

Introduction

Types of data that provide when and where data were collected are called spatial-temporal data. The spatial component is the location or where the data

were recorded and the temporal component is when the data were recorded, at a discrete time. There are three different types of spatial data, geostatistical data, lattice data, and point patterns. Each of these three types of data has an important statistical characteristic: the observations that are closer in space and time are more similar than those that are further apart. This implies that the data are not statistically independent as there is dependence in both time and space, temporal and spatial dependence. Spatial-temporal binary models, models that have binary response variables, account for both spatial and temporal dependence in the data. Two types of models exist to deal with the dependence structures, (1) Spatial-Temporal Generalized Linear Models and (2) Autologistic Models. In this dissertation, the two types of models will be specified with respect to a spatial lattice.

The following section provides an overview of types of spatial data.

Spatial data

Spatial data are observations that occur over locations or within specific regions. Spatial data can be categorized into three different categories, geostatistical data, point pattern data, and lattice data (Cressie, 1993). The three types of spatial data differ based upon how the observations are recorded. In addition, the approaches for modeling the three types of spatial data differ.

Geostatistical Data

Geostatistical data are used when the data are spatially continuous in a region but the data are taken at fixed locations (Cressie, 1993). The data are used in geostatistics, a branch of spatial statistics that was developed by Matheron (1963) in the early 1980s, to predict probability distributions of ore grades for mining from disciplines of geology, mining engineering, mathematics and statistics. The main goal of geostatistical data is to summarize spatial correlation and draw inferences from the data. A popular method to conduct these types of analyses is by kriging, an interpolation tool based upon linear least squares estimation algorithms developed in the 1980s (Cressie, 1993). A common example is to map climatological temperatures of a region based upon a small sample of observed temperatures from regional weather stations.

Point Patterns

Spatial point pattern data occur when the locations of events are the variables of interest. Point patterns consist of a finite number of locations observed in a spatial region. Point patterns are used to answer whether the pattern from the locations exhibits complete spatial randomness, clustering or regularity. The goal is to estimate parameters of the random set and the point process (see Cressie, 1993; Laslett, McBratney, Pahl & Hutchinson, 1987) and identify, quantify and model the inherent spatial pattern of the data.

Lattice Data

A lattice, $\mathcal{L}$, is a finite collection of spatial sites, $s = 1, \cdots, n$, which exist on either a regular or irregular lattice structure. In spatial analysis, the lattice refers to a countable collection of regular or irregular sites linked to spatial neighborhood information. A lattice is constructed based upon the data where each site is indexed as $Z(\boldsymbol{s})$. A regular lattice has an inherent structure in which all of the sites are of equivalent size i.e. a grid. An irregular lattice, however, can have a different structure for each of the sites. For example, state counties. Sites that share a common boundary or are within a distance, d, between sites (e.g. Euclidean) are called neighbors. The neighbors of the s^{th} site create a neighborhood structure $\mathcal{N}$. For instance, in the regular lattice case the neighborhood structure for site s can be defined as the four adjacent borders whereas on an irregular lattice the neighborhood structure is generally defined using a Euclidean distance between sites or adjacent sites. A neighborhood structure for the sites, $\mathcal{S}$, is defined as

$$\mathcal{N} = \{\mathcal{N}_s | \forall s \in \mathcal{S}\},$$

where $\mathcal{N}_s$ is the set of sites that neighbor site s but does not neighbor itself ($s \notin \mathcal{N}_s$).

The neighbors of the sth site are ordered according to the adjacency of the sth site, the further away a site i is to another site j, the higher the order. The order of the neighborhood, first order, second order, etc. is based upon a distance, d, $2d$, etc., respectively, of a neighbor from a site, s. The neighbors of the k^{th} order

are denoted as $\mathcal{N}_s^{(k)}, k = 1, \cdots K$, where K is the largest order of the neighborhood, such that $\mathcal{N}_s = \cup_{i=1}^{K} \mathcal{N}_s^{(k)}$. The spatial dependence among the sites are defined by a spatial lattice process, $\boldsymbol{Z(s)} : \boldsymbol{s} \in \mathcal{L}$, which uses a Markov random field to define the spatial dependence among the $\mathcal{S}$ sites. In this scenario, a model can be built from the spatial dependence between counties.

To better display the neighborhoods of an irregular lattice, Figure 1 shows a map of Colorado counties numbered in alphabetical order.

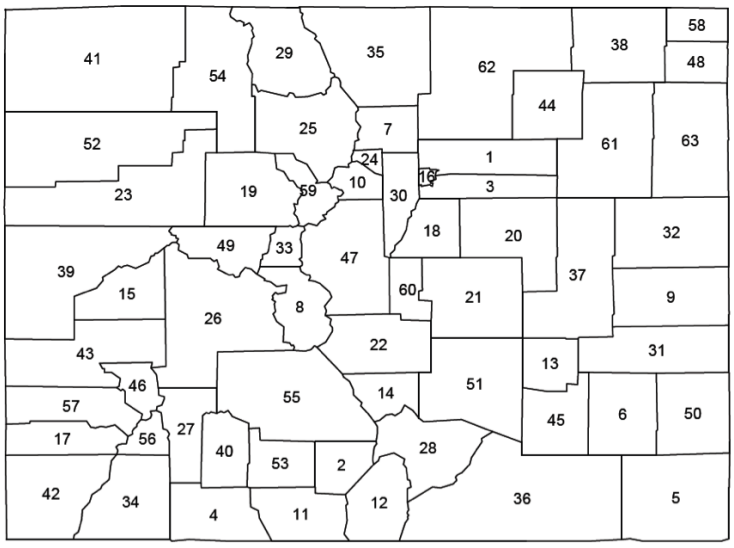


Figure 1. Map of Colorado counties numbered in alphabetical order

In this scenario a spatial lattice, $\mathcal{L}_N$ of the sth county, and the neighborhood set $\mathcal{N}_s^{(k)}$ can be specified as

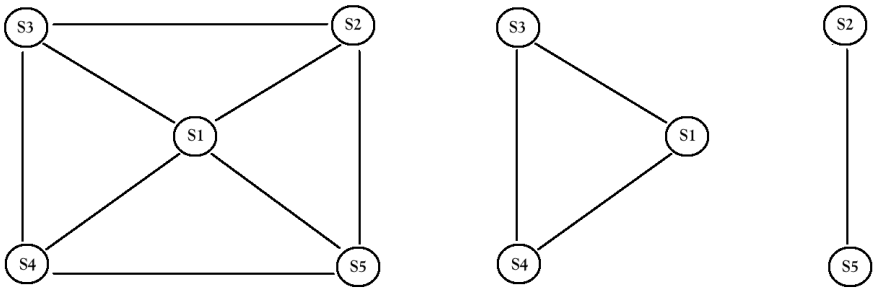
$$\mathcal{L}_N \equiv \{\mathcal{N}_s : s = 1, \cdots, 63\}$$

$$\mathcal{N}_s^{(k)} \equiv \text{Spatial neighborhoods of } k^{th} \text{ order for } \mathcal{N}_s, s = 1, \cdots, 63$$

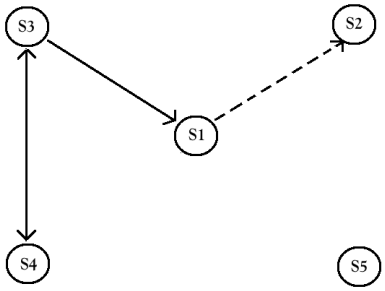
For example, Boulder County is labeled site 7. The first order spatial neighbors of site 7, based upon adjacency, are sites $\{1, 24, 25, 30, 35, 62\}$. Given a neighborhood structure spatial dependence exists among the sites. In terms of the Colorado counties the neighborhoods associated with site 7 are more closely related than characteristics in site 42, which is several orders of neighborhoods away.

A non-stationary grid, on the other hand, might depend upon other types of distance, i.e. Euclidean distance. A special non-stationary lattice structure is social networks. In this case, peer networks defined creates a grid based upon the nature of relationship or proximity an individual is to another. As social networks can change over time the spatial structure of the data also changes. An example of a grid based upon three different levels is displayed in Figure 2.

Figure 2(a) displays the spatial framework at the school level. A solid line connects all the students to each other. Figure 2(b) displays a possible spatial relationship among sites where the solid line represents students who are in the same grade. In this scenario, students S1, S3, and S4 are in one grade and S2 and S5 are in a different grade. Figure 2(c) displays a possible relationship among friends within a school. The solid lines represent friends within the same grade and the dotted line represents friends across grade levels within a school. In this scenario, S1 has friends S2 and S3 where S2 is a friend in a different grade level. S3 has a reciprocal relationship with S4 and a non-reciprocal friend relationship with S1. However, S5 is not friends with any others, S1-S4. Notice that school level peers



(a) School Level Peers (b) Grade Level Peers



(c) Friend Level Peers

Figure 2. Peer Network Spatial Framework

will likely change from one year to the next so the relationship of S1 to S2 will be reflected on a non-stationary spatial grid.

The purpose of analyzing spatial lattice data is to draw inference and identify relationships between data locations. This is different from geostatistical data as no possibility of a response between data locations can occur. A common example is mountain pine beetle outbreaks (MPB), in which the spatial lattice shows presence or absence of a particular beetle outbreak in Colorado.

The focus of this dissertation is on spatial-temporal binary data measured repeatedly over space and time on a stationary lattice and non-stationary peer network spatial frame.

Spatial Generalized Linear Mixed Models (SGLMM)

The Spatial Generalized Linear Mixed Model (SGLMM) was proposed by Besag (1991) for areal data. It is a hierarchal model that introduces spatial dependence through a latent Gaussian Markov random field. The model, initially introduced for count data, has been extended to binary applications for regression and prediction in such fields as ecology, geology, forestry, and health. It uses a conditional approach that incorporates the unobserved spatial process as a random effect within the mean function and models the conditional mean and variance as a function of both fixed effects and random effects. The random effects are derived from an unobserved spatial process from a Gaussian Markov random field (GMRF), commonly referred to as spatial autoregressive models.

The SGLMM for areal data developed by Besag (1991) is constructed as follows. Let $G = (I, E)$ be an underlying graph, where $I = \{1, 2, \cdots, n\}$ are sites from an area of interest and E are edges (each of which is a pair of sites, (i, j), that represents the proximity of areas i and j). Let an adjacency matrix, $\mathbf{A}$ be an $n \times n$ matrix representing the graph, G, given by $\text{diag}(\mathbf{A}) = \mathbf{0}$ and $\mathbf{A}_{ij} = \mathbf{1}\{(i, j) \in E, i \neq j\}$, where $\mathbf{1}$ denotes an indicator function. Let $\boldsymbol{Z} = (Z_1, \cdots, Z_n)'$ denote a random

field where Z_i is the random variable associated with site i. The SGLMM model is given by

$$g(E(Z_i|\boldsymbol{\beta}, \mathbf{W}_i)) = \boldsymbol{X_i}\boldsymbol{\beta} + \mathbf{W}_i \tag{1}$$

where g is a link function, $\boldsymbol{X}_i$ is the i^{th} row of the design matrix, $\boldsymbol{\beta}$ is a p-dimensional vector of parameters and $\mathbf{W}_i$ is a spatial random effect associated with site i. For a binary spatial model, the link function, g, is the logit function, $\text{logit}(p) = \log\{p/(1-p)\}$. The random effects, $\mathbf{W}_i$ are assumed to follow a spatial autoregressive model by either a (1) simultaneous autoregressive (SAR) or (2) conditional autoregressive (CAR) model.

The spatial autoregressive model uses an autoregressive model similar to time series models where the data at a particular site, i, given its neighboring sites is a linear combination of the neighboring sites. The autoregression in the spatial model induces a dependence structure for the data. Spatial autoregressive models create spatial correlation in the data by autoregression. The proximity at site, i, to a spatial location, j, is used to define a neighborhood structure among the data rather than specifying the spatial autocorrelation structure directly.

Simultaneous Autoregressive (SAR) Model

The SAR model was introduced by Whittle (1954) to deal with the spatial variation in crop fields. It is a generalized linear model with a spatial random effect, specified by a link function g, where the spatial errors follow a spatial Gaussian process. The SAR model is specified as follows. Let

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}, \tag{2}$$

$$\boldsymbol{W} = \mathbf{B}\boldsymbol{W} + \boldsymbol{\epsilon},$$

then

$$(\mathbf{I} - \mathbf{B})(\boldsymbol{Y^*} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{Y}^*$ denotes the link function $g(E(\boldsymbol{Y}|\boldsymbol{\theta}, \mathbf{W}))$ with respect to count, binary, or continuous data, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$, and $\mathbf{B}$ is a matrix of spatial dependence parameters with the diagonal entries $b_{ii} = 0$ as a site cannot be a neighbor with itself. For a binary response, $\boldsymbol{Y}$, the link function g is the logit function, $\text{logit}(p) = \log\{p/(1-p)\}$.

The variance-covariance matrix, $\boldsymbol{\Sigma}_{SAR}$, of $\boldsymbol{Y}^*$ can be derived from (3) as

$$\begin{aligned} \boldsymbol{\Sigma}_{SAR} &= \text{Var}[\boldsymbol{Y^*}] \\ &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}_\epsilon(\mathbf{I} - \mathbf{B})'^{-1}, \end{aligned} \tag{4}$$

assuming that $(\mathbf{I} - \mathbf{B})^{-1}$ exists. Then

$$\boldsymbol{Y}^* \sim N(\boldsymbol{X}\boldsymbol{\beta}, (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}_\epsilon(\mathbf{I} - \mathbf{B})'^{-1}). \tag{5}$$

The autoregression of the model induces a general covariance structure of the data, $\boldsymbol{Y}^*$, that is defined indirectly through $\mathbf{B}$ and the choice of $\boldsymbol{\Sigma}_\epsilon$. In this model, the n autoregressions in the model occur simultaneously at each data location. In the case that autoregression does not exist then $b_{ij} = 0$, which reduces the model (2) to a generalized linear model with uncorrelated errors.

To better understand the spatial dependence and autocorrelation, the SAR model is commonly written as

$$\begin{aligned} \boldsymbol{Y}^* &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W} \\ \boldsymbol{W} &= \rho\mathbf{C}\boldsymbol{W} + \boldsymbol{\epsilon} \end{aligned} \tag{6}$$

where $\mathbf{B} = \rho\mathbf{C}$, ρ represents the spatial correlation, and $\mathbf{C}$ is a spatial proximity matrix. For example, if the spatial locations are on a regular lattice, $\mathbf{C}$ is a matrix of 1's and 0's based upon a specified neighborhood structure i.e. queen neighborhood structure (the four nearest neighbors to the left, right, upper and lower sites). If however, the spatial locations are on an irregular grid, $\mathbf{C}$ can be defined by shared borders, centroid distances or other measures of distance.

From (6) the model can be rewritten as

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{C})^{-1}\boldsymbol{\epsilon} \tag{7}$$

$$= \boldsymbol{X}\boldsymbol{\beta} - \rho\mathbf{C}\boldsymbol{X}\boldsymbol{\beta} + \rho\mathbf{C}\boldsymbol{Y}^* + \boldsymbol{\epsilon}. \tag{8}$$

From (7) the autoregression in the linear model induces spatial autocorrelation through the term $(\mathbf{I} - \rho\mathbf{C})^{-1}\boldsymbol{\epsilon}$. From (8) the two additional terms, $\rho\mathbf{C}\boldsymbol{X}\boldsymbol{\beta}$ and

$\rho\mathbf{C}\boldsymbol{Y}^*$, are called the spatially lagged variables. For the SAR model to be well defined, the matrix $(\mathbf{I} - \rho\mathbf{C})$ must be non-singular.

Conditional Autoregressive (CAR) Model

The conditional autoregressive model is defined by a set of conditional probability distributions for each response, Y_i, given the observed values of all the other observations. It models the function, $f(Y_i|\boldsymbol{Y}_{-i})$, where $\boldsymbol{Y}_{-i}$ denotes the vector of all responses except the response associated with site i, and continues in this manner for each response in turn rather than simultaneously.

The CAR model depends only on the neighboring sites, j, of site i i.e. Y_i depends on Y_j only if location j is in the neighborhood set, $\mathcal{N}_i$, of i. The CAR model uses the data, $\boldsymbol{Y}$, to induce the distribution of the error terms rather than letting $\boldsymbol{W}$ induce a distribution on $\boldsymbol{Y}$ as follows. Let

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}, \tag{9}$$

or

$$(\mathbf{I} - \mathbf{B})\boldsymbol{Y}^* = \boldsymbol{W}, \tag{10}$$

where $\boldsymbol{Y}^*$ denotes the link function $g(E(Y|\boldsymbol{\theta}, \mathbf{W}))$ with respect to count, binary, or continuous data, and $\mathbf{B}$ is defined as in the SAR model, a matrix of spatial dependence parameters with $b_{ii} = 0$. For a binary response, $\boldsymbol{Y}$, the link function g is the logit function, $\text{logit}(p) = \log\{p/(1-p)\}$.

The variance-covariance matrix, $\boldsymbol{\Sigma}_{CAR}$ of $\boldsymbol{Y^*}$ is defined as

$$\begin{aligned} \boldsymbol{\Sigma}_{CAR} &= \text{Var}[\boldsymbol{Y^*}] \\ &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}_W \end{aligned} \tag{11}$$

where $\boldsymbol{\Sigma}_W = \text{diag}[\sigma_1^2, \cdots, \sigma_n^2]$. Assuming that $(\mathbf{I} - \mathbf{B})^{-1}$ exists the distribution of $\boldsymbol{Y^*}$ for the CAR model is

$$\boldsymbol{Y}^* \sim N(\boldsymbol{X\beta}, (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}_W). \tag{12}$$

The CAR model depends on the mean and the variance with respect to the neighborhood set, whereas the SAR model simultaneously describes the autoregression that occurs at each data location. The SAR model is commonly used in the econometrics literature with respect to continuous data (Lee, 2007; Lee & Liu, 2010), and binary data (Klier & McMillen, 2008; Pinkse & Slade, 1998) using generalized method of moments or generalized least squares estimation. The CAR model, on the other hand, is generally the preferred model due to it dependence on the mean and variance of each site and is easier to specify (Cressie, 1993; Cressie & Wikle, 2011; Gelfand, Diggle, Guttorp & Fuentes, 2010).

While the use of CAR and SAR models for SGLMM is very flexible and widely used, it has two major shortcomings: (1) variance inflation due to spatial confounding of the fixed and random effects (Clayton et al. 1993; Reich et al. 2006), and (2) computational challenges in high dimensional latent variables due to the CAR and SAR models. Although methods have been developed to help resolve these problems such as reducing the dimension of the spatial random effects, the

results comparing centered automodels to the new methods were shown to be comparable in computation time and parameter estimates Hughes, Haran & Caragea (2011). The benefit of using the automodel is that it does not impose a Gaussian Markov random field structure for the spatial random effects but rather uses the nature of the neighbors to create a non-distributional Markov random field. In this dissertation a type of automodel for binary data called the autologistic model will be used.

Autologistic Regression Models

In 1974, Besag proposed the automodels that are constructed from distributions of the exponential family i.e. Poisson, Normal, Bernoulli, and Binomial distributions called the autopoisson, autonormal, autologistic, and autobinomial, respectively. In particular, the autologistic model is used for spatial binary data and has applications in many fields such as ecology (Huffer & Wu, 1998), health, and geography. The autologistic model is constructed from a Markov random field used to analyze the spatial dependencies in the data from a conditional distribution dependent upon each individual site on a lattice. It is a conditional probability model that describes the probability distribution of a random variable conditioned on the neighboring sites. A joint distribution with respect to all the sites is constructed from the Hammersley-Clifford Theorem. The idea of the automodels is to capture the spatial dependence among random variables directly from the response distribution rather than hierarchically as in the SGLMM. As Markov random fields,

Hammersley-Clifford Theorem and Automodels are vital components to the construction of the autologistic regression model an overview of the methods is provided.

Markov Chain

A Markov Chain, constructed by A.A. Markov in 1907, is a type of chance process that transitions from one state to another from a discrete, finite or countable, number of possible states denoted as $S = s_1, \cdots s_t$. The process starts in one of these states and moves from one state to another, called a step. At each step, the chain retains the information of the states, or moments in time, and maps the steps of each of the states. If the chain is currently in state s_i, it randomly moves to state s_j at the next step based upon a probability p_{ij}. The conditional probability distribution of the current state only depends on the current state, and not on the sequence of states that the chain was in before the current state. In other words, the future is conditionally independent of the past given the present state of the process.

Markov chains are used for temporal and spatial data when there is dependence among the sites where $Z_i(t)$ denotes a discrete random variable corresponding to site i at time t. If the Z's are independent then dependence on the previous responses does not exist. Markov chains with respect to the temporal part of the sites, $t = 1, \cdots, T$ can be defined based upon conditional or joint probabilities as follows.

Suppose that $Z(t)$ is a random process for the response at time t and that the data are observed at $t = 0, 1, \cdots T$ times at a spatial location in one dimension space. A Markov chain based upon conditional probabilities for discrete chains Z is defined by

$$P(Z(t)|Z(1), \cdots Z(t-1)) = P(Z(t)|Z(t-1)) \tag{13}$$

Equivalently the joint probabilities are characterized by

$$P(Z(2), \cdots, Z(t)|Z(1)) = \prod_{i=2}^{t} P_i(Z(t)|Z(t-1)) \forall\ i \geq 2 \tag{14}$$

In (14), P_t is a function that only depends upon $Z(t)$ and $Z(t-1)$. To show that (13) is equivalent to (14) is fairly straightforward. The conditional probability approach is the most common method for defining spatial dependence between sites and will be used henceforth.

Markov Random Fields

Markov random field theory is a branch of probability theory that provides a convenient and consistent way to analyze spatial dependencies on a lattice. It is an extension of a Markov chain with the additional property that it can be used with n-dimensional spatial data. The Markov random field uses the Markov property, a memoryless property of a stochastic process, for a set of random variables. Its properties provide a way to model context-dependent entities such as correlated features due to the spatial dependence. The field exists on an s-dimensional random process defined on a discrete lattice, $\mathcal{L}$. Markov random field properties assume that Z_s is

a Markov chain, which is a finite set, such that the conditional distribution of Z_s depends only on the neighbors at s sites. In other words, Markov random fields do not rely on a complete ordering of the spatial locations, but it is based upon which are neighbors of a particular location (Besag, 1974; Cressie, 1993).

The following describes the properties of a Markov random field. Let $Z = \{Z_1, \cdots Z_s\}$ be a set of observed response variables located on the spatial sites, $\mathcal{S}$. A realization for each of the observations, Z_i, has a value z_i located on the lattice, $\mathcal{L}$. The set of observations, Z, represent a random field within the lattice, where the probability that a random value Z_i takes a value z_i is denoted as $P(Z_i = z_i)$. The set Z is a Markov random field on $\mathcal{S}$ with respect to a neighborhood structure $\mathcal{N}$ if and only if the following two conditions are satisfied.

(i) *Positivity Condition* (Hammersley & Clifford, 1971)

Let $\{\boldsymbol{s}_i : i = 1, \cdots, n\}$ be a set of locations on a lattice where a discrete variable $\boldsymbol{Z}$ is observed. Define the support of the probability distribution of $\boldsymbol{Z}$ to be $\Omega \equiv \{\boldsymbol{z} : P(\boldsymbol{z}) > 0\}$ where $\Omega \subset \mathcal{L}$ and $\boldsymbol{z} \equiv (z(\boldsymbol{s}_1), \cdots, z(\boldsymbol{s}_n))$. Let Ω_i be defined as $\Omega_i \equiv [\boldsymbol{z}(\boldsymbol{s_i}) : P(\boldsymbol{z}(\boldsymbol{s}_i)) > 0]$.

(ii) *Consistency of Conditional Probabilities*

The conditional probabilities at each site, $P(z(\boldsymbol{s}_i)|z(\boldsymbol{s}_j), j \neq i))$ can be calculated from the joint probabilities, $P(\boldsymbol{z})$. The joint and the conditional probabilities can be related by Besag's Lemma (Besag, 1974) otherwise known as the factorization theorem, which is stated as follows. Suppose that the vari-

ables $Z(\boldsymbol{s}_i), i = 1, \cdots, n$ have a joint probability mass function $P(\boldsymbol{z})$ which satisfies the positivity condition for any two realizations $\boldsymbol{z}$ and $\boldsymbol{y}$ of $\boldsymbol{Z}$. The joint probabilities can be related to the conditional probabilities as

$$\frac{P(\boldsymbol{z})}{P(\boldsymbol{y})} = \prod_{i=1}^{n} \frac{P(z(\boldsymbol{s}_i)|z(\boldsymbol{s}_1), \cdots z(\boldsymbol{s}_{i-1}), y(\boldsymbol{s}_{i+1}), \cdots, y(\boldsymbol{s}_n))}{P(y(\boldsymbol{s}_i)|z(\boldsymbol{s}_1), \cdots z(\boldsymbol{s}_{i-1}), y(\boldsymbol{s}_{i+1}), \cdots, y(\boldsymbol{s}_n))} \tag{15}$$

for $\boldsymbol{z}, \boldsymbol{y} \in \Omega$. For example, if $n = 2$ then

$$\frac{P(\boldsymbol{z})}{P(\boldsymbol{y})} = \frac{P(z(\boldsymbol{s}_1)|z(\boldsymbol{s}_1), y(\boldsymbol{s}_2))}{P(y(\boldsymbol{s}_1)|z(\boldsymbol{s}_1), y(\boldsymbol{s}_2))} \cdot \frac{P(z(\boldsymbol{s}_2)|z(\boldsymbol{s}_2), z(\boldsymbol{s}_1), y(\boldsymbol{s}_3), y(\boldsymbol{s}_2))}{P(y(\boldsymbol{s}_2)|z(\boldsymbol{s}_1), y(\boldsymbol{s}_3), y(\boldsymbol{s}_2))}$$

The positivity condition ensures that each term in the denominator is non-zero. When both of the above conditions hold then a Markov random field exists. In the case that the positivity condition holds, the joint probability, $P(z(\boldsymbol{s}))$, of any random field is uniquely determined by its local conditional probabilities (Besag, 1974). The second condition states that in Markov random fields only the neighboring sites have direct interactions with each other, otherwise defined as a clique, which are central to the representation of $\boldsymbol{Z}$.

A clique is defined to be a set of sites that consist either of a single site or of sites that are neighbors of each other. For example, if $n = 3$, one possible set of cliques from the sites, s_1, s_2, s_3 have the neighbors, $s_1 \in N(s_2)$, $s_1 \in N(s_3)$, but $s_2 \notin N(s_3)$. Therefore, the cliques are $\{s_1\}, \{s_2\}, \{s_1, s_2\}, \{s_1, s_3\}$.

A negpotential function is used to construct a consistent conditional probability. The negpotential function utilizes the relationship of (15) when $\boldsymbol{y} \equiv \boldsymbol{0}$. The negpotential function (Besag, 1974; Cressie, 1993) is defined as follows. Without loss of generality, assume that zero can occur at each site, $\boldsymbol{0} \in \Omega$. This ensures that

under the positivity condition a realization of zeros may exist. The negpotential function,Q , for $\mathbf{z} \in \Omega$ is then defined as follows:

$$Q(\mathbf{z}) \equiv \ln\left(\frac{P(\mathbf{z})}{P(\mathbf{0})}\right), \tag{16}$$

For example, if the Markov random field is Gaussian, then $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$ and the negpotential function, $Q(\mathbf{z})$, is

$$Q(\mathbf{z}) = -\frac{1}{2}\mathbf{z}'\mathbf{\Sigma}\mathbf{z}, \tag{17}$$

where , $\mathbf{z} \in \mathcal{R}^n$. The negpotential function can be rearranged in terms of its probabilities and in the discrete case is represented as

$$P(\mathbf{z}) = \frac{\exp\left(Q(\mathbf{z})\right)}{\sum_{\mathbf{w}\in\Omega}\exp(Q(\mathbf{w}))}, \tag{18}$$

whereas for the continuous case the summation is replaced by an integral. This property is important as the numerator in (18) can be defined based upon a probability distribution. However, the denominator represents a normalizing constant that must be accounted for in estimation. Further explanation of the negpotential function and how it pertains to the autologistic models will discussed in the traditional spatial-temporal autologistic and centered spatial-temporal autologistic model sections.

The negpotential function yields two important properties (Cressie, 1993)

(i) Given any $\{\boldsymbol{z} \in P(\boldsymbol{z}) > 0\}$, the realization $\boldsymbol{z}(s_i)$ can be written as

$(z_1, \cdots, z_{i-1}, 0, z_{i+1}, \cdots, z_n)$. Therefore, a ratio of probabilities from the factorization theorem will be

$$\exp\{Q(\boldsymbol{z}) - Q(\boldsymbol{z}_i)\} = \frac{P(\boldsymbol{z})}{P(\boldsymbol{z}_i)} = \frac{P(\boldsymbol{z}(\boldsymbol{s_i})|\boldsymbol{z}(\boldsymbol{s_j}), j \neq i)}{P(\boldsymbol{0}|\boldsymbol{z}(\boldsymbol{s_j}), j \neq i)}. \tag{19}$$

(ii) The negpotential function Q can be expanded uniquely on $\{\boldsymbol{z} : P(\boldsymbol{z}) > 0\}$ and is of the form

$$Q(\boldsymbol{z}) = \sum_{1 \leq i \leq n} z(\boldsymbol{s_i}) G_i(z(\boldsymbol{s}_i)) + G_{ij} + G_{ijk} + \cdots + G_{1\cdots n} \tag{20}$$

where

$$G_{ij} = \sum \sum_{1 \leq i < j \leq n} z(\boldsymbol{s_i}) z(\boldsymbol{s_j}) G_{ij}(z(\boldsymbol{s_i}), z(\boldsymbol{s_i})), \tag{21}$$

$$G_{ijk} = \sum \sum_{1 \leq i < j < k \leq n} \sum z(\boldsymbol{s_i}) z(\boldsymbol{s_j}) z(\boldsymbol{s_k}) G_{ijk}(z(\boldsymbol{s_i}), z(\boldsymbol{s_i}), z(\boldsymbol{s_k})), \tag{22}$$

$$\vdots$$

$$G_{1,\cdots,n} = z(\boldsymbol{s_i}) \cdots z(\boldsymbol{s_n}) G_{1\cdots n}(z(\boldsymbol{s_i}), \cdots, z(\boldsymbol{s_n})) \tag{23}$$

All arguments in which $Q(\mathbf{z})$ is evaluated have the support Ω. For example, $z(\boldsymbol{s_i}) G_i(z(\boldsymbol{s}_i)) \equiv Q(0, \cdots, 0, z(\mathbf{s}_i), 0, \cdots 0)$ and

$$z(\boldsymbol{s_i}) z(\boldsymbol{s_j}) G_{ij}(z(\boldsymbol{s_i}), z(\boldsymbol{s_i})) \equiv Q(\mathbf{s}_i, \mathbf{s}_j) - Q(\mathbf{s}_i) - Q(\mathbf{s}_j)$$

where

$$Q(\mathbf{s}_i, \mathbf{s}_j) = Q(0, \cdots, 0, z(\mathbf{s}_i), 0, \cdots, 0, z(\mathbf{s}_j), 0, \cdots, 0)$$

$$Q(\mathbf{s}_i) = Q(0, \cdots, 0, z(\mathbf{s}_i), 0, \cdots, 0, 0, 0, \cdots, 0)$$

$$Q(\mathbf{s}_j) = Q(0, \cdots, 0, 0, 0, \cdots, 0, z(\mathbf{s}_j), 0, \cdots, 0).$$

The higher order terms are equivalent to (20).

The expansion of the negpotential functions, (20), is unique but the functions $\boldsymbol{G}$ in the equation are not uniquely specified. Define $\boldsymbol{G}(z(\boldsymbol{s}_i)z(\boldsymbol{s}_j), \cdots) \equiv 0$ when $z(\boldsymbol{s}_i) = 0, z(\boldsymbol{s}_j) = 0, \cdots, z(\boldsymbol{s}_n) = 0$ then the negpotential function, Q, remains unique. The properties (i) and (ii) of the negpotential function infer that the expansion for $Q(\boldsymbol{z})$ consists of conditional probabilities. The G functions, however general, are important as they do not involve any neighborhood information. Instead, the Hammersley-Clifford Theorem presented in the next section contains the neighborhood structure and states that depending upon the cliques of the Markov Random Field many of the G functions will be equal to zero.

Hammersley-Clifford Theorem

Suppose that $\boldsymbol{Z}$ is a random process distributed according to a Markov random field with neighborhood structure $\{N(\mathbf{s}_i), i = 1, \cdots n\}$ on $\boldsymbol{\Omega}$ that satisfies the positivity condition. Then, the negpotential function, Q, given by (18), must satisfy the following property

$$\text{If sites } i, j, \cdots, n \text{ do not form a clique, then } G_{ij,\cdots n} \equiv 0$$

The result of the Hammersley-Clifford Theorem's restriction of the G functions being defined only on a clique, means that the G functions only depend upon the size of the clique (Cressie, 1993). The theorem is important in modeling as the conditional specification will only contain a few nonzero G functions.

The Hammersley-Clifford Theorem also results in an expansion of the G-functions in terms of the conditional probabilities. Cressie (1993) shows that for $i = 1, \cdots, n$,

$$\begin{aligned} z(\boldsymbol{s}_i)G_i(z(\boldsymbol{s}_i)) &= \ln\left[\frac{P(z(\boldsymbol{s}_i)|z(\boldsymbol{s}_j) = 0 : j \neq i)}{P(0(\boldsymbol{s}_i)|z(\boldsymbol{s}_j) = 0 : j \neq i)}\right] \\ &= \ln\left[\frac{f_i(Z_i^*|\boldsymbol{Z}_i^*)}{f_i(Z_i|\boldsymbol{Z}_i^*)}\right]. \end{aligned} \tag{24}$$

The pairwise interaction term of sites, $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ can also be defined as

$$\begin{aligned} &z(\boldsymbol{s}_i)z(\boldsymbol{s}_j)G_{ij}(z(\boldsymbol{s}_i), z(\boldsymbol{s}_j)) \\ &\quad = \ln\left(\frac{P(z(\boldsymbol{s}_i)|z(\boldsymbol{s}_j), 0(\boldsymbol{s}_k) : k \neq i, j)}{P(z(\boldsymbol{s}_i) = 0|z(\boldsymbol{s}_j), z(\boldsymbol{s}_k) = 0 : k \neq i, j)}\frac{P(z(\boldsymbol{s}_i) = 0|z(\boldsymbol{s}_k) : k \neq i)}{P(z(\boldsymbol{s}_i)|z(\boldsymbol{s}_k) = 0 : k \neq i)}\right) \\ &\quad = \ln\left(\frac{f_i(Z_i|Z_j, \boldsymbol{Z}_{i,j}^*)f_i(Z_i^*|\boldsymbol{Z}_i^*)}{f_i(Z_i^*|Z_j, \boldsymbol{Z}_{i,j}^*)f_i(Z_i|\boldsymbol{Z}_i^*)}\right). \end{aligned} \tag{25}$$

The probability density function $f(Z)$ is analogous to the probability form, $P(z(\boldsymbol{s}_i))$, specified in the pairwise interactions, where Z_i^* denotes $z(\mathbf{s}_i) = 0$.

The consistency condition from the Markov random field on the conditional probabilities can be expressed as conditions such that the G functions are well defined. The pairwise interaction terms from the negpotential function are invariant to whether the conditional probabilities from the i's or the j's site's exist in (24). The theorem specifies the most general form in which $Q(z)$ can take to provide a valid probability structure.

Automodels

The Hammersley-Clifford, Negpotential functions, Q, and G-functions are all essential in the construction of a large class of models called the *auto spatial models* of Besag (1974). The auto spatial models or automodels are specified as follows. For $i = 1, \cdots, n$,

$$P(Y(\boldsymbol{s}_i)|\boldsymbol{Y}_{-i}) \equiv \exp\left[A_i(\boldsymbol{Y}_{-i})B_i(Y_(\boldsymbol{s}_i)) + C_i(Y(\boldsymbol{s}_i)) + D_i(\boldsymbol{Y}_{-i})\right], \tag{26}$$

where $A_i(\boldsymbol{Y}_{-i})$ and $D_i(\boldsymbol{Y}_{-i})$ are functions of the site i's observed neighboring values and $B_i(Y_(\boldsymbol{s}_i))$ and $C_i(Y(\boldsymbol{s}_i))$ are specified by a particular exponential family distribution. Assume (26) and pairwise-only dependence, a subset of G functions corresponding to cliques contain three or more sites are all equivalent to 0, between sites then

$$A_i(\boldsymbol{Y}_{-i}) = \alpha_i + \sum_{j=1}^{n} \theta_{ij}\{B_j(Y(\boldsymbol{s}_j)) - B_j(0)\} \tag{27}$$

for $i = 1, \cdots, n$ where $\theta_{ii} = 0, \theta_{ij} = \theta_{ji}$ and $\theta_{ik} = 0$ for $k \notin N_i$. This implies that $A_i(\boldsymbol{Y}_{-i})$ is a linear function of the parameters corresponding to site i and the spatial dependence is conveyed through θ_{ij}.

It can be seen from the negpotential functions, (20) that up to an additive constant,

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n} [\alpha_i\{B_i(y(\boldsymbol{s}_i)) - B_j(0)\} + C_i(y(\boldsymbol{s}_i))] + \sum\sum \theta_{ij}\{B_j(y(\boldsymbol{s}_i)) - B_i(0)\}\{B_j(y(\boldsymbol{s}_i)) - B_j(0)\} \tag{28}$$

the automodel given by (26) imply that the negpotential function, $Q(\cdot)$ is truncated. The neighborhoods, $N(\boldsymbol{s}_i)$ in (26) and (27) can be specified as

$$N(\boldsymbol{s}_i) = \{\boldsymbol{s}_j : \theta_{ij} \neq 0\} \tag{29}$$

therefore the θ_{ij}'s have a direct interpretation as spatial dependence parameters. To model spatial binary data, the automodel form is called the autologistic model, which will be the focus of this dissertation.

Traditional Binary Autologistic Model

The traditional binary autologistic model was proposed by Besag (1972, 1974). It models the spatial dependence among random variables directly through conditional distributions for binary data. The traditional autologistic model is defined as follows.

Let $\boldsymbol{Y}$ be a random field, where $Y_i \in \{0, 1\}$ represents the response at the i^{th} grid on a lattice for $i = 1, \cdots, I$. The full conditional distribution for the traditional autologistic model is given by

$$\log\left(\frac{P(Y_i = 1)}{P(Y_i = 0)}\right) = \boldsymbol{X}_i\boldsymbol{\beta} + \sum_{l=1}^{L}\sum_{j\neq i}\phi_{ij}^{(l)}Y_j, \tag{30}$$

where $\boldsymbol{X}_i$ is the i^{th} row of the design matrix, $\boldsymbol{\beta}$ are the regression parameters, l is the order of neighborhoods with respect to the i^{th} site and $\boldsymbol{\phi} = \phi_{ij}$ are the spatial dependence parameters such that $\phi_{ij} \neq 0$ if and only if Y_i and Y_j are neighbors. The sum in (30) is the autocovariate that models the dependence between Y_i and

the remainder of the sites on the lattice, Y_{-i}. In the SGLMM the autocovariate is replaced by a latent field of random effects based upon the i^{th} element.

Let $\phi_{ij} = \phi \boldsymbol{I}_{\{i \sim j\}}$ where $\boldsymbol{I}_{(.)}$ represents the indicator function and $\sim$ denotes the neighbor relation such that $\phi > 0$. If only pairwise dependencies are assumed then the underlying graph has a clique number 2 (Cressie, 1993). In this instance, by the Hammersley-Clifford Theorem, the joint distribution of $\boldsymbol{Y}$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$ is

$$P(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{\exp(Q(\boldsymbol{Y}|\boldsymbol{\theta}))}{\sum_{Z \in \Omega} \exp(Q(\boldsymbol{Z}|\boldsymbol{\theta}))}, \tag{31}$$

where the sample space, Ω, is $\{0,1\}^I$ for a lattice with I points. Therefore the joint distribution is

$$P(Y|\boldsymbol{\theta}) = c(\boldsymbol{\theta})^{-1} \exp\left(\sum_i Y_i \boldsymbol{X}_i \boldsymbol{\beta} + \frac{\phi}{2} \sum_{i,j} \mathbf{1}_{\{i \sim j\}} Y_i Y_j \right) \tag{32}$$

$$= c(\boldsymbol{\theta})^{-1} \exp\left(\boldsymbol{Y'X\beta} + \frac{\phi}{2} \boldsymbol{Y'AY} \right), \tag{33}$$

where $\boldsymbol{A}$ is an $I \times I$ adjacency matrix, $\boldsymbol{A}_{ij} = \boldsymbol{I}_{\{i \sim j\}}$ and $c(\boldsymbol{\theta})^{-1}$ is a normalizing constant that may be intractable (Brook, 1964; Cressie, 1993).

Traditional Spatial-Temporal Binary Autologistic Model

The binary spatial-temporal autologistic model is used when there are dichotomous responses. It models the binary data on a spatial lattice repeatedly over time while accounting for the spatial and temporal dependence simultaneously (Zhu et al., 2005).

The spatial-temporal binary autologistic model proposed by Zhu et al. (2005) adds an additional temporal component to the traditional autologistic regression model. It models the response variable, Y_{it}, at times $1, \cdots, t$, using a conditional distribution of $\boldsymbol{Y_t}$ that depends on the previous S times, $t-1, \cdots, t-S$. For each time, t, it is assumed that the response variable follows a Markov random field under a specified spatial neighborhood. A logistic regression model is used, where Y_{it} is assumed to be Bernoulli due to the binary response. The conditional probability of success is denoted as

$$p_{it} = P(Y_{it} = 1 | Y_{jt} : J \in \mathcal{N}_i, Y_{it'} : t' = t-1, \cdots, t-S)$$

for $t = S+1, \cdots, T$. The probabilities, p_{it}, are modeled using a logit link and the following systematic component

$$\eta_{it} = \sum_{k=0}^{K} \beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} Y_{jt}\right] + \sum_{s=1}^{S} \gamma_s Y_{i,t-s} \tag{34}$$

$$\boldsymbol{\eta_t} = \boldsymbol{X_t\beta} + \frac{\phi}{2}\boldsymbol{AY_t} + \gamma \boldsymbol{Y_{t-s}}, \tag{35}$$

where $\boldsymbol{\eta}_t$ is the vector notation with respect to all sites, $\boldsymbol{\beta}$ are the regression coefficients, ϕ is a spatial autoregressive coefficient, γ is a temporal autoregressive coefficient, $\boldsymbol{A}$ is an $I \times I$ adjacency matrix, and $\boldsymbol{A}_{ij} = \boldsymbol{I}_{\{i \sim j\}}$. The probability, p_{it} is defined as

$$p_{it} = \frac{\exp\{\eta_{it}\}}{1 + \exp\{\eta_{it}\}}. \tag{36}$$

The joint distribution over $i = 1, \cdots I$ sites for a given time point, t based upon the Hammersley-Clifford Theorem (Cressie, 1993) is

$$p(\boldsymbol{Y}_t|\boldsymbol{Y}_t') = \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z\in\Omega}\exp(Q(\boldsymbol{Z}_t;\boldsymbol{\beta},\phi,\gamma))} \tag{37}$$

$$= c(\boldsymbol{Y}_{t-1},\cdots,\boldsymbol{Y}_{t-S}:\boldsymbol{\beta},\phi,\gamma)^{-1}\times\exp\sum_{i=1}^{I}Y_{it}\eta_{it} \tag{38}$$

$$= c(\boldsymbol{Y}_t;\boldsymbol{\theta})^{-1}\times\exp(\boldsymbol{Y_t'X_t\beta}+\frac{\phi}{2}\boldsymbol{Y_t'AY_t}+\gamma\boldsymbol{Y_t'Y_{t-s}}), \tag{39}$$

where $c(\boldsymbol{Y}_t;\boldsymbol{\theta})^{-1}$ is an unknown normalizing constant.

Centered Spatial-Temporal Binary Autologistic Regression Model

Parameter interpretation for the spatial-temporal autologistic model is not straightforward when regression coefficients are incorporated in the model. In the presence of positive spatial and temporal dependence the conditional expectation of Y_{it} given its neighbors for a non-centered model is,

$$E(Y_{it}|Y_{i't'}:(i',t'\in\mathcal{N}_{it})) = \frac{\exp\{\eta_{it}\}}{1+\exp\{\eta_{it}\}} \tag{40}$$

which is an increasing function based upon the standard logit,

$$\frac{\exp\{\boldsymbol{X\beta}\}}{1+\exp\{\boldsymbol{X\beta}\}} \tag{41}$$

where η_{it} is specified as (34) in the spatial temporal autologistic model.

The expectation of Y_{it} under independence with non-zero spatial or temporal neighbors can never decrease. However, if most of the neighbors with respect to a particular site, i, are 0 then this assumption is unreasonable as it biases the realizations on the spatial field toward 1. This means that the odds of $Y_i = 1$ in the non-centered model relative to the independence model, standard logit model, increases

for any nonzero neighbors, and can never decrease. The minimum value of the odds ratio for the traditional model is one, which is realized only if all neighboring observations are zero. This can only be true at all locations for a spatial field with constant value zero. This is due to the non-negative autocovariates that cause spatial confounding when the spatial and/or temporal dependence is large. Therefore, if there are varying levels of spatial dependence the interpretation of the parameters is difficult.

Wang & Zheng (2013) proposed a centered spatial-temporal autologistic model, which centers the parameters in the model to help alleviate this problem with parameter interpretation. The idea of a centered parameterization for non-Gaussian Markov random field models was proposed by Kaiser & Cressie (1997), who considered a Winsorized Poisson conditional model. Kaiser & Caragea (2009) later investigated centered parameterization for Markov random field models from an exponential family. A centered parameterization was developed for the spatial autologistic regression model, which was shown to overcome the interpretation difficulties that the uncentered model has, hence it is the preferred model.

The centered spatial-temporal autologistic regression model has the same properties of the traditional spatial-temporal model as it models the data repeatedly over time on a spatial lattice while accounting for covariates and the spatial and temporal dependence simultaneously (Zhu et al., 2005). It is altered from the non-centered model by centering the response in the model as follows.

Under a regularity condition of pairwise-only dependence, a centered spatial-temporal autologistic regression model is defined by the full conditional Bernoulli distribution where p_{it} is modeled by a logit function

$$p(Y_{it}|Y_{i't'} : (i', t' \neq (i, t))) = \frac{\exp\{Y_{it}\eta_{it}\}}{1 + \exp\{Y_{it}\eta_{it}\}} \tag{42}$$

with systematic component η_{it}

$$\eta_{it} = \sum_{k=0}^{K} \beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} Y^*_{jt}\right] + \sum_{s=1}^{S} \gamma_s Y^*_{i,t-s} \tag{43}$$

The systematic component can also be written in vector notation with respect to all sites as

$$\boldsymbol{\eta}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \frac{\phi}{2}\boldsymbol{A}\boldsymbol{Y}^*_{\boldsymbol{t}} + \gamma\boldsymbol{Y}^*_{t-s} \tag{44}$$

where $\boldsymbol{\beta}$ are the regression coefficients, ϕ is a spatial autoregressive coefficient, γ is a temporal autoregressive coefficient, $\boldsymbol{A}$ is an $I \times I$ adjacency matrix, $\boldsymbol{A}_{ij} = \boldsymbol{I}_{\{i\sim j\}}$ and Y^*_{it} denotes the centered response for the i^{th} site and the t^{th} time point,

$$Y^*_{it} = Y_{it} - p_{it}. \tag{45}$$

The center p_{it} is the probability of $Y_{it} = 1$ under independence,

$$p_{it} = \frac{\exp\{x'_{it}\boldsymbol{\beta}\}}{1 + \exp\{x'_{it}\boldsymbol{\beta}\}}. \tag{46}$$

The conditional expectation of Y_{it} given its neighbors and covariates is,

$$E(Y_{it}|Y_{i't'} : (i', t') \in N_{it}) = \frac{\exp\{\eta_{it}\}}{1 + \exp\{\eta_{it}\}} \tag{47}$$

Suppose the spatial and temporal coefficients are positive. If the conditional expectation of the centered model is compared to the independence model then

$$E(Y_{it}|Y_{i't'} : (i', t') \in N_{it}) > p_{it} \tag{48}$$

when

$$\frac{\phi}{2}\boldsymbol{A}\boldsymbol{Y}_{\boldsymbol{t}}^{*} + \gamma \boldsymbol{Y}_{t-s}^{*} > \frac{\phi}{2}\boldsymbol{A}\boldsymbol{p}_{\boldsymbol{t}}^{*} + \gamma \boldsymbol{p}_{t-s}^{*}, \tag{49}$$

where $\boldsymbol{p}_t$ and $\boldsymbol{p}_{t-s}$ are the expected number of non-zero spatial and temporal neighbors under the independence model (46). If the spatial parameter is equal to zero, then the observed number of non-zero spatial neighbors is greater than the expected number of non-zero neighbors under independence, i.e. $\boldsymbol{Y}_t > \boldsymbol{p}_t$. Similarly, if the temporal value is equal to zero then the conditional expectation of $\boldsymbol{Y}_t$ increases over $\boldsymbol{p}_t$ only when $Y_{i,t-s} > Y_{i,t-s}$. In other words, the observed number of non-zero temporal neighbors is greater than the expected number of non-zero temporal neighbors under independence.

The joint distribution of the full spatial-temporal process, Y_{it}, of the centered model is specified by the Hammersley-Clifford Theorem (Cressie, 1993). If the conditional distribution (42) is well defined, the joint likelihood is

$$\boldsymbol{L}(\boldsymbol{\theta}) = p(\boldsymbol{Y}_t|\boldsymbol{Y}_t') = c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\beta}, \phi, \gamma)^{-1} \times \exp \sum_{i=1}^{I} Y_{it}^{*}\eta_{it} \tag{50}$$

$$= c(\boldsymbol{Y}_t; \boldsymbol{\theta})^{-1} \times \exp(\boldsymbol{Y}_{\boldsymbol{t}}'\boldsymbol{X}_t\boldsymbol{\beta} + \frac{\phi}{2}\boldsymbol{Y}_{\boldsymbol{t}}'\boldsymbol{A}\boldsymbol{Y}_{\boldsymbol{t}}^{*} + \gamma \boldsymbol{Y}_{\boldsymbol{t}}'\boldsymbol{Y}_{t-s}^{*}). \tag{51}$$

This model is similar to the uncentered parameterization, where $c(\boldsymbol{Y}_t; \boldsymbol{\theta})^{-1}$ is an unknown normalizing constant.

Statistical Inference for Binary Spatial Models

Inference for spatial or spatial-temporal models depends upon the type of spatial or spatial-temporal model. The spatial generalized linear mixed model is considered a marginal model as it represents the spatial term as a random component. The autologistic model, on the other hand, is a conditional model where the spatial component in the model is conditional on a Markov random field causing estimation to be different than SGLMM. The autologistic regression model has a normalizing constant that does not have a closed form, hence direct maximization of the likelihood function is not straightforward.

Statistical inference for the spatial generalized linear mixed model (SGLMM) is based on the idea that the spatial term is modeled as a random effect. The most common types of estimation for the SGLMM model include quasi-likelihood (QL), penalized quasi-likelihood estimating equations and Bayesian hierarchical models (Gotway & Stroup, 1997; Schabenberger & Gotway, 2004). Lin et al. (2009) proposed a marginal model using quasi-likelihood (QL) estimating equations for statistical inference for a non-separable spatial-temporal binary data, which allows separate modeling of regression and dependence of the response variables. In 2010, Lin extended his work on QL estimating equations for separable spatial-temporal binary data. Generalized method of moments estimation has also been used for spatial models with respect to the spatial autoregressive model for both continuous and binary data (Lee, 2007; Lee & Liu, 2010; Pinkse & Slade, 1998; Klier

& McMillen, 2008). In this section a brief review of maximum pseudo-likelihood, MCML, and Bayesian inference for the autologistic model and GMM for the SGLMM are presented.

Monte Carlo Inference for the Autologistic Model

Inference for the traditional autologistic model presented in the autologistic model section is conducted by pseudolikelihood, maximum likelihood (ML) and Bayesian inference following Besag (1975); Geyer (1994); Møller, Pettitt, Reeves & Berthelsen (2006) respectfully. In addition, Zheng & Zhu (2008) explored these methods in context for the traditional spatial-temporal model, Hughes et al. (2011) for the centered autologistic model and Wang & Zheng (2013) compared the centered spatial-temporal model to PL, Bayesian and expectation maximum likelihood (EML) methods. Sampling methods such as Gibbs samplers and Perfect samplers are common methods used for Monte Carlo sampling. The methods are used to estimate of the unknown normalizing constants that occur in the joint distribution of the autologistic models. In this dissertation, the focus is on pseudolikelihood, Monte Carlo maximum likelihood, and Bayesian methods using perfect sampling and Gibbs sampling for the autologistic models that are specified in Appendix C.

Maximum Pseudolikelihood (MPL)

Maximum pseudolikelihood was proposed for the autologistic model by Besag (1974). The estimate, $\tilde{\boldsymbol{\theta}}$, of the parameter vector is the value that maximizes the product of conditional likelihoods,

$$\tilde{\boldsymbol{\theta}} = \text{argmax } \ell_{PL}(\boldsymbol{\theta}),$$

where

$$\ell_{PL}(\boldsymbol{\theta}) = \sum_i \log \frac{\exp(Y_i(\eta_{it}))}{1 + \exp(Y_i \eta_{it})}, \tag{52}$$

and η_{it} is specified by an autologistic model from the autologistic model section.

Maximum pseudo-likelihood does not have the same value as the true likelihood estimates, however, Besag (1975) showed that the maximum pseudo-likelihood estimators converge to the maximum likelihood estimators as the lattice size tends to ∞.

The pseudolikelihood function can be maximized to obtain the maximum pseudolikelihood estimator (MPLE) of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \gamma)'$ using standard statistical software packages such as *glm* in R or *proc logistic* in SAS. The standard errors and approximate confidence intervals can be estimated by a parametric bootstrap as follows.

Step 1: Find the joint conditional probability distribution, $P(Y_i = 1|\boldsymbol{Y}_{-i}, \boldsymbol{\theta})$ from the autologistic model.

Step 2: Compute the MPLE for the sample.

A parametric bootstrap can be used to compute the standard error of the MPL estimates that proceeds as follows. Generate M random samples from the joint probability distribution from one of the autologistic models from the likelihood function evaluated at MPL estimates using a perfect simulation sample using a Gibbs sampler. Compute the MPL estimates for each sample, $\tilde{\boldsymbol{\theta}}^{(1)}, \cdots, \tilde{\boldsymbol{\theta}}^{(M)}$. Quantiles of the bootstrap sample are then used to construct approximate confidence intervals for the elements of $\boldsymbol{\theta}$. The perfect samples are used to generate the first Monte Carlo sample and guarantee that the MCML estimation targets a monotone autologistic regression model (Propp & Wilson, 1996; Møller, 1999).

The benefit of maximum pseudolikelihood estimation is that it is consistent and asymptotically normal (Guyon, 1995). However, when spatial and/or temporal dependence is strong it tends to be statistically inefficient (Gumpertz, Graham & Ristaino, 1997; Huffer & Wu, 1998; Zheng & Zhu, 2008; Hughes et al., 2011). In these cases alternative estimation methods are preferable.

Monte Carlo Maximum Likelihood (MCML)

In the case of strong spatial and/or temporal dependence an alternative estimation approach is Monte Carlo maximum likelihood (MCML). Estimation for the traditional autologistic model was developed by Huffer & Wu (1998) and used for the non-centered and centered spatial-temporal autologistic model (see Zhu et al., 2005; Zheng & Zhu, 2008; Zhu, Zheng, Carroll & Aukema, 2008; Wang &

Zheng, 2013). In this section the MCML estimation for the autologistic models is explained.

MCML approximates an expectation from the sample mean of a function of simulated random variables, which represent the normalizing constants. It is an iterative process using a Markov chain that randomly samples from a specific probability distribution, $P(\boldsymbol{Y}_t|\boldsymbol{\theta})$, through simulation to estimate the value of the parameters. The samples are neither independent nor identically distributed; however the probability converges in distribution to the actual distribution as if the samples are iid (Huffer & Wu, 1998). The parameter values are then found by averaging the simulated estimates. By the law of large numbers, the estimated parameters from MCML will be close to the true parameter values. Monte Carlo simulation is a way of making sure that certain values will have more impact on the parameter being estimated, by the "important" values being sampled more frequently. In this way the variance of the parameter estimators, $\hat{\boldsymbol{\theta}}$, can be reduced.

While MPL estimates and MCML estimates are both based upon approximating, $P(\boldsymbol{Y}_t|\boldsymbol{\theta})$, MPL sidesteps the intractable normalizing constant using bootstrap samples whereas MCML approximates a ratio of the true and approximate parameter values, $c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})$. Monte Carlo maximum likelihood for the autologistic models proceeds as follows. First, find the log-likelihood of the conditional distri-

bution. The likelihood function for $\boldsymbol{\theta}$ from the full conditional distribution, $P(\boldsymbol{Y}_t|\boldsymbol{\theta})$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \mathcal{L}(\boldsymbol{\theta}, \mathbf{Y}_{S+1}, \cdots \mathbf{Y}_T|\mathbf{Y}_1, \cdots \mathbf{Y}_S) \\ &= \prod_{t=S+1}^{T} \left[c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta})\right]^{-1} \times \prod_{t=S+1}^{T} \exp Y_{it}\eta_{it} \\ &= \prod_{t=S+1}^{T} \left[c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta})\right]^{-1} \times \exp \sum_{t=S+1}^{T} Y_{it}\eta_{it}.\end{aligned} \tag{53}$$

From (53), the log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = -\log\left[\sum_{i=S+1}^{T} c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta})\right] + \sum_{i=S+1}^{T} \boldsymbol{Z_t\theta}, \tag{54}$$

where

$$\boldsymbol{Z_t\theta} = \boldsymbol{Y_t'X}_t\boldsymbol{\beta} + \frac{\phi}{2}\boldsymbol{Y_t'AY_t^*} + \gamma\boldsymbol{Y_t'Y}_{t-s}^*. \tag{55}$$

The log likelihood, (54), is difficult to maximize due to the normalizing constant, $c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-S}, \boldsymbol{\theta})$, that does not have a closed form. To resolve this issue, MCML is employed to approximate a ratio of log-likelihoods,

$$\ell(\boldsymbol{\theta}) - \ell(\tilde{\boldsymbol{\theta}}) = \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})'\boldsymbol{Z}_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \tilde{\boldsymbol{\theta}})}, \tag{56}$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_0, \cdots, \tilde{\theta}_{K+L+S})'$ is a reference parameter and the samples, $Z_1, \cdots, Z_M$ are perfect samples from $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. As the difference in the log likelihood functions is approximately zero, (56) can be rewritten as

$$\sum_{t=S+1}^{T} \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \tilde{\boldsymbol{\theta}})} = \sum_{t=S+1}^{T} \frac{\exp\{\boldsymbol{\theta' Z_t}\}}{\exp\{\tilde{\boldsymbol{\theta}}\boldsymbol{' Z_t}\}}, \tag{57}$$

where the expectation of the density is

$$\frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \tilde{\boldsymbol{\theta}})} = E_{\tilde{\boldsymbol{\theta}}} \left[\frac{\exp\{\boldsymbol{\theta}' \boldsymbol{Z_t}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z_t}\}} \right] \tag{58}$$

for a time point t. The log-likelihood, (56), is maximized using Monte Carlo estimation techniques to obtain $\hat{\boldsymbol{\theta}}$, the Monte Carlo estimate of $\boldsymbol{\theta}$ from the following,

$$\frac{1}{M} \sum_{m=1}^{M} \frac{\exp\{\boldsymbol{\theta}' \boldsymbol{Z}_{\boldsymbol{t}}^{(\boldsymbol{m})}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z}_{\boldsymbol{t}}^{(\boldsymbol{m})}\}}, \tag{59}$$

where M is the number of iterations using a Monte Carlo Gibbs sampler such that $m = 1, \cdots M$, and $\boldsymbol{Z}_t^{(m)}$ is the m^{th} set of Monte Carlo samples of the response vector, $\boldsymbol{Y}_t$ and $Z_1, \cdots, Z_m$ are perfect samples from the model $\boldsymbol{\theta} = \boldsymbol{\phi}$. According to Geyer (1994), $\hat{\boldsymbol{\theta}}$ converges to the maximum likelihood estimate, $\boldsymbol{\theta}$, of $\boldsymbol{\theta}$ as $M \to \infty$. The log likelihood ratio, (56), becomes

$$\begin{aligned}
\ell(\boldsymbol{\theta}) - \ell(\tilde{\boldsymbol{\theta}}) &= \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \boldsymbol{Z}_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \tilde{\boldsymbol{\theta}})} \\
&\approx \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \boldsymbol{Z}_t - \sum_{t=S+1}^{T} \log \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\exp\{\boldsymbol{\theta}' \boldsymbol{Z}_{\boldsymbol{t}}^{(\boldsymbol{m})}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z}_{\boldsymbol{t}}^{(\boldsymbol{m})}\}} \right] \\
&\equiv \ell(\boldsymbol{\theta}; \tilde{\boldsymbol{\theta}}).
\end{aligned} \tag{60}$$

In this case, the log-likelihood, $\ell(\tilde{\boldsymbol{\theta}})$ is free of $\boldsymbol{\theta}$ so it is possible to maximize (60) with respect to $\boldsymbol{\theta}$ and obtain the maximum likelihood estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. The sampling variances are obtained by taking the diagonal elements of, $\mathcal{I}_M^{-1} = [- \bigtriangledown^2 \ell_M(\hat{\boldsymbol{\theta}}_M)]^{-1}$, which converges to the Fisher information matrix, $\mathcal{I}^{-1}$.

Generally, estimation of the reference parameter $\tilde{\boldsymbol{\theta}}$ is done by maximum pseudolikelihood (MPL) as it tends to be the natural choice. However, when MPL

is far away from the MCML estimate, the MCML estimate can be more difficult to obtain, therefore stochastic approximation algorithms such as those of Gu & Zhu (2001) may be preferable (Zhu et al., 2008).

Expectation-Maximization Pseudolikelihood (EMPL)

In 2013, Wang proposed an expectation-maximization psuedolikelihood (EMPL) in place of MPL due to MPL's lack of efficiency when spatial and/or temporal dependence is strong. It is composed of a combination of an expectation-maximization (EM) algorithm and a Newton-Raphson algorithm. The EMPL algorithm (Wang 2012) proceeds as follows.

Step 1: Initialization Preselect parameters, $\boldsymbol{\theta}_0$ and set $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_0$.

Step 2: Expectation Given $\hat{\boldsymbol{\theta}}^{(l-1)}$,

(1) Compute p_{it}^{l-1}, the expectation of Y_{it} from an independent logistic regression model.

(2) *Compute the centered response for the l^{th} iteration, $Y_{it}^{*} = Y_{it} - p_{it}^{l-1}$.

Step 3: Maximization Obtain $\hat{\boldsymbol{\theta}}^{l}$ by maximizing the log likelihood function of the conditional distribution,

$$\sum_{it} \log \left[\frac{\exp Y_{it}^{*} \eta_{it}}{1 + \exp \eta_{it}} \right], \tag{61}$$

by the Newton-Raphson algorithm.

*Only completed in the centered autologistic model.

Step 4: Convergence criteria Repeat steps 2 and 3 until $|\hat{\boldsymbol{\theta}}^{l} - \hat{\boldsymbol{\theta}}^{l-1}| < \delta$, then $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{l}$, where $\hat{\boldsymbol{\theta}}$ is the EMPL estimate of $\boldsymbol{\theta}$, and δ a preselected precision parameter.

A parametric bootstrap can be used to compute the standard error of the EMPL estimates, which proceeds as follows. Generate M random samples from the joint probability distribution from one of the autologistic models from the likelihood function evaluated at EMPL estimates using a perfect simulation sample using a Gibbs sampler. Compute the EMPL estimates for each sample, $\tilde{\boldsymbol{\theta}}^{(1)}, \cdots, \tilde{\boldsymbol{\theta}}^{(M)}$. Quantiles of the bootstrap sample are then used to construct approximate confidence intervals for the elements of $\boldsymbol{\theta}$. The perfect samples are used to generate the first Monte Carlo sample and guarantee that the MCML estimation targets a monotone autologistic regression model (Propp and Wilson 1996, Møller 1999).

A natural choice for the starting values, $\boldsymbol{\theta}_0$, is the maximum pseudolikelihood estimate from the traditional autologistic model or independent logistic regression model. The choice of the starting value can impact the convergence of the model as well as the data generation size and inherent structure of the data. If the starting value is far away from the actual value, computation time increases.

Monte Carlo Expectation-Maximization Likelihood (MCEML)

Wang & Zheng (2013) proposed a expectation-maximization pseudolikelihood (EMPL) and Monte Carlo expectation-maximization likelihood estimator

in lieu of MPL and MCMC maximum likelihood, respectively. The Monte Carlo expectation-maximization likelihood estimator was found to be comparable to Bayesian inference and MCML as the estimators were consistent and were comparable in run times. However, the EMPL was not found to be suitable for the centered model with high spatial and/or temporal dependence (Wang & Zheng, 2013). The Monte Carlo expectation-maximization likelihood estimator developed by Wang & Zheng (2013) is found in a similar manner to MCML methods as it uses perfect sampling or perfect sampling with Gibbs sampler and a reference parameter $\boldsymbol{\psi}$. The Monte Carlo expectation-maximization likelihood estimator is constructed as follows.

Let $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_0, \cdots, \tilde{\theta}_{p+2})'$ be a reference parameter for the centered spatial-temporal binary autologistic model where Z_ψ is specified similarly to the reference parameter in MCML, the centered values based upon $\tilde{\boldsymbol{\theta}}$. The likelihood function is specified as

$$c^*(\tilde{\boldsymbol{\theta}})\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \frac{c^*(\tilde{\boldsymbol{\theta}})}{c^*(\boldsymbol{\theta})} \exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\boldsymbol{\theta}}) \tag{62}$$

$$= \exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\boldsymbol{\theta}}) E_{\tilde{\boldsymbol{\theta}}} \left[\frac{\exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\boldsymbol{\theta}})}{\exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\tilde{\boldsymbol{\theta}}})} \right]^{-1} \tag{63}$$

Generating M Monte Carlo samples of $\boldsymbol{Y}$ from the likelihood function evaluated at $\boldsymbol{\psi}$ the expectation becomes

$$E_{\tilde{\boldsymbol{\theta}}} \left[\frac{\exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\boldsymbol{\theta}})}{\exp(\boldsymbol{\theta}' \boldsymbol{Z}^*_{\boldsymbol{\psi}})} \right]^{-1} \approx M^{-1} \sum_{m=1}^{M} \exp(\boldsymbol{\theta}' \boldsymbol{Z}^{*(m)}_{\boldsymbol{\theta}} - \boldsymbol{Z}^{*(m)}_{\tilde{\boldsymbol{\theta}}}) \tag{64}$$

By (62) and (64), a Monte Carlo Markov Chain approximation of the rescaled likelihood is

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta' Z^*_\theta})}{\exp(\tilde{\boldsymbol{\theta}} \boldsymbol{Z}_{\tilde{\boldsymbol{\theta}}})} M^{-1} \left[\sum_{m=1}^{M} \exp(\boldsymbol{\theta' Z^{*(m)}_\theta} - \boldsymbol{Z^{*(m)}_{\tilde{\theta}}}) \right]^{-1} \tag{65}$$

therefore, the log-likelihood ratio is

$$\ell(\boldsymbol{\theta}) - \ell(\tilde{\boldsymbol{\theta}}) = \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \boldsymbol{Z}^*_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \tilde{\boldsymbol{\theta}})}, \tag{66}$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_0, \cdots, \tilde{\theta}_{K+L+S})'$ is the reference parameter. Once again, Monte Carlo estimation techniques are used to approximate the ratio of normalizing constants,

$$\begin{aligned} \ell(\boldsymbol{\theta}) - \ell(\tilde{\boldsymbol{\theta}}) &= \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \boldsymbol{Z}^*_t - \sum_{t=S+1}^{T} \log \frac{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\theta})}{c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \tilde{\boldsymbol{\theta}})} \\ &\approx \sum_{t=S+1}^{T} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \boldsymbol{Z}^*_t - \sum_{t=S+1}^{T} \log \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\exp\{\boldsymbol{\theta' Z^{*(m)}_t}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z^{*(m)}_t}\}} \right] \\ &\equiv \ell(\boldsymbol{\theta}; \tilde{\boldsymbol{\theta}}). \end{aligned} \tag{67}$$

To estimate (68) the MCEML estimates are found by combining an EM algorithm and a Newton-Raphson algorithm, which differ from MCML estimates. The MCEML algorithm developed by Wang & Zheng (2013) is defined as follows.

Step 1: Initialization Preselect parameters, $\boldsymbol{\theta}_0$ and set $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_0$.

(1) Choose a reference parameter vector, $\tilde{\boldsymbol{\theta}}$, and generate M Monte Carlo samples of $\boldsymbol{Y}$ from the likelihood function evaluated at $\tilde{\boldsymbol{\theta}}$.

(2) Compute $p^{l-1}_{it\tilde{\theta}}$, the expectation of Y_{it} from an independent logistic regression model evaluated at $\tilde{\boldsymbol{\theta}}$.

(3) *Compute the centered response, $Y_{it}^{*} = Y_{it} - p_{it}^{l-1}$. for M Monte Carlo samples evaluated at $\tilde{\boldsymbol{\theta}}$.

Step 2: Expectation Given $\hat{\boldsymbol{\theta}}^{(l-1)}$,

(1) Compute p_{it}^{l-1}, the expectation of Y_{it} from an independent logistic regression model.

(2) *Compute the centered response for the l^{th} iteration, $Y_{it}^{*} = Y_{it} - p_{it}^{l-1}$.

(3) *Compute the centered response for M Monte Carlo samples (generated at step 1) at the l^{th} iteration, $Y_{it}^{*} = Y_{itm} - p_{it}^{l-1}$.

Step 3: Maximization Obtain $\hat{\boldsymbol{\theta}}^{l}$ by maximizing

$$\boldsymbol{\theta}' \boldsymbol{Z}^{*}_{\boldsymbol{\theta}^{(l-1)}} - \tilde{\boldsymbol{\theta}} \boldsymbol{Z}^{*}_{\tilde{\boldsymbol{\theta}}} - \log \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\exp\{\boldsymbol{\theta}' \boldsymbol{Z}^{*(m)}_{\boldsymbol{\theta}^{(l-1)}}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z}^{*(m)}_{\tilde{\boldsymbol{\theta}}}\}} \right], \tag{68}$$

evaluated at the m^{th} Monte Carlo sample of $\boldsymbol{Y}$ (generated at initialization) by the Newton-Raphson algorithm.

Step 4: Convergence criteria Repeat steps 2 and 3 until $|\hat{\boldsymbol{\theta}}^{l} - \hat{\boldsymbol{\theta}}^{l-1}| < \delta$, then $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{l}$, where $\hat{\boldsymbol{\theta}}$ is the MCEML estimate of $\boldsymbol{\theta}$, and δ a preselected precision parameter.

The standard errors can be found from the diagonal elements of the Fisher information matrix of the original data approximated by MCEML estimation.

*Only completed in the centered autologistic model.

Bayesian Inference

The intractable normalizing functions in the autologistic model initially made Bayesian analyses impossible. However, Møller et al. (2006) constructed an auxiliary-variable MCMC algorithm that makes analysis of the normalizing constant possible by contracting a proposal distribution such that the normalizing constant cancels out the Metropolis-Hastings ratio. This method only requires that independent realizations are drawn from the unnormalized density for any $\boldsymbol{\theta}$ value, which can be done by means of perfect sampling or perfect sampling with Gibbs sampling.

Bayesian inference for the spatial-temporal autologistic model was developed by Zheng & Zhu (2008). The parameter estimates are obtained using a Metropolis-Hastings (MH) algorithm to generate Monte Carlo samples of the parameter vector, $\boldsymbol{\theta}$ from the posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{Y})$ where $\boldsymbol{Y} = (\boldsymbol{Y}_1', \cdots, \boldsymbol{Y}_T')'$ denotes all the data over each site and each time. The MH algorithm based upon Zheng & Zhu (2008) proceeds as follows.

Step 1: Select a starting parameter vector, $\boldsymbol{\theta}^{(0)}$ and set $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_0$.

Step 2: Propose a parameter vector, $\boldsymbol{\theta}^*$, according to a proposal distribution, $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$,$N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}\{\sigma_0^2, \cdots, \sigma_p^2, \sigma_{k+1}^2, \sigma_{k+2}^2\}$ is a diagonal variance matrix chosen from a specified normal prior distribution where k is the number of covariates.

Step 3: Compute the MH acceptance probability

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = \min\left\{\frac{p(\boldsymbol{\theta}^*)\ell(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta})\ell(\boldsymbol{\theta})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}, 1\right\}, \tag{69}$$

where $p(\boldsymbol{\theta})$ denotes a prior distribution for $\boldsymbol{\theta}$ and $\frac{\ell(\boldsymbol{\theta}^*)}{\ell(\boldsymbol{\theta})}$ is the likelihood of normalizing constants. A proposed value α^* is accepted as the new value with probability given by the minimum of 1 or the Metropolis-Hastings ratio.

Step 4: Repeat Steps 1 to 3 M times.

At the $(m+1)^{th}$ step, let $\boldsymbol{\theta}^{m+1} = \boldsymbol{\theta}$ with probability $\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta})$. Otherwise, $\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}$. If the ratio of the new-to-old likelihoods is greater than 1, but a random uniform draw is even smaller, then accept the new values. Otherwise, reject the new values and keep the old values for each of the l iterations.

To evaluate the likelihood ratio, $\frac{\ell(\boldsymbol{\theta}^*)}{\ell(\boldsymbol{\theta})}$, a ratio of unknown normalizing constants, methods similar to MCML are used. The likelihood function can be rewritten as

$$\ell(\boldsymbol{\theta}) = c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y_{t-S}}; \boldsymbol{\theta})^{-1} \exp(\boldsymbol{\theta'Z}). \tag{70}$$

With a preselected parameter vector, $\tilde{\boldsymbol{\theta}}$, generate M Monte Carlo samples from the joint distribution, $p(\mathbf{Y}_2, \cdots, \mathbf{Y_{T-1}}|\boldsymbol{Y}_1, \boldsymbol{Y}_T; \tilde{\boldsymbol{\theta}})$ evaluated at $\tilde{\boldsymbol{\theta}}$ at the beginning of the MH algorithm. The ratio of normalizing constants can be approximated from

equation (59) in MCML. Therefore, by (59) and (70), the likelihood ratio can be approximated by,

$$\frac{\ell(\boldsymbol{\theta}^*)}{\ell(\boldsymbol{\theta})} = \frac{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{Z})}{\exp(\boldsymbol{\theta}'\boldsymbol{Z})} \times \frac{c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})}{c(\boldsymbol{\theta}^*)/c(\tilde{\boldsymbol{\theta}})} \tag{71}$$

$$\approx \exp((\boldsymbol{\theta}^* - \boldsymbol{\theta})'\boldsymbol{Z}) \frac{M^{-1}\sum_{m=1}^{M}\exp\{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})'\boldsymbol{Z}_t^{(m)}\}}{M^{-1}\sum_{m=1}^{M}\exp\{(\boldsymbol{\theta}^* - \tilde{\boldsymbol{\theta}})'\boldsymbol{Z}_t^{(m)}\}} \tag{72}$$

where $c(\boldsymbol{\theta}) = c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \boldsymbol{\theta})^{-1}, c(\tilde{\boldsymbol{\theta}}) = c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \tilde{\boldsymbol{\theta}})^{-1}$ and $c(\boldsymbol{\theta}^*) = c(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{\mathbf{t-S}}; \boldsymbol{\theta}^*)^{-1}$. The M Monte Carlo samples of $\boldsymbol{Y}$ are generated from perfect simulation and perfect sampling with Gibbs sampler.

For the MH algorithm, a good choice for the reference parameter, $\tilde{\boldsymbol{\theta}}$ is the MPL estimates. An adaptive MCMC method is used to update the mean and variance of the samples based upon a set number of iterations. The proposal distributions are adjusted such that the acceptance rate is around 20-40% as suggested by Gelman, Carlin, Stern & Rubin (2003). Once the acceptance rate is steady within the range, based upon trace plots, the initial sample are discarded, the burn-in period. Inferences are made on the rest of the samples.

A common problem with Bayesian analysis is the proposal distributions have to be found sequentially for each iteration, making it the most computationally expensive compared to MCML, MCEML. To obtain a small efficiency gain, the proposal distributions can be updated simultaneously, allowing a correlation structure to exist between the $\boldsymbol{\theta}$ parameters. For example, a normal random walk MH algorithm can be implemented so that the proposal for $\boldsymbol{\theta}$ is multivariate normal of dimension $k+2$, i.e. $\boldsymbol{\theta}^{*(m+1)}|\boldsymbol{\theta}^{*(m)} \sim \mathcal{N}(\boldsymbol{\theta}^{*(m)}, \boldsymbol{V})$. Despite the small increase in

efficiency, Bayesian inference is impractical for large lattices when using the perfect sampler (Møller et al., 2006).

Generalized Method of Moments

Generalized method of moments (GMM) estimation was developed by Hansen (1982) as a generalization of the method of moments. The idea of GMM is to construct a set of estimating equations for an unknown parameter vector, $\boldsymbol{\theta}$, by making sample moments match population moments. It is a nice alternative to maximum likelihood as it does not require full specification of the distribution of the data, but rather uses specified moments derived from an underlying model.

Generalized method of moments for spatial models have been developed using a simultaneous autoregressive model. In a continuous setting, Lee (2007); Lee & Liu (2010) proposed and derived the asymptotic properties of GMM estimation methods for a spatial autoregressive model and a mixed regressive spatial autoregressive, which allows covariates to interact with the spatial terms as an alternative for quasi-likelihood methods. Lee's construction of GMM involved the use of instrumental variables, a common method in econometrics where the instrumental variables are selected such that they are uncorrelated with the error. In practice, regression parameters are commonly chosen as instrumental variables, however, for spatial models a variation of a neighborhood matrix is commonly used (Lee, 2007; Lee & Liu, 2010).

For binary spatial models, GMM estimation was initially proposed by Pinkse & Slade (1998) as an alternative to maximum likelihood and quasi-likelihood approaches. Pinkse & Slade (1998) used the SAR model with a binary response and applied a probit transformation and Klier & McMillen (2008) applied a logit link. The GMM estimation for the model does not rely on the normality assumption, which is common with maximum likelihood based methods with respect to the SAR model. The GMM estimator for the binary spatial model is designed with spatially dependent errors:

$$g(E(Y|\boldsymbol{\theta}, \mathbf{W})) = \boldsymbol{X\beta} + \boldsymbol{W},$$
$$\boldsymbol{W} = \rho \mathbf{C}\boldsymbol{W} + \boldsymbol{\epsilon}, \tag{73}$$

where $g(E(Y|\boldsymbol{\theta}, \mathbf{W}))$ represents the link function from either a logit or probit model, $\boldsymbol{C}$ is a positive definite matrix defined based upon spatial neighbors, and $\boldsymbol{\epsilon}$ is a vector of normally distributed independent and identically distributed errors. The covariance matrix of the model is equivalent to (4) presented in the simultaneous autoregressive (SAR) model section.

The model structure implies autocorrelation for $\boldsymbol{\epsilon}$ unless the spatial correlation is 0, $\rho = 0$. The parameter vector, $\boldsymbol{\theta}$, is estimated by GMM where the generalized error residuals are $u_i = Y_i - P_i$, where

$$P_i = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

The SAR model with spatial lagged variables defines $\boldsymbol{W} = (\mathbf{I} - \rho\mathbf{C})^{-1}\boldsymbol{\epsilon}$. The GMM estimator is the set of values for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \rho)$ that minimize

$$Q_n(\boldsymbol{\theta}) = \boldsymbol{u'ZMZ'u}, \tag{74}$$

where $\boldsymbol{Z}$ is a matrix of instruments and $\boldsymbol{C}$ is a positive definite matrix defined based upon spatial neighbors. The iterative process of GMM with respect to the binary SAR model has the following steps.

Step 1: Propose initial values, $\hat{\boldsymbol{\theta}}^{(0)}$, and calculate $\hat{\boldsymbol{u}}$ and the derivative, $\boldsymbol{G}$, with respect to $\boldsymbol{\theta}$. The derivatives of $\boldsymbol{G}$ with respect to $\boldsymbol{\theta}$ are

$$G_{\beta i} = \frac{\partial P}{\partial \boldsymbol{\beta}} P_i(1 - P_i)X_i\beta, \tag{75}$$

$$G_{\rho i} = \frac{\partial P}{\partial \boldsymbol{\rho}} P_i(1 - P_i)X_i\beta\Lambda_{ii}, \tag{76}$$

where $\boldsymbol{\Lambda}$ is an $n \times n$ matrix $(\mathbf{I} - \rho\mathbf{C})^{-1}\mathbf{C}(\mathbf{I} - \rho\mathbf{C})^{-1}(\mathbf{I} - \rho\mathbf{C})^{-1}$ which reduces to $\mathbf{C}$ when $\rho = 0$, no spatial correlation.

Step 2: Regress $\boldsymbol{G}$ on the instrumental variable, $\boldsymbol{Z}$. The predicted values are $\hat{\boldsymbol{G}}$.

Step 3: Construct new estimates for $\boldsymbol{\theta}$, where $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{(\hat{G}'\hat{G})^{-1}\hat{G}'u}$.

Step 4: Iterate until convergence is acquired.

In the case that $\boldsymbol{M} = (\boldsymbol{Z'Z})^{-1}$, the GMM estimator reduces to two-stage least squares estimation. The covariance matrix is given by

$$\boldsymbol{V} = (\mathbf{I} - \rho\mathbf{C})^{-1}(\mathbf{I} - \rho\mathbf{C})'^{-1}. \tag{77}$$

A more efficient model was proposed by Klier & McMillen (2008), which linearizes the Pinkse & Slade (1998) model around a convenient starting point. The linearization makes estimation of binary spatial models computationally faster, yet makes interpretations more difficult due to an extra transformation. It is compared to the Pinkse model with respect to a logit link. The model is linearized around a convenient starting point, the standard logit model (Greene, 2002). The linearization of u_i is centered on initial estimates from the standard logit. The generalized residuals are defined as $u_i = u_i^{(0)} - G(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)})$ where $u_i^{(0)}$ are the standard logit residuals and $\boldsymbol{\theta^{(0)}}$ are the initial estimates. The quadratic form of the GMM estimators, $\boldsymbol{u'ZMZ'u}$, are constructed from the linearized residuals from the following two steps.

Step 1: Estimate the model parameters by a standard logit. The estimated values are $\hat{\boldsymbol{\beta}}$. Calculate the residuals, u_0 and the partial derivatives with respect to $\boldsymbol{\beta}$ and ρ defined as

$$G_{\beta i} = \frac{\partial P}{\partial \boldsymbol{\beta}} P_i(1 - P_i)X_i \tag{78}$$

$$G_{\rho i} = \frac{\partial P}{\partial \boldsymbol{\rho}} P_i(1 - P_i)\mathbf{C}_i\hat{\boldsymbol{\beta}} \tag{79}$$

Step 2: Regress $\boldsymbol{G_\beta}$ and $\boldsymbol{G}_\rho$ on the instrumental variable, $\boldsymbol{Z}$. The predicted values are $\hat{\boldsymbol{G}}_{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{G}}_\rho$. Regress again the initial residuals on $\hat{\boldsymbol{G}}_{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{G}}_\rho$. The resulting coefficients are the estimated values of $\boldsymbol{\theta}$.

The linearized model provides accurate estimates as long as ρ is small. However, when there is strong spatial dependence the linearized model lacks efficiency (Klier & McMillen, 2008).

The primary advantage to using maximum likelihood based estimation such as maximum pseudolikelihood, Monte Carlo maximum likelihood, and Bayesian methods presented in the estimation methods section is the potential for efficiency. However, in spatial models the true structure of the model is rarely known as the selection of neighbors are chosen arbitrarily. For example, neighbors can be chosen based upon adjacency, such as rook, bishop and queen moves as on a chess board, or based upon Euclidean distance such as is commonly used for non-stationary models. In this case, efficiency is less likely to be certain as the true model is not clearly defined whereas generalized method of moments estimation does not impose the same restrictions.

In general, generalized method of moments (GMM) estimation is a promising alternative to maximum likelihood techniques as it has two main advantages over maximum likelihood. The first advantage is GMM does not rely on a potentially inaccurate assumption of normally distributed errors, commonly found in SGLMM. The second advantage is when two-stage least squares are used the estimation does not require calculating the determinants and inverses of $n \times n$ matrices when the model is linearized as in Klier & McMillen (2008).

CHAPTER III

GENERALIZED METHOD OF MOMENTS ESTIMATION

Introduction

Likelihood based estimation methods for the spatial-temporal binary models have a normalizing constant that cannot be evaluated using standard maximum likelihood methods. The complexity of the spatial and temporal terms in the binary model tends to lead to biased parameter estimates when standard estimation techniques such as maximum likelihood methods are used. In Chapter 2, alternative estimation methods based upon Monte Carlo sampling and maximum likelihood, i.e. Monte Carlo maximum likelihood, Monte Carlo maximum expectation, and Bayesian estimation have been presented to obtain less biased estimates. However, with the added complexity of spatial and temporal dependence and associations between observations the likelihood-based methods may not be the most efficient estimation techniques. In complex spatial-temporal binary models the likelihood-based methods can have convergence issues when there is strong spatial and/or temporal dependence, are dependent upon good initial starting values, and the computation time can significantly increases for large spatial frames. These problems arise due

to the need for the full probability distribution to be defined and the intractable normalizing constant.

Maximum likelihood based estimation such as maximum pseudolikelihood, Monte Carlo maximum likelihood, and Bayesian methods presented in Chapter 2 are used for their efficiency. However, the efficiency becomes questionable when the true structure of the model is not known. In spatial modeling the true structure of the model is rarely known as the selection of neighbors are chosen arbitrarily. For example, neighbors can be chosen base upon adjacency, such as rook, bishop and queen moves as on a chess board, or based upon Euclidean distance is commonly used for non-stationary models. In this case, efficiency is less likely to be certain as the true model is not clearly defined whereas generalized method of moments estimation does not impose the same restrictions. In addition, ML estimation methods have been developed for spatial-temporal and centered spatial-temporal models around a stationary spatial grid where the neighbors are defined the same from one time point to another. Yet, if the spatial data frame changes from one time point to another the true structure of the model is harder to define. In this case, likelihood methods may not perform as well requiring additional models and estimation techniques such as GMM to be developed to account for the difference in spatial structure year to year.

In general, generalized method of moments (GMM) estimation has two main advantages over maximum likelihood. The first advantage is GMM does not rely on a potentially inaccurate assumption of normally distributed errors, commonly found

using the spatial generalized linear mixed models (SGLMM). The second advantage is when two-stage least squares are used the estimation does not require calculating the determinants and inverses of $n \times n$ matrices when the model is linearized as in Klier & McMillen (2008). In addition, it is expected that GMM will have a higher convergence rate and be computationally faster compared to maximum likelihood methods.

Generalized method of moments for Spatial-Temporal models were introduced in Chapter 2 with respect to the simultaneous autoregressive process due to its ease of use and Gaussian Markov field neighborhood structure. These GMM methods were all based upon instrumental variables where the error, $\boldsymbol{\epsilon}$, needs to be uncorrelated with the proposed value, $\boldsymbol{Z}$. A disadvantage of instrumental variables is that the working correlation is chosen arbitrarily as any type of regression coefficient can be specified as long as it is uncorrelated with the error. Optimal instruments were proposed by Lee (2007) such that the GMM estimators were asymptotically consistent and efficient, yet the instruments are with respect to continuous data. In the binary spatial model, GMM is based upon either the probit and tobit model or using a logit link but linearizing the model to obtain faster estimates. However, GMM has only been developed for the spatial generalized linear mixed model and does not take into account time nor use the underlying structure of the data as in the autologistic model.

The purpose of this research is to develop an alternative to traditional likelihood methods using generalized method of moments (GMM) for a centered spatial-

temporal autologistic model with stationary and non-stationary data. The centered spatial-temporal autologistic model will be used for stationary data and a modified centered spatial-temporal model will be constructed to deal with non-stationary spatial frame i.e. social network framework. GMM estimation does not have the restriction that the full conditional distribution be defined, but rather can be specified by the first two moments. The GMM approach that is proposed has two major advantages over spatial GMM methods that have been developed. First, GMM has only been developed for spatial models with respect to the spatial generalized linear mixed model and not a spatial-temporal model. The GMM approach that is proposed is with respect to the spatial-temporal centered autologistic model that captures the spatial and temporal dependence directly using a binary Markov random field to capture the natural structure of binary data through neighbor definitions. Second, the working correlation structure for the GMM approach is specified based on a set of estimating equations specified from moment conditions constructed according to a neighborhood structure to deal with the spatial and temporal dependence of the binary data without using instrumental variables, which tend to be chosen arbitrarily.

The Binary Spatial-Temporal Moving Grid Model section recalls the centered spatial-temporal binary autologistic model and presents a modified version where the spatial frame changes for every time point, a moving grid. The Generalized Method of Moments section presents the generalized method of moments approach for spatial-temporal binary models for a stationary and non-stationary

grid. Three approaches are explained and discussed with respect to specified spatial frames.

Binary Spatial-Temporal Moving Grid Model

The centered spatial-temporal autologistic model has a spatial grid that is stationary for each time point. However, in some cases the spatial grid changes at each time point. For example, social interactions have a tendency to change over a period of time therefore causing the spatial structure to change in conjunction with the social network. In addition, methods to assess mountain pine beetle damage are based upon polygons that move at each time point.

A new model is proposed to better handle the changes to the spatial frame at different time points. A modified version of the centered binary spatial-temporal autologistic regression model specified in Chapter 2 is used where distance measurements are incorporated in the spatial and temporal components such that changes in the grid structure are accounted for. The modified centered spatial temporal binary autologistic model uses a conditional distribution of $\boldsymbol{Y_t}$ to model the response variable, Y_{it} at times $1, \cdots, t$ for site $i = 1, \cdots, I$, that depends on the previous, S, times, $t-1, \cdots, t-S$. For each time, t, it is assumed that the response variable follows a Markov random field under a specified spatial neighborhood. A logistic regression model is used, where Y_{it} is assumed Bernoulli due to the binary response.

The conditional probability of success is specified where the probabilities, p_{it}, are modeled using a logit link

$$p(Y_{it}|Y_{i't'} : (i',t') \neq (i,t)) = \frac{\exp\{Y_{it}\eta_{it}\}}{1+\exp\{Y_{it}\eta_{it}\}}, \tag{80}$$

for $t = S+1, \cdots, T$ and the following systematic component

$$\eta_{it} = \sum_{k=0}^{K} \beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} \frac{1}{d_{ij}} Y^*_{jt}\right] + \sum_{s=1}^{S} \frac{1}{d_{ij}} \gamma_s Y^*_{i,t-s} \tag{81}$$

$$\boldsymbol{\eta}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \frac{\boldsymbol{\phi}}{2}\boldsymbol{A_d}\boldsymbol{Y}^*_t + \gamma \boldsymbol{D}\boldsymbol{Y}^*_{t-s},$$

where $\boldsymbol{\beta}$ are the regression coefficients, ϕ is a spatial autoregressive coefficient, γ is a temporal autoregressive coefficient, $\boldsymbol{D}$ is a distance matrix constructed by the Euclidean distance site i is to site j, $\boldsymbol{A}_d$ is an $I \times I$ adjacency matrix constructed by the Euclidean distance, d, and $\boldsymbol{A}_{ij} = (1/d_{ij})\boldsymbol{I}_{\{i\sim j\}}$. The centered response, $Y^*_{i,t}$ is constructed by the ith site and tth time point,

$$Y^*_{it} = Y_{it} - p_{it}, \tag{82}$$

where p_{it} is the probability of $Y_{it} = 1$ under independence as mentioned in the autologistic models in Chapter 2 for the centered model.

The moving grid adds an additional component, d_{ij}, to the centered spatial-temporal binary autologistic model constructed by Wang & Zheng (2013). It represents the Euclidean distance between the current i^{th} site and any other site, j, and is added to both the spatial and the temporal term. The weight for the spatial component is added due to the irregularity of the spatial frame. The temporal term is assigned a weight due to the varying spatial frame each year. Since the spatial

frame moves for each time, the previous time responses, $Y^*_{i,t-s}$ are found based upon the minimum Euclidean distance to the current years response. Previous years' observations that are closest to the current time point are assigned higher weights and those further away are assigned lower weights.

The centering of the model from Chapter 2 indicates that the odds of $Y_{ij} = 1$ increases if the number of positive neighbors is greater than the number of positive neighbors expected under the independence model, and decreases if the number of positive neighbors is less than the number of positive neighbors expected under the independence model.

The joint distribution of the model, for sites $i = 1, \cdots I$ for a given time point, t, is defined in the same way as in the spatial-temporal binary autologistic model section based upon the Hammersley-Clifford Theorem (Cressie, 1993) as follows,

$$\begin{aligned} p(\boldsymbol{Y}_t : \boldsymbol{Y}'_t) &= \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z\in\Omega}\exp(Q(\boldsymbol{Z}_t;\boldsymbol{\theta},\phi,\gamma))} \\ &= c(\boldsymbol{Y}_{t-1},\cdots,\boldsymbol{Y}_{t-S}:\boldsymbol{\theta},\phi,\gamma)^{-1} \times \exp\sum_{i=1}^{I} Y_{it}\eta_{it} \\ &= c(\boldsymbol{Y}_t,\boldsymbol{\theta})^{-1} \times \exp\left(\boldsymbol{Y_t'X\beta} + \frac{\boldsymbol{\phi}}{2}\boldsymbol{Y_t'A_dY_t^*} + \gamma\boldsymbol{Y_t'DY^*_{t-s}}\right), \end{aligned}$$

where $c(\boldsymbol{Y}_t,\boldsymbol{\theta})^{-1}$ is a normalizing constant without a closed form representation. In Chapter 2, psuedolikelihood, Monte Carlo maximum likelihood, Bayesian and expectation-maximization methods are used to estimate the parameters. In the next section a GMM approach is introduced as an alternative method to maximum likelihood methods for the centered spatial-temporal binary autologistic model.

Generalized Method of Moments

Generalized method of moments (GMM) estimation is an alternative method to likelihood methods as it does not require complete knowledge of the distribution of the data; instead it uses moment methods derived from an underlying model. In place of using a likelihood, GMM is used to estimate the p-dimensional parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \gamma)'$ from a set of $N = I \times T$ moment conditions where I represents the number of sites and T is the number of time points selected. A general form of GMM with respect to space and time proceeds as follows.

Suppose that for a vector of responses, denoted as $\boldsymbol{Y} = (Y_1, \cdots, Y_N)$, the $p \times 1$ parameter estimates, $\boldsymbol{\theta}$, are unknown. The goal of the estimation problem is to find the true value of the parameters, $\boldsymbol{\theta}$, or at least a reasonably close estimate by using moment conditions. The moment conditions, denoted as $\boldsymbol{g(Y, \theta)}$, are defined such that

$$E[\boldsymbol{g(Y, \theta)}] = \mathbf{0}, \tag{83}$$

where $\boldsymbol{g(Y, \theta)}$ is a $q \times 1$ function of $\boldsymbol{\theta}$ and $\boldsymbol{Y}$ and the components of $E[\boldsymbol{g(Y, \theta)}]$ exist and are finite for all values of $\boldsymbol{\theta}$. The moment conditions represent q equations for p unknowns where $q > p$. In the case of ordinary method of moments $\hat{\boldsymbol{\theta}}$ would be estimated by setting the following sample moment conditions,

$$G_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} g(\boldsymbol{Y_i}, \hat{\boldsymbol{\theta}}), \tag{84}$$

to zero and solving for $\hat{\boldsymbol{\theta}}$. However, since $q > p$, the system is over-identified therefore no solution exists. To overcome this difficulty, Hansen (1982) proposed using a positive definite weight matrix, $\boldsymbol{W}_N$, to create a quadratic form,

$$Q_N(\boldsymbol{\theta}) = G_N(\boldsymbol{\theta})'\boldsymbol{W}_N G_N(\boldsymbol{\theta}) \tag{85}$$

which can be minimized by

$$\hat{\boldsymbol{\theta}}_{GMM} = \text{argmin}_\theta Q_N(\boldsymbol{\theta}) \tag{86}$$

such that the parameter estimates can be found. The idea is to minimize the quadratic form such that $E[\boldsymbol{g}(\boldsymbol{Y}, \hat{\boldsymbol{\theta}})]$ is as close to zero as possible. In the case of spatial dependence as recalled from the generalized method of moments section in Chapter 2, the choice of the sample moment conditions, $G(\boldsymbol{\theta}) = \boldsymbol{u'Z}$, were based upon the spatial autoregressive binary model from a set of arbitrarily chosen instrumental variables, $\boldsymbol{Z}$, such that the error and the instrumental variable were uncorrelated. However, the instrumental variable method did not account for time, only space.

A new approach for GMM accounts for spatial and temporal dependence using the spatial temporal binary autologistic model. In this method the moment conditions, denoted as $G_{IT}(\boldsymbol{\theta})$, are specified by a spatial-temporal frame. It uses the structure of the spatial neighborhood to construct moment conditions rather than arbitrarily selecting an instrumental variable. It captures the spatial dependence among the sites through the spatial neighborhood and the temporal dependence is captured from the previous years' observations. The alternative GMM method creates a quadratic form for each year conditional on the previous year and

then minimizes the sum of all the years' quadratic forms. The alternative GMM method is specified as follows. Let the quadratic form for the spatial-temporal binary autologistic model be denoted as

$$Q_{IT}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{G}_{I1}^T(\boldsymbol{\theta}) & \cdots & \boldsymbol{G}_{IT}^T(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{I1} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{W}_{IT} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{G}_{I1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{G}_{IT}(\boldsymbol{\theta}) \end{bmatrix}, \tag{87}$$

where the valid sample moment conditions with respect to all the sites, $\boldsymbol{G}_{I1}(\boldsymbol{\theta}), \cdots, \boldsymbol{G}_{IT}(\boldsymbol{\theta})$, are defined as

$$\boldsymbol{G}_{It}(\boldsymbol{\theta}) = \left(\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' (\boldsymbol{J} - \boldsymbol{N})^* (\boldsymbol{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})). \tag{88}$$

The matrix, $(\boldsymbol{J} - \boldsymbol{N})^*$, is the set of non-neighbors that are uncorrelated with each site of the response, $\boldsymbol{J}$ is a matrix composed of 1's and $\boldsymbol{N}$ is a neighbor matrix that is defined based upon spatial locations the set of residuals are $(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it}(\boldsymbol{\theta}))$. Traditionally, the matrix $(\boldsymbol{J} - \boldsymbol{N})^*$, is an $I \times I$ matrix leaving only the sites that are not neighbors with site i as 1's and all others 0 values. However, GMM will be constructed in multiple ways where $(\boldsymbol{J} - \boldsymbol{N})^*$ will represent an expanded matrix of neighbors. Details of the methods are to follow. The residuals, $(\boldsymbol{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))$, with respect to all I sites are the generalized logit residuals where

$$\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\eta}_t)}{1 + \exp(\boldsymbol{\eta}_t)}. \tag{89}$$

In (87), note that Q_{IT} is a sum of quadratic functions for each year as

$$Q_{I1} = \boldsymbol{G}_{I1}(\boldsymbol{\theta})'\boldsymbol{W}_{I1}^{-1}\boldsymbol{G}_{I1}(\boldsymbol{\theta}) \tag{90}$$

$$\vdots$$

$$Q_{IT} = \boldsymbol{G}_{IT}(\boldsymbol{\theta})'\boldsymbol{W}_{IT}^{-1}\boldsymbol{G}_{IT}(\boldsymbol{\theta}).$$

The quadratic functions are constructed with respect to the current year's response, Y_t, and the previous year's response, Y_{t-1}, as constructed in the systematic component in the spatial-temporal autologistic model. For example, for $t = 1$, the quadratic function, Q_{I1} will only be calculated with respect to the current year as there is not a previous years' response in the systematic component. In the second year, $t = 2$, Q_{I2} will take into account the current years response and the previous years' response, Y_{t-1} is contained within the residuals for the second years' sample moment condition, G_{I2}. After the quadratic form is constructed with respect to each year, the quadratic form, Q_{IT}, is minimized such that $Q_{IT} = \sum_{t=1}^{T} Q_{It}$ from combinations of the $q > p$ available moment conditions.

The neighbor matrix, $\boldsymbol{N}$ of dimension $I \times I$ can be specified in one of two ways. The first neighbor matrix is constructed from a stationary grid where the spatial frame stays the same each year and the second is constructed based upon a non-stationary grid. The choice of the neighbor matrix is determined based upon the following definitions.

1. Consider a stationary lattice where neighborhoods are based upon those closest to site i using a rook, queen, or bishop design. The maximum number

of neighbors based upon any one of the designs is four for a regularly spaced grid. For example, a rook format would look as follows:

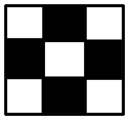


Figure 3. Rook neighbor structure

2. Consider a non-stationary lattice where the neighbor structure of the spatial frame changes with respect to the time points. In this scenario, weights will be defined based upon Euclidean distance as presented in the binary spatial-temporal moving grid model section. Peer network associations, will be based upon relative distance or strength of relationship one peer has to another.

Spatial-Temporal Generalized Method of Moments Approaches

The number of moment conditions can vary depending upon different specifications of a neighborhood matrix as noted by $(\boldsymbol{J} - \boldsymbol{N})^*$. The number of non-neighbors from the neighbor matrix can be large depending upon how the neighbor matrix is specified. For example, a 10×10 grid creates a 100×100 adjacency neighbor matrix. Row 1 denotes the neighbors with site 1, which has 2 neighbors, and therefore will be $100 - 2 = 98$ non-neighbors. Site 2 has 3 neighbors, 97 non-neighbors, etc. As the moment conditions are dependent upon the number of non-neighbors, the number of moment conditions increases substantially depending

upon the size of the grid. Due to this potential problem, alternative GMM methods have been specified in three ways.

Method 1: Let the sample moment conditions for each year be specified as $\boldsymbol{G}_t(\boldsymbol{\theta}) = \boldsymbol{D}'_{\text{exp}} (\boldsymbol{J} - \boldsymbol{N})_{\text{exp}} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))_t$. $\boldsymbol{D}_{\text{exp}}$, is an expanded derivative matrix where the derivative with respect to each site, i, is repeated ($I-$# neighbors of site i) times in a diagonal matrix, where I is the number of sites. Let $\boldsymbol{D}_k$ denote the matrix of derivatives with respect to each site and parameter, k. For one time, let the expanded derivative matrix be denoted as

$$\boldsymbol{D}_{\text{exp}} = \begin{bmatrix} \boldsymbol{D}_1 \\ - - \\ \boldsymbol{D}_2 \\ - - \\ \vdots \\ - - \\ \boldsymbol{D}_k \end{bmatrix} \tag{91}$$

where

$$\boldsymbol{D}_1 = \begin{bmatrix} \boldsymbol{d}_{11} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{d}_{I1} \end{bmatrix}, \tag{92}$$

$$\vdots$$

$$\boldsymbol{D}_k = \begin{bmatrix} \boldsymbol{d}_{1k} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{d}_{Ik} \end{bmatrix}.$$

The dimension of each of the derivative matrices, $\boldsymbol{D}_1, \cdots, \boldsymbol{D}_k$, is $(\sum_i NN_i \times \sum_i NN_i)$ where $NN_i = (I - \#$ neighbors of site i), the number of non-neighbors for each site, i. The submatrices, $\boldsymbol{d}_{11}, \cdots \boldsymbol{d}_{I1}, \cdots \boldsymbol{d}_{1k}, \cdots, \boldsymbol{d}_{Ik}$, are diagonal matrices constructed for each site based upon the number of non-neighboring sites, NN_i as follows

$$\boldsymbol{d}_{11} = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}_{1t}(\boldsymbol{\theta})}{\partial \theta_1}\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(\frac{\partial \boldsymbol{\mu}_{1t}(\boldsymbol{\theta})}{\partial \theta_1}\right) \end{bmatrix} \tag{93}$$

$$\vdots$$

$$\boldsymbol{d}_{Ik} = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}_{It}(\boldsymbol{\theta})}{\partial \theta_k}\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(\frac{\partial \boldsymbol{\mu}_{It}(\boldsymbol{\theta})}{\partial \theta_k}\right) \end{bmatrix}$$

For example, a 3×3 lattice has 9 sites. The first site has two neighbors, therefore the derivative with respect to the first site and first parameter, $\boldsymbol{d}_{11}$ will be repeated on the diagonal $9 - 2 = 7$ times, site 2 has three neighbors so $\boldsymbol{d}_{21}$ will be repeated on the diagonal $9 - 3 = 6$ times, etc. The number of rows of the expanded derivative matrix for a 3×3 grid will be $k \times \sum_i NN_i = k \times 57$ and the number of columns will be 57.

The expanded neighbor matrix, $(\boldsymbol{J}-\boldsymbol{N})_{\text{exp}}$ is a stacked matrix denoted as

$$(\boldsymbol{J}-\boldsymbol{N})_{\text{exp}} = \begin{bmatrix} \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{M}_I \end{bmatrix}, \tag{94}$$

where each row containing non-neighbors for site i are expanded such that each row only contains one non-zero value.

For example, if $(\boldsymbol{J}-\boldsymbol{N}) = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$ then

$$\boldsymbol{M}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \boldsymbol{M}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{95}$$

$$\boldsymbol{M}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{M}_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{96}$$

Note that the number of rows of $(\boldsymbol{J}-\boldsymbol{N})_{\text{exp}}$ is $\sum_{i=1}^{I} NN_i$ which equals the number of columns of $\boldsymbol{D}_{\text{exp}}$. The number of sample moment conditions, $\boldsymbol{G}_t(\boldsymbol{\theta}) = \boldsymbol{D}'_{\text{exp}}(\boldsymbol{J}-\boldsymbol{N})_{\text{exp}}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\theta}))_t$, for one year using this method results in $k \times \sum_{i=1}^{I} NN_i$ moment conditions. To re-iterate for the 3×3 grid example, method 1 results in $57 = \sum_{i=1}^{I} NN_i \times k$ moment conditions. A 10×10 grid will have approximately $8090 \times k$ moment conditions per year, which becomes computationally challenging due to the calculation of matrix inverses. As the grid size increases this method

starts to become infeasible to compute with standard computers. For example the 10×10 grid with multiple years is not able to be held in memory on an 8 core windows computer and a 5×5 grid simulation with 1000 samples would take over a week to produce estimates.

Method 2: In this case, properties of method 1 are used but in this method the number of non-neighbors is reduced. It only includes the values that are within an l^{th} order of neighbors. The first order neighbors are the sites that are directly

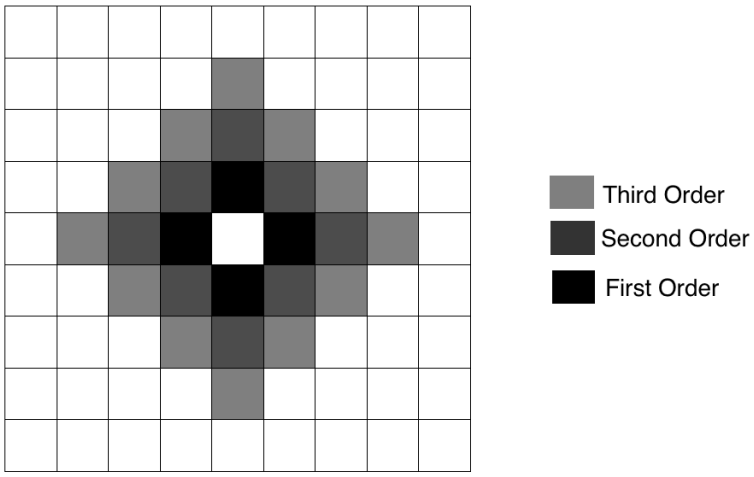


Figure 4. Order of neighborhoods

adjacent to site i, the second order neighbors are the sites that are adjacent to the first order neighbors but not including the first order neighbors, etc. Figure 4 displays a 9×9 grid example of 1st, 2nd, and 3rd order neighbors.

The moment conditions in this method are reduced as only the neighbors within the l^{th} neighborhood are considered when constructing the moment conditions. Method 1, on the other hand, considers all possible conditions among non-neighboring sites. The number of sample moment conditions, $\boldsymbol{G}_t(\boldsymbol{\theta}) = \boldsymbol{D}'_{\exp}\left(\boldsymbol{J} - \boldsymbol{N}\right)_{\exp}\left(\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta})\right)_t$, for one year using this method results in $k \times \sum_{i=1}^{I} NN_{il}$ moment conditions, where NN_{il} is the number of non-neighbors for site i to the l^{th} spatial neighborhood. For example, a 10×10 grid including only neighbors within a spatial neighborhood of order 5, will have approximately $800 \times k$ moment conditions per year, a reduction of one order of magnitude from the approximately $8090 \times k$ moment conditions from method 1.

Method 3: Let the sample moment conditions for each year be specified as $\boldsymbol{G}_t(\boldsymbol{\theta}) = \boldsymbol{D}'(\boldsymbol{J} - \boldsymbol{N})(\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))_{\boldsymbol{t}}$. In this method, $\boldsymbol{D}$, is an expanded derivative matrix with respect to each site, i, and parameter θ_k. For one time, let the expanded derivative matrix, $\boldsymbol{D}$ be denoted as

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_1 \\ - - \\ \boldsymbol{D}_2 \\ - - \\ \vdots \\ - - \\ \boldsymbol{D}_k \end{bmatrix} \tag{97}$$

where

$$\boldsymbol{D_1} = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}_{1t}(\boldsymbol{\theta})}{\partial \theta_1}\right) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \left(\frac{\partial \boldsymbol{\mu}_{It}(\boldsymbol{\theta})}{\partial \theta_1}\right) \end{bmatrix}$$

$$\vdots \tag{98}$$

$$\boldsymbol{D_k} = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}_{1t}(\boldsymbol{\theta})}{\partial \theta_k}\right) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \left(\frac{\partial \boldsymbol{\mu}_{It}(\boldsymbol{\theta})}{\partial \theta_k}\right) \end{bmatrix}$$

The dimension of each of the derivative matrices $\boldsymbol{D}_1, \cdots, \boldsymbol{D}_k$ is an $I \times I$ matrix where the derivatives with respect to each site are on the diagonal. The full derivative matrix, therefore, has $I \times k$ rows and I columns.

The matrix $(\boldsymbol{J} - \boldsymbol{N})$, is an $I \times I$ matrix leaving only the sites that are not neighbors with site i a value of 1 and all other sites are given a value of 0. Each moment condition in this method is summed over the non-neighboring sites ($I -$ # neighbors of site i). This method results in $I \times k$ moment conditions, a significant reduction in estimating conditions from method 1. For example, method 1 has approximately $8090 \times k$ moment conditions per year, method 2 will have approximately $800 \times k$ conditions whereas method 3 will have $100 \times k$ moment conditions per year. If the spatial grid is relatively large then method 3 is the preferred choice due to the reduction in moment conditions.

The relative importance of each of the original moment conditions in the GMM estimator is determined by how much information it provides about $\boldsymbol{\theta}$, which is measured by $\left(\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'(\boldsymbol{J} - \boldsymbol{N})$ and the weight matrices, $\boldsymbol{W}_{I1}, \cdots, \boldsymbol{W}_{IT}$. The weight matrices, $\boldsymbol{W}_{I1}, \cdots, \boldsymbol{W}_{IT}$, are found from the inverse of the covariance matrix of the moment conditions, $\boldsymbol{g}(\boldsymbol{Y}, \hat{\boldsymbol{\theta}})$, defined as $\boldsymbol{V}^{-1} = Cov(\boldsymbol{g}(\boldsymbol{Y}, \hat{\boldsymbol{\theta}}))^{-1}$. Hansen (1982) showed that they are the optimal choice weight matrices. The idea is that the covariance matrix will assign less weight to the sample moment conditions that have large variances versus those with smaller variances (Qu, Lindsay & Li, 2000). The optimal weight matrices and $\hat{\boldsymbol{\theta}}_{GMM}$ can be found by a using a two-step GMM algorithm described in Chapter 4.

CHAPTER IV

ESTIMATION METHOD COMPARISONS

Introduction

This chapter describes the simulation methods used to compare Monte Carlo maximum likelihood, Monte Carlo expectation-maximization and generalized method of moments estimation developed in chapter 3. The description is followed by a comparison of two different lattice structures, a 5×5 and 10×10 grid and two real datasets. The comparison demonstrates how the proposed GMM methods from chapter 3 perform on a small simulation and two real datasets and is compared to existing methods. In addition, the modified centered spatial-temporal binary autologistic model is demonstrated on one of the datasets.

To make comparisons, a small simulation study simultaneously comparing the estimation methods (MCEML, MCML, and GMM) is constructed. The simulation study presented draws comparisons of these three methods via examination of computation time, convergence, precision, and bias associated with the corresponding estimators. The objective of the simulation study is to answer the following questions:

1. How does the computation time compare across the estimation methods?

2. How does MCEML, MCML, and GMM compare under different types/strengths of spatial dependence and lattice size?

3. Do the estimation methods converge?

4. What is the precision and bias of the estimation methods?

To address the questions, multiple lattices sizes (5×5, and 10×10) and two different spatial dependence values were constructed in the simulation study. The specifics of the simulation study are discussed in the following section. The specific values for spatial dependence and the covariate chosen for the model were chosen based upon the study by Wang & Zheng (2013) to best make comparisons to existing estimation methods.

The simulation study is carried out in R, with most of the program written by the author. The *glm* function is used to obtain maximum pseudolikelihood and expectation-maximization pseudolikelihood estimates and both *optim* and *maxNR* are used to carry out numerical optimization. The functions written by the author include construction of a perfect sample, Gibbs sampler with perfect sampler, Monte Carlo Markov Chain maximum likelihood, generalized method of moments, and Monte Carlo expectation-maximization programs.

The remainder of the chapter is divided into the following sections. The second section explains and describes the cases in the simulation study and how each case is implemented with respect to GMM, MCML, and MCEML. The third section presents and compares estimation methods from the simulation study. The

fourth section presents results based upon two data examples. Lastly, the simulation and data set results are summarized.

Simulation

The centered spatial-temporal binary autologistic model is used to compare estimation methods for simulated data from two different lattice sizes, 5×5 and 10×10, a small and moderate size lattice. To investigate the potential problems when spatial and/or temporal dependence is high, two different spatial dependence levels, 0.1 (low) and 0.9 (high), are assessed in the simulation study. One continuous covariate generated from $X_{it} \sim N(3, 1)$ is considered in the model which varies across space and time. The intercept, slope of the covariate, and temporal component are fixed as $(\beta_0, \beta_1, \gamma)' = (1, -0.5, 0.5)'$. The temporal component is fixed as 0.5 to indicate a moderate temporal effect. The spatial parameter, ϕ, is allowed to vary at values $\phi = 0.1$ and $\phi = 0.9$ to reflect different degrees of spatial dependence.

The simulated binary data for the centered spatial-temporal autologistic model are drawn from the conditional distribution defined as

$$\log\left(\frac{P(Y_{it}=1)}{P(Y_{it}=0)}\right) = \beta_0 + X_{1it}\beta_1 + \frac{\phi}{2}\sum_{N_i} Y^*_{it} + \gamma Y^*_{i,t-1}, \tag{99}$$

where

$$Y^*_{it} = Y_{it} - p_{it}. \tag{100}$$

$Y^*_{i,t}$ denotes the centered response for the ith site and tth time point, β_0 is the intercept, β_1 is the slope of the covariate, ϕ is the spatial autoregressive coefficient and γ

is the temporal autoregressive coefficient. The probability, p_{it}, is the expectation of Y_{it} under the independent logistic regression model,

$$p_{it} = \frac{\exp(\beta_0 + X_{1it}\beta_1)}{1 + \exp(\beta_0 + X_{1it}\beta_1)}. \tag{101}$$

For each combination of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \gamma)'$ and initial values, $M = 1000$ datasets are generated from the centered spatial-temporal binary autologistic regression model using a perfect simulation sample with a Gibbs sampler. The choice of fixed values, covariate, spatial dependence levels and number of iterations are chosen based upon the study by Wang & Zheng (2013) to make comparisons to existing estimation methods for the centered binary autologistic spatial-temporal model.

Simulation Algorithm

The simulation study is carried out in R, using samples generated from perfect simulation samples with a Gibbs sampler. The idea is to start at a value close to the true value via perfect sampling and then generate the rest of the samples from a Gibbs sampler. This helps to ensure that the samples are relatively good realizations of the true data based upon set covariates. The following algorithm generates samples for each case:

Step 1: For each spatial dependence, $\phi = 0.1, 0.9$, and lattice size, 5×5, 10×10, draw $M = 1000$ samples from the centered spatial-temporal binary autoregressive model, (99), with spatial, temporal and covariates using a perfect simulation sample with Gibbs sampler.

Step 2: Fit the model using Monte Carlo maximum likelihood, Monte Carlo expectation-maximization and generalized method of moments (method 2 and method 3) using

1. Initial values specified as a vector of zeros, $\mathbf{0} = (0, 0, 0, 0)$.
2. Initial values specified by maximum pseudolikelihood (MPL) or expectation-maximization pseudolikelihood (EMPL) methods.

Step 3: For each iteration, record if the model converges and computation time.

Step 4: Summarize the M samples and record the precision and parameter estimates.

Initial testing of the simulated values is conducted by comparing simulated values generated using a perfect simulation sample written in R to values generated from the *ngspatial* package in R (R Core Team, 2013), constructed for maximum pseudolikelihood estimation of the spatial binary autologistic model. It is verified that the simulated samples from the program written in R by the author produced the same percentage of 1's as the *autologistic* program in the *ngspatial* package (Hughes & Cui, 2013). After verification, the temporal term is added to the perfect sampler program where samples are generated via the centered spatial-temporal autologistic model. Next, the simulated samples are tested using maximum pseudolikelihood estimation and verification that the parameter estimates are close to the original parameter estimates in the low spatial dependence case. Verification with high spatial

dependence is not conducted as MPL often struggles with high spatial/temporal dependence (Huffer & Wu, 1998).

Generalized Method of Moments Algorithm

In Chapter 2, parameter estimation for the likelihood-based methods (MCML, MCEML) was defined. Recall that MCML and MCEML generate M samples from a ratio of log-likelihoods and then maximizes the log-likelihood function to find the parameter estimates. The process for finding the parameter estimates for GMM, $\hat{\boldsymbol{\theta}}_{GMM}$, however, are found using a two-step GMM algorithm. The following algorithm describes the process for obtaining the parameter estimates for GMM, a two-stage process.

Step 1: Obtain initial estimates, $\hat{\boldsymbol{\theta}}^{(1)}$, by Maximum Pseudolikelihood.

Step 2: Find an estimate for the optimal weight matrix, $\hat{\boldsymbol{W}}_{IT}$, based upon $\hat{\boldsymbol{\theta}}^{(1)}$.

Find the GMM estimator

$$\hat{\boldsymbol{\theta}}_{GMM} = \text{argmin}_{\theta} \boldsymbol{G}_{IT}(\boldsymbol{\theta})' \hat{\boldsymbol{W}}_{IT} \boldsymbol{G}_{IT}(\boldsymbol{\theta}).$$

where

$$\hat{\boldsymbol{W}}_{IT} = \left(g(\boldsymbol{Y}_t, \hat{\boldsymbol{\theta}}^{(1)}) g(\boldsymbol{Y}_t, \hat{\boldsymbol{\theta}}^{(1)})' \right)^{-1}.$$

Simulation Statistical Performance

The 1000 simulated data samples for the centered spatial-temporal binary autologistic model are estimated using MCML, MCEML, and GMM methods for

comparison. The parameter estimates and convergence generated from *optim* in R are recorded for each sample. To evaluate the bias and precision of the parameter estimates of the simulated data sets, the average and standard deviation of all the datasets estimates are reported and denoted as estimate and standard error (SE) based upon only the convergent data sets. The sections labeled Computational Demand and Convergence and Estimation Bias and Precision describe the results for the different size lattices, 5×5 and 10×10. The results are based upon computation time, convergence and parameter estimate comparison.

A 5×5 grid is constructed to assess how the generalized method of moments approaches described in Chapter 3 compare to MCML and MCEML. The first generalized method of moments approach (Method 1) involves constructing moment conditions for every non-neighboring site. However, this method results in a set of moment conditions that is significantly larger than the number of observations. Method 1 resulted in a number of moment equations that is approximately 460 times the amount of observations, therefore it is not an advisable method to use (Hansen, 2001). As mentioned in Chapter 3, Method 1 is over identified. The second method uses properties of Method 1, but reduces the number of neighbors to the fourth order neighbors, whereas method 3 uses a linear combination to combine all non-neighboring sites for each moment condition. Additionally, a stationary 10×10 grid was constructed to compare GMM method 3 to the likelihood-based methods MCML and MCEML. Method 2 was not feasible for the 10×10 grid as

R ran out of memory on a personal computer when the moment conditions for each year are stored.

For each simulated data set, the EMPLEs and MPLEs, are computed using the estimates from the independent logistic regression model as initial values. The initial estimates for MCEML and MCML using EMPL and MPL, respectively, are compared to an arbitrary starting point chosen as a vector of 0's. The initial estimates for GMM estimation are chosen based upon MPL and a vector of 0's for comparison.

Computation demand and convergence. Comparison of computation time and convergence for MCML, MCEML, GMM_2 and GMM_3 are displayed in Table 1 and Table 2 for a 5×5 and 10×10 lattice, different values of ϕ, and different initial values. For MCML and MCEML a total of 7500 Monte Carlo samples are generated where the first 500 samples are discarded for burn-in and every 5th draw was kept to obtain 1000 samples. The computing time was based on R programs written by the author and run on an Intel i7-2600k, 4.5 gHz personal computer.

Convergence and computation time 5×5 ***Grid.*** Table 1 displays the convergence rates and computation time for the 5×5 lattice. MCML and GMM_2 demonstrate to be the most computationally similar with little change in time based upon initial starting values. Interestingly, GMM_3 is the quickest estimation method by a reduction of almost five times that of the second fastest method, MCML. It is not surprising to find that MCEML takes the longest computation time as it first generates a sample of centered responses, calculates a new parameter

value, and then updates the responses for each sample. The complexity comes from generating Monte Carlo samples from the reference parameter, updating the centered response with respect to each Monte Carlo sample at each expectation step, and maximizing the result. The process is repeated in the EM algorithm until the l^{th} $\boldsymbol{\theta}$ value compared to the $(l-1)$th sample reaches a specified tolerance level resulting in additional computation time. This justifies why MCEML is on the order of three times the amount of time it takes GMM_2 to estimate the parameters. The computation time for GMM_2 is longer than GMM_3 due to the large dimension of the moment conditions and covariance structure that needs to be inverted for the initial weight matrix, $\hat{W}_{IT}$. Although the method is similar for GMM_3, the number of moment conditions is greatly reduced due to the summation of non-neighbors for each site.

MCML, GMM_2 and GMM_3 all have high proportions of convergence. All three estimation methods have convergence rates above approximately 97.5%, with MCML converging for almost every sample with only approximately every 3 or 4 samples per 1000 that did not converge. The two GMM methods, GMM_2 and GMM_3, performed equally as well in terms of convergence. Convergence rates of MCEML are shown to be much lower than MCML, GMM_2 and GMM_3 with a convergence rate around 91% for the starting values of 0 and about 94% when EMPL estimates are used. The convergence rates also varied depending upon the level of spatial dependence for MCEML. In the case of high spatial dependence the proportion of convergence for MCEML is not as high as the proportion of convergence for

Table 1

Computation time and convergence comparison for different initial starting values and values of ϕ for a 5×5 stationary lattice.

Estimation Method	Lattice $i \times i$	Spatial ϕ	Temporal γ	Initial Values	Computation Time (sec)	Proportion of Convergence
MCML	5×5	0.1	0.5	MPL	54.01	0.998
	5×5	0.9	0.5	MPL	57.11	0.996
	5×5	0.1	0.5	**0**	78.04	0.998
	5×5	0.9	0.5	**0**	85.18	0.997
MCEML	5×5	0.1	0.5	EMPL	395.49	0.960
	5×5	0.9	0.5	EMPL	415.01	0.937
	5×5	0.1	0.5	**0**	400.05	0.921
	5×5	0.9	0.5	**0**	417.41	0.918
GMM_2	5×5	0.1	0.5	MPL	97.61	0.980
	5×5	0.9	0.5	MPL	98.68	0.978
	5×5	0.1	0.5	**0**	98.04	0.978
	5×5	0.9	0.5	**0**	100.88	0.980
GMM_3	5×5	0.1	0.5	MPL	12.04	0.981
	5×5	0.9	0.5	MPL	9.77	0.979
	5×5	0.1	0.5	**0**	10.54	0.980
	5×5	0.9	0.5	**0**	9.76	0.981

low spatial dependence. This result is not seen in GMM_2, GMM_3 or MCML. It is important to note that although MCEML converges using the Newton-Raphson algorithm, approximately 46.5% of the time the number of iterations, set to 100 for the EM algorithm, is exceeded.

The starting values of MPL and EMPL versus a vector of zeros has little impact on the computation time and convergence for MCML, GMM_2, and GMM_3. The convergence rate for MCEML is impacted by the initial estimates the most as there is approximately a 1% decrease in the convergence rate. In terms of computation time MCML and MCEML estimation result in the largest difference in computation time by approximately 15 seconds for MCML and 10 seconds for MCEML and only about a 3 second difference for either GMM_2 and GMM_3.

Convergence and computation time 10×10 ***Grid.*** Table 2 displays the convergence and computation time for the 10×10 grid. The change in lattice size results in a dramatic increase in computation time for both MCML and GMM methods. The results for GMM_2 are not reported as the time to compute the GMM estimates for one iteration took over one hour. Based upon this discovery, GMM_2 is not included in the 10×10 study. In addition, R ran out of memory from the multiple matrix inversions based upon each year's moment equations storage despite using the R package *BigMemory* on a personal computer. Estimation for GMM_3 is still possible but demonstrated an increase in computation time. In the 10×10 case MCML is found to be the fastest method computationally by a reduction of approximately five times that of GMM_3.

In the 10×10 lattice, the fastest method is MCML followed by GMM_3, MCEML computation time took twice as long as GMM_3. The computation time with respect to spatial dependence levels showed an increase in time for MCML and

Table 2

Computation time and convergence comparison for different initial starting values and values of ϕ for a 10×10 stationary lattice.

Estimation Method	Lattice $i \times i$	Spatial ϕ	Temporal γ	Initial Values	Computation Time (sec)	Proportion of Convergence
MCML	10×10	0.1	0.5	MPL	128.33	0.996
	10×10	0.9	0.5	MPL	132.45	0.997
	10×10	0.1	0.5	**0**	130.33	0.995
	10×10	0.9	0.5	**0**	137.98	0.996
MCEML	10×10	0.1	0.5	EMPL	991.71	0.912
	10×10	0.9	0.5	EMPL	1148.60	0.875
	10×10	0.1	0.5	**0**	987.01	0.896
	10×10	0.9	0.5	**0**	1162.91	0.899
GMM_3	10×10	0.1	0.5	MPL	566.86	0.983
	10×10	0.9	0.5	MPL	511.66	0.978
	10×10	0.1	0.5	**0**	542.02	0.982
	10×10	0.9	0.5	**0**	501.66	0.988

MCEML when the spatial dependence was high. GMM_3 slightly decreased by approximately 10% in computation time when there is large spatial dependence.

MCML and GMM_3 both suggest a high proportion of convergence with rates above approximately 97.5%. MCEML however, only converges approximately 87.5% of the time, 10% lower than MCML and GMM_3. Again, it is important to note that although MCEML converges based upon the Newton-Raphson algorithm, approximately 53% of the time the EM algorithm exceeds the number of iterations

before hitting a specified tolerance. Spatial dependence did not appear to have an impact on any of the estimation methods.

Convergence and computation summary. Overall, the results of GMM estimation compared relatively well to the other estimation methods in terms of computation time and convergence. On the 5×5 lattice, GMM_2 took longer to compute than GMM_3 due to the number of moment conditions and yet still showed similar convergence rates. The different values of spatial dependence did not have a significant impact on the convergence rates on either lattice size with the exception of MCEML. However, MCML and MCEML exhibit a slightly longer computation time for both lattice sizes when the spatial dependence is high. The increase in computation time is due to the additional time needed for the perfect sampler's chains to coalesce from the higher spatial dependence. In the next section, the parameter estimates bias and precision are presented and discussed.

Estimation Methods Performance. In the simulation study, parameter estimates bias and the precision were recorded for each of the different spatial components and initial values. The precision and estimates were found from the standard deviation and average, respectfully, of the 1000 simulation data sets denoted as Estimate and SE in the tables to follow. The four different estimation methods, MCEML, MCML, GMM_2, and GMM_3 are compared with respect to varying spatial dependence, starting values and lattice size.

Bias and Precision for the 5×5 ***Grid.*** Two different GMM methods, GMM_2 and GMM_3 are compared to existing methods MCML and MCEML for the

5×5 grid. As can be seen from Table 3, Table 4, Figure 5 and Figure 6, the parameter estimates bias for GMM_2 and MCML perform similarly for the intercept, β_0, spatial and temporal terms in the model. GMM_2 and MCML display almost no bias in the parameter estimates when $\phi = 0.1$ or $\phi = 0.9$ in the spatial and temporal terms. The precision of the estimates based upon one standard error from the parameter estimates for GMM_2 also suggest tighter intervals compared to MCML with the exception of the slope, β_1. These results are consistent across the varying spatial dependence values, $\phi = 0.1$ and $\phi = 0.9$. Interestingly, MCEML and GMM_3 performed equally as well in terms of bias. Although, more bias is exhibited in MCEML with regard to the intercept term compared to the other estimation methods, the spatial term is relatively close to the true estimate. Figures 5 and 6 suggest that the standard error intervals of GMM_3 display the largest variation for low and high spatial dependence as compared to other estimation methods. However, the bias decreased when the spatial dependence is high. The greatest bias is shown when there is high spatial dependence for the MCEML estimates as compared to MCML, GMM_2 and GMM_3 estimation methods.

Depending upon the different levels of spatial dependence, the MCEML estimated coefficients on the slope and temporal terms fluctuated signs from the true parameter values. In the case of high spatial dependence, the signs are the same as the true values. However, when there is low spatial dependence for MCEML the coefficients are opposite of the true values exhibiting the greatest amount of bias. The sign of the coefficient is also different than the true value for the temporal term

Table 3

Parameter estimation with starting values from MPL and EMPL on a 5×5 grid with 5 time points.

		MCEML		MCML		GMM_2		GMM_3	
Actual	Covariate	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
1.0	β_0	-0.916	3.246	1.093	1.316	1.604	1.708	0.595	4.660
-0.5	β_1	0.368	0.567	-0.538	0.244	-1.284	0.907	-0.670	1.846
0.1	ϕ	0.008	1.062	0.126	0.738	0.4734	0.375	1.180	3.327
0.5	γ	-0.649	2.437	0.491	0.492	0.378	0.368	-0.577	5.151
1.0	β_0	-1.046	5.249	0.934	1.480	1.008	0.308	0.156	3.488
-0.5	β_1	-0.330	1.093	-0.628	0.279	-0.772	1.006	-0.589	2.740
0.9	ϕ	1.082	2.356	1.134	0.766	0.602	0.628	1.554	6.135
0.5	γ	0.094	1.523	0.433	3.132	0.491	0.585	-0.009	5.330

when estimation was done via GMM_3. However, the difference in sign for the temporal term is consistent across the levels of spatial dependence. This could be attributed to the fact that the set of moment conditions are first found independently of each year and based upon a set of linear combinations of the non-neighbors.

The initial value, MPL and EMPL, versus a vector of zeros has an impact on all of the parameter estimates with respect to the low spatial dependence case. The bias of MCEML increases when the initial values are not EMPL for the intercept, spatial and temporal terms in the model, whereas the bias for MCEML is similar when the initial values are EMPL or 0 for the spatial term. The standard error for MCEML also decreases when EMPL starting values are used ver-

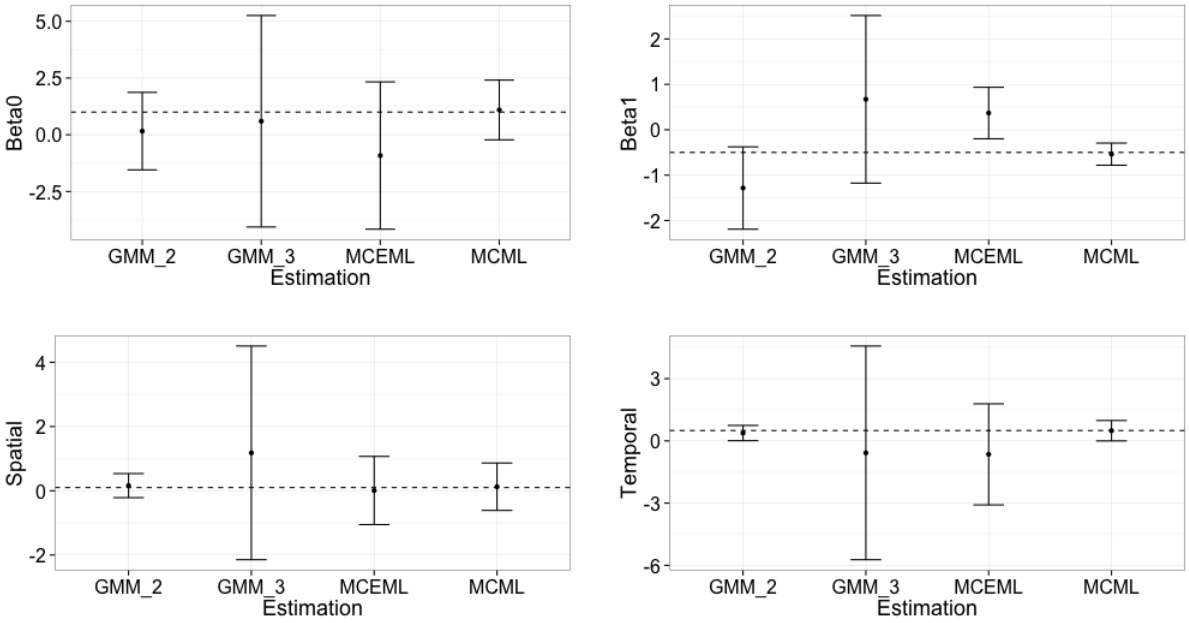


Figure 5. 5×5 grid parameter estimates and one standard deviation error bars with $\phi = 0.10$ and initial values EMPL or MPL.

sus a vector of zeros. Initial values also impact the bias of parameter estimates of GMM_2, the starting value using MPL led to more accurate estimates of the parameters. However, the standard errors for both sets of initial values did not change. This trend is not displayed across the estimation methods as bias and precision for GMM_3 did not change for the low or high spatial dependence cases and little changed for the MCEML parameter estimates. As can be seen from Figures 7 and 8, the effect the initial values have on MCML slightly increases the variability in the estimates for the temporal term, but the estimate displays very little bias.

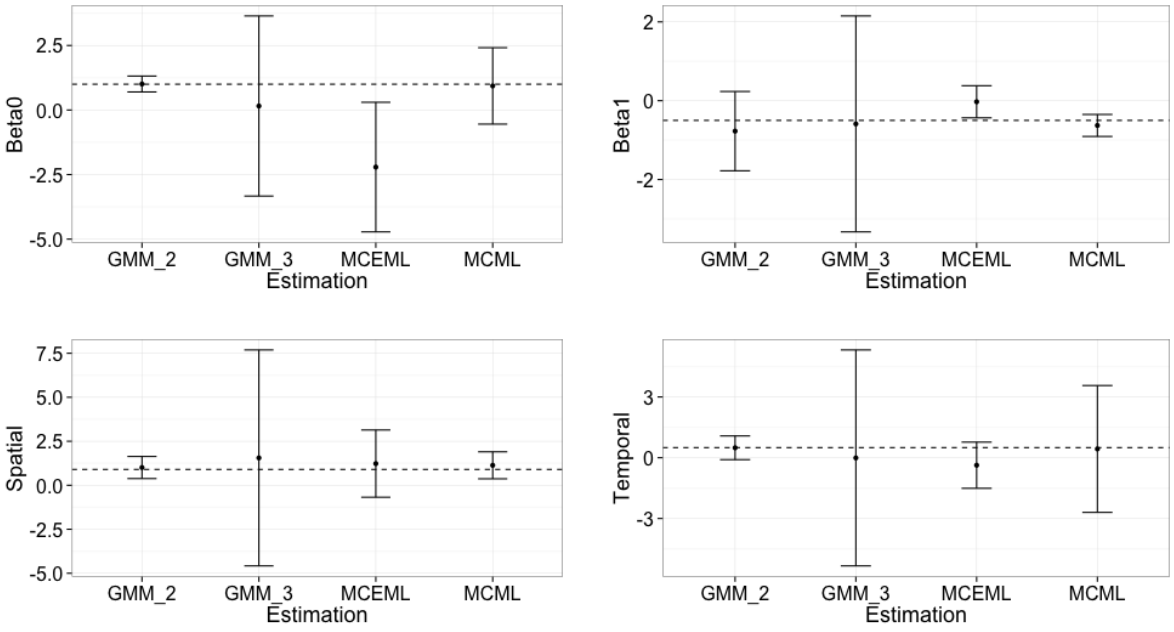


Figure 6. 5×5 grid parameter estimates and one standard deviation error bars with $\phi = 0.90$ and initial values EMPL or MPL.

Bias and Precision for the 10×10 ***Grid.*** In the 10×10 lattice the precision and parameter estimates for three different estimation methods GMM_3, MCML and MCEML are compared. Table 5, Table 6 and Figures 9-12 display the results for different levels of spatial dependence and initial values. As in the 5×5 lattice size, the MCML parameter estimates has very little bias as the estimates are very close to the true values with the exception of the spatial term for low spatial dependence. MCEML, on the other hand, did not perform well with low spatial dependence as it displays a lot of bias and variability. The only case that the MCEML estimate was within the one standard deviation range, for low spa-

Table 4

Parameter estimation with a vector of zeros for starting values on a 5×5 *grid with 5 time points,* $\phi = 0.10$ *and* $\phi = 0.90$.

		MCEML		MCML		GMM_2		GMM_3	
Actual	Covariate	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
1.0	β_0	-0.186	0.444	0.952	1.316	2.709	2.374	0.128	6.010
-0.5	β_1	-4.387	1.844	-0.429	0.244	-1.827	0.951	-0.781	1.852
0.1	ϕ	0.237	0.446	0.233	0.772	-0.823	0.473	1.270	4.771
0.5	γ	2.664	0.485	0.488	0.492	-0.299	3.625	-0.611	4.507
1.0	β_0	-2.210	2.51	0.934	1.480	1.008	0.308	0.156	3.488
-0.5	β_1	-0.027	0.408	-0.628	0.279	-0.772	1.006	-0.589	2.740
0.9	ϕ	1.230	1.910	1.134	0.766	0.602	0.628	1.554	6.135
0.5	γ	-0.370	1.140	0.433	3.132	0.491	0.585	-0.009	5.330

tial dependence, is for the spatial term. All other MCEML parameter estimates have either the wrong sign for the coefficient or are well outside the one standard deviation error bars. In the high spatial dependence case, however, MCEML does contain the true value within the one standard deviation error bars for all the coefficients, despite the amount of bias in the intercept and spatial term. MCEML estimation exhibits the same fluctuation in signs as the 5×5 case depending upon the level of spatial dependence. In the case of high spatial dependence, displayed in Table 5, Figures 10 and 12, the signs of the MCEML estimates are the same as the true values. However, when there is low spatial dependence the coefficients are opposite of the true values.

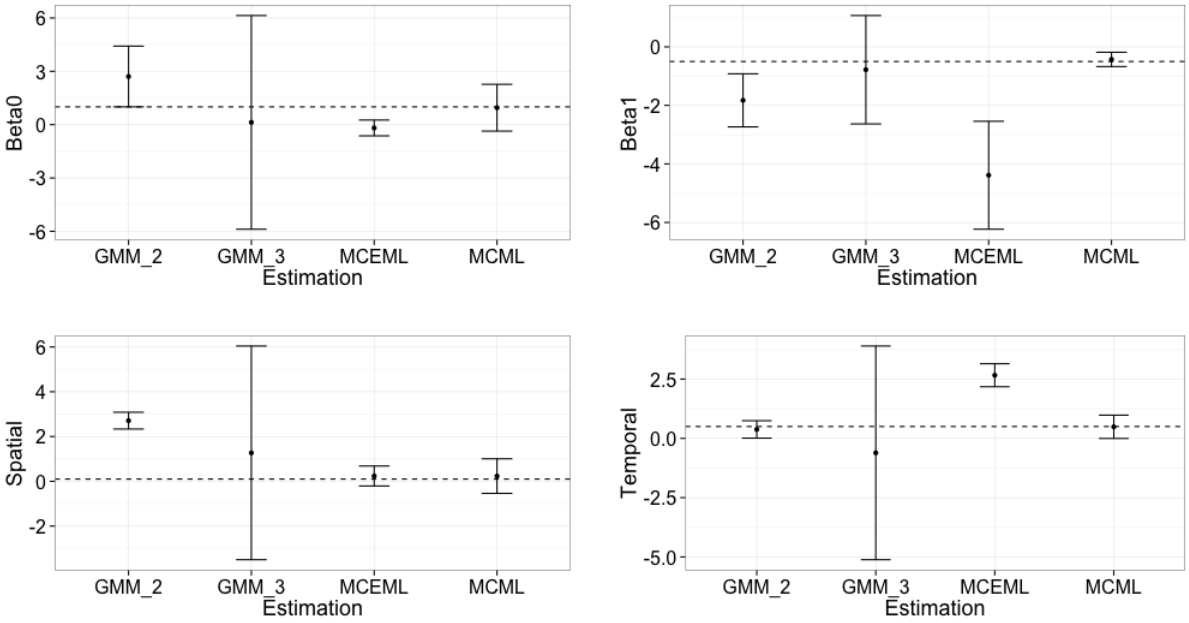


Figure 7. 5×5 grid parameter estimates and one standard deviation error bars with $\phi = 0.10$ and initial values vector of zeros.

GMM_3 estimation exhibits more variation in the low spatial dependence case than in the high spatial dependence case. In the low spatial dependence case, the parameter estimates display more bias for the spatial and temporal terms than the slope and intercept. However, the precision of the estimates are much greater than the 5×5 case. In the situation that there was high spatial dependence, the bias in the GMM estimates is reduced and the true parameter values are within the one standard deviation error bar. The sign of the coefficient for the temporal term differs from the true value for the temporal term, however as in the 5×5 case is consistent across the levels of spatial dependence.

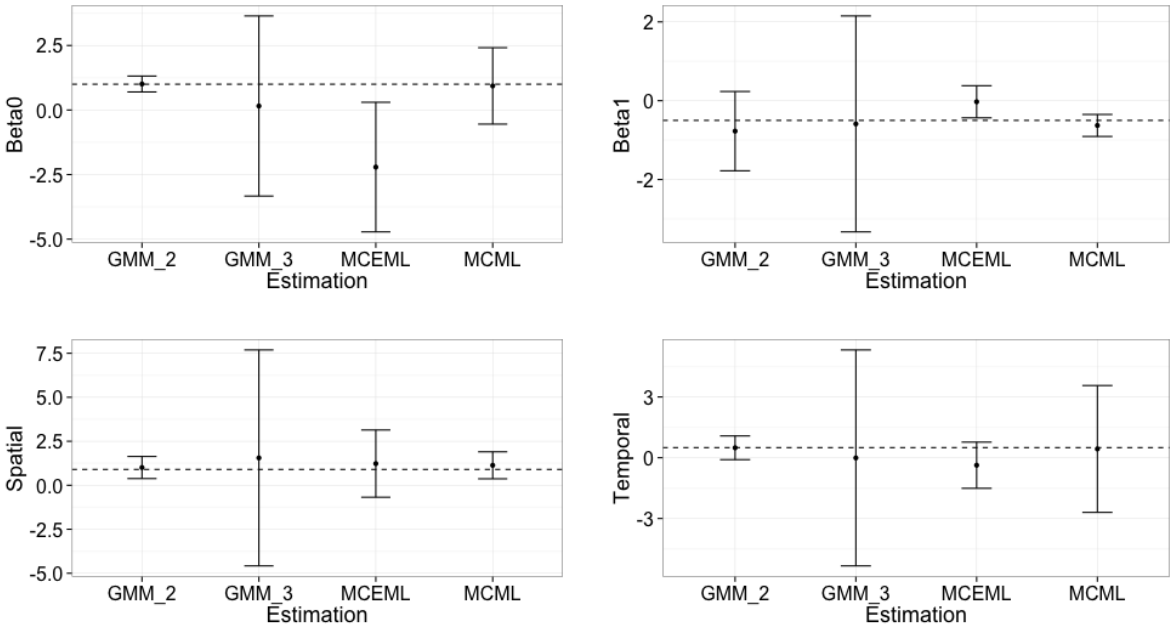


Figure 8. 5×5 grid parameter estimates and one standard deviation error bars with $\phi = 0.90$ and initial values vector of zeros.

The initial values, MPL, and EMPL, versus a vector of zeros have an impact on all of the parameter estimates with respect to the low spatial dependence case. Interestingly, the bias for MCEML decreases when the starting values are 0 for high spatial dependence. In the case of low spatial dependence and zero starting values, the greatest amount of bias occurs in the MCEML estimates for the spatial and temporal terms. In addition, the standard error bar ranges for MCEML do not change for the low spatial case but exhibit a decrease in size for the intercept and temporal term with high spatial dependence. Initial values do not have much of an effect on MCML or GMM_3 estimation with the exception of the spatial term. In

Table 5

Parameter estimation with starting values from MPL and EMPL on a 10×10 *grid with 5 time points.*

		MCEML		MCML		GMM_3	
Actual	Covariate	Estimate	SE	Estimate	SE	Estimate	SE
1.0	β_0	-0.426	0.693	0.999	0.813	1.106	0.653
-0.5	β_1	-0.164	0.166	-0.505	0.109	-0.560	0.284
0.1	ϕ	0.011	0.354	0.120	0.435	0.495	0.374
0.5	γ	-0.119	0.233	0.490	0.222	-0.284	0.774
1.0	β_0	-0.069	1.678	0.982	0.796	1.281	0.575
-0.5	β_1	-0.332	1.589	-0.508	0.124	-0.813	0.448
0.9	ϕ	0.384	0.757	0.921	0.332	0.915	0.550
0.5	γ	0.471	0.723	0.554	0.234	-0.203	0.341

both estimation methods, the standard error intervals widened for the low spatial dependence case but in the high spatial dependence case the intervals stay the same across the different initial values. Both methods display high precision and little bias as the true parameter estimates are close to the estimated values.

Simulation Estimation Summary

Overall, GMM_2 performs well for both levels of spatial dependence. The parameter estimates are close to the true values and are within one standard error of the true values when MPL is used for initial values. The performance of GMM_2 is worse when arbitrary starting values are used. GMM_2 shows a decrease in bias the

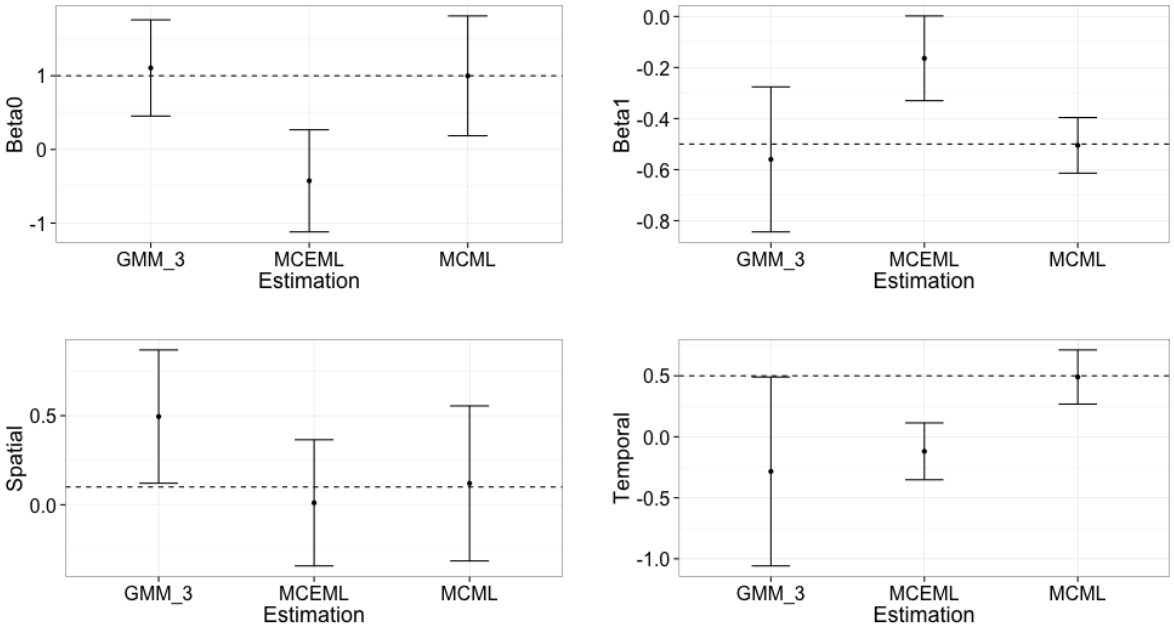


Figure 9. 10×10 grid parameter estimates and one standard deviation error bars with $\phi = 0.10$ and initial values EMPL or MPL.

greater the spatial dependence as well as a decrease in the variability in the estimates the larger the lattice. This suggests that the MCML and GMM_2 simulation results are comparable. GMM_3 tends to work better for large lattices and larger spatial dependence as the parameter estimates tend to be more accurate than for the low spatial dependence case. The initial values are important in estimating the parameters for MCEML. The direction of the signs change depending upon the initial starting values as well as for low and high spatial dependence levels. Further discussion is presented in Chapter 5.

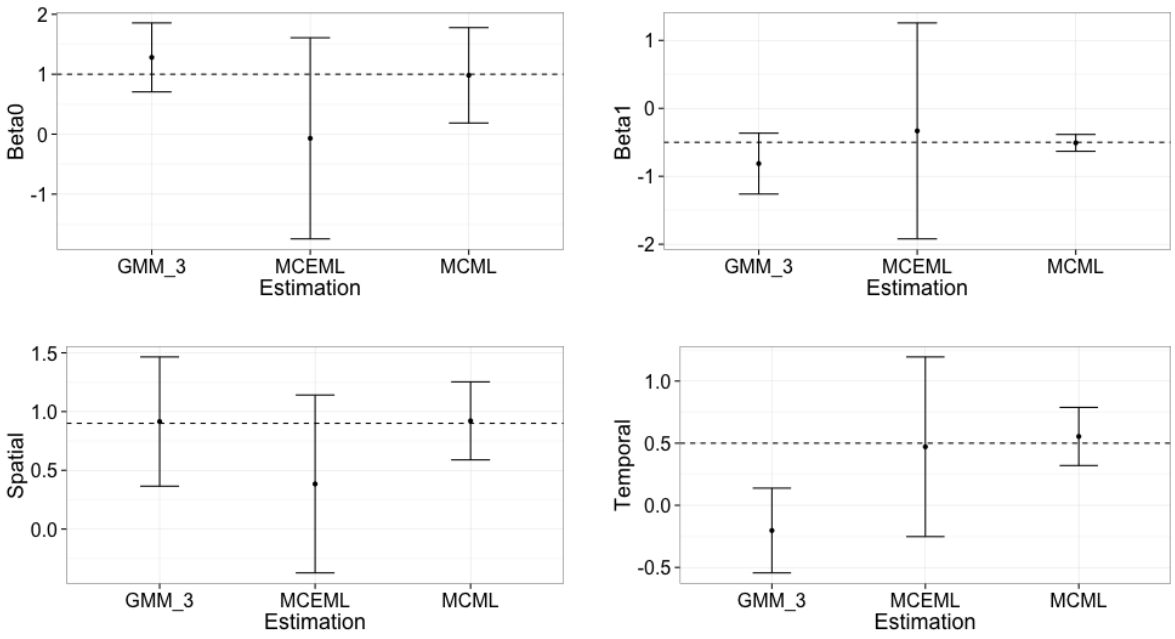


Figure 10. 10×10 grid parameter estimates and one standard deviation error bars with $\phi = 0.90$ and initial values EMPL or MPL.

Dataset Estimation Comparisons

Chapter 1 introduced two datasets chosen to compare GMM methods developed in Chapter 3 to MPL, MCEML, and MCML estimation. The data are tested using the modified *autologistic* function with added temporal component for MCEML, MCML, and GMM for binary spatial-temporal data.

In this dissertation, the National Longitudinal Study of Adolescent Health (Add Health) data and Rocky Mountain Pine Beetle data are used to demonstrate the different estimation techniques in terms of convergence, computation time and bias. The data are analyzed and compared via the different estimation methods.

Table 6

Parameter estimation with a vector of zeros for starting values on a 10×10 *grid with 5 time points,* $\phi = 0.10$ *and* $\phi = 0.90$.

		MCEML		MCML		GMM_3	
Actual	Covariate	Estimate	SE	Estimate	SE	Estimate	SE
1.0	β_0	0.042	0.248	0.974	0.822	1.181	0.711
-0.5	β_1	-0.367	1.586	-0.522	0.233	-0.532	0.277
0.1	ϕ	0.373	0.526	0.186	0.477	0.301	0.443
0.5	γ	0.483	0.706	0.446	0.311	-0.147	0.813
1.0	β_0	-1.720	8.729	1.022	0.796	1.197	0.602
-0.5	β_1	-0.485	1.918	-0.531	0.251	-0.785	0.513
0.9	ϕ	1.510	4.477	0.932	0.401	1.130	0.490
0.5	γ	-0.550	2.810	0.552	0.269	-0.244	0.434

The first method uses a stationary lattice structure and the second uses a non-stationary lattice where the spatial frame can change from year-to-year. The results of the two different datasets with respect to the estimation methods are described along with the datasets.

The first dataset, the Rocky Mountain Aerial Detection Survey (ADS) data, similar to Zhu et al. (2008) British Columbia mountain pine beetle data for the binary spatial-temporal autologistic model and Wang & Zheng (2013) southern pine beetle data for the centered spatial-temporal binary autologistic model are used to make comparisons among estimation methods. The data contains observations over space and time. The second dataset, the National Longitudinal Study of Adolescent

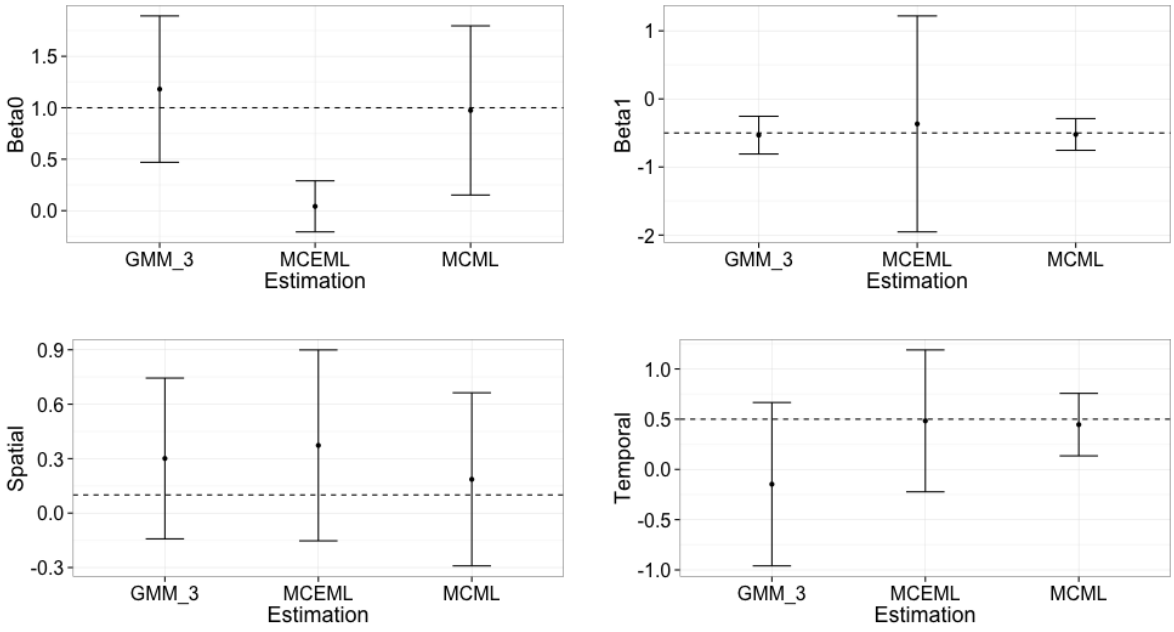


Figure 11. 10×10 grid parameter estimates and one standard deviation error bars with $\phi = 0.10$ and initial values vector of zeros.

Health (Add Health) data from the North Carolina Population Center is the motivation for the modified centered spatial-temporal binary autologistic model due to the change in peer network structure year-to-year. The estimation methods with different neighborhood structures are demonstrated and compared.

Rocky Mountain Forest Service Data Comparison

The Rocky Mountain Pine Beetle data is a public dataset collected by the United States Forest Service through aerial survey methods across the Rocky Mountain Region from 2001-2010. The data are composed of categories of damage-causing

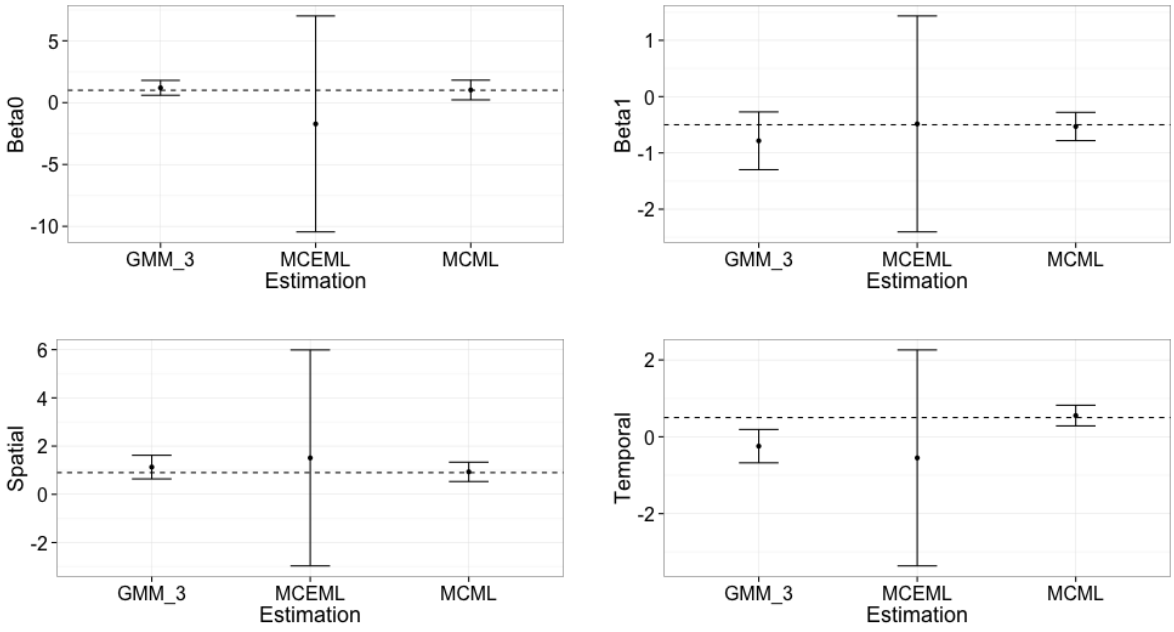


Figure 12. 10×10 grid parameter estimates and one standard deviation error bars with $\phi = 0.90$ and initial values vector of zeros.

agents, host tree species (trees in which the damage has taken place), number of dead trees, numbers of acres, surveyor identifications and numbers of dead trees per acre. A stationary grid of dimension 42×55 is constructed for each year. A representation of the study region is provided in Figure 13 from 2001 to 2010 and the binary representation of the data is displayed in Figure 14.

A binary response is noted based upon presence or absence of the damage-causing agent, bark beetle, at each spatial location. For each site i the August mean maximum temperature in degrees Celsius, January mean minimum temperature in degrees Celsius, mean annual precipitation in inches, and elevation in feet are cal-

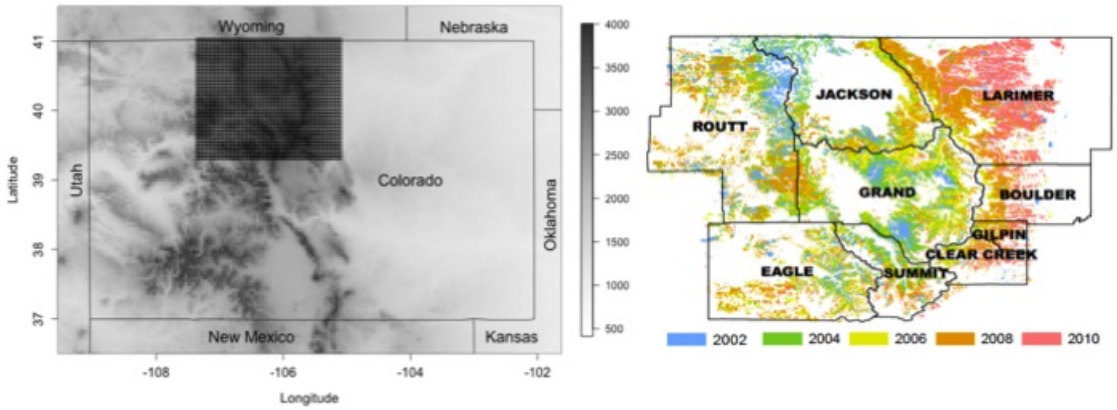


Figure 13. Study area in Colorado on a 42 by 55 grid, MPB Damage 2001-2010

culated. Weather variables are taken from the PRISM dataset, which is publicly available at `http://www.prism.oregonstate.edu/`. The Parameter-elevation Relationships on Independent Slopes Model (PRISM) data estimates monthly weather data over a contiguous grid at a resolution of 0.0416 decimal degrees latitude and longitude (∼4 km) cells (Daly et al., 2002) and aligns with the resolution of the gridded MPB data. The weather variables are adjusted to account for a one-year lag between infestation and the time MPB damage is detected in the ADS. That is, for August mean maximum and mean annual precipitation data is used from 1999 to 2008 and for January mean minimum temperatures data from 2000 to 2009.

The parameters of the mountain pine beetle data are estimated using likelihood-based methods, MCML, and MCEML and GMM_3. Table 7 provides the computation time using MPL, MCML, MCEML and GMM_3. For MCML and MCEML

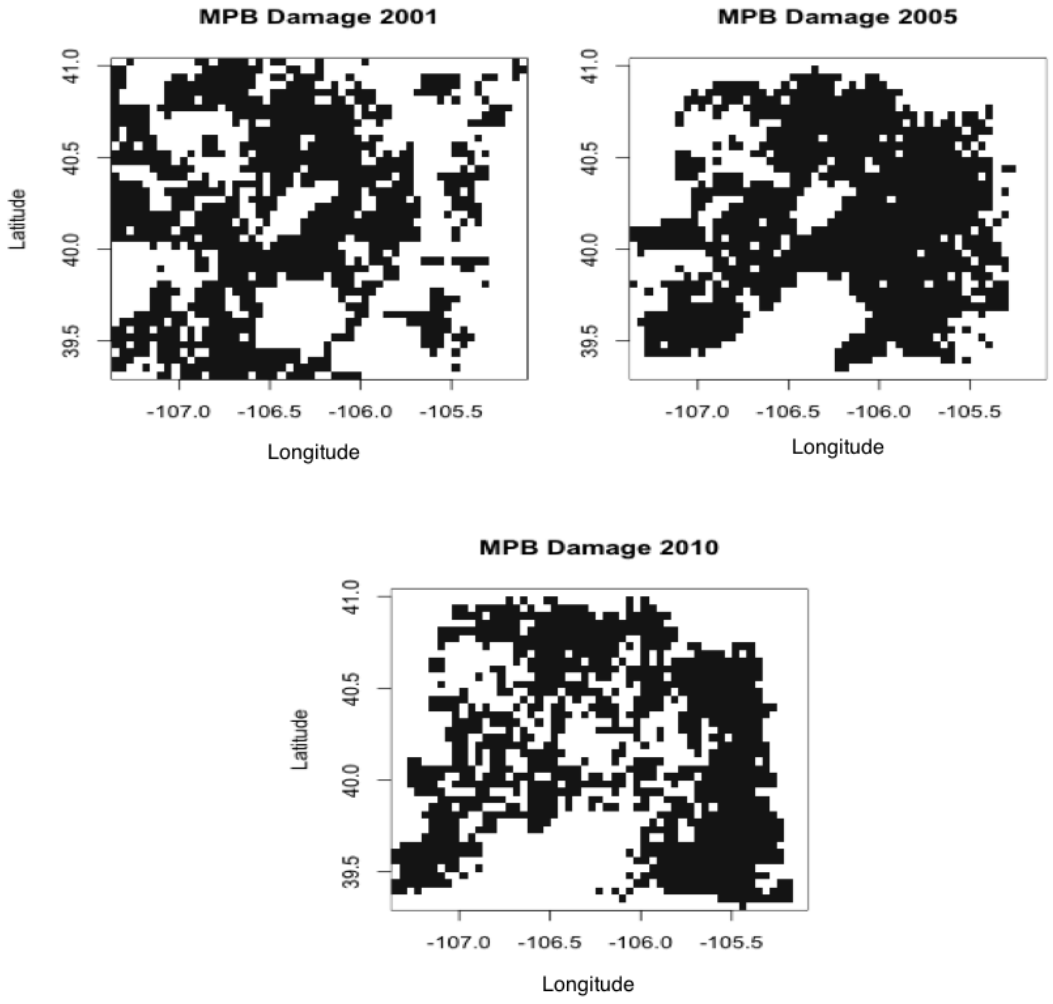


Figure 14. MPB Damage in years 2001, 2005, and 2010. Black grid cells represent MPB damage and white grid cells represent no damage.

the reference parameters are chosen from maximum psuedolikelihood. A total of 100,000 Monte Carlo samples are generated for each method with the first 1000 samples discarded for burn-in. The computation time between the different estimation methods vary. MCEML takes the longest computation time due to the ex-

tra steps associated with the model hitting a set tolerance. MCML is the fastest method taking about 1/3 the time as GMM_3 and 1/2 the time of MCEML. All of the methods converge and MCEML does not exceed the maximum iterations as it hits the specified tolerance before the maximum iterations are met.

Table 7

MPB data convergence and computation time comparisons

Estimation Method	Initial Value	Computation Time (sec)	Convergence (Y/N)
MCML	MPL	10944.32	Yes
MCML	(0,0,0,0)	15948.44	Yes
MCEML	EMPL	20335.01	Yes
MCEML	(0,0,0,0)	27582.26	Yes
GMM_3	MPL	15674.94	Yes
GMM_3	(0,0,0,0)	20671.27	Yes

Table 8 displays the parameter estimate results based upon different estimation techniques. The results suggest that MCML and GMM_3 have similar interpretations. In the MPB example, GMM_3 and MCML both suggest that as the January minimum temperatures increase the probability of mountain pine beetle outbreaks increases by about the same rate. The same result is noted in the mean August maximum temperatures. The spatial coefficient, ϕ, indicates that there is moderate spatial dependence among the sites as the probability of MPB damage within the grid cell increases and γ indicates a strong temporal dependence.

Table 8

Parameter estimation comparison for MPB Damage data, initial values MPL or EMPL

Covariate	MCML	GMM_3	MCEML
Intercept	-1.870	-0.720	-1.020
January Minimum Temperature	0.790	0.640	0.531
August Maximum Temperature	-0.050	-0.080	-0.002
Precipitation	-0.290	-0.420	-0.311
Elevation	-0.250	-0.270	-0.240
Spatial, ϕ	0.470	0.620	0.396
Temporal, γ	1.380	1.470	1.220

In general, the estimates are slightly larger for MCML but GMM_3 has very similar estimates. GMM_3 and MCML both indicate that as the precipitation increases the probability of mountain pine beetle damage decreases. All of the approaches suggest that there is a positive spatial and temporal dependence for the mountain pine beetle outbreaks. As there was high temporal dependence, MPL was a good initial value, but not the most appropriate estimation method (Huffer & Wu, 1998). Although the interpretation for the estimates using MCEML is the same as MCML and GMM_3, the computation time is much greater than the other two methods therefore MCEML is not be the preferred method.

In Table 9 parameter estimates are shown for initial estimates set to 0. Changing the initial value to a vector of 0's has an impact on all of the estimation meth-

Table 9

Parameter estimation comparison for MPB Damage data, initial values vector of zeros

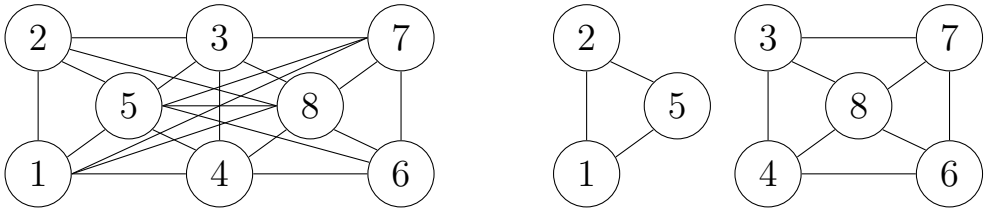
Covariate	MCML	GMM_3	MCEML
Intercept	-1.870	-1.562	-5.853
January Minimum Temperature	-0.095	0.237	0.328
August Maximum Temperature	0.592	0.455	0.327
Precipitation	-0.385	-0.267	0.147
Elevation	-1.026	0.486	-0.886
Spatial, ϕ	0.285	0.106	-0.592
Temporal, γ	0.538	0.732	0.433

ods. The signs on the coefficient change for the January minimum temperatures and August maximum temperatures for MCML inferring that the probability of mountain pine beetle outbreaks decrease as the minimum January temperature increases. However, MCML still indicates positive spatial and temporal dependence for the mountain pine beetle data. MCEML's interpretation of the estimates change for the August maximum temperatures, the probability of mountain pine beetle outbreaks increase as August minimum temperatures increase, opposite of the EMPL results. The largest difference in the MCEML results is denoted in the spatial term, it changes from positive spatial dependence to negative spatial dependence. GMM_3 also changes signs for the August maximum temperatures, but the interpretation for the spatial and temporal terms remain the same, positive spatial

and temporal dependence. Results of different starting values indicate that MPL and EMPL should be used as initial values when possible.

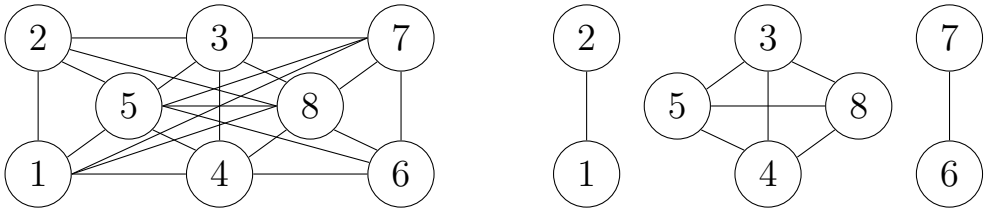
National Adolescent Health Data Comparison

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States assessed over three different time periods. The survey combines data on social, economic, psychological and physical well being based upon social contexts such as friendships, family, and neighborhoods. It contains detailed information about the respondents' characteristics and peer group networks over the course of several years. Three waves of data are available, Wave I, Wave II, and Wave III, which are from the years 1994 to 2002. The Add Health data is used to assess the effect of peer networks at multiple levels (friend and school) using a binary variable, smoke (yes or no) using a spatial neighborhood constructed based upon relationships one peer has to another. An example of peer network structures for the first and second wave of the Add Health data are displayed in Figure 16.



(a) School 1 Peers

(b) Grade Level Peers



(c) School 2 Peers (d) Grade Level Peers

Figure 15. Peer Network Spatial Framework

In this dissertation, Add Health data from the eastern region of the country consisting of about 500 observations are selected to evaluate peer relationships based upon friendships within grades. A participant is denoted as a neighbor if they identified subject i as a friend within the same grade, denoted as a reciprocal relationship, any other associations were denoted as non-neighbors. The spatial weight matrix was recalculated for each wave to account for different correlation structures that occur across waves (i.e. school relocation or ended friendship).

The modified centered spatial-temporal binary autologistic model specified in Chapter 3 is fit using three different estimation techniques, MCEML, MCML and GMM_3. The modified model is applied due to the change in spatial frame from year-to-year. Initial starting values are also assessed in this data example as interpretation of the parameters may vary due to the modified model. The variables that are chosen in the model, friends smoking (yes/no), and gender of the student (male/female), are selected based upon the idea that peer influences could influence the probability an adolescent smokes.

Table 10 displays the computation time and convergence for different initial starting values for MCEML, MCML and GMM_3. The results show that all of the methods converge with respect to the different starting values. The computation time of MCML and GMM_3 are shown to be comparable, this is due to the number extra number of Monte Carlo samples generated for MCML. An interesting discovery is that MCEML suggests to be the fastest method in this data situation especially with respect to initial values of $(0, 0, 0, 0)$. The computation time shows little change between initial values for MCML and GMM_3.

Table 10

Add Health data convergence and computation time comparisons

Estimation Method	Initial Value	Computation Time (sec)	Convergence (Y/N)
MCML	MPL	6182.07	Yes
MCML	(0,0,0,0)	6112.84	Yes
MCEML	EMPL	2539.40	Yes
MCEML	(0,0,0,0)	500.03	Yes
GMM_3	MPL	6105.21	Yes
GMM_3	(0,0,0,0)	6097.18	Yes

The results of the parameter estimates are displayed in Table 11 and Table 12 with different initial starting values. The parameter interpretations using MPL and EMPL as starting values are similar for GMM_3 and MCML. The interpretation for the effect of friends are different for MCEML compared to MCML and GMM_3. MCEML infers that the probability of an adolescent smoking increases when friends

smoke, where MCML and GMM_3 both agree that the probability of an adolescent smoking decreases when friends smoke. Although the interpretation is the same across the estimation method for gender, the probability of an adolescent smoking decreases if one is male, the spatial terms do not agree. MCML and GMM_3 indicate a positive friend within grade peer association for smoking while MCEML indicates that there is a negative friend within grade peer association. The probability that an adolescent smokes has the same interpretation for MCML and GMM_3, but draw different conclusions with MCEML estimation.

Table 11

Parameter estimates for the binary variable smoke (yes=1, no=0) initial values MPL or EMPL

Covariate	MCML	GMM_3	MCEML
Intercept, β_0	0.024	0.097	0.0120
Friends Smoke, β_1	-0.066	-0.026	0.029
Gender (Male), β_2	-0.032	-0.075	-0.014
Friends(Grade), ϕ	0.092	0.219	-0.220
Temporal, γ	2.520	2.397	2.738

The results from Table 12 show that the initial starting value has a large impact on MCEML. The parameter estimates do not change much from the initial estimates of $(0, 0, 0, 0)$. This is likely due to the estimates hitting the specified tolerance right away, which explains why the computation time is greatly reduced in this case. GMM_3 also suggests to be sensitive to initial values. The interpretation

of the coefficients change based upon the initial starting values. In the case where MPL values are used friends have less influence on smoking whereas friends smoking increases the probability of smoking when the parameter values are $(0, 0, 0, 0)$. Similar results occur for temporal and gender coefficients in the model.

Table 12

Parameter estimates for the binary variable smoke (yes=1, no=0) initial values vector of zeros.

Covariate	MCML	GMM_3	MCEML
Intercept, β_0	0.263	0.054	0.000
Friends Smoke, β_1	0.068	0.023	0.000
Gender (Male), β_2	-0.295	0.010	0.000
Friends(Grade), ϕ	0.070	0.039	0.000
Temporal, γ	2.520	-0.030	0.100

As can be seen in Table 12, the initial values have an impact on the interpretation of the estimates across the estimation methods. For example, the coefficients for MCML change signs for friends smoking but the spatial and temporal coefficients stay the same. The coefficients for GMM_3 also change sign for both friends smoking and gender. If the initial starting values were zero starting values, influences on the probability of an adolescent smoking can be incorrect. Despite the differences in the friends smoking and gender coefficients for GMM_3 and MCML, both suggest that there is a small positive peer relationship of friends within the same grade. MCEML does not demonstrate much change from the initial start-

ing values even with a tolerance set to 0.001 for the EM steps. The small estimates seen from MPL and EMPL starting values likely impacted the MCEML results as it also converges fast indicating the tolerance is met quickly. Since the interpretation based upon initial starting values differs across the methods, there is not strong evidence that GMM_3 has a better interpretation than MCML when the starting values are misspecified.

Estimation Methods Summary

In general, estimation by MCEML took twice as long as GMM_3 and MCML for the Rocky Mountain Forest Service data, a large spatial grid. However, MCEML is computationally faster for the Add Health data but the interpretation of the estimates differ from MCML and GMM_3. In both data examples the interpretations of the parameter estimates for MCML and GMM_3 are the same. The estimates for MCEML are only similar to MCML and GMM_3 in the Rocky Mountain Forest Service data. Convergence is not a problem for any of the estimation techniques, although it is important to point out that MCML and MCEML need to have a reference value that is close to the true value. In the data examples specified above MPL values are used as a reference parameter. However, MCEML estimates can diverge to infinity if there is strong spatial and/or temporal dependence.

CHAPTER V

CONCLUSIONS

The generalized method of moments estimation approach developed in this dissertation used properties of Markov Random fields, spatial-temporal properties from the centered spatial-temporal binary autologistic model, and characteristics of stationary and non-stationary data. The methodology developed for the generalized method of moments approaches focused on the centered spatial-temporal binary autologistic model and a modified spatial-temporal model where the neighborhood structure changes each year. Three different generalized method of moments approaches were constructed, a traditional generalized method of moments that creates a set of moment conditions for every combination of non-neighbors and parameters, a reduced neighbor approach which only created moment conditions if the non-neighbors were within a certain spatial neighborhood, and an approach that sums the moment conditions over all the non-neighboring sites.

The literature on spatial-temporal binary models showed that work has not been done on alternative estimation methods using generalized method of moments for spatial-temporal binary data on a stationary and non-stationary grid. Likelihood-based estimation methods i.e. maximum pseudolikelihood and Markov

Chain Monte Carlo maximum likelihood are well known in the spatial-temporal field unlike generalized method of moments. There is a need to explore alternative estimation methods as likelihood-based methods are computationally expensive for large lattices due to the need to generate Monte Carlo samples. They can also have convergence failure or be sensitive to initial values. A few articles have investigated a spatial only GMM technique using marginal models but have not included a temporal term.

The GMM approaches developed in this dissertation have two major advantages over spatial GMM methods that have been developed. First, GMM has only been developed for spatial models with respect to the spatial generalized linear mixed model and not a spatial-temporal model. The GMM approaches developed in this dissertation use the conditional spatial-temporal centered binary autologistic model. This model captures the spatial and temporal dependence directly therefore the natural structure of binary data is captured through neighbor definitions. Second, the working correlation structure for the GMM approach is specified based on a set of moment conditions constructed from combinations of the non-neighbors of a particular site and uncorrelated with the systematic component of the model which accounts for both spatial and temporal dependence in the binary responses.

The alternative approaches using generalized method of moments estimation developed in this dissertation provide methods for analyzing binary spatial-temporal data without the use of Monte Carlo estimation. The research conducted focused on the centered spatial-temporal binary autologistic model, a conditional

model where each site is conditioned on all the neighboring sites and the current year's value depends upon the previous years' value. The centered spatial-temporal binary autologistic model was chosen due to its ability to better interpret covariates in the model, which are common in ecological and health data.

To answer the first research question, how generalized method of moments is specified, three different generalized method of moments approaches were proposed based upon varying lattices sizes. The first generalized method of moments approach, a more traditional approach, was found not to be feasible for moderate to large lattices as the inversion of the weight matrix with respect to the moment conditions was too large to hold in memory on a personal computer. Two other alternative approaches were constructed and tested on 1000 simulated datasets to assess convergence, bias and precision in the estimates. These two GMM approaches were specified with respect to the size of the spatial lattice and varying levels of spatial dependence.

A simulation study was conducted to answer the second research question, comparing two different GMM approaches, GMM_2 and GMM_3, to existing estimation methods, MCML and MCEML. Comparisons among the three different estimation methods were made using low and high spatial dependence values, different initial values, and two differently sized lattices. The estimation method results varied depending upon the level of spatial dependence and initial starting values. Research questions 2-5 found that alternative GMM methods are useful in certain data situations. In the case that the data had low spatial dependence and the lat-

tice size was small GMM_2 showed to have a small bias and smaller standard errors than other estimation methods. Even though GMM_2 was found to be computationally slower than the other estimation techniques it converged for 95% of the data sets and showed comparable results to MCML. In a data situation where there is high spatial dependence and a small lattice size, GMM_2 also had the least amount of variability and bias for the majority of the parameter estimates. The specification of the moment conditions for each non-neighboring site, using GMM_2 resulted in the least amount of bias for the spatial component. In high spatial dependence situations, the neighboring sites have more weight on the current site than non-neighboring sites. By constructing moment conditions based upon non-neighboring sites within a specific order of spatial neighborhood, GMM_2 can weigh the spatial dependence appropriately. If computation cost is of concern, then MCML is a viable alternative due to its low variability and bias. If the data have low spatial dependence and the lattice is moderate in size, MCML showed to be the preferred method over GMM_3, however as computer power becomes less of a problem, GMM_2 would be a good alternative. Lastly, if the lattice is large and the spatial dependence is high, MCML and GMM_3 were shown to be comparable and are reasonable choices as all the parameter estimates were within one standard error of the true parameter value.

In terms of initial values, if the initial values are unknown or there is not a good starting point, the results of the simulation showed that GMM_3 did not perform as well as MCML, but is a better choice than MCEML. Estimation by

MCEML showed that the direction of the signs changed depending upon the level of spatial dependence. The GMM_3 estimates with different starting values were more biased than MCML however, most of the coefficients were within one standard error of the true parameter estimate. Therefore, in certain cases GMM performs as well as MCML and is less sensitive to initial values compared to MCEML. This result was also verified in the additional two data sets used to compare estimation techniques.

The Rocky Mountain Forest Service data models mountain pine beetle outbreaks over space and time. The dataset is similar to the British Columbia mountain pine beetle data used for spatial-temporal binary autologistic model (Zheng & Zhu, 2008). The Adolescent Health data provided by the North Carolina Population Research Center, was used to model peer network structures based upon levels of relationships. In general, estimation by MCEML took twice as long as GMM_3 and MCML when the spatial grid was large in size as in the mountain pine beetle data. When the lattice was moderate in size as in the Add Health data, MCEML was computationally faster however the interpretation of the estimates differed from MCML and GMM_3. In both data examples the interpretation of the parameter estimates for MCML and GMM_3 were similar. The Add Health data added an additional component to the centered spatial-temporal binary autologistic model, a moving grid where different spatial weights could be assigned each year due to peer network structures. The results comparing the different estimation methods were similar when the starting values were based upon MPL or EMPL. The parame-

ter estimates varied more when the starting value was 0 however, MCEML struggled the most. When the parameter estimates were found using traditional starting values, MPL and EMPL, the values were relatively small with the exception of the temporal component. The fact that the final parameter estimates based upon MCML and GMM_3 were close to zero indicates that when starting values of 0 were used for MCEML estimation the difference in the estimates from one step to the next were very small. Therefore, the tolerance limit was met within the first few iterations of the MCEML algorithm giving final estimates the same as or close to the starting values. Although the tolerance limit was arbitrarily chosen as 0.001, this may not have been suitable for the Add Health example due to the values of the parameter estimates.

Overall, initial starting values did not impact the convergence or computation time for GMM_2 and GMM_3 but were slightly more variable in the parameter estimates when the initial values were far away from the true values. While GMM_3 showed the most variation in the parameter estimates for the simulated data, it performed well for the two real data sets. It performed just as well as MCML and in the simulation, took approximately the same computation time. MCML's computation time is impacted by the amount of burn-in and Monte Carlo chains sampled, whereas the both GMM_2 and GMM_3 have a more consistent computation time as GMM does not use Monte Carlo methods. An explanation for why the GMM_3 standard errors were so large may be due to the large reduction of the number of moment conditions. In GMM_2, the moment conditions were constructed from

all the non-neighboring sites within a specified neighborhood order, therefore accounting for the individual non-neighbor effect. GMM_3, on the other hand, created moment conditions based upon the sum of all the non-neighboring sites potentially creating more variability between the moment conditions ultimately affecting the standard errors.

In the GMM methods specified in this dissertation, the working correlation structure may be misspecified causing the temporal term in GMM_3 to be opposite than the true value. Consideration of alternative estimation techniques may be given for modeling spatial-temporal binary data using quadratic inference functions (Qu et al., 2000). Estimation by quadratic inference functions does not involve direct maximization of the correlation in the parameters and is optimal if the working correlation structure is misspecified. It uses linear combinations of a set of basis matrices constructed used for the inverse of the working correlation matrix, rather than moment conditions.

The moving grid constructed for the peer network structure for the Add Health data is also another area for future investigation. Although the moving grid appeared to be effective for relating friendships from one year to another, simulation using random graphs can help verify the moving grid's efficiency. As the spatial structure changes for each year a new random graph can be constructed for each year and then tied back to the previous years response using the minimum Euclidean distance to the current year's response.

Another area for investigation is the extension of the centered spatial-temporal autologistic model for multinomial response data. The Rocky Mountain data has multiple categories of damage, which lends itself to spatial-temporal multinomial modeling. In this model, the responses can be nominal or ordinal where the groups or categories can have different spatial dependence structures. As the spatial dependence of the categories may be of importance separate spatial effects will be accounted for in the model.

Overall, the generalized method of moments approaches presented in this dissertation produced similar results to MCML estimation and displayed lower bias and shorter computation time then MCEML estimation. The outcomes of these GMM approaches for spatial-temporal data encourage further work in this area. GMM was constructed based upon a conditional model where a site's observation was conditional on all the neighboring sites. However, GMM is generally constructed for marginal models. A different approach for GMM estimation for the spatial-temporal binary model is to use a marginal model. In this type of model the spatial and temporal dependence will be captured indirectly using a latent spatial structure or spatial correlation from kriging estimates. The potential model could account for the negative temporal value that occurs in the simulated data for GMM_3 by accounting for space and time in the correlation structure.

The development of generalized method of moments in this dissertation covered three different approaches. The outcome of the methods provides avenues for further work on how the moment conditions can be specified for a spatial-temporal

binary model especially as computation time increases. The methodology used for GMM_2 and GMM_3 indicate that even though GMM estimation depends on the type of data situation, likelihood-based methods produced similar results. However, computation of GMM methods is easier to implement in statistical software compared to MCML and MCEML as it does not need the additional Monte Carlo computation. The generalized method of moments methods developed in this dissertation creates additional options to evaluating spatial-temporal binary data. A possible solution to this problem is to combine the reduced non-neighborhood structure from GMM_2 and pairing combinations of the summations of all the non-neighbors for each site from GMM_3. To reduce the number of moment conditions, values that have low spatial weight can be specified by GMM_3 methods whereas sites that have high spatial dependence can be specified by a spatial order of non-neighbors as in GMM_3. The total number of moment conditions will be reduced and may help account for the negative values in the temporal effect for GMM_3.

REFERENCES

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*, 192–236.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, *24*, 179–195.

Besag, J. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *32*(1), 1–20.

Besag, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*, 75–83.

Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, *51*(3/4), 481–483.

Cressie, N. A. (1993). *Statistics for Spatial Data, revised edition*. Wiley, New York.

Cressie, N. A. & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken, New Jersey.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate Research*, *22*, 99–113.

Gelfand, A., Diggle, P. J., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*. CRC press.

Geyer, C. (1994). On the convergence of monte carlo maximum likelihood calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*, 261–274.

Gotway, C. A. & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 157–178.

Greene, W. H. (2002). *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Gu, M. G. & Zhu, H.-T. (2001). Maximum likelihood estimation for spatial models by markov chain monte carlo stochastic approximation. *Journal of the Royal Statistical Society. Series B (Methodological)*, *63*, 339–355.

Gumpertz, M. L., Graham, J. M., & Ristaino, J. B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*, 131–156.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer, New York.

Hammersley, J. M. & Clifford, P. (1971). Markov fields on finite graphs and lattices. *Unpublished*.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.

Hansen, L. P. (2001). *Generalized Method of Moments Estimation: A Time Series Perspective (published title "Method of Moments") Eds. of Methodology: Statistics, International Encyclopedia of the Social and Behavior Sciences*. Pergamon: Oxford.

Huang, F. & Ogata, Y. (2002). Generalized pseudo-likelihood estimates for markov random fields on lattice. *Annals of the Institute of Statistical Mathematics*, *54*(1), 1–18.

Huffer, F. W. & Wu, H. (1998). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, *54*, 509–524.

Hughes, J. & Cui, X. (2013). *ngspatial: Classes for Spatial Data*. Minneapolis, MN. R package version 1.0-3.

Hughes, J., Haran, M., & Caragea, P. C. (2011). Autologistic models for binary data on a lattice. *Environmetrics*, *22*(7), 857–871.

Kaiser, M. S. & Caragea, P. C. (2009). Exploring dependence with data on spatial lattices. *Biometrics*, *65*(3), 857–865.

Kaiser, M. S. & Cressie, N. (1997). Modeling poisson variables with positive spatial dependence. *Statistics and Probability Letters*, *35*, 423–432.

Klier, T. & McMillen, D. P. (2008). Clustering of auto supplier plants in the united states. *Journal of Business & Economic Statistics*, *26*(4), 460–471.

Laslett, G., McBratney, A., Pahl, P., & Hutchinson, M. (1987). Comparison of several spatial prediction methods for soil ph. *Journal of Soil Science*, *38*(2), 325–341.

Lee, L.-F. (2007). The method of elimination and substitution in the gmm estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *140*(1), 155–189.

Lee, L.-F. & Liu, X. (2010). Efficient gmm estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*, *26*(01), 187–230.

Lin, P.-S. (2010). Estimating equations for separable spatial-temporal binary data. *Environmental and Ecological Statistics*, *17*, 543–557.

Lin, P.-S., Lee, H.-Y., & Clayton, M. (2009). A comparison of efficiencies between quasi-likelihood and pseudo-likelihood estimates in non-separable spatial-temporal binary data. *Journal of Statistical Planning and Inference*, *139*, 3310–3318.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*(8), 1246–1266.

Mitchell, R. G. & Preisler, H. K. (1991). Analysis of spatial patterns of lodgepole pine attacked by outbreak populations of the mountain pine beetle. *Forest Science*, *37*, 1390–1408.

Møller, J. (1999). Perfect simulation of conditionally specified models. *Biometrika*, *93*(2), 451–458.

Møller, J., Pettitt, A. N., Reeves, R., & Berthelsen, K. K. (2006). An efficient markov chain monte carlo method for distributions with intractable nor-

malising constants. *Journal of the Royal Statistical Society, Series B, Methodological*, *61*, 251–264.

Negron, J. F. & Popp, J. B. (2004). Probability of ponderosa pine infestation by mountain pine beetle in the colorado front range. *Forest Ecology and Management*, *191*, 17–27.

Pinkse, J. & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, *85*(1), 125–154.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, *9*(1-2), 223–252.

Qu, A., Lindsay, B. G., & Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, *87*(4), 823–836.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Schabenberger, O. & Gotway, C. A. (2004). *Statistical methods for spatial data analysis*. CRC Press.

Wang, Z. & Zheng, Y. (2013). Analysis of binary data via a centered spatial-temporal autologistic regression model. *Environmental and Ecological Statistics*, 1–21.

Waring, R. H. & Pitman, G. B. (1985). Modifying lodgepole pine stands to change susceptibility to mountain pine beetle attack. *Ecology*, *66*, 889–897.

Whittle, P. (1954). On stationary processes in the plane. *Biometrica*, *41*, 434–449.

Zheng, Y. & Zhu, J. (2008). Markov chain monte carlo for a spatial-temporal autologistic regression model. *Journal of Computational and Graphical Statistics*, *17*(1), 123–137.

Zhu, J., Huang, H.-C., & Wu, J. (2005). Modeling spatial-temporal binary data using markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics*, *10*(2), 212–225.

Zhu, J., Zheng, Y., Carroll, A. L., & Aukema, B. H. (2008). Autologistic regression analysis of spatial-temporal binary data via monte carlo maximum likelihood. *Journal of Agricultural, Biological, and Environmental Statistics*, *13*(1), 84–98.

APPENDIX A

DERIVATION OF THE NEGPOTENTIAL FUNCTION

The negpotential function for the centered spatial-temporal autologistic model in a regression setting is derived. In this derivation, we assume pairwise-only dependencies, i.e. that the underlying graph has clique number 2.

Let the conditional density for Y be given by a Bernoulli distribution,

$$f_{it}(Y_{it}|Y_{t(-i)}, \boldsymbol{\theta}) = p_{it}^{Y_{it}}(1-p_{it})^{1-Y_{it}} \tag{102}$$

where the probability for site i is

$$p_{it} = \frac{\exp(\sum_{k=0}^{K}\beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L}\phi_l \sum_{\mathcal{N}_i}(Y-\mu)_{jt}\right] + \sum_{s=1}^{S}\gamma_s(Y-\mu)_{i,t-s})}{1+\exp(\sum_{k=0}^{K}\beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L}\phi_l \sum_{\mathcal{N}_i}(Y-\mu)_{jt}\right] + \sum_{s=1}^{S}\gamma_s(Y-\mu)_{i,t-s})}. \tag{103}$$

and the probability over all sites is

$$\boldsymbol{p}_t = \frac{\exp(\boldsymbol{X_t\beta} + \boldsymbol{\phi A(Y_t-\mu_t)} + \boldsymbol{\gamma(Y_{t-s}-\mu_{t-s})})}{1+\exp(\boldsymbol{X_t\beta} + \boldsymbol{\phi A(Y_t-\mu_t)} + \boldsymbol{\gamma(Y_{t-s}-\mu_{t-s})})}.$$

The center $\boldsymbol{\mu}_t$ is the the probability of $Y=1$ under independence,

$$\boldsymbol{\mu}_t = \frac{\exp\{\boldsymbol{X\beta}\}}{1+\exp\{\boldsymbol{X\beta}\}}. \tag{104}$$

The joint distribution over $i = 1, \cdots I$ sites for a given time point, t is defined by the Hammersley-Clifford theorem (Cressie, 1993) as follows. Suppose that the probability structure is only dependent upon contributions from cliques containing no more than two sites then the negpotential function (Besag, 1974) is

$$Q(Y) = \sum G(Y_{it}) + \sum G_{itjt}(Y_{it}, Y_{jt}), \tag{105}$$

where $G(Y_{it}) \equiv 0$ unless the sites i and j are neighbors as there are pairwise-only dependencies between sites. For some suitable $\boldsymbol{Y^*} \in \boldsymbol{\Omega}$,

$$G_{it}(Y_{it}) = \log\left(\frac{f_{it}(Y_{it}|\boldsymbol{Y}^*_{jt})}{f_{it}(Y^*_{it}|\boldsymbol{Y}^*_{jt})}\right) \tag{106}$$

and

$$G_{ijt}(Y_{it}, Y_{jt}) = \mathbf{1}_{i\sim j} \log\left(\frac{f_{it}(Y_{it}|Y_{jt}, \boldsymbol{Y}^*_{it,jt}) f_{it}(Y^*_{it}|\boldsymbol{Y}^*_{it})}{f_{it}(Y^*_{it}|Y_{jt}, \boldsymbol{Y}^*_{it,jt}) f_{it}(Y_{it}|\boldsymbol{Y}^*_{it})}\right) \tag{107}$$

based upon the factorization theorem. As the negpotential is a log-likelihood, the log of the conditional probability distribution is

$$\begin{aligned}
\log f_{it}(Y_{it}|Y_{-it}, \boldsymbol{\theta}) &= Y_{it} \ln p_{it} + (1 - Y_{it}) \ln(1 - p_{it}) \\
&= Y_{it} \ln p_{it} + \ln(1 - p_{it}) - Y_{it} \ln(1 - p_{it}) \\
&= Y_{it} \ln\left(\frac{p_{it}}{1 - p_{it}}\right) + \ln(1 - p_{it}) \qquad (108) \\
&= Y_{it}\eta_{it} - \ln(1 + \exp\{\eta_{it}\}) \qquad (109)
\end{aligned}$$

where $\eta_{it} = \sum_{k=0}^{K} \beta_k X_{itk} + \frac{1}{2}\left[\sum_{l=1}^{L} \phi_l \sum_{\mathcal{N}_i} Y^*_{jt}\right] + \sum_{s=1}^{S} \gamma_s Y^*_{i,t-s}$. The equations (108) and (109) can be shown to be equivalent using a logit transformation where $\ln\left(\frac{p_{it}}{1-p_{it}}\right) = \eta_{it}$ and $\ln(1 - p_{it}) = \frac{1}{1+\exp\{\eta_{it}\}}$. The G functions can now be defined using Equation (108), $\boldsymbol{Y^*} = \mathbf{0}$ and substituting η_{it} into the function as

$$G_{it}(Y_{it}) = Y_{it}\left(\sum_{k=0}^{K} \beta_k X_{kit} + \sum_{s=1}^{S} \gamma Y_{i,t-s}\right), \tag{110}$$

and

$$G_{ijt}(Y_{it}, Y_{jt}) = \frac{1}{2}\left[Y_{it}\left(\sum_{l=1}^{L} \phi \sum_{j\in\mathcal{N}_i^{(l)}} Y_{jt}\right)\right]. \tag{111}$$

Now, the negpotential function is

$$Q(\boldsymbol{Y}|\boldsymbol{\theta}) = \sum G(Y_{it}) + \sum G_{ijt}(Y_{it}, Y_{jt}) \tag{112}$$

$$= c(\boldsymbol{\theta})^{-1} \exp\left(\boldsymbol{Y_t'X_t\beta} + \frac{\phi}{2}\boldsymbol{Y_t'AY_t} + \gamma\boldsymbol{Y_t'Y_{t-s}}\right) \tag{113}$$

The joint distribution over $i = 1, \cdots I$ sites for a given time point, t based upon the Hammersley-Clifford theorem (Cressie, 1993) is

$$p(\boldsymbol{Y}_t|\boldsymbol{Y}_t') = \frac{\exp(Q(\boldsymbol{Y}_t))}{\sum_{Z\in\Omega}\exp(Q(\boldsymbol{Z}_t;\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\gamma}))} \tag{114}$$

$$= c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\beta}, \phi, \gamma)^{-1} \times \exp\sum_{i=1}^{I} Y_{it}\eta_{it} \tag{115}$$

where $c(\boldsymbol{Y}_{t-1}, \cdots, \boldsymbol{Y}_{t-S} : \boldsymbol{\beta}, \phi, \gamma)^{-1}$ is an unknown normalizing constant.

APPENDIX B

SAMPLING METHODS

Perfect Sampling

Perfect sampling, developed by Propp & Wilson (1996), is a common method used in the case where estimation of parameters is difficult such as dealing with a normalizing constant as in the autologistic model. The algorithm draws an exact sample during each iteration, increasing the accuracy of the inference compared to other algorithms i.e gibbs sampling (Murray, 2006).

Perfect sampling is based upon the property that the likelihood of two ratios can be written as an expectation of a density. This changes the probability such that estimation is easier to calculate and understand. The perfect sampler is based on coupling from the past (CFTP) developed by Propp & Wilson (1996). The idea is to use CFTP and repeatedly use the same sampler for generating lower and upper Markov chains started increasingly further back in time until a pair of upper and lower chains coalesce at time 0. The perfect simulation sampler as the result of the chain is then returned. A CFTP sampler for the autologistic model can be constructed as follows.

First, simulate values from an autologistic model

$$P(Y_i = 1|\boldsymbol{Y}_{-i}, \boldsymbol{\theta}) = p_{it} = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})}, \tag{116}$$

where η_{it} represents the systematic component from an autologistic model as specified in section 2.4. The model is useful if the spatial component in the model is

greater than or equal to zero, which implies that the cdf, F_i of Y_j is decreasing in $\sum_{N_i} Y_j$ such that

$$F_i(y) \begin{cases} 0 & \text{if } y < 0 \\ 1 - p_i & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1. \end{cases} \tag{117}$$

This property holds as a non-negative spatial component is required, therefore p_i is increasing in $\sum_{N_i} \boldsymbol{Y}_j$ and so F_i is also increasing in $\sum_{N_i} Y_j$. CFTP uses Markov chains from a stationary distribution and generates random variables, $\boldsymbol{Y}$ from the conditional distribution. The draws from the conditional distribution are exact samples, samples drawn from the true conditional distribution versus Monte Carlo sampling methods. Based upon Møller (1999), the CFTP proceeds as follows.

Let $\boldsymbol{L}_T(t,i)$ and $\boldsymbol{U}_T(t,i)$ denote the i^{th} observations at time t of the lower and upper Markov chains, respectively, where these chains were started at some time T in the past. Fix $T < 0$ and set the lower chain, $\boldsymbol{L}_T(T,\cdot) = 0$ and the upper chain to $\boldsymbol{U}_T(T,\cdot) = 1$. Update the chains based upon

$$\boldsymbol{L}_T(t,i) = F_i^{-1}(\boldsymbol{R}(t,i))|\boldsymbol{L}_T(t,1:i-1), \boldsymbol{L}_T(t-1,i+1:n) \tag{118}$$

$$\boldsymbol{U}_T(t,i) = F_i^{-1}(\boldsymbol{R}(t,i))|\boldsymbol{U}_T(t,1:i-1), \boldsymbol{U}_T(t-1,i+1:n) \tag{119}$$

where the $\boldsymbol{R}(t,i)$ are independent and standard uniform variates and

$$F_i^{-1}(p) = \begin{cases} 0 & \text{if } p \leq 1 - p_i \\ 1 & \text{if } p > 1 - p_i. \end{cases} \tag{120}$$

If $\boldsymbol{L}_T$ and $\boldsymbol{U}_T$ unite at time $t_0 \leq 0$ return $\boldsymbol{L}_T(T,\cdot)$ as a sample from the joint distribution. Otherwise, double the time, T, and start the chain again. The new uniform

variates are used for $T, T+1, \cdots, (T/2)-1$, but for $(T/2), \cdots, -1$, reuse the previously generated uniform variates.

Perfect Simulation with Gibbs Sampling

Perfect sampling requires more computation time than Gibbs sampling, however it can guarantee that the sample is drawn from the exact target distribution during each iteration. One combination of Monte Carlo sampling methods is to combine Gibbs sampling with perfect simulation samples called Gibbs sampled started at perfect simulation samples (PGS) (Zheng & Zhu, 2008; Wang & Zheng, 2013). The idea is the first sample is drawn from PS to guarantee that it is from the target distribution. From the initial sample, use a Gibbs sampler to generate the other independent samples, causing the other subsequent samples to be from the target distribution exactly.

Gibbs sampling is a special case of Metropolis-Hastings sampling, using conditional distributions to construct Markov chain moves at each iteration rather than using the full conditional distribution. This is useful when the full conditional distribution is unknown or is difficult to sample from directly, but the conditional distribution is well known and easier to sample from.

The Gibbs sampling algorithm generates a parameter from the distribution for each iteration, conditional on the current values of the other parameters in the model. In this case, k random variables are simulated sequentially from k conditional distributions to generate a single k-dimensional vector using the fully joint

distribution denoted as $\boldsymbol{Y} = \{y_1, \cdots y_n\}$ from a joint distribution $p(y_1, \cdots y_n)$. A Markov chain is used to create a sequence of samples from Gibbs samplers where the stationary distribution of the Markov chain converges to the target joint distribution (Wang & Zheng, 2013). The Markov chain with Gibbs sampling is generated by repeatedly applying the following update steps.

Step 1: Begin with some initial value $\boldsymbol{Z}^{(0)}$ for each variable.

Step 2: From each sample $i = 1, \cdots k$, sample each variable $z_j^{(i)}$ from the conditional distribution, $p(z_j | z_1^{(i)}, \cdots, z_n^{(i-1)})$.

Step 3: Calculate the ratio of two normalizing constants

$$r = \frac{1}{M} \sum_{m=1}^{M} \frac{\exp\{\boldsymbol{\theta' Z_t^{(m)}}\}}{\exp\{\tilde{\boldsymbol{\theta}}' \boldsymbol{Z_t^{(m)}}\}} \tag{121}$$

Step 3: If the ratio, $r \leq 1$ then set the m^{th} sample to new value.

Step 4: Otherwise if ratio, $r > 1$ then stay at initial value.

The combination of PS and Gibbs sampling creates a sample that does not require a burn-in period and samples are considered independent as they are from the target distribution. In this dissertation, Monte Carlo samples will be generated from PGS as it is computationally faster than PS. The Monte Carlo sampling methods are used to investigate parameter estimation and statistical inference for the autologistic models.

APPENDIX C

R CODE FOR GMM, MCEML, MPL, AND MCML

```
#####################################
#  Spatial-Temporal Model Simulation
#####################################
library(ngspatial)

setwd("/Users/Kimberly/Desktop/DisRFiles/R/")
WD <- getwd()
source("STsimGibbsFunctions.R")

#########################################################
#  Simulation 1
#  beta0 = 1, beta1 = -0.5, theta = 0.1, gamma = 0.5
#########################################################
#Test runs
n <-5
A <- adjacency.matrix(n)
tp <- 6
G <- n*n
set.seed(2342)
X1 <- matrix(1, nrow=tp*G)
X2 <- matrix(rnorm(tp*G, 3, 1))
X <-as.matrix(cbind(X1, X2))
X <- array(X, dim=c(G,tp,2))

theta <- c(1, -0.5, 0.1, 0.5)
M <-2000
simtheta01tm1 <- array(NA, dim=c(G, M, tp))
k<-2354

time <- proc.time()
for (i in 1:M){
  set.seed(k)
  simtheta01tm1[,i,] <- STrauto(X, A, theta, tp)
    cat("Dataset", i, "of" ,M)
    k <- k+2
}
proc.time()-time
#Simulated data file: simtheta01tm1 theta = c(1, -0.5, 0.1, 0.5)
save(simtheta01tm1, file="/Users/kimberly/Desktop/simdatatheta01tm9.
   RData")

#########################################################
#  Simulation 2
#  beta0 = 1, beta1 = -0.5, theta = 0.9, gamma = 0.5
#########################################################
theta <- c(1, -0.5, 0.9, 0.5)

simtheta09tm1 <- array(NA, dim=c(G, M, tp))
k<-2354

time <- proc.time()
for (i in 1:M){
```

```
  set.seed(k)
  simtheta09tm1[,i,] <- STrauto(X, A, theta, tp)
    cat("Dataset", i, "of" ,M)
    k <- k+2
}
proc.time()-time
save(simtheta09tm1, file="/Users/kimberly/Desktop/ngspatial/
   simtheta09tm1.RData")

#########################################################
#  Simulation 3, 10 by 10 grid
#  beta0 = 1, beta1 = -0.5, theta = 0.1, gamma = 0.5
#########################################################
#Test runs
n <-10
A <- adjacency.matrix(n)
tp <- 6
G <- n*n
X1 <- matrix(1, nrow=tp*G)
X2 <- matrix(rnorm(tp*G, 3, 1))
X <-as.matrix(cbind(X1, X2))
X <- array(X, dim=c(G,tp,2))

theta <- c(1, -0.5, 0.1, 0.5)
M <-2000
simtheta01 <- array(NA, dim=c(G, M, tp))
k<-2354
thin <-10
burn <-100

for (i in 1:M){
  set.seed(k)
  simtheta01[,i,] <- STrauto(X, A, theta, tp)
    cat("Dataset", i, "of" ,M)
    k <- k+2
}
save(simtheta01, file="/Users/kimberly/Desktop/ngspatial/
   simtheta01tm25.RData")


#########################################################
#  Simulation 4, 10 by 10 grid
#  beta0 = 1, beta1 = -0.5, theta = 0.9, gamma = 0.5
#########################################################
#Test runs
theta <- c(1, -0.5, 0.9, 0.5)
M <-2000
simtheta09 <- array(NA, dim=c(G, M, tp))


for (i in 1:M){
  set.seed(k)
```

```
  simtheta09[,i,] <- STrauto(X, A, theta, tp)
    cat("Dataset", i, "of" ,M)
    k <- k+2
}

#################################
#   GMM Method 2 for Spatio-Temporal Model
#################################
# Modified 02-15-2014

library(ngspatial)
library(MASS)

#Initialization##

GMMtheta <- function(X, Y, tp, theta, JAexp, x, optit){

quad.phi <- function(phi){

  Inv <- weightmat(theta, X, Y, tp, JAexp, x)
  Winv1 <- Inv$Winv1
  Winv <- Inv$Winv

  Q <- rep(NA, tp)
  ncov <- dim(X)[3]+2

  DD1 <- matrix(NA, nrow=n, ncol=ncov-1)

  G <- matrix(NA, nrow=ncov*n, ncol=(tp-1))
  DD <- matrix(NA, nrow=n, ncol=ncov)

    psi <- phi[ncov-1]
    gamma <- phi[ncov]
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = phi[-((ncov-1):ncov)]

      for(i in 1:dim(X)[3]){
        Xbeta = Xbeta+X[,,i]*beta[i]
    }
    mu = exp(Xbeta)/(1+exp(Xbeta))

    #Setting up the centered current responses
    Yt = A %*%(Y- mu)

    Ytm1 =Y[,-tp]-mu[,-tp]
    eta1 <- Xbeta[,1] + psi * Yt[,1]
    eta <-Xbeta[,-1] + psi * Yt[,-1] + gamma*(Ytm1)
    eta <- cbind(eta1, eta)

    pt = exp(eta)/(1+exp(eta))
    residuals <- (Y-pt)
```

```
    for (t in 1:tp){
      if (t==1){
        DG <- dGspat(X[,1,], Yt[,1], pt[,1], x)
        Nresid <- JAexp%*%residuals[,1]
        for (i in 1:(ncov-1)){
          DD1[,i] <- DG[,i]*Nresid
        }
        G1 <- (matrix(c(DD1[,1], DD1[,2], DD1[,3]), nrow =(ncov-1)*n
           ))
        Q[t] <- t(G1)%*%Winv1%*%G1
      }
      else{
        DG <- dG(X[,t,], Yt[,t], Ytm1[,(t-1)], pt[,t], x)
        Nresid <- JAexp%*%residuals[,t]
        for (i in 1:ncov){
          DD[,i] <- DG[,i]*Nresid
        }
        G[,(t-1)] <- (matrix(c(DD[,1], DD[,2], DD[,3], DD[,4]), nrow
            =ncov*n))
        Q[t] <- t(G[,(t-1)])%*%Winv[,,(t-1)]%*%G[,(t-1)]
      }
    }
    Q <- sum(Q)
  }
  psi <- optim(theta, quad.phi,control = list(maxit = optit))
  convergence <- psi$convergence
  coefficients <- psi$par
  object <- list(coefficients =coefficients, convergence =
     convergence)
  object
}



dG <- function(X, Yt, Ytm1, pt, x){
  d_xs <- matrix(NA, ncol=dim(X)[1], nrow= (dim(X)[2]))
  #d_int <- pt*(1+pt)
    for (i in 1:(dim(X)[2])){
      d_xs[i,] <- (pt*(1-pt)*X[,i])
    }
  d_spat <- (pt*(1-pt)*Yt)
  d_temp <- (pt*(1-pt)*Ytm1)
  dg <- rbind(d_xs, d_spat, d_temp)

  p=dim(dg)[1]
  for(j in 1:dim(dg)[1]){
    der <- dg[j,]
    d=0
    for (i in 1:dim(dg)[2]){
      nder=rep(der[i], x[i])
      d = append(d, nder)
    }
```

```
    dd <- d[-1]
    assign(paste0("D", j), dd)
    }
  cbind(D1, D2, D3, D4)
}



dGspat <- function(X, Yt, pt, x){
  d_xs <- matrix(NA, ncol=dim(X)[1], nrow= (dim(X)[2]))

    for (i in 1:(dim(X)[2])){
      d_xs[i,] <- (pt*(1-pt)*X[,i])
    }
  d_spat <- (pt*(1-pt)*Yt)
  dg <- rbind(d_xs, d_spat)

  p=dim(dg)[1]
  for(j in 1:(dim(dg)[1])){
    der <- dg[j,]
    d=0
    for (i in 1:dim(dg)[2]){
      nder=rep(der[i], x[i])
      d = append(d, nder)
    }
    dd <- d[-1]
    assign(paste0("D", j), dd)
    }

  cbind(D1, D2, D3)
}



weightmat <- function(theta, X, Y, tp, JAexp, x){
    n <- sum(x)
    ncov <- dim(X)[3]+2
    ng <- ncov*n

    DD1 <- matrix(NA, nrow=n, ncol=ncov-1)

    G <- matrix(NA, nrow=ncov*n, ncol=tp-1)
    DD <- matrix(NA, nrow=n, ncol=ncov)
    Winv <- array(NA, dim=c(ncov*n, ncov*n, tp-1))

    psi <- theta[ncov-1]
    gamma <- theta[ncov]
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = theta[-((ncov-1):ncov)]

      for(i in 1:dim(X)[3]){
        Xbeta = Xbeta+X[,,i]*beta[i]
```

```
    }
    mu = exp(Xbeta)/(1+exp(Xbeta))

    Yt = A %*%(Y- mu)

    Ytm1 =Y[,-tp]-mu[,-tp]
    eta1 <- Xbeta[,1] + psi * Yt[,1]
    eta <-Xbeta[,-1] + psi * Yt[,-1] + gamma*(Ytm1)
    eta <- cbind(eta1, eta)

    pt = exp(eta)/(1+exp(eta))
    residuals <- (Y-pt)

    for (t in 1:tp){
      if (t==1){
        Deriv <- dGspat(X[,1,], Yt[,1], pt[,1], x)
        Nresid <- JAexp%*%residuals[,1]
        for (i in 1:(ncov-1)){
          DD1[,i] <- Deriv[,i]*Nresid
        }
        G1 <- (1/nrow(Yt))*(matrix(c(DD1[,1], DD1[,2], DD1[,3]),
           nrow =(ncov-1)*n))
        Winv1 <-ginv(G1%*%t(G1))
      }
      else{
        Deriv <- dG(X[,t,], Yt[,t], Ytm1[,(t-1)], pt[,t], x)
        Nresid <- JAexp%*%residuals[,t]
        for (i in 1:ncov){
          DD[,i] <- Deriv[,i]*Nresid
        }
        G[,(t-1)] <- (matrix(c(DD[,1], DD[,2], DD[,3], DD[,4]), nrow
            =ncov*n))
        Winv[,,(t-1)] <-ginv(G[,(t-1)]%*%t(G[,(t-1)]))
      }
    }

    object= list(Winv1 =Winv1, Winv =Winv)
    object
}

expandmatrix <- function(x){
  nrow <- sum(colSums(x))
  ematrix <- matrix(0, nrow, ncol=dim(x)[2])
  j=1
  for (i in 1:(dim(x)[2])){
    r1<-which(x[i,]==1)
    l[i] <-length(r1)
  }
  ematrix
}
```

```
##################################
#   GMM Method 3 for Spatio-Temporal Model
##################################
library(ngspatial)

GMMthetaM3 <- function(X, Y, tp, theta, A, optit){

################################
#Quadratic Form
################################
quad.phi <- function(phi){

  #Initializing data
  ncov <- dim(X)[3]+2
  n <- dim(Y)[1]
  I <- diag(nrow(Y))
  J <- matrix(1, nrow=nrow(Y), ncol=nrow(Y))

  Inv <- weightmat(theta, X, Y, tp, A, J, I)
  Winv1 <- Inv$Winv1
  Winv <- Inv$Winv

  D1 <- matrix(NA, nrow=n, ncol=ncov-1)

  G <- matrix(NA, nrow=ncov*n, ncol=(tp-1))
  D <- matrix(NA, nrow=n, ncol=ncov)

  Q <- rep(NA, tp-1)

    psi <- phi[ncov-1]
    gamma <- phi[ncov]
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = phi[-((ncov-1):ncov)]

      for(i in 1:dim(X)[3]){
        Xbeta = Xbeta+X[,,i]*beta[i]
    }
    mu = exp(Xbeta)/(1+exp(Xbeta))

    Yt = A %*%(Y- mu)

    Ytm1 =Y[,-tp]-mu[,-tp]
    eta1 <- Xbeta[,1] + psi * Yt[,1]
    eta <-Xbeta[,-1] + psi * Yt[,-1] + gamma*(Ytm1)
    eta <- cbind(eta1, eta)

    pt = exp(eta)/(1+exp(eta))
    pt[is.na(pt)]<-1
    residuals <- (Y-pt)

    for (t in 1:tp){
      if (t==1){
```

```
        DG1 <- dGspat(X[,1,], Yt[,1], pt[,1])
        Nresid <- ((J-A)%*%residuals[,1])
        for (i in 1:(ncov-1)){
          D1[,i] <- DG1[i,]*Nresid
        }
        G1 <- (matrix(c(D1[,1], D1[,2], D1[,3]), nrow =(ncov-1)*n))
        Q[t] <- t(G1)%*%Winv1%*%G1
      }
      else{
        DG <- dG(X[,t,], Yt[,t], Ytm1[,(t-1)], pt[,t])
        Nresid <- ((J-A)%*%residuals[,t])
        for (i in 1:ncov){
          D[,i] <- DG[i,]*Nresid
        }
        G[,(t-1)] <- (matrix(c(D[,1], D[,2], D[,3], D[,4]), nrow =
           ncov*n))
        Q[t] <- t(G[,(t-1)])%*%Winv[,,(t-1)]%*%G[,(t-1)]
      }
    }
    Q <- sum(Q)
  }
  psi <- optim(theta, quad.phi,control = list(maxit = optit))
  convergence <- psi$convergence
  coefficients <- (psi$par)
  object <- list(coefficients =coefficients, convergence =
     convergence)
  object
}



dG <- function(X, Yt, Ytm1, pt){
  d_xs <- matrix(NA, ncol=dim(X)[1], nrow= (dim(X)[2]))
  #d_int <- pt*(1+pt)
    for (i in 1:(dim(X)[2])){
      d_xs[i,] <- (pt*(1-pt)*X[,i])
    }
  d_spat <- (pt*(1-pt)*Yt)
  d_temp <- (pt*(1-pt)*Ytm1)
  dg <- rbind(d_xs, d_spat, d_temp)
  dg
}



dGspat <- function(X, Yt, pt, x){
  d_xs <- matrix(NA, ncol=dim(X)[1], nrow= (dim(X)[2]))
  #d_int <- pt*(1+pt)
    for (i in 1:(dim(X)[2])){
      d_xs[i,] <- (pt*(1-pt)*X[,i])
    }
  d_spat <- (pt*(1-pt)*Yt)
```

```
  dg <- rbind(d_xs, d_spat)
}



weightmat <- function(theta, X, Y, tp, A, J, I){
    n <- dim(Y)[1]
    ncov <- dim(X)[3]+2

    DD1 <- matrix(NA, nrow=n, ncol=ncov-1)
    G <- matrix(NA, nrow=ncov*n, ncol=tp-1)
    DD <- matrix(NA, nrow=n, ncol=ncov)
    Winv <- array(NA, dim=c(ncov*n, ncov*n, tp-1))

    psi <- theta[ncov-1]
    gamma <- theta[ncov]
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = theta[-((ncov-1):ncov)]

      for(i in 1:dim(X)[3]){
        Xbeta = Xbeta+X[,,i]*beta[i]
    }
    mu = exp(Xbeta)/(1+exp(Xbeta))

    #Setting up the centered current responses
    Yt = A %*%(Y- mu)

    Ytm1 =Y[,-tp]-mu[,-tp]
    eta1 <- Xbeta[,1] + psi * Yt[,1]
    eta <-Xbeta[,-1] + psi * Yt[,-1] + gamma*(Ytm1)
    eta <- cbind(eta1, eta)

    pt = exp(eta)/(1+exp(eta))
    pt[is.na(pt)]<-1
    residuals <- (Y-pt)

    for (t in 1:tp){
      if (t==1){
        Deriv1 <- dGspat(X[,1,], Yt[,1], pt[,1])
        Nresid <- ((J-A)%*%residuals[,1])
        for (i in 1:(ncov-1)){
          DD1[,i] <- (Deriv1[i,])*Nresid
        }
        G1 <- (1/nrow(Yt))*(matrix(c(DD1[,1], DD1[,2], DD1[,3]),
            nrow =(ncov-1)*n))
        Winv1 <- ginv(G1%*%t(G1))
      }
      else{
        Deriv <- dG(X[,t,], Yt[,t], Ytm1[,(t-1)], pt[,t])
        Nresid <- ((J-A)%*%residuals[,t])
        for (i in 1:ncov){
```

```
          DD[,i] <- (Deriv[i,])*Nresid
        }
        G[,(t-1)] <- (matrix(c(DD[,1], DD[,2], DD[,3], DD[,4]), nrow
            =ncov*n))
        Winv[,,(t-1)] <- ginv(G[,(t-1)]%*%t(G[,(t-1)]))
      }
    }
    object= list(Winv1 =Winv1, Winv =Winv)
    object
}


############################
#  MCEML Spatial Temporal Model
############################

########################
#  Centered likelihood
########################

library(maxLik)


EMZact.centered <- function(Y, X, A, pit.y){
  tp <- dim(Y)[2]
  p <- dim(X)[3]
  dimx <- dim(X)[3]+2
  Z.obs <- matrix(NA, ncol=tp-1, nrow = dimx)
  Ycent <- Y-pit.y

  for (t in 1:(ncol(Y))){
    for (i in 1:dim(X)[3]){
      Z.obs[i,(t-1)] <- sum(t(Ycent[,t])%*%X[,t,i])
    }
    Z.obs[p+1,(t-1)] <-  0.5*sum(A%*%Ycent[,t])
    Z.obs[p+2,(t-1)] <-  sum(t(Ycent[,t]%*%Ycent[,t-1]))
  }
  Z.obs
}




act.pit <- function(Y, X, newtheta, A){
  tp <- dim(Y)[2]
  dimx <- dim(X)[3]+2
  Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = newtheta[-((dimx-1):dimx)]
  p <- dim(X)[3]
   for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
   }
```

```
    mu = exp(Xbeta)/(1+exp(Xbeta))
  mu
}

##########################################
#  MCEML.Gibbscenter (simulated observations)
##########################################
##Simulated Data ###
MCEML.Gibbscenter <- function(Y.cur, A, M, tp, X, pit){
    k <- 0
    p <- dim(X)[3]
    dimx <- dim(X)[3]+2
    Z <- array(NA, dim = c(M, dimx, (tp-1)))

    for (j in 1:M){
      Ycent <- Y.cur[,j,]-pit
      for (t in 2:tp){
          for (m in 1:p){
            Z[j,m,(t-1)] <- sum(t(Ycent[,t])%*%X[,t,m])
          }
        Z[j,(dim(X)[3]+1),(t-1)] <- 0.5*sum(A%*%Ycent[,t])
        Z[j,(dim(X)[3]+2),(t-1)] <- sum(t(Ycent[,t])%*%(Ycent[,(t-1)
           ]))
      }

    }
  Z
}

pit <- function(Y.cur, A, newtheta,M,tp, X){
    dimx <- dim(X)[3]+2
    Xbeta = matrix(0,nrow=dim(X)[1], ncol=tp)
      beta = newtheta[-((dimx-1):dimx)]
    p <- dim(X)[3]
       for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
    }
      mu = exp(Xbeta)/(1+exp(Xbeta))
    mu
}

stemp <- function(Yt, X, A, theta){
  ncov = length(theta)
  tp <- dim(X)[2]
     Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
     beta = theta[-((ncov-1):ncov)]
     psi = theta[ncov-1]
   gamma = theta[ncov]
     for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
   }
     mu = exp(Xbeta)/(1+exp(Xbeta))
```

```
     summu = A %*% mu
    for (k in 2:ncol(Yt)){
      #k=2
      Y[,k]= Yt[,(k-1)]
      for (j in 1:nrow(X)){
          sumY_j =A[j,] %*% Yt[,k]
          eta <-Xbeta[j,k] + psi * (sumY_j - summu[j,k]) + gamma*(Yt
              [j,(k-1)]-mu[j,(k-1)])
            prob <- exp(eta)/(1+exp(eta))
            Yt[j, k] <- rbinom(1,1, prob)
            }
      }
     Yt
}

gibbs.STsample <- function(X, A, theta, nit, thin, burn){
  kp <- 0
  tp <- dim(X)[2]
  Yt.draws <- array(NA, dim=c(nrow(X), nit, tp))
  Ynew <- STrauto(X, A, theta, tp)
  for (i in 1:(burn+nit*thin)){
    Ycur <- stemp(Ynew, X, A, theta)
    Ynew <- Ycur
  if (i%in%seq(burn+thin,burn+thin*nit, by=thin)){
    kp <- kp+1
    Yt.draws[,kp,]<- Ycur
    }
  }
  Yt.draws
}


##########################################
#  Srautologistic1: Spatial model only no time terms
#
##########################################

Srautologistic1 = function(X, A, theta){
  p <- length(theta)
    beta = theta[-p]
    psi = theta[p]
    n = nrow(X)
  Xbeta <- X%*%beta
    mu = exp(Xbeta)/(1+exp(Xbeta))
    summu = A %*% mu
    L = matrix(0, nrow=n)  #Lower bound
    U = matrix(1, nrow=n)  #Upper bound
    Tt = 2
    R = matrix(0, nrow=Tt, ncol=n)
    t = 0
    restart = F
    repeat{
```

```
          if (t == Tt && sum(U - L) == 0){
            L
            break
          }
          t=t+1
          if (t > Tt){
            L <- rep(0, length(L))
            U <- rep(1, length(U))
        insert <- matrix(0, nrow=Tt, ncol=n)
               R <-  rbind(insert, R)
               Tt= Tt*2
               t = 1
               restart = T
          }
          if (! restart || (restart && t <= Tt / 2)){
            R[t,] = runif(n,0,1)

            }
          for (i in 1:n){
              sumL_i =A[i,] %*% L
              q = 1 / (1 + exp(Xbeta[i] + psi * (sumL_i - summu[i]))
                 )
              L[i]=ifelse(R[t,i] > q, 1, 0)

                sumU_i = A[i,] %*% U
              q = 1 / (1 + exp(Xbeta[i] + psi * (sumU_i - summu[i]))
                 )
              U[i]=ifelse(R[t,i] > q, 1, 0)

        }
        #cat("sample is:",t)
      }
      L
 }


####################
##   MCEML Program
####################
########################

#Initializing Packages

library(maxLik)
library(MASS)
library(ngspatial)



#MCML and MPL tests
setwd("/Users/kimberly/Desktop/DisRFiles/R/MCEML")
WD <- getwd()
```

```
source("MCEML ProgramNRtest.R")

#Step 1: Find the initial values
#MCML and MPL tests
setwd("/Users/kimberly/Desktop/DisRFiles/R/MPL")
WD <- getwd()
source("MPL_STModel02102014.R")

########################
# MCEML Maximization
########################

#Maximization Steps
MCEML.Max <- function(Zpsi, psi, thetahat0, M, thin, burn, tp, X, Y,
    A, optit, initial, NIT){
  YM <-gibbs.STsample(X, A, psi, M, thin, burn)#keep this value for
     future iterations
  iter = 0
  repeat{
    iter= iter+1
    tol.low <- 0.001

    pit.y <- act.pit(Y, X, thetahat0, A)
    Zactual <- EMZact.centered(Y, X, A, pit.y)
    new.pit <- pit(YM, A, thetahat0, M, tp, X)
    Zthetahat <- MCEML.Gibbscenter(YM, A, M, tp, X, new.pit)

  par1 <-MCEML.Main(initial, psi, 1000, Zthetahat, Zactual, Zpsi, M,
     tp)
  thetahat <- par1$par.est
  convergence <- par1$convergence
  final <- list(thetahat = thetahat, convergence=convergence)

    if (iter >=NIT){
      cat("# exceeded the maximum iteration number", fill=TRUE)
      break
      }else if (sum(abs(thetahat-thetahat0))<=(4*tol.low)){
      break
      }else {
        thetahat0 <- thetahat
      }
  }
  final
}

##########################################
#  MCEML.Main
#
##########################################
MCEML.Main = function(initial, psi, optit, Zthetahat, Zactual, Zpsi,
    M,tp){
  expz <- matrix(NA, nrow=M, ncol=tp-1)
```

```
  loglike = function(theta){
    for (k in 1:(tp-1)){
      for (m in 1:M){
        expz[m,k] = exp(t(theta)%*%Zthetahat[m,,k]-t(psi)%*%Zpsi[m,,
           k])
      }
    }
    -(sum(exp(t(theta-psi)%*%(Zactual)))*sum(colMeans(expz)))

  }
  est = maxNR(loglike, start = initial, iterlim=optit)
  par.est <- est$estimate
  convergence <- est$code
  final <- list(par.est = par.est, convergence=convergence)
}

####################
#  MCML
####################
#########################
#  Centered likelihood
#########################
## Define sufficient statistics from observed data
#####################################################
#  Zact.centered
#
#####################################################
Zact.centered <- function(Y, X, theta, A){
  tp <- dim(Y)[2]
  dimx <- dim(X)[3]+2
  Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = theta[-((dimx-1):dimx)]
  p <- dim(X)[3]
   for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
   }
    mu = exp(Xbeta)/(1+exp(Xbeta))
    summu = A %*% mu
    Ycent <- Y-mu
  p <- dim(X)[3]
  Z.obs <- matrix(NA, ncol=tp-1, nrow = dimx)
  for (t in 1:(ncol(Y))){
    for (i in 1:dim(X)[3]){
      Z.obs[i,(t-1)] <- sum(t(Ycent[,t])%*%X[,t,i])
    }
    Z.obs[p+1,(t-1)] <-  0.5*sum(A%*%Ycent[,t])
    Z.obs[p+2,(t-1)] <-  sum(t(Ycent[,t]%*%Ycent[,t-1]))
  }
  Z.obs
}
```

```
##########################################
#  MCML.Gibbscenter (simulated observations)
##########################################

MCML.Gibbscenter <- function(A, theta, nit, thin, burn, tp, X){
    k <- 0
    p <- dim(X)[3]
    dimx <- dim(X)[3]+2
    Z <- array(NA, dim = c(nit, length(theta), (tp-1)))
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
      beta = theta[-((dimx-1):dimx)]
    p <- dim(X)[3]
       for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
    }
      mu = exp(Xbeta)/(1+exp(Xbeta))
      summu = A %*% mu
    #Ytm1 <- array(NA, dim=c(100, tp-1, nit))
    Y.cur <-gibbs.STsample(X, A, theta, nit, thin, burn)
    #Ycent <- Y.cur-summu
    #Ytm1 <- Y.cur[,,-tp]
    #Y.cur <- Y.cur[,,-1]
    for (j in 1:nit){
      Ycent <- Y.cur[,j,]-mu
      for (t in 2:tp){
        #t <-2
          for (m in 1:p){
            Z[j,m,(t-1)] <- sum(t(Ycent[,t])%*%X[,t,m])#dim tp by
                ncov
          }
        Z[j,(dim(X)[3]+1),(t-1)] <- 0.5*sum(A%*%Ycent[,t])
        Z[j,(dim(X)[3]+2),(t-1)] <- sum(t(Ycent[,t])%*%(Ycent[,(t-1)
           ]))
      }
    }
  Z
}

##########################################
#  MCML.Main
##########################################

MCML.Main = function(Y, X, tp, M, A, thin, burn, initial, psi,optit)
   {
  #psi <- glm(Y[,tp] ~ X[,tp,]+A%*%Y[,tp]+Y[,tp-1] - 1, family =
     binomial)$coef
  Z.act <- Zact.centered(Y, X, psi, A)
  Z <- MCML.Gibbscenter(A, psi, M, thin, burn, tp, X)
  expz <- matrix(NA, nrow=M, ncol=tp-1)
  loglike <- rep(NA, tp-1)
  ll = function(theta){
    for (k in 1:(tp-1)){
```

```
      for (m in 1:M){
        expz[m,k] = exp(t(theta-psi)%*%Z[m,,k])
      }
    }
    loglike = -(sum(t(theta-psi)%*%(Z.act))-sum(log(colMeans(expz)))
       )
  }
  #est = optim(initial, method="BFGS", ll, hessian=TRUE) #Just to
     compare
  est =optim(initial, ll,control = list(maxit = optit))
  est
}

stemp <- function(Yt, X, A, theta){
  ncov = length(theta)
  tp <- dim(X)[2]
     Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
     beta = theta[-((ncov-1):ncov)]
     psi = theta[ncov-1]
   gamma = theta[ncov]
     for(i in 1:dim(X)[3]){
     Xbeta = Xbeta+X[,,i]*beta[i]
   }
     mu = exp(Xbeta)/(1+exp(Xbeta))
     summu = A %*% mu
    for (k in 2:ncol(Yt)){
      Y[,k]= Yt[,(k-1)]
      for (j in 1:nrow(X)){
          sumY_j =A[j,] %*% Yt[,k]
          eta <-Xbeta[j,k] + psi * (sumY_j - summu[j,k]) + gamma*(Yt
             [j,(k-1)]-mu[j,(k-1)])
            prob <- exp(eta)/(1+exp(eta))
            Yt[j, k] <- rbinom(1,1, prob)
            }
      }
     Yt
}

gibbs.STsample <- function(X, A, theta, nit, thin, burn){
  kp <- 0
  tp <- dim(X)[2]
  Yt.draws <- array(NA, dim=c(nrow(X), nit, tp))
  Ynew <- STrauto(X, A, theta, tp)
  for (i in 1:(burn+nit*thin)){
    Ycur <- stemp(Ynew, X, A, theta)
    Ynew <- Ycur
  if (i%in%seq(burn+thin,burn+thin*nit, by=thin)){
    kp <- kp+1
    Yt.draws[,kp,]<- Ycur
    }
  }
  Yt.draws
```

```
}

################################################
#STrauto: ST model with gamma the temporal terms
################################################

STrauto = function(X, A, theta, tp){
  p = length(theta)
    Xbeta = matrix(0,nrow=nrow(X), ncol=tp)
    beta = theta[-((p-1):p)]
    psi = theta[p-1]
  gamma = theta[p]
    n = nrow(X)
    for(j in 1:dim(X)[3]){
    Xbeta = Xbeta+X[,,j]*beta[j]
  }
    mu = exp(Xbeta)/(1+exp(Xbeta))
    summu = A %*% mu
    L = matrix(0, nrow= n)  #Lower bound
    U = matrix(1, nrow=n)  #Upper bound
  Y1 <- Srautologistic1(X[,1,], A, theta[-p])
  Y <- matrix(Y1, nrow = n, ncol=tp)

    for (k in 2:tp){
      L = matrix(0, nrow= n)  #Lower bound
      U = matrix(1, nrow=n)  #Upper bound
      Tt = 2
      R = matrix(0, nrow=Tt, ncol=n)
      t = 0
      restart = F
        repeat{
          if (t == Tt && sum(U - L) == 0){
        L
            break
          }
          t=t+1
          if (t > Tt){
            L <- rep(0, length(L))
            U <- rep(1, length(U))
        insert <- matrix(0, nrow=Tt, ncol=n)
              R <-  rbind(insert, R)
               Tt= Tt*2
               t = 1
               restart = T
          }
          if (! restart || (restart && t <= Tt / 2)){
            R[t,] = runif(n,0,1)
            }
          for (i in 1:n){
              sumL_i =A[i,] %*% L
              q = 1 / (1 + exp(Xbeta[i,k] + psi * (sumL_i - summu[i,
                 k])+gamma*(Y[i,(k-1)]-mu[i,(k-1)])))
```

```
              L[i]=ifelse(R[t,i] > q, 1, 0)


                sumU_i = A[i,] %*% U
              q = 1 / (1 + exp(Xbeta[i,k] + psi * (sumU_i - summu[i,
                 k])+gamma*(Y[i,(k-1)]-mu[i,(k-1)])))
              U[i]=ifelse(R[t,i] > q, 1, 0)
          }
        #cat("sample is:",t)
        }
        Y[,k]<-L
      }
    Y
}

################################################
#Srautologistic: Spatial model only no time terms
################################################

Srautologistic = function(X, A, theta){
  p <- length(theta)
    beta = theta[-p]
    psi = theta[p]
    n = nrow(X)
  Xbeta <- X%*%beta
    mu = exp(Xbeta)/(1+exp(Xbeta))
    summu = A %*% mu
    L = matrix(0, nrow=n)  #Lower bound
    U = matrix(1, nrow=n)  #Upper bound
    Tt = 2
    R = matrix(0, nrow=Tt, ncol=n)
    t = 0
    restart = F
    repeat{
          if (t == Tt && sum(U - L) == 0){
            L
            break
          }
          t=t+1
          if (t > Tt){
            L <- rep(0, length(L))
            U <- rep(1, length(U))
        insert <- matrix(0, nrow=Tt, ncol=n)
               R <-  rbind(insert, R)
               Tt= Tt*2
               t = 1
               restart = T
          }
          if (! restart || (restart && t <= Tt / 2)){
            R[t,] = runif(n,0,1)

            }
```

```
          for (i in 1:n){
              sumL_i =A[i,] %*% L
              q = 1 / (1 + exp(Xbeta[i] + psi * (sumL_i - summu[i]))
                  )
              L[i]=ifelse(R[t,i] > q, 1, 0)

                sumU_i = A[i,] %*% U
              q = 1 / (1 + exp(Xbeta[i] + psi * (sumU_i - summu[i]))
                  )
              U[i]=ifelse(R[t,i] > q, 1, 0)

        }
      }
      L
 }


##############################
# Spatial-Temporal autologistic model
##############################

##########################################
#  autologistic function
##########################################


autologistic.logPL = function(X, A, Y, theta)
{
    p = length(theta)
    beta = theta[-p]
    psi = theta[p]
    Xbeta = X %*% beta
    mu = exp(Xbeta) / (1 + exp(Xbeta))
    logPL = Xbeta + psi * (A %*% (Y - mu))
    logPL = t(Y) %*% logPL - sum(log(1 + exp(logPL)))
    logPL
}

##########################################
#  STautologistic function
##########################################


STautologistic.logPL = function(X, A, Y, theta)
 {
  Xbeta = matrix(0,nrow=nrow(Y), ncol=ncol(Y))
  ncov <- dim(X)[2]
  p = length(theta)
  beta <- theta[1:(p-2)]
  psi = theta[p-1]
  gamma = theta[p]
```

```
    Ys <- Y[,2:(ncol(Y))]
    Ytm1 <- Y[,1:(ncol(Y)-1)]
  for(j in 1:dim(X)[3]){
    Xbeta = Xbeta +X[,,j]*beta[j]
  }
  #Centering the model
     mut = exp(Xbeta[2:ncol(Y)]) / (1 + exp(Xbeta[2:ncol(Y)]))
     mutm1 = exp(Xbeta[1:(ncol(Y)-1)]) / (1 + exp(Xbeta[1:(ncol(Y)
        -1)]))

     slogPL = autologistic.logPL(X[,1,], A, Y[,1], theta[-p])
     stlogPL = Xbeta[,2:ncol(Y)] +psi * (A %*% (Ys - mut)) + gamma*(
        Ytm1 - mutm1)

     stlogPL <- sum(diag(t(stlogPL) %*%Ys) - colSums(log(1 + exp(
        stlogPL))))
     stlogPL <- slogPL+stlogPL
     stlogPL
}


STautologistic.objective = function(params, X, A, Y)
{
     p = length(params)
     -STautologistic.logPL(X, A, Y, c(params[1:(p-2)], exp(params[(p
        -1):p])))
}

autologistic.fit = function(X, A, Y, optit = 1000){
  TP <- ncol(Y)-1
  start <- matrix(NA, TP, dim(X)[3])
    for (i in 1:TP){
      start[i,] = glm(Y[,i] ~ X[,i,]- 1, family = binomial)$coef
      }
  start = colMeans(start)
    opt = try(optim(c(start, 0, 0), STautologistic.objective, X = X,
        A = A, Y = Y, control = list(maxit = optit)), silent = TRUE)
     if (class(opt) == "try-error")
     {
      coefficients = NULL
      fitted.values = NULL
      linear.predictors = NULL
      residuals = NULL
      convergence = NULL
      message = opt[1]
     }
     else
     {
         convergence = opt$convergence
    if (convergence == 1)
      message = "optim iteration limit 'optit' was reached."
    else if (convergence == 10)
```

```
      message = "The Nelder-Mead simplex degenerated."
    else
      message = NULL
    p = dim(X)[3] + 2
        coefficients = opt$par

        coefficients[(p-1):p] = exp(coefficients[(p-1):p])
        names(coefficients)[p-1] = "psi"
        names(coefficients)[p] = "gamma"
        Xbeta = matrix(0,nrow=nrow(Y), ncol=(ncol(Y)-1))
    X <- X[,-1,]
    Ys <- Y[,2:(ncol(Y))]
      Ytm1 <- Y[,1:(ncol(Y)-1)]
        for(j in 1:dim(X)[3]){
      Xbeta = Xbeta+X[,,j]*coefficients[-((p-1):p)][j]
    }
         mu = exp(Xbeta)
         mu = mu / (1 + mu)
         autocovariate = A %*% (Ys - mu)
         autoregressive = (Ytm1 - mu)
         linear.predictors = Xbeta + coefficients[p-1] *
            autocovariate+coefficients[p]*autoregressive
         fitted.values = exp(linear.predictors)
         fitted.values = fitted.values / (1 + fitted.values)
         residuals = Y[,-1] - fitted.values
     }
     object = list(coefficients = coefficients, fitted.values=fitted
        .values, linear.predictors=linear.predictors, residuals=
        residuals, convergence=convergence, message=message, optit=
        optit)
     class(object) = "autologistic"
     object
 }


##########################################
#  Finding the Monte Carlo Standard Errors
##########################################

autologistic.bmse = function(mat)
{
  if (sum(is.na(mat)) > 0)
  {
    warning("The sample contains NAs.")
    bmse = NA
  }
  else
  {
    mat = as.matrix(mat)
    p = ncol(mat)
    bmse = c()
    for (i in 1:p)
```

```
    {
      temp = bmse.st(mat[, i])
      if (temp == -1)
        temp = NA
      bmse = c(bmse, temp)
    }
  }
  bmse
}



bmse.st <- function(vals)
{
  N = length(vals)
    if (N < 10){
        result = -1
        }
    else
    {
        b = floor(sqrt(N))
        a = floor(N / b)
        temp <- rep(NA, b)
        Ys <- rep(NA, a)
        for (i in 1:a)
        {
            temp = vals[((i-1) * b):(((i-1) + 1) * b - 1)]
            Ys[i] = mean(temp)
        }
        muhat = mean(Ys)
        sigmahatsq = b * sum((Ys - muhat)^2) / (a - 1)
        result = sqrt(sigmahatsq / N)
    }
  result
}


STautologistic.boothelper = function(dummy, X, A, theta, optit, tp)
{
    Y = STrauto(X, A, theta, tp) #replace with Gibbs sampler
    fit = autologistic.fit(X, A, Y, optit)
    fit
}

STautologistic.bootstrap = function(X, A, theta, bootit, optit, tp)
{
  boot.sample = data.frame(matrix(, bootit, length(theta)))
    for (j in 1:bootit)
    {
      fit = STautologistic.boothelper(NULL, X, A, theta, optit, tp)
      temp = rep(NA, length(theta))
      if (is.null(fit$convergence) || fit$convergence != 0)
```

```
          warning(fit$message)
      else
          temp = fit$coef
      boot.sample[j, ] = temp
      }
      boot.sample
}

STautologistic = function(X, A, Y, optit = 1000, bootit = 1000, tp)
{
  fit = autologistic.fit(X, A, Y, optit)
  if (fit$convergence != 0)
    stop(fit$message, " Consider increasing the value of 'optit'.")
    # if (! is.numeric(bootit) || ! is.wholenumber(bootit) || bootit
        < 1)
      # stop("'bootit' must be a positive whole number.")
    fit$sample = STautologistic.bootstrap(X, A, fit$coef, bootit,
       optit, tp)
    fit$mcse = autologistic.bmse(fit$sample)
    fit$iter = bootit
  fit
}

summary.autologistic = function(object, alpha = 0.05, digits = 4,
   ...)
{
  cat("\nCall:\n")
  print(object$call)
  p = length(object$coef)
    if (is.null(object$sample) || (sum(is.na(object$sample)) > 0))
  {
    warning("The sample is NULL or contains NAs.")
      coef.table = cbind(object$coef, NA, NA, NA)
  }
  else
  {
    ci = matrix(, p, 2)
      for (j in 1:p)
        ci[j, ] = quantile(object$sample[, j], c(alpha / 2, 1 -
           alpha / 2), na.rm = TRUE)
    coef.table = cbind(object$coef, ci, object$mcse)
  }
    colnames(coef.table) = c("Estimate", "Lower", "Upper", "MCSE")
  rownames(coef.table) = names(object$coef)
  cat("\nCoefficients:\n")
  print(signif(coef.table, digits))
  cat("\nNumber of iterations:", object$iter, "\n\n")
}
```