Semiparametric Varying Coefficient Models for Matched Case-Crossover Studies

Ana Maria Ortega-Villa

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

Inyoung Kim, Chair
Christopher Franck
Feng Guo
George Terrell

October 1, 2015
Blacksburg, Virginia

KEYWORDS: Bayesian Nonparametric; Conditional logistic regression; Matched case-control study; Regression splines; Spatial-temporal data; Varying Coefficient Model;

Semiparametric Time Varying Coefficient Model for Matched Case-Crossover Studies

Ana Maria Ortega Villa

(ABSTRACT)

Semiparametric modeling is a combination of the parametric and nonparametric models in which some functions follow a known form and some others follow an unknown form. In this dissertation we made contributions to semiparametric modeling for matched case-crossover data.

In matched case-crossover studies, it is generally accepted that the covariates on which a case and associated controls are matched cannot exert a confounding effect on independent predictors included in the conditional logistic regression model. Any stratum effect is removed by the conditioning on the fixed number of sets of the case and controls in the stratum. However, some matching covariates such as time, and/or spatial location often play an important role as an effect modification. Failure to include them makes incorrect statistical estimation, prediction and inference. Hence in this dissertation, we propose several approaches that will allow the inclusion of time and spatial location as well as other effect modifications such as heterogeneous subpopulations among the data.

To address modification due to time, three methods are developed: the first is a parametric approach, the second is a semiparametric penalized approach and the third is a semiparametric Bayesian approach. We demonstrate the advantage of the one stage semiparametric approaches using both a simulation study and an epidemiological example of a 1-4 bi-directional case-crossover study of childhood aseptic meningitis with drinking water turbidity.

To address modifications due to time and spatial location, two methods are developed: the first one is a semiparametric spatial-temporal varying coefficient model for a small number of locations. The second method is a semiparametric spatial-temporal varying coefficient model, and is appropriate when the number of locations among the subjects is medium to large. We demonstrate the accuracy of these approaches by using simulation studies, and when appropriate, an epidemiological example of a 1-4 bi-directional case-crossover study.

Finally, to explore further effect modifications by heterogeneous subpopulations among strata we propose a nonparametric Bayesian approach constructed with Dirichlet process priors, which clusters subpopulations and assesses heterogeneity. We demonstrate the accuracy of our approach using a simulation study, as well a an example of a 1-4 bi-directional case-crossover study.

To my amazing parents Beatriz and Luis.

Acknowledgments

I would like to express my deepest gratitude to my my advisor, Dr. Inyoung Kim for her mentoring, guidance, support and encouragement as well as for always conveying her love of statistics, throughout my time at Virginia Tech. Dr. Kim is an incredible advisor, always willing to help and encourage her students, as well as always there to offer support for problems whether related to statistics or not. I hope one day I can provide to my students at least a fraction of what she has provided for me.

I would also like to thank my committee members, Dr. Christopher Franck, Dr. Feng Guo and Dr George Terrell for their help, comments and positive input. I would also like to thank Dr. Birch and Dr. Smith for their motivation, support and belief that I could be a good instructor and allowing me the possibility to pursue teaching while in the department.

Lastly I would like to thank my amazing parents Beatriz and Luis, for showing me the importance of education at an early age, and sharing with me their love for their job, their research and more importantly their teaching. For inviting me to their classrooms and showing me the beauty of a well prepared lecture and of classroom interaction between student and professor. But most importantly for their unending love and support for me, and allowing me to follow my dreams. Thank you to Gavin Corral, for being there since the very first day I started the Statistics program, for always reminding me that I could do anything I put my mind to, for helping me get through that demanding first year. Thank you Gavin, for coming to all my talks and for reading all my papers, and standing by me all this time.

I would have not made it here without the help of all of you!

Sincerely,

Ana Maria Ortega Villa

Fall 2015

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Background

1.1.1 The Turbidity and Aseptic Meningitis Relationship

Aseptic meningitis is an inflammation of the layers of the brain associated with non bacterial elements. In children aseptic meningitis is the most common infection of the central nervous system and is one of the main causes of children mortality and morbidity. Park et al. (2003) found that contact with contaminated water is one of the principal methods of transmission of the disease. Additionally Kim et al. (2003) showed that drinking water turbidity is associated with the emergence of aseptic meningitis in children under 15 years of age.

1.1.2 Semiparametric Regression Models

Semiparametric regression incorporates flexibility in the estimation of nonlinear relationships inside regression analyses by combining parametric and nonparametric modeling. These

semiparametric models are more flexible in the sense that the relationships are not determined a priori but are gathered from the data. Our approach to semiparametric regression is that of Rupert et al. (2003), incorporating penalized regression, more specifically a varying coefficient model to matched case-crossover studies.

1.1.3 Semiparametric Varying Coefficient Models

In this dissertation we are implementing semiparametric varying coefficient models in which the effect of one covariate conditional on a fixed value of a second covariate is modeled by a linear relationship. Let X be a covariate, for given values of another covariate (called modifier covariate) T has a linear relationship with the mean of a response variable Y. The respective varying coefficient model is:

$$y_i = \beta(t_i)x_i + \varepsilon_i$$

$\beta(t_i)$ is modeled using penalized linear splines, allowing the slope coefficient to be smooth functions of T.

$$\beta(t_i) = \alpha_0 + \alpha_1 t_i + \sum_{k=1}^{L} \alpha_{1+k}(t_i - \kappa_k)_+$$

where $\kappa_1, \ldots, \kappa_L$ are knots over the t_i values.

The finalized model is

$$y_i = \Big(\alpha_0 + \alpha_1 t_i + \sum_{k=1}^{L} \alpha_{1+k}(t_i - \kappa_k)_+\Big)x_i + \varepsilon_i$$

1.1.4 Matched Case-Crossover Studies

A matched case-crossover study is a special case of matched case-control study. An important feature of the matched case-crossover study is that the control information for each patient is based on his or her own exposure experience. The analysis of matched case-crossover study can be viewed as a stratified data analysis of retrospective, self-matched follow-up studies, each with a sample size of one, see Navidi (1998). A matched case-crossover design uses a sample from a study population of individuals all of whom have experienced the outcome of interest. When the measurements are taken on each subject in an exposed and unexposed setting, each subject acts as their own control. Here we are interested in the situation where there is one measurement for the case period, and M measurements for the control periods.

The analysis of the study with binary responses is based on an extension of the classical prospective logistic regression model where each subject is considered its own stratum. For $1-M$ matched case-crossover studies, the model with the binary outcomes (Y_1, Y_2, ..., Y_{M+1}) and continuous covariate (X_1, X_2, ..., X_{M+1}) with a scalar X_j, $j = 1, ..., M$, matching covariate V, and stratum level S is

$$\Pr(Y_j = 1|X_j, V, S) = \frac{1}{1 + \exp[-(\mathrm{X_j}\beta + \mathrm{q(V, S)}\}]},$$

where $q(\bullet)$ is an arbitrary function that includes the intercept and unknown effects of the strata and matching covariate. The stratifying variable is the individual patient in a matched case-crossover study.

It is generally accepted that covariates for which cases and controls are matched cannot exert a confounding effect on independent predictors included in analysis. This is because a conditional logistic regression model removes any stratum effect and matching covariates, i.e. $q(\bullet)$ disappears (Hosmer et al., 1989) conditioning on the one case and M controls in

the stratum. Therefore, the retrospective conditional logistic regression model with binary outcomes $(Y_1, Y_2, \ldots, Y_{M+1})$ for 1 case-M controls is then expressed as

$$\Pr\Big(Y_1 = 1 | X_1, ..., X_{M+1}, S, V, \sum_{j=1}^{M+1} Y_j = 1\Big) = \frac{1}{1 + \sum_{j=2}^{M+1} \exp\{(X_j - X_1)\beta\}}. \tag{1.1}$$

Hence, the conditional logistic regression model is not able to detect any effects due to the matching covariates by stratum variable.

1.1.5 Spatial Data

Spatial Data can be considered to be of one of three forms: point-referenced data, in which we consider our variables as random vectors of data at a specific location which is considered to vary continuously across an area, areal data in which our area of interest is partitioned into a finite number of units with clear boundaries, and point pattern data in which the interest lies in the spatial pattern of occurrences over an area.

Our proposed spatial-temporal method is focused on point-referenced data in which we consider our response and covariate to be observed at a specific location which is assumed to vary continuously. In point-reference data it is assumed that the covariance between two random variables observed in two locations depends on the distance between them.

1.1.6 Spatial Covariance Functions

In Spatial Statistics the covariance function $C(s_i, s_j)$ describes the spatial covariance of a random field, say $Y(s)$, on a domain D at specific locations s_i and s_j. Commonly the covariance function has a parametric form. Some common covariance structures for spatial data are given by Table 1.1. Ee are using a Gaussian covariance structure in our spatial

simulation and estimation models.

	Covariance Function	Parameters
Exponential	$\sigma^2 \exp\left(\frac{-\|\|d\|\|}{\rho}\right)$	$\sigma^2 > 0,\ \rho > 0$
Gaussian	$\sigma^2 \exp\left(\frac{-\|\|d\|\|^2}{\rho}\right)$	$\sigma^2 > 0,\ \rho > 0$
Spherical	$\sigma^2\left(1 - \frac{3\|\|d\|\|}{2m} + \frac{\|\|d\|\|^3}{2m^3}\right)$	$\sigma^2 > 0,\ \|\|d\|\| \leq m$

Table 1.1: Some covariance functions; d=distance between locations s_i and s_j; σ^2 is the scale parameter for the overall variability of the process; ρ is the dependence range; m is the maximum range of dependence

1.1.7 Bayesian Nonparametric Models

Bayesian Nonparametric Models (BNP) provide solutions to a variety of problems like determining the number of clusters in a mixture model, the number of factors in a factor analysis, the number of variables to keep in a model, cluster identification, among others, by fitting a single model that can adapt to the data (Hjort et al., 2010).

In Chapter 4 we are concerned with the determination of functional clusters among the data. In traditional clustering approaches, we are required to specify the number of clusters before analyzing the data. The BNP approach estimates the number of clusters needed to analyze the data, as well as allows future data to manifest new clusters that have been previously unseen.

In this dissertation cluster selection is done using the Dirichlet Process (DP) along with the Pölya urn representation. The Dirichlet Process (Antoniak, C., 1974; Ferguson, 1973) places a prior on an infinite-dimensional space of random measures, by providing a nonparametric prior specification over the class of possible distribution functions. The details of the DP are presented in Chapter 4.

1.2 Motivation

The main goal in this dissertation is to make contributions to semiparametric modeling for matched case-crossover studies. We do this by addressing the problem of detecting effect modifications arising from time, time and spatial location of the subjects, or heterogeneous subpopulations among the data.

In order to detect modifications arising from time, we propose three different alternatives to detect modifications due to time, of the effect of a covariate over the presence of disease. The first alternative is a two stage parametric approach in which the first stage consists of individually estimating the effect coefficients for each particular time period and the second stage consists of fitting a polynomial regression of these estimated coefficients and time. The second alternative is a one stage semiparametric penalized likelihood based approach in which we incorporate a semiparametric time varying coefficient model to the analysis and conduct estimations using the Newton-Raphson method and a penalized likelihood approach. The third alternative is a fully Bayesian semiparametric time varying coefficient model using Adaptive Rejection Metropolis Sampling within Gibbs Sampling.

In order to detect modifications arising from time and spatial location we propose two methods under the Bayesian hierarchical model framework. The first method is appropriate when there is a small number of locations among the subjects. This method incorporates the estimation of a semiparametric time varying coefficient model, along with its first derivative. The second method is a two-stage spatial-temporal varying coefficient model. The first stage consists in estimating a covariance structure precision parameter using the empirical semivariogram. The second stage consists of using this precision parameter as a hyper parameter in a fully Bayesian approach modeling spatial dependencies under a Gaussian covariance structure.

In order to detect modifications arising from time and heterogeneous subpopulations among the data we propose a functional nonparametric Bayesian method in order to determine functional clusters of time varying coefficients among strata. The main interest in this method is assessing in which way the time varying coefficient function changes between clusters. To achieve this goal we propose the functional nonparametric Bayesian time varying coefficient model. The time varying coefficient model is estimated using regression splines, clustering of time varying coefficients among strata is assessed using a Dirichlet process, and parameter estimation is done using adaptive rejection Metropolis sampling (ARMS) within Gibbs sampling.

1.3 Overview

The rest of this dissertation is organized as follows. In Chapter 2, we introduce three varying time coefficient models to assess the effect modification by time, along with a comparison between them. In Chapter 3, we present two alternatives for spatial-temporal varying coefficient models In Chapter 4, we present the functional nonparametric Bayesian time varying coefficient model. In Chapter 5 we present a general review of the work that we have presented as well as directions for future research.

Chapter 2

Semiparametric Time Varying Coefficient Model for Matched Case-Crossover Studies

2.1 Background

The use of conditional logistic regression in the analysis of matched case-crossover studies does not permit the the estimation of the effect of matching covariates. However, in some cases an effect modification arising from time exists and it is of interest to assess it in order to make correct statistical estimation, prediction and inference.

In the application example, a 1-4 bi-directional matched case-crossover study, our main research goal is to assess the relationship between drinking water turbidity and the presence or absence of aseptic meningitis in children under fifteen years of age. Additionally we are interested in determining whether an effect modification arising from time exists.

Chapter 2. Semiparametric Time Varying Coefficient Model for Matched Case-Crossover Studies

The example data comes from a a bidirectional matched case-crossover study design to assess the risk of hospital admissions from aseptic meningitis after exposure to turbidity increase in drinking water. For each subject, the case period was matched to 4 control periods which are 35 and 42 days before and after the case day. For each patient, thirty one case measurements were collected. These measurements are associated with the latent period of the disease, leading to hospital admission. For each of the thirty one case measurements two control periods (35 and 42 days) prior to the case period and another two control periods (35 and 42 days) posterior to the case period are matched. This scheme was used in Kim et al. (2003).

By matching the day of the week for the case and control periods, potential confounders from the variation of the day of the week were controlled and control days were selected further than the possible latent period. Furthermore, we investigate how drinking water turbidity is associated with an increased risk of aseptic meningitis in children. We also estimate the time interval between increase in drinking water turbidity and the peak risk of aseptic meningitis

The methods for assessing the functional relationship between time varying coefficient and matched binary outcomes are quite limited. To do this we propose a time varying coefficient model in order to evaluate effect modification and make correct statistical inference. Three methods are developed: one is a parametric approach and the other two are semiparametric approaches. Our parametric time varying coefficient approach is a two-stage method developed by using both conditional logistic and polynomial regressions, while our other two semiparametric time varying coefficient approaches are both one-stage methods developed by using regression splines. The first semiparametric is a penalized likelihood based approach and the second, is a full Bayesian approach. Our proposed approaches allow us to not only detect the parametric relationship between the predictor and binary outcomes, but also evaluate parametric or semiparametric relationships between the predictor and time.

The reminder of this chapter is organized as follows. In Section 2.2, we introduce time variation to the analysis of matched case-crossover studies. In Section 2.3 we present a parametric two stage approach to assess effect variation due to time. In Section 2.4, we describe a semiparametric one-stage penalized approach. In Section 2.5, we present a semiparametric one-stage Bayesian approach. In Section 2.7 we present two simulation studies; one is to compare the parametric two stage approach and the semiparametric one stage penalized approach and the other is to compare the semiparametric one stage penalized approach and the Bayesian approach. Our simulation results suggest that the semiparametric one stage Bayesian approach performs better than other two approaches in terms of mean squared error (MSE) and R^2. It is also more efficient and flexible than the penalized approach in terms of estimating the smoothing parameter. In Section 2.8, we apply our approaches to the 1-4 bi-directional case-crossover study. Concluding remarks are provided in Section 2.9.

2.2 Time variation in Matched case-crossover studies

We consider $1-M$ matched case-crossover study with kth stratum, $k = 1, \ldots, N$, where each stratum is the individual subject. Let x_{1k} represent the value of the covariate of interest for the case of stratum k, and x_{2k}, x_{3k}, x_{4k} and $x_{M+1,k}$ the values for the controls of stratum k.

The conditional logistic regression model with time varying coefficient for kth stratum can be written as

$$\begin{aligned} P\Big(Y_{1k}=1|x_{1k},...,x_{M+1,k},k,V,\sum_{j=1}^{M+1}Y_{jk}=1\Big) &= \frac{\exp\{x_{1k}\beta(t)+q(\bullet)\}}{\sum_{j=1}^{M+1}\exp\{x_{jk}\beta(t)+q(\bullet)\}} \\ &= \frac{1}{1+\sum_{j=2}^{M+1}\exp\{(x_{jk}-x_{1k})\beta(t)\}} \\ &\equiv l_k\{\beta(t)\} \end{aligned} \tag{2.1}$$

where $\beta(t)$ is a function of time t, $t = 1, \ldots, T$.

This time varying coefficient model replaces the parameter β in (1.1) by the unknown function $\beta(t)$. We consider $\beta(t)$ to be of either parametric or nonparametric form.

2.3 A Parametric two-stage approach

The function $\beta(t)$ can be estimated using a parametric function of t. We can use a two-stage approach described as the follows:

Step 1: For each t, fit the model (2.1) using the conditional likelihood $l_k\{\beta(t)\}$ and full conditional likelihood $l\{\beta(t)\}$ which is

$$l\{\beta(t)\} = \prod_{k=1}^{N} l_k\{\beta(t)\}, \tag{2.2}$$

where $x^*_{jk} = x_{jk} - x_{1k}$ for $j = 2, ..., M+1$; we obtain estimated parameter, $\hat{\beta}(t)$, for given t, using Newton-Raphson algorithm with Gradient and Hessian functions described in the Appendix A.1.

Step 2: Fit the pth polynomial regression of $\hat{\beta}(t)$ on t, where $p = 1, ..., P_{\max}$, and select the best the p polynomial model using AIC.

This two stage parametric approach can detect not only parametric relationship between the predictor and binary outcomes but also evaluate the parametric relationship between the predictor and time. It is also simple and easy to use. However, it only evaluates the parametric relationship between the predictor and time. To allow the estimation of nonparametric relationships, we further develop semiparametric one stage approaches.

2.4 Semiparametric one-stage penalized approach

We model $\beta(t)$ nonparametrically so that we estimate $\beta(t)$ via a pth order regression splines.

For a pth order regression spline with a truncated power series basis and with L knots, we define a pth order regression spline with a truncated power series basis and with L knots, $\{1, t, t^2, \ldots, t^p, (t-\xi_1)_+^p, (t-\xi_2)_+^p, \ldots, (t-\xi_L)_+^p\}$, where $(u)_+^p = u^p I(u \geq 0)$ and knots $\xi_1 < \ldots < \xi_L$. Knots can be selected a priori, e.g., at sample quantiles of observed t's which are scaled in [0,1]. Let $\alpha_0, \alpha_1, \ldots, \alpha_p$ be the regression coefficients and $\alpha_{p+1}, \ldots, \alpha_{p+L}$ be the regression coefficients of the truncated power series basis. With this setup, $\beta(t)$ can be written as

$$\beta(t) = \alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p. \tag{2.3}$$

The conditional likelihood $l_k\{\beta(t)\}$ for stratum k and the full conditional likelihood are

$$\begin{aligned} \Pr(y_{1k} = 1 | \beta(t), x_{1k}, \ldots, x_{M+1,k}, \sum_{j=1}^{M+1} y_{jk} = 1) &= \frac{\exp\{x_{1k}\beta(t) + q(\bullet)\}}{\sum_{j=1}^{M+1} \exp\{x_{jk}\beta(t) + q(\bullet)\}} \\ &= \frac{1}{1 + \sum_{j=2}^{M+1} \exp\{x_{jk}^* \beta(t)\}} \\ &\equiv l_k\{\beta(t)\} \end{aligned} \tag{2.4}$$

and

$$L\{\beta(t)\} = \prod_{k=1}^{N} l_k[\{\beta(t)\}],$$

respectively.

We estimate all unknown parameters using penalized log-likelihood approach. Let λ be a

smoothing parameter. The penalized log likelihood is

$$\begin{aligned} l\{\beta(t)\} &\equiv \log[L\{\beta(t)\}] \\ &= -\sum_{k=1}^{N} \log\left[1+\sum_{j=2}^{M+1}\exp\Big\{(\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{1+j}(t-\xi_l)_+^p)x_{jk}^*\Big\}\right]+\frac{\lambda}{2}\sum_{l=1}^{L}\alpha_{l+1}^2. \end{aligned}$$

The Newton-Raphson method is used to estimate $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}=(\alpha_0,\alpha_1,\ldots,\alpha_p,\alpha_{p+1},\ldots,\alpha_{p+L})$. The Gradient vector and Hessian matrix are summarized in the Appendix A.2. The smoothing parameter λ is estimated by using a grid search within a preselected range. We choose the smoothing parameter minimizing the AIC or AICc criteria,

$$\begin{aligned} AIC &= 2(L+1)-2\ln[L\{\beta(t)\}] \\ AICc &= AIC+\frac{2(L+p+1)(L+p+1+1)}{N-(L+1)-1} \\ &= AIC+\frac{2(L+p+1)(L+p+2)}{N-L-p-2}, \end{aligned}$$

where L represents the number of knots, p represents the order of the polynomial regression and N represents the number of strata.

This one stage penalized approach allows us to not only detect the parametric relationship between the predictor and binary outcome but also simultaneously evaluate the semiparametric relationship between the predictor and time. Hence it is more flexible than the parametric two stage approach. However, the performance of this approach is affected by the choice of smoothing parameter which is estimated using a grid search within preselected ranges. Hence it is important to select appropriate range of smoothing parameter values.

2.5 Semiparametric one-stage Bayesian approach

The alternative one-stage approach is developed under a Bayesian framework. Instead of using a grid search for the smoothing parameter within preselected range, we specify the prior distributions and therefore, estimate the smoothing parameter in the whole domain of the prior distribution. Hence the semiparametric one stage Bayesian approach can estimate parameters more accurately and efficiently than the one stage penalized approach.

Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2})^T$, where $\boldsymbol{\alpha_1} = (\alpha_0, \ldots, \alpha_p)^T$ and $\boldsymbol{\alpha_2} = (\alpha_{p+1}, \ldots, \alpha_{p+L})^T$. We first specify the prior distributions of parameters as the follows:

$$\begin{aligned}
\begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_p \end{pmatrix} &\sim N(\mathbf{0}, \phi_1^{-1} I), \\
\begin{pmatrix} \alpha_{p+1} \\ \vdots \\ \alpha_{p+L} \end{pmatrix} &\sim N(\mathbf{0}, \phi_\alpha^{-1} I), \\
\phi_\alpha &\sim \text{Gamma}(u, v),
\end{aligned}$$

where ϕ_1^{-1} is a fixed large number and u and v are uninformative hyperparameters of the gamma distribution.

Using these prior specifications, our joint likelihood ($\mathcal{L}$) is then expressed as

$$\begin{aligned}
\mathcal{L} &= \left[\prod_{k=1}^{N} \frac{1}{1 + \sum_{j=2}^{M+1} \exp\Big\{ \big(\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{p+l} (t - \xi_j)_+^p \big) x_{jk}^* \Big\}} \right] \\
&\times \left\{ \frac{\phi_1^p}{(\sqrt{2\pi})^p} \exp\Big(- \frac{\phi_1 \boldsymbol{\alpha_1}^T \boldsymbol{\alpha_1}}{2} \Big) \right\} \left\{ \frac{\phi_\alpha^L}{(\sqrt{2\pi})^L} \exp\Big(- \frac{\phi_\alpha \boldsymbol{\alpha_2}^T \boldsymbol{\alpha_2}}{2} \Big) \right\}
\end{aligned}$$

$$\times \quad \left\{ \frac{1}{\Gamma(u)v^u} \phi_\alpha^{u-1} \exp(-\frac{\phi_\alpha}{v}) \right\}. \tag{2.5}$$

The full conditional likelihoods for α_i, $i = 1, \ldots, p$, are

$$p(\alpha_i|-) \quad \propto \quad \frac{\phi_1 \exp\Big(-\frac{\phi_1 \alpha_i^2}{2} \Big)}{\prod_{k=1}^{N} \left[1 + \sum_{j=2}^{M+1} \exp\Big\{ (\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{p+l} (t - \xi_l)_+^p) x_{jk}^* \Big\} \right]}. \tag{2.6}$$

The full conditional likelihoods for α_{p+i}, $i = 1, \ldots, L$, are

$$p(\alpha_{p+i}|-) \quad \propto \quad \frac{\phi_\alpha \exp\Big(-\frac{\phi_\alpha \alpha_{p+i}^2}{2} \Big)}{\prod_{k=1}^{N} \left[1 + \sum_{j=2}^{M+1} \exp\Big\{ (\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{p+l} (t - \xi_l)_+^p) x_{jk}^* \Big\} \right]}. \tag{2.7}$$

The full conditional likelihoods for ϕ_α is

$$\begin{aligned} p(\phi_\alpha|-) \quad &\propto \quad \phi_\alpha^{(u+L)-1} \exp\left\{ -\phi_\alpha \Big(\sum_{l=1}^{L} \frac{\alpha_{p+l}^2}{2} + \frac{1}{v} \Big) \right\} \\ &= \quad \text{Gamma}\left\{ u + L, \Big(\sum_{l=1}^{L} \frac{\alpha_{p+l}^2}{2} + \frac{1}{v} \Big)^{-1} \right\}. \end{aligned} \tag{2.8}$$

Since the full conditional distributions of parameters except for ϕ_α do not have closed forms, we use Adaptive Rejection Metropolis Sampling (ARMS) within Gibbs sampling as proposed by Gilks et al. (1995). The detailed procedures of our ARMS are provided in the Appendix A.3. ARMS algorithm for Matched Case Crossover studies is summarized in Algorithm 1.

2.6 Test hypothesis for time varying coefficient

Our test of interest is whether the time varying coefficient is significant or not. Our null hypothesis (H_0) and alternative hypothesis (H_1) can be expressed as

$$H_o: \ \beta(t) = \beta \ \text{ vs } \ H_1: \ \beta(t) \neq \beta, \tag{2.9}$$

where β is a constant parameter.

To evaluate these hypotheses we implement Bayes Factor (Jeffreys, 1961; Kass and Raftery, 1995) defined by

$$B_{10} = \frac{P(\boldsymbol{D}|H_1)}{P(\boldsymbol{D}|H_0)},$$

where $\boldsymbol{D}$ represents the data, $P(\boldsymbol{D}|H_0)$ and $P(\boldsymbol{D}|H_1)$ are the marginal likelihoods under H_0 and H_1, respectively. The marginal likelihoods are calculated as follows

$$P(\boldsymbol{D}|H_0) \ = \ \int P(\boldsymbol{D}|\beta, H_0)p(\beta|H_0)d\beta,$$

where β is the constant parameter under H_0 and $p(\beta|H_0)$ is its prior distribution, and

$$P(\boldsymbol{D}|H_1) \ = \ \int P(\boldsymbol{D}|\beta(t), H_1)p(\beta(t)|H_1)d\beta,$$

where $\beta(t)$ is estimated by regression splines (2.3) and $p(\beta(t)|H_1)$ represents the set of prior distributions presented in section 2.5.

Large values of BF represent evidence in favor of H_1. This indicates that H_1 is more strongly supported by the data under consideration than H_0. By following Jeffreys (1961), we interpret the value of BF as weak if $BF \leq 3$, positive if $3 < BF \leq 12$, strong if $12 < BF \leq 150$ and decisive if $BF > 150$.

Difficulty lies in calculating $P(\boldsymbol{D}|H_0)$ and $P(\boldsymbol{D}|H_1)$, therefore the estimation of the Bayes Factor is obtained following procedures presented by Weinberg (2012) as follows:

Step 1: Set $Z_0 = P(\boldsymbol{D}|\mathrm{H}_0)$ and $Z_1 = P(\boldsymbol{D}|\mathrm{H}_1)$

Step 2: Set $i = 0, 1$, $n = 1, \ldots N'$ the number of MCMC samples, $L_{in} = P(\boldsymbol{D}|\theta_{in})$, θ_{1n} represents the n-th sample of $\theta_1 = \{\alpha_0, \ldots, \alpha_p, \alpha_{p+1}, \ldots, \alpha_{p+L}\}$ which is the set of parameters under the alternative hypothesis, and θ_{2n} represents the n-th sample of the $\theta_2 = \{\beta\}$ which is the set of parameters under the null hypothesis

Step 3: Calculate estimates for Z_0 and Z_1 based on the MCMC samples using

$$Z_i = \left(\frac{1}{N'} \sum_{n=1}^{N'} \frac{1}{L_{in}} \right)^{-1}$$

where

$$L_{1n} = \frac{1}{\prod_{k=1}^{N} \left[1 + \sum_{j=2}^{M+1} \exp\left\{ (\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{p+l} (t - \xi_l)_+^p) x_{jk}^* \right\} \right]}$$

$$L_{0n} = \frac{1}{\prod_{k=1}^{N} \left[1 + \sum_{j=2}^{M+1} \exp\left\{ \beta x_{jk}^* \right\} \right]}$$

2.7 Simulation

We conducted two simulation studies. One is to compare the parametric two-stage approach with the semiparametric one stage penalized approach in Section 2.7.1, while the other is to compare the semiparametric one stage penalized approach with semiparametric one stage Bayesian approach in Section 2.7.2. For each simulation study, two approaches are compared to evaluate how much each approach accurately estimates parameters and correctly predicts the probability at time $t+1$. Our semiparametric approaches are obtained from the linear regression splines model ($p = 1$).

Additionally, we conducted Bayes Factor estimations and interpretations for all the simulated datasets for the Semiparametric Bayesian alternative.

2.7.1 Comparison between the parametric two stage and semi-parametric penalized one stage approaches

Simulation setting

We consider a 1-4 matched case-crossover study which is motivated by our childhood aseptic meningitis example. We set the number of stratum as 100 and the number of time periods as 100 for each stratum, that is, $K = 100$ and $T = 100$ for each dataset.

The covariates X were generated from a $N(0, 1)$ distribution. Three different functions for $\beta(t)$ are used: (1) $\beta(t) = 0.5t$, (2)$\beta(t) = t^2$, and (3) $\beta(t) = \cos(\pi t)$. The 1-4 matched case-controls are generated from the prospective model

$$Pr(Y_k = 1|X_k, t, q) = \frac{\exp\{X_k\beta(t) + q(\bullet)\}}{1 + \exp\{X_k\beta(t) + q(\bullet)\}} \tag{2.10}$$

where $q(\bullet)$ is unknown effect including intercept; $q(\bullet)$ can be generated from any distribution. However, it is disappeared under the condition $\sum_{j=1}^{5} Y_j = 1$ for each stratum k, $k = 1, \ldots, N$ regardless of distribution of $q(\bullet)$.

For each function $\beta(t)$, we simulated 100 data sets with $N = 100$ and $T = 100$. We then apply the two approaches. We compare these two approaches in terms of mean squared error (MSE) and R^2 which is the square of correlation between $\beta(t)$ and $\hat{\beta}(t)$. This MSE is calculated as follows:

$$\text{MSE} = \text{Bias}^2 + \text{Variance}$$

where

$$\begin{aligned} \text{Bias}^2 &= \frac{1}{100}\left[\sum_{d=1}^{100}\frac{1}{100}\sum_{t=1}^{100}\{\beta(t)-\hat{\beta}_d(t)\}^2\right], \\ \text{Variance} &= \frac{1}{100}\left[\sum_{d=1}^{100}\frac{1}{100-1}\sum_{t=1}^{100}\{\hat{\beta}_d(t)-\bar{\hat{\beta}}(t)\}^2\right]. \end{aligned}$$

Here $\hat{\beta}_d(t)$ represents the estimated $\beta(t)$ using the d simulated data and

$$\bar{\hat{\beta}}(t) = \frac{1}{100}\sum_{d=1}^{100}\hat{\beta}_d(t).$$

Before conducting the complete simulation, we first performed a small simulation to choose the possible number of knots and smoothing parameters. The estimation of $\beta(t) = 0.5t$ and t^2 is not sensitive to the number of knots and smoothing parameter values because the functions are smooth enough. We selected the number of knots in the range of [25, 50, 75] which are the 25, 50 and 75% of the total number of time periods. The grid range of smoothing parameter λ is [0.1, 0.5, 1, 5, 10, 15, 30, 50, 75, 100, 500, 1000]. However, the estimations for $\beta(t) = \cos(\pi t)$ are sensitive. Hence we selected the smaller grid of the number of knot and the range of smoothing parameter around the Bayesian results. After we fixed the number of knots and location of knots using sample quantile, we then select the smoothing parameters using AIC and AICc. The number of knots and smoothing parameters for semiparametric one-stage penalized approach are summarized in Table 2.1.

$\beta(t)$	Number of Knots (L)	Smoothing Parameter (λ)
$0.5t$	25	15
t^2	25	100
$\cos(\pi t)$	5	0.04

Table 2.1: Number of knots (L) and smoothing parameter (λ) using small simulation for the semiparametric one-stage penalized approach

Evaluation of Estimation Accuracy

We performed 100 simulations under the setting described in section 2.7.1. The parametric two stage approach was performed using the conditional logistic and then quadratic regression for three different functions of $\beta(t)$. For the semiparametric one stage approach, we fixed the number knots and chose the location of knot using sample quantiles. The smoothing parameter λ was chosen using AIC and AICc.

Figure 2.1 shows the MSE, R^2, and an estimated $\hat{\beta}(t)$, when the true $\beta(t)$ is $0.5t$. Figure 2.1 (a) shows the MSE values for both methods: solid line represent one stage approach with symbol circle and dashed line with symbol $\times$ represents two stage approach. Figure 2.1 (b) shows the R^2 values for both methods. Figure 2.1 (c) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches. As shown in Figure 2.1 (a) and (b), the semiparametric one stage approach provides smaller MSE and larger R^2 values, suggesting the outperformance of the semiparametric one stage approach.

We also observe similar results using $\beta(t) = t^2$. Figure 2.2 shows the MSE, R^2, and estimated $\hat{\beta}(t)$ when true $\beta(t)$ is t^2. Figure 2.2 (a) shows the MSE values for both methods: solid line represent one stage approach with symbol circle and dashed line with symbol $\times$ represents two stage approach. Figure 2.2 (b) shows the R^2 values for both methods. Figure 2.2 (c) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches. As shown in Figure 2.2 (a) and (b), the semiparametric one stage approach provides smaller MSE and larger R^2 values, suggesting the outperformance of the semiparametric one stage approach.

For simulation using $\beta(t) = \cos(\pi t)$, Figure 2.3 presents the results obtained. Similar to the results shown in Figures 2.1 and 2.2, part (a) and part (b) shows a favorable performance of the one-stage semiparametric method. Part (c) shows an example of the fit in which both methods perform well, obtaining a smoother fit from the one-stage semiparametric method. Table 2.2 presents a summary of the results presented in Figures 2.1 - 2.3.

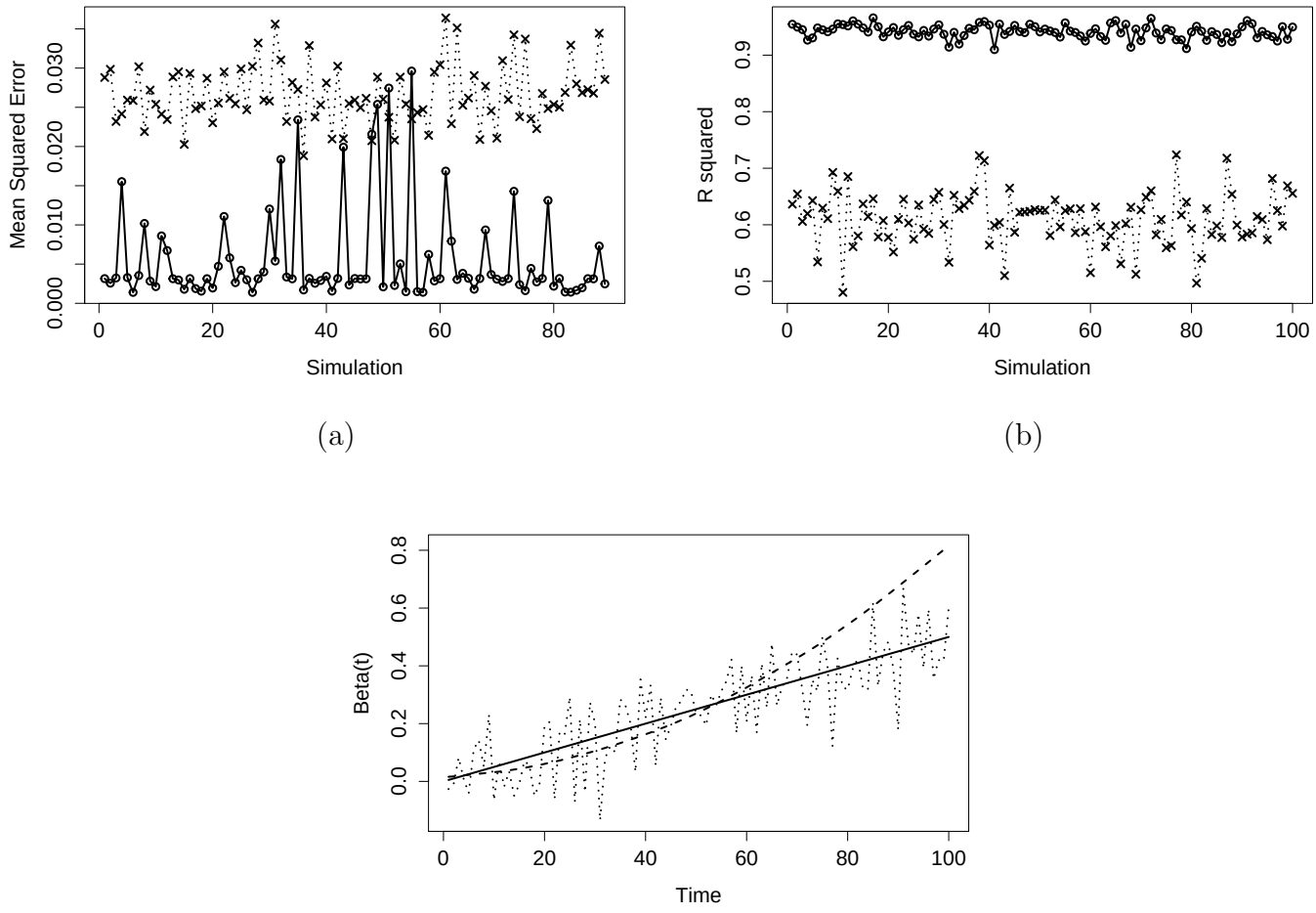


Figure 2.1: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = 0.5t$; solid line represent one stage approach with symbol circle and dashed line with symbol $\times$ represents two stage approach; (a) shows the MSE values for both methods; (b) shows the R^2 values for both methods; (c) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches.

Evaluation of Prediction Accuracy

We also compare the two approaches in terms of prediction accuracy. We forecast $P(Y_{1,t+1} = 1|X_{1,t}, ..., X_{M+1,t}, t+1, \sum_{j=1}^{M+1} Y_{j,t+1} = 1)$ as a measure of prediction accuracy. We consider the

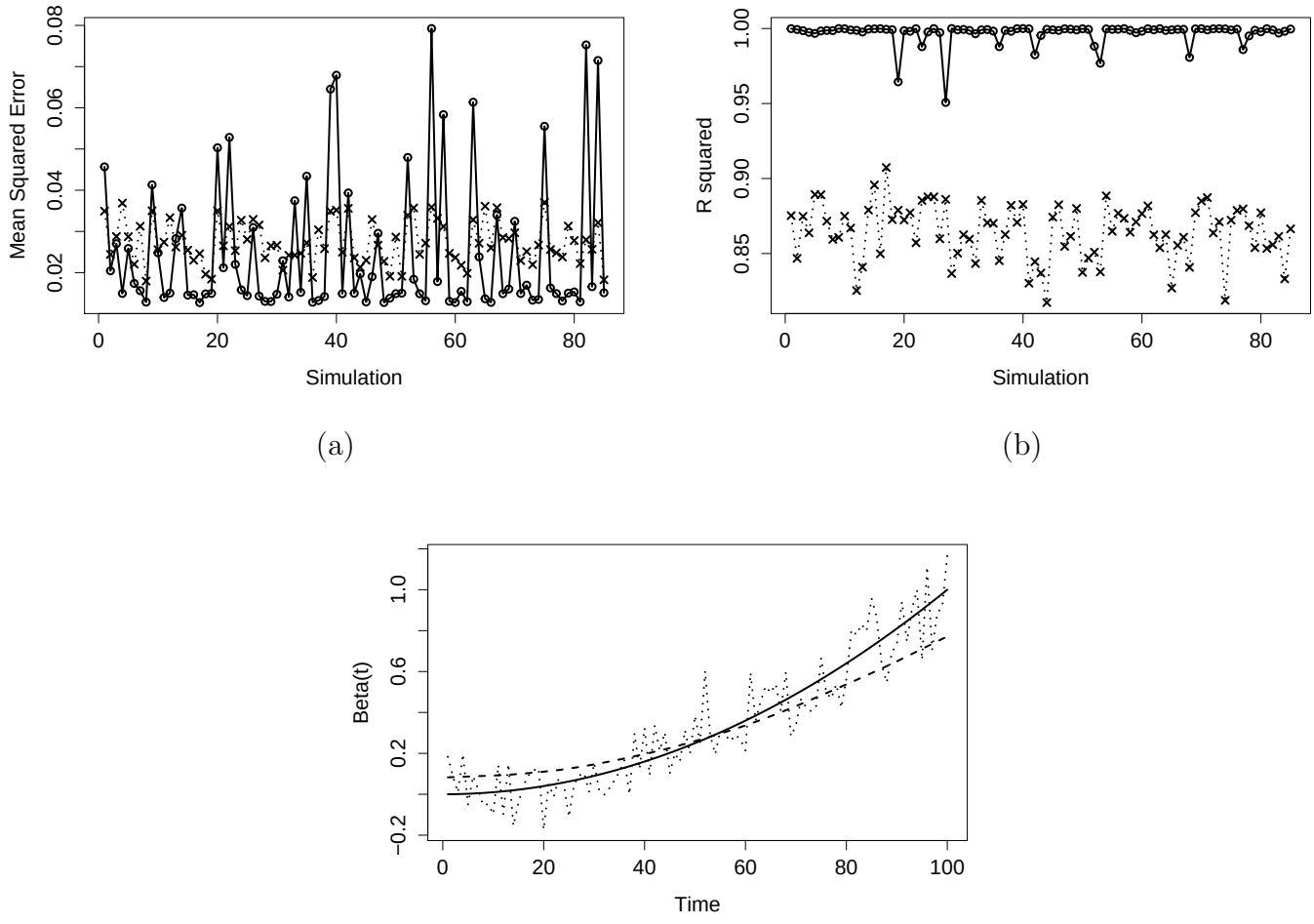


Figure 2.2: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = t^2$; solid line represent one stage approach with symbol circle and dashed line with symbol $\times$ represents two stage approach; (a) shows the MSE values for both methods; (b) shows the R^2 values for both methods; (c) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches.

true forecast probability for a particular stratum k as follows:

$$\Pr(Y_{1k,t+1} = 1 | X_{1k,t}, ..., X_{M+1k,t}, \sum_{j=1}^{M+1} Y_{jk,t+1} = 1) = \frac{1}{1 + \sum_{j=2}^{M+1} \exp\{x^*_{jk,t}\beta(t+1)\}}, \qquad (2.11)$$

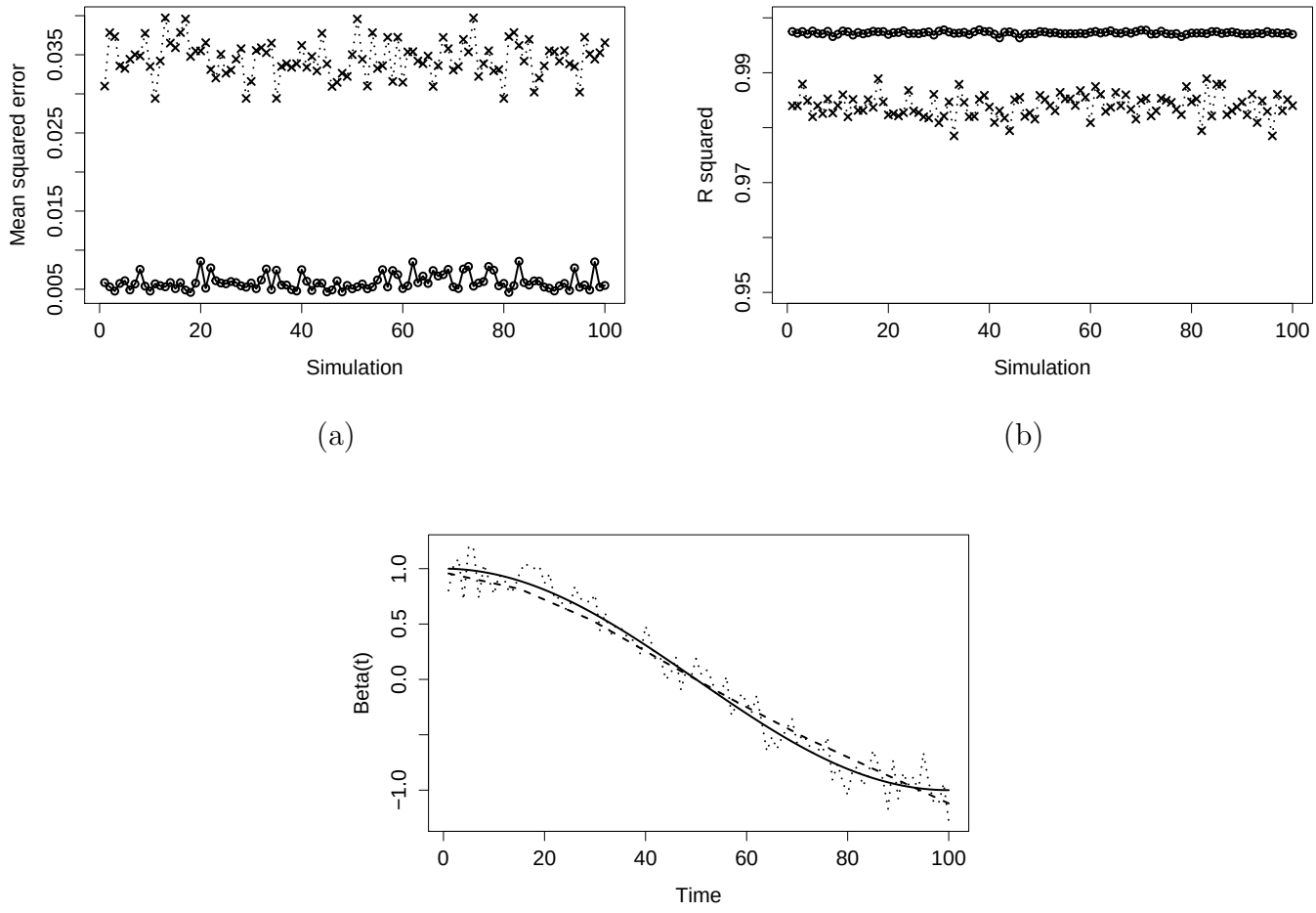


Figure 2.3: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = \cos(\pi t)$; solid line represent one stage approach with symbol circle and dashed line with symbol $\times$ represents two stage approach; (a) shows the MSE values for both methods; (b) shows the R^2 values for both methods; (c) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches.

where $\beta(t+1)$ represents the function $\beta(t)$ evaluated at $t+1$.

Two stage approach computes the forecast probability as

$$\Pr(Y_{1k,t+1} = 1 | X_{1k,t}, ..., X_{M+1k,t}, \sum_{j=1}^{M+1} Y_{jk,t+1} = 1) = \frac{1}{1 + \sum_{j=2}^{M+1} \exp\{x^*_{jk,t} \hat{\beta}(t+1)\}},$$

$\beta(t)$	Method	MSE	R^2
$0.5t$	Semiparametric One Stage	0.0097	0.9890
	Parametric Two Stage	0.0272	0.6670
t^2	Semiparametric One Stage	0.0298	0.9970
	Parametric Two Stage	0.0314	0.8670
$\cos(\pi t)$	Semiparametric One Stage	0.0059	0.9972
	Parametric Two Stage	0.0345	0.9839

Table 2.2: Average MSE and average R^2 for the semiparametric one-stage penalized method and the parametric two-stage method using 100 simulated data sets

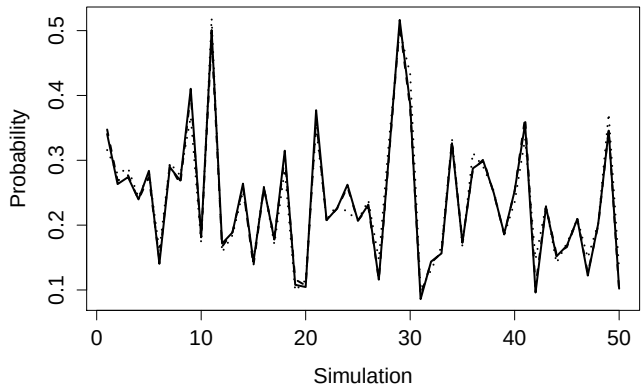
where $\hat{\beta}(t)$ is estimated using polynomial regression, while the one stage approach calculates

$$\Pr(Y_{1,t+1}=1|X_{1,t},...,X_{M+1,t},\sum_{j=1}^{M+1}Y_{jk,t+1}=1)=\frac{1}{1+\sum_{j=2}^{M+1}\exp\{(\hat{\alpha}_0+\hat{\alpha}_1(t+1)+\sum_{l=1}^{L}\hat{\alpha}_{1+l}[(t+1)-\xi_l]_+)x^*_{j,t}\}}. \quad (2.12)$$

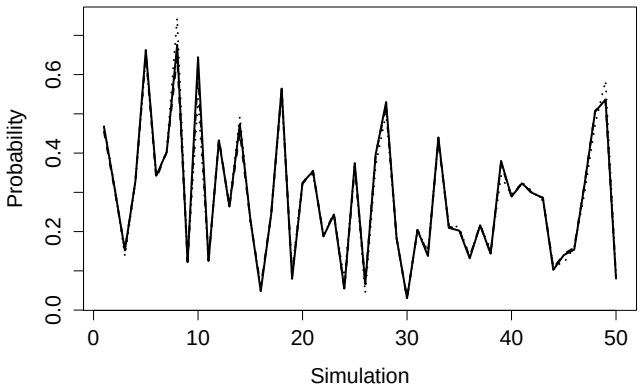
Figure 2.4 presents the results of forecasted probabilities using both approaches for 50 simulation data sets. Figure 2.4 (a), (b), and (c) show the forecasted probabilities for $\beta(t)=0.5t$, $\beta(t)=t^2$, and $\beta(t)=\cos(\pi t)$, respectively. The solid line, the dotted line, the dashed lines represent the true probability, forecasted probability using the two-stage parametric method, forecasted probability using the one-stage semiparametric method.

We summarize the results of forecasted probabilities using both approaches for 50 simulation data sets. We provide root mean squared prediction bias (RMSPB) and mean absolute prediction bias (MAPB) given by

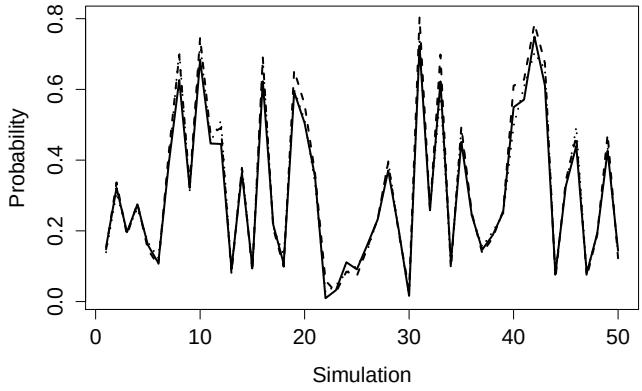
$$\text{RMSPB}=\sqrt{\frac{1}{50}\sum_{d=1}^{50}(\hat{p}_{t+1}-p_{t+1})^2}$$



(a)



(b)



(c)

Figure 2.4: Forecasted probabilities using one and two stages approaches for fifty simulated data sets; (a), (b), and (c) showed the forecasted probabilities for $\beta(t) = 0.5t$, $\beta(t) = t^2$, and $\beta(t) = \cos(\pi t)$ respectively; The solid, the dotted, the dashed lines represent the true probability, forecasted probability using the two-stage parametric method, forecasted probability using the one-stage semiparametric method.

and

$$\text{MAPB} = \frac{1}{50}\sum_{d=1}^{50} |\hat{p}_{t+1} - p_{t+1}|$$

where $p_{t+1} = \Pr(Y_{1k,t+1} = 1 | X_{1k,t}, ..., X_{M+1k,t}, \sum_{j=1}^{M+1} Y_{jk,t+1} = 1)$ and $\hat{p}_{t+1}$ is the estimated forecast probability.

Table 2.5 shows our semiparametric one stage approach provides better prediction accuracy than two stage approach.

2.7.2 Comparison between semiparametric one stage penalized and Bayesian approaches

Evaluation of Estimation Accuracy

We performed a simulation study to compare two approaches: the penalized and Bayesian one stage approaches. The simulation study was conducted under the simulation settings described in Section 2.7.1. Using a small simulation, we first chose the number of knots for the three different $\beta(t)$ functions. This result is summarized in Table 2.3. We then conducted 50 simulations and calculated MSE and R^2.

$\beta(t)$	Number of Knots (L)
$0.5t$	5
t^2	10
$\cos(\pi t)$	5

Table 2.3: Number of knots (L) using small simulation for the semiparametric one-stage Bayesian approach

In Table 2.4 we present the numerical results for the MSE and the smoothing parameters. It can be seen that Bayesian approach has smaller MSE so that it performed better than the penalized approach in terms of MSE.

We also summarize the MSE results in Figures 2.10-2.7 for $\beta(t) = 0.5t$, $\beta(t) = t^2$, and $\beta(t) = \cos(\pi t)$, respectively. The dashed line represents Bayesian approach and the dotted

$\beta(t)$	Semiparametric One stage	Smoothing parameter	L	MSE
$0.5t$	Penalized	$\lambda = 15$	25	0.0110
	Bayesian	$\phi_\alpha = 0.05$	5	0.0003
t^2	Penalized	$\lambda = 100$	25	0.0143
	Bayesian	$\phi_\alpha = 0.097$	10	0.0038
$\cos(\pi t)$	Penalized	$\lambda = 0.04$	5	0.0034
	Bayesian	$\phi_\alpha = 0.048$	5	0.0002

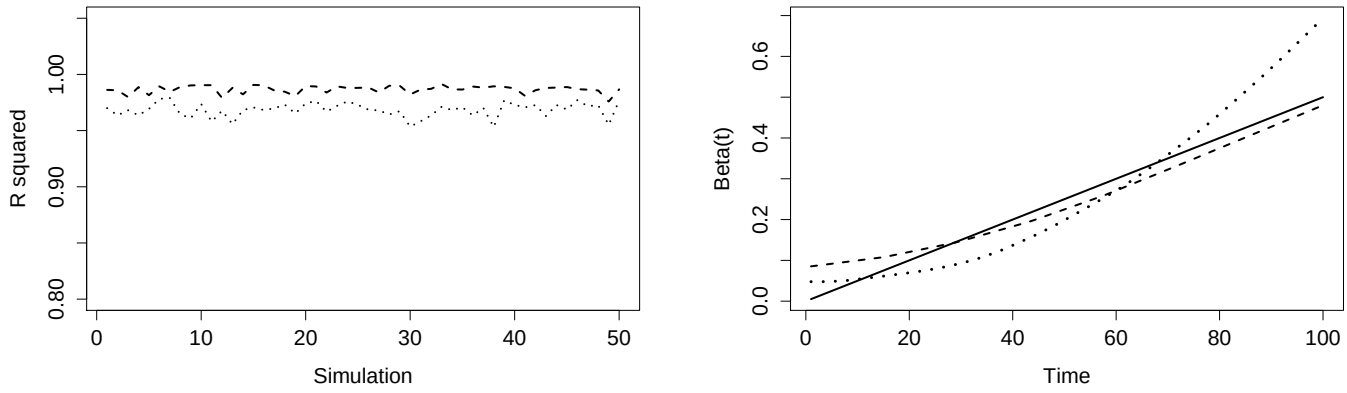
Table 2.4: Average MSE of semiparametric one stage penalized and Bayesian approaches with smoothing parameter and the number of knots (L) using 50 simulated data sets

line represents penalized approach. Figure 2.10 (a) shows the R^2 values of semiparametric one stage Bayesian and penalized approaches: the dashed line represents Bayesian approach and the dotted line represents penalized approach. Figure 2.10 (b) showed the estimated function $\hat{\beta}(t)$ when true $\beta(t) = 0.5t$. Similar figures are displayed in Figure 2.6 (a) and (b) when true $\beta(t) = t^2$ and in Figure 2.7 for $\beta(t) = \cos(\pi t)$. Overall, semiparametric one stage Bayesian approach performs better than semiparametric one stage penalized approach in terms of MSE and R^2.

Evaluation of Prediction Accuracy

As we described in Section 2.7.1, we forecast this probability $\Pr(Y_{1,t+1}|X_{1,t}, \ldots X_{M+1,t}, t+1, \sum_{j=1}^{M+1} Y_{jk,t+1} = 1)$ to evaluate prediction accuracy of two approaches. The true probability is the same as before used in equation (2.11), and the one-stage semiparametric method probability for both the penalized and Bayesian approaches are also calculated as before shown in equation (2.12).

Figure 2.8 presents the results of forecast probabilities using penalized and Bayesian approaches for fifty simulated data sets. Figure 2.8 (a), (b), and (c) showed the forecast probabilities for $\beta(t) = 0.5t$, $\beta(t) = t^2$, and $\beta(t) = \cos(\pi t)$, respectively. The solid, the



(a) (b)

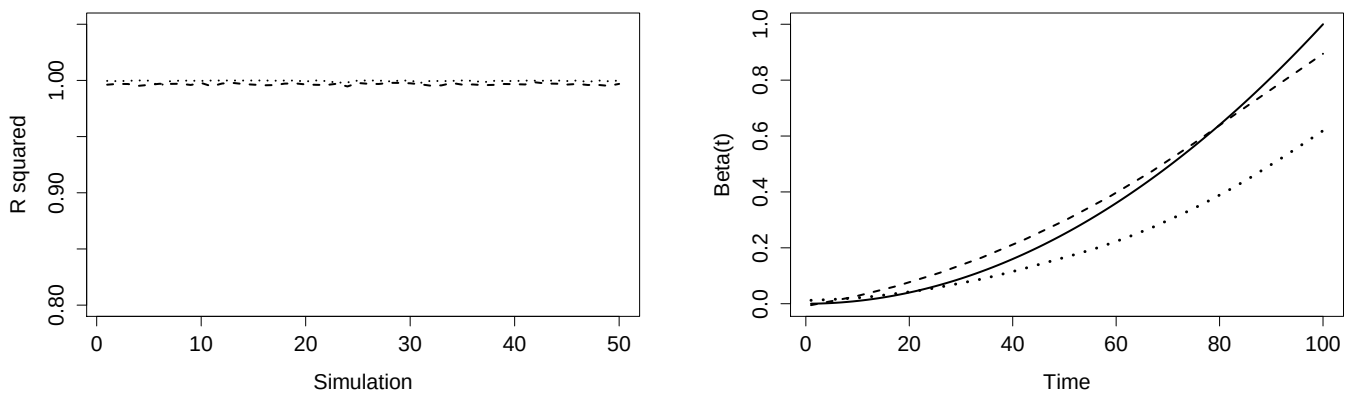
Figure 2.5: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = 0.5t$; dashed line represent one stage Bayesian approach and dotted line represents one stage penalized approach; (a) shows the R^2 values for both methods; (b) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches and true function which is solid line

dotted, the dashed lines represent the true probability, forecasted probability using the one stage penalized method, forecasted probability using the one-stage Bayesian method. Figure 2.8 (a)-(c) show a competitive performance of both methods which are close to the true probability values. This result is summarized in Table 2.5 which shows our two semiparametric one stage approaches are comparable and both better prediction accuracy than two stage approach.

Bayesian Inference based on Bayes Factor

We simulated fifty datasets under our three different $\beta(t)$ functions. For each of these functions we followed the procedure:

1. Calculate the Bayes Factor for each of the 100 strata as outlined in Section 2.6.



(a) (b)

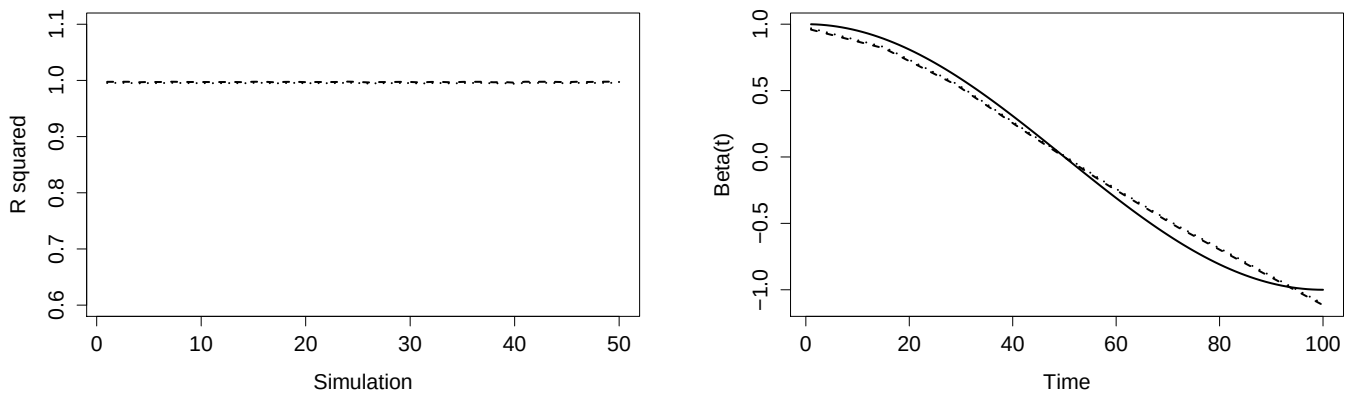
Figure 2.6: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = t^2$; dashed line represent one stage Bayesian approach and dotted line represents one stage penalized approach; (a) shows the R^2 values for both methods; (b) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches and true function which is solid line

$\beta(t)$	Prediction Bias	Parametric	Semiparametric Penalized	Semiparametric Bayesian
$0.5t$	RMSPB	0.0194	0.0053	0.0058
	MAPB	0.0147	0.0039	0.0042
t^2	RMSPB	0.0135	0.0088	0.0092
	MAPB	0.0135	0.0051	0.0052
$cos(\pi t)$	RMSPB	0.0271	0.0035	0.0033
	MAPB	0.0186	0.0026	0.0025

Table 2.5: Summary of root mean squared prediction bias (RMSPB) and mean absolute prediction bias (MAPB) of three approaches using 50 simulated data sets: RMSPB $= \sqrt{\frac{1}{50}\sum_{d=1}^{50}(\hat{p}_{t+1} - p_{t+1})^2}$ and MAPB $= \frac{1}{50}\sum_{d=1}^{50}|\hat{p}_{t+1} - p_{t+1}|$; $p_{t+1} = \Pr(Y_{1k,t+1} = 1|X_{1k,t}, ..., X_{M+1k,t}, \sum_{j=1}^{M+1} Y_{jk,t+1} = 1)$ and $\hat{p}_{t+1}$ is the estimated forecast probability.

2. The Bayes Factor for the dataset will be the average of the 100 calculated Bayes Factors.

3. Assess the average Bayes Factor with Jeffreys (1961) suggestion.



(a) (b)

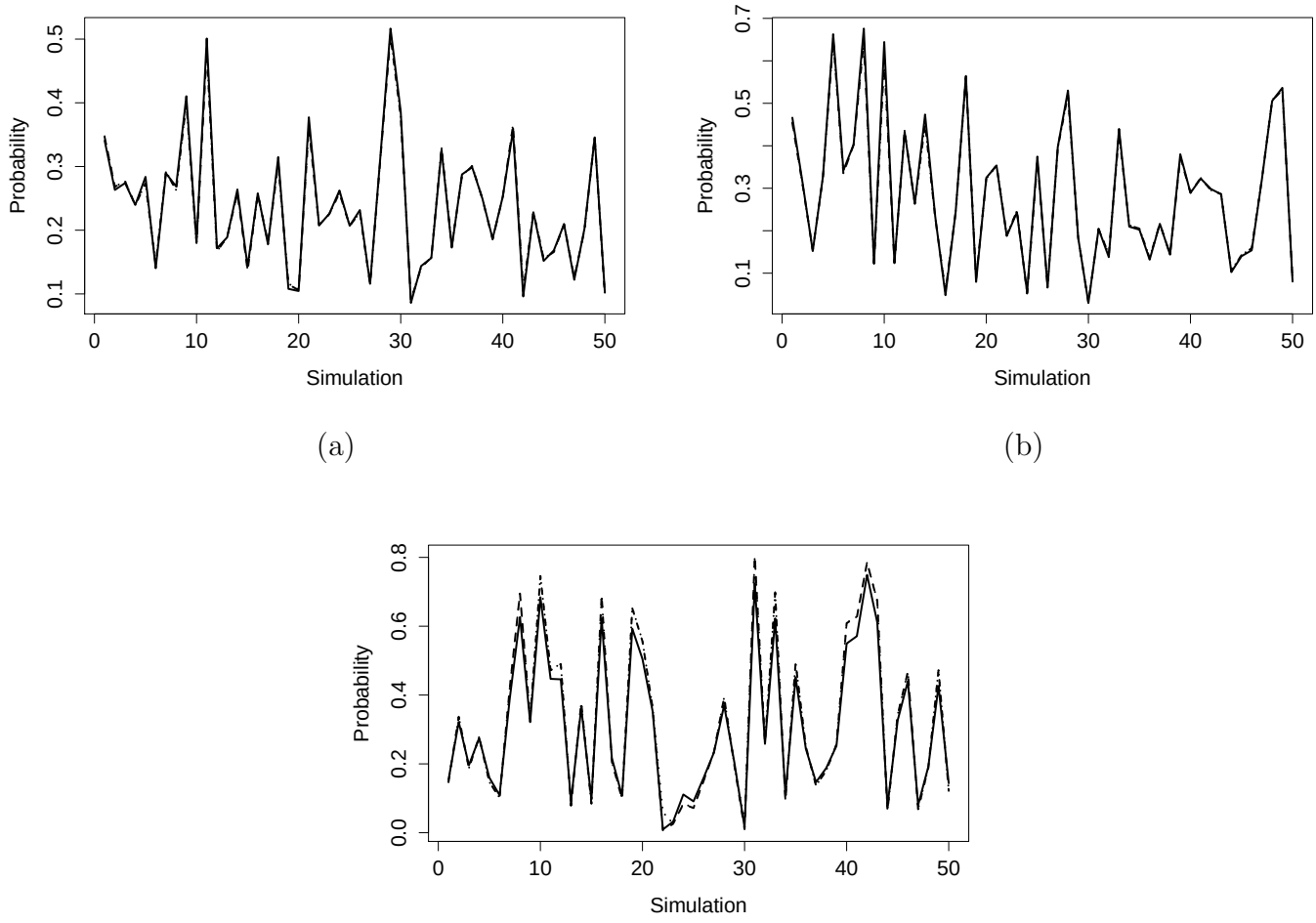
Figure 2.7: Comparison between parametric two stage and semiparametric one stage approaches for Estimation Accuracy using true function $\beta(t) = \cos(\pi t)$; dashed line represent one stage Bayesian approach and dotted line represents one stage penalized approach; (a) shows the R^2 values for both methods; (b) shows the estimated function, $\hat{\beta}(t)$ obtained from both approaches and true function which is solid line.

4. The probability of rejection of H_0 is calculated by counting how many times among the 100 strata the BF was rejected.

The estimated probability is summarized in Table 2.6. This result suggests that semiparametric one-stage Bayesian approach has a high power to make the correction decision.

Function under H_1	Probability of Rejection of H_0
$\beta(t) = 0.5t$	96%
$\beta(t) = t^2$	100%
$\beta(t) = \cos(\pi t)$	100%

Table 2.6: The estimated power of semiparametric one stage Bayesian approach using Bayes Factor



(a)

(b)

(c)

Figure 2.8: Forecasted probabilities using two semiparametric one stage approaches for fifty simulated data sets; (a), (b), and (c) showed the forecasted probabilities for $\beta(t) = 0.5t$, $\beta(t) = t^2$, and $\beta(t) = \cos(\pi t)$ respectively; The solid, dotted, the dashed lines represent the true probability, forecasted probability using the one stage penalized method, forecasted probability using the semiparametric one-stage Bayesian method.

2.8 Example: $1-4$ bi-directional matched case-crossover study

As we mentioned in Section 2.1, the principal interest of our $1-4$ bidirectional case-crossover example is to understand the relationship between the effect of drinking water turbidity and the risk of aseptic meningitis in children as well as to explore the importance of potential effect modification on the turbidity effect by time. The analysis of a case-crossover study is a stratified data analysis of retrospective, self-matched follow-up studies, each with a sample size of one. In this study, the stratifying variable is the individual child who is less than 15 years old, diagnosed in one of two years and lives on one of two urban communities in South Korea. Our potential effect modification is time effect and the covariate of interest is water turbidity.

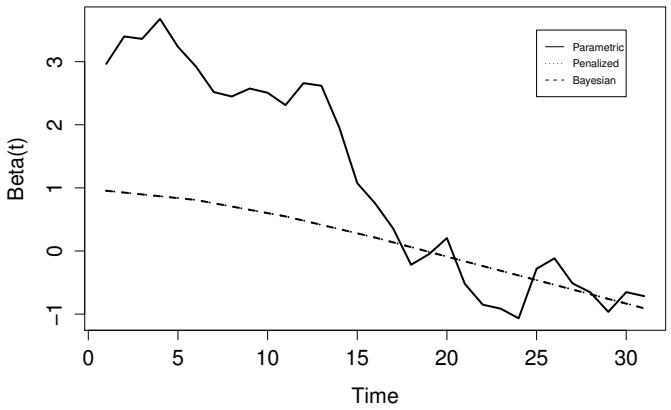


Figure 2.9: Estimated $\hat{\beta}(t)$ using aseptic meningitis data: $\hat{\beta}(t)$ are obtained from the parametric two-stage method (solid line), the semiparametric one-stage method (dotted line) and the semiparametric one-stage Bayesian method (dashed line).

We estimated the varying coefficients model using three methods: a two-stage parametric approach and two semiparametric one-stage approaches. The parametric two stage approach

is obtained using quadratic regression in the second Step. The semiparametric one stage penalized approach is performed using $L = 16$ and $\lambda = 5$. The semiparametric one stage Bayesian approach is also performed using $L = 16$ knots. The estimated function $\hat{\beta}(t)$s are displayed in Figure 2.9, where the solid line represents the parametric two-stage method, the dashed line represents the semiparametric one-stage Bayesian method and the dotted line represents the semiparametric one-stage method.

We also predict the forecast probability using the following procedures;

Step 1: For each time value do the following:

Step 2: Delete time t and name it t^*.

Step 2.1: Estimate $\hat{\beta}(t)$ using both approaches using all the data except for t^*.

Step 2.2: Estimate $\hat{\beta}(t^*)$ with the correct approach:

- Parametric Approach: Using the polynomial regression coefficients (in this case quadratic regression $\hat{\theta}_0,\ \hat{\theta}_1,\ \hat{\theta}_2$):

$$\hat{\beta}(t^*) \quad = \quad \hat{\theta}_0 + \hat{\theta}_1 t^* + \hat{\theta}_2 {t^*}^2$$

- Semiparametric Approaches: Using the estimated coefficients for the regression splines:

$$\hat{\beta}(t^*) \quad = \quad \hat{\alpha_0} + \hat{\alpha_1} t^* + \sum_{l=1}^{L} \hat{\alpha_{1+l}} (t^* - \xi_l)_+$$

Step 2.3: For t^* calculate $Pr(Y_{1k,t^*} = 1 | X_{1k,t^*}, \ldots, X_{M+1k,t^*}, \sum_{j=1}^{M+1} Y_{jk,t^*} = 1)$ for

each stratum;

$$Pr(Y_{1k,t^*} = 1 | X_{1k,t^*}, \ldots, X_{M+1k,t^*}, \sum_{j=1}^{M+1} Y_{jk,t^*} = 1) = \frac{1}{1 + \sum_{j=2}^{M+1} \exp\{x^*_{jk,t^*} \hat{\beta}(t^*)\}}$$

Step 2.4: Calculate the success probability for time t^*, which is the number of forecasted probabilities greater than 0.2 (threshold) divided by the total number of strata for both methods; we use 0.2 because 1 out of 5 is the probability of case in a 1-4 matched case-control study;

Step 3: The success probability for each method is the average of the success probabilities of all considered time periods.

The estimated success probabilities of three approaches are summarized in Table 2.7. Showing the outperformance of the semiparametric time varying coefficient models. Semiparametric one-Stage Penalized and Bayesian approaches provide 0.9275 and 0.9354, while parametric Two-Stage approach provides 0.7097.

Method	Predicted Success Probability
Parametric Two-Stage	0.7097
Semiparametric One-Stage Penalized	0.9275
Semiparametric One-Stage Bayesian	0.9354

Table 2.7: Predicted success probabilities of aseptic meningitis data

Figure 2.10 presents the forecasted probabilities for each of the forecasted times for both semiparametric methods, the dashed line with symbol $\times$ represents the Bayesian alternative and the dotted line with symbol $\oplus$ represents the penalized alternative showing very similar results for both methods. This Figure 2.10 suggests that the peak risk of aseptic meningitis was 4th day after the admission. The risk increases in an interval [1,4], decreases in an interval [5,15] and disappears after day 15.

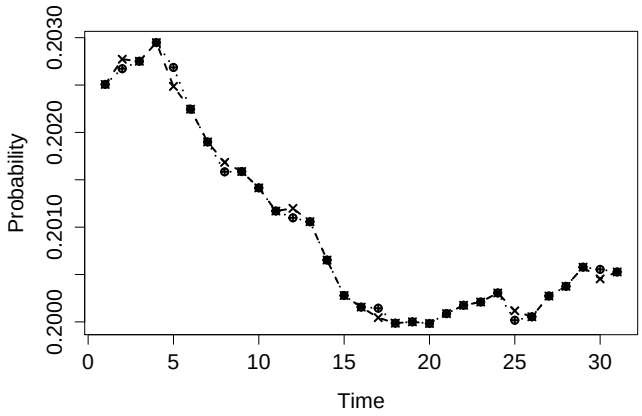


Figure 2.10: Forecasted probabilities for each of the forecasted times for both semiparametric methods, the dashed line with symbol $\times$ represents the Bayesian alternative and the dotted line with symbol $\oplus$ represents the penalized alternative.

In addition, we also perform a test of whether the time varying coefficient is significant

$$\mathrm{H}_0 : \beta(\mathrm{t}) = \beta \text{ vs } \mathrm{H}_1 : \beta(\mathrm{t}) \neq \beta$$

with the procedure outlined in section 2.6. The estimated Bayes Factor is 5.3808 which corresponds to a Positive evidence against Ho, concluding that $\beta(t)$ depends on time. Therefore, the effect of drinking water turbidity on aseptic meningitis in children, changes over time.

2.9 Discussion

In this chapter, we have proposed a parametric two stage approach and two semiparametric one stage approaches. They allow us to detect relationships between a predictor X and the relative risk of disease or binary outcome, as well as determine effect modifications by time. The parametric alternative is a two-stage method in which the relationships of interest are estimated for particular time periods and time dependencies are posteriorly estimated by

a polynomial regression. The semiparametric alternative is a semiparametric time varying coefficient model where the time dependencies are model by regression splines in either a penalized likelihood or a Bayesian way.

Simulation results show that the semiparametric one stage penalized one has performed better than parametric two stage approach and is also comparable with semiparametric one stage Bayesian approach. The performance of semiparametric one stage penalized approach is highly affected by the choice of smoothing parameter. On the other hand, semiparametric one stage Bayesian approach automatically estimates this smoothing parameter under Bayesian framework. Hence, the semiparametric one stage approach is more efficient and flexible. Because of this, the semiparametric one stage Bayesian approach performs better than its penalized counterpart in terms of MSE, R^2, and efficiency of choosing smoothing parameter. Our example also shows the prediction improvement using the Bayesian approach.

Simulation results show that the semiparametric one stage penalized method has performed better than parametric two stage approach and is also comparable with semiparametric one stage Bayesian approach. The performance of the semiparametric one stage penalized approach is highly affected by the choice of smoothing parameter. On the other hand, the semiparametric one stage Bayesian approach automatically estimates this smoothing parameter under Bayesian framework. Hence, the semiparametric one stage approach is more efficient and flexible.

The semiparametric one stage Bayesian approach performs better than its penalized counterpart in terms of MSE, R^2, and efficiency of choosing smoothing parameter. Our example also shows the prediction improvement using the Bayesian approach. Additionally, the Bayesian approach allows to test hypothesis of interest via Bayes Factors, providing more advantages over the other alternatives.

Chapter 2. Semiparametric Time Varying Coefficient Model for Matched Case-Crossover Studies

Since our varying coefficient model has been developed to detect the parametric relationship between the predictor and binary outcomes as well as simultaneously evaluate nonparametric relationships between the predictor and time. We will further develop our varying coefficients model to allow both nonparametric relationship between the predictor and binary outcomes and between the predictor and time. Furthermore we will also incorporate spatial dependence into our time varying coefficient models because our data is also obtained from different spatial locations.

Chapter 3

Semiparametric Spatial-Temporal Varying Coefficient Model for Matched Case-Crossover Studies

3.1 Background

The methods for assessing the functional relationship between time varying coefficients, spatial dependencies and matched binary outcomes are extremely limited. In order to solve this problem we propose two semiparametric spatial-temporal varying coefficient model alternatives in order to evaluate the effect modification and make correct statistical inference. The choice between these two proposed alternatives will be determined by the number of geographical locations among the subjects. When the number of locations is small, around two or three it is recommended to use alternative 1, a semiparametric spatial-temporal varying coefficient model for a small number of locations. When the number of locations is medium to large, then it is recommended to use the semiparametric spatial-temporal varying coefficient

model.

In both methods we propose a semiparametric approach using regression splines to estimate the time varying coefficient portion. When the number of locations is small, we evaluate the effect modification due to spatial location by estimating the time varying coefficient model presented in Chapter 2 for each individual region. In addition we propose the examination of the relationship between the estimated function and its first derivative with respect to time in order to assess effect modification due to time and spatial location among the subjects.

When the number of locations is medium to large we propose a random effect to assess the spatial effect with the use of the empirical semivariogram. Our approach allows us to detect the parametric relationship between the predictor and binary outcomes and evaluate parametric or semiparametric relationships between the predictor and time as well as effect modification due to geographical location.

The rest of this chapter is organized in two main sections: In section 3.2, we propose the semiparametric spatial-temporal varying coefficient model for a small number of locations. In section 3.3, we propose the semiparametric spatial-temporal varying coefficient model. Each section will contain all the elements of each method along with a discussion.

3.2 Semiparametric Spatial-Temporal Varying Coefficient Model for a Small Number of Locations

3.2.1 Background

The use of conditional logistic regression in the analysis of matched case-crossover studies does not support the estimation of any effect modification due to time or spatial location.

Chapter 3. Semiparametric Spatial-Temporal Varying Coefficient Model for Matched Case-Crossover Studies

The main interest of this section is to determine whether an effect modification due to time or spatial location exist for a small number of location among the subjects.

The application example, 1-4 bi-directional matched case-crossover study, in which the interest lies in assessing the relationship between drinking water turbidity and the presence or absence of aseptic meningitis in children, gathers data from two provinces in South Korea. In this section, he goal is to assess whether the effect of drinking water turbidity over the presence or absence of the disease changes between regions, and furthermore if there is an effect modification due to time.

The methods for assessing the functional relationship between time varying coefficients, spatial dependencies and matched binary outcomes are limited. To address this problem we propose a semiparametric spatial-temporal varying coefficient model for a small number of locations among the subjects. This model can be used to evaluate effect modification by time and spatial location so that we can make correct statistical inference. This model is designed to be implemented when the number of spatial locations is small. We estimate a spatial-temporal varying coefficient model and its derivative model using regression splines. It is important to estimate these two models simultaneously. We estimate them under the Bayesian hierarchical model framework. The estimated function and its derivative function is then studied to determine differences in effect modification due to time and spatial location.

The remainder of this section is organized as follows: In section 3.2.2, we propose our approach, the spatial-temporal varying coefficient model for small number of locations. In section 3.2.3, we describe our Bayesian hierarchical model framework and explain how to estimate the model and its derivative function. In section 3.2.4 we present testing procedures for the existence of features. In section 3.2.5, we present the simulation study and results. Section 3.2.6 presents the estimation of the 1-4 bidirectional matched case-crossover study. Section 3.2.7 contains concluding remarks.

3.2.2 Spatial-Temporal Varying Coefficient Model for a Small Number of Spatial Locations

We consider a $1-M$ matched case-crossover study with k-th stratum, $k = 1, ..., N$. We consider the r-th region, $r = 1, 2, ..., R$. Let $x_{1k,r}$ represent the value of the covariate of interest for the case of stratum k in region r, and $x_{2k,r}$, $x_{3k,r}, \ldots,$ $x_{M+1,k,r}$ represent the covariate values for the controls of stratum k in region r.

The spatial-temporal varying coefficient model replaces the parameter β in (1.1) by the unknown function $\beta_r(t)$ for given r. We model $\beta_r(t)$ semiparametrically so that we estimate $\beta_r(t)$ via $p(>1)$-th order regression splines. We define pth order regression spline with a truncated power series basis and with L knots, $\{1, t, t^2, \ldots, t^p, (t-\xi_1)_+^p, (t-\xi_2)_+^p, \ldots, (t-\xi_L)_+^p\}$, where $(u)_+^p = u^p I(u \geq 0)$ and knots $\xi_1 < \ldots < \xi_L$. Knots can be selected a priori, for example, at sample quantiles of observed t's which are scaled in [0,1]. Let $\alpha_{0,r}, \alpha_{1,r}, \ldots, \alpha_{p,r}$ be the regression coefficients and $\alpha_{p+1,r}, \ldots, \alpha_{p+L,r}$ be the regression coefficients of the truncated power series basis. With this setup, $\beta_r(t)$ can be written as

$$\beta_r(t) = \alpha_{0,r} + \alpha_{1,r}t + \cdots + \alpha_{p,r}t^p + \sum_{l=1}^{L} \alpha_{p+l,r}(t-\xi_l)_+^p.$$

Since p is larger than 1, we can take the first derivative of $\beta_r(t)$ with respective t so that we can obtain its derivative function,

$$\frac{d}{dt}\beta_r(t) \equiv \beta_r'(t) \quad = \quad \alpha_{1,r} + 2\alpha_{2,r}t + \cdots + p\alpha_{p,r}t^{p-1} + p\sum_{l=1}^{L} \alpha_{p+l,r}(t-\xi_l)_+^{p-1}.$$

Estimation of these two functions simultaneously are of our main fundamental interest. We estimate them under the Bayesian hierarchical framework. We note that Frequentist approach (Jarrow et al., 2004), using penalized splines usually estimated the function first

and then estimated the derivative of the estimated function. This two step procedure caused some problems because it used the same smoothing parameter in both functions, although the smoothing function of the derivative may be different from that of the function. The derivative estimates are typically more noisy than estimate of function. However, the Bayesian approach can estimate these two functions simultaneously using hierarchical framework.

3.2.3 Bayesian Hierarchical Model Framework

Let $\boldsymbol{\alpha_{1,r}} = (\alpha_{0,r}, \ldots, \alpha_{p,r})^T$, $\boldsymbol{\alpha_{2,r}} = (\alpha_{p+1,r}, \ldots, \alpha_{p+L,r})^T$, and $\boldsymbol{\alpha_r} = (\boldsymbol{\alpha_{1,r}}, \boldsymbol{\alpha_{2,r}})^T$, with $r = 1, ..., R$. We first specify the prior distributions of parameters as follows:

$$\begin{aligned}
\boldsymbol{\alpha_{1,r}} &\sim N(\mathbf{0}, \phi_1^{-1} I), \\
\boldsymbol{\alpha_{2,r}} &\sim N(\mathbf{0}, (\phi_{\alpha,r})^{-1} I), \\
\phi_{\alpha,r} &\sim \text{Gamma}(u_\alpha, v_\alpha), \text{ for all } r = 1, ..., R,
\end{aligned}$$

where ϕ_1^{-1} is a fixed large number, u_α and v_α, are the hyperparameters of the Gamma distribution.

The joint likelihood for each region ($\mathcal{L}_r$) is expressed as:

$$\begin{aligned}
\mathcal{L}_r \propto & \left[\prod_{k=1}^{N} \frac{1}{1 + \sum_{j=2}^{M+1} \exp\Big\{ (\alpha_{0,r} + \alpha_{1,r} t + \ldots + (\alpha_{p,r}) t^p + \sum_{l=1}^{L} \alpha_{p+l,r} (t - \xi_j)_+^p)(x_{jk,r})^* \right] \\
\times & \left\{ \exp\Big(- \frac{\phi_1 \boldsymbol{\alpha_{1,r}}^T \boldsymbol{\alpha_{1,r}}}{2} \Big) \right\} \left\{ (\phi_{\alpha,r})^L \exp\Big(- \frac{\phi_{\alpha,r} \boldsymbol{\alpha_{2,r}}^T \boldsymbol{\alpha_{2,r}}}{2} \Big) \right\} \left\{ (\phi_{\alpha,r})^{u_\alpha - 1} \exp\Big(- \frac{\phi_{\alpha,r}}{v_\alpha} \Big) \right\},
\end{aligned}$$

where $(x_{jk,r})^* = x_{jk,r} - x_{1,r}$.

The full conditional likelihoods for $\alpha_{q,r}$, $q = 1, \ldots, p$, are

$$p(\alpha_{q,r}|-) \propto \frac{\phi_1 \exp\Big(- \frac{\phi_1 (\alpha_{q,r})^2}{2} \Big)}{\prod_{k=1}^{M+1} \Big[1 + \sum_{j=2}^{M+1} \exp\Big\{ (\alpha_{0,r} + \alpha_{1,r} t + \ldots + (\alpha_{p,r}) t^p + \sum_{l=1}^{L} \alpha_{p+l,r} (t - \xi_j)_+^p \Big) (x_{jk,r})^* \Big]}.$$

The full conditional likelihoods for $\alpha_{p+q,r}$, $q = 1, \ldots, L$, are

$$p(\alpha_{p+q,r}|-) \propto \frac{\phi_{\alpha,r} \exp\Big(- \frac{\phi_{\alpha,r} (\alpha_{p+q,r})^2}{2} \Big)}{\prod_{k=1}^{M+1} \Big[1 + \sum_{j=2}^{M+1} \exp\Big\{ (\alpha_{0,r} + \alpha_{1,r} t + \ldots + \alpha_{p,r} t^p + \sum_{l=1}^{L} \alpha_{p+l,r} (t - \xi_j)_+^p \Big) (x_{jk,r})^* \Big]}.$$

The full conditional likelihood for the smoothing parameters $\phi_{\alpha,r}$ is

$$\begin{aligned} p(\phi_{\alpha,r}|-) &\propto (\phi_{\alpha,r})^{(u_\alpha + L) - 1} \exp\Big\{ - \phi_{\alpha,r} \Big(\sum_{l=1}^{L} \frac{(\alpha_{p+l,r})^2}{2} + \frac{1}{v_\alpha} \Big) \Big\} \\ &= \text{Gamma}\Big\{ u_\alpha + L, \Big(\sum_{l=1}^{L} \frac{(\alpha_{p+l,r})^2}{2} + \frac{1}{v_\alpha} \Big)^{-1} \Big\}. \end{aligned}$$

Since the full conditional distributions of parameters except for $\phi_{\alpha,r}$ do not have closed forms, we use Adaptive Rejection Metropolis Sampling (Gilks et al., 1995) within Gibbs Sampling(ARMS). The detailed procedures of our ARMS algorithm are provided in the Appendix A.3. The ARMS algorithm for Matched Case Crossover studies is summarized in Algorithm 1.

Let $\boldsymbol{\alpha}_{\boldsymbol{1,r}}^{(\boldsymbol{s})} = \big(\alpha_{0,r}^{(s)}, \ldots, \alpha_{p,r}^{(s)}\big)^T$, $\boldsymbol{\alpha}_{\boldsymbol{2,r}}^{(\boldsymbol{s})} = \big(\alpha_{p+1,r}^{(s)}, \ldots, \alpha_{p+L,r}^{(s)}\big)^T$ with $r = 1, \ldots, R$ corresponding to the region of interest and $s = 1, \ldots, S$ corresponding to the MCMC sample point obtained from the ARMS procedure. The $\beta_r(t)$ and $\beta_r'(t)$, $r = 1, \ldots, R$, are of our fundamental interest.

We construct the following two functions simultaneously:

$$\begin{aligned} \beta_{r,s}(t) &= \alpha_{0,r}^{(s)} + \alpha_{1,r}^{(s)} t + \cdots + \alpha_{p,r}^{(s)} t^p + \sum_{l=1}^{L} \alpha_{p+l,r}^{(s)} (t - \xi_l)_+^p \\ \beta_{r,s}'(t) &= \alpha_{1,r}^{(s)} + 2\alpha_{2,r}^{(s)} t + \cdots + p\alpha_{p,r}^{(s)} t^{p-1} + p \sum_{l=1}^{L} \alpha_{p+l,r}^{(s)} (t - \xi_l)_+^{p-1} \end{aligned}$$

Then the $\beta_r(t)$ and $\beta'_r(t)$ are estimated as follows using MCMC samples, for each time $t = 1, \ldots, T$,

$$\begin{aligned} \beta_r(t) &\approx \frac{1}{S}\sum_{s=1}^{S}\beta_{r,s}(t), \\ \beta'_r(t) &\approx \frac{1}{S}\sum_{s=1}^{S}\beta'_{r,s}(t). \end{aligned}$$

Let $\boldsymbol{\beta_r(t)} = \{\beta_{r,1}(t), \ldots, \beta_{r,S}(t)\}$ and $\boldsymbol{\beta'_r(t)} = \{\beta'_{r,1}(t), \ldots, \beta'_{r,S}(t)\}$. The $(1-\gamma)100\%$ Bayesian credible bands for $\beta_r(t)$ and $\beta'_r(t)$ can be calculated using the $\gamma/2$ and $1-\gamma/2$ quantiles of $\boldsymbol{\beta_r(t)}$ and $\boldsymbol{\beta'_r(t)}$ respectively.

3.2.4 Testing for the Existence of Features

In this section, we explain how to test whether an effect modification arising from time exists, as well as determining if this effect modification changes in spatial location. We propose the empirical procedures under the Bayesian paradigm for $1-M$ matched case-crossover studies.

Testing for the Existence of Modification by Time

We conduct the following approach to assess the presence of effect modification due to time.

Step 1: For each region, estimate the $\boldsymbol{\alpha}_r$ coefficients for $\beta_r(t)$, $r = 1, \ldots, R$;

Step 2: Use the estimated $\boldsymbol{\alpha}_r$ coefficients to find the first derivative of $\beta_r(t)$ with respect to time and denote it as $\beta'_r(t)$;

Step 3: Find a $(1-\gamma)100\%$ Bayesian credible interval at each time period for $\beta'_r(t)$;

Step 4: Determine the existence of a time varying function according to the following criteria: If the credible bounds $\beta'_r(t)$ are above zero, the $\beta_r(t)$ function is a monotonic increasing function in regions; If the credible bounds $\beta'_r(t)$ are below zero, the $\beta_r(t)$ function is a

monotonic decreasing function in regions; If the credible band covers a portion of the zero line, there is no significant increasing or decreasing pattern of the $\beta_r'(t)$ function.

Testing for Existence of Modification by Spatial Location

We conduct the following approach to determine whether there is a difference between the $\beta_r(t)$s for $r = 1, \ldots, R$.

Step 1: Find the paired differences between the $\beta_r(t)$, for example $\beta_2(t) - \beta_1(t)$;

Step 2: For each set of paired differences, find a $(1-\gamma)100\%$ Bayesian credible interval at each time period;

Step 3: Calculate the proportion of these credible intervals that do not include zero.

3.2.5 Simulation Study

Simulation Setting

We consider a 1-4 matched case-crossover study motivated by our childhood aseptic meningitis example. We set the number of stratum to 100 ($N = 100$), the number of time periods to 100 ($T = 100$) and two locations ($R = 2$). The covariates X were generated from a standard normal distribution N(0,1).

The 1-4 matched case-controls are generated from the prospective model

$$Pr(Y_k = 1|X_k, t, r, q) = \frac{\exp\{X_k\beta_r(t) + q(\bullet)\}}{1 + \exp\{X_k\beta_r(t) + q(\bullet)\}},$$

where $q(\bullet)$ is an unknown effect which disappears under the condition $\sum_{j=1}^{5} Y_j = 1$ for each stratum k, $k = 1, \ldots, N$ regardless of the distribution of $q(\bullet)$. For the estimation and method evaluation,

we simulated 50 datasets. We consider two cases. Case 1 is for estimating true positive rate and Case 2 is for estimating false positive rate.

- Case 1: $K = 100$, $R = 2$, $\beta_1(t) = \cos(\pi t)$, $\beta_2(t) = t^2$.
- Case 2: Same setting as Case 1 except for $\beta_1(t) = \beta_2(t) = t^2$

Testing of the Existence of Time Varying Coefficients under Case 1

It is of interest to examine whether increasing and decreasing patterns in the estimated $\beta_r(t)$ are significant, or not. Figure 3.1 presents an example of the estimated $\beta_r(t)$ with credible bounds. The shaded area corresponds to 99% credible bounds, the dashed lines correspond to the 95% credible bounds and the solid line corresponded to the estimated function. Parts (a) and (b) correspond to the estimated $\beta_1(t)$ and $\beta_2(t)$ functions respectively. Parts (c) and (d) correspond to the estimated $\beta_1'(t)$ and $\beta_2'(t)$ functions respectively.

When examining Figure 3.1 (a) and (b), it appears that $\beta_1(t)$ is decreasing, which means there would be a negative effect modification of our covariate over the presence or absence of disease for subjects in region one. $\beta_2(t)$ shows an increasing pattern, which means there would be a positive effect modification of our covariate over the presence or absence of disease for subjects in region two.

In order to assess whether these patterns are significant we further examine and obtain Figure 3.1 (c) and (d). We see that for all time t, $\beta_1'(t) < 0$ and $\beta_2(t) > 0$, this means that $\beta_1(t)$ is significantly decreasing and $\beta_2(t)$ is significantly increasing. Therefore, the effect modifications that we suspected are significant. This figure illustrate that for the entire time $\beta_1'(t)$ is significantly decreasing and $\beta_2'(t)$ is significantly decreasing.

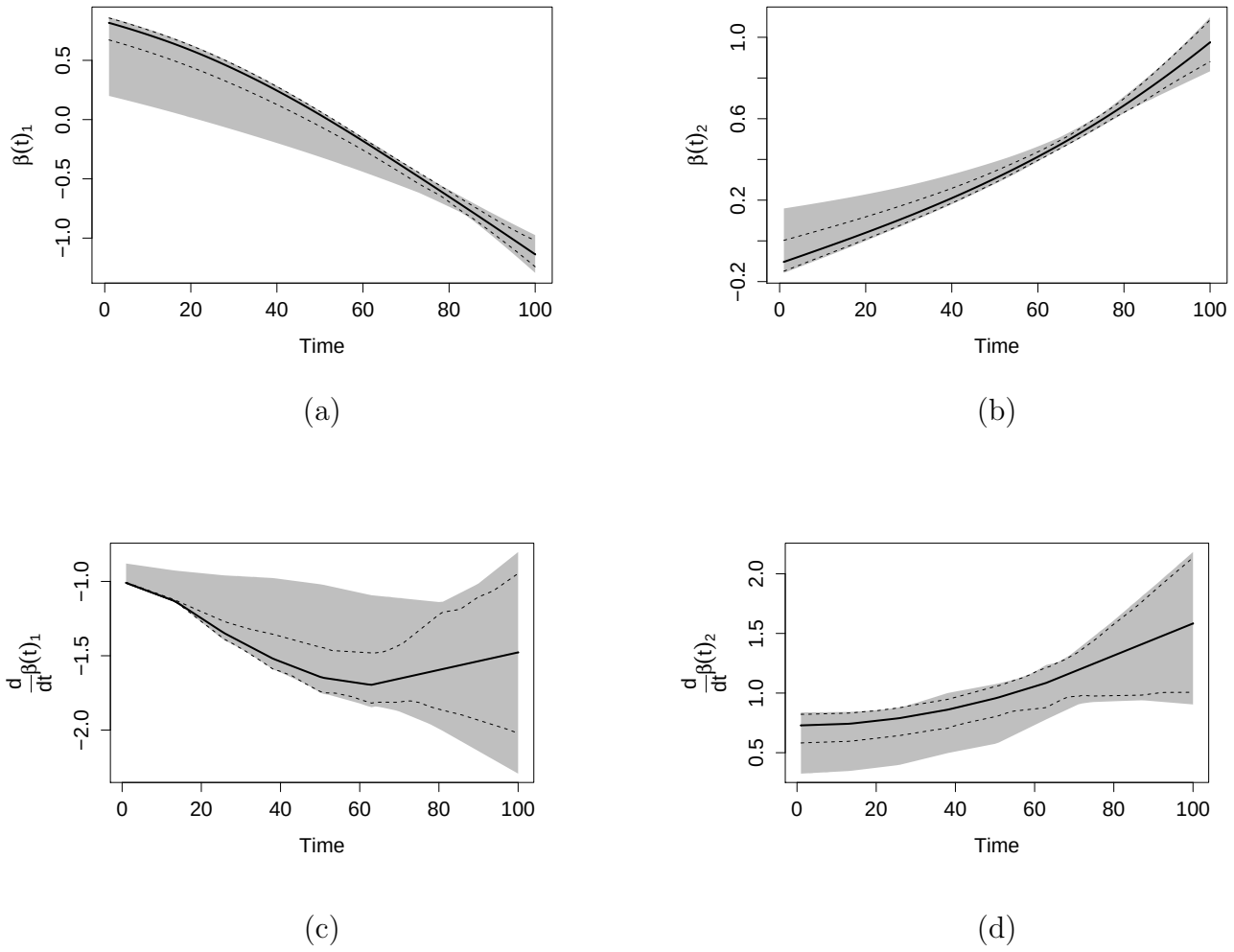


Figure 3.1: Example of the estimated $\beta_r(t)$ with credible bounds. The shaded area corresponds to 99% credible bounds, the dashed lines correspond to the 95% credible bounds and the solid line corresponded to the estimated function. (a) and (b) correspond to the estimated $\beta_1(t)$ and $\beta_2(t)$ functions respectively. (c) and (d) correspond to the estimated $\beta_1'(t)$ and $\beta_2'(t)$ functions respectively, where $\beta_r'(t) = \frac{d}{dt}\beta_r(t)$, $r = 1, 2$.

Testing of the Existence of Modification by Spatial Locations under Case 1

In order to assess the difference of the time varying coefficient functions between regions we consider the proportion of credible intervals for $\beta_2(t) - \beta_1(t)$ per each simulated dataset that do not include zero. Table 3.1 presents the average proportion of 95% and 99% credible intervals that do not

include zero. These high (97.08% and 93.20%) proportion values support the significant of the difference between two regions, $\beta_2(t) \neq \beta_1(t)$.

	95%	99%
$\beta_2(t) - \beta_1(t)$	0.9708	0.9320
$\beta_2'(t) - \beta_1'(t)$	1	1

Table 3.1: Average proportions of credible intervals that do not include zero, to test the difference between regions when $\beta_1(t) = t^2$ and $\beta_2(t) = cos(\pi t)$.

Another quantity of interest to assess the difference of the time varying coefficient functions between regions is the proportion of credible intervals for $\beta_2'(t) - \beta_1'(t)$ per each simulated dataset that do not include zero. Table 3.1 presents the average proportion of 95% and 99% credible intervals that do not include zero. These high proportions (100% for both) support the significant of the difference between two regions, $\beta_2'(t) \neq \beta_1'(t)$. This result also supports the significance of $\beta_2(t) \neq \beta_1(t)$.

Testing of the Existence of Modification by Spatial Locations under Case 2

In order to assess the difference of the time varying coefficient functions between regions, we consider the proportion of intervals for $\beta_2(t) - \beta_1(t)$ and $\beta_2'(t) - \beta_1'(t)$ that contain zero. These proportions will represent the false positive rate which is the average proportion of times in which a difference in regions is detected, when in fact, no difference is present. Table 3.2 summarizes these proportions, presenting low values of the false positive rate for both $\beta_2(t) - \beta_1(t)$ and $\beta_2'(t) - \beta_1'(t)$.

	95%	99%
$\beta_2(t) - \beta_1(t)$	0.0844	0.0242
$\beta_2'(t) - \beta_1'(t)$	0.0108	0.0056

Table 3.2: Average proportions of credible intervals that do not include zero, to test the difference between regions when $\beta_1(t) = \beta_2(t) = t^2$.

These values mean that 8% of the 95% credible intervals do not include zero (support the difference in regions hypothesis), and that 2% of the 99% credible intervals do not include zero. Low false

positive rates are evidence for the accuracy of the method.

3.2.6 Example: $1-4$ bi-directional matched case-crossover study Estimation

The main interest of our $1-4$ bidirectional case-crossover example is to understand the relationship between the effect of drinking water turbidity and the risk of aseptic meningitis in children as well as to explore the presence of potential effect modification on the turbidity effect by time and spatial location. The analysis of a case-crossover study is a stratified data analysis of retrospective, self-matched follow-up studies, each with a sample size of one. In this study, the stratifying variable is the individual child who is less than 15 years old, diagnosed in one of two years and lives on one of two urban communities in South Korea. Our potential effect modification includes a time effect and a spatial modification, where the covariate of interest is water turbidity. We applied the testing approaches presented in Section 3.2.4, by first estimating the time varying coefficients $\beta_1(t)$ and $\beta_2(t)$ for the two regions, respectively.

For testing of differences between two regions, we obtain Figure 3.2 which presents a graphical representation of this estimation with the solid line representing $\beta_1(t)$ and the dashed line representing $\beta_2(t)$. This figure shows a visible difference in the relationship between the coefficients and time, $\beta(t)$, for each region. It seems like both time varying coefficient functions are decreasing, however, the estimated function $\beta_2(t)$ drops more dramatically than for $\beta_1(t)$. Figure 3.2 (b) presents the estimated $\beta_2(t) - \beta_1(t)$.

Upon inspection of this figure we concluded that $\beta_1(t)$ and $\beta_2(t)$ are significantly different. Figure 3.2 (c) presents the estimated $\beta_2'(t) - \beta_1'(t)$ along with their credible bands. This figure supports the conclusion obtained from (b), $\beta_1(t)$ and $\beta_2(t)$ are significantly different. We also created 95 and 99% credible intervals for $\beta_2(t) - \beta_1(t)$ and for $\beta_2'(t) - \beta_1'(t)$. The proportions of these credible intervals that do not include zero are presented in Table 3.3. These high proportion values (93.55%

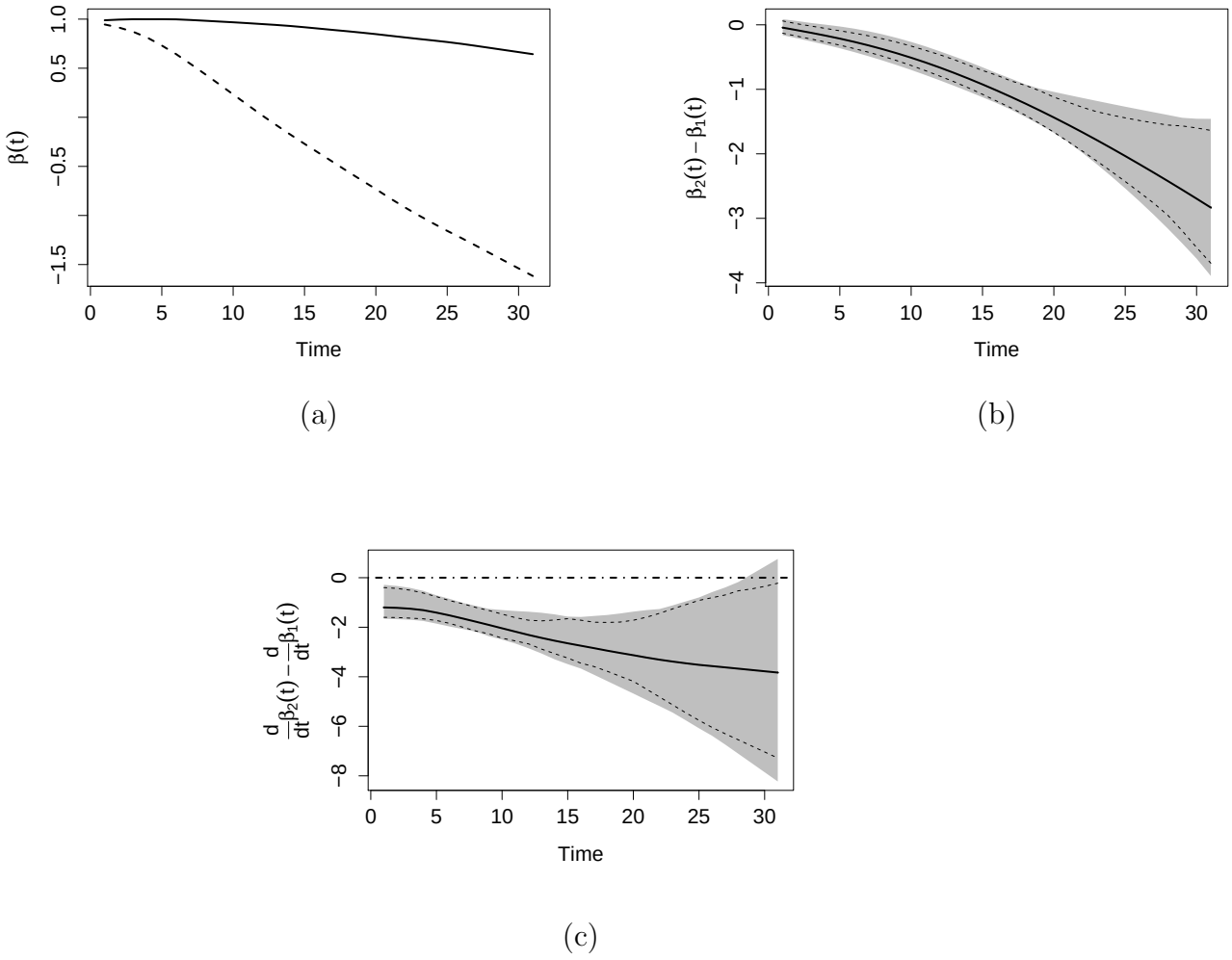


Figure 3.2: (a) Estimated Varying Coefficient Functions for each region. The solid line represents $\beta_1(t)$ (region 1) and the dashed line represents $\beta_2(t)$ (region 2). (b) Estimated $\beta_2(t) - \beta_1(t)$ and confidence bounds. (c) Estimated $\beta_2'(t) - \beta_1'(t)$ and confidence bounds, where $\beta_r'(t) = \frac{d}{dt}\beta_r(t)$, $r = 1, 2$.

and 87.10%) support the claim $\beta_2(t) \neq \beta_1(t)$. This result means that the spatial location of the subjects modifies the effect of water turbidity and binary outcomes of presence or absence of aseptic meningitis in children.

For testing of existence of effect modification due to time, we generated Figure 3.3 which presents

	95%	99%
$\beta_2(t) - \beta_1(t)$	0.9355	0.8710
$\beta_2'(t) - \beta_1'(t)$	1	0.9032

Table 3.3: Average proportions of credible intervals that do not include zero, to test the difference between regions for the 1-4 bidirectional matched case-crossover example.

the results of the estimation of $\beta_1(t)$ (a), $\beta_1'(t)$ (c), $\beta_2(t)$ (b) and $\beta_2'(t)$ (d). By examining these two functions together, we can conclude that even though $\beta_1(t)$ seems to be decreasing (part (a) only), the fact that the credible bounds for $\beta_1'(t)$ include zero for all t shows that this decrease is not significant. Therefore, in region one the effect of water turbidity on the presence of aseptic meningitis in children does not vary over time. However, by examining t (b) and (d) together, we conclude that $\beta_2(t)$ is decreasing, and since $\beta_2'(t)$'s credible bounds do not include zero for most values of t, this decrease is significant. Therefore, in region two the effect of water turbidity on the presence of aseptic meningitis in children does varies over time.

3.2.7 Discussion

In this section, we have proposed a semiparametric spatio-temporal varying coefficient model, which is appropriate when the number of spatial locations is small. This method allows us to detect relationships between a predictor X and binary outcome, as well as determine effect modifications by time and/or spatial location.

Simulation results show that the method performs well in estimating unknown $\beta(t)$ functions. Additionally, simulation results show that the method is successful at detecting differences in effect modification for subjects located in different regions, as well as detecting if the obtained functions vary through time significantly. We present an application for a 1-4 bidirectional matched case-crossover study of aseptic meningitis in children in two provinces in South Korea. We applied our approach and concluded that the effect modification of water turbidity on the presence or absence of disease varies according to the location of the subject. Moreover, we concluded that there is no

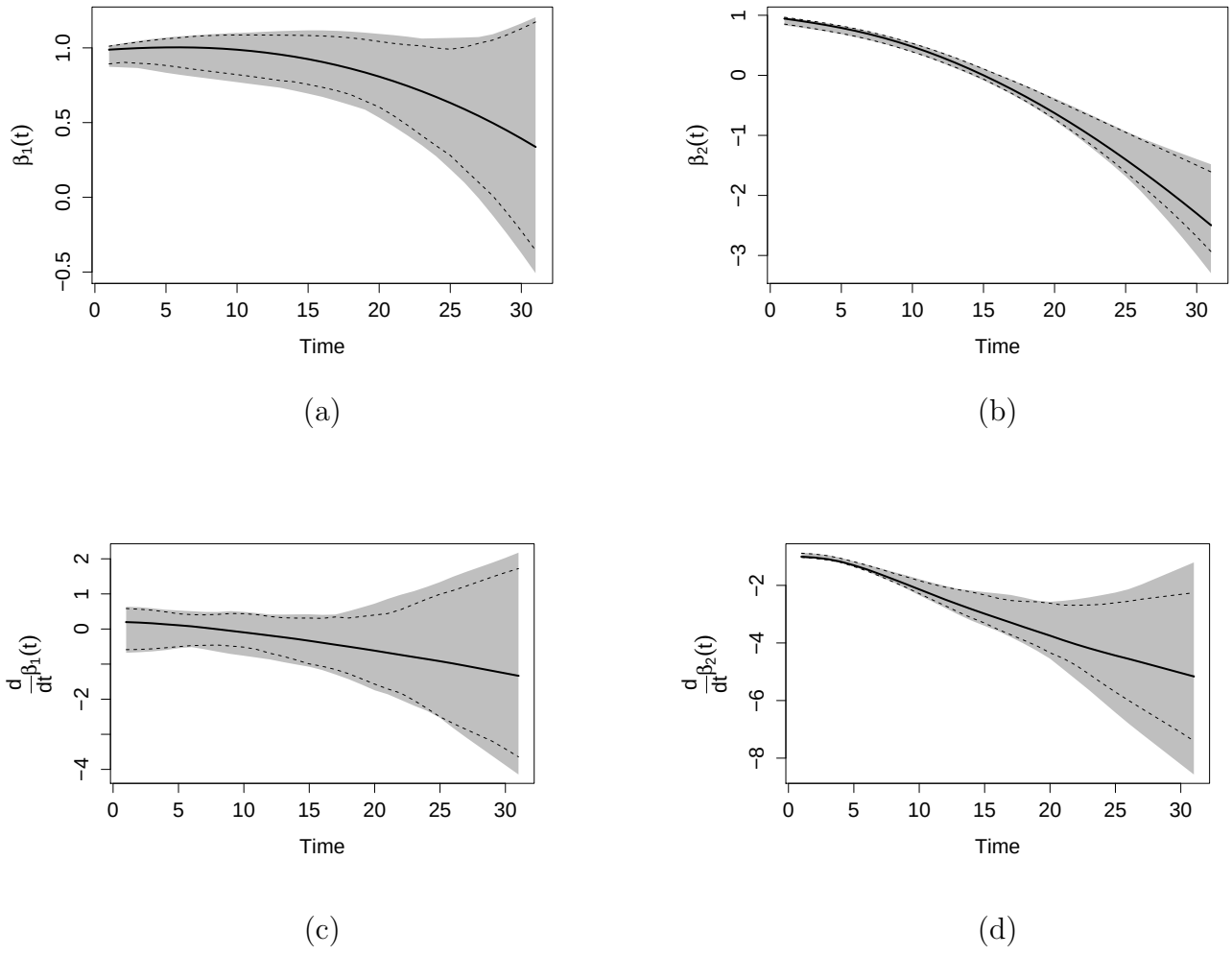


Figure 3.3: Estimated $\beta_r(t)$ with credible bounds in the 1-4 Bidirectional Matched Case-Crossover Study. The shaded area corresponds to 99% credible bounds, the dashed lines correspond to the 95% credible bounds and the solid line corresponded to the estimated function. (a) and (b) correspond to the estimated $\beta_1(t)$ and $\beta_2(t)$ functions respectively. (c) and (d) correspond to the estimated $\beta_1'(t)$ and $\beta_2'(t)$ functions respectively, where $\beta_r'(t) = \frac{d}{dt}\beta_r(t)$, $r = 1, 2$.

effect modification by time for subjects in region one, and there is an effect modification due to time, present for subjects in region two.

We will further develop our approach for the case when there is a large number of locations present in the next section.

3.3 Spatial-Temporal Varying Coefficient Model

3.3.1 Background

In Section 3.2 we proposed a method that allows the assessment of effect modification due to time and spatial location for a small number of locations among the subjects of a matched case-crossover study. In this section we further develop the spatial-temporal approach for the case in which we have a medium to large number of locations.

The proposed method is created to model semiparametrically the effect modification by time using regression splines, and the effect modification by spatial effect using a Gaussian random process. This model allows us to not only detect the parametric relationship between the predictor and binary outcomes but also evaluate a semiparametric relationship between the predictor and time as well as an association between the predictor and spatial effect. We refer to this model as semiparametric spatio-temporal varying coefficient model. We develop the procedure to estimate an empirical semivariogram in order to incorporate spatial dependence to matched case-crossover study estimation.

This section is organized as follows: In section 3.3.2, we propose the framework for our approach. In section 3.3.4 we present two different semivariogram estimation methods: a logistic regression-semivariogram and a conditional logistic regression-semivariogram. In section 3.3.5, we present the framework and results of our simulation study. In general, our simulation results suggest that the method perform well at estimating time varying functions as well as detecting effect modification due to spatial location among the subjects. Section 3.3.6 contains concluding remarks.

3.3.2 Spatial-Temporal Varying Coefficient Model

In this section, we propose a semiparametric spatial-temporal varying coefficient model. We consider a $1 - M$ matched case-crossover study with kth stratum, $k = 1, \ldots, N$. Let x_{1k} represent

the value of the covariate of interest for the case of stratum k, and x_{2k}, x_{3k}, $\ldots x_{M+1,k}$ the covariate values for the controls of stratum k. Let Z_1 represent the model matrix corresponding to the locations of the case, and Z_2, Z_3, $\ldots Z_{M+1}$ represent the model matrices corresponding to the locations of the controls, where the k-th row corresponds to information for stratum k. In this way, Z_i is created by row vectors $\mathbf{z_{i1}}, \mathbf{z_{i2}}, \ldots, \mathbf{z_{iK}}$, where $\mathbf{z_{ik}}$ represents location information for stratum $k = 1, \ldots, K$, and $i = 1$ represents location information for the case and $i = 2, \ldots, M+1$ represents location information for the controls.

To illustrate the construction of the $Z_1 \ldots X_{M+1}$ consider a 1-2 matched matched case-crossover study with 4 subjects and 3 locations as presented in Table 3.4.

Subject	Case	Control	Control
1	Location 1	Location 1	Location 1
2	Location 1	Location 1	Location 2
3	Location 3	Location 2	Location 3
4	Location 1	Location 3	Location 3

Table 3.4: Example for the construction of the $Z_{1k} \ldots X_{M+!,k}$ model matrices.

Let $Z_1 = (\mathbf{z_{11}}, \mathbf{z_{12}}, \mathbf{z_{13}}, \mathbf{z_{14}})^T$, $Z_2 = (\mathbf{z_{21}}, \mathbf{z_{22}}, \mathbf{z_{23}}, \mathbf{z_{24}})^T$ and $Z_3 = (\mathbf{z_{31}}, \mathbf{z_{32}}, \mathbf{z_{33}}, \mathbf{z_{34}})^T$, where

$$Z_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, Z_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, Z_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let us consider the parametric association between a covariate and binary outcomes and the semiparametric association between the time varying coefficient and binary outcomes. We replace the parameter β in (1.1) by the unknown function $\beta(t)$ and also incorporate spatial information. The conditional logistic regression model with time varying coefficient and spatial random effects for

kth stratum can then be written as

$$\begin{aligned} P\Big(Y_{1k}=1|x_{1k},...,x_{M+1,k},k,\sum_{j=1}^{M+1}Y_{jk}=1\Big) &= \frac{\exp\{x_{1k}\beta(t)+Z_{1k}\boldsymbol{\gamma}+q(\bullet)\}}{\sum_{j=1}^{M+1}\exp\{x_{jk}\beta(t)+Z_{jk}\boldsymbol{\gamma}+q(\bullet)\}} \\ &= \frac{1}{1+\sum_{j=2}^{M+1}\exp\{(x_{jk}-x_{1k})\beta(t)+(Z_{jk}-Z_{1k})\boldsymbol{\gamma}\}} \\ &\equiv l_k\{\beta(t),\boldsymbol{\gamma}\} \end{aligned} \tag{3.1}$$

where $\beta(t)$ is a function of time t, $t=1,\ldots,T$, and $\boldsymbol{\gamma}$ is the Gaussian random vector associated with spatial location.

The function $\beta(t)$ and $\boldsymbol{\gamma}$ are modeled semiparametrically. The function $\beta(t)$ can have semiparametric form via regression splines. For a pth order regression spline with a truncated power series basis and with L knots, we define a pth order regression spline with a truncated power series basis and with L knots, $\{1,t,t^2,\ldots,t^p,(t-\xi_1)_+^p,(t-\xi_2)_+^p,\ldots,(t-\xi_L)_+^p\}$, where $(u)_+^p = u^pI(u\geq 0)$ and knots $\xi_1<\ldots<\xi_L$. Knots can be selected a priori, e.g., at sample quantiles of observed t's which are scaled in [0,1]. Let $\alpha_0,\alpha_1,\ldots,\alpha_p$ be the regression coefficients and $\alpha_{p+1},\ldots,\alpha_{p+L}$ be the regression coefficients of the truncated power series basis. With this setup, $\beta(t)$ can be written as

$$\beta(t)=\alpha_0+\alpha_1t+\ldots+\alpha_pt^p+\sum_{l=1}^{L}\alpha_{1+l}(t-\xi_l)_+^p.$$

The spatial information ($\boldsymbol{\gamma}$) is modeled using Gaussian random process as follows:

$$\begin{aligned} \boldsymbol{\gamma} &\sim MVN(\mathbf{0},\frac{1}{\phi_s}\Sigma), \\ \Sigma_{ii'} &= \exp(-\phi_\rho \parallel \boldsymbol{s}_i-\boldsymbol{s}_{i'}\parallel^2) \end{aligned} \tag{3.2}$$

where ϕ_s is the precision parameter for the spatial random effect, ϕ_ρ is the covariance structure precision parameter, and $\parallel \boldsymbol{s}_i-\boldsymbol{s}_{i'}\parallel$ represents the distance between location i and i', which are calculated using Great Circle distances, $i,i'=\{1,2,....n_L\}$ with n_L being the total number of locations. Longitudinal and latitudinal information are used to calculate these distances.

3.3.3 Bayesian Hierarchical Model Framework

Let $\boldsymbol{\alpha_1} = (\alpha_0, \ldots, \alpha_p)^T$ and $\boldsymbol{\alpha_2} = (\alpha_{p+1}, \ldots, \alpha_{p+L})^T$. Define $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2})^T$.

We first specify the prior distributions of parameters as follows:

$$\begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_p \end{pmatrix} \sim N(\mathbf{0}, \phi_1^{-1} I),$$

$$\begin{pmatrix} \alpha_{p+1} \\ \vdots \\ \alpha_{p+L} \end{pmatrix} \sim N(\mathbf{0}, \phi_\alpha^{-1} I),$$

$$\boldsymbol{\gamma} \sim MVN(\mathbf{0}, \frac{1}{\phi_s}\Sigma),$$

$$\begin{aligned} \phi_\alpha &\sim \text{Gamma}(u_\alpha, v_\alpha), \\ \phi_s &\sim \text{Gamma}(u_s, v_s) \end{aligned}$$

where ϕ_1^{-1} is a fixed large number, u_α, u_s, v_α and v_s are the hyperparameters of the Gamma distributions, and Σ is the Gaussian covariance structure given by:

$$\Sigma_{ii'} = \exp(-\phi_\rho \parallel \boldsymbol{s}_i - \boldsymbol{s}_{i'} \parallel^2) \tag{3.3}$$

where $\parallel \boldsymbol{s}_i - \boldsymbol{s}_{i'} \parallel$ represents the distance between location i and i', where $i, i' = \{1, 2, ...n_L\}$, with n_L being the total number of locations.

For ϕ_ρ, we do not specify the prior distribution because ϕ_s and ϕ_ρ are not identifiable under one of the following situations: (i) $\phi_s \to 0$, (ii) $1/\phi_\rho \to 0$ and $\phi_s \sim O(\phi_\rho^c)$ for any positive c; or (iii)$\phi_\rho \to 0$.

To avoid this identification situation, we estimate the covariance structure parameter ϕ_ρ using the semivariogram. This estimation is presented in section 3.3.4.

Once the covariance structure parameter ϕ_ρ is estimated, our joint likelihood ($\mathcal{L}$) is then expressed as

$$\begin{aligned}\mathcal{L} \quad \propto \quad & \left[\prod_{k=1}^{N} \frac{1}{1+\sum_{j=2}^{M+1}\exp\Big\{(\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{p+l}(t-\xi_j)_+^p)x_{jk}^*+z_{jk}^*\boldsymbol{\gamma}\Big\}}\right] \\ \times \quad & \left\{\exp\Big(-\frac{\phi_1 \boldsymbol{\alpha_1}^T\boldsymbol{\alpha_1}}{2}\Big)\right\}\left\{\phi_\alpha^L \exp\Big(-\frac{\phi_\alpha \boldsymbol{\alpha_2}^T\boldsymbol{\alpha_2}}{2}\Big)\right\}\left\{\phi_\alpha^{u_\alpha-1}\exp\Big(-\frac{\phi_\alpha}{v_\alpha}\Big)\right\} \\ \times \quad & \left\{\phi_s^{u_s-1}\exp\Big(-\frac{\phi_s}{v_s}\Big)\right\}\left\{\phi_s^{n_L/2}\exp\Big(\frac{-\phi_s\boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma}}{2}\Big)\right\} \end{aligned} \tag{3.4}$$

where $x_{jk}^* = x_{jk} - x_{1k}$ and $z_{jk^*} = z_{jk} - z_{1k}$.

The full conditional likelihoods for α_q, $q = 1, \ldots, p$, are

$$p(\alpha_q|-) \quad \propto \quad \frac{\phi_1 \exp\Big(-\frac{\phi_1\alpha_q^2}{2}\Big)}{\prod_{k=1}^{N}\left[1+\sum_{j=2}^{M+1}\exp\Big\{(\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{p+l}(t-\xi_l)_+^p)x_{jk}^*+z_{jk}^*\boldsymbol{\gamma}\Big\}\right]} \tag{3.5}$$

The full conditional likelihoods for α_{p+q}, $q = 1, \ldots, L$, are

$$p(\alpha_{p+q}|-) \quad \propto \quad \frac{\phi_\alpha \exp\Big(-\frac{\phi_\alpha\alpha_{p+q}^2}{2}\Big)}{\prod_{k=1}^{N}\left[1+\sum_{j=2}^{M+1}\exp\Big\{(\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{p+l}(t-\xi_l)_+^p)x_{jk}^*+z_{jk}^*\boldsymbol{\gamma}\Big\}\right]} \tag{3.6}$$

The full conditional likelihood for $\boldsymbol{\gamma}$ is

$$p(\boldsymbol{\gamma}|-) \quad \propto \quad \frac{\exp(\phi_s(-\boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma}))}{\prod_{k=1}^{N}\left[1+\sum_{j=2}^{M+1}\exp\Big\{(\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{p+l}(t-\xi_l)_+^p)x_{jk}^*+z_{jk}^*\boldsymbol{\gamma}\Big\}\right]} \tag{3.7}$$

The full conditional likelihoods for the smoothing parameters ϕ_α and ϕ_s are:

$$p(\phi_\alpha|-) \quad \propto \quad \phi_\alpha^{(u_\alpha+L)-1}\exp\left\{-\phi_\alpha\Big(\sum_{l=1}^{L}\frac{\alpha_{p+l}^2}{2}+\frac{1}{v_\alpha}\Big)\right\}$$

$$= \text{Gamma}\left\{u_\alpha + L, \left(\sum_{l=1}^{L} \frac{\alpha_{p+l}^2}{2} + \frac{1}{v_\alpha}\right)^{-1}\right\} \tag{3.8}$$

$$\begin{aligned} p(\phi_s|-) &\propto \phi_s^{(u_s+n_L/2)-1}\exp\left\{-\phi_s\left(\frac{\boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma}}{2} + \frac{1}{v_s}\right)\right\} \\ &= \text{Gamma}\left\{u_s + n_L/2, \left(\frac{\boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma}}{2} + \frac{1}{v_s}\right)^{-1}\right\} \end{aligned} \tag{3.9}$$

Since the full conditional distributions of parameters except for ϕ_α and ϕ_s do not have closed forms, we use Adaptive Rejection Metropolis Sampling (Gilks et al., 1995). The detailed procedures of our ARMS algorithm are provided in the Appendix B.1. ARMS algorithm for Matched Case Crossover studies is summarized in Algorithm 2.

3.3.4 Semivariogram Estimation

In the model, we do not specify the prior distribution because ϕ_s and ϕ_ρ due to identifiability problems. To overcome this situation we estimate the covariance structure parameter ϕ_ρ using a semivariogram. We propose an empirical way to estimate semivariogram for matched case-crossover studies. The empirical semivariogram is calculated by following procedures:

Step 1: Fit the conditional logistic regression under the model

$$Pr(Y_{1k} = 1|x_{1k}, \ldots, x_{M+1,k}, k, \sum_{j=1}^{M+1} Y_{jk} = 1) = \frac{\exp(\theta \mathrm{x_{1k}})}{\sum_{j=1}^{M+1}\exp(\theta \mathrm{x_{jk}})}, \text{ where } k = 1, \ldots, N,$$

Notice: time is ignored for this estimation.

Step 2: Compute $\hat{p}_k$ which is given by

$$\hat{p}_k = \frac{\exp(\hat{\theta}\mathrm{x}_{1\mathrm{k}})}{\sum_{j=1}^{M+1}\exp(\hat{\theta}\mathrm{x}_{\mathrm{jk}}))}, \text{ where } k = 1, \ldots, N,$$

Step 3: Compute $r_k(s_i)$ which represents the residual for stratum k of the case at location s_i

$$r_k(s_i) = \frac{1-\hat{p}_k}{\sqrt{\hat{p}_k(1-\hat{p}_k)}},$$

Step 4: Calculate the empirical semivariogram given by

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{(i,i')|\boldsymbol{h}_{ij}=\boldsymbol{h}}\left[\left\{r_k(s_i) - r_k(s_{i'})\right\}^2\right]$$

where $N(\boldsymbol{h})$ is the total number of pairs for which $\boldsymbol{h}_{ij} = \boldsymbol{h}$, $\boldsymbol{h}_{ij} = s_i - s_{i'}$ and $r(s_i)$ and $r(s_{i'})$ represent the set of all residuals from locations s_i and $s_{i'}$, respectively including all strata $k = 1 \ldots N$.

Step 5: The semivariogram is characterized primarily by three measures: (1) the nugget which represents variability that cannot be explained by the distance between observations; (2) the sill, which represents the maximum observed variability of the data and (3) the range, which represents the value at which the semivariogram stops increasing, i.e. the distance at which two observations are unrelated. The estimation of ϕ_ρ is done by calculating the inverse of the range ρ at which 90% of the sill is reached.

3.3.5 Simulation Study

In this section we conduct the simulation to investigate the estimation accuracies of the model for $\beta(t)$

Simulation Setting

We consider a 1-4 matched case-crossover study. We set the number of stratum (N) as 100 and the number of time periods (T) as 100 for each stratum, that is, $N = 100$ and $T = 100$. We also consider $n_L = 9$ locations corresponding to the nine provinces in South Korea. The latitude and longitude of these locations are presented in the Appendix B.2.

The covariates X were generated from a $N(0, 1)$ distribution. Three different functions for $\beta(t)$ are used: (1) $\beta(t) = 0.5t$, (2) $\beta(t) = t^2$, and (3) $\beta(t) = \cos(\pi t)$. Values of ϕ_s are considered to be $\phi_s = \{1,\ 2,\ 3\}$ and $\phi_\rho = 2$. Spatial random effects are generated from multivariate normal distribution, $\boldsymbol{\gamma} \sim MVN(\mathbf{0}, \frac{1}{\phi_s}\Sigma)$ where Σ is the spatial covariance structure given by (3.2). Distances are calculated using the Great Circle Distance method under an ellipsoid earth mechanism. The Great Circle Distance method calculates the shortest distance between two points on the surface of an ellipsoid, measured along the surface of the ellipsoid, taking into account the radius of the earth.

The 1-4 matched case-controls are generated from the prospective model

$$Pr(Y_k = 1|X_k, t, q) = \frac{\exp\{X_k\beta(t) + Z_k\boldsymbol{\gamma} + q(\bullet)\}}{1 + \exp\{X_k\beta(t) + Z_k\boldsymbol{\gamma} + q(\bullet)\}},$$

where $q(\bullet)$ is unknown effect including intercept and can be generated from any distribution. However, $q(\bullet)$ disappears under the condition $\sum_{j=1}^{5} Y_j = 1$ for each stratum k, $k = 1, \ldots, N$ regardless of the distribution of $q(\bullet)$.

$\beta(t)$	Number of Knots (L)
$0.5t$	5
t^2	10
$\cos(\pi t)$	5

Table 3.5: Number of knots (L) for estimations in simulated data sets

For each function $\beta(t)$, we simulated 50 data sets and applied our proposed approach. Table 3.5 summarizes the number of knots used during estimations for each of the $\beta(t)$ functions. We then determine the accuracy of our approach in terms of the mean squared error (MSE) and R^2 which is the square of correlation between $\beta(t)$ and $\hat{\beta}(t)$. This MSE is calculated as follows:

$$\text{MSE} = \text{Bias}^2 + \text{Variance}$$

where

$$\begin{aligned} \text{Bias}^2 &= \frac{1}{100}\left[\sum_{d=1}^{50}\frac{1}{50}\sum_{t=1}^{100}\{\beta(t)-\hat{\beta}_d(t)\}^2\right], \\ \text{Variance} &= \frac{1}{100}\left[\sum_{d=1}^{50}\frac{1}{50-1}\sum_{t=1}^{100}\{\hat{\beta}_d(t)-\bar{\hat{\beta}}(t)\}^2\right]. \end{aligned}$$

Here $\hat{\beta}_d(t)$ represents the estimated $\beta(t)$ using the d simulated data and

$$\bar{\hat{\beta}}(t) = \frac{1}{50}\sum_{d=1}^{50}\hat{\beta}_d(t).$$

Simulation Results

We studied three different simulation cases. Case 1 is for evaluating the performance of the method without the use of a semivariogram, that is known ϕ_ρ. Case 2 and Case 3 are considered to evaluate the performance of the method with all parameters unknown.

Case 1: Known ϕ_ρ

Case 2: Unknown ϕ_ρ using empirical semivariogram under conditional logistic regression.

Case 3: Unknown ϕ_ρ using under logistic regression ignoring stratum effect.

For Case 1, we estimated all parameters by fixing the number of knots to the quantities presented in Table 3.5. We chose the location of knots using sample quantiles of the data, and set $\phi_\rho = 2$, the true value of the parameter used to generate the data.

For Case 2 and Case 3 we performed estimations with the settings for number and location of knots as in Case 1, with the difference that ϕ_ρ is estimated using semivariogram under both conditional logistic and logistic methods respectively.

Figure 3.4 presents a comparison between the true and estimated function using the proposed method. Part (a) presents $\beta(t) = 0.5t$, part (b) presents $\beta(t) = t^2$ and part (c) presents $\beta(t) = cos(\pi t)$. In this figure the solid line represents the true function and the dashed line represents the estimated function. This figure shows that the proposed method is successful at capturing the true structure of all the simulated functions.

Table 3.6 summarizes the results of the estimation of precision parameters ϕ_α and ϕ_s. This table shows the accuracy of the estimation of ϕ_s for all simulated functions, and different values of $\phi_s = \{1,\ 2,\ 3\}$. The average squared bias for ϕ_s is 0.0199. Table 3.7 presents accuracy measures for all simulated functions under different values of ϕ_s. The results

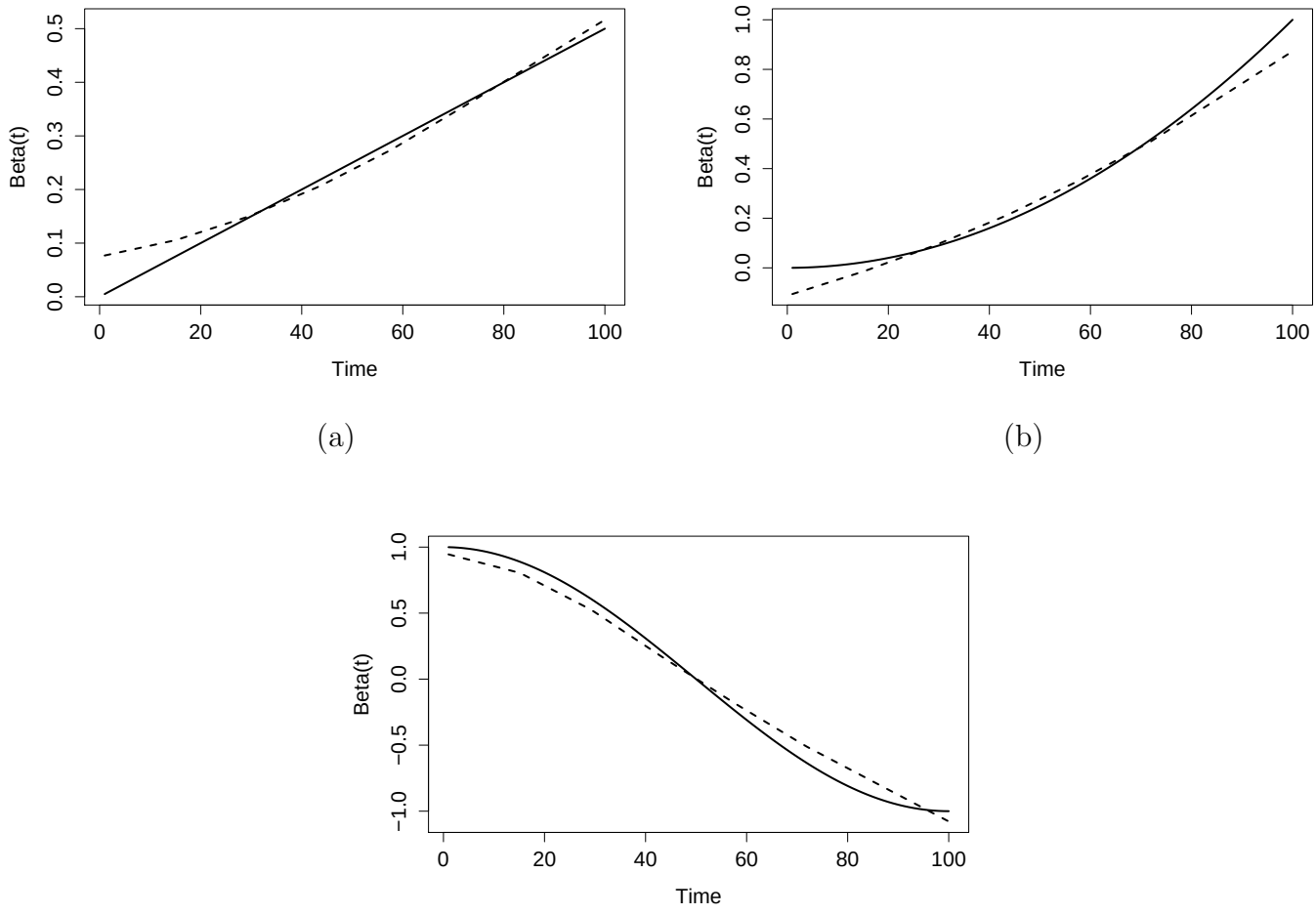


Figure 3.4: Example of the estimated $\beta(t)$ under the known spatial covariance parameter: (a) β(t)= $0.5t$, (b) β(t)= t^2 and (c) β(t)= $\cos(\pi t)$. Solid line represents true function, dashed line represents estimated function.

presented in this table, show that our approach performs well in estimating unknown $\beta(t)$ from the large $R^2 \geq 0.97$ and small MSE (≤ 0.0005).

For Case 2, we estimate ϕ_ρ using semivariogram estimation under conditional logistic regression. Table 3.8 summarizes the results of the estimation of precision parameters ϕ_α ϕ_s, and ϕ_ρ. In this table we can see the accuracy of our method, since the estimated values for

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
$\hat{\phi}_\alpha$	0.0505	0.0499	0.0598	0.0507	0.0519	0.0601	0.0499	0.0594	0.0502
$\hat{\phi}_s$	1.0966	1.8325	2.8729	0.9657	1.8901	2.7961	1.1324	1.7740	2.9503

Table 3.6: Estimated precision parameter under the known ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$.

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
MSE	0.0004	0.0004	0.0004	0.0004	0.0005	0.0004	0.0005	0.0004	0.0001
R^2	0.9715	0.9788	0.9830	0.9915	0.9817	0.9905	0.9935	0.9940	0.9947

Table 3.7: Average MSE and R^2 under the known spatial covariance parameter ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$.

ϕ_s, and ϕ_ρ are very close to the true values. Furthermore, the average squared bias for ϕ_s is 0.0191 and for ϕ_ρ is 0.0277. These estimations are competitive with the results obtained from the logistic-regression-semivariogram method.

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
$\hat{\phi}_\alpha$	0.0590	0.0552	0.0499	0.0521	0.0470	0.0501	0.0493	0.0499	0.0490
$\hat{\phi}_s$	0.8822	2.0900	2.6200	0.9912	1.9801	3.0513	0.9739	2.0156	3.0334
$\hat{\phi}_\rho$	1.8260	1.7998	1.7789	1.9880	2.0415	2.1858	1.9161	1.7057	1.9826

Table 3.8: Estimated precision parameter under the unknown spatial covariance parameter ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$ with conditional logistic regression semivariogram estimation.

Table 3.9 presents the accuracy measures for the estimation of our three functions under different values of ϕ_s. This table suggests that our method performs well in estimating the unknown $\beta(t)$ functions in terms of R^2 and MSE. It shows large R^2 values, with all $R^2 > 0.95$ and small MSE values, with all MSE ≤ 0.0005. These estimations are competitive with the results obtained from the logistic-regression-semivariogram method.

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
MSE	0.0003	0.0003	0.0004	0.0004	0.0005	0.0002	0.0005	0.0002	0.0001
R^2	0.9904	0.9677	0.9873	0.9901	0.9914	0.9851	0.9602	0.9754	0.9845

Table 3.9: Average MSE and R^2 under the unknown spatial covariance parameter ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$ with conditional logistic regression semivariogram estimation.

For Case 3, we estimate ϕ_ρ using semivariogram estimation under logistic regression. Table 3.10 summarizes the results of the estimation of precision parameters ϕ_α ϕ_s, and ϕ_ρ. In this table we can see the accuracy of our method, since the estimated values for ϕ_s, and ϕ_ρ are very close to the true values. Furthermore, the average squared bias for ϕ_s is 0.0154 and for ϕ_ρ is 0.0368.

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
$\hat{\phi}_\alpha$	0.0490	0.0502	0.0497	0.0501	0.0490	0.0502	0.0495	0.0489	0.0499
$\hat{\phi}_s$	0.8839	2.1400	2.8200	0.9611	1.9179	2.8583	0.9971	1.7956	2.9484
$\hat{\phi}_\rho$	1.7264	1.7799	1.8782	1.7780	2.1467	2.1858	2.0077	1.7057	2.0274

Table 3.10: Estimated precision parameter under the unknown spatial covariance parameter ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$ with logistic regression semivariogram estimation.

Table 3.11 presents the accuracy measures for the estimation of our three functions under different values of ϕ_s. This table suggests that our method performs well in estimating the unknown $\beta(t)$ functions in terms of R^2 and MSE. It shows large R^2 values, with all $R^2 > 0.95$ and small MSE values, with all MSE ≤ 0.0004.

Figure 3.5 presents three estimated functions obtained from our method under the logistic-regression-semivariogram: part (a) for $\beta(t) = 0.5t$, part (b) for $\beta(t) = t^2$ and part (c) for $\beta(t) = cos(\pi t)$. In this figure the solid line represents the true function and the dashed

	$\beta(t) = 0.5t$			$\beta(t) = t^2$			$\beta(t) = cos(\pi t)$		
ϕ_s	1	2	3	1	2	3	1	2	3
MSE	0.0004	0.0003	0.0004	0.0004	0.0004	0.0002	0.0003	0.0002	0.0002
R^2	0.9634	0.9737	0.9825	0.9699	0.9805	0.9801	0.9912	0.9984	0.9915

Table 3.11: Average MSE and R^2 under the unknown spatial covariance parameter ϕ_ρ case using the Spatio-Temporal Varying coefficient model with semivariogram for $\beta(t)$ with logistic regression semivariogram estimation.

line represents the estimated function. This figure shows that the proposed method under unknown ϕ_ρ is just as successful at capturing the true structure of all the simulated functions, as the method with assumed ϕ_ρ.

3.3.6 Discussion

In this section, we have proposed a semiparametric spatial-temporal varying coefficient model which allows us to detect relationships between a predictor X and binary outcome, as well as determine effect modifications by time and/or spatial location.

We proposed an alternative to evaluate semiparametric relationships between the predictor and time, and parametric relationships between the predictor and binary outcome under two different semivariogram estimation methods. The first one allows the construction of a semivariogram with residuals from a logistic regression, and the second one constructs a semivariogram with residuals from a conditional logistic regression. Estimations under both methods are satisfactory, however, the logistic regression method is faster and easier to implement.

Simulation results show that the method under both semivariogram estimations, perform well in estimating the unknown function $\beta(t)$, spatial parameters ϕ_ρ, and ϕ_s as well as finding a suitable smoothing parameter for the semiparametric estimation ϕ_α, providing satisfactory

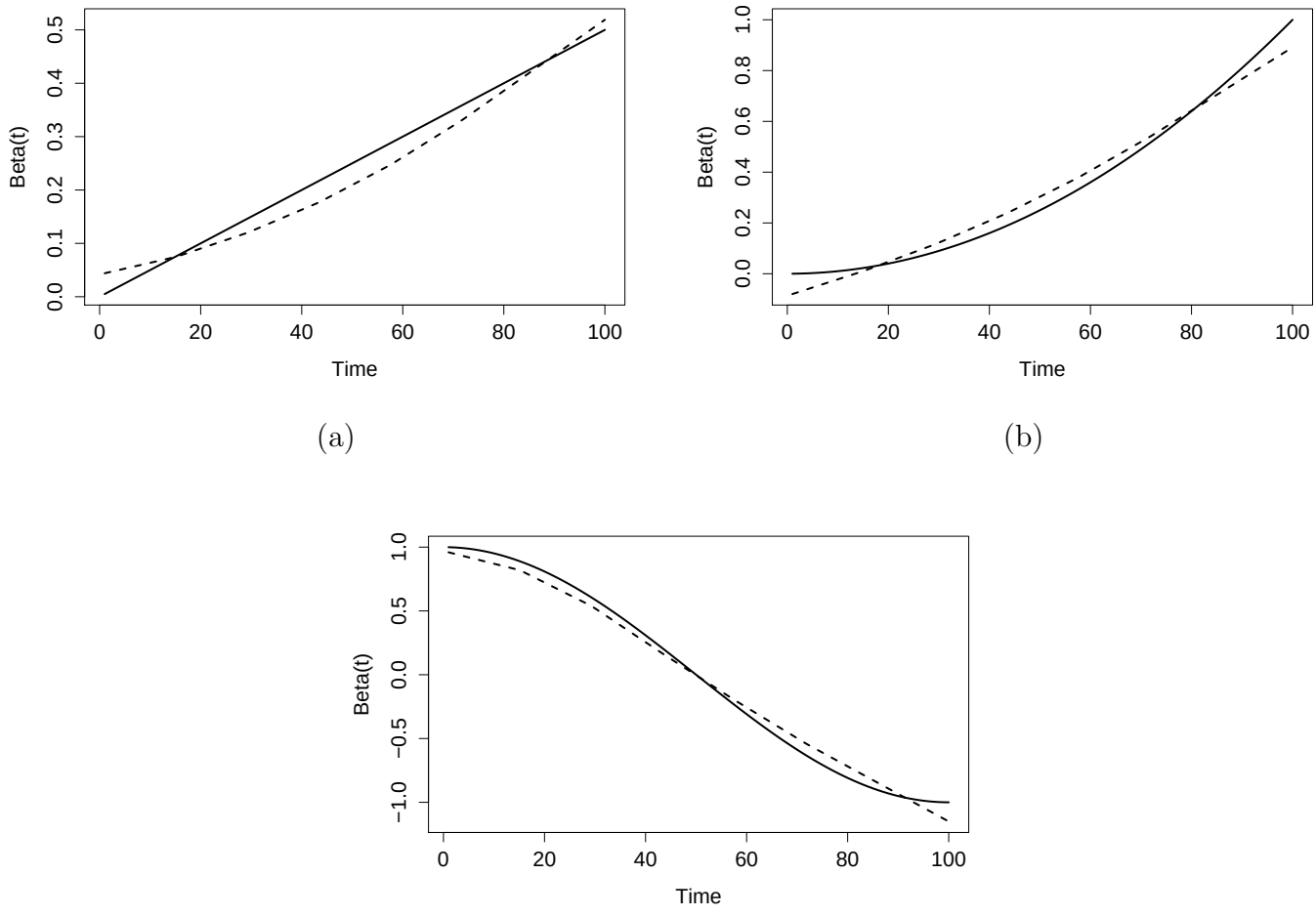


Figure 3.5: Example of the estimated $\beta(t)$ under the unknown spatial covariance parameter: (a) β(t)= $0.5t$, (b) β(t)= t^2 and (c) β(t)= $\cos(\pi t)$. Solid line represents true function, dashed line represents estimated function.

results in terms of accuracy of the estimation, that solve the problem at hand.

This research can be expanded by developing a method that allows us to detect semiparametric relationships between the predictor and binary outcomes, and semiparametric relationships between the predictor and time, as well as detection of effect modification due to spatial location. This can be accomplished by modeling the $\beta(t)$ function using the method

proposed in this paper, and in addition, modeling the relationship between the covariate and binary response in a semiparametric way, using regression splines.

Chapter 4

Functional Nonparametric Bayesian Time Varying Coefficient Model for Matched Case-Crossover Studies

4.1 Background

The methods for assessing the functional relationship between time varying coefficients and matched binary outcomes while assessing for a clustering structure within the data for matched case-crossover studies are very limited. Our main goals in this chapter are (1) to understand the functional relationship between drinking water turbidity and the relative risk of of aseptic meningitis in children, (2) to explore the modification in these effects by association with the potential latent period and (3) to determine functional clusters of time varying coefficients among strata. We are interested in assessing in which way the time varying coefficient function changes between clusters. To achieve these three goals, we propose a nonparametric Bayesian approach constructed with Dirichlet process priors, which

automatically cluster subpopulations among strata and assess heterogeneity.

We develop our nonparametric Bayesian approach under semiparametric time varying coefficient models in matched case-crossover studies. The time varying coefficient model is estimated using regression splines, clustering of time varying coefficients among strata is assessed using a Dirichlet process, and parameter estimation is done using adaptive rejection Metropolis sampling (ARMS) within Gibbs sampling.

The rest of this chapter is organized as follows: In section 4.2, we propose our approach, the functional semiparametric Bayesian time varying coefficient model in matched case-crossover studies. Section 4.3 contains the MCMC sampling scheme. In section 4.4 we provide the detailed procedure for cluster determination. In section 4.5 we present the simulation study framework and accuracy results. Section 4.6 presents the estimation of the 1-4 bidirectional matched case-crossover study. Section 4.7 contains concluding remarks.

4.2 Functional Semiparametric Bayesian Approach

We consider the conditional logistic regression model with time varying coefficient for kth stratum from equation 2.1, in which we model $\beta(t)$ in the conditional logistic regression semiparametrically via a pth order regression splines.

4.2.1 Semiparametric Bayesian Hierarchical Framework

Let $\boldsymbol{\alpha_1}^{(k)} = (\alpha_0^{(k)}, \alpha_1^{(k)}, \ldots, \alpha_p^{(k)})^T$ be the coefficients associated with the parametric portion of the time varying coefficient function for stratum $k = 1, \ldots, N$, and $\boldsymbol{\alpha_2}^{(k)} = (\alpha_{p+1}^{(k)}, \ldots, \alpha_{p+L}^{(k)})^T$ be the coefficients associated with the nonparametric portion of the time varying coefficient function for stratum k. Also let $\boldsymbol{\alpha}^{(k)} = \left(\boldsymbol{\alpha_1}^{(k)}, \boldsymbol{\alpha_2}^{(k)}\right)$. Our main interest is to estimate

functional clusters of time varying coefficients among strata.

Prior Distributions

In order to estimate functional clusters of time varying coefficients among strata, we specify the prior distribution of the parameters $\boldsymbol{\alpha}^{(k)}$ in the semiparametric time varying coefficient function for each stratum as Dirichlet Process (DP) priors (Ferguson, 1973; Ray and Mallick, 2006), which leads to mixtures of Dirichlet processes.

That is, the prior distribution of $\boldsymbol{\alpha}^{(k)}$ is following an unknown distribution, denoted by F. The DP provides a nonparametric prior specification over the class of possible distribution functions. The DP has two parameters F_0 and γ, where γ is known as the concentration parameter and F_0 is the base measure.

The base measure, F_0, is a distribution used to approximate the true nonparametric shape of F. This means that F_0 is a parametric function, that depends on its own known parameters. The concentration parameter (γ), a scalar, reflects our prior belief about the similarity between the nonparametric distribution F and the base function F_0.

We can express the prior distribution of $\boldsymbol{\alpha}^{(k)}$ as follows:

$$\begin{aligned} \boldsymbol{\alpha}^{(k)} &\sim F, \\ F &\sim DP\big(\gamma, F_0\big), \end{aligned}$$

where $\gamma > 0$ is concentration parameter and F_0 is the base measure.

As $\gamma \to \infty$, $F \to F_0$, so that the base measure is the parametric prior distribution for $\boldsymbol{\alpha}^{(k)}$. Also if $\boldsymbol{\alpha}^{(k)} \equiv \boldsymbol{\alpha}$, the same is also true, i.e., $F = F_0$. Where $\boldsymbol{\alpha}$ is a common set of coefficients for the semiparametric spline function.

We set the prior distribution of γ as a discrete uniform distribution $\gamma \sim Unif[0.01, 0.1, 1, 10, 100]$. The base measure F_0 is specified as $F_0 = N(\mathbf{0}, \phi_k^{-1}\boldsymbol{I})$, where the ϕ_ks play a role as smoothing parameters. We set $\phi_k = \phi_1$, for $\boldsymbol{\alpha_1}^{(k)}$ and $\phi_k = \phi_{k,\alpha}$ for $\boldsymbol{\alpha_2}^{(k)}$. Here ϕ_1 is a fixed small number, for global pattern estimation but set $\phi_{k,\alpha} \sim$ Gamma(u, v) for local pattern estimation. Here, u and v are the hyperparameters of the Gamma distribution chosen to be uninformative.

Our prior specifications are summarized as follows:

$$\begin{aligned}
\boldsymbol{\alpha}^{(k)} &\sim F, \\
F &\sim DP\big(\gamma, F_0\big), \\
\gamma &\sim \text{Unif}[0.01, 0.1, 1, 10, 100], \\
F_0 &= N(\mathbf{0}, \phi_k^{-1}\boldsymbol{I}), \\
\phi_{k,\alpha} &\sim \text{Gamma(u, v)}.
\end{aligned}$$

4.2.2 Posterior Distribution of the $\boldsymbol{\alpha}$ Coefficients under unknown F

We subsequently use the Pölya urn representation (Blackwell and MacQueen (1973)) to obtain the clustered sample points. The conditional posterior distributions of $\boldsymbol{\alpha}^{(k)}$ is derived as the following mixture distribution:

$$\begin{aligned}
p(\boldsymbol{\alpha}^{(k)}|, k \neq l) \quad \propto \quad & \left\{ \sum_{l \neq k} \frac{1}{\prod_{t=1}^{T} \left\{ 1 + \sum_{j=2}^{M} \exp\big[(x_{jkt} - x_{1kt})\beta_l(t)\big] \right\}} \right\} \delta_{\boldsymbol{\alpha}^{(l)}} \\
& + \left\{ \gamma \int \int \dots \int \frac{N(\alpha_0; 0, \phi_1^{-1}) \prod_{j=1}^{M+1} N(\alpha_j|0, \phi_k^{-1})}{\prod_{t=1}^{T} \left\{ 1 + \sum_{j=2}^{M} \exp\big[(x_{jkt} - x_{1kt})\beta_k(t)\big] \right\}} d\alpha_0 d\alpha_1 \dots d\alpha_{M+1} \right\}
\end{aligned}$$

$$\times N(\alpha_0; 0, \phi_1^{-1}) \prod_{j=1}^{M+1} N(\alpha_j|0, \phi_k^{-1}) \frac{1}{\prod_{t=1}^{T} \left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta_k(t)\right]\right\}}.$$

Let q_0 and q_z be denoted as follows:

$$\begin{aligned} q_0 &= \int\int \dots \int \frac{N(\alpha_0; 0, \phi_1^{-1}) \prod_{j=1}^{M+1} N(\alpha_j|0, \phi_k^{-1})}{\prod_{t=1}^{T} \left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta_k(t)\right]\right\}} d\alpha_0 d\alpha_1 \dots d\alpha_{M+1} \\ q_z &= \frac{1}{\prod_{t=1}^{T} \left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta_k(t)\right]\right\}} \end{aligned}$$

where $z = 1, \dots, k-1$.

With the previous notation, the Pölya urn is then constructed as follows:

$$\boldsymbol{\alpha}^{(k)} = \begin{cases} \tilde{\boldsymbol{\alpha}} & \text{with probability } \frac{\gamma q_0}{\gamma q_0 + q_1 + \dots + q_{k-1}} \\ \boldsymbol{\alpha}^{(1)} & \frac{q_1}{\gamma q_0 + q_1 + \dots + q_{k-1}} \\ \vdots & \\ \boldsymbol{\alpha}^{(k-1)} & \frac{q_{k-1}}{\gamma q_0 + q_1 + \dots + q_{k-1}} \end{cases} \tag{4.1}$$

where $\tilde{\boldsymbol{\alpha}}$ are sampled from the full conditional distributions described in next section. In this way, groups of clusters are obtained from the sampling of identical coefficients.

4.2.3 Posterior Distribution of the $\boldsymbol{\alpha}$ Coefficients under $F = F_0$

Here $\tilde{\boldsymbol{\alpha}}$ are sampled from the full conditional distributions of $\boldsymbol{\alpha}$ under $F = F_0$. The full conditional distributions are summarized as follows:

- Full conditional distribution of α_i, $i = 0, 1, \ldots, p$ of $\tilde{\boldsymbol{\alpha}}_{\mathbf{1}}$:

$$p(\alpha_i|-) \quad \propto \quad \frac{\exp\left(\frac{-\phi_1 \alpha_i^2}{2}\right)}{\prod_{k=1}^{N}\prod_{t=1}^{T}\left\{1+\sum_{j=2}^{M}\exp\left[(\mathrm{x_{jkt}}-\mathrm{x_{1kt}})\beta(\mathrm{t})\right]\right\}};$$

- Full conditional distribution α_{p+i}, $i = 1, \ldots, L$ of $\tilde{\boldsymbol{\alpha}}_{\mathbf{2}}$:

$$p(\alpha_{p+i}|-) \quad \propto \quad \frac{\exp\left(\frac{-\phi_{\mathrm{k},\alpha} \alpha_{\mathrm{p+i}}^2}{2}\right)}{\prod_{k=1}^{N}\prod_{t=1}^{T}\left\{1+\sum_{j=2}^{M}\exp\left[(\mathrm{x_{jkt}}-\mathrm{x_{1kt}})\beta(\mathrm{t})\right]\right\}};$$

- Full conditional distribution of $\phi_{k,\alpha}$:

$$\phi_{k,\alpha} \quad \propto \quad \mathrm{Gamma}\left[\mathrm{L}+\mathrm{u}, \left(\sum_{\mathrm{l}=1}^{\mathrm{L}} \alpha_{\mathrm{p+l}}^2/2 + 1/\mathrm{v}\right)^{-1}\right].$$

Since the full conditional distributions of parameters except for the precision parameters ($\phi_{k,\alpha}$) do not have closed form, we use Adaptive Rejection Metropolis Sampling (ARMS) (Gilks et al., 1995) within Gibbs Sampling. The detailed procedures of our ARMS algorithm are provided in the Appendix B.1. The ARMS algorithm for Matched Case Crossover studies is summarized in Algorithm 1.

4.3 MCMC Sampling Scheme

In this section, we describe how to sample $\{\tilde{\boldsymbol{\alpha}}, \boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(k-1)}\}$, q_0, and q_z in order to sample the coefficients associated with stratum k, $\boldsymbol{\alpha}^{(k)}$, using the frame of Pölya urn (4.1). The MCMC sampling scheme is summarized in the following steps:

Step 1: Initiate $\boldsymbol{\alpha}^{(1)}$ as $\tilde{\boldsymbol{\alpha}}$ obtained from the full conditional distributions described in section 4.2.3.

Step 2: Sample γ from a prior distribution Unif$[0.01, 0.1, 1, 10, 100]$;

Step 3: Approximately calculate q_0 using ARMS samples by following procedures:

Step 3.1: Sample the coefficients $\alpha_{i,k,s}$ which is the sth ARMS for the parametric coefficients at a given stratum k, $i = 0, 1, \ldots, p$,

$$p(\alpha_{i,k,s}|-) \propto \frac{\exp\left(\frac{-\phi_1 \alpha_{i,k}^2}{2}\right)}{\prod_{t=1}^{T}\left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta(t)\right]\right\}};$$

Step 3.2: Sample the coefficients $\alpha_{p+i,k,s}$, $i = 1, \ldots, L$ from ARMS,

$$p(\alpha_{p+i,k,s}|-) \propto \frac{\exp\left(\frac{-\phi_{k,\alpha} \alpha_{p+i,k}^2}{2}\right)}{\prod_{t=1}^{T}\left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta(t)\right]\right\}};$$

Step 3.3: Construct $\beta_{k,s}(t)$, which is semiparametric time varying coefficient function at iteration s of the MCMC scheme and stratum k,

$$\beta_{k,s}(t) = \alpha_{0,k,s} + \alpha_{1,k,s} t + \ldots + \alpha_{p,k,s} t^p + \sum_{l=1}^{L} \alpha_{p+l,k,s} (t - \xi_l)_+^p;$$

Step 3.4: Calculate $f(Y|\boldsymbol{\alpha})_s^{(k)}$ approximately using ARMS sampling

$$f(Y|\boldsymbol{\alpha})_s^{(k)} \approx \frac{N(\alpha_{0,k,s}; 0, \phi_1^{-1}) \prod_{j=1}^{M+1} N(\alpha_{j,k,s}|0, \phi_k^{-1})}{\prod_{t=1}^{T}\left\{1 + \sum_{j=2}^{M} \exp\left[(x_{jkt} - x_{1kt})\beta_{k,s}(t)\right]\right\}};$$

Step 3.5: Repeat Step 3.1-Step 3.5, for $s = 1, \ldots, S$;

Step 3.6: Calculate q_0 approximately using MC;

$$q_0 \approx \frac{1}{S}\sum_{s=1}^{S} f(Y|\boldsymbol{\alpha})_s^{(k)}$$

Step 4: Obtain q_z, $z = 1, \ldots, k-1$;

$$q_z = \frac{1}{\prod_{t=1}^{T}\left\{1+\sum_{j=2}^{M}\exp\left[(x_{jkt}-x_{1kt})\beta_z(t)\right]\right\}},$$

where $\beta_z(t) = \alpha_0^{(z)} + \alpha_1^{(z)}t + \ldots + \alpha_p^{(z)}t^p + \sum_{l=1}^{L}\alpha_{p+l}^{(z)}(t-\xi_l)_+^p$, $\boldsymbol{\alpha}_1^{(z)} = (\alpha_0^{(z)}, \alpha_1^{(z)}, \ldots, \alpha_p^{(z)})$ and $\boldsymbol{\alpha}_2^{(z)} = (\alpha_{p+1}^{(z)}, \ldots, \alpha_{p+L}^{(z)})$, as defined in section 4.2.1;

Step 5: Calculate probabilities for Pölya urn (4.1);

Step 6: With probability of $\gamma q_0/(\gamma q_0 + q_1 + \cdots + q_{k-1})$, $\boldsymbol{\alpha}^{(k)}$ is $\tilde{\boldsymbol{\alpha}}$; with probability of $q_{k-1}/(\gamma q_0 + q_1 + \cdots$ $\boldsymbol{\alpha}^{(k)}$ is $\tilde{\boldsymbol{\alpha}}^{(k-1)}$;

Step 7: Sample $\phi_{k,\alpha}$ from the following Gamma distribution

$$\phi_{k,\alpha} \propto \text{Gamma}\left[\text{L}+\text{u}, \left(\sum_{\text{l}=1}^{\text{L}}\left(\alpha_{\text{p+l}}^{(\text{k})}\right)^2/2 + 1/\text{v}\right)^{-1}\right].$$

Step 8: Repeat Step 2 - Step 7 until the MCMC chain converges.

4.4 Cluster Determination

The obtained functional clustering estimated from the Pölya urn often yields a large number of clusters which may not be informative. Hence in this section, we provide a procedure to determine an informative clustering scheme.

Let $\alpha_{i,k} = \{\alpha_{i,k,1}, \ldots, \alpha_{i,k,S}\}$ represent the set of all sampled values for coefficient $\alpha_i^{(k)}$ where $i = 1, \ldots, p$ of stratum and $k = 1, \ldots, N$. Let $\alpha_{p+l,k} = \{\alpha_{p+l,k,1}, \ldots, \alpha_{p+l,k,S}\}$ represent the set of all sampled values for coefficient $\alpha_{p+l}^{(k)}$ where $l = 1, \ldots, L$ of stratum and $k = 1, \ldots, N$. All these values are sampled based on the Pölya urn from section 4.3. We further define $\hat{\alpha}_i^{(k)}$ and $\hat{\alpha}_{p+l}^{(k)}$ be the estimates of $\alpha_i^{(k)}$ and $\alpha_{p+l}^{(k)}$. Then their estimates are calculated by taking the sample means, that is,

$$\hat{\alpha}_i^{(k)} = \frac{1}{S}\sum_{s=1}^{S} \alpha_{i,k,s}$$
$$\hat{\alpha}_{p+l}^{(k)} = \frac{1}{S}\sum_{s=1}^{S} \alpha_{p+l,k,s}$$

where $i = 1, \ldots, p$ and $l = 1, \ldots, L$.

Let $\hat{\boldsymbol{\alpha}}^{(\boldsymbol{k})}$ be the row vector containing all coefficient estimates for stratum k, that is, $\hat{\boldsymbol{\alpha}}^{(\boldsymbol{k})} = \left(\hat{\alpha}_0^{(k)}, \hat{\alpha}_1^{(k)}, \ldots, \hat{\alpha}_p^{(k)}, \hat{\alpha}_{p+1}^{(k)}, \hat{\alpha}_{p+L}^{(k)}\right)$. By collectin all the estimated stratum coefficients $\hat{\boldsymbol{\alpha}}^{(\boldsymbol{k})}$, $k = 1, \ldots, N$, we then obtain matrix $A = \left(\hat{\boldsymbol{\alpha}}^{(\boldsymbol{1})}, \hat{\boldsymbol{\alpha}}^{(\boldsymbol{2})}, \ldots, \hat{\boldsymbol{\alpha}}^{(\boldsymbol{N})}\right)^T$, where A is $N \times (1 + p + L)$ matrix.

We then obtain the number of cluster by the following the procedure:

Step 1: Apply the k-means clustering algorithm (Hartigan, 1975) to matrix Ausing the following the k means algorithm:

Step 1.1 Consider a set number of clusters, k^*, where $k^* = 1, 2, \ldots, \sqrt{N/2}$ and initialize k^* means, denoted by $\boldsymbol{m_1}$, $\boldsymbol{m_2}, \ldots, \boldsymbol{m_{k^*}}$, which are initialized by randomly choosing k^* random centers;

Step 1.2 Assign each stratum to the cluster that has the smallest within-cluster sum of squares (WCSS), which is defined as the squared value of the euclidean distance

of the observation and the mean of the assigned cluster. That is, calculate

$$WCSS_r = ||\boldsymbol{\alpha}^{(k)} - \boldsymbol{m_r}||^2$$

for $r = 1, \ldots, k^*$, and assign the stratum to the cluster with minimum $WCCSS_r$.

Step 1.3 One the stratum is assigned to a particular cluster, say cluster r^*, calculate the new mean of the cluster, $\boldsymbol{m_{r^*}}$;

Step 1.4 Repeat Steps 1.2 and 1.3 until all points have been assigned;

Step 1.5 Repeat steps 1.1-1.4 until convergence which is reached with the assignments to clusters no longer change;

Step 2: Create a plot of the total within groups sum of squares in the y-axis and number of clusters extracted in the x-axis; once the clustering assignment is done for each of the k^* clusters, proceed the following steps:

Step 2.1 For each k^* clustering assignment, with $k^* = 1, 2, \ldots, \sqrt{N/2}$, calculate the total within cluster sum of squares (TWCSS).

$$TWCSS = \sum_{r=1}^{k^*} WCSS_r$$

Step 2.2 Plot the number of clusters extracted, (i.e., k^*), in the x-axis and the total within groups sum of squares in the y-axis;

Step 3: The final number of clusters is determined by examining the plot and choosing the number of clusters at which the plot presents a bend, followed by a flat trend.

4.5 Simulation Study

We conducted simulations to understand the performance of our functional semiparametric Bayesian approach. Two simulation cases are considered in terms of the number of clusters. We consider a 1-4 matched case-crossover study motivated by our childhood aseptic meningitis example. We set the number of stratum to 100 ($K = 100$) and the number of time periods to 100 ($T = 100$). The covariates X, were generated from a standard normal distribution N(0,1).

The 1- 4 matched case-controls are generated from the prospective model

$$Pr(Y_k = 1|X_k, t, q) = \frac{\exp\{X_k\beta(t) + q(\bullet)\}}{1 + \exp\{X_k\beta(t) + q(\bullet)\}},$$

where $q(\bullet)$ is an unknown effect which disappears under the condition $\sum_{j=1}^{5} Y_j = 1$ for each stratum k, $k = 1, \ldots, N$ regardless of the distribution of $q(\bullet)$.

The following two cases were simulated:

Case 1: K=100, T=100, the number of cluster = 2, where 50% of the strata are in cluster 1 and 50% in cluster 2.

$$\beta(t) = \begin{cases} cos(\pi t) & \text{if cluster 1} \\ t^2 & \text{if cluster 2} \end{cases}$$

Case 2: K=100, T=100, the number of cluster = 3, where 33% of the strata are in cluster

1, 33% in cluster 2 and 34% are in cluster 3.

$$\beta(t) = \begin{cases} cos(\pi t) & \text{if cluster 1} \\ t^2 & \text{if cluster 2} \\ 0.25t & \text{if cluster 3} \end{cases}$$

We consider these two cases because one cluster has the nonparametric curve but the other(s) has(have) the parametric curve. We randomly assign the cluster label into strata. For the estimation and method evaluation, we simulated 50 datasets of 1-4 matched case-crossover studies.

4.5.1 Case 1: two clusters

For each stratum, we first apply the semiparametric time varying coefficient model from Chapter 2 which is a parametric Bayesian approach under $F = F_0$. Figure 4.1 presents an example of the estimated time varying coefficients using a parametric Bayesian approach. By examining the individual estimations, it is evident that there are two underlying cluster structures, one group of functions increasing, and the other group of functions decreasing, which is expected given the data generating structures $\beta(t) = \cos(\pi t)$ and $\beta(t) = t^2$.

We then applied our functional semiparametric Bayesian approach. Figure 4.2 displays the plot used to determine the number of clusters. This plot presents the graphical representation of the total within sum of squares against the number of extracted clusters. Two clusters are selected because it is the location at which the plot presents a bend, followed by a flat trend.

In table 4.1, we can see that at least in 90% of the generated datasets the correct number of

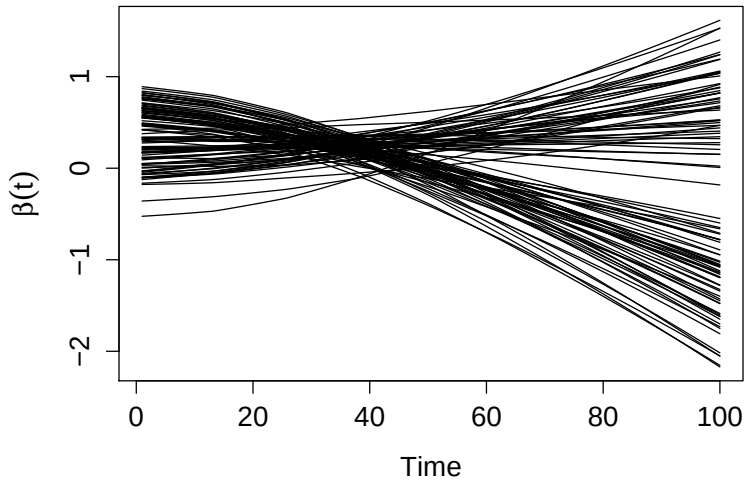


Figure 4.1: Simultaneous representation of $\beta(t)$ for an example dataset from Case 1 simulations.

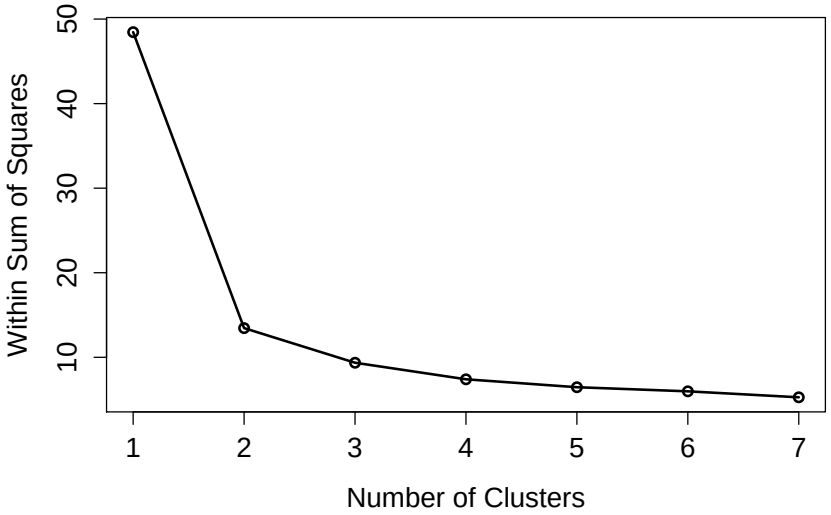


Figure 4.2: Graphical representation of the total within sum of squares vs. the number of extracted clusters for an example dataset from Case 1 simulations.

clusters were selected. For those datasets the minimum success cluster probability allocation is 0.92 for $\gamma = 0.01$, and increases as the value of γ increases because the true number of

cluster is small. When γ is small, a large number of automatically selected clusters is often created. These simulation results suggest that our functional semiparametric Bayesian time varying coefficient model is pretty accurate at capturing the underlying cluster structure.

γ	The proportion of case where correct number of cluster was estimated	The proportion of strata that were correctly assigned
0.01	0.90	0.9211
0.1	0.94	0.9554
1	0.94	0.9375
10	0.98	0.9682
100	0.98	0.9850

Table 4.1: Proportions of cases where the correct number of clusters was estimated and of strata that were correctly assigned using our functional semiparametric Bayesian approach when the true number of cluster is three and the true strata is assigned under simulation Case 1.

The estimated the time varying coefficient function obtained from using our functional semiparametric approach is displayed in Figure 4.3. In part (a) and (c), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric estimation of the time varying function in the cluster. In part (b) and (d) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric estimation of the time varying function in the cluster. Parts (a) and (b) present the estimated functions for cluster 1 and parts (b) and (d) present the estimated functions for cluster 2. It is clear that the underlying structure of the simulated data was successfully captured for this dataset.

4.5.2 Simulation Case 2: Three Clusters

We randomly selected a dataset from this simulation set up to present as an example. Figures 4.4-4.6 present an example estimation for this selected dataset. Figure 4.4 presents the

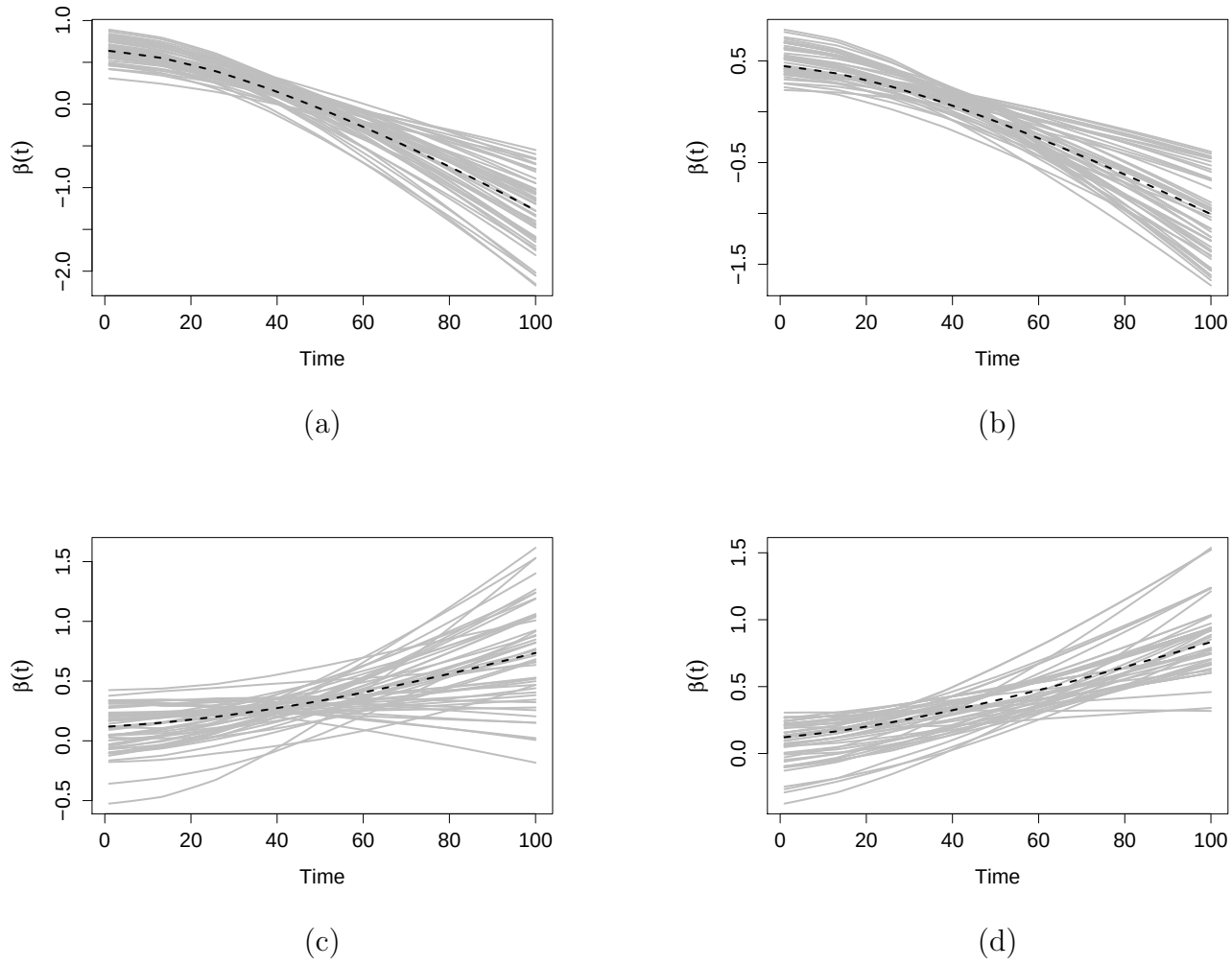


Figure 4.3: Estimation of the time varying coefficient function for each of the two extracted clusters using parametric and nonparametric Bayesian approaches: In part (a) and (c), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric final estimation of the time varying function in the cluster. In part (b) and (d) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric final estimation of the time varying function in the cluster. Parts (a) and (b) present the estimated functions for cluster 1 and parts (b) and (d) present the estimated functions for cluster 2.

simultaneous representation of the parametric estimation of the dataset. In this figure, we can see that there are three underlying cluster structures in the data.

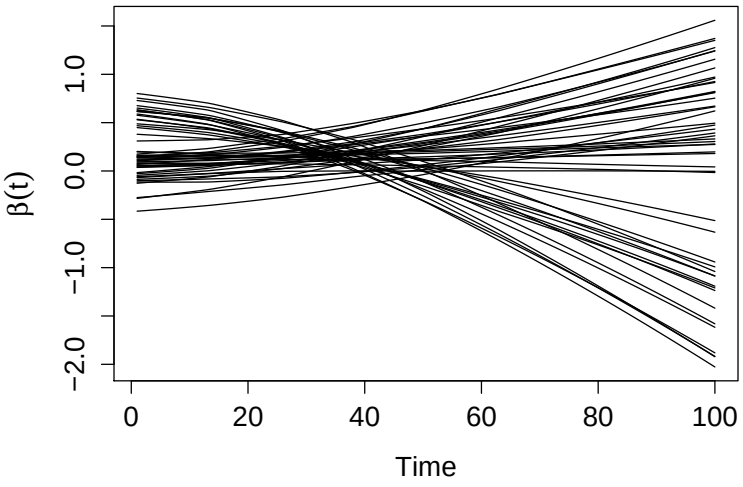


Figure 4.4: An example of varying coefficient $\beta(t)$ generated from simulation case 2.

We then applied our functional semiparametric Bayesian approach. Figure 4.5 presents the plot used to determine the number of clusters for this example dataset. Based on this figure, we should extract three clusters, since at this location the plot presents a bend, followed by a flat trend.

Accuracy of the estimation was also evaluated using the proportion of correct cluster allocation. Table 4.2 provides the proportion of number of cases in which the correct number of clusters (3) was selected. Table 4.2 also presents the proportions of strata that were correctly assigned to the clusters, for the datasets in which the correct number of clusters was selected.

From Table 4.2, we can notice that at least in 80% of the generated datasets the correct number of clusters was selected. For those datasets the minimum success cluster probability allocation is 0.76 for $\gamma = 0.01$ because small values of γ create the larger number of automatically selected clusters because the true number of cluster is small. The success cluster probability is 0.92 when $\gamma = 100$ because the true number of cluster is small. Therefore

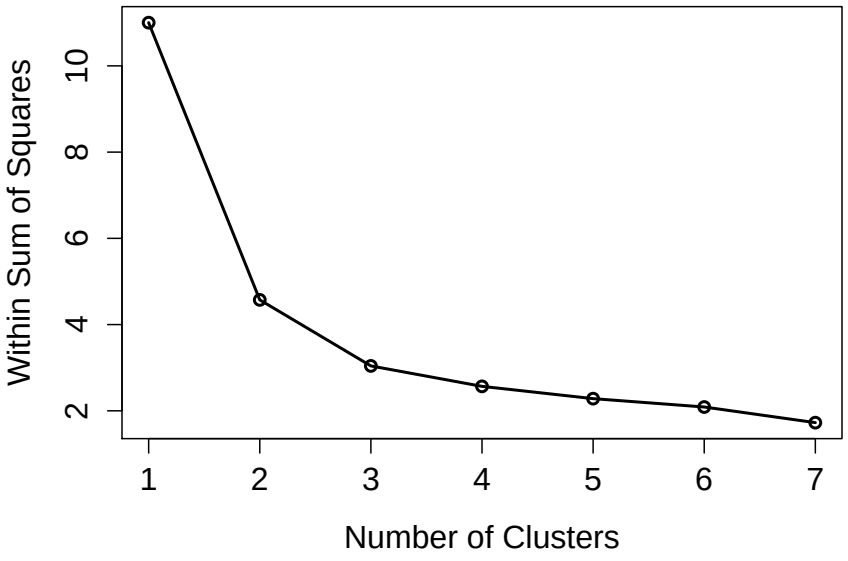


Figure 4.5: Plot of the total within sum of squares vs. the number of extracted clusters obtained from our functional Bayesian approach using a simulated dataset under simulation Case 2.

γ	The proportion of cases where correct number of cluster was estimated	The proportion of strata that were correctly assigned
0.01	0.80	0.7628
0.1	0.82	0.8065
1	0.86	0.8186
10	0.90	0.8645
100	0.92	0.9203

Table 4.2: Proportions of cases where the correct number of clusters was estimated and of strata that were correctly assigned using our functional semiparametric Bayesian approach when the true number of cluster is three and the true strata is assigned under simulation Case 2

the accuracy increases as the value of γ increases. These results show that our functional semiparametric Bayesian approach accurately captures the underlying cluster structure of the data.

Figure 4.6 presents the estimated time varying coefficient function obtained from our func-

tional semiparametric Bayesian approach for each of the three extracted clusters. In parts (a), (c) and (e), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric final estimation of the time varying function in the cluster. In parts (b), (d) and (f) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric final estimation of the time varying function in the cluster. Part (a) and (b) present the estimated functions for cluster 1, part (c) and (d) present the estimated functions for cluster 2 and parts (e) and (f) present the estimated functions for cluster 3. Based on this figure, we conclude that the underlying structure of the simulated dataset was captured using our approach.

Therefore our simulation results support that our functional semiparametric approach can determine the functional clusters among time varying coefficients, understand the functional relationship between predictors and clustered binary outcomes, and assess heterogeneity.

4.6 Example: 1-4 Bi-directional Matched Case-Crossover Study Estimation

Our main goals of interest in this matched case-crossover study are (1) to understand the functional relationship between drinking water turbidity and the relative risk of of aseptic meningitis in children, (2) to explore the modification in these effects by association with the potential latent period and (3) to determine functional clusters of time varying coefficients among strata. We are also interested in assessing in which way the time varying coefficient function changes between clusters. In this study, the stratifying variable is the individual child who is less than 15 years old, diagnosed in one of two years and lives on one of two

urban communities in South Korea. There are 92 children. Our potential effect modification includes a time effect and a potential underlying clustered structure which may be due to two urban communities.

The motivation behind the development of our functional semiparametric Bayesian approach in this example dataset comes from the application of the parametric Bayesian approach. The total number of strata is 92. For each stratum, we apply the parametric Bayesian approach and estimate a semiparametric time varying coefficient model presented in Chapter 2 is estimated for each stratum. Figure 4.7 presents an example of the estimated time varying coefficient for nine randomly selected strata out of 92. This figure 4.7 shows that there are several curve patterns present in the time varying coefficient functions. Figure 4.7 (a), (e), (g) and (h) seem to have an increasing time varying function, while Figure 4.7 (b) and (c) seem to have a curved pattern that increases and later decreases. Since the parametric Bayesian approach requires to estimate time varying coefficient for each strata and provides large number functions, it is difficult to summarize a potential underlying clustered structure.

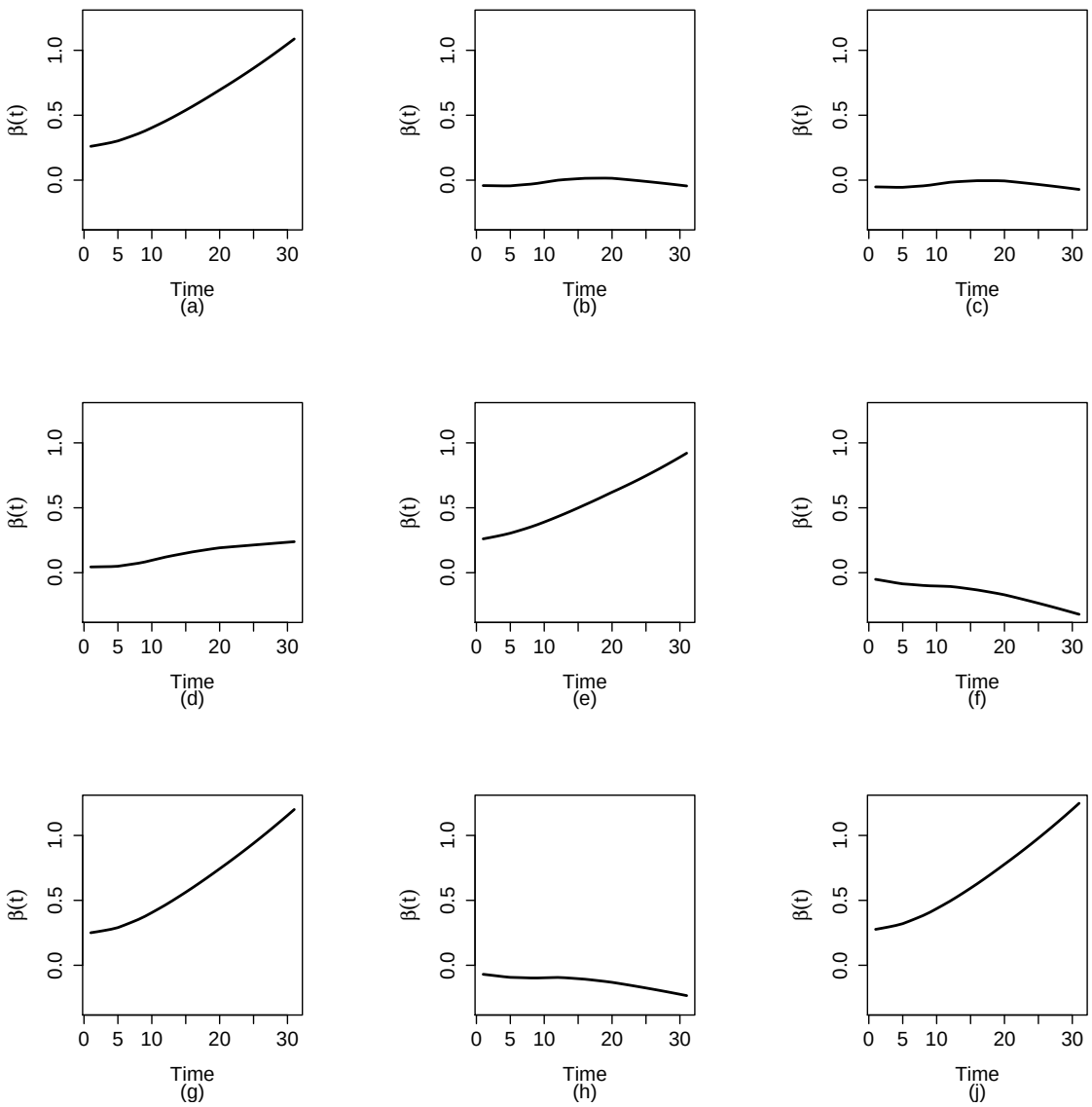


Figure 4.7: Parametric estimation of $\beta(t)$ for nine randomly chosen strata for the example data of 1-4 bi-directional matched case-crossover design.

Figure 4.8 presents the estimation of the parametric approach for ninety two strata simultaneously. Upon examination of this figure, we can observe that there are several underlying cluster structures in the data. Based on this examination we applied our functional semi-parametric Bayesian approach which consists of first using the Pölya urn sampling scheme

and then calculating the k-means clustering approach.

Figure 4.9 displays the plot used to determine the number of clusters for the ninety two selected strata. This plot is a graphical representation of the total within sum of squares against the number of extracted clusters. The final number of clusters is selected, by examining the plot and choosing the number of clusters at which the plot presents a bend, followed by a flat trend. According to this criteria, we select three time varying coefficient clusters for these selected strata. The proportions of three clusters are 31.5%, 54.35%, and 14.15%. Cluster 1 consists of 29 strata from different regions and years. Cluster 2 contains 50 strata, all the region 1 collected in year 2002 and most of region one collected in year 2001. Cluster 3 consists of 13 strata, from different regions but all from year 2002.

Figure 4.10 exhibits the estimated time varying coefficient function for each of the extracted clusters. In parts (a), (c) and (e), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric final estimation of the time varying function in the cluster. In parts (b), (d) and (f) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric final estimation of the time varying function in the cluster. Part (a) and (b) present the estimated functions for cluster 1, part (c) and (d) display the estimated functions for cluster 2 and parts (e) and (f) present the estimated functions for cluster 3.

Figure 4.10 provides three distinct trends among the estimated time varying coefficient functions, each one captured by its respective final estimated function. In cluster 1, $\hat{\beta}(t)$ is increasing until 15th days and then decreasing. This means that the risk of aseptic meningitis is increasing until 15th days but afterwards it decreases. In cluster 2, $\hat{\beta}(t)$ is overall rapidly decreasing which means that the risk of aseptic meningitis is decreasing a time passes. However, in cluster 3, $\hat{\beta}(t)$ is increasing which means that the risk of aseptic meningitis is increasing as days go by.

Hence, these three clusters are considered as the effect modification. Without using our functional semiparametric Bayesian approach, we can not detect these modifications. Therefore, our functional semiparametric Bayesian approach allows us to achieve our three goals and provide the following three results. First we found the functional relationship between drinking water turbidity and the relative risk of of aseptic meningitis in children using the time varying coefficient function as time passes. Second, since the estimated time varying coefficients among three clusters are different, these three clusters can represent the modification in these effects by association with the potential latent period. Third, cluster one contains individuals for which the effect of water turbidity over the presence of aseptic meningitis in children seems to remain constant over time, cluster two contains individuals for which the effect of water turbidity over the presence of aseptic meningitis seems to decrease over time, and cluster three contains individuals for which the effect of water turbidity over the presence of aseptic meningitis seems to increase over time.

4.7 Discussion

In this chapter, we have proposed a functional semiparametric Bayesian time varying coefficient model for matched case-crossover studies. This method allows us to detect parametric relationships between a covariate of interest and the binary outcome of presence or absence of disease, semiparametric relationships between the covariate and time, as well as obtain functional clusters of time varying coefficients among strata. Our model was developed under a nonparametric Bayesian approach constructed with Dirichlet process priors, which automatically cluster subpopulations among strata and assess heterogeneity.

The results of the presented simulation studies show that the method performs well in estimating the unknown time varying functions, by capturing the underlying structure of the

data. Simulation results also show, that the proposed method performs well at identifying underlying clustering structures by determining the correct number of clusters and assigning the correct strata to the extracted clusters.

We present an application for a 1-4 bi-directional matched case-crossover study in which the condition of interest is aseptic meningitis in children under fifteen years of age, collected in South Korea, in which the covariate of interest is water turbidity. The estimation of a reduced subset of ninety two patients yielded the construction of three clusters. Cluster one contains individual for which the effect of water turbidity over the presence of aseptic meningitis in children seems to remain constant over time , cluster two contains individuals for which the effect of water turbidity over the presence of aseptic meningitis seems to decrease over time, and cluster three contains individuals for which the effect of water turbidity over the presence of aseptic meningitis seems to increase over time.

We note that although our approach is developed under the nonparametric Bayesian approach and has allowed us to automatically estimate the cluster, it often provides a large number of clusters which may not be informative. Hence we use the k-means clustering approach to determine an informative clustering scheme. This may be influenced by the estimation of γ, which is based on the prior distribution, and not the posterior distribution. We may further use a Gamma prior for γ to further study the determination of the number of clusters for this approach.

This problem may be solved by using weighted Dirichlet process (WDP). We will conduct further research, to develop our approach using weighted Dirichlet process (WDP), without using k-means clustering. However, the extension of our approach to WDP will require heavy computational burden and derivation of marginal likelihood. We consider further research on WDP to be worthwhile.

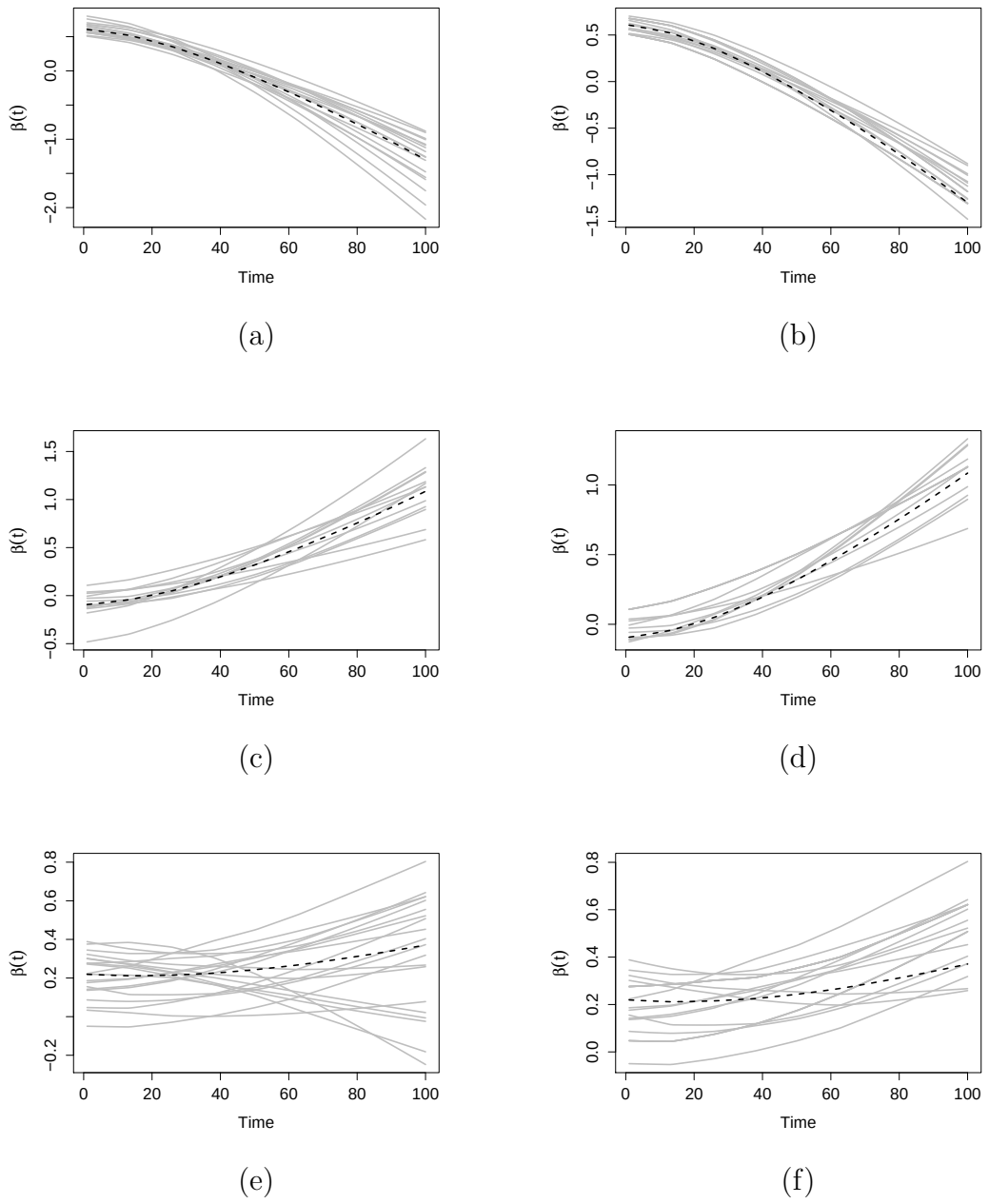


Figure 4.6: Estimation of the time varying coefficient function for each of the three extracted cluster using parametric and nonparametric Bayesian approaches for simulations in Case 2: In parts (a), (c) and (e), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric final estimation of the time varying function in the cluster. In parts (b), (d) and (f) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric final estimation of the time varying function in the cluster. Part (a) and (b) present the estimated functions for cluster 1, part (c) and (d) present the estimated functions for cluster 2 and parts (e) and (f) present the estimated functions for cluster 3.

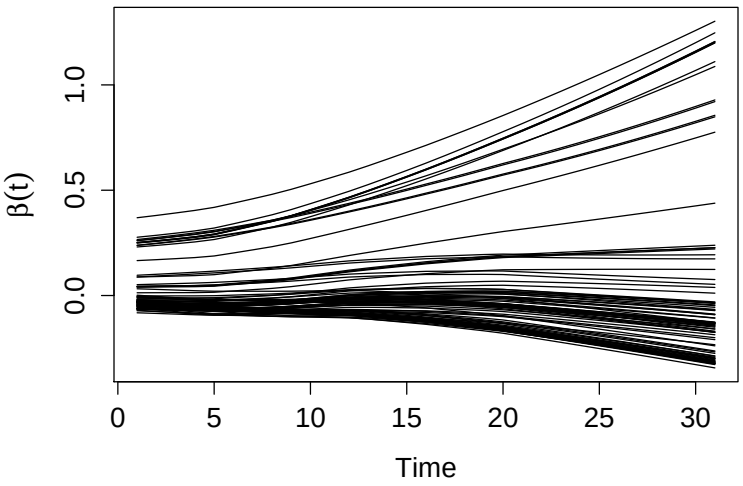


Figure 4.8: Simultaneous representation of $\beta(t)$ for ninety two strata from the example data of 1-4 bi-directional matched case-crossover design.

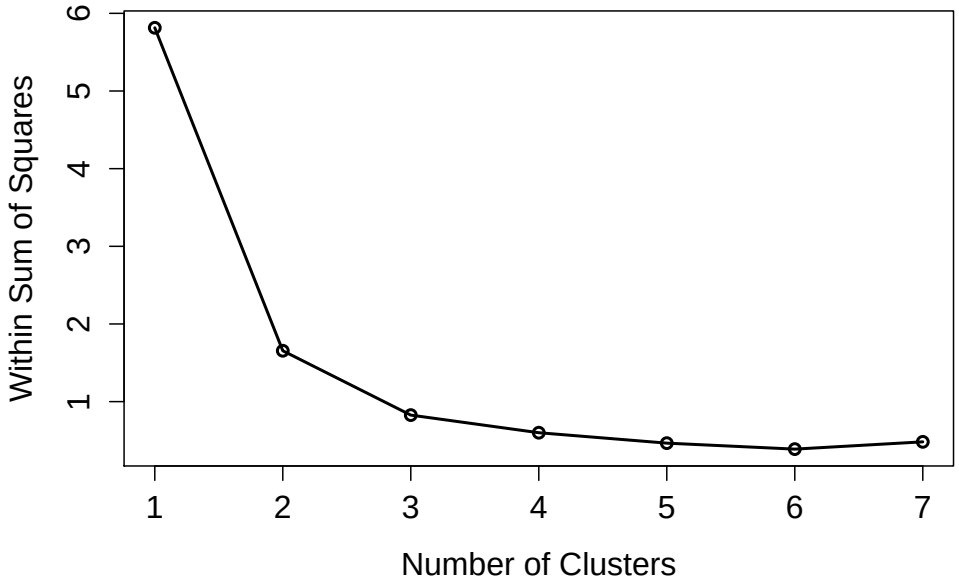


Figure 4.9: Plot of the total within sum of squares vs. the number of extracted clusters for ninety two strata in the example data of 1-4 bi-directional matched case-crossover design.

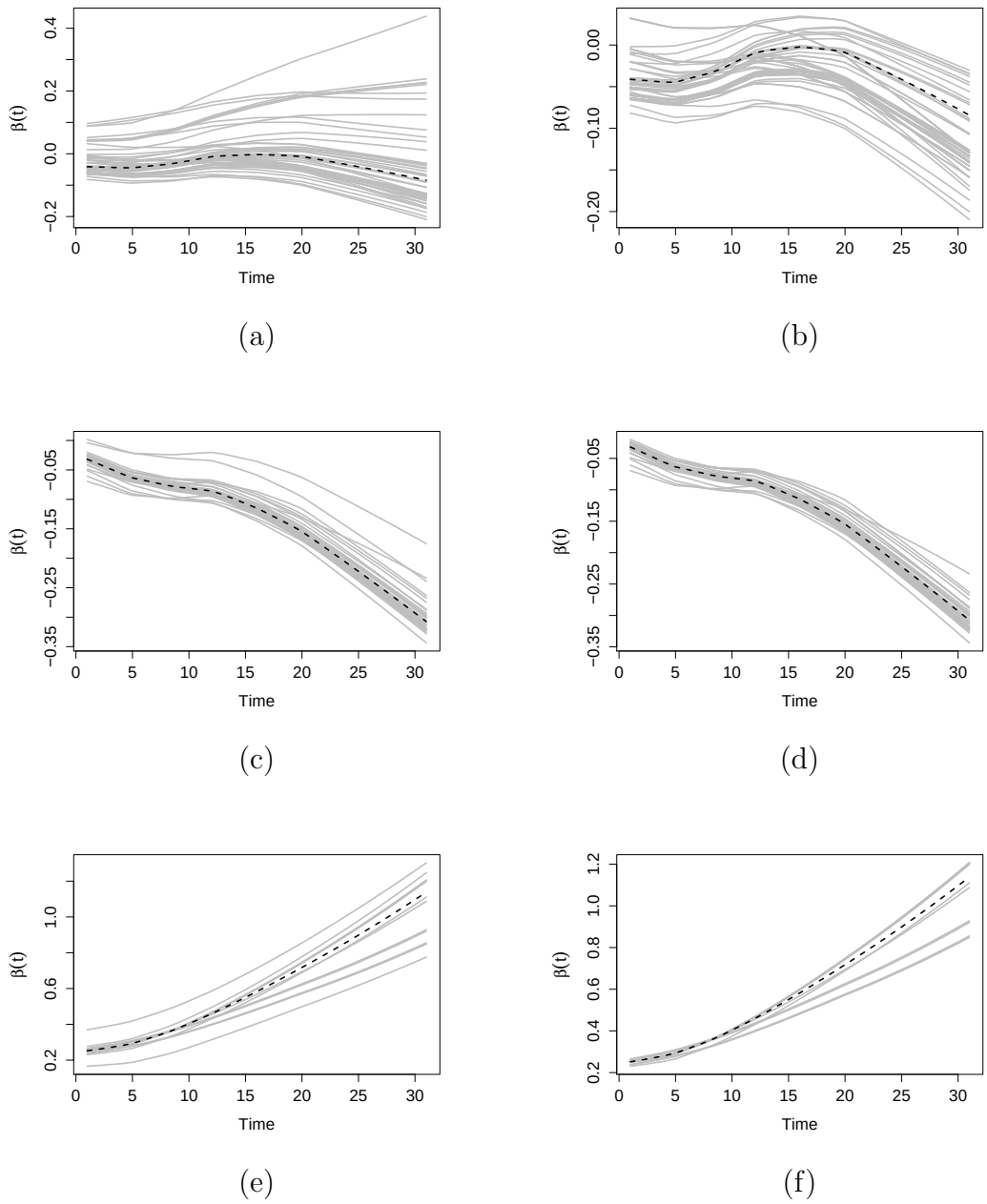


Figure 4.10: Estimation of the time varying coefficient function for each of the three extracted clusters for the example data of 1-4 bi-directional matched case-crossover design. In parts (a), (c) and (e), the gray lines represent the parametric estimations for the strata in the cluster, and the dashed black line represents the nonparametric final estimation of the time varying function in the cluster. In parts (b), (d) and (f) the gray lines represent the nonparametric estimations for the strata in the cluster, and the black line represents the nonparametric final estimation of the time varying function in the cluster. Part (a) and (b) present the estimated functions for cluster 1, part (c) and (d) present the estimated functions for cluster 2 and parts (e) and (f) present the estimated functions for cluster 3.

Chapter 5

General Conclusions and Further Research

Major conclusions and contributions of this dissertation are summarized in this chapter and possible future research areas are introduced.

5.1 General Conclusions and Contributions

In Chapter 2, we proposed three different methodologies to detect relationships between a predictor and the relative risk of disease or binary outcome, as well as detect effect modifications by time. We presented model comparison in terms of estimation accuracy as well as prediction accuracy.

The first method is a parametric two-stage approach. The first stage consists on fitting a conditional logistic regression at each time period. The end result of this stage is an effect coefficient for each time. The second stage, consists on fitting a p-th polynomial

regression regression of the coefficients (response) on time (predictor). Parametric tests can be performed to determine whether an effect modification from time exists.

This two-stage approach can detect parametric relationships between the predictor and binary outcomes and parametric relationships between the predictor an time. An advantage of this method is that it is simple and easy to use. However, a disadvantage of this method is that it only detects parametric relationships between the predictor and time.

The second method, is a semiparametric one-stage penalized approach. The effect of a covariate over the binary outcome is evaluated via p-th order regression splines with a truncated power series basis. All unknown parameters are estimated using a penalized log-likelihood approach. The smoothing parameter associated with the penalty is obtained by using a grid search within a preselected rang, minimizing the AIC and AICc criteria.

This one-stage penalized approach allows us to detect parametric relationships between the predictor and binary outcomes and simultaneously evaluate the semiparametric relationship between the predictor and time. The advantage of this method is that is more flexible than the parametric two-stage approach. However, the disadvantage is that its performance affected by the choice of smoothing parameter. In order to overcome this disadvantage, we further develop the semiparametric one-stage Bayesian approach.

The third method, is a semiparametric one-stage approach developed under the Bayesian hierarchical model framework. We also evaluate the effect of a covariate over the binary outcome is evaluated via p-th order regression splines with a truncated power series basis. However, the choice of smoothing parameter is selected in the entire domain of the prior distribution, allowing the estimation of parameters to be more accurate and efficient than the one-stage penalized approach.

In addition to the effect estimation, we also provided hypothesis testing for the time varying

coefficient, via Bayes Factors. This hypothesis test allows us to determine whether an effect modification arising from time exists. The advantages of this method include the more accurate and flexible estimation of the parameters as well as the ability to conduct hypothesis tests via Bayes Factors.

Simulation results showed that the semiparametric time varying coefficient model under the Bayesian framework performs best in terms of estimation and prediction accuracy. Additionally the Bayesian alternative automatically estimates the smoothing parameter value in the domain of the prior distribution providing this approach with more flexibility than its penalized counterpart.

Additionally we provided estimations of a 1-4 bi-directional matched case-crossover study in which the covariate of interest is water turbidity and the condition of interest aseptic meningitis in children under 15 years of age. The results of the Bayesian approach showed that there is a modification in the effect of water turbidity over the presence or absence of aseptic meningitis in children. Additionally, we concluded that the peak risk of disease was the fourth day of hospital admission.

In Chapter 3, we proposed two different methodologies to detect relationships between a predictor and the relative risk of disease or binary outcome, as well as detect effect modifications by time and spatial location. The choice between these methods depends on the number of locations from which the data was collected.

When the number of locations among the subjects is small, we propose a semiparametric spatial-temporal model for a small number of locations in which the effect modification due to time is explored by using the semiparametric time varying coefficient model under the Bayesian framework presented in Chapter 2. The effect modification due to spatial location among the subjects is explored by examining the relationship between the estimated function

and its first derivative with respect to time, within and between the different locations.

We provided the details of testing procedures to determine whether an effect modification due to time exists. The time effect test is conducted by examining the first derivative of the semiparametric time varying coefficient function with respect to time and its credible bounds. We also provided details for testing procedures to determine whether an effect modification due to spatial location exists. This spatial effect test is conducted by examining credible bounds of the pairwise differences between the semiparametric time varying coefficient functions at the different locations.

Simulation results showed that the proposed method performed well in estimating the unknown time varying coefficient functions, determining whether this functions vary through time significantly, as well as detecting differences in effected modification for subjects located in different regions. We also considered the example data of 1-4 bi-directional matched case-crossover study in which the covariate of interest is water turbidity and the condition of interest is aseptic meningitis in children under 15 years of age located in two different provinces in South Korea. We applied our approach and concluded that there is an effect modification arising from the spatial location among the subjects. For region 1, there is no effect modification due to time, however, for region 2, an effect modification due to time exists.

When the number of locations among the subjects is medium to large, we propose a semiparametric spatial-temporal model in which the effect modification due to time is also explored using the semiparametric time varying coefficient model from Chapter 2, but the effect of spatial location among the subjects is examined by using a random effect and the estimation of an empirical semivariogram.

This spatial-temporal semiparametric varying coefficient model was developed under the

Bayesian hierarchical model framework. Due to identifiability issues, the precision parameter associated with the Gaussian covariance structure of the random effect is estimated using the empirical semivariogram using the residuals of either a conditional logistic regression or a logistic regression.

This method allows the detection of the parametric relationship between the predictor and binary outcomes, the semiparametric relationship between the predictor and time and potential effect modifications due to spatial location among the subjects.

Simulation results show that the spatial-temporal semiparametric varying coefficient model performs well in estimating the unknown time varying function, spatial parameters and finding a suitable smoothing parameter for the semiparametric estimation. Additionally, simulation results show that empirical semivariogram estimations using conditional logistic and logistic regression present similar results. However, the logistic regression estimation method is faster and easier to implement than the conditional logistic, for empirical semivariogram estimation.

In chapter 4, we proposed a functional semiparametric Bayesian time varying coefficient model for matched case crossover studies. The main goals of this method are to understand the functional relationship between a covariate of interest and the binary outcome that represents the presence or absence of disease, to explore effect modification due to time as well as to determine functional clusters if time varying coefficients among strata.

The proposed method is a nonparametric Bayesian approach constructed with Dirichlet process priors with a Pölya Urn sampling scheme, in combination with a k-means clustering procedure for the semiparametric time varying coefficient model which is developed using p-th order regression splines.

Simulation results showed that the proposed model performs well at estimating the unknown

time varying functions by capturing the underlying structure of the data and at identifying the correct number of clusters and assigning the correct strata to the extracted clusters. In addition, we applied the proposed method to the example data of1-4 bi-directional matched case crossover study, in which the condition of interest is aseptic meningitis and the covariate of interest is drinking water turbidity. The application of this approach yielded the identification and assignment of the data to three clusters, each one with a specific structure of time varying coefficient function.

Section 5.2 presents different ways in which the research presented in this dissertation can be extended.

5.2 Further Research

The work in this dissertation can be extended in several ways:

1. *Extension of the semiparametric spatial-temporal model to be more flexible*:

   This flexible model can be developed to detect nonparametric relationships between the predictor and binary outcomes and between the predictor and time. The nonparametric spatial-temporal model provides the ability to detect only parametric relationships between our predictor and binary outcome as well as parametric and nonparametric relationships between our predictor and time through $\beta(t)x_{jk}$, $j = 1, \ldots, M+1$, $t = 1, \ldots T$. Hence this extension will allow us to also detect nonparametric relationships between our predictor and binary outcome as well as between our predictor and time.

   By replacing $\beta(t)x_{jk}$ by the following nonparametric function $m(\cdot, \cdot)$,

$$m(x_{jk}, t_k) = \beta_1(t)x_{jk} + \cdots + \beta_{p_x}(t)x_{jk}^{p_x} + \sum_{l_x=1}^{L_x} \beta_{p+l_x}(t)(x_{jk} - \xi_{l_x}^x)_+^{p_x}$$

where

$$\beta_i(t_k) = \alpha_0^{(i)} + \alpha_1^{(i)} t_k + \cdots + \alpha_{p_i}^{(i)} t^{p_i} + \sum_{r=1}^{R_i} \alpha_{p_i+1}^{(i)} (t_k - \xi_r^{(i)})_+.$$

Here p_x denotes the order of the polynomial regression for X, $i = 1, \ldots, p_x$, p_i represents the order of the polynomial regression for each $\beta_i(t_k)$, $\xi_{l_x}^x$ are the corresponding knots for X. $\beta_i(t_k)$ represents the semiparametric time varying coefficient function as defined in 2 with knots $\xi_r^{(i)}$, parametric coefficients $\alpha_0^{(i)}, \alpha_1^{(i)}, \ldots, \alpha_{p_i}^{(i)}$ and nonparametric coefficients $\alpha_{p_i+1}^{(i)}, \alpha_{p_i+2}^{(i)} \ldots, \alpha_{p_i+R_i}^{(i)}$.

We have developed this method under Bayesian hierarchical model framework and have studied this model via a small simulation which is not included in this dissertation. Further intensive simulation works need to be done to understand the performance of this model in different situations before applying this model to an example dataset.

2. *Derivation of the relationship between prospective and retrospective models for matched case-crossover data*:

   We have developed our semiparametric time varying coefficient model under the retrospective model. However, this model can be developed under the prospective model too. We have not yet studied whether they are equivalent and under which conditions this equivalency holds. Theoretical derivation is required to provide this relationship and is further validated using simulation studies.

3. *Sensitivity analysis of prior distributions on γ in the nonparametric Bayesian approach*:

   We note that although our approach is developed under the nonparametric Bayesian approach and is allowed us to automatically estimate the cluster, it often provide the large number of clusters which may not be informative. Hence we use the k means clustering to determine an informative cluster. This may be, because the estimation of

γ is based on a prior distribution, not the posterior distribution. We may further use the Gamma prior of γ to further study the determination of the number of clusters.

4. *Extension of our functional semiparametric Bayesian approach to be more flexible:* We need to refine the clustering scheme presented in the functional semiparametric Bayesian time varying coefficient model for matched case-crossover studies. We used the k-means clustering procedure to estimate the informative number of cluster. Our further research is to develop our approach using weighted Dirichlet process (WDP) without using k mean clustering. However, the extension of our approach to WDP will require a heavy computational burden and derivation of marginal likelihood. Further research will using WDP will be worthwile.

5. *Extension of our functional semiparametric Bayesian approach to the functional spatial-temporal semiparametric Bayesian approach:* We have developed the functional semiparametric Bayesian approach under semiparametric time varying coefficient model. However it will be very useful if we further develop it under spatial-temporal model. The spatially and temporally correlated data is very popular. As a result, the research is rapidly growing. Hence the development of our approach to be used in spatial-temporal data is worthwhile future research.

Bibliography

Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2, 11521174.

Albert, P. S. and McShane, L. M. (1995) A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data. *Biometrics*, 51, 2, 627-638.

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Polya urn Schemes. *The Annals of Statistics*, 353-355.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 2, 209230.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistics Association*, 85, 398-409.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 6, 721-741.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995) Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 455-472.

BIBLIOGRAPHY

Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.

Hartigan, J. A. (1975). *Clustering Algorithms* New York: John Wiley and Sons.

Hastie, T. J. and Tibshirani, R. J. (1993) Varying coefficients models. *Journal of the Royal Statistical Society Series B*, 55, 757-96.

Hjort, N., Holmes, C., Müller, P., and Walker, S. G. (2010) *Bayesian nonparametrics: principles and practice*. Cambridge University Press.

Hosmer Jr, D. W., and Lemeshow S. (1989). *Applied logistic regression*. John Wiley and Sons.

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model. *Journal of the American Statistical Association*, 99, 465.

Jeffreys, H. (1961). Theory of Probability (3rd ed.), Oxford, U.K.: Oxford University Press.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773-795.

Kim, H., Cheong, H. K., Park, S. K., and Bae, G. R.. (2003). Drinking Water Turbidity and Aseptic Meningitis in Children in An Urban Community in Korea: Isee-567. *Epidemiology*, 14(5), S110.

Metropolis, M., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1087-1093.

BIBLIOGRAPHY

Navidi W. (1998).Bidirectional case-crossover designs for exposure with time trends. *Biometrics*, 54, 596-605.

Park, S. K., Park, B., Ki, M., Kim, H., Lee, K., Jung, C., ... and Cheong, H. K. (2010). Transmission of seasonal outbreak of childhood enteroviral aseptic meningitis and hand-foot-mouth disease. *Journal of Korean medical science*, 25(5), 677-683.

Ray, S. and Mallick, B. (2006). Functional clustering by Bayesian wavelet methods. . *Journal of the Royal Statistical Society*, Series B, 68, 2, 305-332.

Ruppert, David, Matt P. Wand, and Raymond J. Carroll (2003). Semiparametric regression. No. 12. Cambridge university press, 2003.

Weinberg, M. D. (2012). Computing the Bayes Factor from a Markov Chain Monte Carlo Simulation of the Posterior Distribution. *Bayesian Analysis*, 7, 3, 737-770.

Appendix A

Gradient, Hessian and ARMS for Time Varying Coefficient Model

In Section A.1-A.2, we summarize the Gradient and Hessian matrix for parametric two stages and semiparametric one stage approach, respectively. In Section A.3, we provide the arms procedure for semiparametric one stage Bayesian approach.

A.1 Gradient and Hessian for two stage approach

Gradient $l'(\beta)$ and Hessian value $l''(\beta)$ with respect to β for the parametric two-stage approach are calculated as follows;

$$\begin{aligned} l'(\beta) &= \sum_k x_{1k} - \sum_k \left\{ \frac{\sum_{j=1}^{M} x_{jk}\exp(x_{jk}\beta)}{\sum_{j=1}^{M}\exp(x_{jk}\beta)} \right\}; \\ l''(\beta) &= -\sum_k \left[\frac{\sum_{l=1}^{M-1} \left\{ \sum_{j=l+1}^{M} \exp(x_{jk}\beta)\left(x_{lk}^2 - 2x_{lk}x_{jk} + x_{jk}^2\right) \right\} \exp(x_{lk}\beta)}{\left\{ \sum_{j=1}^{M}\exp(x_{jk}\beta) \right\}^2} \right]. \end{aligned}$$

We estimate parameters via Newton Raphson method with these gradient and Hessian value.

A.2 Gradient and Hessian for one stage approach

On the other hand, Gradient and Hessian with respect to the parameters for the one-stage semiparametric approach are the following. Gradient and Hessian with respect to α_0 are the following

$$
\begin{aligned}
l'(\alpha_0) &= \sum_{k=1}^{N} \frac{\sum_{l=2}^{M} \exp[\{\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p\} x_{lk}^*] x_{lk}^*}{1 + \sum_{l=2}^{M} \exp[\{\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p\} x_{lk}^*]} \\
l''(\alpha_0) &= \sum_{k=1}^{N} \left\{ \frac{\sum_{l=2}^{M-1} \left(\sum_{z=l+1}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{zk}^*\} (x_{lk}^2 \right.}{\left(1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{lk}^*\}\right)^2} \right. \\
&\quad \left. \frac{-2 x_{lk} x_{zk} + x_{zk}^2) + x_{lk}^2 \Big) \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{lk}^*\}}{\left(1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{lk}^*\}\right)^2} \right\}
\end{aligned}
$$

Gradient and Hessian with respect to α_1 are the following

$$
\begin{aligned}
l'(\alpha_1) &= \sum_{k=1}^{N} \frac{\left[\sum_{l=2}^{M} \left\{\exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p] x_{lk}^*\}\right\} x_{lk}^*\right] t_k}{1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p] x_{lk}^*\}} \\
l''(\alpha_1) &= \sum_{k=1}^{N} \left\{ \frac{\left[\sum_{l=2}^{M-1} \left(\sum_{z=l+1}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p] x_{zk}^*\} (x_{lk}^2 \right.\right.}{\left(1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p +] x_{lk}^*\}\right)^2} \right. \\
&\quad \left. \left. \frac{-2 x_{lk} x_{zk} + x_{zk}^2) + x_{lk}^2 \big) \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p] x_{lk}^*\}\right] t_k^2}{\left(1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p] x_{lk}^*\}\right)^2} \right\}
\end{aligned}
$$

Gradient and Hessian with respect to α_{1+l}, $l = 1, ..., L$ are the following.

$$
l'(\alpha_{1+l}) = \sum_{k=1}^{N} \frac{\left[\sum_{l=2}^{M} \left\{\exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{lk}^*\}\right\} x_{lk}^*\right] (t_k - \xi_j)_+}{1 + \sum_{l=2}^{M} \exp\{[\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{l=1}^{L} \alpha_{1+l}(t-\xi_l)_+^p] x_{lk}^*\}} + \alpha_{1+l}\lambda
$$

$$l''(\alpha_1) = \sum_{k=1}^{N} \left\{ \frac{\left[\sum_{l=2}^{M-1}\left(\sum_{z=l+1}^{M}\exp\{[\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{1+l}(t-\xi_l)_+^p]x_{zk}^*\}(x_{lk}^2-2x_{lk}x_{zk}\right.\right.}{\left(1+\sum_{l=2}^{M}\exp\{[\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{1+l}(t-\xi_l)_+^p]x_{lk}^*\}\right)^2}$$
$$\frac{\left.\left.+x_{zk}^2)+x_{lk}^2\right)\exp\{\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{1+l}(t-\xi_l)_+^p]x_{lk}^*\}\right](t_k-\xi_j)_+^2}{\left(1+\sum_{l=2}^{M}\exp\{[\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{l=1}^{L}\alpha_{1+l}(t-\xi_l)_+^p]x_{lk}^*\}\right)^2} \right\} + \lambda$$

We estimate parameters via Newton Raphson method with these gradient vector and Hessian matrix.

A.3 Adaptive Rejection Metropolis Sampling within Gibbs Sampling for Semiparametric one stage Bayesian approach

We implemented Adaptive Rejection Metropolis Sampling within Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith 1990) in order to obtain sequential samples from the posterior distributions of our parameters: $(\alpha_0, \ldots \alpha_p, \alpha_{1+l}, ; l = 1, \ldots, L)$ and ϕ_α. During each iteration of the Gibbs sampler, each of the parameters listed above is updated by a sample drawn from its full conditional distribution. The full conditional distributions for the α parameters provided in section 2.5 are not of closed form and therefore we implemented Adaptive Rejection Metropolis Sampling (ARMS) within the Gibbs sampler to draw samples from their full conditional distributions.

ARMS Gilks et al. (1995) is an extension of Adaptive Rejection Sampling (Gilks and Wilde 1992) with the inclusion of a Metropolis-Hastings (Metropolis *et al.*, 1953) step, using ARS to generate a proposal distribution and evaluating it using Metropolis-Hastings acceptance probabilities. $\{\alpha_0, \ldots \alpha_p, \alpha_{1+l}, ; l = 1, \ldots, L), \phi_\alpha\}$ is the set of variables to be sampled

using the Gibbs sampler. In general let us represent each individual α variable by α^* as this procedure will be repeated for each one. The goal is to sample α^* from its full conditional distribution which is proportional to $f(\alpha^*)$, where $f(\alpha)$ is

$$f(\alpha) = \frac{\phi \exp\left(\frac{-\phi^2\alpha^2}{2}\right)}{\prod_{k=1}^{N}\left(1+\sum_{j=2}^{M}\exp\left[\left\{\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{j=1}^{L}\alpha_{1+j}(t-\xi_j)_+^p\right\}x_{jk}^*\right]\right)}.$$

Here we set $\phi = \phi_1$ for $\alpha_0, \ldots \alpha_p$ and $\phi = \phi_\alpha$ for α_{p+i}, $i = 1, \ldots, L$.

Let α_c^* represent the current value of α^* at a given iteration of the Gibbs sampler. We want to replace α_c^* with a new value α_M^* sampled from f, where f is the posterior distribution of α^*. We create a new function

$$h_n(\alpha^*) = \max\Big(L_{i,i+1}(\alpha^*, S_n), \min\{L_{i-1,i}(\alpha^*, S_n), L_{i+1,i}(\alpha^*, S_n)\}\Big),$$

where $S_n = \{\alpha_i^*,\ i = 0, \ldots, n+1\}$ is the current set of abscissas in ascending order and α_0^* and α_{n+1}^* are the lower and upper limits of the the domain of f. $L_{i,i'}(\alpha^*, S_n)$ represents the straight line through points $[\alpha_i^*, \ln f(\alpha_i^*)]$ and $[\alpha_{i'}^*, \ln f(\alpha_{i'}^*)]$. Therefore, a sampling density $g_n(\alpha^*)$ is given by

$$\begin{aligned} g_n(\alpha^*) &= \frac{1}{m_n}\exp\big\{h_n(\alpha^*)\big\}, \\ m_n &= \int \exp\big\{h_n(\alpha^*)\big\}. \end{aligned}$$

Algorithm 1: ARMS Algorithm for Matched Case-Crossover Studies

Data: Input 1-M Matched Case-Crossover Data

Result: Samples α^* from the following loop:

Step 0: Initialize n and S_n;

for *The total number of Samples* **do**

Step 1: Sample α^* from $g_n(\alpha^*)$;

Step 2: Sample U from uniform(0,1);

Step 3: **if** $U > f(\alpha^*)/exp\big(h_n(\alpha^*)$ **then**

set $S_{n+1} = S_n \cup \{\alpha^*\}$; Relabel points in S_{n+1} in ascending oder, increment n and go back to Step 2. (Rejection step) ;

else

Set $\alpha_A^* = \alpha^*$(Acceptance step);

end

Step 4: Sample U from uniform(0,1);

Step 5: **if**

$$U > min\left[1, \frac{f(\alpha_A^*)min\{f(\alpha_c^*), exp(h_n(\alpha_c^*))\}}{f(\alpha_c^*)min\{f(\alpha_A^*), exp(h_n(\alpha_A^*))\}}\right]$$

then

$\alpha_M^* = \alpha_c$ (Metropolis-Hastings rejection step);

else

$\alpha_M^* = \alpha_A$ (Metropolis-Hastings acceptance step);

end

Step 6: Return α_M^*.;

end

Appendix B

ARMS and Spatial information for Spatial-Temporal Varying Coefficient Model

In Section B.1, we provide the arms procedure for semiparametric one stage Bayesian approach.

B.1 Adaptive Rejection Metropolis Sampling within Gibbs Sampling for Semiparametric one stage Bayesian approach

We implemented Adaptive Rejection Metropolis Sampling within Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith 1990) in order to obtain sequential samples from the

posterior distributions of our parameters: $(\alpha_0, \ldots \alpha_p, \alpha_{p+l}, ; l = 1, \ldots, L)$, $\boldsymbol{\gamma}$, ϕ_α and ϕ_s. During each iteration of the Gibbs sampler, each of the parameters listed above is updated by a sample drawn from its full conditional distribution. The full conditional distributions for the α and $\boldsymbol{\gamma}$ parameters provided in section 3.3 are not of closed form and therefore we implemented Adaptive Rejection Metropolis Sampling (ARMS) within the Gibbs sampler to draw samples from their full conditional distributions.

ARMS Gilks et al. (1995) is an extension of Adaptive Rejection Sampling (Gilks and Wilde 1992) with the inclusion of a Metropolis-Hastings (Metropolis *et al.*, 1953) step, using ARS to generate a proposal distribution and evaluating it using Metropolis-Hastings acceptance probabilities. $\{\alpha_0, \ldots \alpha_p, \alpha_{p+l}, ; (l = 1, \ldots, L), \boldsymbol{\gamma}, \phi_\alpha, \ \phi_s\}$ is the set of variables to be sampled using the Gibbs sampler. In general let us represent each individual α variable by α^* as this procedure will be repeated for each one. The goal is to sample α^* from its full conditional distribution which is proportional to $f(\alpha^*)$, where $f(\alpha)$ is

$$f(\alpha) = \frac{\phi \exp\left(\frac{-\phi^2\alpha^2}{2}\right)}{\prod_{k=1}^{N}\left(1 + \sum_{j=2}^{M} \exp\left[\left\{\alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p + \sum_{j=1}^{L} \alpha_{1+j}(t-\xi_j)_+^p\right\} x_{jk}^* + z_{jk}^* \boldsymbol{\gamma}\right]\right)}.$$

Here we set $\phi = \phi_1$ for $\alpha_0, \ldots \alpha_p$ and $\phi = \phi_\alpha$ for α_{p+i}, $i = 1, \ldots, L$.

Let α_c^* represent the current value of α^* at a given iteration of the Gibbs sampler. We want to replace α_c^* with a new value α_M^* sampled from f, where f is the posterior distribution of α^*. We create a new function

$$h_n(\alpha^*) = \max\Big(L_{i,i+1}(\alpha^*, S_n), \min\{L_{i-1,i}(\alpha^*, S_n), L_{i+1,i}(\alpha^*, S_n)\}\Big),$$

where $S_n = \{\alpha_i^*, \ i = 0, \ldots, n+1\}$ is the current set of abscissas in ascending order and α_0^* and α_{n+1}^* are the lower and upper limits of the the domain of f. $L_{i,i'}(\alpha^*, S_n)$ represents the

straight line through points $[\alpha_i^*, \ln f(\alpha_i^*)]$ and $[\alpha_{i'}^*, \ln f(\alpha_{i'}^*)]$. Therefore, a sampling density $g_n(\alpha^*)$ is given by

$$\begin{aligned} g_n(\alpha^*) &= \frac{1}{m_n}\exp\big\{h_n(\alpha^*)\big\}, \\ m_n &= \int \exp\big\{h_n(\alpha^*)\big\}. \end{aligned}$$

Also, we need to sample $\boldsymbol{\gamma}$ from its full conditional distribution which is proportional to $f(\boldsymbol{\gamma})$, where $f(\boldsymbol{\gamma})$ is

$$f(\boldsymbol{\gamma}) = \frac{\phi_s \exp\Big(\frac{-\phi_s \boldsymbol{\gamma}\Sigma^{-1}\boldsymbol{\gamma}}{2}\Big)}{\prod_{k=1}^{N}\left(1+\sum_{j=2}^{M}\exp\left[\Big\{\alpha_0+\alpha_1 t+\ldots+\alpha_p t^p+\sum_{j=1}^{L}\alpha_{1+j}(t-\xi_j)_+^p\Big\}x_{jk}^*+z_{jk}^*\boldsymbol{\gamma}\right]\right)}.$$

Let $\boldsymbol{\gamma}_c$ represent the current value of $\boldsymbol{\gamma}$ at a given iteration of the Gibbs sampler. We want to replace $\boldsymbol{\gamma}_c$ with a new value $\boldsymbol{\gamma}_M$ sampled from f where f is the posterior distribution of $\boldsymbol{\gamma}$.

We create a new function

$$h_n(\boldsymbol{\gamma}) = \max\Big(L_{i,i+1}(\boldsymbol{\gamma}, S_n), \min\{L_{i-1,i}(\boldsymbol{\gamma}, S_n), L_{i+1,i}(\boldsymbol{\gamma}, S_n)\}\Big),$$

where $S_n = \{\boldsymbol{\gamma}_i,\ i = 0, \ldots, n+1\}$ is the current set of abscissas in ascending order and $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_{n+1}$ are the lower and upper limits of the the domain of f. $L_{i,i'}(\boldsymbol{\gamma}, S_n)$ represents the straight line through points $[\boldsymbol{\gamma}_i, \ln f(\boldsymbol{\gamma}_i)]$ and $[\boldsymbol{\gamma}_{i'}, \ln f(\boldsymbol{\gamma}_{i'})]$. Therefore, a sampling density

$g_n(\boldsymbol{\gamma})$ is given by

$$\begin{aligned} g_n(\boldsymbol{\gamma}) &= \frac{1}{m_n}\exp\big\{h_n(\boldsymbol{\gamma})\big\}, \\ m_n &= \int \exp\big\{h_n(\alpha^*)\big\}. \end{aligned}$$

Let θ be either α^* or $\boldsymbol{\gamma}$.

Algorithm 2: ARMS Algorithm for Matched Case-Crossover Studies

Data: Input 1-M matched Case-Crossover data

Result: Samples θ from the following loop:

Step 0: initialize n and S_n;

for *Total Number of Samples* **do**

Step 1: Sample θ from $g_n(\theta)$;

Step 2: Sample U from uniform(0,1);

Step 3: **if** $U > f(\theta)/exp\big(h_n(\theta)$ **then**

set $S_{n+1} = S_n \cup \{\theta\}$; Relabel points in S_{n+1} in ascending oder, increment n and go back to Step 2. (Rejection step) ;

else

Set $\theta_A = \theta$ (Acceptance step);

end

Step 4: Sample U from uniform(0,1);

Step 5: **if**

$$U > min\left[1, \frac{f(\theta_A)min\{f(\theta_c), exp(h_n(\theta_c))\}}{f(\theta_c)min\{f(\theta_A), exp(h_n(\theta_A))\}}\right]$$

then

$\theta_M = \theta_c$ (Metropolis-Hastings rejection step);

else

$\theta_M = \theta_A$ (Metropolis-Hastings acceptance step);

end

Step 6: Return θ_M.;

end

B.2 Nine provinces in South Korea and their Latitude and Longitude Coordinates

Province	Latitude	Longitude
North Chungcheong	36.6175641	127.2893383
South Chungcheong	36.5212735	126.7988864
Gangwon	37.8665198	128.2039048
Gyeonggi	37.598322	127.1173206
North Gyeongsang	35.3565811	128.6892794
Souh Gyeongsang	35.2210613	128.3910462
Jeju	33.3354988	126.5690003
North Jeolla	35.7275835	126.9387574
South Jeolla	34.7255374	126.8139992