Statistical Monitoring and Modeling
for Spatial Processes

Matthew J. Keefe

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Christopher T. Franck, Co-Chair
Marco A.R. Ferreira, Co-Chair
William H. Woodall
Ronald D. Fricker, Jr.

February 28, 2017
Blacksburg, Virginia

Keywords: Bayesian Analysis, Objective Priors, Risk-adjustment, Spatial Statistics,
Statistical Process Monitoring
Copyright 2017, Matthew J. Keefe
Statistical Monitoring and Modeling
for Spatial Processes

Matthew J. Keefe

ABSTRACT

Statistical process monitoring and hierarchical Bayesian modeling are two ways to learn more about processes of interest. In this work, we consider two main components: risk-adjusted monitoring and Bayesian hierarchical models for spatial data. Usually, if prior information about a process is known, it is important to incorporate this into the monitoring scheme. For example, when monitoring 30-day mortality rates after surgery, the pre-operative risk of patients based on health characteristics is often an indicator of how likely the surgery is to succeed. In these cases, risk-adjusted monitoring techniques are used. In this work, the practical limitations of the traditional implementation of risk-adjusted monitoring methods are discussed and an improved implementation is proposed. A method to perform spatial risk-adjustment based on exact locations of concurrent observations to account for spatial dependence is also described. Furthermore, the development of objective priors for fully Bayesian hierarchical models for areal data is explored for Gaussian responses. Collectively, these statistical methods serve as analytic tools to better monitor and model spatial processes.
Many current scientific applications involve data collection that has some type of spatial component. Within these applications, the objective could be to monitor incoming data in order to quickly detect any changes in real time. Another objective could be to use statistical models to characterize and understand the underlying features of the data. In this work, we develop statistical methodology to monitor and model data that include a spatial component. Specifically, we develop a monitoring scheme that adjusts for spatial risk and present an objective way to quantify and model spatial dependence when data is measured for areal units. Collectively, the statistical methods developed in this work serve as analytic tools to better monitor and model spatial data.
Dedication

To my wife Jess, for her constant loving support.
Acknowledgements

I would like to express my gratitude for my co-advisor, Dr. Chris Franck, for initially getting me involved in research and helping me to develop my technical skills which have undoubtedly strengthened my research skills and passion for the field of statistics. I consider you not only a mentor, but a good friend that I am comfortable talking to about anything. I would also like to thank my co-advisor Dr. Marco Ferreira, for continually teaching me new techniques and expanding my knowledge of statistics.

Furthermore, I am grateful for the remaining members of my committee Dr. William Woodall and Dr. Ron Fricker. I attribute my ability to effectively read and comprehend papers in statistics to Dr. Woodall and his excellent teaching of Quality Control. I also appreciate the guidance provided by Dr. Ron Fricker and all that he has done as department head.

I would like to thank my co-authors, Justin B. Loda and Ahmad E. Elhabashy for their contributions to the research on the improved implementation of the risk-adjusted Bernoulli CUSUM. Also, I am thankful for the supportive faculty, staff, and graduate students in the Virginia Tech Department of Statistics.

Finally, I would like to thank all of my family for their love and support. Thank you to Ralph, Nancy, Marty, Gary, Donna, Kacey, Kelly, and all of my closest friends. Thank you Jess, for always loving me, supporting me in all of my decisions, caring for me when I broke my arm, and most importantly for being the most amazing wife I could ever ask for. I love you.
Contents

1 Introduction 1

1.1 Background .................................................. 1
1.2 Statistical Process Monitoring .......................... 3
  1.2.1 Overview .................................................. 3
  1.2.2 Risk-adjusted Monitoring ............................ 3
1.3 Modeling Spatial Data ....................................... 4
  1.3.1 Types of Spatial Data .................................. 4
  1.3.2 Modeling Areal Data .................................. 5
  1.3.3 Prior Distributions for Hierarchical Spatial Models .... 6
1.4 Dissertation Outline ......................................... 7

2 Improved Implementation of the Risk-adjusted Bernoulli CUSUM Chart to Monitor Surgical Outcome Quality 10

2.1 Introduction .................................................. 12
2.2 Methods ......................................................... 13
  2.2.1 Risk-adjusted Bernoulli CUSUM Procedure ........ 13
2.2.2 Proposed Approach .................................................. 14
2.2.3 Proposed Implementation ............................................. 15
2.2.4 Illustration .............................................................. 17
2.2.5 Simulation Study ....................................................... 19
2.3 Results ................................................................. 20
2.4 Discussion ............................................................. 22

3 Monitoring Foreclosure Rates with a Spatially Risk-adjusted Bernoulli CUSUM Chart for Concurrent Observations ................................................. 27

3.1 Introduction ............................................................. 29
3.2 CUSUM and risk-adjustment procedure .................................. 31
  3.2.1 Standard CUSUM .................................................. 31
  3.2.2 Risk-adjusted Bernoulli CUSUM .................................. 32
  3.2.3 Modified risk-adjusted Bernoulli CUSUM ......................... 34
  3.2.4 Spatial risk-adjustment ............................................. 35
3.3 Simulation study ......................................................... 37
3.4 Case study: monitoring foreclosure rates ................................... 39
  3.4.1 Data description ................................................... 39
  3.4.2 Estimating risk of foreclosure .................................... 42
  3.4.3 Results ............................................................ 44
3.5 Discussion ............................................................. 47

4 On the Formal Specification of Sum-zero Constrained Intrinsic Conditional
## Autoregressive Models

4.1 Introduction ....................................................... 58
4.2 Intrinsic Conditional Autoregressive Model .................. 59
4.3 Obtaining the Sum-zero Constrained Intrinsic Conditional Autoregressive Model 60
4.4 Examples .......................................................... 64
4.5 Discussion ......................................................... 65

## Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors

5.1 Introduction ....................................................... 70
5.2 Model Specification ............................................... 72
  5.2.1 Model ......................................................... 72
  5.2.2 Intrinsic CAR as the limit of a proper GMRF ............ 74
  5.2.3 Remarks ...................................................... 76
5.3 Objective Priors .................................................. 77
  5.3.1 Likelihood Functions .......................................... 77
  5.3.2 Reference Prior ................................................ 80
  5.3.3 Independence Jeffreys Prior .............................. 82
  5.3.4 Jeffreys-rule Prior ........................................... 83
  5.3.5 Remarks ...................................................... 83
5.4 Sampling from the Posterior Distribution .................... 84
5.5 Simulation Study .................................................. 85
5.5.1 Comparison of Priors ............................................. 85
5.5.2 Simulation Design ................................................. 89
5.5.3 Simulation Results ............................................... 90
5.6 Case Study ..................................................... 91
  5.6.1 Data Description ........................................... 91
  5.6.2 Modeling .................................................. 94
5.7 Discussion ..................................................... 97

6 Conclusions and Future Work ........................................ 122
  6.1 Conclusions .................................................. 122
  6.2 Future Work ................................................ 123
  6.2.1 Non-Gaussian Response CAR Models ................. 123
  6.2.2 Bayesian Process Monitoring ......................... 127
  6.3 Final Remarks .............................................. 128
List of Figures

2.1 Distribution of the number of days lived for patients who died within 30 days of surgery (1992-1993) .................................................. 16

2.2 Illustration of proposed implementation of the risk-adjusted Bernoulli CUSUM chart. Circled triangles indicate patients who died resulting in CUSUM statistics that were updated on the given day. .................................................. 18

2.3 ARL (in days) by number of operations per day for (a) $R = 2$, (b) $R = 3$, (c) $R = 4$, and (d) $R = 5$ (based on 1,000 simulated charts) .................. 23

3.1 Sale type distributions for Wayne County, Michigan for (a) 2005, (b) 2006, (c) 2007 and (d) yearly percentage breakdown ......................... 40

3.2 Boxplots for the number of (a) foreclosures and (b) regular sales by day of the week for Wayne County, Michigan from 1 January 2005 through 30 June 2005, our Phase I sample. Panel (c) shows the number of transactions per week for Phase I and II prior to when the chart signals (see Results). This illustrates that aggregation at the week level demonstrates no significant shift in the distribution of the number of concurrent observations between Phase I and II. .................................................. 42

3.3 Kernel density estimates for (a) observed foreclosures, (b) regular sales, and (c) estimated foreclosure probability .......................... 43
3.4 Risk-adjusted and non-risk-adjusted Bernoulli CUSUM control charts for monitoring foreclosure rates in Wayne County for (a) 1 July 2005 to 10 June 2014 and zoomed in for (b) 1 July 2005 to 1 January 2008 ... 45

3.5 Risk-adjusted and non-risk-adjusted Bernoulli CUSUM control charts for monitoring foreclosure rates in Wayne County for 1 July 2005 to July 2006 ... 46

3.6 Weekly proportion of foreclosures in Wayne County, Michigan for 1 January 2005 through 10 June 2014 ... 48

3.7 Likelihood-based cross validation function for various bandwidths ... 51

5.1 (a) Raw plot of $\pi(\tau_c)$ and (b) plot of $\log \pi(\tau_c)$ vs. $\log_{10} \tau_c$ for the reference, CARBayes, and NB priors. The CARBayes and NB priors both approach $\infty$ as $\tau_c \to 0$, while the reference prior does not. Panel (c) shows the Kullback-Leibler divergence per observation between the spatial model and the independent data model across values of $\tau_c$, indicating that large values of $\tau_c$ correspond to a model for nearly independent data. ... 88

5.2 Frequentist coverage and $\log_{10}$ average interval length (IL) for $\tau_c$ for $n = 100$ (top row), $n = 49$ (middle row), and $n = 25$ (bottom row). Reference prior shows favorable performance in terms of frequentist coverage and average interval length. ... 92

5.3 $\log_{10}$MSE for the posterior median of $\tau_c$ for (a) $n = 100$, (b) $n = 49$, and (c) $n = 25$. Reference prior leads to favorable performance in terms of estimation of $\tau_c$. Note that since the y-axis is on the $\log_{10}$ scale, the difference in MSE of the posterior median of $\tau_c$ between the reference prior and the CARBayes prior is considerable. ... 93
5.4 Map of (a) 2012 SMRs for foreclosure rates, (b) 2012 unemployment rates (\%), (c) posterior median $E[\text{SMR}]$, and (d) posterior standard deviation of $E[\text{SMR}]$ in Ohio counties
List of Tables

2.1 Example Data Set ................................................. 17

2.2 ARL comparison (in days) for traditional and proposed implementation schemes 21

3.1 Run length performance for varying number of concurrent observations. ARL and standard deviation of run length (SDRL) are presented. ............... 38

3.2 Possible control limits with corresponding in-control average run lengths (ARLs) 47

5.1 Posterior Summaries for Foreclosure Rate Case Study for Bayesian Analyses Using Reference, CARBayes, and NB Priors ........................................... 96

A1 Simulation results for $\tau_c$ with $n = 25$, $p = 1$, and first-order neighborhood . 108

A2 Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and first-order neighborhood . 108

A3 Simulation results for $\tau_c$ with $n = 100$, $p = 1$, and first-order neighborhood . 108

A4 Simulation results for $\tau_c$ with $n = 25$, $p = 6$, and first-order neighborhood . 108

A5 Simulation results for $\tau_c$ with $n = 49$, $p = 6$, and first-order neighborhood . 109

A6 Simulation results for $\tau_c$ with $n = 100$, $p = 6$, and first-order neighborhood . 109

A7 Simulation results for $\tau_c$ with $n = 25$, $p = 1$, and second-order neighborhood 109

A8 Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and second-order neighborhood 109
A9 Simulation results for $\tau_c$ with $n = 100$, $p = 1$, and second-order neighborhood 110
A10 Simulation results for $\tau_c$ with $n = 25$, $p = 6$, and second-order neighborhood 110
A11 Simulation results for $\tau_c$ with $n = 49$, $p = 6$, and second-order neighborhood 110
A12 Simulation results for $\tau_c$ with $n = 100$, $p = 6$, and second-order neighborhood 110
A13 Simulation results for $\sigma^2$ with $n = 25$, $p = 1$, and first-order neighborhood 111
A14 Simulation results for $\sigma^2$ with $n = 49$, $p = 1$, and first-order neighborhood 111
A15 Simulation results for $\sigma^2$ with $n = 100$, $p = 1$, and first-order neighborhood 111
A16 Simulation results for $\sigma^2$ with $n = 25$, $p = 6$, and first-order neighborhood 111
A17 Simulation results for $\sigma^2$ with $n = 49$, $p = 6$, and first-order neighborhood 112
A18 Simulation results for $\sigma^2$ with $n = 100$, $p = 6$, and first-order neighborhood 112
A19 Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and second-order neighborhood 112
A20 Simulation results for $\tau_c$ with $n = 100$, $p = 1$, and second-order neighborhood 112
A21 Simulation results for $\sigma^2$ with $n = 100$, $p = 1$, and second-order neighborhood 113
A22 Simulation results for $\sigma^2$ with $n = 25$, $p = 6$, and second-order neighborhood 113
A23 Simulation results for $\sigma^2$ with $n = 49$, $p = 6$, and second-order neighborhood 113
A24 Simulation results for $\sigma^2$ with $n = 100$, $p = 6$, and second-order neighborhood 113
A25 Simulation results for $\beta$ with $n = 25$, $p = 1$, and first-order neighborhood . . 114
A26 Simulation results for $\beta$ with $n = 49$, $p = 1$, and first-order neighborhood . . 114
A27 Simulation results for $\beta$ with $n = 100$, $p = 1$, and first-order neighborhood . 114
A28 Simulation results for $\beta$ with $n = 25$, $p = 6$, and first-order neighborhood . . 114
A29 Simulation results for $\beta$ with $n = 49$, $p = 6$, and first-order neighborhood . . 115
A30 Simulation results for $\beta$ with $n = 100$, $p = 6$, and first-order neighborhood 115
A31 Simulation results for $\beta$ with $n = 25$, $p = 1$, and second-order neighborhood 115
A32 Simulation results for $\beta$ with $n = 49$, $p = 1$, and second-order neighborhood 115
A33 Simulation results for $\beta$ with $n = 100$, $p = 1$, and second-order neighborhood 116
A34 Simulation results for $\beta$ with $n = 25$, $p = 6$, and second-order neighborhood 116
A35 Simulation results for $\beta$ with $n = 49$, $p = 6$, and second-order neighborhood 116
A36 Simulation results for $\beta$ with $n = 100$, $p = 6$, and second-order neighborhood 116
Chapter 1

Introduction

1.1 Background

Many current data problems involve spatial effects that should be incorporated into statistical monitoring and modeling techniques. Sometimes the objective is to continuously monitor a process in order to quickly detect changes. Other times, the objective is to model a process in order to better characterize and understand its underlying features. Furthermore, monitoring and modeling techniques can frequently be combined in order to use available data to model the process, followed by surveillance of the process based on the model. In this research, the focus is on risk-adjusted monitoring and modeling of data sources that contain a spatial component.

The motivation for much of this research comes from a partnership with the Virginia Center for Housing Research and Metrostudy, a Hanley Wood Company, through which we obtained a comprehensive data set containing approximately 54 million publicly available house transaction records from municipalities across the United States between 2005 and 2014. The data include very fine spatial and temporal information, including latitude, longitude, and sale date of each closing record. Additionally, all transactions in our database
are categorized into four sale types: New Sale, Regular Resale, Real-Estate Owned (REO) Sale, and Foreclosure. We group New Sales and Regular Resales together as regular sales, where a New Sale refers to a newly built house and a Regular Resale refers to a previously owned house that is resold. A REO Sale often occurs after a homeowner defaults on their mortgage and a real-estate company takes ownership of the house which is then sold. For the sake of the analyses considered in this work, Foreclosures and REO sales are combined together and treated as undesirable events.

Given the characteristics of this data, we expect to see both changes over time and throughout the spatial domain. Thus, it is of interest to develop statistical methods to (a) monitor foreclosure rates and (b) model the spatial distribution of foreclosures across a region of interest. There are many ways to approach the development statistical techniques in the context of this data. We pursue both risk-adjusted monitoring techniques and fully Bayesian hierarchical modeling for spatially dependent data. Specifically, we extend risk-adjusted monitoring techniques to accommodate spatial dependence, as well as develop an objective manner to choose prior distributions for a sophisticated Bayesian hierarchical model for spatial data. The remainder of this chapter reviews existing statistical techniques that are the foundation for this research.

First, we review related work involving statistical process monitoring. Specifically, we focus on risk-adjusted monitoring, most of which pertains to the monitoring of surgical outcome quality. Then, since our motivating data lends itself to spatial analysis techniques, we provide an overview of the types of spatial data and existing work on modeling techniques for spatial data aggregated by region, commonly referred to as areal data. Then, we review the topic of objective priors for these fully Bayesian spatial models, since this is the approach considered in this work. Following the literature review in this chapter, Chapters 2, 3, 4, and 5 are self-contained manuscripts that contain the details and development of the statistical methodology used for the monitoring and modeling of spatial processes. Chapter 6 is a discussion of the dissertation as a whole followed by a description of the potential avenues for future work.
1.2 Statistical Process Monitoring

1.2.1 Overview

Broadly, statistical process monitoring is a suite of statistical techniques by which practitioners can actively assess whether or not their process is stable. Typically, control charts are used to rapidly detect shifts in a process parameter in order to alert the practitioner that a change in the process has occurred. Once a shift has been detected, the practitioner can further investigate the process to determine the reason for the process shift (Montgomery, 2007).

One frequently used control chart that is easily adopted for data from various distributions and can be implemented to quickly detect small shifts in a process parameter is the cumulative sum (CUSUM) control chart, first introduced by Page (1954). In contrast to the Shewhart chart, which only considers information from the most recently sampled observation, the CUSUM chart accumulates information from the observations over time.

1.2.2 Risk-adjusted Monitoring

While CUSUM charts exist for monitoring the parameters of a wide variety of distributions, one commonly used CUSUM chart is the Bernoulli CUSUM, which is used to continuously monitor the rate for binary outcomes of a particular event (Reynolds and Stoumbos, 1999). When using the Bernoulli CUSUM, it is assumed that the observations in the process being monitored all are independent trials with the same probability of “success”. However, in some situations we can use information about the population from which the observations are coming to improve this monitoring scheme. Often in medical settings when surgical outcome quality is being monitored, surgery patients can have drastically different health statuses before they undergo surgery. A more appropriate monitoring scheme for this situation is the risk-adjusted Bernoulli CUSUM control chart proposed by Steiner et al. (2000). The
risk-adjusted Bernoulli CUSUM chart appropriately adjusts the standard Bernoulli CUSUM chart to account for varying prior risks. This extension results in larger increases in the value of the one-sided CUSUM statistic for unexpected adverse events than for expected adverse events. Many other types of risk-adjusted monitoring schemes have been developed. For a thorough review of these methods, see Grigg and Farewell (2004) and Woodall et al. (2015).

In order to obtain predicted probabilities of the event of interest, it is necessary to fit a risk-adjustment model to what is referred to as Phase I data. Phase I data is used to define how the process should normally operate. An overview of Phase I issues and methods was given by Jones-Farmer et al. (2014). One commonly used method of obtaining predicted probabilities is to fit a logistic regression model to the Phase I data, however, there are many other possible techniques for doing this as described in Chapter 3.

1.3 Modeling Spatial Data

1.3.1 Types of Spatial Data

There are three major categories of spatial data: point referenced data, areal data, and point process data. Consider a $d$-dimensional region $D \subset \mathbb{R}^d$. Point referenced data is characterized by fixed locations within $D$, where each location has a response measurement. For example, weather stations located throughout a country might be used to measure the amount of annual rainfall across the region. If the region of interest is completely partitioned into a finite number of subregions, where each subregion has a measured response, this is called areal data. Typically, a structure defining which subregions are “neighbors” is imposed on this type of data. For example, disease incidence rates might be recorded for the counties of a state and “neighbors” might be defined as those counties sharing a border. Finally, point process data is characterized by locations of events that are random within the region of interest according to some underlying process. Additionally, if corresponding measurements
are taken for each random location, this is called a marked point process. For example, the 
epicenter locations of earthquakes in a region are point process data. If the magnitudes of 
the earthquakes are also recorded, this is then a marked point process. For a full description 
of these types of spatial data, see Cressie (1993). The introduction and work in Chapters 4 
and 5 focus on areal data, although this work could be used to develop similar techniques 
for other types of spatial data.

1.3.2 Modeling Areal Data

When working with areal data, it is very natural to think that the measurement in a given 
subregion is similar to the measurements taken at its neighboring subregions. One common 
approach for accommodating this type of spatial dependence is using a conditional autore-
gressive (CAR) model. This model specification was first introduced by Besag (1974). The 
idea behind the conditional autoregressive model is that the full conditional distribution of 
the response for a given subregion depends on its neighboring subregions. By using Brook’s 
lemma (Brook, 1964), the joint distribution of the response within each subregion can be 
obtained from the conditional specification given all other subregions. In its proper form, 
the CAR model can be used as a model for the observed data, however, the intrinsic CAR 
model proposed by Besag et al. (1991) cannot be used to model the observed data. This 
model is frequently used as a model for a spatially varying random effect to be included in 
a hierarchical Bayesian model in order to estimate the underlying response surface.

Typically, an intrinsic conditional autoregressive model is defined through proper conditional 
distributions for each spatial random effect that imply an improper joint distribution. Con-
sider the frequently used intrinsic CAR model for \( \omega = (\omega_1, \omega_2, \ldots, \omega_n)^T \) specified by its 
Gaussian conditional distributions

\[
(\omega_i | \omega_{-i}) \sim N \left( \frac{\sum_{j=1}^{n} g_{ij} \omega_j}{h_i}, \frac{1}{\tau_i h_i} \right)
\]  

(1.1)
where \( \omega_{-i} \) is the vector of the CAR elements for all subregions except subregion \( i \) and \( \tau_{\omega} > 0 \) is a precision parameter. In addition, \( g_{ij} \geq 0 \) is a measure of how similar subregions \( i \) and \( j \) are, \( g_{ij} = g_{ji} \), and \( h_i = \sum_{j=1}^{n_i} g_{ij} \). Notice that the mean of the conditional distributions is just a weighted average of the neighboring subregions. Conditional variance for subregions is more neighbors is smaller than the conditional variance for subregions with less neighbors. While these conditional distributions are proper distributions, intrinsic autoregressions (or intrinsic CARs) carry the notion that they are improper “densities” that are somehow made proper by imposing a constraint. The most frequently used constraint is a sum-zero constraint that ensures that the sum of the spatial random effects is equal to zero. For more details on the intrinsic CAR model, see Besag and Kooperberg (1995).

### 1.3.3 Prior Distributions for Hierarchical Spatial Models

In practice, a CAR component is commonly used to model a spatial random effect in a Bayesian hierarchical model. In this situation, many will typically use vague naïve prior distributions for the parameters of the CAR model. For examples, see Bernardinelli et al. (1995), Best et al. (1999), Bell and Broemeling (2000), and Lee (2013). The disadvantage of using naïve priors that follow known distributions is that it is often difficult to choose informative priors that are meaningful based on subject matter expertise. An alternative to choosing naïve prior distributions is to use “objective” or “default” priors. Three common choices are the Jeffreys prior (Jeffreys, 1961), the independence Jeffreys prior, and the reference prior (Bernardo, 1979). The Jeffreys prior for an unknown parameter \( \theta \) is given by \( |I(\theta)|^{1/2} \), where \( I(\theta) \) denotes the Fisher information matrix for \( \theta \). Similarly, the independence Jeffreys prior is obtained by assuming that the parameters are independent a priori and the Jeffreys rule is used to obtain the marginal prior of each of the parameters. The reference prior is obtained by factoring the joint prior distribution based on parameter importance and then using the Jeffreys rule algorithm based on integrated likelihoods. Several objective priors for commonly used spatial models will be explored in this work.
1.4 Dissertation Outline

The remainder of this dissertation contains four manuscripts that are at different stages of production. Chapter 2 provides a practical approach for implementing the risk-adjusted Bernoulli CUSUM chart that improves the time until detection of a signal. This work is a manuscript that has been accepted for publication in the *International Journal for Quality in Health Care* (Keefe et al., 2017b). Chapter 3 describes a method for spatial risk-adjustment in conjunction with risk-adjusted monitoring for concurrent observations. This chapter has been published in the *Journal of Applied Statistics* (Keefe et al., 2017a). Chapter 4 is a note that addresses the specification of intrinsic CAR models obtained by considering a limiting case of a proper CAR model that has been constrained. The motivation for Chapter 4 comes from a reviewer’s comment that was received regarding the manuscript that comprises Chapter 5 of this dissertation. The intent is to submit this note to the Miscellanea section of *Biometrika*. Chapter 5 proposes an objective Bayesian analysis for Gaussian hierarchical models with intrinsic CAR priors used to obviate the need of choosing an appropriate prior distribution for an advanced spatial model for areal data. This manuscript has been reviewed by *Bayesian Analysis* once and is currently being revised for resubmission. Chapter 6 summarizes all of the work included in this dissertation and describes some future research opportunities related to the material covered.
Bibliography


Chapter 2

Improved Implementation of the Risk-adjusted Bernoulli CUSUM Chart to Monitor Surgical Outcome Quality

MATTHEW J. KEEFE$^1$, JUSTIN B. LODA$^1$, AHMAD E. ELHABASHY$^{23}$, WILLIAM H. WOODALL$^1$

Accepted for Publication in the International Journal for Quality in Health Care

$^1$Department of Statistics, Virginia Tech, Blacksburg, VA 24061
$^2$Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061
$^3$Production Engineering Department, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt
Abstract

Many medical institutions are monitoring surgical outcome quality. The most popular method for the surveillance of binary surgical outcomes is the risk-adjusted Bernoulli cumulative sum (CUSUM) chart. This technique adjusts for each patient’s preoperative risk, thus taking into consideration the heterogeneity of patients when monitoring their outcomes. While this monitoring approach is becoming increasingly more popular, there is an opportunity for a practical improvement that incorporates available data in a more timely manner, leading to a dramatic effect on the time until detection of process deterioration. In this paper, we propose a simple approach that incorporates outcome information sooner and provide comparisons of the traditional implementation of the risk-adjusted Bernoulli CUSUM chart and our proposed implementation to demonstrate improved performance.

Keywords: Cumulative sum, risk-adjustment, statistical process monitoring, surgical performance


2.1 Introduction

Statistical process monitoring techniques are becoming more widely used in healthcare applications. In particular, methods for monitoring surgical outcomes are used to detect deterioration in surgical performance as quickly as possible to avoid undesirable consequences. For a comprehensive review of monitoring techniques for surgical outcome quality, see Woodall et al. (2015).

Frequently in healthcare applications, patients’ preoperative risks vary widely across the population. To account for the heterogeneity across patients, many risk-adjusted monitoring techniques have been developed. These techniques incorporate information about each patient’s potential risk factor characteristics, such as age, gender, and health status, into the calculation of a statistic that is monitored. For example, if a patient who was old and unhealthy died shortly after surgery, this result would be more likely than an instance where a young, healthier patient died. Risk-adjusted monitoring techniques use this information about each individual to allow for meaningful monitoring of surgical outcomes. We propose an implementation scheme for the risk-adjusted Bernoulli cumulative sum (CUSUM) chart that significantly improves the time until detection of process deterioration.

The risk-adjusted Bernoulli CUSUM chart proposed by Steiner et al. (2000) is a control chart that can monitor 30-day mortality rates prospectively, where each patient has a predicted probability of 30-day mortality based on a risk-adjustment model. This approach can be used to monitor the rate of other adverse events, not just mortality. When discussing our proposed implementation for risk-adjusted monitoring, we focus exclusively on the risk-adjusted Bernoulli CUSUM chart since it has the strongest theoretical justification and is the most popular approach (Steiner and Woodall, 2016).

When monitoring surgical outcomes, the outcome of interest is usually based on some prespecified period of time after surgery. For example, Steiner et al. (2000) considered death within 30 days after surgery. When outcomes such as this one are used, there is a period
of time during which the outcome of the patient may be unknown. Specifically, for patients who survive the entire time period (e.g., 30 days), their outcomes are unknown until the end of the time period. However, for patients who do experience the adverse event sooner than the end of the time period (e.g., death occurs within 30 days after surgery), the outcome is obtained earlier. The standard risk-adjusted Bernoulli CUSUM method monitors patients in the order in which they undergo surgery, despite the fact that information about many of their outcomes is known sooner than 30 days. For example, if we are monitoring 30-day mortality rate and a patient dies one day after surgery, the traditional implementation of the risk-adjusted Bernoulli CUSUM chart does not incorporate this outcome information into the chart until 29 days later. We propose an implementation scheme for the risk-adjusted Bernoulli CUSUM chart that incorporates patients’ surgical outcomes as soon as they are available, rather than waiting the length of the pre-specified time window to incorporate their information. Our proposed implementation considers all outcomes immediately.

2.2 Methods

2.2.1 Risk-adjusted Bernoulli CUSUM Procedure

The risk-adjusted Bernoulli CUSUM chart proposed by Steiner et al. (2000) is capable of monitoring binary outcomes while adjusting for prior risk of the adverse event occurring. A risk-adjustment model is fit to a Phase I sample so that predicted probabilities of the adverse event of interest (e.g., 30-day mortality) can be calculated. Phase I data is typically a historical sample collected that characterizes how the process being monitored operates under stable conditions. For a comprehensive overview of Phase I and its importance in statistical process monitoring, see Jones-Farmer et al. (2014). When monitoring surgical quality, a logistic regression model is typically fit with covariate information about the patients in order to obtain the predicted probability of the adverse event of interest.
The risk-adjusted CUSUM chart is designed to detect a shift in an odds ratio $R$ from $R_0$ to $R_1 > R_0$, where typically $R_0 = 1$. We let $p_t$ represent the predicted probability of the event of interest for the $t^{th}$ observation. Thus, the odds of the event can be calculated as $p_t / (1 - p_t)$. Therefore, under the in-control odds $R_0$, the odds of the event of interest are given by $R_0 p_t / (1 - p_t)$. Likewise, under $R_1$, the odds of the event of interest are given by $R_1 p_t / (1 - p_t)$. Thus, the corresponding in-control and out-of-control probabilities are given by

\[
p_{0t} = \frac{R_0 p_t}{1 - p_t + R_0 p_t} \quad \text{and} \quad p_{1t} = \frac{R_1 p_t}{1 - p_t + R_1 p_t},
\]

respectively. This leads directly to the calculation of the score for the risk-adjusted Bernoulli CUSUM given by

\[
W_t = \begin{cases} 
\log \left( \frac{1 - p_t + R_0 p_t}{1 - p_t + R_1 p_t} \right) & \text{if } y_t = 1 \\
\log \left( \frac{1 - p_t + R_0 p_t}{1 - p_t + R_1 p_t} \right) & \text{if } y_t = 0
\end{cases}
\]

(2.2)

The CUSUM statistics are given by

\[
S_t = \max(0, S_{t-1} + W_t),
\]

(2.3)

where $S_0 = 0$ and a signal is given when $S_t > h$ for $h > 0$.

Currently, the observations would be indexed in the order in which the patients undergo surgery. However, when operating under this assumption, the outcomes for those patients who do not survive are observed earlier than 30 days, but for patients who survive, outcomes are not observed until the end of the 30-day time window.

### 2.2.2 Proposed Approach

We show that it is beneficial to incorporate patient outcome information as soon as it is obtained, rather than waiting until the end of the waiting period associated with the outcome. To illustrate this result, we propose a simple implementation scheme for the risk-adjusted Bernoulli CUSUM chart that incorporates patients’ surgical outcomes as soon as they are
To illustrate our proposed implementation scheme of the risk-adjusted Bernoulli CUSUM chart, we will use the same data set from a United Kingdom center for cardiac surgeries as was used by Steiner et al. (2000). The data set consists of 6,994 patients from the years 1992 through 1998 and contains descriptive information such as surgery date, pre-operative Parsonnet score, and the number of days before any patient mortality. The Parsonnet score is a single value used to characterize a patient’s overall health status (Parsonnet et al., 1989). The first two years of data (1992-1993) were taken as Phase I data and were used to fit the following logistic regression model for risk-adjustment,

\[
\text{logit}(p_t) = -3.68 + 0.077X_t, \quad (2.4)
\]

where \(X_t\) is the Parsonnet score of patient \(t\) and \(p_t\) is the pre-operative risk of mortality within 30 days of surgery for this patient.

In this set of data, for those patients who died within 30 days after surgery, the distribution of days lived is heavily right-skewed. Figure 2.1 shows that the majority of patients from the first two years of data who died within 30 days, died within a week after surgery. Practically, if a patient dies 5 days after surgery, it does not seem reasonable to wait up to an additional 25 days to include the outcome in the monitoring scheme. Figure 2.1 illustrates why it could be very helpful to incorporate outcome information as soon as it is available. If deaths occur closer to the 30-day point, then use of our proposed method would not be as advantageous.

### 2.2.3 Proposed Implementation

We propose an implementation scheme where we continue to consider the patients in the order of their surgical operations as is done in the traditional implementation (Steiner et al., 2000). Although there is no way to know in advance if a particular patient is going to live or die, we initially assume in advance that all patients are going to survive until the
end of the 30-day time window, and then, as the operations occur and the actual adverse outcomes are obtained, the chart is updated accordingly, rather than waiting 30 days. Using this implementation, all patients would be assumed to survive until determined otherwise. Specifically, the chart would be updated when patients who were assumed to survive actually die within 30 days after surgery. Thus, as soon as a patient dies, we would incorporate this outcome into the chart immediately. In this manner, some previously charted CUSUM statistics would dynamically change throughout the monitoring process.

The proposed implementation is a recursive process that would be initiated every time an adverse event occurs. In this manner, there would be a 30-day moving window in which the CUSUM statistics could be updated. For instance, if we are now considering the outcomes obtained on the 45th day of monitoring, then this backtracking window would update all outcomes considered from the 16th day to the 45th day. Patients who were not operated on within the last 30 days no longer need to have their control chart values updated. It is important to realize that at the end of the monitoring process once all outcome information has been obtained, the CUSUM chart using the proposed implementation will be identical to that with the traditional implementation. The dynamic updating of the CUSUM statistics through time allows our proposed implementation to detect process deterioration sooner than the traditional implementation.
Table 2.1: Example Data Set

<table>
<thead>
<tr>
<th>Patient</th>
<th>Operation Day</th>
<th>Survived 30 days?</th>
<th>Days Lived</th>
<th>Outcome Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td>30+</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>No</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>No</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>Yes</td>
<td>30+</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>Yes</td>
<td>30+</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>No</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>No</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Yes</td>
<td>30+</td>
<td>33</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>No</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>Yes</td>
<td>30+</td>
<td>35</td>
</tr>
</tbody>
</table>

2.2.4 Illustration

As an illustration of the proposed implementation, consider the small, artificial data set of surgical outcomes with an associated 30-day mortality response given in Table 2.1. Each row corresponds to a patient. Information regarding whether or not each patient survived 30 days after surgery, as well as the number of days lived after surgery was recorded. The number of days lived for patients who survived 30 days after surgery was recorded as 30+. The day on which outcome information was obtained is calculated based on each patient’s operation day and the number of days they lived after surgery.

Using the proposed implementation scheme, each plotted CUSUM statistic corresponds to a patient. However, the chart is updated through time for each day that new outcome information is obtained. Figure 2.2 illustrates how the chart is implemented to update information as it is obtained. On Day 1, all patients operated on thus far are assumed to survive 30 days, and thus all CUSUM statistics are zero. The CUSUM statistics for the third patient and all subsequent patients are updated on Day 3 when it was learned that the third patient died. All patients after the third patient on Day 3 are still assumed to survive 30 days, and thus result in a decreasing trend of CUSUM statistics. On Day 4, patients 2 and 6 died and the CUSUM statistics are updated. Finally, on Day 7 outcome information for patients 7 and 9 was obtained and the CUSUM statistics were updated again. Hence, previously plotted
CUSUM statistics are dynamically updated as outcome information is obtained, similar to reliability monitoring schemes involving dynamically changing observations (Yashchin, 2012). Also, note that the resulting CUSUM chart for Day 7 is identical to the CUSUM chart obtained after Day 35 using the traditional implementation. For this illustration, a control limit of $h = 2.5$ is used, but another control limit could be used in practice. If the proposed implementation is used, the chart signals on Day 7, whereas if the traditional implementation is used, the chart would not signal until Day 34. Clearly, detection time can be reduced by using the proposed implementation.
2.2.5 Simulation Study

A simulation study was conducted using the UK cardiac surgery data to compare the in-control and steady state out-of-control performance of the proposed implementation scheme with that of the traditional implementation of the Bernoulli risk-adjusted CUSUM chart. For both implementations, each CUSUM statistic plotted on the control chart corresponds to a patient, where patients are ordered by the date of their operation. It is more informative to consider the average run length (ARL) in number of days, rather than in number of patients. It is important to note that our proposed implementation will never signal before the traditional implementation in terms of the number of patients. When implementing the chart, however, the time until a signal can be determined by the number of days since monitoring began. The benefit in our implementation scheme is clearly seen in recognizing deterioration sooner, in terms of number of days, rather than number of patients, due to removing the 30 day wait time restriction. For our simulation study, the number of operations for a given day was drawn with replacement from the empirical distribution of the Phase I data and varies from one to eight. We also considered simulations where we fixed the number of operations per day to assess performance of the proposed implementation as the number of operations per day increases.

The simulation procedure used to compare our proposed implementation method to the traditional implementation method can be described in the following steps. For each simulated patient $t$,

1. Sample with replacement a Parsonnet score from the in-control empirical distribution.
2. Use Equation (2.4) to determine the predicted probability, $p_t$, of death within 30 days of surgery. Adjust $p_t$ based on the assumed odds ratio $R$.
3. Generate a Bernoulli random variable with probability of “success” $p_t$.
4. With the outcome obtained in Step 3, calculate the CUSUM statistic using Equation (2.3) with $R_1 = 2$. 
5. Repeat steps 1-4 until $S_t > h$.

To be consistent with the work of Steiner et al. (2000), we used an upper control limit of $h = 4.5$ which produces an in-control ARL, in terms of the number of patients, of approximately 7,400 (see Tian et al. (2015) for details). We set up the control chart so that it is designed to detect a shift of $R_1 = 2$. For the out-of-control simulations, the process was initially simulated as in-control under the baseline model with $R = 1$ for the first 50 patients to achieve steady state conditions and the odds of death within 30 days was shifted after patient 50. We considered various values of the odds ratio $R$ between 1 and 10. Additionally, we considered values of $R = 2, 3, 4, 5$ for fixed numbers of operations per day of 1 through 10. Each ARL simulation result is based on 1,000 simulated control charts.

### 2.3 Results

The proposed implementation scheme shows improved detection time of process deterioration in terms of days. Typically, monitoring schemes are compared using ARL, where the in-control ARLs of the schemes are set to be equal and the out-of-control ARLs are observed for different size shifts. The results of the simulation study for various odds ratios $R$ are provided in Table 2.2. We note that in our case, the in-control ARLs of the two implementation schemes are close, but not quite equal. With the proposed implementation, the in-control ARL is on average 15.8 days less than with the traditional implementation. Any false alarm obtained by the traditional method would likely be obtained at an earlier time by our proposed implementation scheme. As the size of the odds ratio $R$ increases, the average time until detection for the proposed implementation improves more. For this application, we can signal process deterioration up to 29 days sooner; however, if the waiting period for the response of interest were more than 30 days, more time could be saved. For example, in organ transplantation applications the monitoring of survival times is frequently used rather than the risk-adjusted Bernoulli CUSUM because the waiting period after transplantation is
Table 2.2: ARL comparison (in days) for traditional and proposed implementation schemes

<table>
<thead>
<tr>
<th>$R$</th>
<th>Traditional</th>
<th>Proposed</th>
<th>Difference (days)</th>
<th>% Reduction Relative to Traditional Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2080.9</td>
<td>2065.1</td>
<td>15.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1.5</td>
<td>175.5</td>
<td>159.0</td>
<td>16.5</td>
<td>9.4</td>
</tr>
<tr>
<td>2</td>
<td>79.0</td>
<td>61.5</td>
<td>17.5</td>
<td>22.2</td>
</tr>
<tr>
<td>2.5</td>
<td>60.4</td>
<td>41.8</td>
<td>18.6</td>
<td>30.8</td>
</tr>
<tr>
<td>3</td>
<td>51.1</td>
<td>31.8</td>
<td>19.3</td>
<td>37.7</td>
</tr>
<tr>
<td>3.5</td>
<td>47.1</td>
<td>27.4</td>
<td>19.7</td>
<td>41.8</td>
</tr>
<tr>
<td>4</td>
<td>44.5</td>
<td>24.3</td>
<td>20.2</td>
<td>45.4</td>
</tr>
<tr>
<td>4.5</td>
<td>42.4</td>
<td>21.9</td>
<td>20.5</td>
<td>48.4</td>
</tr>
<tr>
<td>5</td>
<td>41.0</td>
<td>19.8</td>
<td>21.1</td>
<td>51.6</td>
</tr>
<tr>
<td>5.5</td>
<td>39.8</td>
<td>18.5</td>
<td>21.4</td>
<td>53.6</td>
</tr>
<tr>
<td>6</td>
<td>39.2</td>
<td>17.6</td>
<td>21.6</td>
<td>55.2</td>
</tr>
<tr>
<td>6.5</td>
<td>38.6</td>
<td>16.7</td>
<td>21.9</td>
<td>56.8</td>
</tr>
<tr>
<td>7</td>
<td>37.9</td>
<td>15.6</td>
<td>22.3</td>
<td>58.9</td>
</tr>
<tr>
<td>7.5</td>
<td>37.3</td>
<td>14.9</td>
<td>22.4</td>
<td>60.1</td>
</tr>
<tr>
<td>8</td>
<td>37.0</td>
<td>14.5</td>
<td>22.5</td>
<td>60.7</td>
</tr>
<tr>
<td>8.5</td>
<td>36.6</td>
<td>13.9</td>
<td>22.6</td>
<td>61.9</td>
</tr>
<tr>
<td>9</td>
<td>36.3</td>
<td>13.5</td>
<td>22.8</td>
<td>62.7</td>
</tr>
<tr>
<td>9.5</td>
<td>36.0</td>
<td>13.0</td>
<td>23.0</td>
<td>63.8</td>
</tr>
<tr>
<td>10</td>
<td>35.7</td>
<td>12.5</td>
<td>23.2</td>
<td>65.0</td>
</tr>
</tbody>
</table>

usually one year (Biswas and Kalbfleisch, 2008; Collett et al., 2009; Neuberger et al., 2010). In this case, our proposed implementation makes the risk-adjusted Bernoulli CUSUM more similar to methods with continuous updating schemes that consider time until event data and would lead to significantly improved performance in terms of days until detection of a process change.

Another important aspect to notice is that the time until detection of process deterioration will depend on the number of operations performed per day. Figure 2.3 shows that the out-of-control ARL improves for both the traditional and proposed implementations of the risk-adjusted Bernoulli CUSUM as the number of operations per day increases, yet the difference in days until detection between the two implementations does not change with the number of operations per day. However, with the proposed implementation one always
detects the shift sooner than with the traditional implementation. As expected, larger shifts in the process result in lower ARLs for both methods. In the limiting case, if all adverse events occurred immediately (i.e. death on the first day) the improvement in detection for the proposed implementation would be exactly 29 days.

2.4 Discussion

With the traditional implementation of the risk-adjusted Bernoulli CUSUM chart, one monitors patient by patient in the order of operation with a waiting period to determine the outcome. Practically, it is inefficient to wait a specific time period, such as 30 days, if information about some of the outcomes is available much sooner. We have proposed a more practical and appealing implementation scheme for the risk-adjusted Bernoulli CUSUM chart that incorporates outcome information as soon as it is available. We have illustrated that the proposed implementation significantly improves the time until detection of deterioration in the process, especially when most adverse outcomes occur toward the beginning of the waiting period.

Performance of our proposed monitoring scheme in terms of time until detection of process deterioration is limited by the effect of estimation error inherent in fitting the risk-adjustment model. Furthermore, this method is intended for binary outcomes that have a waiting period required before obtaining the outcome. If deaths occur closer to the end of the waiting period, then use of our proposed method would not be as advantageous, but would still perform as well as the traditional approach.

Other practical issues regarding the implementation of the risk-adjusted Bernoulli CUSUM chart have been addressed. For example, Tian et al. (2015) discussed the impact of varying patient populations and its effect on chart performance. As a consequence, an appropriate monitoring scheme to account for varying patient populations is the use of dynamic probability control limits (Zhang and Woodall, 2015). Additionally, situations arise in which
Figure 2.3: ARL (in days) by number of operations per day for (a) $R = 2$, (b) $R = 3$, (c) $R = 4$, and (d) $R = 5$ (based on 1,000 simulated charts)
observations happen concurrently and there is no way to determine the exact order (Keefe et al., 2016). Furthermore, Paynabar et al. (2012) explored the importance of including other covariate information, such as surgeon information, into the risk-adjustment procedure. Jones and Steiner (2012) studied the effect of Phase I estimation error on the performance of the risk-adjusted Bernoulli CUSUM chart. Also, Tang et al. (2015) developed a risk-adjusted CUSUM chart for multi-responses, in cases where the response is not binary, but rather has several categories. This multi-response technique was further developed by Zhang et al. (2017) to include dynamic probability control limits. Our proposed implementation could be applied directly to accommodate dynamic probability control limits, concurrent observations, covariate information, or multi-responses in order to improve performance.
Bibliography


Chapter 3

Monitoring Foreclosure Rates with a Spatially Risk-adjusted Bernoulli CUSUM Chart for Concurrent Observations

Matthew J. Keefe\(^1\), Christopher T. Franck\(^1\), William H. Woodall\(^1\)

Paper Published in *Journal of Applied Statistics*  

\(^1\)Department of Statistics, Virginia Tech, Blacksburg, VA, USA
Abstract

Frequently in process monitoring, situations arise in which the order that events occur cannot be distinguished, motivating the need to accommodate multiple observations occurring at the same time, or concurrent observations. The risk-adjusted Bernoulli cumulative sum (CUSUM) control chart can be used to monitor the rate of an adverse event by fitting a risk-adjustment model, followed by a likelihood ratio-based scoring method that produces a statistic that can be monitored. In our paper, we develop a risk-adjusted Bernoulli CUSUM control chart for concurrent observations. Furthermore, we adopt a novel approach that uses a combined mixture model and kernel density estimation approach in order to perform risk-adjustment with regard to spatial location. Our proposed method allows for monitoring binary outcomes through time with multiple observations at each time point, where the chart is spatially adjusted for each Bernoulli observation’s estimated probability of the adverse event. A simulation study is presented to assess the performance of the proposed monitoring scheme. We apply our method using data from Wayne County, Michigan between 2005 and 2014 to monitor the rate of foreclosure as a percentage of all housing transactions.

Keywords: Cumulative sum, mortgage default, kernel density estimation, spatial risk-adjustment, statistical process monitoring, Wayne County
3.1 Introduction

Process monitoring methods have predominantly been applied to industrial settings where items being manufactured are typically homogeneous. However, in certain instances, such as medical applications, the subjects’ baseline risks often vary widely. To accommodate the heterogeneous collection of subjects, risk-adjusted monitoring techniques have been developed. Risk-adjustment procedures using the difference between observed and expected outcomes were devised by Lovegrove et al. (1997), Lovegrove et al. (1999), and Poloniecki et al. (1998). Risk-adjusted Bernoulli cumulative sum (CUSUM) charts (Steiner et al., 2000) and Shewhart charts (Cook et al., 2003) have also been developed. Techniques that use sequential probability ratio tests for monitoring scenarios that require risk-adjustment were introduced by Spiegelhalter et al. (2003). For an overview of these standard risk-adjusted monitoring techniques, see Grigg and Farewell (2004) and Woodall et al. (2015). Sego et al. (2009) proposed a risk-adjusted CUSUM chart to monitor survival times and various researchers have introduced risk-adjusted exponentially weighted moving average-based methods (Grigg and Spiegelhalter, 2007; Steiner and Jones, 2010; Cook et al., 2011).

We will extend the risk-adjusted Bernoulli CUSUM method proposed by Steiner et al. (2000), who provided justification for a CUSUM chart that monitors 30-day mortality rates for one patient at a time where each patient has a predicted probability of 30-day mortality estimated before the surgical procedure. In general, the method can be used to monitor the rate of any adverse event of interest.

We propose a spatially risk-adjusted Bernoulli CUSUM chart that (a) uses kernel density estimation and mixture models to perform spatial risk-adjustment, and (b) accommodates concurrent observations as an extension of the method of Steiner et al. (2000). We apply our proposed method to housing transactions classified as either regular sales or foreclosures. Each housing transaction has its own respective predicted probability of being a foreclosure. We refer to our proposed chart as the spatially risk-adjusted Bernoulli CUSUM chart for concurrent observations, where a group of observations for which the order of occurrence
cannot be distinguished are together considered concurrent observations.

The rise of foreclosure rates in the USA is a well-known consequence of the Great Recession. With the collapse of the housing market, many homeowners defaulted on their mortgages causing foreclosure rates to increase across the country. Quercia and Stegman (1992) provided an overview of research studying the process of mortgage default and foreclosure. Additionally, the literature suggests geographical consequences of the collapse of the sub-prime mortgage boom and housing bubble that were both local and widespread (Christophers, 2009; Martin, 2011). This drastic shift in the rate of foreclosure in the housing market motivates the need to monitor foreclosure rates in order to detect a shift in the market.

There have been many analyses of mortgage default and foreclosure. For example, Campbell and Cocco (2011) implemented dynamic models to determine a household’s decision to default on its mortgage. Hazard modeling is another approach that has been used to describe the mortgage default process (Deng et al., 2000; Ciochetti et al., 2003; Clapp et al., 2006). Schwartz and Torous (1993) and Pennington-Cross (2010) used generalized linear models to study foreclosure trends. Foreclosure records within the housing market provide an abundant source of data that is available for the development of statistical methods and economic analysis. We apply our spatially risk-adjusted monitoring technique for concurrent events to housing transactions in Wayne County, Michigan to monitor foreclosure rates. Our method can be used to detect an increase in the rate of the event, allowing for an early warning that the process being monitored has shifted from its usual state of equilibrium.

The remainder of our paper is organized in the following manner. In Section 2, we discuss (a) the CUSUM monitoring procedure using the standard non-risk-adjusted method, (b) the risk-adjusted method, (c) the proposed extension that accommodates concurrent observations, and (d) the proposed method of spatial risk-adjustment. In Section 3, we describe a simulation study used to study the performance of the proposed monitoring scheme. Section 4 contains an illustration of how our proposed spatially risk-adjusted monitoring scheme for concurrent observations would be implemented using closing records from Wayne County,
Michigan for the years 2005-2014. Finally, in Section 5, we provide conclusions and discuss the merits and caveats of our proposed method.

3.2 CUSUM and risk-adjustment procedure

3.2.1 Standard CUSUM

The standard CUSUM monitoring scheme was first introduced by Page (1954). This procedure, commonly referred to as the tabular CUSUM, is used to detect a shift in the process parameter of interest, denoted \( \theta \). To implement a standard CUSUM control chart, one would observe the independent random variables \( X_1, X_2, ... \), and monitor the statistics given by

\[
S_t = \max(0, S_{t-1} + W_t), \quad t = 1, 2, ...
\]

such that \( S_0 = 0 \) and \( W_t \) is the likelihood ratio-based score for the observation(s) at time \( t \) given by

\[
W_t = \log \left( \frac{f(x_t; \theta_1)}{f(x_t; \theta_0)} \right), \quad (3.2)
\]

where \( f(x_t; \theta_0) \) is the likelihood function under the in-control process parameter \( \theta_0 \) and \( f(x_t; \theta_1) \) is the likelihood function under the out-of-control process parameter \( \theta_1 \). The goal of the CUSUM monitoring scheme is to detect a shift in the process parameter from \( \theta_0 \) to \( \theta_1 \).

For a one-sided CUSUM control chart designed to detect increases in the process parameter, the chart signals and the process is deemed out of control if \( S_t > h^+ \), where \( h^+ \) is the control limit chosen to produce a chart with a desired in-control average run length (ARL). The in-control ARL is defined as the average number of statistics plotted on the chart until one is above the control limit, assuming the process is operating under \( \theta_0 \).

Using this likelihood ratio-based approach, Reynolds and Stoumbos (1999) proposed the Bernoulli CUSUM monitoring scheme, which allows continuous monitoring of binary out-
comes. The resulting likelihood ratio-based value used to detect a shift from the in-control probability of the event of interest $p_0$ to $p_1 > p_0$ can be calculated by

$$W_t = y_t \log \left( \frac{p_1}{p_0} \right) + (1 - y_t) \log \left( \frac{1 - p_1}{1 - p_0} \right),$$  \hspace{1cm} (3.3)$$

where $y_t$ is the outcome of the $t^{th}$ observation that is 0 or 1, depending whether or not the $t^{th}$ observation corresponds to the event of interest. While this type of CUSUM chart is appropriate for monitoring the rate of binary outcomes over time, it does not consider the risk of the event occurring when calculating $W_t$. It is reasonable to assume that just as medical patients have varying risks of adverse events, each mortgage has a prior risk of foreclosure. Observing a foreclosure in an area where foreclosures are rare provides more evidence of a shift in the market than a foreclosure observed in an area where the foreclosure risk is high. The risk-adjusted Bernoulli CUSUM control chart is appropriate for this situation.

### 3.2.2 Risk-adjusted Bernoulli CUSUM

The risk-adjusted Bernoulli CUSUM chart developed by Steiner et al. (2000) appropriately adjusts the standard Bernoulli CUSUM chart to account for varying prior risks. This extension results in larger increases in the value of the one-sided CUSUM statistic for unexpected adverse events than for expected adverse events.

In order to obtain predicted probabilities of the event of interest, it is necessary to fit a risk-adjustment model to what is referred to as Phase I data. Phase I data is used to define how the process should operate under stability with in-control parameter $\theta_0$. An overview of Phase I issues and methods was given by Jones-Farmer et al. (2014). Typically, a logistic regression model is fit with independent variables that may help determine whether or not an event is likely to occur in order to estimate pre-event probabilities. Other techniques for estimating risks for each observation, such as kernel density estimation (Rosenblatt, 1956) for events and non-events in conjunction with calculation of membership probabilities from
a mixture of the two estimated densities could also be used. This non-parametric binary regression technique was used by Kelsall and Diggle (1998) to estimate spatial variation in the risk of disease.

A key difference between the non-risk-adjusted chart and the risk-adjusted chart is that each observation now has its own predicted probability of the event occurring. Thus, we can no longer consider a shift in the probability of the event from the in-control \( p_0 \) to \( p_1 > p_0 \). Instead, it is particularly useful to express the desired shift to detect in terms of an odds ratio. Suppose we want to detect a shift in the odds expressed as an odds ratio from \( R_0 \) to \( R_1 > R_0 \), where \( R_0 \) is frequently chosen to be one. We denote the estimated probability of the event of interest at time \( t \) as \( p_t \). Thus, the odds of the event are given by

\[
\text{Odds(event)} = p_t : (1 - p_t). \tag{3.4}
\]

With this, the odds of the event under \( R_0 \) can be calculated as

\[
R_0 p_t : (1 - p_t), \tag{3.5}
\]

and under \( R_1 \) as

\[
R_1 p_t : (1 - p_t). \tag{3.6}
\]

The corresponding in-control and out-of-control probabilities are given by

\[
p_{0t} = \frac{R_0 p_t}{1 - p_t + R_0 p_t} \quad \text{and} \quad p_{1t} = \frac{R_1 p_t}{1 - p_t + R_1 p_t}. \tag{3.7}
\]

Substituting the results from Equation (3.7) for \( p_{0t} \) and \( p_{1t} \) into the standard Bernoulli CUSUM statistic given in Equation (3.3) for \( p_0 \) and \( p_1 \), respectively, produces \( W_t \) for the risk-adjusted Bernoulli CUSUM chart of Steiner et al. (2000) given by

\[
W_t = y_t \log \left( \frac{p_{1t}}{p_{0t}} \right) + (1 - y_t) \log \left( \frac{1 - p_{1t}}{1 - p_{0t}} \right), \tag{3.8}
\]
which is equivalent to

\[ W_t = \begin{cases} 
\log \left[ \frac{(1-p_t + R_0 p_t) R_t}{(1-p_t + R_1 p_t) R_0} \right] & \text{if } y_t = 1 \\
\log \left[ \frac{1-p_t + R_0 p_t}{1-p_t + R_1 p_t} \right] & \text{if } y_t = 0 
\end{cases} \]  \quad (3.9) 

One limitation of this method is that it is based on the assumption that there is only one observation at each time point. In a situation where we know the exact order of events with certainty, the current risk-adjusted Bernoulli CUSUM is acceptable. However, in cases where there are multiple observations at each time point, the existing method is no longer appropriate. For example, one could imagine that on any given day, multiple housing transactions occur. With no information other than the transaction date, it is impossible to identify the order that houses sell or foreclose within a day. In addition, this level of data aggregation may not be large enough to be of any practical consequence. Thus, we propose a useful extension of the risk-adjusted Bernoulli CUSUM chart that is an analogous method that permits monitoring binary outcomes in the case where there are multiple observations at each time point and each observation has its own estimated probability of the adverse event of interest.

### 3.2.3 Modified risk-adjusted Bernoulli CUSUM

Following the approach of Page (1954), we use the log-likelihood ratio to calculate an appropriate CUSUM statistic. Thus, for the risk-adjusted Bernoulli CUSUM chart for concurrent observations has \( W_t \) as follows

\[ W_t = \log \left[ \frac{\prod_{i=1}^{n_t} p_{1ti}^{y_{ti}} (1 - p_{1ti})^{1-y_{ti}}}{\prod_{i=1}^{n_t} p_{0ti}^{y_{ti}} (1 - p_{0ti})^{1-y_{ti}}} \right], \]  \quad (3.10) 

where \( n_t \) is the number of observations at time \( t \), \( y_{ti} \) equals 1 for an event and 0 otherwise for the \( i^{th} \) observation at time \( t \) \((i = 1, 2, ... n_t)\), and \( p_{0ti} \) and \( p_{1ti} \) are the predicted probabilities
under $R_0$ and $R_1$ for the $i^{th}$ observation at time $t$, respectively, using

\[ p_{0ti} = \frac{R_0 p_{ti}}{1 - p_{ti} + R_0 p_{ti}} \quad \text{and} \quad p_{1ti} = \frac{R_1 p_{ti}}{1 - p_{ti} + R_1 p_{ti}}. \] (3.11)

Substitution into Equation (3.10) and simplification leads to

\[ W_t = \sum_{i=1}^{n_t} \left[ y_{ti} \log \left( \frac{p_{1ti}}{p_{0ti}} \right) + (1 - y_{ti}) \log \left( \frac{1 - p_{1ti}}{1 - p_{0ti}} \right) \right]. \] (3.12)

Note that if $R_0$ is chosen to be 1, as is typically the case, then $p_{0ti}$ is simply the estimated probability of the event based on the risk-adjustment model. Also, notice that the result given in Equation (3.12) is equivalent to the sum of all of the values for the likelihood ratio-based CUSUM scores for the concurrent observations. So, with this procedure one intuitively combines all of the CUSUM scores for one time point and then plots the CUSUM statistic.

It is important to realize that if this is not done and the observations at any given time point are arbitrarily ordered within that time point, the behavior of the control chart would depend on the order in which the concurrent observations are considered. For example, consider the situation where multiple housing transactions occur on a single day. If the observations on that day are ordered such that all of the foreclosures are recorded prior to the regular housing transactions, the chart would more likely signal prematurely. However, by implementing our modification, we can appropriately monitor foreclosure rates using a risk-adjusted Bernoulli CUSUM chart for concurrent observations.

\subsection*{3.2.4 Spatial risk-adjustment}

There are many possible ways to estimate the prior risk of the event of interest, such as the use of logistic regression with explanatory variables that affect the probability of the event, as in Steiner et al. (2000). However, when the location of the events within a spatial domain are available, the locations can be used to spatially estimate the probability of an
event occurring using the Phase I sample. Our proposed technique that uses mixture models and kernel density estimation has the potential to extend process monitoring to a host of economic, epidemiological, and other societal problems that involve spatial dependence.

Briefly, kernel density estimation is a method used to approximate the distribution of a random variable, much like a histogram, but with a smoother estimate across the domain of the random variable (Rosenblatt, 1956). Gaussian densities centered at each observation with variance determined by a user-specified bandwidth parameter are defined as kernels. The average of the kernels across the sample space approximates the true density of the distribution that generated the observed data. For a spatial region characterized by latitude and longitude, we use two-dimensional Gaussian kernel density estimation, using a bandwidth parameter chosen by a $k$-fold cross validation adaptation of the likelihood-based approach provided by Kelsall and Diggle (1998). Separate kernel density estimates are computed for events and non-events, and the probability of the event is computed across the spatial domain by considering a mixture of the two distributions given by

$$f = \pi_1 f_1 + (1 - \pi_1) f_2,$$  \hspace{1cm} (3.13)

where $f_1$ is the unknown spatial distribution of the event of interest, $f_2$ is the unknown spatial distribution of the non-events, $0 < \pi_1 < 1$ is the mixture weight representing the proportion of all observations categorized as events of interest and hence $(1 - \pi_1)$ is the mixture weight representing the proportion of all observations that are non-events. This mixture model approach considers the kernel density estimate for events relative to the mixture kernel density estimate to compute the probability of an event across the region of interest. From McLachlan and Peel (2004), the probability of an observation belonging to the density $f_1$, denoted by $q$, given the observation’s geographic longitude $l_1$ and latitude $l_2$, the mixture weight, and $f_2$ is given by

$$q(l_1, l_2) = \frac{\pi_1 f_1(l_1, l_2)}{\pi_1 f_1(l_1, l_2) + (1 - \pi_1) f_2(l_1, l_2)}.$$

\hspace{1cm} (3.14)
Furthermore, since the proportion of observations belonging to $f_1$ and $f_2$ are known from the Phase I data, the estimated probability of the event of interest given a location $(l_1, l_2)$ is

$$\hat{q}(l_1, l_2) = \frac{\hat{\pi}_1 \hat{f}_1(l_1, l_2)}{\hat{\pi}_1 \hat{f}_1(l_1, l_2) + (1 - \hat{\pi}_1) \hat{f}_2(l_1, l_2)},$$

(3.15)

where $\hat{f}_1$ and $\hat{f}_2$ are the estimated densities for events and non-events, respectively, obtained by using kernel density estimation. Here $\hat{\pi}_1$ is the estimated mixture weight corresponding to the proportion of observations categorized as events which is estimated by calculating the proportion of observations in the Phase I data that are events. More details on the bandwidth selection of the kernel density estimation can be found in the appendix. This new approach to risk-adjustment within the process monitoring framework allows spatial dependence to be incorporated into the monitoring scheme in order to appropriately adjust for regions at higher and lower risk of the event of interest.

### 3.3 Simulation study

To determine whether variation in the number of concurrent observations at each time point affects chart performance, a simulation study was conducted to compare fixed and varying numbers of concurrent observations. The in-control ($R_1 = 1$) and steady state out-of-control ($R_1 = 2$) performance of the risk-adjusted Bernoulli CUSUM chart for concurrent observations was examined. All results are based on 10,000 simulated charts that used a fixed control limit $h^+ = 4$. The number of concurrent observations $n_t$ was fixed at 5, 50, and 250, in addition to varying uniformly over the intervals from 2 to 10, 2 to 100, and 2 to 500. Predicted probabilities of the event were randomly generated from a Uniform(0,1) distribution.

The results in Table 3.1 indicate that introducing variation in the number of concurrent observations produces no discernible change in the first two moments of the run length distribution of the monitoring scheme. We note that a drastic systematic shift in the number
Table 3.1: Run length performance for varying number of concurrent observations. ARL and standard deviation of run length (SDRL) are presented.

<table>
<thead>
<tr>
<th>Shift ($R_1$)</th>
<th>$n_t = 5$</th>
<th>$n_t = 50$</th>
<th>$n_t = 250$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARL</td>
<td>SDRL</td>
<td>ARL</td>
</tr>
<tr>
<td>1</td>
<td>544.1</td>
<td>531.7</td>
<td>257.0</td>
</tr>
<tr>
<td>2</td>
<td>17.3</td>
<td>11.1</td>
<td>2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shift ($R_1$)</th>
<th>$n_t = 2$ to $10$</th>
<th>$n_t = 2$ to $100$</th>
<th>$n_t = 2$ to $500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARL</td>
<td>SDRL</td>
<td>ARL</td>
</tr>
<tr>
<td>1</td>
<td>505.1</td>
<td>495.1</td>
<td>302.4</td>
</tr>
<tr>
<td>2</td>
<td>14.2</td>
<td>9.4</td>
<td>2.3</td>
</tr>
</tbody>
</table>

of concurrent observations would change the performance of the chart. For example, an increase in the number of concurrent observations may elicit a premature signal. This is a disadvantage of any monitoring scheme with a fixed control limit. As pointed out by Tian et al. (2015) and Tang et al. (2015), if the risk population changes, the chart performance is not guaranteed. To resolve this, a separate control chart could be used to monitor the number of concurrent observations at each time point if one is concerned about a systematic shift in this quantity. As in other process monitoring methods, we assume that the Phase I sample collected is representative of the process being monitored. Therefore, the distribution of the number of concurrent observations in Phase II would be expected to be similar to that of Phase I. If this distribution has changed, the chart would need to be recalibrated. Additionally, the ability of the chart to detect a shift in the process remains intact even when there is variation in the number of concurrent observations throughout the monitoring process.
3.4 Case study: monitoring foreclosure rates

3.4.1 Data description

The rise of foreclosure rates beginning in 2006 and spreading through the housing market in the following years in the USA is a well-known consequence of the Great Recession. Immergluck (2008) noted that despite the increase of foreclosure risk in late 2006, awareness of this issue was not brought to the attention of planners and policymakers until the year 2007. Specifically, Dewar et al. (2015) discussed the impact of foreclosures on housing disinvestment in Detroit, Michigan. Due to the pronounced effect that foreclosures had on the housing stock in Detroit, we chose to focus on housing transactions in Wayne County, Michigan between the years of 2005 and 2014. Figure 3.1 illustrates the changing nature of foreclosures in Wayne County for 2005-2007.

The data set was obtained from our partnership with Metrostudy, a Hanley Wood Company and contains house closing records during the years 2005-2014. The Wayne County, Michigan data are a subset of 370,517 records from a larger database of approximately 54 million publicly available records from municipalities across the USA. The data include very fine spatial and temporal information, including latitude, longitude, and sale date of each closing record. Since many of the houses close on the same date and exact locations of each house are known, our spatially risk-adjusted Bernoulli CUSUM for concurrent observations is more appropriate than the standard risk-adjusted Bernoulli CUSUM.

In order to apply our spatially risk-adjusted method, it is necessary to obtain a model that provides predicted probabilities of foreclosure. These probabilities, or prior risks of foreclosure, can be calculated from a model fit to the Phase I data. The Phase I data contain measurements on a process over a baseline time period that practitioners can use to learn more about the process and use to justify the construction of a suitable process monitoring method. For our data, we used the first six months of 2005 as our Phase I data, since this time period is known to be before the Great Recession. The empirical foreclosure
Figure 3.1: Sale type distributions for Wayne County, Michigan for (a) 2005, (b) 2006, (c) 2007 and (d) yearly percentage breakdown
rate, as a percentage of the total number of sales, between 1 January 2005 and 30 June 2005 was approximately 38.7%.

All transactions in our database were categorized into four sale types-New Sale, Regular Resale, Real-Estate Owned (REO) Sale, and Foreclosure. We grouped New Sales and Regular Resales together as regular sales, where a New Sale refers to a newly built house and a Regular Resale refers to a previously owned house that is resold. A REO Sale is a sale that often occurs after a homeowner defaults on their mortgage and a real-estate company takes ownership of the house which is then sold. In order to more accurately estimate the foreclosure rate, we linked all REO Sales that occurred within two years of a foreclosure at the same address and then treated all of them as single undesirable events. All REO Sales not linked to foreclosures were also treated as undesirable events. This helps to reduce likely double counting of events where a REO Sale following a foreclosure at the same address pertains to the same mortgage default case. Also, this accounts for mortgage default cases for which a REO Sale record is present, but the foreclosure record is not. For the sake of monitoring foreclosure rates in Wayne County, we assumed that the first six months of 2005 represent an appropriate baseline process. We also restricted our analysis to only single family home transactions, which comprise the majority of the market at approximately 92% of all transactions.

In addition to using the Phase I data to construct a model for risk-adjustment, it is important to explore this subset of the data in order to better understand any existing features in the data. Within any given week, the number of closings appear to follow a pattern such that foreclosures and regular sales occur more frequently on certain days of the week. Figure 3.2 indicates that foreclosures (a) predominantly occur on Wednesdays, while regular sales (b) occur most frequently on Fridays. Thus, the rate of foreclosure as a percentage of all transactions would be higher on Wednesdays and lower on Fridays.

Due to the cyclic weekly pattern of the housing transactions, we chose to aggregate all records occurring in the same week. Thus, we treated all house closings that occurred in the same
Figure 3.2: Boxplots for the number of (a) foreclosures and (b) regular sales by day of the week for Wayne County, Michigan from 1 January 2005 through 30 June 2005, our Phase I sample. Panel (c) shows the number of transactions per week for Phase I and II prior to when the chart signals (see Results). This illustrates that aggregation at the week level demonstrates no significant shift in the distribution of the number of concurrent observations between Phase I and II.

week as concurrent observations. Although some information may be lost when aggregating, this removes some of the dependence over time. For a discussion of aggregation and its effects on process monitoring, see Schuh et al. (2013). Furthermore, Figure 3.2(c) shows that after aggregating by week, the distribution of the number of concurrent observations prior to when the chart signals remains the same in Phase I and II. This removes the concern that a shift in the volume of weekly transactions is affecting the chart performance. This ensures that the distribution of the number of concurrent observations is consistent and will not affect the performance of the monitoring scheme, as was discussed in Section 3.

We implemented our spatially risk-adjusted Bernoulli CUSUM control chart for concurrent observations to monitor the rate of foreclosure using these records. Our method would be especially useful for policymakers who could then track foreclosure rates. With the ability to appropriately monitor foreclosure rates, these individuals would be able to have an early warning that an issue in the housing market is developing.

### 3.4.2 Estimating risk of foreclosure

We use the proposed kernel density estimation and mixture model approach using the Phase
Figure 3.3: Kernel density estimates for (a) observed foreclosures, (b) regular sales, and (c) estimated foreclosure probability

I sample described in Section 3.2.4 to spatially estimate the probability of a housing transaction being a foreclosure. Figure 3.3 shows the kernel density estimates for (a) observed foreclosures, (b) regular sales, and (c) foreclosure probability. For a comprehensive review of mixture models, see McLachlan and Peel (2004). Details on the selection of the bandwidth for the kernel density estimation approach for this case study can be found in the appendix.

Areas in Figure 3.3 taking on values toward the high end of the spectrum indicate areas in which there are more foreclosures in (a) and areas in which there are more regular sales in (b). Figure 3.3 (c) provides the estimated spatial risk of foreclosure across Wayne County.
Some parts of southwestern Wayne County have high risk of foreclosure due to the lack of data in that region as seen in Figure 3.1. These high risk areas in regions where little data are available are a result of extrapolation of the mixture model.

Using this method provides a predicted probability of foreclosure for every ordered pair of latitude and longitude in Wayne County and thus a predicted risk of foreclosure for all housing transactions for July 2005 through June 2014. Using these probabilities, we can, as an illustration, retrospectively monitor the trend of foreclosure rates in Wayne County using our proposed spatially risk-adjusted Bernoulli CUSUM control chart for concurrent observations. For this analysis, housing transactions that occurred in the same week were considered concurrent observations. We choose to detect a shift in the odds ratio from $R_0 = 1$ to $R_1 = 2$, where this choice corresponds to a doubling of the odds of foreclosure.

### 3.4.3 Results

Figure 3.4 provides the risk-adjusted Bernoulli CUSUM chart for concurrent observations for the entire timespan of the data (a) and zoomed in for transactions that occurred up to the beginning of 2008 (b). A CUSUM statistic is plotted for each week instead of each plotted statistic corresponding to a specific transaction. Figure 3.4 also provides both the risk-adjusted and non-risk-adjusted control charts. The non-risk-adjusted control chart is based on the assumption that the predicted probability of foreclosure for every housing transaction is the baseline rate of 0.387 calculated based on the Phase I data. Likely foreclosures that occur shift the risk-adjusted chart statistic upwards less than the non-risk-adjusted chart statistic and unlikely regular sales that occur shift the risk-adjusted chart statistic down more than the non-risk-adjusted chart statistic. Similarly, unlikely foreclosures that occur shift the risk-adjusted chart statistic upwards more than the non-risk-adjusted chart statistic and likely regular sales that occur shift the risk-adjusted chart statistic down less than the non-risk-adjusted chart statistic. For example, as seen in Figure 3.1, foreclosures do exhibit strong spatial dependence. If an additional house forecloses in a neighborhood already
Figure 3.4: Risk-adjusted and non-risk-adjusted Bernoulli CUSUM control charts for monitoring foreclosure rates in Wayne County for (a) 1 July 2005 to 10 June 2014 and zoomed in for (b) 1 July 2005 to 1 January 2008.

Ravaged by foreclosure, this is less alarming than if a foreclosure occurs in a neighborhood with fewer foreclosures. Based on the resulting spatially risk-adjusted Bernoulli CUSUM chart for concurrent observations, it is clear that the foreclosure rate in Wayne County is consistently stable for the second half of 2005. Shortly after 2005 ended, the foreclosure rate begins to increase, with a steep increase of the CUSUM statistic toward the end of 2006 leading into 2007. Immergluck (2008) suggested that while many believe the impact of the Great Recession on the housing market notably began in 2007, the consequences were actually beginning to reveal themselves earlier. Our proposed spatially risk-adjusted Bernoulli CUSUM method appropriately detects the foreclosure rate deviating from the assumed stable market conditions. Furthermore, a closer look at the first year of monitoring in Figure 3.5 shows a brief spike in foreclosure rates in November 2005 followed by a short period of stability before the Great Recession began to take effect on the housing market.

Typically, a control limit $h^+$ is chosen for the CUSUM control chart in order to determine when the process is no longer stable. It is standard to choose the control limit based on the
Figure 3.5: Risk-adjusted and non-risk-adjusted Bernoulli CUSUM control charts for monitoring foreclosure rates in Wayne County for 1 July 2005 to July 2006

desired in-control ARL. Control limits farther from zero will cause the chart to take a longer time to signal a shift in the process, but result in a lower false alarm rate. It is usually desirable to have an in-control ARL that is considered a long time in the application to which the chart is being applied. In this application, due to the initial spike of the CUSUM statistic in November 2005, any control limit with a reasonable corresponding in-control ARL signals that the process has shifted at this time. Table 3.2 provides various control limits and their corresponding in-control ARLs (in weeks) estimated using resampling of the Phase I data. For this case-study, we randomly selected one Sunday, Monday, Tuesday, etc. with replacement from the Phase I sample and aggregated the observations on those days to create a week’s worth of concurrent observations. Weekly concurrent observations were sampled until the chart signaled in order to obtain in-control ARLs for each control limit seen in Table 3.2. In practice, similar Phase I resampling methods could be used to obtain the desired in-control ARL and thus, an appropriate control limit.

A control limit chosen to be $h^+ = 10$, results in a very large in-control ARL of 2713 weeks, or
Table 3.2: Possible control limits with corresponding in-control average run lengths (ARLs)

<table>
<thead>
<tr>
<th>$h^+$</th>
<th>ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>722.22</td>
</tr>
<tr>
<td>2</td>
<td>811.37</td>
</tr>
<tr>
<td>3</td>
<td>925.97</td>
</tr>
<tr>
<td>4</td>
<td>1039.37</td>
</tr>
<tr>
<td>5</td>
<td>1189.22</td>
</tr>
<tr>
<td>6</td>
<td>1381.43</td>
</tr>
<tr>
<td>7</td>
<td>1591.61</td>
</tr>
<tr>
<td>8</td>
<td>1872.91</td>
</tr>
<tr>
<td>9</td>
<td>2227.36</td>
</tr>
<tr>
<td>10</td>
<td>2712.74</td>
</tr>
</tbody>
</table>

approximately 52 years. Because the nature of this data is such that any reasonable control limit will result in a signal immediately at the first spike in November 2005, we present our figures without control limits. In addition to the control chart, Figure 3.6 provides the weekly rate of foreclosures as a proportion of all transactions for the entire period of the study. The solid vertical line indicates where Phase I ends and monitoring begins. The solid horizontal line corresponds to the baseline rate of foreclosures calculated from the Phase I sample. If a control limit $h^+ = 10$ was chosen, signals that the process is no longer stable would have occurred at the vertical dashed lines in Figure 3.6. The time of these signals is consistent with our understanding of the housing bubble and its relation to the Great Recession as described by Bezemer (2010). Using the proposed monitoring method for foreclosure rates could provide an early warning to those studying the housing market.

### 3.5 Discussion

The impact that the Great Recession had on the housing market and specifically on foreclosure rates was extremely unfortunate for many homeowners during that time. It would be useful for real estate planners and policymakers to be able to monitor foreclosure rates in order to understand and react to changes in the market. We propose a monitoring scheme
Figure 3.6: Weekly proportion of foreclosures in Wayne County, Michigan for 1 January 2005 through 10 June 2014

which implements a spatially risk-adjusted Bernoulli CUSUM chart for concurrent observations to monitor the foreclosure rate. We demonstrated that our method is capable of detecting early out-of-control behavior of a process, specifically the foreclosure rate within the housing market. Without adjusting for the spatial risk of foreclosure, monitoring of foreclosure rates could potentially be misleading and result in a signal of a process shift that did not actually occur.

While our method provides a framework for monitoring binary data with concurrent observations, we assume that the distribution of the number of concurrent observations at each time point is the same for both Phase I and II of the monitoring process. Our simulation study illustrates that variation in the number of concurrent observations is not an issue, unless there is a systematic shift in the distribution of the number of concurrent observations. If there is concern that this type of systematic shift is possible, a separate control chart could be used to monitor the number of concurrent observations. A potential technique that could appropriately adapt the control limit to the different numbers of concurrent observations at each time point is an implementation of dynamic probability control limits for the
risk-adjusted Bernoulli CUSUM chart proposed by Zhang and Woodall (2015).

Additionally, the performance of a control chart to detect a process change is reliant on the Phase I sample. Similar to Tian et al. (2015) and Tang et al. (2015), we assume that the Phase I sample used is representative of the process being monitored in Phase II. If there is a systematic shift in the distribution of the number of concurrent observations, chart performance is not guaranteed. For our application, we chose housing transactions occurring in the first six months of 2005 as our Phase I data. While only using half of a year might not be representative of a true baseline sample that includes the seasonality of the housing market, the lack of further historical data prior to 2005 prohibited us from using a longer time period as our Phase I sample.

The foreclosure case study has a few limitations. First, the transaction dates in the data can vary depending on the jurisdiction from the precise date that the homeowner signed the closing documents. We work with the data understanding that there could be minor fluctuations in the transaction dates. Second, the spatio-temporal data used dates back only as far as 2005. A longer historical record might enable a more formal treatment of an appropriate model to estimate the risk of foreclosure during the Phase I period. It is possible that the true underlying foreclosure rate, as a percentage of all housing transactions, during a stable baseline market is something other than a constant in time. Third, during the Phase I sample, an overall decrease in the rate of foreclosure is apparent (Figure 3.6) rather than a constant rate. This contributes to a CUSUM score of zero for the first four months monitoring (Figure 3.5). More historical data from previous years may enable the use of a more representative Phase I sample and risk-adjustment model to assess risk during this period. It should be noted that since the overall rate of foreclosure is decreasing during our Phase I sample, the chart would signal sooner, providing a conservative early warning. Despite these minor issues, the methodology for spatially risk-adjusted monitoring of concurrent binary observations is a useful approach to accommodate events that may occur simultaneously.

The proposed method is recommended for the monitoring of any binary outcome where
concurrent observations are drawn from a heterogeneous population with each observation having a different risk of the event of interest. The spatial risk-adjustment is particularly useful when the exact location of the observations is known and one would expect there to be inherent spatial dependence. The monitoring of housing market processes is an application that could provide insight into the market and help policymakers identify changes in the market.

**Appendix**

**Bandwidth selection for kernel density estimation**

An ideal method for choosing the bandwidth parameter $b$ for the kernel density estimation as shown by Kelsall and Diggle (1998) is a likelihood-based cross validation method given by

$$CV(b) = \left[ \prod_{i=1}^{n} \hat{q}_b^{-i}(l_{1i}, l_{2i})^{y_i} \left(1 - \hat{q}_b^{-i}(l_{1i}, l_{2i})\right)^{1-y_i} \right]^{-1/n},$$

(3.16)

where $n$ is the number of observations, $\hat{q}_b^{-i}(l_{1i}, l_{2i})$ is the leave-one-out estimated foreclosure probability at the $i^{th}$ location $(l_{1i}, l_{2i})$. By minimizing Equation (3.16) with respect to the bandwidth $b$, an optimal bandwidth can be selected. Due to the large number of observations in the Phase I data, we used a modified criterion that randomly divides the data into $k$ subsets and uses a $k$-fold cross validation metric given by

$$CV_2(b) = \left[ \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \hat{q}_b^{-i}(l_{1j}, l_{2j})^{y_j} \left(1 - \hat{q}_b^{-i}(l_{1j}, l_{2j})\right)^{1-y_j} \right\} \right]^{-1/k},$$

(3.17)

where $n_i$ is the number of observations in the $i^{th}$ subset and $\hat{q}_b^{-i}(l_{1j}, l_{2j})$ is the estimated foreclosure probability for the $j^{th}$ observation in the $i^{th}$ subset based on the kernel density estimation with the $i^{th}$ subset left out. By choosing $k = 10$ and minimizing the quantity given in Equation (3.17) with respect to $b$, we obtained a bandwidth of $b = 0.00007$. Figure
Figure 3.7: Likelihood-based cross validation function for various bandwidths

3.7 illustrates how $b$ was chosen given the value of $CV_2$ for various values of $b$. 
Bibliography


Chapter 4

On the Formal Specification of
Sum-zero Constrained Intrinsic
Conditional Autoregressive Models

Matthew J. Keefe ¹, Christopher T. Franck ¹,
Marco A.R. Ferreira ¹
Manuscript in Preparation

¹Department of Statistics, Virginia Tech, Blacksburg, VA 24061
Abstract

We propose a formal specification of sum-zero constrained intrinsic conditional autoregressive models. Practitioners often use these models as priors for Gaussian spatial random effects in Bayesian hierarchical models. Typically, an intrinsic conditional autoregressive model is defined through proper conditional distributions for each spatial random effect that imply an improper joint distribution. To ensure propriety of the distribution, in practice a sum-zero constraint is informally imposed on the vector of spatial random effects within a Markov chain Monte Carlo algorithm in what is known as centering-on-the-fly. While ingenious, this mathematically informal way to impose the sum-zero constraint obscures the actual joint density of the spatial random effects. Here, we formally define a sum-zero constrained intrinsic conditional autoregressive model as the limit of a sum-zero constrained proper conditional autoregressive model. Finally, we show that for a broad class of sequences of proper conditional autoregressive models, the resulting sum-zero constrained intrinsic conditional autoregressive model is unique.

Keywords: Areal data, Conditional autoregressive, Intrinsic autoregressive, Spatial statistics
4.1 Introduction

Intrinsic conditional autoregressive models are often used as prior distributions for spatial random effects in Bayesian hierarchical models. These autoregressive models are useful in applications where a neighbor-based notion of proximity is considered, such as disease mapping (Clayton and Kaldor, 1987; Bell and Broemeling, 2000; Moraga and Lawson, 2012; Goicoa et al., 2016), image restoration (Besag et al., 1991), complex survey data (Mercer et al., 2015), and neuroimaging (Liu et al., 2016). Because the intrinsic conditional autoregressive model is not a proper distribution, to make it proper in practice, a sum-zero constraint is imposed by a technique called centering-on-the-fly in which the vector of sampled spatial random effects is re-centered around its mean after each iteration of the Markov chain Monte Carlo algorithm being used. Even though it works well in practice, centering-on-the-fly obscures the actual joint density of the spatial random effects. As a consequence, the informal imposition of the sum-zero constraint prevents the development of formal methodology, such as objective Bayes methods for intrinsic conditional autoregressive-based hierarchical models.

The main contribution of this work is a procedure to formally define sum-zero constrained intrinsic conditional autoregressive models by first imposing the sum-zero constraint on a proper conditional autoregressive model and then by taking the limit to obtain an intrinsic conditional autoregressive model.

An important issue is that of uniqueness of intrinsic conditional autoregressive models when defined as limits of proper conditional autoregressive models. In particular, Lavine and Hodges (2012) have demonstrated that several approaches to obtain an intrinsic conditional autoregressive model as the limit of a proper conditional autoregressive model lead to differing limiting likelihood ratios, and thus arbitrary inference. However, Lavine and Hodges (2012) do not impose the sum-zero constraint. In contrast, we show that by first imposing the sum-zero constraint and then taking the limit for a broad class of sequences of proper conditional autoregressive models, the resulting sum-zero constrained intrinsic conditional autoregressive model is unique.
4.2 Intrinsic Conditional Autoregressive Model

The most commonly used conditional autoregressive model can be represented as a Gaussian Markov random field. Consider a geographical region of interest that is partitioned into \( n \) disjoint subregions. Additionally, suppose that a neighborhood structure is imposed on the region of interest such that \( \{N_j; j = 1, \ldots, n\} \) denotes the set of subregions that are neighbors of subregion \( j \). Now, consider \( \omega = (\omega_1, \ldots, \omega_n)^T \) to be a random vector corresponding to subregions \( 1, \ldots, n \) that is to be modeled by the conditional autoregressive model. Under these conditions, the joint density of \( \omega \) is

\[
p(\omega) \propto \exp\left\{ -\frac{\tau}{2} \omega^T H \omega \right\},
\]

(4.1)

where \( \tau > 0 \) is a precision parameter. If \( H \) were positive definite, then the joint distribution of \( \omega \) would be a proper Gaussian Markov random field. However, as is frequently the case for areal data, \( H \) is not positive definite. This leads to an improper joint distribution for \( \omega \). We define \( H \) as a known matrix determined by the neighborhood structure of the region of interest, such that

\[
(H)_{ij} = \begin{cases} 
  h_i, & \text{if } i = j \\
  -g_{ij}, & \text{if } i \in N_j \\
  0, & \text{otherwise},
\end{cases}
\]

(4.2)

where \( g_{ij} \geq 0 \) is a measure of how similar subregions \( i \) and \( j \) are, \( g_{ij} = g_{ji} \), and \( h_i = \sum_{j=1}^{n} g_{ij} \). Thus, \( H \) is symmetric and positive semi-definite. We assume that all of the subregions that completely partition the region of interest are connected (i.e. any two subregions may be connected by a path). As a consequence, \( H \) has only one eigenvalue equal to 0 with corresponding eigenvector \( n^{-1/2}1_n \), where \( 1_n \) is the \( n \times 1 \) vector of ones.
4.3 Obtaining the Sum-zero Constrained Intrinsic Conditional Autoregressive Model

In order to obtain the sum-zero constrained intrinsic conditional autoregressive model, we consider a sequence of three steps. In Step 1, we consider a proper conditional autoregressive model with propriety parameter $\lambda$ for the random vector $\phi^{**}$. In Step 2, we formally impose the sum-zero constraint on $\phi^{**}$ to obtain a sum-zero constrained proper conditional autoregressive model for the random vector $\phi^*$. Finally, in Step 3, we consider the limit as $\lambda \to 0$ that leads to a sum-zero constrained intrinsic conditional autoregressive model for the random vector $\phi$.

Step 1: Begin with a Proper Conditional Autoregressive Model

Consider the class of proper conditional autoregressive models with density

$$p(\phi^{**}) \propto \exp \left\{ -\frac{\tau}{2} \phi^{**T} \Sigma^{-1}_{\lambda} \phi^{**} \right\},$$

where $\Sigma^{-1}_{\lambda} = \lambda K + H$ such that $K$ is a diagonal matrix $K = \text{diag}(k_1, \ldots, k_n)$ with $k_i \geq 0$ for all $i$ and $k_i > 0$ for at least one $i$ for $i = 1, \ldots, n$. As a result of the following lemma, $\Sigma^{-1}_{\lambda}$ is invertible for $\lambda > 0$.

**Lemma 4.1.** Consider the positive semi-definite matrix $H$, a scalar $\lambda > 0$, and a diagonal matrix $K = \text{diag}(k_1, \ldots, k_n)$ with $k_i \geq 0$ for all $i$ and $k_i > 0$ for at least one $i$ for $i = 1, \ldots, n$. Then $\lambda K + H$ is positive definite.

**Proof of Lemma 4.1.** Since $H$ is positive semi-definite, $z^T H z \geq 0$ for any nonzero real vector $z$. Furthermore,

$$z^T (\lambda K + H) z = z^T H z + z^T (\lambda K) z.$$
Consider the two cases (1) \( z = a1_n \) and (2) \( z \neq a1_n \) for \( a \in \mathbb{R} \), where \( 1_n \) is the \( n \times 1 \) vector of ones. In case (1), because \( n^{-1/2}1_n \) is the eigenvector of \( H \) corresponding to its null eigenvalue, \( z^T H z = 0 \) and \( z^T (\lambda K) z > 0 \). In case (2), \( z^T H z > 0 \) and \( z^T (\lambda K) z \geq 0 \). Thus, \( z^T (\lambda K + H) z > 0 \). Therefore, \( \lambda K + H \) is positive definite.

**Step 2: Formally Impose the Sum-zero Constraint**

Consider the centering matrix \( P = (I_n - n^{-1}1_n1_n^T) \) where \( 1_n \) is the \( n \times 1 \) vector of ones and \( I_n \) is the \( n \times n \) identity matrix. To impose the sum-zero constraint, let \( \phi^* = P\phi^{**} \) which leads to

\[
\phi^* \sim N\left(0, \tau^{-1}\Sigma_{\phi\lambda}\right), \tag{4.4}
\]

where \( \Sigma_{\phi\lambda} = P\Sigma_{\lambda}P^T \). By pre-multiplying \( \phi^{**} \) by the centering matrix \( P \), we have constrained the sum of the elements of the vector \( \phi^* \) to be zero.

**Step 3: Take the Limit as \( \lambda \to 0 \)**

Finally, taking the limit as \( \lambda \to 0 \) leads to the sum-zero constrained intrinsic conditional autoregressive model given by the singular Gaussian distribution

\[
\phi \sim N\left(0, \tau^{-1}\Sigma_{\phi}\right), \tag{4.5}
\]

where \( \Sigma_{\phi} = \lim_{\lambda \to 0} \Sigma_{\phi\lambda} \).

In order to obtain this limit, it is particularly useful to express the matrix \( H \) using its spectral decomposition given by \( H = QDQ^T \), where \( Q = (q_1, \ldots, q_n) \) is a matrix comprised of columns which are the normalized eigenvectors of \( H \) and \( D = \text{diag}(d_1, \ldots, d_n) \) where \( d_1 \geq \cdots \geq d_{n-1} > d_n = 0 \) are the ordered eigenvalues of \( H \). Finally, we note that \( d_n = 0 \) has a corresponding eigenvector \( q_n = n^{-1/2}1_n \) (Ferreira and De Oliveira, 2007; De Oliveira and Ferreira, 2011).
Using the spectral decomposition of $H$, we can write $\Sigma_{\phi\lambda} = PQE^{-1}Q^TP^T$, where $E = \lambda Q^T K Q + D$. Additionally, we consider the following lemma:

**Lemma 4.2.** Consider the centering matrix $P$ and the matrix $Q$ which has columns that are the normalized eigenvectors of $H$. Then $PQ = (q_1, \ldots, q_{n-1}, 0)$.

**Proof of Lemma 4.2.**

\[
PQ = \begin{pmatrix} I - \frac{1}{n} 1_n 1_n^T \end{pmatrix}(q_1, \ldots, q_n) = (q_1, \ldots, q_n) - \frac{1}{n} 1_n (0, \ldots, 0, \sqrt{n}) = (q_1, \ldots, q_{n-1}, 0)
\]

\[\square\]

**Theorem 4.1.** Consider the proper conditional autoregressive model of Equation (4.3). For all $K$ such that $K$ is a diagonal matrix $K = \text{diag}(k_1, \ldots, k_n)$ with $k_i \geq 0$ for all $i$ and $k_i > 0$ for at least one $i$ for $i = 1, \ldots, n$, as long as the sum-zero constraint is mathematically imposed prior to considering the limit as $\lambda \to 0$, the resulting intrinsic conditional autoregressive model is a singular Gaussian distribution with covariance matrix $\tau^{-1} \Sigma_{\phi} = \tau^{-1} H^+$, where $H^+$ is the Moore-Penrose pseudoinverse of $H$ and therefore does not depend on $K$.

**Proof of Theorem 4.1.** Let $F = Q^T K Q$ and $E = \lambda F + D$. Denote a partition of a $n \times n$ symmetric matrix $A$ by

\[
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix},
\]

where $A_{11} = A_{1:(n-1),1:(n-1)}$ is a $(n - 1) \times (n - 1)$ matrix, $A_{12} = A_{1:(n-1),n}$ is a $(n - 1) \times 1$ vector, and $A_{22} = A_{n,n}$ is a scalar. Partitioning $E$ such that

\[
E = \begin{bmatrix} \lambda F_{11} + D_{11} & \lambda F_{12} \\ \lambda F_{12}^T & \lambda F_{22} + D_{22} \end{bmatrix},
\]
and using a standard matrix algebra results for partitioned matrices (Rencher and Schaalje, 2008, p. 23), we obtain

\[ E^{-1} = L = \begin{bmatrix} L_{11} & L_{12} \\ L_{12}^T & L_{22} \end{bmatrix}, \]

where the elements of \( L \) are given by

\[ L_{22} = (E_{22} - E_{12}^T E_{11}^{-1} E_{12})^{-1}, \]
\[ L_{12} = -L_{22} E_{11}^{-1} E_{12}, \]
\[ L_{11} = E_{11}^{-1} + L_{22} E_{11}^{-1} E_{12} E_{12}^T E_{11}^{-1}. \]

Thus, using Lemma 4.2,

\[ \Sigma_{\phi \lambda} = PQE^{-1}Q^TP^T \]
\[ = (q_1, \ldots, q_{n-1}, 0)E^{-1}(q_1, \ldots, q_{n-1}, 0)^T \]
\[ = TL_{11}T^T \]

where \( T = (q_1, \ldots, q_{n-1}) \) is a \( n \times (n - 1) \) matrix.

We can consider the limit as \( \lambda \to 0 \) to obtain the sum-zero constrained intrinsic conditional autoregressive model.

\[
\lim_{\lambda \to 0} \Sigma_{\phi \lambda} = \lim_{\lambda \to 0} TL_{11}T^T = T \left( \lim_{\lambda \to 0} L_{11} \right) T^T
\]
\[ = \lim_{\lambda \to 0} T \left\{ (\lambda F_{11} + D_{11})^{-1} + (\lambda F_{22} - \lambda F_{12}^T (\lambda F_{11} + D_{11})^{-1} \lambda F_{12})^{-1} \right. \]
\[ \times \left. [(\lambda F_{11} + D_{11})^{-1} \lambda F_{12} \lambda F_{12}^T (\lambda F_{11} + D_{11})^{-1}] \right\} T^T \]
\[ = TD_{11}^{-1}T^T = (q_1, \ldots, q_{n-1}) \text{diag} (d_1^{-1}, \ldots, d_{n-1}^{-1}) (q_1, \ldots, q_{n-1})^T \]
\[ = H^+. \]
Therefore, regardless of the matrix $K$, the resulting distribution for the sum-zero constrained intrinsic conditional autoregressive model obtained in Equation (4.5) is a singular Gaussian distribution given by

$$\phi \sim N \left(0, \tau^{-1} H^+ \right),$$

where $H^+$ is the Moore-Penrose pseudoinverse of $H$ (Penrose, 1955).

4.4 Examples

For comparison, we will consider the same examples proposed in Sections 3.3.2 and 3.3.3 of Lavine and Hodges (2012). Lavine and Hodges (2012) find that in their setup, each of these examples lead to distinct limits. In both examples, Lavine and Hodges (2012) consider models for two subregions and let $s$ be a vector modeled by an intrinsic conditional autoregressive model given by $p(s \mid \tau) \propto \exp(-\frac{1}{2} s^T W s)$, where

$$W = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$ 

In their Section 3.3.2, Lavine and Hodges (2012) make the distribution of $s$ proper by adding a positive quantity $\delta$ to the diagonal of the precision matrix such that $(s \mid \tau) \sim N(0, \tau^{-1} W_\delta)$, where $W_\delta^{-1} = W + \delta I_2$, where $I_2$ is the $2 \times 2$ identity matrix. In this situation, we can apply our Theorem 4.1 with $K = I_2$, $H = W$, and $\lambda = \delta$ to obtain the resulting singular Gaussian distribution with covariance matrix $\tau^{-1} H^+$ that results from imposing the sum-zero constraint followed by taking the limit as $\lambda \to 0$.

In their Section 3.3.3, they make the distribution of $s$ proper by adding a different quantity to the diagonal of the precision matrix such that $(s \mid \tau) \sim N(0, \tau^{-1} W_\delta)$, where $W_\delta^{-1} = W + (\delta/\tau) I_2$. In this example, we can apply our Theorem 4.1 with $K = \tau^{-1} I_2$, $H = W$, and $\lambda = \delta$. The resulting limiting distribution obtained after imposing the sum-zero constraint
and taking the limit as $\lambda \to 0$ is the same singular Gaussian distribution obtained in the previous example.

Thus, while Lavine and Hodges (2012) find that these two examples lead to different limits, our approach that first imposes the sum-zero constraint and then takes the limit of the proper conditional autoregressive model leads to the same intrinsic conditional autoregressive model.

### 4.5 Discussion

We provide a formal method to obtain sum-zero constrained intrinsic conditional autoregressive models that leads to a unique singular Gaussian distribution with zero mean vector and singular covariance matrix equal to the Moore-Penrose pseudoinverse of $H$.

Our result has both practical and theoretical implications. On the practical side, our result allows simulation of sum-zero constrained spatial random effects directly from their joint multivariate full conditional distribution, speeding up convergence of the Markov chain Monte Carlo algorithm. On the theoretical side, our result allows for the foundation of theoretically justified methods for models that use intrinsic conditional autoregressive models, such as the derivation of objective Bayes priors for models with Gaussian spatial random effects that are assigned intrinsic conditional autoregressive priors (Keefe et al., 2017). Going forward, our result provides a framework for researchers to develop methods for other models using conditional autoregressive components. For example, future work could use our result to investigate objective priors for other models that incorporate intrinsic conditional autoregressive effects, such as spatial generalized linear mixed models and spatio-temporal models for areal data.
Bibliography


Chapter 5

Objective Bayesian Analysis for
Gaussian Hierarchical Models with
Intrinsic Conditional Autoregressive
Priors

Matthew J. Keefe ¹, Christopher T. Franck ¹,
Marco A.R. Ferreira ¹
Manuscript Under Revision for Bayesian Analysis

¹Department of Statistics, Virginia Tech, Blacksburg, VA 24061
Abstract

Bayesian hierarchical models are commonly used for modeling spatially correlated areal data. However, choosing appropriate prior distributions for the parameters in these models is necessary and sometimes challenging. In particular, an intrinsic conditional autoregressive (CAR) hierarchical component is often used to account for spatial association. Vague proper prior distributions have frequently been used for this type of model, but this requires the careful selection of suitable hyperparameters. In this paper, we derive several objective priors for the Gaussian hierarchical model with an intrinsic CAR component and discuss their properties. We show that the independence Jeffreys and Jeffreys-rule priors result in improper posterior distributions, while the reference prior results in a proper posterior distribution. We present results from a simulation study that compares frequentist properties of Bayesian procedures that use several competing priors, including the derived reference prior. We demonstrate that using the reference prior results in favorable coverage, interval length, and mean squared error. Finally, we illustrate our methodology with an application to 2012 housing foreclosure rates in the 88 counties of Ohio.

Keywords: Conditional autoregressive, hierarchical models, objective prior, reference prior, spatial statistics
5.1 Introduction

Bayesian hierarchical models with intrinsic conditional autoregressive (CAR) priors are used for many statistical models for spatially dependent data in applications such as disease mapping (Clayton and Kaldor; Bell and Broemeling, 2000; Moraga and Lawson, 2012; Goicoca et al., 2016), image restoration (Besag et al., 1991), complex survey data (Mercer et al., 2015), and neuroimaging (Liu et al., 2016). The use of CAR specifications for modeling areal data was first introduced by Besag (1974), followed by the intrinsic CAR model as a prior for a latent spatial process proposed by Besag et al. (1991). Bayesian methods have been the dominating paradigm for spatial models including a CAR component (Sun et al., 1999; Hodges et al., 2003; Reich et al., 2006; Banerjee et al., 2014). Often in practice, it is difficult to subjectively choose informative priors for the parameters of the CAR component that are meaningful based on relevant domain knowledge. As a result, practitioners frequently use vague naïve priors (Bernardinelli et al., 1995; Best et al., 1999; Bell and Broemeling, 2000; Lee, 2013) that in the case of spatial models may lead to poor performance, such as slow MCMC convergence (Natarajan and McCulloch, 1998), unacceptably wide credible intervals, or larger mean squared error for estimation of parameters. To solve these problems, we introduce a novel objective prior that eliminates the need to subjectively choose priors for the parameters of the CAR prior when no previous knowledge is available. Our objective prior can serve as an automatic prior for the popular Gaussian hierarchical model with an intrinsic CAR prior for the spatial random effects that would enable an automatic Bayesian analysis.

To address concerns with the use of subjective proper priors, research has been conducted to explore objective Bayesian analysis of various spatial models. In particular, Berger et al. (2001) have introduced objective priors for geostatistical models for spatially correlated data, which has been further developed by De Oliveira (2007) who has derived objective priors for a similar model that includes measurement error. The development of objective priors for CAR models for the observed data has been explored by Ferreira and De Oliveira (2007),...
De Oliveira (2012), and Ren and Sun (2013). While many use CAR models directly for observed data, sometimes it is preferred to use a CAR prior for spatial random effects to allow for a smoother spatial process. More recently, Ren and Sun (2014) have derived objective priors for autoregressive models incorporated into the model as a latent process. Ren and Sun (2014) have focused on the use of proper CAR priors for the latent process component. However, in practice spatial random effects are usually assigned intrinsic CAR priors (Best et al., 2005). In this work, we consider objective priors for hierarchical models that use intrinsic CAR priors for the spatial random effects.

In the spatial statistics literature, intrinsic autoregressions (or intrinsic CARs) carry the notion that they are improper “densities” that are somehow made proper by imposing a constraint. The most frequently used constraint is a sum-zero constraint that ensures that the sum of the spatial random effects is equal to zero. In this paper, the intrinsic CAR prior for the spatial random effects with the sum-zero constraint is actually a proper joint density that is singular.

We provide further details and explanations in the following sections. In Section 5.2, we introduce notation and describe the Gaussian hierarchical model with unstructured random effects and intrinsic CAR spatial random effects. We also describe how to express the constrained intrinsic CAR prior as the limit of a proper Gaussian Markov random field (GMRF). We derive explicit expressions for the reference prior, the independence Jeffreys prior, and the Jeffreys-rule prior for this model and discuss posterior propriety in Section 5.3. We describe the details of the Markov chain Monte Carlo (MCMC) algorithm we use to simulate from the posterior distribution in Section 5.4. In Section 5.5, we present results of a simulation study to assess the frequentist properties of the Bayesian analyses using several competing priors. To illustrate and compare our proposed method to other common approaches, we conduct an analysis using the proposed reference prior and two frequently used subjective priors to model 2012 foreclosure rates in the 88 counties of Ohio in Section 5.6. Finally, we provide our conclusions and discussion of future work in Section 5.7. For clarity of exposition, the proofs of the theoretical results are provided in the Appendix.
5.2 Model Specification

5.2.1 Model

Consider a contiguous geographical region of interest that is partitioned into $n$ disjoint subregions that collectively make up the entire region of interest. For example, a state could be divided into several counties. Additionally, suppose that a neighborhood structure is considered for the region of interest such that $\{N_j; j = 1, \ldots, n \}$ denotes the set of subregions that are neighbors of subregion $j$. Typically, subregions that share a boundary are considered neighbors. Within this framework, consider the following model:

$$Y = X\beta + \theta + \phi,$$  \hspace{1cm} (5.1)

where $Y$ is the $n \times 1$ vector containing the response variable, $X$ is a $n \times p$ matrix of covariates, and $\beta$ is the $p \times 1$ vector of fixed effect regression coefficients. Furthermore, we assume that $\theta = (\theta_1, \theta_2, \ldots, \theta_n)^T$ is a vector of unstructured random effects defined such that $\theta_i$ are independent and normally distributed with mean 0 and variance $\sigma^2$ for $i = 1, 2, \ldots, n$. Finally, $\phi = (\phi_1, \phi_2, \ldots, \phi_n)^T$ is a vector of spatial random effects that is assigned a sum-zero constrained intrinsic CAR prior (i.e. $\sum_{i=1}^{n} \phi_i = 0$) (Besag and Kooperberg, 1995). We describe how to formally induce this constraint in Section 5.2.2. Additionally, we assume that $\theta$ and $\phi$ are independent a priori.

Typically, intrinsic CAR models are defined by their conditional distributions, which are proper distributions. Consider the frequently used intrinsic CAR model for $\omega = (\omega_1, \ldots, \omega_n)^T$ specified by its conditional distributions

$$p(\omega_i|\omega_{\sim i}) \propto \exp \left\{ -\frac{\tau_\omega}{2} \left[ \sum_{i=1}^{n} \omega_i^2 h_i - 2 \sum_{i<j} \omega_i \omega_j g_{ij} \right] \right\}, \hspace{1cm} (5.2)$$

where $\omega_{\sim i}$ is the vector of the CAR elements for all subregions except subregion $i$ and $\tau_\omega > 0$. 

is a precision parameter. In addition, \( g_{ij} \geq 0 \) is a measure of how similar subregions \( i \) and \( j \) are, \( g_{ij} = g_{ji} \), and \( h_i = \sum_{j=1}^{n} g_{ij} \). Alternatively, we may write the joint density for \( \omega \) as

\[
 p(\omega) \propto \exp \left\{ -\frac{\tau_\omega}{2} \omega^T H \omega \right\},
\]

where \( H \) is a symmetric, positive semi-definite precision matrix defined as

\[
 (H)_{ij} = \begin{cases} 
 h_i, & \text{if } i = j \\
 -g_{ij}, & \text{if } i \in N_j \\
 0 & \text{otherwise},
\end{cases}
\]

The matrix \( H \) is assumed to be fixed and known, as its structure is typically chosen as a modeling decision. One common choice for the similarity measure is \( g_{ij} = 1 \) if subregions \( i \) and \( j \) are neighbors, and \( g_{ij} = 0 \) if subregions \( i \) and \( j \) are not neighbors. This binary similarity measure implies that \( h_i \) is the number of neighbors of subregion \( i \). Note that Equation (5.3) is a multivariate normal kernel specified by its precision matrix \( \tau_\omega H \). Furthermore, note that the matrix \( H \) is singular with one eigenvalue equal to zero and corresponding eigenvector \( n^{-1/2}1_n \), where \( 1_n \) denotes a \( n \times 1 \) vector of ones. As a consequence, the density given in Equation (5.3) does not change if we substitute \( \omega \) by \( \omega + a1_n \), where \( a \) is any real constant. Because of that, Equation (5.3) is an improper “density”. To make this density proper, practitioners usually assume the sum-zero constraint \( \sum_{i=1}^{n} \omega_i = 0 \). This constraint is frequently imposed in an informal manner with the vector of spatial random effects being recentered around its own mean at the end of each MCMC iteration. While ingenious, this mathematically informal way to impose the sum-zero constraint obscures the actual joint prior density of \( \omega \) under the constraint. Consequently, a hierarchical model with spatial random effects defined by Equation (5.3) with the sum-zero constraint \( \sum_{i=1}^{n} \omega_i = 0 \) imposed within the MCMC algorithm does not yield itself to a formal objective Bayes analysis.

To enable a formal objective Bayes analysis, in Section 5.2.2 we consider an intrinsic CAR
expressed as the limit of a proper GMRF, rather than starting with an intrinsic CAR. We impose the sum-zero constraint before taking the limit. The resulting joint density for the spatial random effects after taking the limit is proper.

5.2.2 Intrinsic CAR as the limit of a proper GMRF

In the following development, we consider the following notation to clearly demonstrate the transition between different CAR models. First, we consider a proper CAR for the random vector $\phi^{**}$, then impose the sum-zero constraint to obtain the random vector $\phi^*$, and finally consider the limiting case that leads to the sum-zero constrained intrinsic CAR for the random vector $\phi$. The intrinsic CAR prior is a limiting case of a proper CAR prior. Specifically, we express the CAR prior using a signal-to-noise ratio parametrization as a proper GMRF with parameter $\lambda$

$$\phi^{**} \sim N\left(0, \frac{\sigma^2}{\tau_c} \Sigma_\lambda \right), \quad (5.5)$$

where $\tau_c > 0$ is an unknown parameter and $\Sigma_\lambda^{-1} = \lambda I_n + H$ with $I_n$ being the $n \times n$ identity matrix and $H$ is defined as in Equation (5.4).

Note that rather than using a CAR model directly for the data, we consider a hierarchical model with both spatial random effects and unstructured random effects. Because of this, we use a signal-to-noise ratio parametrization of the CAR prior where the variance of the spatial random effects is expressed as a function of the variance of the unstructured random effects, as seen in expression (5.5). When $\lambda \to 0$, $\phi^{**}$ in Equation (5.5) approaches an intrinsic CAR (Besag et al., 1991; Besag and Kooperberg, 1995). In practice, the intrinsic CAR spatial random effects are constrained to sum to zero to ensure propriety of the prior. This is often performed within the posterior sampling algorithm by centering the sampled random effects after each iteration. Rather than taking this approach, we first consider the proper GMRF in (5.5), then we impose the sum-zero constraint, and finally we take the limit as $\lambda \to 0$ to
obtain a constrained intrinsic CAR.

To impose the sum-zero constraint for the CAR prior in Equation (5.5), we let \( \phi^* = P\phi^{**} \), where \( P = (I_n - n^{-1}1_n1_n^T) \) is a centering matrix. After imposing the constraint, the distribution of \( \phi^* \) is given by

\[
\phi^* \sim N \left( 0, \frac{\sigma^2}{\tau_c} \Sigma_{\phi\lambda} \right), \tag{5.6}
\]

where \( \Sigma_{\phi\lambda} = P\Sigma_{\lambda}P^T \).

Now, suppose that the spectral decomposition of \( H \) is given by \( H = QDQ^T \) where \( Q = (q_1, q_2, \ldots, q_n) \) is a \( n \times n \) matrix comprised of columns which are the normalized eigenvectors of \( H \) and \( D = \text{diag}(d_1, d_2, \ldots, d_n) \) where \( d_1 \geq d_2 \geq \cdots \geq d_n \) are the ordered eigenvalues of \( H \). Additionally, if all of the subregions that completely partition the region of interest are connected (i.e. there is a path connecting any two subregions), then \( H \) has rank \( n - 1 \). Consequently, 0 is an eigenvalue of \( H \) with multiplicity 1 (e.g. see De Oliveira and Ferreira (2011)). As a result, \( d_{n-1} > d_n = 0 \) and \( d_n \) has a corresponding eigenvector \( q_n = n^{-1/2}1_n \).

**Lemma 5.1.** If all of the subregions that completely partition the region of interest are connected, then \( \Sigma_{\phi\lambda} \) has rank \( n - 1 \) and the spectral decomposition of \( \Sigma_{\phi\lambda} \) is \( \Sigma_{\phi\lambda} = QM_\lambda Q^T \), where \( M_\lambda = \text{diag}((\lambda + d_1)^{-1}, \ldots, (\lambda + d_{n-1})^{-1}, 0) \). Finally, \( \Sigma_{\phi\lambda} \) can be written as

\[
\Sigma_{\phi\lambda} = \sum_{i=1}^{n-1} \frac{1}{\lambda + d_i} q_iq_i^T. \tag{5.7}
\]

**Proof.** See Appendix

Now we take the limit as \( \lambda \to 0 \) to obtain the final density for \( \phi \) which appears in our model from Equation (5.1) given by

\[
\phi \sim N \left( 0, \frac{\sigma^2}{\tau_c} \Sigma_{\phi\lambda} \right), \tag{5.8}
\]
where $\Sigma_\phi = \lim_{\lambda \to 0} \Sigma_{\phi \lambda} = Q M Q^T$ and $M = \text{diag} \left( d_1^{-1}, \ldots, d_n^{-1}, 0 \right)$. Hence, $\Sigma_\phi = H^+$ is the Moore-Penrose pseudoinverse of $H$ (Penrose, 1955). In a related paper, Keefe et al. (2017a) have generalized this result for a broad class of proper CAR models that all result in the same sum-zero constrained intrinsic CAR model. The resulting constrained intrinsic CAR prior for $\phi$ in Equation (5.8) is a singular Gaussian distribution, which is amenable for an objective Bayes analysis. Thus, it follows that the model for the response $Y$ with the spatial random effects $\phi$ integrated out is

\begin{equation}
(Y | \beta, \sigma^2, \tau_c) \sim N \left( X \beta, \sigma^2 \left( I_n + \tau_c^{-1} \Sigma_\phi \right) \right).
\end{equation}

(5.9)

5.2.3 Remarks

Note that by using this parametrization, we assume a signal-to-noise ratio parametrization for the variance components of the random components of the model, as is done in De Oliveira (2007). One of the benefits of the signal-to-noise ratio parametrization is that it leads to simpler expressions for the objective priors. So, the unknown model parameters for this hierarchical spatial model for areal data are $\eta = (\beta, \sigma^2, \tau_c) \in \mathbb{R}^p \times (0, \infty) \times (0, \infty)$. Of particular interest is the parameter $\tau_c$, which controls the strength of spatial dependence. Conditional on the neighborhood structure chosen by the user, $\Sigma_\phi$ is fixed and the correlation between the response for two neighboring subregions $Y_i$ and $Y_j$ is given by

\begin{equation}
\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\tau_c^2 + \tau_c \sigma_{ii} + \tau_c \sigma_{jj} + \sigma_{ii} \sigma_{jj}}},
\end{equation}

(5.10)

where $\sigma_{ij}$ denotes the element of $\Sigma_\phi$ located in the $i^{th}$ row and $j^{th}$ column and $\sigma_{ii}$ and $\sigma_{jj}$ denote the $i^{th}$ and $j^{th}$ diagonal elements of $\Sigma_\phi$, respectively. Since $\rho_{ij}$ is a decreasing function of $\tau_c$, as $\tau_c \to 0$, $\rho_{ij}$ increases implying that the spatial dependence also increases. As is commonly the case in stochastic processes, an increase in correlation is often accompanied by an increase in marginal variance. Take for example the first-order autoregressive model
(AR(1)) with error variance $\delta^2$ and autoregressive coefficient $0 < \rho < 1$. The correlation function for lag $h$ is $\text{Corr}(x_t, x_{t-h}) = \rho^h$ and the marginal variance is $\delta^2/(1 - \rho^2)$. Therefore, for the widely used AR(1) model, increase in $\rho$ corresponds to increase in both correlation and marginal variance. Note that in our model, $\text{Var}(y_i) = \sigma^2(1 + \tau^{-1}_c \sigma_i)$. So, it is also the case for our model an increase in correlation is accompanied by an increase in marginal variance.

Specifying the model in terms of a proper GMRF, as we have done in (5.9), simplifies application of the results of De Oliveira (2007) to derive several objective priors. Although these priors take on similar forms as those presented by De Oliveira (2007), the properties of the derived priors differ in our case. The reason for the difference is that for the geostatistical models considered by De Oliveira (2007), the matrix $\Sigma_\phi$ is full rank, whereas we consider a hierarchical model for areal data where $\Sigma_\phi$ is singular.

## 5.3 Objective Priors

### 5.3.1 Likelihood Functions

In order to obtain objective priors, such as the reference, independence Jeffreys, and Jeffreys-rule priors for this model, we consider the likelihood and integrated likelihood functions. The likelihood of $\eta$ given the observed response $y$ and the matrix $X$ is given by

$$L(\eta; y, X) \propto (\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp \left[-\frac{1}{2\sigma^2} (y - X\beta)^T \Omega^{-1} (y - X\beta) \right],$$

(5.11)

where $\Omega = I_n + \frac{1}{\tau_c} \Sigma_\phi$ and $|A|$ denotes the determinant of the matrix $A$. All of the objective priors we derive with respect to the integrated likelihoods in the following sections conveniently fall into a class of priors of the form
\[
\pi(\eta) \propto \frac{\pi(\tau_c)}{(\sigma^2)^a}, \quad (5.12)
\]
where \(a \in \mathbb{R}\) is a hyperparameter and \(\pi(\tau_c)\) is referred to as the marginal prior for \(\tau_c\).

In order to obtain the objective priors, it is first necessary to obtain the integrated likelihoods. If we first integrate \(L(\eta; y, X)\) with respect to \(\beta\), we obtain the integrated likelihood of \((\sigma^2, \tau_c)\). This integration leads to the log-integrated likelihood given by

\[
\log L^I(\sigma^2, \tau_c; y, X) = -\frac{1}{2} \left[ (n - p) \log(\sigma^2) + \log(|\Omega|) + \log(|X^T \Sigma^{-1} X|) + \frac{S^2}{\sigma^2} \right], \quad (5.13)
\]
where \(S^2 = y^T \left[ \Omega^{-1} - \Omega^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Omega^{-1} \right] y\). Furthermore, integrating \((\sigma^2)^{-a} \times L^I(\sigma^2, \tau_c; y, X)\) with respect to \(\sigma^2\) yields the integrated likelihood of \(\tau_c\), which is given by

\[
L^I(\tau_c; y, X) \propto \left( |\Omega||X^T \Sigma^{-1} X| \right)^{-1/2} (S^2)^{-\left(\frac{n-p}{p}+a-1\right)}. \quad (5.14)
\]

Unfortunately, it is difficult to calculate the reference, independence Jeffreys, and Jeffreys-rule priors directly using expressions (5.13) and (5.14). To overcome this challenge, it is particularly useful to express (5.13) in a simplified form involving the eigenvalue decompositions of matrices that are functions of the matrices \(X\) and \(\Sigma_{\phi}\). Consider the following lemma (Verbyla, 1990; Dietrich, 1991; Kuo, 1999; De Oliveira, 2007):

**Lemma 5.2.** Suppose \(X_{n \times p}\) is of full rank, with \(n > p\), and \(\Sigma\) is an \(n \times n\) symmetric positive definite matrix. Then there exists a full rank \(n \times (n-p)\) matrix \(L\) with the following properties:

\(\text{(i)}\) \(L^T X = 0\)
\(\text{(ii)}\) \(L^T L = I_{n-p}\)
\(\text{(iii)}\) \(\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} = L (L^T \Sigma L)^{-1} L^T\)
(iv) \( \log(|\Sigma|) + \log(|X^T\Sigma^{-1}X|) = \log(|L^T\Sigma L|) + c \), where \( c \) depends on \( X \) but not on \( \Sigma \).

(v) \( L^T\Sigma L \) is an \((n-p) \times (n-p)\) diagonal matrix with positive diagonal elements.

Here we propose one way to obtain a matrix \( L \) with the above properties by the following steps:

1. Obtain the orthogonal complement of the matrix \( G = X(X^T X)^{-1}X^T \) given by \( G^* = I_n - X(X^T X)^{-1}X^T \).

2. Calculate the matrix \( Q^* \) such that the columns of \( Q^* \) are the normalized eigenvectors corresponding to the non-zero eigenvalues of \( G^* \).

3. Denote the spectral decomposition of \( Q^*^T \Sigma \phi Q^* \) as \( U \Psi U^T \).

4. Compute \( L = Q^*U \).

The resulting \( L \) matrix has properties that allow for convenient simplifications in the integrated likelihoods. Let \( \xi_1 \geq \xi_2 \geq \cdots \geq \xi_{n-p} > 0 \) be the ordered eigenvalues of \( Q^*^T \Sigma \phi Q^* \). Then, by the way we obtain the matrix \( L \), we have that \( L^T \Sigma L = \Psi = \text{diag}(\xi_1, \xi_2, \ldots, \xi_{n-p}) \).

Then the following results hold (De Oliveira, 2007)

\[
\log(|\Omega|) + \log(|X^T \Omega^{-1} X|) = \sum_{j=1}^{n-p} \log(1 + \frac{1}{\tau_c \xi_j}) + c
\]  

(5.15)

\[
S^2 \equiv \sigma_*^2 \sum_{j=1}^{n-p} \left( \frac{1 + \tau_{cs}^{-1} \xi_j}{1 + \tau_c^{-1} \xi_j} \right) Z_j^2,
\]  

(5.16)

where \( \sigma_*^2 \) and \( \tau_{cs} \) are the true values of \( \sigma^2 \) and \( \tau_c \), respectively, and \( \{Z_j^2\} \overset{iid}{\sim} \chi_1^2 \). Using these results, the log-integrated likelihood in (5.13) satisfies

\[
\log L'(\sigma^2, \tau_c; y, X) \equiv -\frac{1}{2} \left[ (n-p) \log(\sigma^2) + \sum_{j=1}^{n-p} \left\{ \log(1 + \frac{1}{\tau_c \xi_j}) + \frac{\sigma_*^2}{\sigma^2} \left( \frac{1 + \tau_{cs}^{-1} \xi_j}{1 + \tau_c^{-1} \xi_j} \right) Z_j^2 \right\} \right].
\]  

(5.17)
The posterior distribution of $\eta$ using a prior of the form given in (5.12) is proper if and only if

$$0 < \int_{0}^{\infty} L^I(\tau_c; y, X) \pi(\tau_c) d\tau_c < \infty. \quad (5.18)$$

**Proposition 5.1.** Consider the model given in (5.9) along with the prior given in (5.12). Then, $L^I(\tau_c; y, X)$ is a continuous function on $(0, \infty)$ and $L^I(\tau_c; y, X) = O(\tau_c^{-a})$ as $\tau_c \to 0$ and $L^I(\tau_c; y, X) = O(1)$ as $\tau_c \to \infty$.

**Proof.** See Appendix

The condition in (5.18) and the tail behavior of the integrated likelihood in Proposition 5.1 provide justification for determining whether or not the proposed priors lead to proper posterior distributions. For example, the tail behavior of the posterior distribution as $\tau_c \to 0$ must be $O(\tau_c^m)$ where $m > -1$ in order to guarantee posterior propriety. Likewise, the tail behavior of the posterior distribution as $\tau_c \to \infty$ must be $O(\tau_c^m)$ where $m < -1$ in order to guarantee posterior propriety.

### 5.3.2 Reference Prior

In order to obtain a reference prior for the model parameters $\eta$, it is necessary to identify an order of importance for the parameters. Here, we consider $\beta$ to be a nuisance parameter, while $(\sigma^2, \tau_c)$ is the parameter vector of interest. Then, we use exact marginalization to find the reference prior (Berger et al., 2001). First, the joint prior distribution of $\eta$ is factored as $\pi^R(\eta) = \pi^R(\beta|\sigma^2, \tau_c)\pi^R(\sigma^2, \tau_c)$. Then given $(\sigma^2, \tau_c)$, the conditional reference prior for $\beta$ is $\pi^R(\beta|\sigma^2, \tau_c) \propto 1$ (Bernardo and Smith, 1994). Finally, we use the Jeffreys-rule algorithm on the integrated likelihood $L^I(\sigma^2, \tau_c; y, X)$ to obtain $\pi^R(\sigma^2, \tau_c)$.

**Theorem 5.1.** Consider the model given in (5.9). Then, the reference prior of $\eta$ is of the
form (5.12) with

\[ a = 1 \text{ and } \pi^R(\tau_c) \propto \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( \frac{\xi_j}{\tau_c + \xi_j} \right)^2 - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( \frac{\xi_j}{\tau_c + \xi_j} \right) \right\}^2 \right]^{1/2}. \]  

(5.19)

Proof. See Appendix

The following proposition and corollary describe the properties of the proposed reference prior and its resulting posterior distribution.

**Proposition 5.2.** Suppose that \( \xi_1, \xi_2, \ldots, \xi_{n-p} \) are not all equal. Then, the marginal reference prior of \( \tau_c \) in (5.19) is a continuous function on \((0, \infty)\) where:

(i) \( \pi^R(\tau_c) = O(1) \) as \( \tau_c \to 0 \)

(ii) \( \pi^R(\tau_c) = O(\tau_c^{-2}) \) as \( \tau_c \to \infty \).

Proof. See Appendix

**Corollary 5.1.** Using the model given by (5.9), we have

(i) The reference prior \( \pi^R(\tau_c) \) given in (5.19) and its resulting posterior \( \pi^R(\eta|y, X) \) are both proper.

(ii) The \( k^{th} \) moment of the marginal reference posterior \( \pi^R(\tau_c|y, X) \) does not exist for \( k \geq 1 \).

Proof. See Appendix

Notice that the reference prior does depend on the covariates \( X \) through the eigenvalues \( \xi_1, \xi_2, \ldots, \xi_{n-p} \) of \( L^T \Sigma_{\phi} L \). We also note that since our proposed reference prior has the same mathematical expression as that of De Oliveira (2007), the results regarding the marginal reference prior, the joint posterior distribution, and the marginal posterior distribution are
all the same. However, the same is not true for the independence Jeffreys and Jeffreys-rule priors that follow. This is due to the fact that, in our case, $L^T \Sigma \phi L$ is still full rank, while $\Sigma \phi$ is not full rank.

### 5.3.3 Independence Jeffreys Prior

The independence Jeffreys prior is obtained by assuming that $\beta$ and $(\sigma^2, \tau_c)$ are independent a priori. Then, the Jeffreys rule is used to find the marginal prior of each parameter, assuming the other parameters are known. We denote the ordered eigenvalues of $\Sigma \phi$ as $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{n-1} > \gamma_n = 0$. It is important to note that in our case, we are dealing with areal data, and thus $\gamma_n = 0$ since $\Sigma \phi$ is of rank $n-1$. This is a key distinction between our work and that of De Oliveira (2007), which leads to different results regarding posterior propriety.

**Theorem 5.2.** Consider the model given in (5.9). Then, the independence Jeffreys prior of $\eta$ is of the form (5.12) with

$$a = 1 \text{ and } \pi^{J1}(\tau_c) \propto \frac{1}{\tau_c} \left[ \sum_{i=1}^{n} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right)^2 - \frac{1}{n} \left\{ \sum_{i=1}^{n} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right) \right\}^2 \right]^{1/2}.$$  

(5.20)

**Proof.** See Appendix

**Proposition 5.3.** The marginal independence Jeffreys prior of $\tau_c$ in (5.20) is a continuous function on $(0, \infty)$ where $\pi^{J1}(\tau_c) = O(\tau_c^{-1})$ as $\tau_c \to 0$.

**Proof.** See Appendix

**Corollary 5.2.** Using the model given by (5.9), the independence Jeffreys prior $\pi^{J1}(\tau_c)$ given in (5.20) and its resulting posterior $\pi^{J1}(\eta|y, X)$ are both improper.

**Proof.** See Appendix
5.3.4 Jeffreys-rule Prior

The Jeffreys-rule prior is obtained by \( \pi^{J2}(\eta) \propto |I(\eta)|^{1/2} \), where \( I(\eta) \) is the Fisher information matrix of \( \eta \).

**Theorem 5.3.** Consider the model given in (5.9). Then, the Jeffreys-rule prior of \( \eta \), is of the form (5.12) with

\[
a = 1 + \frac{p}{2} \quad \text{and} \quad \pi^{J2}(\tau_c) \propto \left( \frac{\prod_{j=1}^{n-p} (1 + \tau_c^{-1} \xi_j)}{\prod_{i=1}^{n} (1 + \tau_c^{-1} \gamma_i)} \right)^{1/2} \cdot \pi^{J1}(\tau_c).
\]

(5.21)

**Proof.** See Appendix \( \square \)

**Proposition 5.4.** The marginal Jeffreys-rule prior of \( \tau_c \) in (5.21) is a continuous function on \((0, \infty)\) where \( \pi^{J2}(\tau_c) = O(\tau_c^{-3/2}) \) as \( \tau_c \to 0 \).

**Proof.** See Appendix \( \square \)

**Corollary 5.3.** Using the model given by (5.9), the posterior distribution obtained by using the Jeffreys-rule prior \( \pi^{J2}(\tau_c) \) given in (5.21) is improper.

**Proof.** See Appendix \( \square \)

Similar to the reference prior, notice that the Jeffreys-rule prior does depend on the covariates \( X \) through the eigenvalues \( \xi_1, \xi_2, \ldots, \xi_{n-p} \) of \( L^T \Sigma_\phi L \).

5.3.5 Remarks

There are a few key distinctions between our results and those of other researchers. Specifically, our reference prior leads to a proper posterior distribution, while the independence Jeffreys and Jeffreys-rule priors lead to improper posterior distributions, unlike De Oliveira
(2007). Although we consider areal data while De Oliveira (2007) has considered geostatistical data, the reference prior in (5.19) has the same mathematical expression as that proposed by De Oliveira (2007) because we choose to formulate our model with the intrinsic CAR component as the limit of a proper GMRF. However, the matrix $\Sigma_\phi$ is not full rank in the areal data setting. Thus, the smallest eigenvalue of $\Sigma_\phi$ is $\gamma_n = 0$, which leads to improper posterior distributions when the independence Jeffreys and Jeffreys-rule priors are used.

Ren and Sun (2014) have defined their proper CAR prior using conditional distributions with a constant variance that does not depend on the number of neighbors a subregion has, which limits utility for many applications. In contrast, the conditional specification of the CAR prior we use has a variance that depends on the number of neighboring subregions. This allows the conditional distribution of $\phi_i$ given $\phi_{\sim i}$ for subregions with more neighbors to have a smaller variance. Furthermore, the objective priors that have been derived by Ren and Sun (2014) do not have explicit mathematical forms. Specifically, the priors of Ren and Sun (2014) are written as functions of determinants of Fisher information matrices that do not resemble our reference prior. In contrast, the expression of our reference prior can be easily implemented with its explicit mathematical form.

5.4 Sampling from the Posterior Distribution

In order to perform posterior inference about the model parameters, we propose a MCMC algorithm. Specifically, we have implemented a MCMC sampler (Gelfand and Smith, 1990) that considers the integrated likelihood given in Equation (5.9) obtained by integrating out the spatial random effects $\phi$. Thus, our MCMC algorithm generates draws from the posterior distribution of $(\beta, \sigma^2, \tau_c)$. Additionally, we simulate $\phi$ using composite sampling.

The use of the signal-to-noise ratio parametrization results in a Fisher information matrix of $\log(\sigma^2)$ and $\log(\tau_c)$ which does not depend on the value of $\sigma^2$. This implies a uniform prior for $\log(\sigma^2)$. Thus, in sampling from the posterior distribution, we jointly propose values for
log(σ²) and log(τc) with a Metropolis-Hastings step using Normal proposal distributions, and accept proposed values based on the appropriate Metropolis-Hastings acceptance probability (Gamerman and Lopes, 2006; Robert and Casella, 2004). The steps of our MCMC algorithm are provided in Algorithm 1. Posterior inference can then be conducted directly from the MCMC samples after discarding an appropriate number of iterations as burn-in. While we choose to use a MCMC sampler, there are numerous alternative methods that could be used to obtain samples from the posterior distribution. For example, De Oliveira (2007) has implemented adaptive rejection Metropolis sampling (Gilks et al., 1995).

Since the matrix Σφ can be conveniently expressed as a function of eigenvectors and eigenvalues of H, as shown in Equation (5.7), several computational equivalences can be used to speed up the MCMC algorithm. Specifically, execution of the MCMC algorithm relies on computation of the determinant and inverse of the matrix (I + τc⁻¹Σφ). In this work, these quantities can be expressed by the following equivalences

\[ (I + \tau_c^{-1} \Sigma_{\phi})^{-1} = Q \text{diag} \left( \frac{1}{\lambda + d_1}, \ldots, \frac{1}{\lambda + d_{n-1}}, 1 \right) Q^T \]  \hspace{1cm} (5.22)

\[ |\sigma^2 (I + \tau_c^{-1} \Sigma_{\phi})| = (\sigma^2)^n \prod_{i=1}^{n-1} \left( 1 + \frac{\tau_c^{-1}}{\lambda + d_i} \right) \]  \hspace{1cm} (5.24)

### 5.5 Simulation Study

#### 5.5.1 Comparison of Priors

We use Monte Carlo simulation to compare the performance of the proposed reference prior with two commonly used vague naïve prior distributions from the literature. Performance is assessed using frequentist properties of Bayesian procedures, including interval coverage...
Algorithm 1 MCMC Algorithm

1. Initialize $\eta^{(0)} = (\beta^{(0)}, \sigma^{2(0)}, \tau^{(0)}_c)$

For $i$ in 1 to $K$

{ 
2. Generate $\log(\sigma^{2*}) \sim N(\sigma^{2(i-1)}, \delta_s)$ and $\log(\tau^{*}_c) \sim N(\tau^{(i-1)}_c, \delta_t)$
3. Compute joint acceptance probability for $\sigma^{2*}$ and $\tau^{*}_c$:
   \[
   \alpha = \min \left[ 1, \frac{P(\eta^*|\text{Data})q(\eta^*)}{P(\eta^{(i)}|\text{Data})q(\eta^{(i)})} \right]
   \]
4. Generate $(\beta^*, y, \sigma^2, \tau_c, X) \sim N_p(\mu^*, \Sigma^*)$, where
   \[
   \mu^* = (X^T(I_n + \tau^{-1}_c(\Sigma_\phi)^{-1}X)^{-1}X^T(I_n + \tau^{-1}_c(\Sigma_\phi)^{-1}X)^{-1}X^T(I_n + \tau^{-1}_c(\Sigma_\phi)^{-1}X)^{-1}y,
   \]
   \[
   \Sigma^* = \sigma^{2(i-1)}(X^T(I_n + \tau^{-1}_c(\Sigma_\phi)^{-1}X)^{-1}X)^{-1}
   \]
}
5. Use composite sampling to generate $\phi^{(i)}$ ($i = 1, \ldots, K$) from its full conditional distribution which is provided in the Appendix.

rate, average interval length (IL), and mean squared error (MSE).

The vague naïve prior distributions we consider come from the CARBayes R package (Lee, 2013) and Best et al. (1999). Both of these approaches assume gamma(shape, rate) prior distributions for the precisions of both unstructured and spatial random effects. Specifically, the CARBayes package (version 4.0) has implemented gamma(0.001, 0.001) prior distributions as the default for both precision parameters. Best et al. (1999) has used gamma(0.001, 0.001) and gamma(0.1, 0.1) prior distributions for the precisions of the unstructured and spatial random effects, respectively. Previously, the CARBayes package used uniform prior distributions which are no longer available. We adopt the prior distributions implied by the CARBayes package and Best et al. (1999) under our signal-to-noise ratio parametrization. For the purpose of reporting, we refer to these methods as CARBayes and NB (lead author initials), respectively.

Since we have adopted a signal-to-noise ratio parametrization in this work, the gamma($\alpha_1, \beta_1$) and gamma($\alpha_2, \beta_2$) prior distributions for the precisions of the unstructured and spatial
random effects, respectively, are re-parametrized. The implied marginal prior for \( \tau_c \) under
the signal-to-noise ratio parametrization is given by

\[
\pi^V(\tau_c) = \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} (\beta_1 + \beta_2 \tau_c)^{-(\alpha_1 + \alpha_2)} \tau_c^{\alpha_2 - 1},
\]

where \( \Gamma(\cdot) \) is the Gamma function.

When comparing priors, it is important to consider not only the regions in the parameter
space that receive the most mass, but also the rate of decay of the tail of each prior. It is
clear from Equation (5.25) that with regard to tail behavior, as \( \tau_c \to \infty \), \( \pi^V(\tau_c) = O(\tau_c^{\alpha_1 - 1}) \)
and as \( \tau_c \to 0 \), \( \pi^V(\tau_c) = O(\tau_c^{\alpha_2 - 1}) \). Thus, for values of \( \alpha_1 \) and \( \alpha_2 \) closer to 0, the vague
naive prior is closer to impropriety. From Proposition 5.2, we can see that the tail behavior
of the reference prior is considerably different than that of the CARBayes and NB priors.
Consequently, the slower decay rate as \( \tau_c \to \infty \) implies that the CARBayes and NB priors
place significantly more mass on large values of \( \tau_c \). Figure 5.1(a) provides a plot of each of
the prior distributions for \( \tau_c \) that we consider. The plot of \( \log \pi(\tau_c) \) versus \( \log_{10} \tau_c \) provided
in Figure 5.1(b) shows that the reference prior places more mass on values of \( \tau_c \) between
0.01 and 56 (CARBayes) and 0.02 and 35 (NB), while the CARBayes and NB priors place
significantly more mass on large values of \( \tau_c \).

In order to better understand the importance of the parameter \( \tau_c \) as it relates to the strength
of spatial dependence, we consider the Kullback-Leibler divergence between the spatial model
and the independent data model as a function of \( \tau_c \). Specifically, if \( p_i \) and \( p_s \) correspond to the
independent and spatial models, respectively, we consider the Kullback-Leibler divergence
given by

\[
KL(p_i \parallel p_s) = \int p_i(x) \log \left( \frac{p_i(x)}{p_s(x)} \right) dx
= -\frac{n}{2} + \frac{1}{2} \log |\Omega| + \frac{1}{2} \sum_{i=1}^{n} (1 + \tau_c^{-1} d_i^{-1})^{-1}.
\]

(5.26)
Figure 5.1: (a) Raw plot of $\pi(\tau_c)$ and (b) plot of log $\pi(\tau_c)$ vs. log$_{10} \tau_c$ for the reference, CARBayes, and NB priors. The CARBayes and NB priors both approach $\infty$ as $\tau_c \to 0$, while the reference prior does not. Panel (c) shows the Kullback-Leibler divergence per observation between the spatial model and the independent data model across values of $\tau_c$, indicating that large values of $\tau_c$ correspond to a model for nearly independent data.
Note that as $\tau_c \to \infty$, $KL(p_i \mid p_s) \to 0$ indicating that the spatial and independent data models are nearly identical for large values of $\tau_c$. Specifically, Figure 5.1(c) shows that for values of $\tau_c < 10$, the Kullback-Leibler divergence per observation is nearly zero, indicating that the spatial model we consider is only relevant for smaller values of $\tau_c$. Hence, the values of $\tau_c$ for which the reference prior places more mass make practical sense, since large values of $\tau_c$ correspond to nearly independent data. The CARBayes and NB priors unnecessarily place significantly more mass on the large values of $\tau_c$. In contrast, the reference prior intuitively places mass on values of $\tau_c$ that are typical of spatially dependent data.

### 5.5.2 Simulation Design

For the design of our simulations, we consider square regions with three different sample sizes of $n = 5^2, 7^2, 10^2$. We fix $\sigma^2 = 2$, while considering values of $\tau_c$ between 0.01 and 100. To explore how the number of neighbors which exert spatial dependence on a subregion affects inferential performance, we investigate both first- and second-order neighborhood structure (i.e., diagonal subregions are/are not considered neighbors, respectively). We also consider $p = 1$ (intercept only) with $\beta = 1$ and $p = 6$ with $\beta = (-3, -2, -1, 1, 2, 3)^T$. All regressor variables are generated from a normal distribution with mean 0 and variance 1.

We generate results based on 1,000 simulated data sets for each combination of these levels of $n$, $\tau_c$, neighborhood structure, and $p$. For each data set, 15,000 MCMC iterations are obtained with the first 5,000 iterations discarded as burn-in. The Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992) for $\tau_c$ and $\sigma^2$ for three MCMC chains using one setting of parameter values was used to determine the number of MCMC iterations needed to obtain convergence. Gelman-Rubin convergence diagnostics for $\tau_c$ and $\sigma^2$ were calculated to be less than 1.01, indicating that 15,000 MCMC iterations is sufficient for our simulation study. The variances of the proposal distributions described in Section 5.4 were chosen to be $\delta_s = \delta_t = 0.5$ yielding an acceptance rate close to 40%. Due to the heavily right-skewed posterior density for $\tau_c$, we consider 95\% highest probability density regions for $\tau_c$, while 95\% equal-tailed
credible intervals are considered for $\sigma^2$ and $\beta$.

Performance results pertaining to $\tau_c$ less than or equal to 10, first-order neighborhood structure, and $p = 6$ are displayed graphically and discussed in Section 5.5.3. For the sake of brevity, performance results pertaining to $\beta$, $\sigma^2$, second-order neighborhood structure, or $p = 1$ are described in Section 5.5.3 and tabulated in the Appendix. Simulation results for values of $\tau_c > 10$ indicate that the performance of all considered priors deteriorates as $\tau_c$ increases (see Simulation Results in Appendix). This deterioration of performance is not a problem, because as illustrated by the Kullback-Leibler divergence shown in Figure 5.1(c), larger values of $\tau_c$ indicate that the observations are nearly independent, which may imply that a model without spatial dependence would be preferred. We focus on values of $\tau_c < 10$, since our reference prior is intended for a model that accounts for spatial dependence.

5.5.3 Simulation Results

Figure 5.2 displays the frequentist interval coverage rate and average IL (on the log$_{10}$ scale) for $\tau_c$ resulting from the reference, CARBayes, and NB priors across a range of values for $\tau_c$ (on the log$_{10}$ scale) for various sample sizes, assuming $p = 6$ and first-order neighborhood structure. Figure 5.2 demonstrates that the reference prior achieves nominal interval coverage and low average interval length as $n$ increases across a range of values for $\tau_c$. Further, the CARBayes and NB priors exhibit a dip in interval coverage for values of $\tau_c$ corresponding to strong spatial dependence. The CARBayes prior yields particularly wide average interval lengths, potentially detracting from the value of its close-to-nominal coverage. These results demonstrate that the reference prior leads to favorable coverage and interval length compared to the two gamma priors.

Mean squared error of the posterior median of $\tau_c$ is shown in Figure 5.3, which demonstrates that use of the reference prior leads to advantageous estimation properties across all sample sizes, especially smaller sample sizes. Use of the CARBayes prior leads to considerably higher MSE than that of the other priors.
All credible intervals for each of the regression coefficients $\beta$ achieve the nominal level for frequentist coverage with comparable interval lengths across all three priors. Note that for the tabulated results in the Appendix for the case of $p = 6$, the frequentist coverage and average interval length are averaged across all six of the regression coefficients since results were comparable for each of the regression coefficients. Additionally, all three priors show reasonable performance in terms of frequentist coverage and average interval length for $\sigma^2$. Results for second-order neighborhood structure are qualitatively similar to those presented for first-order neighborhood structure. Likewise, results for $p = 1$ are qualitatively similar to those for $p = 6$.

In conclusion, the proposed reference prior exhibits a favorable combination of high interval coverage, short average interval length, and low MSE relative to the CARBayes and NB priors, in addition to philosophical appeal and practical convenience relative to vague naïve priors previously used in the literature.

5.6 Case Study

5.6.1 Data Description

To illustrate an objective Bayesian analysis using the model in (5.9) along with the reference prior, we consider a data set containing foreclosure rates as a proportion of all housing transactions for each of the 88 counties in the state of Ohio for the year 2012. This data set is a subset of a larger database of approximately 54 million records from municipalities across the United States between 2005 and 2014 that we obtained from our partnerships with the Virginia Center for Housing Research and Metrostudy, a Hanley Wood Company. The data include very fine spatial and temporal information, including latitude, longitude, and sale date of each closing record. We choose to aggregate the data at the county level to illustrate this analysis. For more details describing how records were classified as foreclosures,
Figure 5.2: Frequentist coverage and $\log_{10}$ average interval length (IL) for $\tau_c$ for $n = 100$ (top row), $n = 49$ (middle row), and $n = 25$ (bottom row). Reference prior shows favorable performance in terms of frequentist coverage and average interval length.
Figure 5.3: $\log_{10}$MSE for the posterior median of $\tau_c$ for (a) $n = 100$, (b) $n = 49$, and (c) $n = 25$. Reference prior leads to favorable performance in terms of estimation of $\tau_c$. Note that since the y-axis is on the $\log_{10}$ scale, the difference in MSE of the posterior median of $\tau_c$ between the reference prior and the CARBayes prior is considerable.
see Keefe et al. (2017b). In addition to foreclosure rates, county unemployment rates are also available as a potential regressor to be used in modeling (Bureau of Labor Statistics, 2012). For each county, the total number of housing transactions and the observed number of foreclosures are available. Analogously to what is typically done in disease mapping, we consider the standardized morbidity ratio (SMR) for each county, defined by $\text{SMR}_i = \frac{O_i}{E_i}$, where $O_i$ is the observed number of foreclosures in county $i$ and $E_i$ is the expected number of foreclosures in county $i$. The expected counts are calculated by $E_i = n_i \left( \frac{\sum_j O_j}{\sum_j n_j} \right)$, where $n_i$ is the total number of housing transactions in county $i$. A map of the SMRs for foreclosure rates in Ohio counties for 2012 is shown in Figure 5.4(a). Figure 5.4(b) shows a map of the unemployment rates in 2012 for each county.

### 5.6.2 Modeling

Using the 2012 foreclosure rate data set, we fit the model given in (5.9) using the reference, CARBayes, and NB priors in order to compare the results using the proposed reference prior to the results using the priors used frequently in practice. We consider $y_i = \log(\text{SMR}_i)$ as the response variable and unemployment rate as a regressor. The posterior distribution was sampled using the MCMC algorithm described in Section 5.4 with 25,000 iterations where the first 5,000 were discarded as burn-in. Gelman-Rubin convergence diagnostics were calculated using four MCMC chains with the reference prior implemented with different starting values for $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\tau_c$ and $\sigma^2$. The 25,000 MCMC iterations with 5,000 burn-in iterations yielded Gelman-Rubin diagnostics that were all below 1.005, indicating acceptable posterior convergence. Posterior summaries, including posterior medians and credible intervals for the parameters are provided in Table 5.1. Note that 95% HPD credible intervals are used for $\tau_c$, while 95% equal-tailed credible intervals are used for $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\sigma^2$.

From Table 5.1, we see that posterior estimates for the fixed effect regression coefficients and $\sigma^2$ are all similar. Furthermore, note that the credible interval for $\tau_c$ is wider when the CARBayes prior is used, while use of the reference and NB priors results in more narrow
Figure 5.4: Map of (a) 2012 SMRs for foreclosure rates, (b) 2012 unemployment rates (%), (c) posterior median $E[\text{SMR}]$, and (d) posterior standard deviation of $E[\text{SMR}]$ in Ohio counties.
Table 5.1: Posterior Summaries for Foreclosure Rate Case Study for Bayesian Analyses Using Reference, CARBayes, and NB Priors

<table>
<thead>
<tr>
<th>Prior</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>$\beta_0$ (intercept)</td>
<td>-0.5561</td>
<td>(-0.9388, -0.1710)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$ (unemployment rate)</td>
<td>0.0488</td>
<td>(0.0008, 0.0967)</td>
</tr>
<tr>
<td></td>
<td>$\tau_c$</td>
<td>0.2519</td>
<td>(0.0017, 0.9781)</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.0432</td>
<td>(0.0046, 0.0854)</td>
</tr>
<tr>
<td>CARBayes</td>
<td>$\beta_0$ (intercept)</td>
<td>-0.5612</td>
<td>(-0.9482, -0.1708)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$ (unemployment rate)</td>
<td>0.0496</td>
<td>(0.0009, 0.0977)</td>
</tr>
<tr>
<td></td>
<td>$\tau_c$</td>
<td>0.1959</td>
<td>(0.0009, 1.0231)</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.0385</td>
<td>(0.0013, 0.0819)</td>
</tr>
<tr>
<td>NB</td>
<td>$\beta_0$ (intercept)</td>
<td>-0.5687</td>
<td>(-0.9516, -0.1802)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$ (unemployment rate)</td>
<td>0.0504</td>
<td>(0.0020, 0.0985)</td>
</tr>
<tr>
<td></td>
<td>$\tau_c$</td>
<td>0.1589</td>
<td>(0.0010, 0.6290)</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.0345</td>
<td>(0.0019, 0.0755)</td>
</tr>
</tbody>
</table>

credible intervals. This supports the findings of our simulation study regarding the average interval length for $\tau_c$. Although the credible interval for $\tau_c$ using the CARBayes prior is only slightly wider in the case study, the simulation study results illustrate the potential danger of using this particular vague prior. While the analysis of foreclosure rates does not result in undesirably wide credible intervals for $\tau_c$ when the CARBayes prior is used, our simulation study shows that other data sets may lead to credible intervals for $\tau_c$ that are orders of magnitude wider than those obtained when the reference prior is used. Unreasonably wide credible intervals for $\tau_c$ could lead practitioners to decide that a spatial model is not necessary for the data, resulting in the incorrect use of a non-spatial model. We thus recommend use of the reference prior.

The estimate of the regression coefficient for unemployment rate is positive and its credible interval does not contain zero in all three cases. This implies that, on average, as unemployment rate increases, foreclosure rates also increase. Consider the posterior distribution of the expected SMR in county $i$ given by $E[\text{SMR}_i] = \exp\{\beta_0 + \beta_1 x_i + \phi_i\}$, where $x_i$ is the unemployment rate of county $i$ for $i = 1, 2, \ldots, n$. Figures 5.4(c) and 5.4(d) show maps of the posterior median and posterior standard deviation, respectively, of the expected SMR values.
computed directly from the results of the MCMC algorithm. These maps show that counties with higher unemployment rates tend to have higher risk of foreclosure than counties with lower unemployment rates.

It is often difficult to choose sensible hyperparameters for the commonly used priors for this type of spatial model. For instance, it might be difficult for a housing market expert to articulate their understanding of the spatial dependence among counties’ foreclosure rates in such a way as to inform prior specification for a Bayesian hierarchical model. The proposed reference prior is an appealing alternative because it is automatic. It does not require the choice of hyperparameters and has favorable performance. This reference prior is useful in situations where researchers are unsure of how to subjectively choose priors for areal data with a nugget effect.

5.7 Discussion

This paper presents a fully Bayesian analysis for a commonly used Gaussian hierarchical model for spatial data. We have derived explicit expressions for several objective priors, including the reference, independence Jeffreys, and Jeffreys-rule priors for the Gaussian hierarchical model with an intrinsic CAR prior for the spatial random effects. We have shown that the reference prior results in a proper posterior distribution, while the independence Jeffreys and Jeffreys-rule priors lead to improper posterior distributions. Furthermore, we have studied frequentist properties of Bayesian procedures using the proposed reference prior and two commonly used priors for this model.

We have determined that the reference prior leads to a combination of favorable frequentist coverage, average interval length, and mean squared error relative to the two commonly used priors. We have demonstrated using the Kullback-Leibler divergence that focus should be placed on small values of $\tau_c$, which correspond to strong spatial dependence. So, while inferential performance deteriorates across all considered priors for larger values of $\tau_c$, this is
not a concern since perhaps a model without spatial dependence would be preferred in this situation. More importantly, the reference prior performs better than the competing priors for smaller values of $\tau_c$ in terms of frequentist coverage, average interval length, and mean squared error.

Additionally, the reference prior approach obviates the need to subjectively specify hyperparameters, allowing for an automatic Bayesian analysis. This philosophy works well for hierarchical modeling when interpretation of hyperparameters and elicitation of meaningful priors is difficult. Often times, practitioners choose to use a naïve prior in the absence of prior information or understanding of the problem. However, while proper priors lead to proper posterior distributions, the vague naïve prior only masks the impropriety issue, rather than solving it (Berger et al., 2006). By contrast, an objective prior like the reference prior proposed here, will lead to a proper posterior distribution and will let the data speak for themselves.

Although much research has been done on objective Bayesian analysis for spatial models, such as geostatistical models (Berger et al., 2001; De Oliveira, 2007), proper CAR models for the observed data (Ferreira and De Oliveira, 2007; De Oliveira, 2012; Ren and Sun, 2013), and proper CAR models for latent process models for areal data (Ren and Sun, 2014), the hierarchical model with an intrinsic CAR prior that we consider here is one of the most popularly used models in applications of disease mapping and other areas of research. All intrinsic CAR models require subjective decisions regarding rigorous specification and additional constraints to ensure propriety (Lavine and Hodges, 2012). Keefe et al. (2017a) have shown that by first imposing the sum-zero constraint and then considering the limit to obtain the sum-zero constrained intrinsic CAR, as we have done here, there is a broad class of proper CAR priors that all result in a unique intrinsic CAR prior. Our development of an appropriate objective Bayesian analysis for this model will hopefully help researchers analyze areal data.

There are many possible avenues for future research. For example, this work could be
extended by developing objective priors for areal models for non-Gaussian responses, such as counts and rates that can be more accessible to disease mapping applications. Other future work may also include the study and development of objective priors for spatio-temporal models for areal data, as well as the effect of the choice of priors for these types of models.

Appendix

Proof of Lemma 5.1.

\[ \Sigma_{\phi \lambda} = P \Sigma \lambda P^T \]
\[ = P [\lambda I + H]^{-1} P^T \]
\[ = P [\lambda I + QDQ^T]^{-1} P^T \]
\[ = P (Q [\lambda I + D] Q^T)^{-1} P^T \]
\[ = PQ [\lambda I + D]^{-1} Q^T P \]

Note that

\[ PQ = (I - \frac{1}{n} 11^T)(q_1, q_2, \ldots, q_n) \]
\[ = (q_1, q_2, \ldots, q_n) - \frac{1}{n} 11^T(q_1, q_2, \ldots, q_n) \]
\[ = (q_1, q_2, \ldots, q_n) - \frac{1}{n} (0, 0, \ldots, 0, \sqrt{n}) \]
\[ = (q_1, q_2, \ldots, q_n) - \frac{1}{n} (0, 0, \ldots, 0, 1) \]
\[ = (q_1, q_2, \ldots, q_{n-1}, 0) \]
Therefore,

\[ \Sigma_{\phi \lambda} = P \Sigma_{\lambda} P^T \]
\[ = (q_1, q_2, \ldots, q_{n-1}, 0) [\lambda I + D]^{-1} (q_1, q_2, \ldots, q_{n-1}, 0)^T \]
\[ = \sum_{i=1}^{n-1} \frac{1}{\lambda + d_i} q_i q_i^T. \]

\[ \square \]

**Proof of Proposition 5.1.** The integrated likelihood of \( \tau_c \) given \( y \) and \( X \) is given by

\[ L^I(\tau_c; y, X) \propto (|\Omega| |X^T \Omega^{-1} X|)^{-1/2} (S^2)^{-\left(\frac{n-p}{2}+a-1\right)}. \]

First, consider the expression \( |\Omega| |X^T \Omega^{-1} X| \).

By Lemma 5.2 and De Oliveira (2007),

\[ \log(|\Omega| + \log(|X^T \Omega^{-1} X|)) = \prod_{j=1}^{n-p} \left( 1 + \frac{1}{\tau_c} \xi_j \right) \]
\[ = O\left(\tau_{c}^{p-n}\right) \text{ as } \tau_c \to 0 \]

Likewise, we get \( O(1) \) as \( \tau_c \to \infty \).

Now, consider \( S^2 = y^T \left[ \Omega^{-1} - \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} \right] y \).

By Lemma 5.2 and De Oliveira (2007),

\[ S^2 \overset{d}{=} \sigma^2 \sum_{j=1}^{n-p} \left( \frac{1 + \tau_{cs}^{-1} \xi_j}{1 + \tau_c^{-1} \xi_j} \right) Z_j^2, \]

where \( \sigma^2 \) and \( \tau_{cs} \) are the true values of \( \sigma^2 \) and \( \tau_c \), respectively, and \( \{Z_j^2\} \overset{iid}{\sim} \chi_1^2. \)
So,

\[ S^2 = O(\tau_c) \text{ as } \tau_c \to 0 \]
\[ S^2 = O(1) \text{ as } \tau_c \to \infty \]

Therefore,

\[ L^I(\tau_c; y, X) = O(\tau_c^{1-a}) \text{ as } \tau_c \to 0 \]
\[ L^I(\tau_c; y, X) = O(1) \text{ as } \tau_c \to \infty \]

Proof of Theorem 5.1. The reference prior given in (5.19) in Section 5.3.2 is obtained by

\[ \pi^R(\eta) \propto |I^I(\sigma^2, \tau_c)|^{1/2}, \text{ where } I^I(\sigma^2, \tau_c) \text{ is the Fisher information matrix with } (i,j) \text{ entry given by} \]

\[ [I^I(\sigma^2, \tau_c)]_{ij} = -E \left\{ \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log L^I(\sigma^2, \tau_c; y, X) \right\}, \theta_1 = \sigma^2, \theta_2 = \tau_c. \]
Based on the expression for the log-integrated likelihood in (5.13) we can obtain the following

\[
[I'(\sigma^2, \tau_c)]_{11} = \frac{n-p}{2\sigma^4},
\]

\[
[I'(\sigma^2, \tau_c)]_{12} = -\frac{1}{2\sigma^2\tau_c^2} \sum_{j=1}^{n-p} \left( \frac{\xi_j}{1 + \frac{1}{\tau_c}\xi_j} \right),
\]

\[
[I'(\sigma^2, \tau_c)]_{21} = [I'(\sigma^2, \tau_c)]_{12},
\]

\[
[I'(\sigma^2, \tau_c)]_{22} = \frac{1}{2\tau_c^4} \sum_{j=1}^{n-p} \left( \frac{\xi_j}{1 + \frac{1}{\tau_c}\xi_j} \right)^2,
\]

where expectations were evaluated for \(\sigma_*^2 = \sigma^2\) and \(\tau_{cs} = \tau_c\). Thus, we can obtain

\[
\pi^R(\sigma^2, \tau_c) \propto \frac{1}{\sigma^2\tau_c} \left[ \sum_{j=1}^{n-p} \left( \frac{\xi_j}{\tau_c + \xi_j} \right)^2 - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( \frac{\xi_j}{\tau_c + \xi_j} \right) \right\}^2 \right]^{1/2}.
\]

Proof of Proposition 5.2. Suppose that \(\kappa_j = 1/\xi_j\) for \(j = 1, 2, \ldots, n-p\) and \(\bar{\kappa} = \frac{1}{n-p} \sum_{j=1}^{n-p} \kappa_j\).

To show this result, we use the second-order Taylor series expansion of \(\frac{1}{x}\) and \(\frac{1}{x^2}\) about 1.

First, we consider the tail behavior as \(\tau_c \to 0\). Here, we evaluate the Taylor series expansions
at \( x = 1 + \kappa_j \tau_c \).

\[
\pi^R(\tau_c) \propto \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( \frac{\tau_c + \xi_j}{\xi_j} \right)^{-2} - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( \frac{\tau_c + \xi_j}{\xi_j} \right)^{-1} \right\} \right]^{2^{1/2}} \\
= \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} (\kappa_j \tau_c + 1)^{-2} - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} (\kappa_j \tau_c + 1)^{-1} \right\} \right]^{2^{1/2}} \\
= \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( 1 - 2\kappa_j \tau_c + 3\kappa_j^2 \tau_c^2 + O(\tau_c^3) \right) - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( 1 - \kappa_j \tau_c + \kappa_j^2 \tau_c^2 + O(\tau_c^3) \right) \right\} \right]^{2^{1/2}} \\
= \frac{\tau_c}{\tau_c} \left\{ \sum_{j=1}^{n-p} (\kappa_j - \bar{\kappa})^2 + O(\tau_c) \right\}^{1/2} \\
= O(1)
\]

Now, we consider the tail behavior as \( \tau_c \to \infty \). Here, we evaluate the Taylor series expansions at \( x = 1 + \xi_j / \tau_c \).

\[
\pi^R(\tau_c) \propto \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( \frac{\tau_c + \xi_j}{\xi_j} \right)^{-2} - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( \frac{\tau_c + \xi_j}{\xi_j} \right)^{-1} \right\} \right]^{2^{1/2}} \\
= \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( 1 - 2\xi_j / \tau_c + 3\xi_j^2 / \tau_c^2 + O(\tau_c^{-3}) \right) - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( 1 - \xi_j / \tau_c + \xi_j^2 / \tau_c^2 + O(\tau_c^{-3}) \right) \right\} \right]^{2^{1/2}} \\
= \frac{1}{\tau_c} \left[ \sum_{j=1}^{n-p} \left( 1 - 2\xi_j / \tau_c + 3\xi_j^2 / \tau_c^2 + O(\tau_c^{-3}) \right) - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left( 1 - \xi_j / \tau_c + \xi_j^2 / \tau_c^2 + O(\tau_c^{-3}) \right) \right\} \right]^{2^{1/2}} \\
= O(\tau_c^{-2})
\]

Therefore, we have

\[
L^I(\tau_c; y, X)\pi^R(\tau_c) = O(\tau_c^{-1}) \times O(1) = O(1) \text{ as } \tau_c \to 0,
\]

\[
L^I(\tau_c; y, X)\pi^R(\tau_c) = O(1) \times O(\tau_c^{-2}) = O(\tau_c^{-2}) \text{ as } \tau_c \to \infty,
\]
which implies that the reference prior and its resulting posterior are proper.

\begin{proof}

Proof of Corollary 5.1. Results (i) and (ii) follow directly from (5.18) and Propositions 5.1 and 5.2.
\end{proof}

Proof of Theorem 5.2. This result follows from Berger et al. (2001) and De Oliveira (2007).

\begin{proof}

Proof of Proposition 5.3. To show this result, we use the second-order Taylor series expansion of $\frac{1}{x}$ and $\frac{1}{x^2}$ about 1, similar to what was done for the proof of Proposition 5.2. Note that since $\gamma_n = 0$, we have the following result

$$\sum_{i=1}^{n} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right)^c = \sum_{i=1}^{n-1} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right)^c,$$

for some constant $c \in \mathbb{R}$. Suppose that $\lambda_i = 1/\gamma_i$. Consider the tail behavior as $\tau_c \to 0$. Here, we evaluate the Taylor series expansions at $x = 1 + \lambda_i \tau_c$.

$$\pi^{d1}(\tau_c) \propto \frac{1}{\tau_c} \left[ \sum_{i=1}^{n} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right)^2 - \frac{1}{n} \left\{ \sum_{i=1}^{n} \left( \frac{\gamma_i}{\tau_c + \gamma_i} \right)^{-1} \right\} \right]^{1/2}$$

$$= \frac{1}{\tau_c} \left[ \sum_{i=1}^{n-1} \left( \frac{\tau_c + \gamma_i}{\gamma_i} \right)^{-2} - \frac{1}{n} \left\{ \sum_{i=1}^{n-1} \left( \frac{\tau_c + \gamma_i}{\gamma_i} \right)^{-1} \right\} \right]^{1/2}$$

$$= \frac{1}{\tau_c} \left[ \sum_{i=1}^{n-1} (1 + \lambda_i \tau_c)^{-2} - \frac{1}{n} \left\{ \sum_{i=1}^{n-1} (1 + \lambda_i \tau_c)^{-1} \right\} \right]^{1/2}$$

$$= \frac{1}{\tau_c} \left[ \sum_{i=1}^{n-1} (1 - 2\lambda_i \tau_c + 3\lambda_i^2 \tau_c^2 + O(\tau_c^3)) - \frac{1}{n} \left\{ \sum_{i=1}^{n-1} (1 - \lambda_i \tau_c + \lambda_i^2 \tau_c^2 + O(\tau_c^3)) \right\} \right]^{1/2}$$

$$= O(\tau_c^{-1})$$

Note that because $\gamma_n = 0$, the constant terms in the summation do not cancel out and a common factor involving $\tau_c$ cannot be factored, as was done for the reference prior. Therefore,
we have

\[ L^I(\tau_c; y, X) \pi^{J_1}(\tau_c) = O(\tau_c^{1-1}) \times O(\tau_c^{-1}) = O(\tau_c^{-1}) \text{ as } \tau_c \to 0, \]

which implies that the independence Jeffreys prior and its resulting posterior are improper. Note that it is not necessary to consider the tail behavior as \( \tau_c \to \infty \), since the above result already demonstrates impropriety. 

Proof of Corollary 5.2. Result follows directly from (5.18) and Propositions 5.1 and 5.3.

Proof of Theorem 5.3. This result follows from Berger et al. (2001) and De Oliveira (2007).

Proof of Proposition 5.4. We know that \( \pi^{J_1}(\tau_c) = O(\tau_c^{-1}) \text{ as } \tau_c \to 0 \). So we only need to consider the expression

\[
h(\tau_c) = \left[ \frac{n-p}{\prod_{j=1}^{n-p} \left( 1 + \frac{1}{\tau_c} \xi_j \right)} \right]^{1/2} \\
\quad = \left[ \prod_{j=1}^{n-p} \left( 1 + \frac{1}{\tau_c} \xi_j \right)/ \prod_{i=1}^{n} \left( 1 + \frac{1}{\tau_c} \gamma_i \right) \right]^{1/2} \times \prod_{i=n-p+1}^{n} \left( \frac{1}{1 + \frac{1}{\tau_c} \gamma_i} \right) \]

\[
= \left[ \frac{O(\tau_c^{-p-n})}{O(\tau_c^{-p-n})} \times O(\tau_c^{p-1}) \right]^{1/2} \\
= O \left( \frac{\tau_c^{p-1}}{\tau_c} \right) 
\]
Therefore, we get

\[ \pi^J(\tau_c) = O(\tau_c^{-1}) \times O\left(\frac{p-1}{\tau_c^2}\right) = O\left(\frac{p-1}{\tau_c^2}\right) \quad \text{as} \quad \tau_c \to 0, \]

\[ L^I(\tau_c; \mathbf{y}, \mathbf{X}) \pi^J(\tau_c) = O(\tau_c^{-p/2}) \times O\left(\frac{p-1}{\tau_c^2}\right) = O(\tau_c^{-3/2}) \quad \text{as} \quad \tau_c \to \infty, \]

which implies that the posterior distribution yielded by the Jeffreys rule prior is improper.

Note that it is not necessary to consider the tail behavior as \( \tau_c \to \infty \), since the above result already demonstrates impropriety.

\[ \square \]

**Proof of Corollary 5.3.** Result follows directly from (5.18) and Propositions 5.1 and 5.4. \( \square \)

**Full Conditional for \( \phi \)**

Taking the approach of Ferreira et al. (2011), assume that \( \mathbf{X}^* \sim N(\mu, \Sigma) \) such that \( \mu \) is \( p \times 1 \) vector and \( \Sigma \) is a \( p \times p \) singular covariance matrix with rank \( q < p \). Then we have the following fact (Muirhead, 2009):

If \( \mathbf{X}^*, \mu, \) and \( \Sigma \) are partitioned such that

\[
\mathbf{X}^* = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},
\]

where \( \mathbf{X}_1, \mathbf{x}_1, \) and \( \mu_1 \) are \( k \times 1 \) vectors, \( \Sigma_{11} \) is a \( k \times k \) matrix and \( \mathbf{x}_1 - \mu_1 \) belongs to the column space of \( \Sigma_{11} \), then the conditional distribution of \( \mathbf{X}_2 \) given \( \mathbf{X}_1 = \mathbf{x}_1 \) is \( N(\mu_{2.1}, \Sigma_{22.1}) \), where \( \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \) and \( \mu_{2.1} = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1) \).

Now, using this result we can let \( \mathbf{X}_1 = \mathbf{y} \) and \( \mathbf{X}_2 = \phi \).
Then, we have

\[ \mu_1 = X\beta \]
\[ \mu_2 = 0 \]
\[ \Sigma_{11} = \sigma^2 \left( I_n + \frac{1}{\tau_c} \Sigma_\phi \right) \]
\[ \Sigma_{12} = \Sigma_{21} = \Sigma_{22} = \frac{\sigma^2}{\tau_c} \Sigma_\phi. \]

Therefore,

\[ (\phi|y, \sigma^2, \tau_c, X, \beta) \equiv X_2|X_1 \sim N(\mu_{2.1}, \Sigma_{22.1}). \]
Table A1: Simulation results for $\tau_c$ with $n = 25$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.975</td>
<td>0.955</td>
<td>1.000</td>
<td>0.998</td>
<td>0.971</td>
<td>0.950</td>
<td>0.650</td>
<td>0.688</td>
<td>0.496</td>
<td>0.338</td>
<td>0.219</td>
<td>0.133</td>
<td>0.089</td>
<td>0.056</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.999</td>
<td>0.984</td>
<td>0.972</td>
<td>0.900</td>
<td>0.300</td>
<td>0.298</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td>0.297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.988</td>
<td>0.995</td>
<td>0.972</td>
<td>0.881</td>
<td>0.926</td>
<td>0.951</td>
<td>0.908</td>
<td>0.733</td>
<td>0.584</td>
<td>0.461</td>
<td>0.338</td>
<td>0.219</td>
<td>0.133</td>
<td>0.089</td>
<td>0.056</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table A2: Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.963</td>
<td>0.957</td>
<td>0.990</td>
<td>0.981</td>
<td>0.909</td>
<td>0.941</td>
<td>0.904</td>
<td>0.301</td>
<td>0.772</td>
<td>0.607</td>
<td>0.302</td>
<td>0.141</td>
<td>0.041</td>
<td>0.026</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.977</td>
<td>0.972</td>
<td>0.960</td>
<td>0.856</td>
<td>0.919</td>
<td>0.955</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.981</td>
<td>0.977</td>
<td>0.921</td>
<td>0.872</td>
<td>0.913</td>
<td>0.949</td>
<td>0.934</td>
<td>0.841</td>
<td>0.584</td>
<td>0.274</td>
<td>0.097</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table A3: Simulation results for $\tau_c$ with $n = 100$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.971</td>
<td>0.986</td>
<td>0.993</td>
<td>0.981</td>
<td>0.951</td>
<td>0.901</td>
<td>0.871</td>
<td>0.846</td>
<td>0.880</td>
<td>0.801</td>
<td>0.760</td>
<td>0.732</td>
<td>0.704</td>
<td>0.656</td>
<td>0.627</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.983</td>
<td>0.952</td>
<td>0.849</td>
<td>0.949</td>
<td>0.919</td>
<td>0.954</td>
<td>0.972</td>
<td>0.968</td>
<td>0.986</td>
<td>0.981</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.886</td>
<td>0.866</td>
<td>0.869</td>
<td>0.827</td>
<td>0.952</td>
<td>0.941</td>
<td>0.894</td>
<td>0.729</td>
<td>0.436</td>
<td>0.138</td>
<td>0.019</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table A4: Simulation results for $\tau_c$ with $n = 25$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.971</td>
<td>0.998</td>
<td>0.999</td>
<td>1.000</td>
<td>0.987</td>
<td>0.946</td>
<td>0.826</td>
<td>0.618</td>
<td>0.390</td>
<td>0.241</td>
<td>0.130</td>
<td>0.073</td>
<td>0.052</td>
<td>0.031</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.976</td>
<td>0.990</td>
<td>0.978</td>
<td>0.939</td>
<td>0.959</td>
<td>0.968</td>
<td>0.973</td>
<td>0.962</td>
<td>0.968</td>
<td>0.961</td>
<td>0.961</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.966</td>
<td>0.990</td>
<td>0.978</td>
<td>0.948</td>
<td>0.985</td>
<td>0.931</td>
<td>0.830</td>
<td>0.620</td>
<td>0.336</td>
<td>0.132</td>
<td>0.032</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table A5: Simulation results for $\tau_c$ with $n = 49$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.001</th>
<th>0.002</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.97</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>CARBayes</td>
<td>0.97</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>NB</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table A6: Simulation results for $\tau_c$ with $n = 100$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.001</th>
<th>0.002</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.96</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>CARBayes</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>NB</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table A7: Simulation results for $\tau_c$ with $n = 25$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.001</th>
<th>0.002</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>CARBayes</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>NB</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table A8: Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.001</th>
<th>0.002</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>CARBayes</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>NB</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Table A9: Simulation results for $\tau_c$ with $n = 100$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.012</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.976</td>
<td>0.994</td>
<td>0.987</td>
<td>0.949</td>
<td>0.941</td>
<td>0.935</td>
<td>0.896</td>
<td>0.802</td>
<td>0.695</td>
<td>0.579</td>
<td>0.481</td>
<td>0.381</td>
<td>0.291</td>
<td>0.226</td>
<td>0.182</td>
<td>0.140</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.978</td>
<td>0.922</td>
<td>0.845</td>
<td>0.789</td>
<td>0.743</td>
<td>0.797</td>
<td>0.888</td>
<td>0.989</td>
<td>0.989</td>
<td>0.989</td>
<td>0.989</td>
<td>0.988</td>
<td>0.985</td>
<td>0.983</td>
<td>0.983</td>
<td>0.988</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.982</td>
<td>0.922</td>
<td>0.850</td>
<td>0.900</td>
<td>0.947</td>
<td>0.977</td>
<td>0.957</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
</tr>
</tbody>
</table>

Table A10: Simulation results for $\tau_c$ with $n = 25$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.012</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.993</td>
<td>1.000</td>
<td>0.993</td>
<td>0.905</td>
<td>0.911</td>
<td>0.711</td>
<td>0.316</td>
<td>0.208</td>
<td>0.103</td>
<td>0.065</td>
<td>0.039</td>
<td>0.027</td>
<td>0.016</td>
<td>0.013</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.988</td>
<td>0.991</td>
<td>0.984</td>
<td>0.966</td>
<td>0.971</td>
<td>0.977</td>
<td>0.973</td>
<td>0.977</td>
<td>0.971</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.990</td>
<td>0.995</td>
<td>0.933</td>
<td>0.959</td>
<td>0.966</td>
<td>0.946</td>
<td>0.946</td>
<td>0.919</td>
<td>0.218</td>
<td>0.064</td>
<td>0.012</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table A11: Simulation results for $\tau_c$ with $n = 49$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.012</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.982</td>
<td>0.997</td>
<td>1.000</td>
<td>0.976</td>
<td>0.955</td>
<td>0.911</td>
<td>0.711</td>
<td>0.316</td>
<td>0.208</td>
<td>0.103</td>
<td>0.065</td>
<td>0.039</td>
<td>0.027</td>
<td>0.016</td>
<td>0.013</td>
<td>0.009</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.986</td>
<td>0.968</td>
<td>0.990</td>
<td>0.906</td>
<td>0.944</td>
<td>0.902</td>
<td>0.972</td>
<td>0.904</td>
<td>0.970</td>
<td>0.966</td>
<td>0.964</td>
<td>0.904</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.989</td>
<td>0.974</td>
<td>0.909</td>
<td>0.906</td>
<td>0.936</td>
<td>0.919</td>
<td>0.904</td>
<td>0.751</td>
<td>0.372</td>
<td>0.170</td>
<td>0.066</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table A12: Simulation results for $\tau_c$ with $n = 100$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.012</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.975</td>
<td>0.995</td>
<td>0.993</td>
<td>0.962</td>
<td>0.945</td>
<td>0.933</td>
<td>0.876</td>
<td>0.757</td>
<td>0.640</td>
<td>0.516</td>
<td>0.403</td>
<td>0.323</td>
<td>0.258</td>
<td>0.191</td>
<td>0.151</td>
<td>0.135</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.978</td>
<td>0.919</td>
<td>0.836</td>
<td>0.882</td>
<td>0.942</td>
<td>0.975</td>
<td>0.989</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.986</td>
<td>0.916</td>
<td>0.851</td>
<td>0.839</td>
<td>0.949</td>
<td>0.961</td>
<td>0.966</td>
<td>0.947</td>
<td>0.857</td>
<td>0.858</td>
<td>0.859</td>
<td>0.859</td>
<td>0.859</td>
<td>0.859</td>
<td>0.859</td>
<td>0.859</td>
</tr>
</tbody>
</table>
### Table A13: Simulation results for $\sigma^2$ with $n = 25$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>0.016</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>Reference IL</td>
<td>0.006</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>0.003</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.001</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
</tbody>
</table>

### Table A14: Simulation results for $\sigma^2$ with $n = 49$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>0.016</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>Reference IL</td>
<td>0.006</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>0.003</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.001</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
</tbody>
</table>

### Table A15: Simulation results for $\sigma^2$ with $n = 100$, $p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>0.016</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>Reference IL</td>
<td>0.006</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>0.003</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.001</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
</tbody>
</table>

### Table A16: Simulation results for $\sigma^2$ with $n = 25$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>0.016</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>Reference IL</td>
<td>0.006</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>0.003</td>
<td>0.078</td>
<td>0.911</td>
<td>0.968</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.001</td>
<td>0.087</td>
<td>0.968</td>
<td>0.997</td>
<td>0.991</td>
<td>0.985</td>
<td>0.939</td>
<td>0.866</td>
<td>0.761</td>
<td>0.663</td>
<td>0.587</td>
<td>0.515</td>
<td>0.429</td>
<td>0.359</td>
<td>0.282</td>
<td>0.238</td>
</tr>
</tbody>
</table>
Table A17: Simulation results for $\sigma^2$ with $n = 49$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>IL</td>
<td>0.751</td>
<td>0.897</td>
<td>0.968</td>
<td>0.979</td>
<td>0.973</td>
<td>0.941</td>
<td>0.931</td>
<td>0.925</td>
<td>0.924</td>
<td>0.922</td>
<td>0.920</td>
<td>0.924</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>IL</td>
<td>0.719</td>
<td>0.906</td>
<td>0.931</td>
<td>0.868</td>
<td>0.923</td>
<td>0.948</td>
<td>0.951</td>
<td>0.960</td>
<td>0.952</td>
<td>0.956</td>
<td>0.953</td>
<td>0.952</td>
<td>0.951</td>
<td>0.957</td>
<td>0.959</td>
</tr>
<tr>
<td>NB Coverage</td>
<td>IL</td>
<td>0.973</td>
<td>0.973</td>
<td>0.941</td>
<td>0.891</td>
<td>0.929</td>
<td>0.938</td>
<td>0.947</td>
<td>0.946</td>
<td>0.932</td>
<td>0.927</td>
<td>0.925</td>
<td>0.929</td>
<td>0.912</td>
<td>0.931</td>
<td>0.921</td>
</tr>
</tbody>
</table>

Table A18: Simulation results for $\sigma^2$ with $n = 100$, $p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>IL</td>
<td>0.820</td>
<td>0.939</td>
<td>0.960</td>
<td>0.953</td>
<td>0.951</td>
<td>0.937</td>
<td>0.941</td>
<td>0.939</td>
<td>0.940</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.927</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>IL</td>
<td>0.976</td>
<td>0.955</td>
<td>0.894</td>
<td>0.859</td>
<td>0.906</td>
<td>0.932</td>
<td>0.958</td>
<td>0.951</td>
<td>0.951</td>
<td>0.949</td>
<td>0.949</td>
<td>0.951</td>
<td>0.953</td>
<td>0.948</td>
<td>0.952</td>
</tr>
<tr>
<td>NB Coverage</td>
<td>IL</td>
<td>0.992</td>
<td>0.997</td>
<td>0.988</td>
<td>0.943</td>
<td>0.923</td>
<td>0.947</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.941</td>
<td>0.942</td>
<td>0.941</td>
<td>0.941</td>
<td>0.939</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Table A19: Simulation results for $\sigma^2$ with $n = 25$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>IL</td>
<td>0.774</td>
<td>0.929</td>
<td>0.990</td>
<td>0.991</td>
<td>0.956</td>
<td>0.912</td>
<td>0.901</td>
<td>0.892</td>
<td>0.894</td>
<td>0.893</td>
<td>0.893</td>
<td>0.893</td>
<td>0.894</td>
<td>0.894</td>
<td>0.891</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>IL</td>
<td>0.945</td>
<td>0.990</td>
<td>0.981</td>
<td>0.947</td>
<td>0.968</td>
<td>0.955</td>
<td>0.944</td>
<td>0.942</td>
<td>0.946</td>
<td>0.942</td>
<td>0.946</td>
<td>0.943</td>
<td>0.940</td>
<td>0.946</td>
<td>0.940</td>
</tr>
<tr>
<td>NB Coverage</td>
<td>IL</td>
<td>0.990</td>
<td>0.958</td>
<td>0.959</td>
<td>0.958</td>
<td>0.949</td>
<td>0.935</td>
<td>0.926</td>
<td>0.919</td>
<td>0.919</td>
<td>0.921</td>
<td>0.917</td>
<td>0.916</td>
<td>0.917</td>
<td>0.918</td>
<td>0.916</td>
</tr>
</tbody>
</table>

Table A20: Simulation results for $\tau_c$ with $n = 49$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage</td>
<td>IL</td>
<td>0.831</td>
<td>0.949</td>
<td>0.978</td>
<td>0.976</td>
<td>0.954</td>
<td>0.934</td>
<td>0.923</td>
<td>0.913</td>
<td>0.914</td>
<td>0.912</td>
<td>0.910</td>
<td>0.916</td>
<td>0.910</td>
<td>0.904</td>
<td>0.910</td>
</tr>
<tr>
<td>CARBayes Coverage</td>
<td>IL</td>
<td>0.909</td>
<td>0.952</td>
<td>0.894</td>
<td>0.887</td>
<td>0.945</td>
<td>0.959</td>
<td>0.956</td>
<td>0.954</td>
<td>0.952</td>
<td>0.955</td>
<td>0.956</td>
<td>0.956</td>
<td>0.951</td>
<td>0.954</td>
<td>0.956</td>
</tr>
<tr>
<td>NB Coverage</td>
<td>IL</td>
<td>0.994</td>
<td>0.962</td>
<td>0.931</td>
<td>0.913</td>
<td>0.903</td>
<td>0.897</td>
<td>0.944</td>
<td>0.942</td>
<td>0.939</td>
<td>0.937</td>
<td>0.938</td>
<td>0.932</td>
<td>0.919</td>
<td>0.931</td>
<td>0.929</td>
</tr>
</tbody>
</table>
Table A21: Simulation results for $\sigma^2$ with $n = 100$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.885</td>
<td>0.948</td>
<td>0.969</td>
<td>0.956</td>
<td>0.941</td>
<td>0.935</td>
<td>0.927</td>
<td>0.927</td>
<td>0.927</td>
<td>0.927</td>
<td>0.924</td>
<td>0.924</td>
<td>0.924</td>
<td>0.924</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.999</td>
<td>0.931</td>
<td>0.864</td>
<td>0.875</td>
<td>0.931</td>
<td>0.935</td>
<td>0.936</td>
<td>0.935</td>
<td>0.935</td>
<td>0.935</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td>0.932</td>
<td></td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.966</td>
<td>0.933</td>
<td>0.877</td>
<td>0.908</td>
<td>0.947</td>
<td>0.955</td>
<td>0.949</td>
<td>0.945</td>
<td>0.944</td>
<td>0.943</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
<td></td>
</tr>
</tbody>
</table>

Table A22: Simulation results for $\sigma^2$ with $n = 25$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.778</td>
<td>0.960</td>
<td>0.997</td>
<td>0.991</td>
<td>0.953</td>
<td>0.911</td>
<td>0.891</td>
<td>0.885</td>
<td>0.885</td>
<td>0.882</td>
<td>0.878</td>
<td>0.883</td>
<td>0.888</td>
<td>0.889</td>
<td>0.887</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.937</td>
<td>0.953</td>
<td>0.972</td>
<td>0.969</td>
<td>0.959</td>
<td>0.951</td>
<td>0.936</td>
<td>0.938</td>
<td>0.939</td>
<td>0.940</td>
<td>0.940</td>
<td>0.938</td>
<td>0.937</td>
<td>0.939</td>
<td>0.933</td>
<td></td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.934</td>
<td>0.968</td>
<td>0.979</td>
<td>0.971</td>
<td>0.942</td>
<td>0.928</td>
<td>0.916</td>
<td>0.913</td>
<td>0.911</td>
<td>0.913</td>
<td>0.912</td>
<td>0.908</td>
<td>0.911</td>
<td>0.916</td>
<td>0.910</td>
<td></td>
</tr>
</tbody>
</table>

Table A23: Simulation results for $\sigma^2$ with $n = 49$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.815</td>
<td>0.937</td>
<td>0.975</td>
<td>0.976</td>
<td>0.944</td>
<td>0.919</td>
<td>0.891</td>
<td>0.894</td>
<td>0.892</td>
<td>0.893</td>
<td>0.889</td>
<td>0.891</td>
<td>0.890</td>
<td>0.889</td>
<td>0.887</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.907</td>
<td>0.943</td>
<td>0.880</td>
<td>0.896</td>
<td>0.937</td>
<td>0.944</td>
<td>0.941</td>
<td>0.942</td>
<td>0.943</td>
<td>0.944</td>
<td>0.944</td>
<td>0.941</td>
<td>0.940</td>
<td>0.940</td>
<td>0.943</td>
<td></td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.904</td>
<td>0.953</td>
<td>0.905</td>
<td>0.917</td>
<td>0.936</td>
<td>0.933</td>
<td>0.928</td>
<td>0.926</td>
<td>0.924</td>
<td>0.923</td>
<td>0.923</td>
<td>0.926</td>
<td>0.925</td>
<td>0.926</td>
<td>0.921</td>
<td></td>
</tr>
</tbody>
</table>

Table A24: Simulation results for $\sigma^2$ with $n = 100$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.880</td>
<td>0.947</td>
<td>0.962</td>
<td>0.954</td>
<td>0.947</td>
<td>0.931</td>
<td>0.916</td>
<td>0.909</td>
<td>0.904</td>
<td>0.904</td>
<td>0.903</td>
<td>0.902</td>
<td>0.901</td>
<td>0.904</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.969</td>
<td>0.992</td>
<td>0.859</td>
<td>0.866</td>
<td>0.919</td>
<td>0.953</td>
<td>0.951</td>
<td>0.947</td>
<td>0.948</td>
<td>0.942</td>
<td>0.947</td>
<td>0.942</td>
<td>0.945</td>
<td>0.947</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.905</td>
<td>0.928</td>
<td>0.943</td>
<td>0.948</td>
<td>0.935</td>
<td>0.944</td>
<td>0.943</td>
<td>0.934</td>
<td>0.933</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
<td>0.934</td>
</tr>
</tbody>
</table>
Table A25: Simulation results for $\beta$ with $n = 25, p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>0.032</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td>0.1</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>0.32</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>3.2</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>10</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>20</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
</tr>
<tr>
<td>30</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>40</td>
<td>0.978</td>
<td>0.978</td>
<td>0.978</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>60</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
</tr>
<tr>
<td>70</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
</tr>
<tr>
<td>80</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>90</td>
<td>0.968</td>
<td>0.968</td>
<td>0.968</td>
</tr>
<tr>
<td>100</td>
<td>0.966</td>
<td>0.966</td>
<td>0.966</td>
</tr>
</tbody>
</table>

Table A26: Simulation results for $\beta$ with $n = 49, p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>0.032</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td>0.1</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>0.32</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>3.2</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>10</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>20</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
</tr>
<tr>
<td>30</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>40</td>
<td>0.978</td>
<td>0.978</td>
<td>0.978</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>60</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
</tr>
<tr>
<td>70</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
</tr>
<tr>
<td>80</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>90</td>
<td>0.968</td>
<td>0.968</td>
<td>0.968</td>
</tr>
<tr>
<td>100</td>
<td>0.966</td>
<td>0.966</td>
<td>0.966</td>
</tr>
</tbody>
</table>

Table A27: Simulation results for $\beta$ with $n = 100, p = 1$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>0.032</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td>0.1</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>0.32</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>3.2</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>10</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>20</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
</tr>
<tr>
<td>30</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>40</td>
<td>0.978</td>
<td>0.978</td>
<td>0.978</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>60</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
</tr>
<tr>
<td>70</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
</tr>
<tr>
<td>80</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>90</td>
<td>0.968</td>
<td>0.968</td>
<td>0.968</td>
</tr>
<tr>
<td>100</td>
<td>0.966</td>
<td>0.966</td>
<td>0.966</td>
</tr>
</tbody>
</table>

Table A28: Simulation results for $\beta$ with $n = 25, p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>0.032</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td>0.1</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>0.32</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
</tr>
<tr>
<td>3.2</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>10</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>20</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
</tr>
<tr>
<td>30</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>40</td>
<td>0.978</td>
<td>0.978</td>
<td>0.978</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>60</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
</tr>
<tr>
<td>70</td>
<td>0.972</td>
<td>0.972</td>
<td>0.972</td>
</tr>
<tr>
<td>80</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>90</td>
<td>0.968</td>
<td>0.968</td>
<td>0.968</td>
</tr>
<tr>
<td>100</td>
<td>0.966</td>
<td>0.966</td>
<td>0.966</td>
</tr>
</tbody>
</table>
Table A29: Simulation results for $\beta$ with $n = 49, p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.968</td>
<td>0.959</td>
<td>0.958</td>
</tr>
<tr>
<td>0.032</td>
<td>0.963</td>
<td>0.952</td>
<td>0.939</td>
</tr>
<tr>
<td>0.1</td>
<td>0.960</td>
<td>0.929</td>
<td>0.919</td>
</tr>
<tr>
<td>0.32</td>
<td>0.947</td>
<td>0.944</td>
<td>0.945</td>
</tr>
<tr>
<td>1</td>
<td>0.947</td>
<td>0.942</td>
<td>0.941</td>
</tr>
<tr>
<td>3.2</td>
<td>0.941</td>
<td>0.949</td>
<td>0.947</td>
</tr>
<tr>
<td>10</td>
<td>0.946</td>
<td>0.948</td>
<td>0.947</td>
</tr>
<tr>
<td>20</td>
<td>0.947</td>
<td>0.950</td>
<td>0.946</td>
</tr>
<tr>
<td>30</td>
<td>0.947</td>
<td>0.951</td>
<td>0.949</td>
</tr>
<tr>
<td>40</td>
<td>0.947</td>
<td>0.951</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>0.947</td>
<td>0.951</td>
<td>0.950</td>
</tr>
<tr>
<td>60</td>
<td>0.948</td>
<td>0.952</td>
<td>0.950</td>
</tr>
<tr>
<td>70</td>
<td>0.948</td>
<td>0.952</td>
<td>0.950</td>
</tr>
<tr>
<td>80</td>
<td>0.948</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>90</td>
<td>0.947</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>100</td>
<td>0.947</td>
<td>0.951</td>
<td>0.950</td>
</tr>
</tbody>
</table>

Table A30: Simulation results for $\beta$ with $n = 100, p = 6$, and first-order neighborhood

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.961</td>
<td>0.953</td>
<td>0.953</td>
</tr>
<tr>
<td>0.032</td>
<td>0.956</td>
<td>0.939</td>
<td>0.939</td>
</tr>
<tr>
<td>0.1</td>
<td>0.953</td>
<td>0.929</td>
<td>0.929</td>
</tr>
<tr>
<td>0.32</td>
<td>0.941</td>
<td>0.944</td>
<td>0.944</td>
</tr>
<tr>
<td>1</td>
<td>0.947</td>
<td>0.942</td>
<td>0.942</td>
</tr>
<tr>
<td>3.2</td>
<td>0.942</td>
<td>0.949</td>
<td>0.949</td>
</tr>
<tr>
<td>10</td>
<td>0.946</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>20</td>
<td>0.949</td>
<td>0.951</td>
<td>0.951</td>
</tr>
<tr>
<td>30</td>
<td>0.950</td>
<td>0.951</td>
<td>0.951</td>
</tr>
<tr>
<td>40</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>60</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>70</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>80</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>90</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>100</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
</tbody>
</table>

Table A31: Simulation results for $\beta$ with $n = 25, p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.985</td>
<td>0.979</td>
</tr>
<tr>
<td>0.032</td>
<td>0.999</td>
<td>0.952</td>
<td>0.949</td>
</tr>
<tr>
<td>0.1</td>
<td>0.976</td>
<td>0.919</td>
<td>0.909</td>
</tr>
<tr>
<td>0.32</td>
<td>0.941</td>
<td>0.917</td>
<td>0.895</td>
</tr>
<tr>
<td>1</td>
<td>0.911</td>
<td>0.921</td>
<td>0.920</td>
</tr>
<tr>
<td>3.2</td>
<td>0.903</td>
<td>0.944</td>
<td>0.943</td>
</tr>
<tr>
<td>10</td>
<td>0.900</td>
<td>0.947</td>
<td>0.947</td>
</tr>
<tr>
<td>20</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>30</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>40</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>60</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>70</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>80</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>90</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>100</td>
<td>0.900</td>
<td>0.950</td>
<td>0.950</td>
</tr>
</tbody>
</table>

Table A32: Simulation results for $\beta$ with $n = 49, p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Reference Coverage</th>
<th>CARBayes Coverage</th>
<th>NB Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.944</td>
<td>0.952</td>
</tr>
<tr>
<td>0.032</td>
<td>0.997</td>
<td>0.897</td>
<td>0.900</td>
</tr>
<tr>
<td>0.1</td>
<td>0.969</td>
<td>0.921</td>
<td>0.920</td>
</tr>
<tr>
<td>0.32</td>
<td>0.941</td>
<td>0.947</td>
<td>0.946</td>
</tr>
<tr>
<td>1</td>
<td>0.927</td>
<td>0.942</td>
<td>0.941</td>
</tr>
<tr>
<td>3.2</td>
<td>0.928</td>
<td>0.944</td>
<td>0.943</td>
</tr>
<tr>
<td>10</td>
<td>0.924</td>
<td>0.949</td>
<td>0.948</td>
</tr>
<tr>
<td>20</td>
<td>0.923</td>
<td>0.951</td>
<td>0.950</td>
</tr>
<tr>
<td>30</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>40</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>50</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>60</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>70</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>80</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>90</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
<tr>
<td>100</td>
<td>0.922</td>
<td>0.950</td>
<td>0.949</td>
</tr>
</tbody>
</table>
Table A33: Simulation results for $\beta$ with $n = 100$, $p = 1$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>1.000</td>
<td>0.990</td>
<td>0.961</td>
<td>0.956</td>
<td>0.954</td>
<td>0.948</td>
<td>0.946</td>
<td>0.945</td>
<td>0.946</td>
<td>0.945</td>
<td>0.945</td>
<td>0.947</td>
<td>0.949</td>
<td>0.948</td>
<td>0.946</td>
<td>0.947</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.915</td>
<td>0.886</td>
<td>0.861</td>
<td>0.817</td>
<td>0.852</td>
<td>0.866</td>
<td>0.809</td>
<td>0.860</td>
<td>0.858</td>
<td>0.858</td>
<td>0.858</td>
<td>0.859</td>
<td>0.858</td>
<td>0.858</td>
<td>0.859</td>
<td>0.857</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.918</td>
<td>0.864</td>
<td>0.860</td>
<td>0.815</td>
<td>0.950</td>
<td>0.954</td>
<td>0.953</td>
<td>0.950</td>
<td>0.952</td>
<td>0.952</td>
<td>0.952</td>
<td>0.953</td>
<td>0.952</td>
<td>0.952</td>
<td>0.952</td>
<td>0.952</td>
</tr>
</tbody>
</table>

Table A34: Simulation results for $\beta$ with $n = 25$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.950</td>
<td>0.955</td>
<td>0.955</td>
<td>0.947</td>
<td>0.945</td>
<td>0.943</td>
<td>0.943</td>
<td>0.942</td>
<td>0.941</td>
<td>0.941</td>
<td>0.941</td>
<td>0.941</td>
<td>0.940</td>
<td>0.941</td>
<td>0.941</td>
<td>0.941</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.956</td>
<td>0.955</td>
<td>0.950</td>
<td>0.945</td>
<td>0.946</td>
<td>0.943</td>
<td>0.945</td>
<td>0.946</td>
<td>0.947</td>
<td>0.947</td>
<td>0.946</td>
<td>0.947</td>
<td>0.946</td>
<td>0.947</td>
<td>0.946</td>
<td>0.947</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.956</td>
<td>0.953</td>
<td>0.946</td>
<td>0.942</td>
<td>0.942</td>
<td>0.943</td>
<td>0.941</td>
<td>0.941</td>
<td>0.941</td>
<td>0.940</td>
<td>0.941</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
<td>0.941</td>
<td>0.940</td>
</tr>
</tbody>
</table>

Table A35: Simulation results for $\beta$ with $n = 49$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.957</td>
<td>0.954</td>
<td>0.952</td>
<td>0.946</td>
<td>0.942</td>
<td>0.943</td>
<td>0.942</td>
<td>0.942</td>
<td>0.939</td>
<td>0.940</td>
<td>0.939</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
<td>0.940</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.947</td>
<td>0.945</td>
<td>0.939</td>
<td>0.938</td>
<td>0.942</td>
<td>0.944</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.946</td>
<td>0.946</td>
<td>0.944</td>
<td>0.945</td>
<td>0.945</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.948</td>
<td>0.946</td>
<td>0.942</td>
<td>0.940</td>
<td>0.942</td>
<td>0.942</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.943</td>
<td>0.942</td>
</tr>
</tbody>
</table>

Table A36: Simulation results for $\beta$ with $n = 100$, $p = 6$, and second-order neighborhood

<table>
<thead>
<tr>
<th>$\tau_c$</th>
<th>0.01</th>
<th>0.032</th>
<th>0.1</th>
<th>0.32</th>
<th>1</th>
<th>3.2</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Coverage IL</td>
<td>0.957</td>
<td>0.955</td>
<td>0.947</td>
<td>0.944</td>
<td>0.942</td>
<td>0.941</td>
<td>0.942</td>
<td>0.942</td>
<td>0.941</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
</tr>
<tr>
<td>CARBayes Coverage IL</td>
<td>0.947</td>
<td>0.936</td>
<td>0.920</td>
<td>0.935</td>
<td>0.944</td>
<td>0.945</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.945</td>
<td>0.945</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
</tr>
<tr>
<td>NB Coverage IL</td>
<td>0.945</td>
<td>0.938</td>
<td>0.938</td>
<td>0.936</td>
<td>0.943</td>
<td>0.944</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
<td>0.946</td>
</tr>
</tbody>
</table>
Bibliography


Chapter 6

Conclusions and Future Work

6.1 Conclusions

In Chapters 2 through 5, we discussed the details of risk-adjusted monitoring and Bayesian hierarchical spatial models for areal data. Relevant literature for the completed work has been discussed. In Chapter 2, an improved implementation scheme for the risk-adjusted Bernoulli CUSUM chart was proposed that incorporates available data in a more timely manner. Then, the details of a spatially risk-adjusted control chart for monitoring concurrent events motivated by the housing transaction data set were provided in Chapter 3. A note that presents a key finding regarding the specification of sum-zero constrained intrinsic CAR models was presented in Chapter 4. In Chapter 5, we developed objective priors for a Bayesian hierarchical model with a Gaussian response variable and an intrinsic CAR component to account for spatial dependence. Properties of fully Bayesian analyses using the derived prior and other commonly used priors were explored. The proposed reference prior was shown to outperform other competing priors in terms of frequentist coverage, average interval length, and mean squared error.

Chapter 2 is a full draft of a manuscript that has been accepted for publication in the
International Journal for Quality in Health Care. Chapter 3 is in its final form and has been published in the Journal of Applied Statistics. Chapter 4 is in preparation for submission to be submitted to the Miscellanea section of Biometrika. Chapter 5 was initially submitted to Bayesian Analysis and is currently being revised for resubmission.

Although this is the conclusion of the dissertation, there are several avenues for future research. In particular, when monitoring a process as is done in Chapters 2 and 3 of this dissertation, parameter estimation is required to determine what is considered to be in-control. As with any parameter estimation, users should be aware of the consequences of estimation error. In order to account for the uncertainty due to estimation, a simple Bayesian framework for process monitoring could be developed that would be robust to estimation error. Furthermore, the development of objective priors for Gaussian hierarchical models with intrinsic CAR components could be extended to accommodate non-Gaussian responses, such as proportions or counts. Research on objective priors for spatial generalized linear mixed models with CAR components would likely have a significant impact, since these types of models are frequently used in practice for disease mapping and neuroimaging. These two possible future research topics are briefly described in the following sections.

6.2 Future Work

6.2.1 Non-Gaussian Response CAR Models

Bayesian hierarchical models for areal data that include conditional autoregressive (CAR) components are frequently used for disease mapping (Clayton and Kaldor, 1987; Besag et al., 1991; Wakefield, 2007). In this case, the response of interest is often an incidence rate or count of cases for some event or disease. With this type of response, it is not ideal to perform transformations on the response variable and then take the approach described in Chapter 5 of this work. Rather, it is more appealing to model incidence rates using some type of
Bayesian hierarchical generalized linear mixed model (GLMM). For an extensive review of disease mapping techniques, see Wakefield (2007).

As was done in Chapter 5, consider a geographical region of interest that is partitioned into \( I \) disjoint subregions that collectively make up the entire region of interest. For example, a state could be divided into several counties. Let \( n_i \) and \( Y_i \) denote the total number of people at risk and the number of observed events, respectively, in region \( i \) for \( i = 1, 2, \ldots, I \). A popular choice of spatial model for disease mapping is the CAR model of Besag et al. (1991). This Poisson Lognormal model is defined by,

\[
Y_i \sim \text{Poisson}(E_i e^{\psi_i}),
\]

\[
\psi_i = x_i^T \beta + \theta_i + \phi_i,
\]

where \( E_i \) is the expected number of cases in subregion \( i \) assuming a constant rate across the entire region, \( x_i^T \beta \) represents a linear model term including covariates, \( \theta_i \) is a random effect to capture extra Poisson variability, and \( \phi_i \) is the spatial random effect upon which the CAR prior is placed. Note that this model often uses \textit{internal standardization} to calculate the \( E_i \) given by,

\[
E_i = n_i \left( \frac{\sum y_i}{\sum n_i} \right).
\]

The random effects \( \theta_i \) are taken to be independent and identically distributed following a normal distribution with mean 0 and precision \( \tau_h \), while the spatial random effects \( \phi_i \) are given by

\[
p(\phi_i|\tau_c, \phi_{j \neq i}) \propto \tau_c^{1/2} \exp \left\{ -\frac{\tau_c}{2} \sum_{i \neq j} w_{ij}(\phi_i - \phi_j)^2 \right\},
\]

where \( w_{ij} = 1 \) if subregions \( i \) and \( j \) are neighbors and \( w_{ij} = 0 \) if subregions \( i \) and \( j \) are not neighbors.

As pointed out by Banerjee et al. (2014), the use of internal standardization leads to a model that is not technically valid since the data appears on both sides of the Poisson model.
specification. Typically, this issue is ignored, however other types of Bayesian GLMM models could also be used, such as the logistic regression model given by

\[ Y_i \sim \text{Binomial} \left( n_i, p_i = \frac{e^{\psi_i}}{1 + e^{\psi_i}} \right), \]

\[ \psi_i = x_i^T \beta + \theta_i + \phi_i. \]

This logistic regression model avoids the internal standardization problem, however, the Poisson Lognormal is the most popular approach. Of course, the logistic regression approach requires \( n_i \) to be known, which is the case in the application described in Chapter 5 in which we modeled foreclosure rates for the counties of Ohio. Ideally, we would like to take a spatial GLMM approach to modeling foreclosures rates.

For the model specification given in Equations (6.1) and (6.4), \( \tau_h \) and \( \tau_c \) are precision parameters and require prior distributions. The choice of these prior distributions can be quite challenging. Bernardinelli et al. (1995) have explored the importance of choosing appropriate priors for this class of models. Typically, vague Gamma priors are placed on the precisions. For examples, see Best et al. (1999), Bell and Broemeling (2000), Reich et al. (2006), and Lee (2013). While in practice, these subjective priors are frequently used, there exists a need for objective or approximate objective style priors for this class of models. Objective priors would alleviate the need to choose hyperparameters of the priors for the precision parameters of these models and would likely lead to comparable, if not improved performance.

Consider a model for a Gaussian response given by

\[ y = X\beta + \theta + \phi, \]

\[ \phi \sim N \left( \mu 1^T, \sigma^2 \xi \Sigma \right), \]

where \( \Sigma^{-1} \) denotes the adjacency matrix. Under this model, we conjecture that the reference prior for \( (\sigma^2, \xi) \) is

\[ \pi(\sigma^2, \xi) \propto \frac{1}{\sigma^2(1 + \xi)^2}. \]
Note that the model in Equation (6.5) is very similar to that of Chapter 5. The concepts described in Chapter 5 should shed light on this area of future work.

If we consider a normal approximation to the binomial distribution, we can explore potential reference priors for the model given in Equation (6.4). Suppose we have \( \logit(p_i) = \psi_i \) and we let \( a = \text{Var}(\logit(\hat{p}_i)) \). Then, the total unstructured variability is given by \( \sigma^2 = a + \tau_h^{-1} \).

Additionally, let \( \xi \) denote a signal-to-noise ratio parametrization given by

\[
\xi = \frac{\tau_c^{-1}}{\sigma^2} = \frac{\tau_h}{\tau_c(a\tau_h + 1)}.
\] (6.7)

Then, performing the necessary transformation allows us to obtain a conjectured reference prior for the precisions of the random effect terms \( (\tau_c, \tau_h) \) given by

\[
\pi(\tau_c, \tau_h) \propto \frac{1}{[\tau_c(a\tau_h + 1) + \tau_h]^2}.
\] (6.8)

We have implemented an MCMC algorithm using this conjectured prior and the results look reasonable. Ideally, future research would focus on formally deriving a mathematical expression for an objective prior. However, this is challenging because in order to obtain objective priors, integrated likelihood functions must be obtained, which is not trivial for a GLMM. The work of approximate inference for GLMMs provided by Breslow and Clayton (1993) and Natarajan and Kass (2000) may be useful in determining an appropriate way to approximate these integrated likelihood functions. Additionally, concepts from the approach of integrated nested Laplace approximation (INLA) (Rue et al., 2009) may also prove useful for this work. Others have also considered concepts that may be useful in this work, including approximate Bayesian inference for spatial GLMMs (Hosseini et al., 2011) and objective priors for generalized linear models (Ibrahim and Laud, 1991).
6.2.2 Bayesian Process Monitoring

In standard process monitoring techniques, a Phase I sample is typically collected. This Phase I sample is used to characterize how the process behaves when it is in-control via estimation of the in-control parameters and to explore features of the process being monitored. An overview of Phase I has been provided by Jones-Farmer et al. (2014). These standard control charting procedures assume that the in-control process parameter is fixed based on its estimated value from the Phase I data. Often times, an adequately sized Phase I sample is not available and it can be difficult to estimate the in-control process parameters. Saleh et al. (2015) and Keefe et al. (2015) have studied the effect of estimation error on the in-control performance of Shewhart and self-starting control charts, respectively. In this case, the uncertainty that is inherent in the estimated in-control process parameters should be accounted for in the monitoring procedure. A potential area for future research is a cumulative sum (CUSUM) monitoring scheme that uses ratios of Bayesian predictive distributions to account for the uncertainty. This approach could make the monitoring scheme more robust to estimation error.

We propose a Bayesian monitoring procedure similar in style to the standard CUSUM control chart. Similar in spirit to Menzefricke (2002), we incorporate predictive distributions into the monitoring procedure in order to account for the uncertainty due to estimation. In our procedure, one monitors the statistics given by

$$B_t = \max(0, B_{t-1} + M_t), \quad t = 1, 2, \ldots,$$

where $M_t$ is similar in style to the standard CUSUM calculation, but constructed using a Bayesian approach with marginal distributions.

Consider a Phase I sample of size $n$ given by $x = (x_1, x_2, \ldots, x_n)$ and consider monitoring for a future observation $y_t$ for $t = 1, 2, \ldots$. Furthermore, we define a parameter that represents
whether the process is in-control or out-of-control given by

\[ \phi_t = \begin{cases} 
0, & \text{if in-control} \\
1, & \text{if out-of-control} 
\end{cases} \]  \hspace{1cm} (6.10)

With this, we define our Bayesian CUSUM statistic \( B_t \) as a function of \( M_t \) given by

\[ M_t = \log \frac{\int f(y_t|\theta_t, \phi_t = 1)f(\theta_t|x, \phi_t = 1)d\theta_t}{\int f(y_t|\theta_t, \phi_t = 0)f(\theta_t|x, \phi_t = 0)d\theta_t} \]  \hspace{1cm} (6.11)

By integrating with respect to \( \theta_t \) in the numerator and denominator, we are accounting for the estimation error in \( \theta_t \) using marginal distributions. This Bayesian monitoring scheme could be compared to the standard CUSUM chart in order to determine whether or not it is more robust to the effects of estimation error.

### 6.3 Final Remarks

Collectively, this dissertation covers useful techniques for statistical monitoring and modeling of spatial processes. We have presented several techniques summarized above, in addition to presenting details of potential future research topics. The information included in this dissertation could serve as a foundation upon which to develop new statistical methodology for both statistical process monitoring and spatial statistics.
Bibliography


