이학박사 학위논문

Time-dynamic Varying Coefficient Models for Longitudinal Data

경시자료 분석을 위한 동적 변수계수모형

2016년 2월

서울대학교 대학원

통계학과

이 경 은

Time-dynamic Varying Coefficient Models for Longitudinal Data

by

Kyeongeun Lee

A Dissertation

submitted in fulfillment of the requirement

for the degree of

Ph.D of Science

in

Statistics

The Department of Statistics

College of Natural Sciences

Seoul National University

February, 2016

Time-dynamic Varying Coefficient Models for Longitudinal Data

지도교수 박 병 욱

이 논문을 이학박사 학위논문으로 제출함.

2015년 10월

서울대학교 대학원

통계학과

이 경 은

이경은의 이학박사 학위논문을 인준함.

2015년 12월

위 원 장	이	재	용	(인)
부위원장	박	병	욱	(인)
위 원	오	희	석	(인)
위 원	임	요	한	(인)
위 원	양	성	준	(인)

Abstract

Kyeongeun Lee

Statistics

The Graduate School
Seoul National University

In this thesis, we propose a varying coefficient model that can be applied to longitudinal or functional data. The varying coefficient model captures the relationship between the response and the covariates with coefficient functions that are affected by smoothing variables. The varying coefficient model is a structured nonparametric model, and it can easily interpret the effects of the covariates. To avoid the curse of dimensionality, we propose the time-dynamic varying coefficient model as a structured nonparametric model. Also, we construct an iterative algorithm for estimation by extending the smooth backfitting method so that the estimator is defined as a projection of the full-dimensional estimator onto the additive function space with L_2-sense. We show that the proposed algorithm achieves uniform convergence with exponential rate and then study the asymptotic property of the estimator. Based on the numerical study of performances, we apply air quality data to the time-dynamic varying coefficient model for a real data analysis.

Keywords: Kernel smoothing, Longitudinal data, Smooth backfitting, Varying coefficient models

Student Number: 2009-20250

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

In regression analysis, we are generally interested in the relationship between the response and the covariates. One of the best known and conventionally used parametric methods, the linear model, takes the form

$$E(Y|\mathbf{X}) = \sum_{k=1}^{d} \beta_k X_k, \tag{1.1}$$

under the assumption that the relationship between the response variable Y and the covariates $\mathbf{X} = (X_1, \cdots, X_d)^\top$ is linear. In (1.1), β_k are unknown coefficients and need to be estimated. In a (traditional) linear model, the regression coefficients β_k are limited by constants; thus, nonparametric models have been proposed to overcome this limitation. However, the estimation of nonparametric models has always suffered from the "curse of dimensionality".

To avoid such a problem, structured nonparametric models have been suggested, one of them being the varying coefficient model proposed by Hastie and Tibshirani (1993). This model has the form

$$E(Y|\mathbf{X},\mathbf{Z}) = \sum_{k=1}^{d} \beta_k(Z_k)X_k, \tag{1.2}$$

where Y is the response and $\mathbf{X} = (X_1, \cdots, X_d)^\top$ and $\mathbf{Z} = (Z_1, \cdots, Z_d)^\top$ are the covariates. In the varying coefficient models, we call $\mathbf{Z}$ the smoothing variables, and the coefficients are the functions of the smoothing variables $\mathbf{Z}$ so that the impact of the covariates $\mathbf{X}$ on the response depends on $\mathbf{Z}$.

In longitudinal data, we are usually interested in how the covariate affects the response over time. The varying coefficient model gets some advantages in that it is a suitable model for longitudinal data in many cases. First, it can be interpreted as easily as a (traditional) linear model and helps us avoid the curse of dimensionality by adopting a systematic form of the nonparametric model. Usually, in related studies, coefficient functions are the function of time, but in this setting, each coefficient function depends on a single smoothing variable or time. Examples of such studies include Hoover, Rice, Wu and Yang (1998), Huang, Wu and Zhou (2004), and Noh and Park (2010). The model (1.2), wherein each coefficient function is affected by each smoothing variable Z_k, can be extended to a model wherein both the response and the covariates are functional or longitudinal data. It is represented as

$$E(Y(t)|\mathbf{Z}(t),\mathbf{X}(t)) = \sum_{k=1}^{d} \beta_k(Z_k(t))X_k(t), \tag{1.3}$$

where $Y(t)$ is the functional response, and both $\mathbf{Z}(t)$ and $\mathbf{X}(t)$ are d-dimensional functional covariates. Unfortunately, in the model (1.3), time t can affect the coefficient function only with the smoothing variable $\mathbf{Z}$ for each point in time. Therefore, we propose a time-dynamic varying coefficient model

$$E(Y(t)|\mathbf{Z}(t), \mathbf{X}(t)) = \sum_{k=1}^{d} \beta_k(t, Z_k(t))X_k(t), \tag{1.4}$$

so that the coefficient function is a two-dimensional function of time and the smoothing variable $\mathbf{Z}$. Using the model (1.4), it is possible to identify the influence of time t on the relationship between the response and the covariates, and if there is no influence, it enables us to find out whether the model can be reduced to the model (1.3) or not.

Lee, Mammen and Park (2012b) proposed smooth backfitting as a method for estimating the varying coefficient model, and we extend this method to a two-dimensional setting. Zhang, Park and Wang (2013) proposed a method for estimating two-dimensional functions, but in the case, they targeted the estimation of the component functions in the additive model instead of the varying coefficient model. In our model (1.4), we estimate β_k, which is the coefficient of covariate X_k. In particular, the varying coefficient model has a special case where all the covariates $\mathbf{X}$ are constant; thus, the additive model is included in the varying coefficient model. The algorithm suggested in this study achieves uniform convergence in terms of both time t and the smoothing variables $\mathbf{Z}$, and we also provide the asymptotic property of the estimator.

1.2 Varying Coefficient Models

Nonparametric methods fail when the dimension of the covariates is high. One way of avoiding the "curse of dimensionality" is by using structured nonparametric models, such as additive models and varying coefficient models. The additive model is given by

$$E(Y|\mathbf{X}) = \sum_{k=1}^{d} \beta_k(X_k),$$

where Y is the response and $\mathbf{X} = (X_1, \cdots, X_d)^\top$ are the covariates. In this case, each component function β_k is the univariate function of each covariate X_k. There are different methods to estimate the additive model: ordinary backfitting by Buja, Hastie and Tibshirani (1989), marginal integration by Linton and Nielsen (1995), regression spline approach by Stone (1985), and smooth backfitting by Mammen, Linton and Nielsen (1999).

The varying coefficient model proposed by Hastie and Tibshirani (1993) has the form

$$E(Y|\mathbf{X}, \mathbf{Z}) = \sum_{k=1}^{d} \beta_k(Z_k)X_k,$$

where Y is the response, $\mathbf{X} = (X_1, \cdots, X_d)^\top$ are the covariates, and $\mathbf{Z} = (Z_1, \cdots, Z_d)^\top$ are the smoothing variables. Then, each coefficient function β_k is the univariate function of each smoothing variable Z_k. In the special case, all coefficient functions depend on a single variable, which is

$$E(Y|\mathbf{X}, Z) = \sum_{k=1}^{d} \beta_k(Z)X_k,$$

as studied by Fan and Zhang (1999) and Cai, Fan and Li (2000). We focus on the model (1.2), which has different smoothing variables for the coefficient functions. To estimate β_k, we apply local constant smoothing and then minimize

$$n^{-1}\sum_{i=1}^{n}\left(Y_i-\sum_{k=1}^{d}\alpha_k X_{ik}\right)^2 K_h(z_1,Z_{i1})\times\cdots\times K_h(z_d,Z_{id})$$

with respect to α_k, $1 \le k \le d$, where $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n$ are random samples from $(Y, \mathbf{X}, \mathbf{Z})$. One way of estimating the coefficient functions is by the marginal integration method, which was studied by Yang et al. (2006) and Zhang and Li (2007), among others. This method gives the estimator $\hat{\beta}_k(z_k) = \int \hat{\alpha}_k(\mathbf{z})w_k(\mathbf{z}_{-k})d\mathbf{z}_{-k}$, which being a multivariate function, is a function of the variables z_k, $1 \le k \le d$. Thus, this method can suffer from the dimensionality problem. The other method is the smooth backfitting method which was introduced by Mammen, Linton and Nielsen (1999). The smooth backfitting is known to be free of the curse of dimensionality and enjoys some advantages over the marginal integration method. Lee, Mammen and Park (2012b) studied the smooth backfitting method for the varying coefficient models. To estimate β_k, we minimize the integrated kernel-weighted sum of squares

$$\int \frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\sum_{k=1}^{d}\alpha_k(z_k) X_{ik}\right)^2 K_{h_1}(z_1,Z_{i1})\times\cdots\times K_{h_d}(z_d,Z_{id})d\mathbf{z}$$

over $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_d)^\top$. The proposed method aims to obtain the estimator by projecting the vector of the full-dimensional kernel-weighted local polynomial estimators of the coefficient functions onto a Hilbert space with a suitable norm. Park et al. (2015) reviewed the varying coefficient models and introduced methodological and theoretical developments. They also extended the developments to other related problems such as confidence intervals, hypothesis testing, quantile estimation, bandwidth selection, and variable selection. They focused on the kernel smoothing technique, which is a powerful technique for estimating structured nonparametric models and is also known to be one way to avoid the curse of dimensionality. The generalized version was studied by Roca-Pardiñas and Sperlich (2009) with a link function. If we consider all covariates $x_k = 1$, then this model is the same as the generalized additive model studied by Yu, Park and Mammen (2008). Lee, Mammen and Park (2012a) studied a flexible generalized varying coefficient model of the form

$$g(m(\mathbf{x})) = x_1 \left(\sum_{k \in I_1} \beta_{1k}(x_k) \right) + \cdots + x_d \left(\sum_{k \in I_d} \beta_{1k}(x_k) \right),$$

where g is a link function, the index sets I_j are known, and each I_j does not include j. This model includes the generalized varying coefficient models and the generalized additive models as special cases. Lee, Mammen and Park (2014) extended the results to the case for quantile estimation.

1.3 Nonparametric Models for Longitudinal Data

Usually, the longitudinal data contain a time variable T, and a response and the covariates are observed over time. In longitudinal data, we are interested in how the effects of the covariates on the response change with the passage of time, and we use the structured nonparametric model for avoiding the dimensionality problem. In this section, we consider the longitudinal response and covariates. For the additive models, we extend the model as

$$E(Y(t)|\mathbf{X}(t)) = \sum_{k=1}^{d} \beta_k(X_k(t)), \tag{1.5}$$

where $Y(t)$ is the functional response, and $\mathbf{X}(t)$ refers to the d-dimensional functional covariates. Lin and Zhang (1999) and You and Zhou (2007) studied generalized additive model using smoothing splines and two-stage estimation, respectively. Carroll, Maity, Mammen and Yu (2009) proposed the smooth backfitting algorithm for additive models with repeatedly measured data and proved that the algorithm achieves the same efficiency component-wise as achieved in the single function problem, and is, in addition, automatically design independent. In the model (1.5), time exerts an effect on the component functions only through the values of the covariates at that time, after which the component functions are time independent. Accordingly, Zhang, Park and Wang (2013) extended the component functions to vary with time, and thus, the model takes the form

$$m(t, \mathbf{x}) = \beta_0(t) + \sum_{k=1}^{d} \beta_k(t, x_k), \tag{1.6}$$

where $\beta_0(\cdot)$ is the overall mean function, and $\beta_k(\cdot, \cdot)$ are the two-dimensional component functions. As an additive model, this model has the advantage of dimension reduction and captures the relationships between longitudinal covariates and response with time. Zhang, Park and Wang (2013) adopted the smooth backfitting method with a local linear smoother for fitting the model (1.6). In the paper, they extended the smooth backfitting algorithm of Mammen, Linton and Nielsen (1999) to the two-dimensional component function setting and established uniform results with respect to time t and the covariates. Moreover, they obtained the oracle bias and variance together with joint asymptotic normality for the estimators of the component functions.

Park, Mammen, Lee and Lee (2015) studied the varying coefficient model with longitudinal data. Usually, they considered the coefficient functions as functions of time, which meant that the coefficient function depends on a single variable. This model takes the form

$$Y(t) = \sum_{k=1}^{d} \beta_k(t) X_k(t) + \varepsilon(t),$$

where β_k are the coefficient functions, and ε is a mean zero stochastic process. Instead of observing the entire covariate and response processes, one observes them at discrete time points, which are random and different for different subjects. Then, for the ith subject, one observes $\mathbf{X}_{ij} \equiv X_i(T_{ij})$ and $Y_{ij} \equiv Y_i(T_{ij})$ at random time points T_{ij}, $j = 1, \cdots, N_i$. Thus, the varying coefficient model for this longitudinal data takes the form

$$Y_{ij} = \sum_{k=1}^{d} \beta_k(T_{ij}) X_{ijk} + \varepsilon_{ij},$$

where $\varepsilon_{ij} \equiv \varepsilon_i(T_{ij})$. Hoover, Rice, Wu and Yang (1998) applied local constant smoothing and studied the method of minimizing

$$\mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(Y_{ij} - \sum_{k=1}^{d} \alpha_k X_{ijk} \right)^2 K_h(t, T_{ij})$$

with respect to α_k, where $\mathcal{N}_s = \sum_{i=1}^{n} n_i$. Wu, Chiang and Hoover (1998) constructed a class of approximate pointwise and simultaneous confidence regions for $\boldsymbol{\beta}(t)$. Wu and Yu (2002) presented several nonparametric estimation and inference methods for varying coefficient models with longitudinal data. They illustrated applications of those methods with epidemiological examples. Noh and Park (2010) studied the Smoothly Clipped Absolute Deviation (SCAD) penalty in the setting of a spline sieve estimation for varying coefficient models with longitudinal data. They proposed a one-step estimator by linearizing the group SCAD (gSCAD) penalty. The estimator has the oracle property in variable selection and estimation. The proposed estimator has a much simpler and better performance in variable selection and estimation than the ordinary gSCAD estimator.

The coefficient functions have been the function of time t. Now, we extend the model (1.2) such that each coefficient function is the function of each smoothing variable for longitudinal data. Then, our model takes the form

$$E(Y(t)|\mathbf{Z}(t), \mathbf{X}(t)) = \sum_{k=1}^{d} \beta_k(t, Z_k(t))X_k(t),$$

where $Y(t)$ is the functional response, and both $\mathbf{Z}(t)$ and $\mathbf{X}(t)$ are the d-dimensional functional covariates. The coefficient function is a two-dimensional function, and we extend the smooth backfitting technique for this setting in the following chapters.

1.4 Outline of the thesis

The remainder of this thesis is organized as follows. In Chapter 2, we redefine our model and propose the smooth backfitting algorithm to estimate the model. In Chapter 3, we give its asymptotic properties. In Chapter 4, we present the simulation results and illustrate the model using real data. The technical details appear in Chapter 5.

Chapter 2

Methodology

2.1 Models

We assume the response and covariates are L^2-stochastic processes on unit intervals. Let $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n$ be random samples from $(Y, \mathbf{X}, \mathbf{Z})$. The proposed model takes the form:

$$Y_i(t) = \sum_{k=1}^{d} \beta_k(t, Z_{ik}(t)) X_{ik}(t) + \delta_i(t),$$

where δ_i are independent and identically distributed (iid) copies of an L^2-stochastic process δ on [0,1] with $E(\delta(t)|\mathbf{Z}(t), \mathbf{X}(t)) = 0$.

For the ith subject, we observe discrete time points T_{ij}, $j = 1, \cdots, N_{in}$, and at these points we observe the response and covariates. We assume that the numbers of time points N_{in} are iid copies of N_n, which depends on n. Thus, we observe $\{(T_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}, Y_{ij}) : 1 \leq i \leq n, 1 \leq j \leq N_{in}\}$ such that

$\mathbf{Z}_{ij} = \mathbf{Z}_i(T_{ij}) = (Z_{ij1}, \cdots, Z_{ijd})^\top$, $\mathbf{X}_{ij} = \mathbf{X}_i(T_{ij}) = (X_{ij1}, \cdots, X_{ijd})^\top$ and $Y_{ij} = Y_i(T_{ij})$. For the observation, the time-dynamic varying coefficient model is given by,

$$Y_{ij} = \sum_{k=1}^{d} \beta_k(T_{ij}, Z_{ijk}) X_{ijk} + \delta_{ij} + e_{ij}, \ 1 \le j \le N_{in}, 1 \le i \le n,$$

where $\delta_{ij} = \delta_i(T_{ij})$ is the stochastic variation of the true longitudinal process, and e_{ij} is the measurement error. We denote $\varepsilon_{ij} \equiv \delta_{ij} + e_{ij}$.

We identify the coefficient functions β_k by minimizing for each t,

$$\begin{aligned} &E\left[\left(Y - \sum_{k=1}^{d} f_k(T, Z_k(T)) X_k(T)\right)^2 \middle| T = t\right] \\ &= \int E\left[\left(Y - \sum_{k=1}^{d} f_k(T, Z_k(T)) X_k(T)\right)^2 \middle| T = t, \mathbf{Z}(T) = \mathbf{z}\right] p(\mathbf{z}|t) d\mathbf{z} \end{aligned}$$

over a space of function tuples $(f_k, 1 \le k \le d)$. The integrated kernel-weighted least-squares criterion is

$$\int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_{in}} \left[Y_{ij} - \sum_{k=1}^{d} f_k(T_{ij}, Z_{ijk}) X_{ijk}\right]^2 K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}) d\mathbf{z},$$

where $\mathcal{N}_s = \sum_{i=1}^{n} N_{in}$. Here, we define the kernel functions. Let K be a baseline kernel function supported on $[-1, 1]$. A boundary corrected kernel is defined by

$$K_h(u; v) = \frac{K((u - v)/h)/h}{\int_{\mathcal{I}} [K((w - v)/h)/h] dw} \cdot I(u, v \in \mathcal{I} \equiv [0, 1]).$$

When referring to smooth backfitting, we often use the boundary corrected kernel because of the property that $\int K_h(u; v) du = 1$. Also, a product kernel for a bandwidth vector $\mathbf{h} = (h_0, h_1, \cdots, h_d)$ is defined by

$$K_{\mathbf{h}}(t, \mathbf{z}; s, \mathbf{u}) = K_{h_0}(t; s) \prod_{i=1}^{d} K_{h_k}(z_k; u_k)$$

and $\int K_{\mathbf{h}}(t, \mathbf{z}; s, \mathbf{u}) d\mathbf{z}_{-k} = K_{h_0, h_k}(t, z_k; s, u_k)$.

Now, we approximate $\beta_k(T_{ij}, Z_{ijk})$ by

$$\beta_k(t, z_k) + \beta_{k,1}(t, z_k) \left(\frac{T_{ij} - t}{h_0} \right) + \beta_{k,2}(t, z_k) \left(\frac{Z_{ijk} - z_k}{h_k} \right),$$

where $\beta_{k,1}/h_0$ and $\beta_{k,2}/h_k$ are the partial derivatives of β_k with respect to the first and second arguments, respectively. Then, we define the true functions $\boldsymbol{\beta}_k(t, z_k) = (\beta_k(t, z_k), h_0 \partial \beta_k(t, z_k)/\partial t, h_k \partial \beta_k(t, z_k)/\partial z_k)^\top$. Define

$$\begin{aligned}
\mathbf{w}_{ijk}(t, z_k) &= \left(1, \frac{T_{ij} - t}{h_0}, \frac{Z_{ijk} - z_k}{h_k} \right)^\top, \\
\mathbf{v}_{ijk}(t, z_k; x_k) &= \mathbf{w}_{ijk}(t, z_k) x_k, \\
\mathbf{v}(t, \mathbf{z}; \mathbf{x})^\top &= (\mathbf{w}_{ij1}(t, z_1)^\top x_1, \cdots, \mathbf{w}_{ijd}(t, z_d)^\top x_d).
\end{aligned}$$

To estimate $\boldsymbol{\beta}_k(t, \cdot)$, we minimize, for each t,

$$\begin{aligned}
L_t(\mathbf{f}) = \int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_{in}} & \left[Y_{ij} - \sum_{k=1}^{d} X_{ijk} \mathbf{w}_{ijk}(t, z_k)^\top \mathbf{f}_k(z_k) \right]^2 \\
& \times K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}) d\mathbf{z}
\end{aligned} \tag{2.1}$$

over $\mathbf{f} \equiv \mathbf{f}(t, \cdot)$ with $L_t(\mathbf{f}) < \infty$, where $\mathbf{f}_k = (f_k, f_{k,1}, f_{k,2})^\top$ and $\mathbf{f}^\top = (\mathbf{f}_1^\top, \cdots, \mathbf{f}_d^\top)$. We consider the within-subject correlation structure in the integrated kernel-weighted least-squares criterion (2.1) to be independent. It results from the local property of the kernel method. Also, Lin and Carroll

(2000) guarantees that it is an efficient approach to ignore the correlation structure entirely for longitudinal data with the kernel method. The function space for the minimization problem is

$$\mathcal{H}(\hat{\mathbf{M}}_t) = \left\{ \mathbf{f} \in L_2(\hat{\mathbf{M}}_t) : \mathbf{f}_k(\mathbf{z}) = (f_k(z_k), f_{k,1}(z_k), f_{k,2}(z_k))^\top, 1 \le k \le d \right\},$$

where

$$L_2(\hat{\mathbf{M}}_t) = \Big\{ \mathbf{f}(\mathbf{z}) = (\mathbf{f}_1(\mathbf{z})^\top, \cdots, \mathbf{f}_d(\mathbf{z})^\top)^\top : \mathbf{f}_k(\mathbf{z}) = (f_k(\mathbf{z}), f_{k,1}(\mathbf{z}), f_{k,2}(\mathbf{z}))^\top, \\ \|\mathbf{f}\|^2_{\hat{\mathbf{M}}_t} = \int \mathbf{f}(\mathbf{z})^\top \hat{\mathbf{M}}(t,\mathbf{z}) \mathbf{f}(\mathbf{z}) d\mathbf{z} < \infty \Big\}$$

and

$$\hat{\mathbf{M}}(t,\mathbf{z}) = \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_{in}} \mathbf{v}(t,\mathbf{z};\mathbf{X}_{ij}) \mathbf{v}(t,\mathbf{z};\mathbf{X}_{ij})^\top K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}).$$

In the function space $\mathcal{H}(\hat{\mathbf{M}}_t)$, each function f_k depends on only one variable, z_k, but there is no such condition in the function space $L_2(\hat{\mathbf{M}}_t)$.

For $\mathbf{f} \in \mathcal{H}(\hat{\mathbf{M}}_t)$, $L_t(\mathbf{f})$ is equivalent to

$$\int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_{in}} \left[Y_{ij} - \mathbf{v}(t,\mathbf{z};\mathbf{X}_{ij})^\top \tilde{\boldsymbol{\beta}}(t,\mathbf{z}) \right]^2 K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}) d\mathbf{z} \\ + \int \left[\tilde{\boldsymbol{\beta}}(t,\mathbf{z}) - \mathbf{f}(\mathbf{z}) \right]^\top \hat{\mathbf{M}}(t,\mathbf{z}) \left[\tilde{\boldsymbol{\beta}}(t,\mathbf{z}) - \mathbf{f}(\mathbf{z}) \right] d\mathbf{z},$$

where the full-dimensional estimator $\tilde{\boldsymbol{\beta}}(t,\mathbf{z})$ is the minimizer of $L_t(\mathbf{f})$ over $\mathbf{f} \in L_2(\hat{\mathbf{M}}_t)$. The explicit form of $\tilde{\boldsymbol{\beta}}(t,\mathbf{z})$ is

$$\tilde{\boldsymbol{\beta}}(t,\mathbf{z}) = \hat{\mathbf{M}}(t,\mathbf{z})^{-1} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{v}(t,\mathbf{z};\mathbf{X}_{ij}) K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}) Y_{ij}.$$

2.2 Backfitting Equation

The proposed estimator $\hat{\boldsymbol{\beta}}(t,\cdot)$ can be defined as the projection of $\tilde{\boldsymbol{\beta}}(t,\cdot)$ onto $\mathcal{H}(\hat{\mathbf{M}}_t)$:

$$\hat{\boldsymbol{\beta}} = \underset{\mathbf{f}\in\mathcal{H}(\hat{\mathbf{M}}_t)}{\arg\min}\|\tilde{\boldsymbol{\beta}} - \mathbf{f}\|^2_{\hat{\mathbf{M}}_t}.$$

Thus, the solution $\hat{\boldsymbol{\beta}}(t,\mathbf{z}) = (\hat{\boldsymbol{\beta}}_1(t,z_1)^\top,\cdots,\hat{\boldsymbol{\beta}}_d(t,z_d)^\top)^\top$ satisfies the following system of equation,

$$\int \hat{\mathbf{M}}_k(t,\mathbf{z})^\top \left[\tilde{\boldsymbol{\beta}}(t,\mathbf{z}) - \hat{\boldsymbol{\beta}}(t,\mathbf{z})\right] d\mathbf{z}_{-k} = \mathbf{0}_3, \quad 1\le k\le d, \tag{2.2}$$

where $\hat{\mathbf{M}}_k(t,\mathbf{z})$ are the $3d\times 3$ matrices defined by $\hat{\mathbf{M}} = \hat{\mathbf{M}}^\top = (\hat{\mathbf{M}}_1,\cdots,\hat{\mathbf{M}}_d)^\top$. The system of equations (2.2) can be expressed as the following backfitting equation:

for $1\le k\le d$,

$$\hat{\boldsymbol{\beta}}_k(t,z_k) = \tilde{\boldsymbol{\beta}}_k(t,z_k) - \sum_{l\neq k}^{d}\int \hat{\boldsymbol{\Psi}}_k(t,z_k)^{-1}\hat{\boldsymbol{\Psi}}_{kl}(t,z_k,z_l)\hat{\boldsymbol{\beta}}_l(t,z_l)dz_l, \tag{2.3}$$

where

$$\tilde{\boldsymbol{\beta}}_k(t,z_k) = \hat{\boldsymbol{\Psi}}_k(t,z_k)^{-1}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_{in}}\mathbf{w}_{ijk}(t,z_k)K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk})X_{ijk}Y_{ij},$$

$$\hat{\boldsymbol{\Psi}}_k(t,z_k) = \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_{in}}\mathbf{w}_{ijk}(t,z_k)\mathbf{w}_{ijk}(t,z_k)^\top K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk})(X_{ijk})^2,$$

$$\hat{\boldsymbol{\Psi}}_{kl}(t,z_k,z_l) = \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_{in}}\mathbf{w}_{ijk}(t,z_k)\mathbf{w}_{ijl}(t,z_l)^\top$$

$$K_{h_0,h_k,h_l}(t,z_k,z_l;T_{ij},Z_{ijk},Z_{ijl})X_{ijk}X_{ijl}, \text{ for } l\neq k.$$

The smooth backfitting equation (2.3) does not need to compute the full-dimensional estimator $\hat{\boldsymbol{\beta}}$. It only needs $\tilde{\boldsymbol{\beta}}_k$, $\hat{\boldsymbol{\Psi}}_k$, $\hat{\boldsymbol{\Psi}}_{kl}$, and the inversion of $\hat{\boldsymbol{\Psi}}_k$. Also $\tilde{\boldsymbol{\beta}}_k$ are not equal to the z_k part of $\hat{\boldsymbol{\beta}}$.

To solve the system of equations (2.3), we need the following an iterative scheme.

Backfitting Algorithm

With initial estimates $\hat{\boldsymbol{\beta}}_k^{[0]}$, we iterate for $r = 1, \cdots$, the following process: For $1 \le k \le d$,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_k^{[r]}(t, z_k) = \tilde{\boldsymbol{\beta}}_k(t, z_k) &- \sum_{l=1}^{k-1} \int \hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1} \hat{\boldsymbol{\Psi}}_{kl}(t, z_k, z_l) \hat{\boldsymbol{\beta}}_l^{[r]}(t, z_l) dz_l \\ &- \sum_{l=k+1}^{d} \int \hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1} \hat{\boldsymbol{\Psi}}_{kl}(t, z_k, z_l) \hat{\boldsymbol{\beta}}_l^{[r-1]}(t, z_l) dz_l.\end{aligned} \tag{2.4}$$

In the next Chapter, we see that the backfitting algorithm converges to the unique solution of the backfitting equation. We show that the backfitting algorithm (2.4) converges to $\hat{\boldsymbol{\beta}}_k$ at a geometric rate and that $\hat{\boldsymbol{\beta}}_k$ are jointly asymptotically normal.

Chapter 3

Asymptotic Properties

3.1 Convergence of Backfitting

We consider the limit of the matrix $\hat{\mathbf{M}}(t, \mathbf{z})$ before we discuss the asymptotic properties. $\hat{\mathbf{M}}(t, \mathbf{z})$ consists of 3×3 blocks

$$\hat{\mathbf{M}}_{k,l}(t, \mathbf{z}) = \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{w}_{ijk}(t, z_k) \mathbf{w}_{ijl}(t, z_l)^\top X_{ijk} X_{ijl} \times K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}), \quad 1 \leq k, l \leq d.$$

The matrices $\hat{\mathbf{M}}_{k,l}(t, \mathbf{z})$ are approximated by

$$\begin{aligned} & E[\mathbf{w}_k(t, z_k) \mathbf{w}_l(t, z_l)^\top X_k X_l K_{\mathbf{h}}(t, \mathbf{z}; T, \mathbf{Z})] \\ \simeq\ & \mathbf{D}_0 E(X_k X_l | T = t, \mathbf{Z}(T) = \mathbf{z}) p(t, \mathbf{z}), \text{ for } k \neq l, \end{aligned}$$

and

$$\mathbf{D} E(X_k^2 | T = t, \mathbf{Z}(T) = \mathbf{z}) p(t, \mathbf{z}), \text{ for } k = l,$$

where $\mathbf{D}_0$ and $\mathbf{D}$ are 3×3 diagonal matrices with the entries $(1, \int u^2K(u)du, 0)$ and $(1, \int u^2K(u)du, \int u^2K(u)du)$, respectively. Then, $\hat{\mathbf{M}}(t,\mathbf{z})$ is approximated by

$$\mathbf{M}(t,\mathbf{z}) \equiv p(t,\mathbf{z})\Big[E(\mathbf{X}\mathbf{X}^\top|T=t,\mathbf{Z}(T)=\mathbf{z})\otimes\mathbf{D}_0 \\ + \operatorname{diag}\big(E(X_k^2|T=t,\mathbf{Z}(T)=\mathbf{z})\big)\otimes(\mathbf{D}-\mathbf{D}_0)\Big],$$

where $\otimes$ denotes the Kronecker product. Also, the limits of the vector of functions $\hat{\mathbf{\Psi}}_k(t,z_k)$ and $\hat{\mathbf{\Psi}}_{kl}(t,z_k,z_l)$ are, for $l\neq k$,

$$\mathbf{\Psi}_k(t,z_k) = \mathbf{D}E(X_k^2|T=t,Z_k(T)=z_k)p_k(t,z_k),$$
$$\mathbf{\Psi}_{kl}(t,z_k,z_l) = \mathbf{D}_0E(X_kX_l|T=t,Z_k(T)=z_k,Z_l(T)=z_l)p_{kl}(t,z_k,z_l).$$

We define the function space $\mathcal{H}(\mathbf{M}_t)$ as $\mathcal{H}(\hat{\mathbf{M}}_t)$ with $\hat{\mathbf{M}}$ being replaced by $\mathbf{M}$

$$\mathcal{H}(\mathbf{M}_t) = \left\{\mathbf{f}\in L_2(\mathbf{M}_t) : \mathbf{f}_k(\mathbf{z}) = (f_k(z_k), f_{k,1}(z_k), f_{k,2}(z_k))^\top, 1\le k\le d\right\},$$

where

$$L_2(\mathbf{M}_t) = \Big\{\mathbf{f}(\mathbf{z}) = (\mathbf{f}_1(\mathbf{z})^\top,\cdots,\mathbf{f}_d(\mathbf{z})^\top)^\top : \mathbf{f}_k(\mathbf{z}) = (f_k(\mathbf{z}), f_{k,1}(\mathbf{z}), f_{k,2}(\mathbf{z}))^\top, \\ \|\mathbf{f}\|^2_{\mathbf{M}_t} = \int \mathbf{f}(\mathbf{z})^\top\mathbf{M}(t,\mathbf{z})\mathbf{f}(\mathbf{z})d\mathbf{z} < \infty\Big\}.$$

For each coefficient function $\hat{\boldsymbol{\beta}}_k$, we define the subspaces of $\mathcal{H}(\hat{\mathbf{M}}_t)$

$$\mathcal{H}_k(\hat{\mathbf{M}}_t) = \Big\{\mathbf{f}\in L_2(\hat{\mathbf{M}}_t) :\ \mathbf{f}_k(\mathbf{z}) = (f_k(z_k), f_{k,1}(z_k), f_{k,2}(z_k))^\top \text{ and} \\ \mathbf{f}_l\equiv 0 \text{ for } l\neq k\Big\},$$

for $1\le k\le d$. Thus, $\mathcal{H}(\hat{\mathbf{M}}_t) = \mathcal{H}_1(\hat{\mathbf{M}}_t)+\cdots+\mathcal{H}_d(\hat{\mathbf{M}}_t)$. We define $\mathcal{H}_k(\mathbf{M}_t)$ likewise.

Let $\hat{\Pi}_k^t$ and Π_k^t denote the projection operator onto $\mathcal{H}_k(\hat{\mathbf{M}}_t)$ and $\mathcal{H}_k(\mathbf{M}_t)$, respectively. Take $\hat{\Phi}_t = (I - \hat{\Pi}_d^t) \times \cdots \times (I - \hat{\Pi}_1^t)$ and $\Phi_t = (I - \Pi_d^t) \times \cdots \times (I - \Pi_1^t)$. For $\mathbf{f} \in L_2(\hat{\mathbf{M}}_t)$, we note that

$$
\begin{aligned}
&(\hat{\Pi}_k^t(\mathbf{f}))_k = \hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1} \int \hat{\mathbf{M}}_k(t, \mathbf{z})^\top \mathbf{f}(\mathbf{z}) d\mathbf{z}_{-k} \quad \text{and} \\
&(\hat{\Pi}_k^t(\mathbf{f}))_l = \mathbf{0} \quad \text{for} \quad l \neq k,
\end{aligned}
$$

where $(\hat{\Pi}_k^t(\mathbf{f}))_l$ denotes the lth 3-dimensional vector of the projection of $\mathbf{f}$ onto $\mathcal{H}_k(\hat{\mathbf{M}}_t)$. Also, $\hat{\Pi}_k^t(\tilde{\boldsymbol{\beta}}(t, \cdot)) = \tilde{\boldsymbol{\beta}}_k(t, \cdot)$. By this definition, we can rewrite the backfitting equation (2.3) as

$$
\hat{\boldsymbol{\beta}}_k = \Pi \left(\tilde{\boldsymbol{\beta}} - \sum_{l \neq k}^{d} \hat{\boldsymbol{\beta}}_l \,\middle|\, \mathcal{H}_k(\hat{\mathbf{M}}_t) \right), \text{ for } 1 \leq k \leq d.
$$

Below, we present the assumptions for the theorems and lemmas.

Assumptions

(A1) $E(\mathbf{X}\mathbf{X}^\top | T = t, \mathbf{Z}(T) = \mathbf{z})$ is continuous, and its smallest eigenvalue is bounded away from zero on $[0,1] \times \mathbb{Z}$, where $\mathbb{Z} = \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_d$ for compact sets $\mathbb{Z}_k$.

(A2) $\sup_{t \in [0,1], \mathbf{z} \in \mathbb{Z}} E(X_k^4 | T = t, \mathbf{Z}(T) = \mathbf{z}) < \infty$ for all $1 \le k \le d$.

(A3) The density function $p(t, \mathbf{z})$ of $(T, \mathbf{Z}(T))$ is supported on $[0,1] \times \mathbb{Z}$, and is continuous and bounded away from zero and infinity on its support.

(A4) $E|\varepsilon_{11}|^\alpha < \infty$ for $\alpha > 3$.

(A5) The kernel K is a bounded and symmetric probability density function supported on $[-1, 1]$ and is Lipschitz continuous.

(A6) $E(\mathbf{X}\mathbf{X}^\top \sigma^2(T, \mathbf{Z}, \mathbf{X}) | T = t, \mathbf{Z}(T) = \mathbf{z})$ is continuous on $[0,1] \times \mathbb{Z}$, where $\sigma^2(T, \mathbf{Z}, \mathbf{X}) = \mathrm{var}(Y | T, \mathbf{Z}(T), \mathbf{X}(T))$.

(A7) The coefficient functions β_k are twice partially continuously differentiable on $[0,1] \times \mathbb{Z}_k$, and $E(X_k X_l | T = t, \mathbf{Z}(T) = \mathbf{z})$ is continuous in t and $\mathbf{z}$ on $[0,1] \times \mathbb{Z}$ for all $1 \le k, l \le d$.

(A8) N_{in} are iid copies of N_n, and $E \exp(N_n)$ is bounded.

(A9) The bandwidths h_k are of order $(nEN_n)^{-1/6}$.

We note that $\|\cdot\|_{\mathbf{M}_t}$ becomes a norm if we assume that

$$\mathbf{f}(t,\mathbf{z})^\top \mathbf{x} = 0 \quad \text{almost surely implies} \quad \mathbf{f} = \mathbf{0}, \tag{3.1}$$

and under this assumption (3.1), m_j are not identifiable. The assumption (3.1) is satisfied if we assume that the smallest eigenvalue of $E(\mathbf{X}\mathbf{X}^\top | T = t, \mathbf{Z}(T) = \mathbf{z})$ is bounded away from zero on $[0,1] \times \mathbb{Z}$. The assumptions (A2) and (A4) are useful for the uniform convergence of the kernel weighted average. The assumptions (A6) and (A7) are useful to prove the asymptotic distribution of the proposed estimator. To control the within subject dependence structure, we assume (A8). The assumption (A9) assumes that the bandwidth takes the optimal rate for two dimensional smoothing.

Theorem 1. *Suppose that (A1)-(A9) hold. With probability tending to 1, there exists a unique solution of smooth backfitting equation (2.3). Moreover, there exist constants* $0 < C < \infty$ *and* $0 < \gamma < 1$ *such that, with probability tending to 1,*

$$\int \sum_{k=1}^{d} \left| \hat{\boldsymbol{\beta}}_k^{[r]}(t, z_k) - \hat{\boldsymbol{\beta}}_k(t, z_k) \right|^2 p_k(t, z_k) dz_k dt \le C \cdot \gamma^r \cdot \Lambda, \tag{3.2}$$

where

$$\Lambda = \sum_{k=1}^{d} \int \left[\left| \tilde{\boldsymbol{\beta}}_k(t, z_k) \right|^2 + \left| \hat{\boldsymbol{\beta}}_k^{[0]}(t, z_k) \right|^2 \right] p_k(t, z_k) dz_k dt.$$

For each $t \in [0,1]$, there exist constants $0 < C < \infty$ and $0 < \gamma < 1$ such that, with probability tending to 1,

$$\int \sum_{k=1}^{d} \left| \hat{\boldsymbol{\beta}}_k^{[r]}(t, z_k) - \hat{\boldsymbol{\beta}}_k(t, z_k) \right|^2 p_k(t, z_k) dz_k \le C \cdot \gamma^r \cdot \Lambda(t),$$

where $\Lambda(t) = \sum_{k=1}^{d} \int \left[\left| \tilde{\boldsymbol{\beta}}_k(t, z_k) \right|^2 + \left| \hat{\boldsymbol{\beta}}_k^{[0]}(t, z_k) \right|^2 \right] p_k(t, z_k) dz_k$. By integrating out the inequality over t, we confirm that the inequality (3.2) of Theorem 1 holds with probability tending to one. The important part in the proof of Theorem 1 is the constant γ, which is not a function of t.

3.2 Asymptotic Distribution

In this section, we discuss the asymptotic distribution of $\hat{\beta}_k$. We consider approximations of the stochastic and deterministic parts of the estimators $\hat{\boldsymbol{\beta}}_k$. Then, we decompose $\tilde{\boldsymbol{\beta}}$ as $\tilde{\boldsymbol{\beta}}^A + \tilde{\boldsymbol{\beta}}^B$, where

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}^A(t, \mathbf{z}) = \hat{\mathbf{M}}(t, \mathbf{z})^{-1} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{v}(t, \mathbf{z}; \mathbf{X}_{ij}) K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}) \\
\times [Y_{ij} - m(T_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})], \\
\tilde{\boldsymbol{\beta}}^B(t, \mathbf{z}) = \hat{\mathbf{M}}(t, \mathbf{z})^{-1} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{v}(t, \mathbf{z}; \mathbf{X}_{ij}) K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}) \\
\times m(T_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}).
\end{aligned}$$

By Theorem 1, we can define $\hat{\boldsymbol{\beta}}_k^A$ as the solutions of (2.3), with $\tilde{\boldsymbol{\beta}}_k$ being replaced by $\tilde{\boldsymbol{\beta}}_k^A = (\hat{\Pi}_k^t \tilde{\boldsymbol{\beta}}^A)_k$, where

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_k^A(t, z_k) = \hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{w}_{ijk}(t, z_k) X_{ijk} K_{h_0, h_k}(t, z_k; T_{ij}, Z_{ijk}) \\
\times [Y_{ij} - m(T_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij})].
\end{aligned}$$

We define $\hat{\boldsymbol{\beta}}_k^B$ likewise. Then, $\hat{\boldsymbol{\beta}}_k(t, z_k) = \hat{\boldsymbol{\beta}}_k^A(t, z_k) + \hat{\boldsymbol{\beta}}_k^B(t, z_k)$ and $\tilde{\boldsymbol{\beta}}_k(t, z_k) = \tilde{\boldsymbol{\beta}}_k^A(t, z_k) + \tilde{\boldsymbol{\beta}}_k^B(t, z_k)$.

First, we give the approximation of $\hat{\boldsymbol{\beta}}_k^A(t, z_k)$.

Lemma 1. *Suppose that (A1)-(A9) hold. Then, for* $1 \le k \le d$*,*

$$\sup_{t \in I_0, z_k \in I_k} \left| \hat{\boldsymbol{\beta}}_k^A(t, z_k) - \tilde{\boldsymbol{\beta}}_k^A(t, z_k) \right| = o_p(n^{-1/3}), \tag{3.3}$$

where $I_0 = [h_0, 1 - h_0]$ *and* $I_k = [\min \mathbb{Z}_k + h_k, \max \mathbb{Z}_k - h_k]$.

We prove the uniform convergence rate of the stochastic part, which is the two-dimensional optimal rate. Next, we use the following lemma for the approximation of $\hat{\boldsymbol{\beta}}_k^B(t, z_k)$.

Lemma 2. *Suppose that (A1)-(A9) hold. Then, for* $1 \le k \le d$*,*

$$\sup_{t \in I_0, z_k \in I_k} \left| \hat{\boldsymbol{\beta}}_k^B(t, z_k) - \boldsymbol{\beta}_k(t, z_k) - \boldsymbol{\theta}_k(t, z_k) \right| = o_p(n^{-1/3}),$$

where $\boldsymbol{\theta}_k(t, z_k) = \theta_k(t, z_k)\mathbf{1}_1$*,* $\mathbf{1}_1 = (1, 0, 0)^\top$*, and*

$$\theta_k(t, z_k) = \frac{1}{2} \left(\int u^2 K(u) du \right) \left[\frac{\partial^2 \beta_k(t, z_k)}{\partial t^2} h_0^2 + \frac{\partial^2 \beta_k(t, z_k)}{\partial z_k^2} h_k^2 \right].$$

We give the theorem for the asymptotic distribution of the estimator $\hat{\beta}_k$ using these two lemmas.

Theorem 2. *Assume that (A1)-(A9) hold and that the bandwidths* h_k *are asymptotic to* $c_k(nEN_n)^{-1/6}$ *for some constants* $0 < c_k < \infty$ *and* $0 \le k \le d$*. Then, for any* t *and* z_k *in* $(0, 1)$ *and in the interior of* $\mathbb{Z}_k$ *for* $1 \le k \le d$*,*

$$(nEN_n)^{1/3} \begin{bmatrix} \hat{\beta}_1(t, z_1) - \beta_1(t, z_1) \\ \vdots \\ \hat{\beta}_d(t, z_d) - \beta_d(t, z_d) \end{bmatrix} \xrightarrow{d} \mathcal{N}\Big(\Theta(t, \mathbf{z}), \mathbf{V}(t, \mathbf{z})\Big),$$

where

$$(\Theta(t,\mathbf{z}))_k = \frac{1}{2}\left(\int u^2K(u)du\right)\left[\frac{\partial^2\beta_k(t,z_k)}{\partial t^2}c_0^2 + \frac{\partial^2\beta_k(t,z_k)}{\partial z_k^2}c_k^2\right],$$

$$\mathbf{V}(t,\mathbf{z}) = \mathrm{diag}\left(\left(\int K^2(u)du\right)^2 \frac{E[X_k^2\sigma^2(T,\mathbf{Z},\mathbf{X})|T=t, Z_k(T)=z_k]}{[E(X_k^2|T=t, Z_k(T)=z_k)]^2 p_k(t,z_k)} c_0^{-1}c_k^{-1}\right).$$

We show that the estimator $\hat{\beta}_k$ has the oracle properties. The oracle property means that the estimator $\hat{\beta}_k$ has the same asymptotic distribution as the assumed estimator that we know all coefficient functions except β_k.

Chapter 4

Numerical Studies

4.1 Simulation Results

We conduct simulation studies to support the result and the finite sample property of the proposed estimator. We generate 100 datasets with sample sizes $n = 200$ and 400. The number of time points are iid, having a discrete uniform distribution on $\{2, 3, \cdots, 10\}$. The time points are iid $\mathcal{U}[0, 1]$. The covariates are generated by

$$X_{ij1} = T_{ij} - 0.5 + \alpha_{i1} + \beta_{i1} T_{ij} + \zeta_{ij1},$$

$$X_{ij2} = (T_{ij} - 0.5)^3 + \alpha_{i2} + \beta_{i2} T_{ij} + \zeta_{ij2},$$

$$Z_{ij1} = 0.2 \exp(-T_{ij}) - 0.13 + \eta_{ij1},$$

$$Z_{ij2} = 0.7(T_{ij} - 0.5)^3 + \eta_{ij2},$$

where $(\alpha_{i1}, \beta_{i1}, \alpha_{i2}, \beta_{i2})$ are iid $\mathcal{N}(\mathbf{0}, \Sigma)$ for a covariance matrix $\Sigma \neq \mathbf{I}$. Moreover, ζ_{ijk} and η_{ijk} are generated independently from $\mathcal{N}(0, 0.3^2)$ and $\mathcal{U}[0, 1]$, respectively, and are independent of $(\alpha_{i1}, \beta_{i1}, \alpha_{i2}, \beta_{i2})$. The error terms are generated by

$$\varepsilon_{ij} = a_i(\cos(2\pi T_{ij}) + \sin(2\pi T_{ij})) + b_{ij},$$

where a_i and b_{ij} are iid $\mathcal{N}(0, 0.1^2)$ and $\mathcal{N}(0, 0.05^2)$.

We consider the two cases,

$$\begin{aligned}
\text{Case 1: } \beta_1(t, z_1) &= 3t \exp(-(t - 0.5)^2 - (z_1 - 1.5)^2), \\
\beta_2(t, z_2) &= 2(z_2 - 0.5)(t - z_2 + 3). \\
\text{Case 2: } \beta_1(t, z_1) &= 2(-\sin(\pi t) + \cos(0.5\pi z_1)), \\
\beta_2(t, z_2) &= 2(\sin(\pi t) * \cos(\pi z_2)).
\end{aligned}$$

Figures 4.1 and 4.2 show the true coefficient functions of Cases 1 and 2. In Figure 4.1 and Figure 4.2, we can see that the functions of Case 2 have more curvature than the functions of Case 1.

Case 1

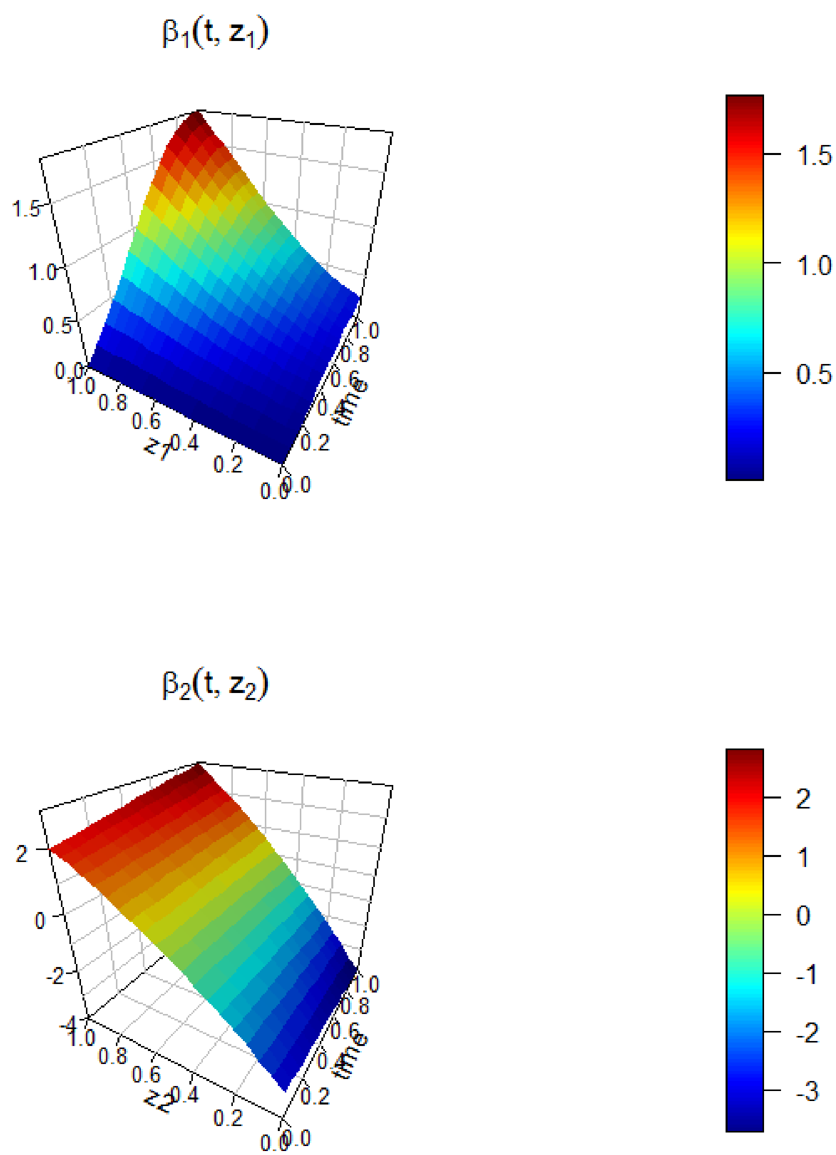

Figure 4.1: The plots are the true coefficient functions of β_1 and β_2 for Case 1.

Case 2

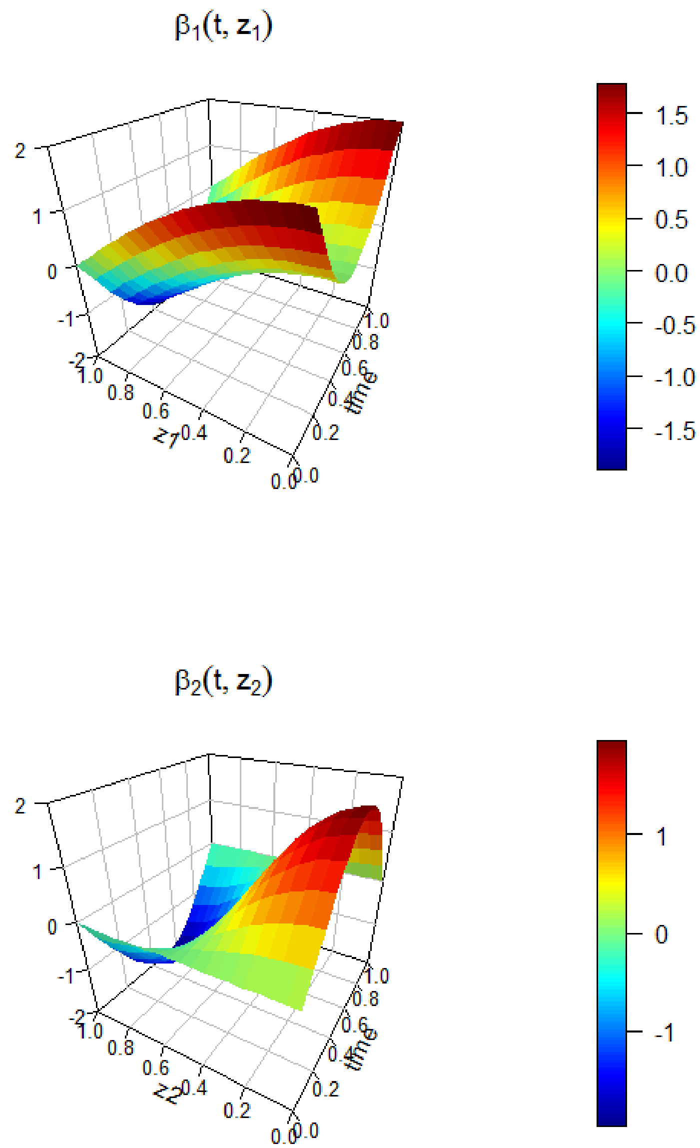

Figure 4.2: The plots are the true coefficient functions of β_1 and β_2 for Case 2.

We use the Epanechnikov kernel $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$ and set the threshold for the convergence of the backfitting algorithm to 10^{-5}. Figure 4.3 and Figure 4.4 show that the average of 100 estimated functions is positive. Also, Table 4.1 and Table 4.2 provide the mean integrated squared error (MISE) of the estimators of the coefficient functions β_k, where MISE is defined by

$$\text{MISE}_k(\hat{\beta}_k) = E\int [\hat{\beta}_k(t, z_k) - \beta_k(t, z_k)]^2 dz_k dt,$$

for the estimator $\hat{\beta}_k$. Figures 4.5 and 4.6 show the squared biases of the estimated functions and the variances of the estimated functions for Cases 1 and 2. These plots and tables show that the errors originate from the variances of the estimated functions, especially at the boundaries. Also, we compare the prediction error with the full-dimensional estimator $\tilde{\boldsymbol{\beta}}$, which is defined on the space $L_2(\hat{\mathbf{M}}_t)$. Table 4.3 provides the average of the squared prediction errors of 100 datasets for each case. The smooth backfitting estimator is always better than the full-dimensional estimator because the dataset is generated by the function space $\mathcal{H}(\hat{\mathbf{M}}_t)$. In the simulations for Cases 1 and 2, the smooth backfitting algorithm converges within 5 and 20 iterations, respectively.

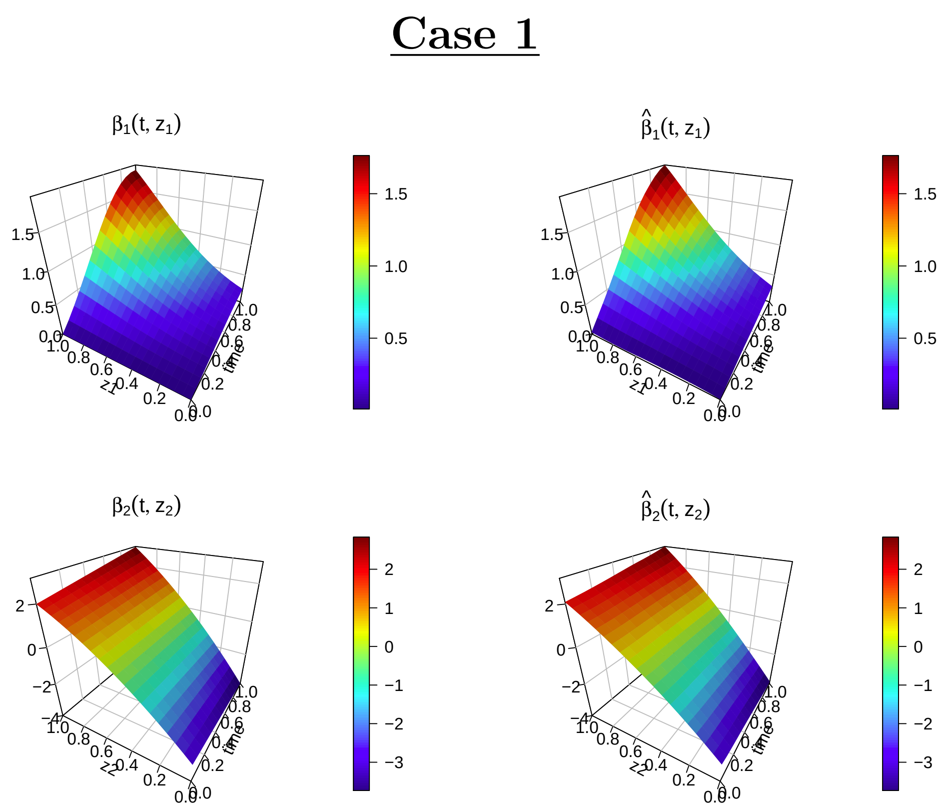

Figure 4.3: The first column plots are the true coefficient functions, while the second column plots are the average of 100 estimated functions for Case 1.

Case 2

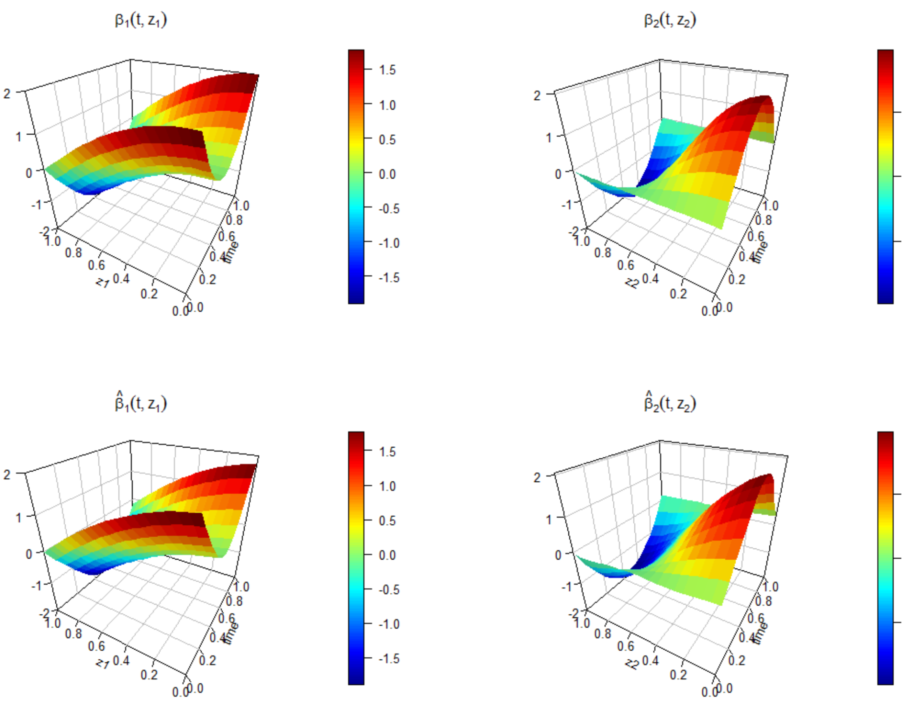

Figure 4.4: The first column plots are the true coefficient functions, and the second column plots are the average of 100 estimated functions for Case 2.

Sample size		Case 1
$n = 200$	MISE_1	0.0051
	MISE_2	0.0047
$n = 400$	MISE_1	0.0031
	MISE_2	0.0036

Table 4.1: The mean integrated squared errors (MISEs) of the estimators $\hat{\beta}_k$ for Case 1

Sample size		Case 2
$n = 200$	MISE_1	0.0245
	MISE_2	0.0442
$n = 400$	MISE_1	0.0170
	MISE_2	0.0335

Table 4.2: The mean integrated squared errors (MISEs) of the estimators $\hat{\beta}_k$ for Case 2

Case 1

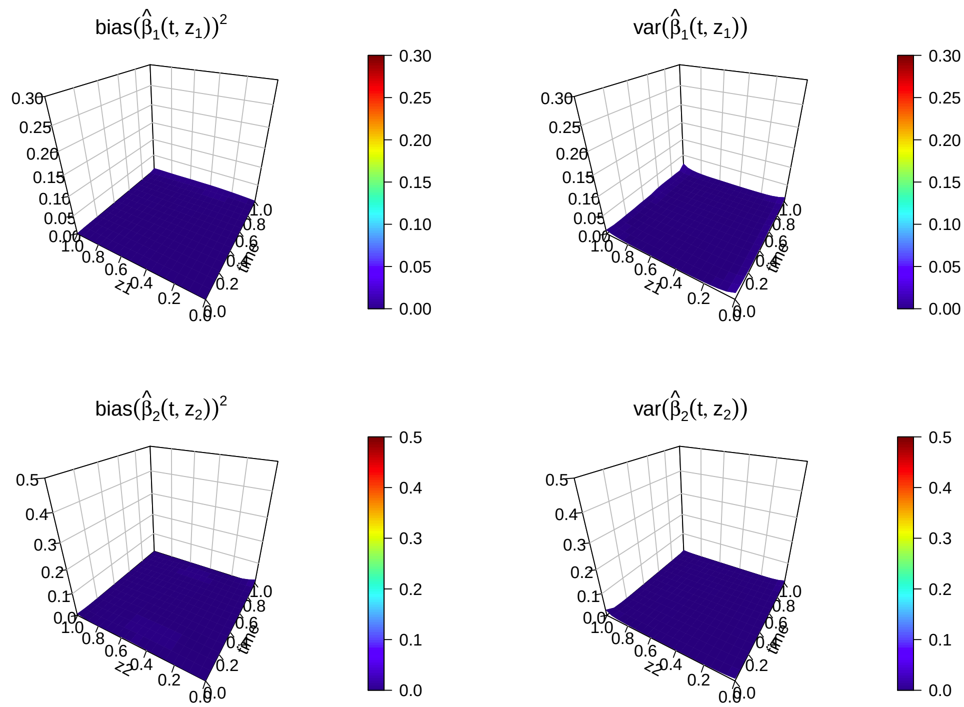

Figure 4.5: The first column plots are the squared biases of the estimated functions, while the second column plots are the variances of the estimated functions for Case 1.

Case 2

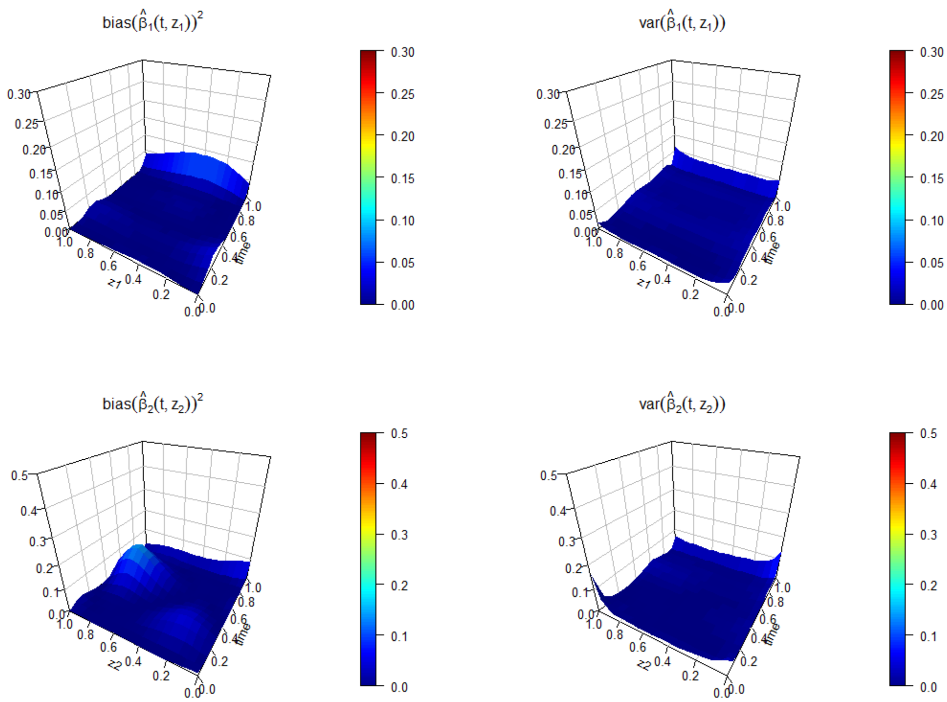

Figure 4.6: The first column plots are the squared biases of the estimated functions, and the second column plots are the variances of the estimated functions for Case 2.

	Sample size	$\hat{\boldsymbol{\beta}}$	$\tilde{\boldsymbol{\beta}}$
Case 1	$n = 200$	0.0147	0.0284
	$n = 400$	0.0141	0.0161
Case 2	$n = 200$	0.0293	0.0506
	$n = 400$	0.0288	0.0311

Table 4.3: The average of the squared prediction errors for the smooth back-fitting estimator and full-dimensional estimator.

4.2 Real Data Analysis

Recently, the government as well as the private have been focusing on air quality changes. It is known that air quality can affect not only public health, life, and industrial activities, but also ecosystems. Conversely, it is known that human and industrial activities affect air quality. In this real data analysis study, we are especially interested in particulate matter (PM_{10}), which consists of solid particles or liquid droplets of diameter 10 μm or less. These particles are of various shapes and sizes as they are generated not only from natural sources, but also by the reactions between gases like sulfur oxides (SOx) and nitrogen oxides (NOx), which are the by-products of traffic and industrial activities. We sourced the air quality data from AIRKOREA (http://www.airkorea.or.kr/), a real-time monitor of air quality and the annual report of Ambient Air Quality in Korea, published by the Ministry of Environment (MOE, 2010-2014). The collected data were measured at hourly intervals in Seoul, Korea, during the period from 2010 to 2014 and are presented as concentrations of air pollutants (PM_{10}, SO_2, CO, O_3, and NO_2).

It is well known that air quality is very sensitive to weather conditions. Accordingly, we collected meteorological data from the National Climate Data Service System (NCDSS, http://sts.kma.go.kr/) and the annual report of the Korea Meteorological Administration (KMA, 2010-2014). The data were measured on an hourly basis in Seoul, Korea, during the period from 2010 to 2014 and consist of surface observations such as temperature, precipitation, humidity, and atmospheric press. Notably, the rainfall in Korea is concentrated in the summer time and lowers the PM_{10} concentration, whereas in spring and winter, the values of PM_{10} are high (Kim et al., 2014). In this study, we focus on the daytime concentration during the spring season and consider the relationships between air quality and the meteorological data. Many researchers have studied the relationship between air quality and meteorological factors, including Habeebullah et al. (2015) for Saudi Arabia and Marković et al. (2008) for Serbia. Munir et al. (2013) applied the generalized additive model to predict PM_{10} concentrations using meteorological data. We use the PM_{10} concentration ($\mu g/m^3$) as the response variable, the concentrations of SO_2 (ppm) and NO_2 (ppm) as the covariates, and temperature (°C) and humidity (%) as the smoothing variables, so that the effects of SO_2 and NO_2 are modified by weather-related actors. The dataset consists of observations measured at 11 time points (from 8 a.m. to 6 p.m.) in spring (March, April, and May) for the period from 2010 to 2014. Thus, there are 416 observations at 11 time points for five variables. We consider the varying coefficient model

$$\mathrm{PM}_{10}(t) = \beta_1(t, \mathrm{temperature}(t))\mathrm{SO}_2(t) + \beta_2(t, \mathrm{humidity}(t))\mathrm{NO}_2(t) + (\mathrm{noise}).$$

We illustrate the estimated coefficient functions in Figures 4.7 and 4.8. Figure 4.7 shows that in the evening time, the covariate SO_2 affects the response of PM_{10} to a higher extent than in the morning. When the temperature is low, the impact of SO_2 on the PM_{10} concentration weakens. Similarly, the NO_2 level affects the response of PM_{10} in morning, and its impact decreases when

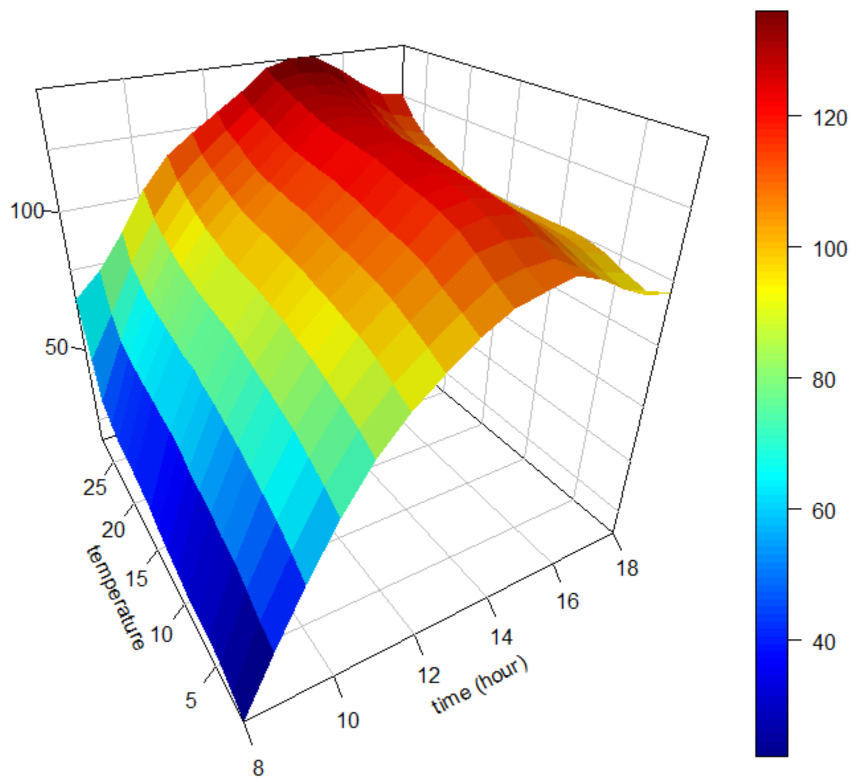

Figure 4.7: This figure plots the estimated coefficient function $\hat{\beta}_1$ with respect to time and temperature.

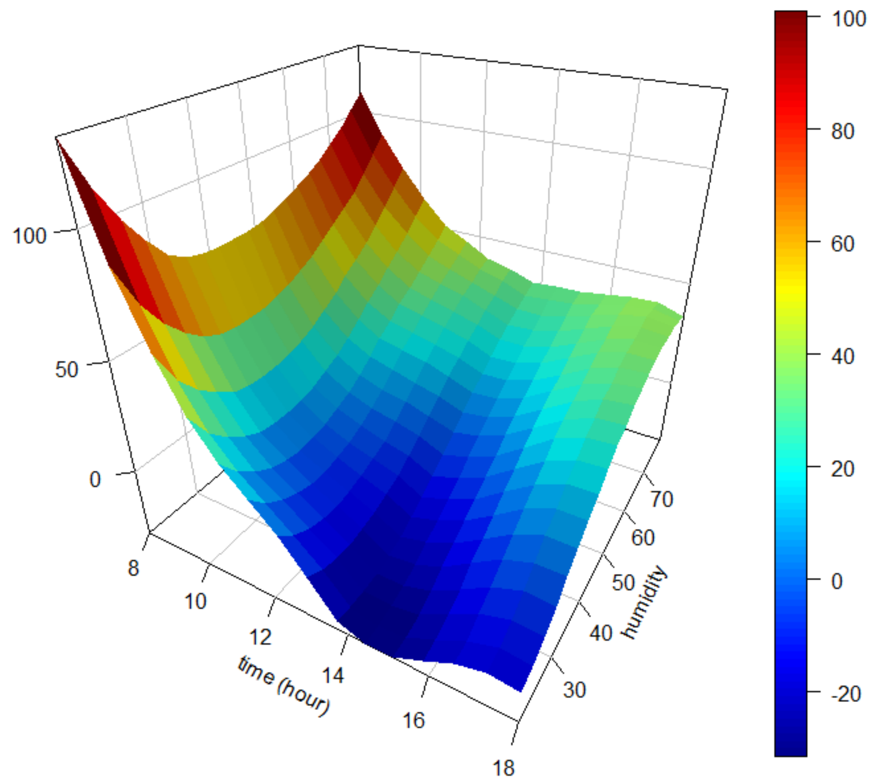

Figure 4.8: This figure plots the estimated coefficient function $\hat{\beta}_2$ with respect to time and humidity.

the humidity is moderate. This result is supported by Yoo et al. (2015). The concentration of PM_{10} and the peaks corresponding to the other air pollutants are due to the increasing traffic and industrial activity in the morning and the planetary boundary layer in the evening.

In this real data analysis, we use temperature and humidity as the smoothing variables, but in fact, precipitation is one of the most important elements of weather-related data pertaining to air pollution. The rainfall washes out the pollutants in the air, which reduces the PM_{10} concentration rapidly. In future studies, precipitation data can be applied to set the lead time. Also, it is well-known and has been illustrated by Yoo et al. (2015) that air quality differs by season. If we consider the seasonal effect and the lead time of rainfall, we can estimate the coefficient functions for the whole year more accurately.

Chapter 5

Technical Details

Proof of Theorem 1

To prove Theorem 1, we show that there exists a function γ such that $0 < \gamma(t) < 1$ and

$$\sup\left\{\|\Phi_t\mathbf{f}\|^2_{\mathbf{M}_t} : \mathbf{f} \in \mathcal{H}(\mathbf{M}_t) \text{ and } \|\mathbf{f}\|^2_{\mathbf{M}_t} \leq 1\right\} \leq \gamma(t) \tag{5.1}$$

for all $t \in [0,1]$. For independent data, (5.1) can be derived by proving that the projection operators $\Pi^t_k : \mathcal{H}_l(\mathbf{M}_t) \to \mathcal{H}_k(\mathbf{M}_t)$ for all $1 \leq k \neq l \leq d$ are Hilbert–Schmidt. [Appendix A.4.2 of Bickel, Klaassen, Ritov and Wellner (1993)]

For our setting, we show

(i) γ is bounded on $[0,1]$,

(ii) $\sup_{t \in I_0} \sup\{\|(\hat{\Phi}^t - \Phi^t)\mathbf{f}\|^2_{\mathbf{M}_t} : \mathbf{f} \in \mathcal{H}(\hat{\mathbf{M}}_t) \text{ and } \|\mathbf{f}\|^2_{\mathbf{M}_t} \leq 1\} \xrightarrow{p} 0$, and

(iii) $\sup_{t\in[0,1]}\sup\{\|(\hat{\Phi}^t-\Phi^t)\mathbf{f}\|^2_{\mathbf{M}_t}:\mathbf{f}\in\mathcal{H}(\hat{\mathbf{M}}_t)$ and $\|\mathbf{f}\|^2_{\mathbf{M}_t}\le 1\}$ is bounded in probability,

as in Zhang et al. (2013).

To prove (i), we first note that

$$\gamma(t)=1-\prod_{k=1}^{d-1}\sin^2(a_k(t)),$$

where $a_k(t)$ is the minimal angle between the two subspaces $\mathcal{H}_k(\mathbf{M}_t)$ and $\mathcal{M}_k(\mathbf{M}_t)\equiv\mathcal{H}_{k+1}(\mathbf{M}_t)+\cdots+\mathcal{H}_d(\mathbf{M}_t)$. Then, we prove $\cos(a_k(t))$ is continuous in t and thus (i) holds.

To prove (ii) and (iii), it suffices to show that for each $1\le k\le d$

$$\sup_{t\in I_0}\sup\left\{\|(\hat{\Pi}^t_k-\Pi^t_k)\mathbf{f}\|^2_{\mathbf{M}_t}:\mathbf{f}\in\mathcal{H}(\hat{\mathbf{M}}_t)\text{ and }\|\mathbf{f}\|^2_{\mathbf{M}_t}\le 1\right\}=o_p(1),$$
$$\sup_{t\in[0,1]}\sup\left\{\|(\hat{\Pi}^t_k-\Pi^t_k)\mathbf{f}\|^2_{\mathbf{M}_t}:\mathbf{f}\in\mathcal{H}(\hat{\mathbf{M}}_t)\text{ and }\|\mathbf{f}\|^2_{\mathbf{M}_t}\le 1\right\}=O_p(1).$$

We know that

$$\begin{aligned}&(\hat{\Pi}^t_k-\Pi^t_k)\mathbf{f}_l\\ =&\int\left(\hat{\boldsymbol{\Psi}}_k(t,z_k)^{-1}\hat{\boldsymbol{\Psi}}_{kl}(t,z_k,z_l)-\boldsymbol{\Psi}_k(t,z_k)^{-1}\boldsymbol{\Psi}_{kl}(t,z_k,z_l)\right)\mathbf{f}_l(z_l)dz_l.\end{aligned}$$

From the Hölder's inequality, it follows that, for $\mathbf{f} \in \mathcal{H}(\mathbf{M}_t)$

$$\|(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}\|_{\mathbf{M}_t} = \|\sum_{l\neq k}^{d}(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}_l\|_{\mathbf{M}_t}$$

$$\begin{aligned}
\leq\ & \sum_{l\neq k}^{d} \|(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}_l\|_{\mathbf{M}_t} \\
=\ & \sum_{l\neq k}^{d} \left[\int \{(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}_l\}^\top \mathbf{\Psi}_k(t, z_k)\{(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}_l\} dz_k\right]^{1/2} \\
\leq\ & \sum_{l\neq k}^{d} \left[\int \Big(\mathbf{\Psi}_l(t, z_l)^{-1}\hat{\mathbf{\Psi}}_k(t, z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t, z_k, z_l)\right. \\
& \qquad - \mathbf{\Psi}_l(t, z_l)^{-1}\mathbf{\Psi}_k(t, z_k)^{-1}\mathbf{\Psi}_{kl}(t, z_k, z_l)\Big)^\top \mathbf{\Psi}_l(t, z_l)\mathbf{\Psi}_k(t, z_k) \\
& \qquad \times \Big(\mathbf{\Psi}_l(t, z_l)^{-1}\hat{\mathbf{\Psi}}_k(t, z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t, z_k, z_l) \\
& \qquad \left. - \mathbf{\Psi}_l(t, z_l)^{-1}\mathbf{\Psi}_k(t, z_k)^{-1}\mathbf{\Psi}_{kl}(t, z_k, z_l)\Big) dz_k dz_l\right]^{1/2} \\
& \quad \times \left[\int \mathbf{f}_l(z_l)^\top \mathbf{\Psi}_l(t, z_l)\mathbf{f}_l(z_l) dz_l\right]^{1/2} \\
\leq\ & \sum_{l\neq k}^{d} \left[\int \|\mathbf{\Psi}_l(t, z_l)^{-1}\hat{\mathbf{\Psi}}_k(t, z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t, z_k, z_l)\right. \\
& \qquad - \mathbf{\Psi}_l(t, z_l)^{-1}\mathbf{\Psi}_k(t, z_k)^{-1}\mathbf{\Psi}_{kl}(t, z_k, z_l)\|^2 \\
& \qquad \left. \times \mathbf{\Psi}_l(t, z_l)\mathbf{\Psi}_k(t, z_k) dz_k dz_l\right]^{1/2} \\
& \quad \times \|\mathbf{f}_l\|_{\mathbf{M}_t} \\
\leq\ & \sum_{l\neq k}^{d} \|\mathbf{f}_l\|_{\mathbf{M}_t}\left[\int \|\mathbf{R}_{kl}(t, z_k, z_l)\|^2 \mathbf{\Psi}_l(t, z_l)\mathbf{\Psi}_k(t, z_k) dz_k dz_l\right]^{1/2},
\end{aligned}$$

where

$$\mathbf{R}_{kl}(t, z_k, z_l) = \boldsymbol{\Psi}_l(t, z_l)^{-1}\hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1}\hat{\boldsymbol{\Psi}}_{kl}(t, z_k, z_l) \\ - \boldsymbol{\Psi}_l(t, z_l)^{-1}\boldsymbol{\Psi}_k(t, z_k)^{-1}\boldsymbol{\Psi}_{kl}(t, z_k, z_l).$$

We can show

$$\sup_{t\in I_0, z_k\in I_k, z_l\in I_l} |\mathbf{R}_{kl}(t, z_k, z_l)| = o_p(1), \\ \sup_{t\in[0,1], z_k\in\mathbb{Z}_k, z_l\in\mathbb{Z}_l} |\mathbf{R}_{kl}(t, z_k, z_l)| = O_p(1). \tag{5.2}$$

This completes the proofs of (ii) and (iii). □

Proof of Lemma 1

By standard techniques of kernel smoothing, we can show the uniform convergence of the kernel weighted average

$$\sup_{t\in[0,1], z_k\in\mathbb{Z}_k} \left| \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk})\varepsilon_{ij} \right| = O_p\left(n^{-1/3}\sqrt{\log n}\right). \tag{5.3}$$

As shown in Section 3.1, $\hat{\Pi}_k^t$ denotes the projection operator onto $\mathcal{H}_k(\hat{\mathbf{M}}_t)$ and $\hat{\Phi}_t = (I - \hat{\Pi}_d^t) \times \cdots \times (I - \hat{\Pi}_1^t)$. Then, we can show that $\hat{\boldsymbol{\beta}} = \sum_{s=0}^{\infty}(\hat{\Phi}_t)^s\hat{\mathbf{r}}$, where $\hat{\mathbf{r}} = (I - \hat{\Phi}_t)\tilde{\boldsymbol{\beta}}$. Also, $\hat{\boldsymbol{\beta}}^A = \sum_{s=0}^{\infty}(\hat{\Phi}_t)^s\hat{\mathbf{r}}^A$, where $\hat{\mathbf{r}}^A = (I - \hat{\Phi}_t)\tilde{\boldsymbol{\beta}}^A$.

If we prove that

$$\sup_{t\in I_0, \mathbf{z}\in I} \left| \sum_{s=1}^{\infty} (\hat{\Phi}_t)^s \hat{\mathbf{r}}^A(t, \mathbf{z}) \right| = o_p(n^{-1/3}), \tag{5.4}$$

then this implies (3.3) for the case $k = 1$. As in the proof of Theorem 2 in Lee, Mammen and Park (2012b), to prove (5.4), it suffices to show that

$$\sup_{t\in I_0, z_k\in I_k} \left| \int \hat{\mathbf{\Psi}}_k(t,z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l \right| = o_p(n^{-1/3}), \tag{5.5}$$

$$\sup_{t\in I_0} \left(\int \left| \int \hat{\mathbf{\Psi}}_k(t,z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l \right|^2 \mathbf{\Psi}_k(t,z_k)dz_k \right)^{1/2} = o_p(n^{-1/3}).$$

The integral on the left-hand side of (5.5)

$$\begin{aligned} &\int \hat{\mathbf{\Psi}}_k(t,z_k)^{-1}\hat{\mathbf{\Psi}}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l \\ &= \int \mathbf{R}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)\mathbf{\Psi}_l(t,z_l)dz_l + \int \mathbf{\Psi}_k(t,z_k)^{-1}\mathbf{\Psi}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l. \end{aligned}$$

By (5.3), it follows that $\sup_{t\in I_0, z_l\in I_l} \left| \tilde{\boldsymbol{\beta}}_l^A(t,z_l) \right| = O_p(n^{-1/3}\sqrt{\log n})$. Then, with (5.2),

$$\sup_{t\in I_0, z_k\in I_k} \left| \int \mathbf{R}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)\mathbf{\Psi}_l(t,z_l)dz_l \right| = o_p(n^{-1/3}),$$

$$\sup_{t\in I_0} \left(\int \left| \int \mathbf{R}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)\mathbf{\Psi}_l(t,z_l)dz_l \right|^2 \mathbf{\Psi}_k(t,z_k)dz_k \right)^{1/2} = o_p(n^{-1/3}).$$

Similarly,

$$\sup_{t\in I_0, z_k\in I_k} \left| \int \mathbf{\Psi}_k(t,z_k)^{-1}\mathbf{\Psi}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l \right| = o_p(n^{-1/3}),$$

$$\sup_{t\in I_0} \left(\int \left| \int \mathbf{\Psi}_k(t,z_k)^{-1}\mathbf{\Psi}_{kl}(t,z_k,z_l)\tilde{\boldsymbol{\beta}}_l^A(t,z_l)dz_l \right|^2 \mathbf{\Psi}_k(t,z_k)dz_k \right)^{1/2} = o_p(n^{-1/3}).$$

This completes the proof of Lemma 1. □

Proof of Lemma 2

We define

$$\tilde{\boldsymbol{\beta}}^{B,1}(t,\mathbf{z}) = \frac{1}{2}\ \hat{\mathbf{M}}(t,\mathbf{z})^{-1}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}(t,\mathbf{z};\mathbf{X}_{ij})K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})$$
$$\times\ \left[\sum_{k=1}^{d}X_{ijk}\left(\frac{T_{ij}-t}{h_0}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial t^2}\right){h_0}^2 + \sum_{k=1}^{d}X_{ijk}\left(\frac{Z_{ijk}-z_k}{h_k}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial {z_k}^2}\right){h_k}^2\right]$$

and $\tilde{\boldsymbol{\beta}}^{B,2}(t,\mathbf{z}) = \tilde{\boldsymbol{\beta}}^{B}(t,\mathbf{z}) - \boldsymbol{\beta}(t,\mathbf{z}) - \tilde{\boldsymbol{\beta}}^{B,1}(t,\mathbf{z})$. We define $\hat{\boldsymbol{\beta}}_k^{B,1}$ and $\hat{\boldsymbol{\beta}}_k^{B,2}$ to be the solution of the backfitting equation (2.3), with $\tilde{\boldsymbol{\beta}}_k$ being replaced by $\hat{\Pi}_k^t\tilde{\boldsymbol{\beta}}^{B,1}$ and $\hat{\Pi}_k^t\tilde{\boldsymbol{\beta}}^{B,2}$.

We can show

$$\sup_{t\in I_0, z_k\in I_k}\hat{\boldsymbol{\beta}}_k^{B,2}(t,\mathbf{z}) = o_p(n^{-1/3}),$$

as in the proof of Lemma 1 and the Taylor expansion of $\beta_k(T_{ij},Z_{ijk})$ at (t,z_k),

$$\beta_k(t,z_k) + \frac{T_{ij}-t}{h_0}\cdot\frac{\partial\beta_k(t,z_k)}{\partial t}h_0 + \frac{Z_{ijk}-z_k}{h_k}\cdot\frac{\partial\beta_k(t,z_k)}{\partial z_k}h_k$$
$$+\ \frac{1}{2}\left[\left(\frac{T_{ij}-t}{h_0}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial t^2}\right){h_0}^2 + \left(\frac{Z_{ijk}-z_k}{h_k}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial {z_k}^2}\right){h_k}^2\right]$$
$$= \mathbf{w}_{ijk}(t,z_k)^\top\boldsymbol{\beta}_k(t,z_k)$$
$$+\ \frac{1}{2}\left[\left(\frac{T_{ij}-t}{h_0}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial t^2}\right){h_0}^2 + \left(\frac{Z_{ijk}-z_k}{h_k}\right)^2\left(\frac{\partial^2\beta_k(t,z_k)}{\partial {z_k}^2}\right){h_k}^2\right].$$

We can prove that

$$
\begin{aligned}
&\int \hat{\mathbf{M}}_k(t,\mathbf{z})^\top \tilde{\boldsymbol{\beta}}^{B,1}(t,\mathbf{z}) d\mathbf{z}_{-k} \\
&= \frac{1}{2} \int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{w}_{ijk}(t,z_k) X_{ijk} K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}) \\
&\qquad \times \sum_{k=1}^{d} \left[X_{ijk} \left(\frac{T_{ij}-t}{h_0} \right)^2 \left(\frac{\partial^2 \beta_k(t,z_k)}{\partial t^2} \right) {h_0}^2 \right. \\
&\qquad\qquad \left. + X_{ijk} \left(\frac{Z_{ijk}-z_k}{h_k} \right)^2 \left(\frac{\partial^2 \beta_k(t,z_k)}{\partial {z_k}^2} \right) {h_k}^2 \right] d\mathbf{z}_{-k} \\
&= \frac{1}{2} E(X_k^2 | T=t, Z_k=z_k) p_k(t,z_k) \left(\int u^2 K(u) du \right) \mathbf{1}_1 \\
&\quad \times \left[\left(\frac{\partial^2 \beta_k(t,z_k)}{\partial t^2} \right) {h_0}^2 + \left(\frac{\partial^2 \beta_k(t,z_k)}{\partial {z_k}^2} \right) {h_k}^2 \right] \\
&\quad + \frac{1}{2} \sum_{l \neq k}^{d} \int E(X_k X_l | T=t, Z_k=z_k, Z_l=z_l) p_{kl}(t,z_k,z_l) \left(\int u^2 K(u) du \right) \mathbf{1}_1 \\
&\quad \times \left[\left(\frac{\partial^2 \beta_l(t,z_l)}{\partial t^2} \right) {h_0}^2 + \left(\frac{\partial^2 \beta_l(t,z_l)}{\partial {z_l}^2} \right) {h_l}^2 \right] dz_l \\
&\quad + o_p(n^{-1/3})
\end{aligned}
$$

uniformly for $t \in I_0$ and $z_k \in I_k$.

We also have

$$
\begin{aligned}
&\int \hat{\mathbf{M}}_k(t,\mathbf{z})^\top \hat{\boldsymbol{\beta}}^{B,1}(t,\mathbf{z}) d\mathbf{z}_{-k} \\
= &\int \sum_{l=1}^{d} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{v}_{ijk}(t,z_k;X_{ijk}) \mathbf{v}_{ijl}(t,z_l;X_{ijl})^\top K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}) \\
&\qquad \times \hat{\boldsymbol{\beta}}_l^{B,1}(t,z_l) d\mathbf{z}_{-k} \\
= &\frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{w}_{ijk}(t,z_k) \mathbf{w}_{ijk}(t,z_k)^\top {X_{ijk}}^2 K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk}) \hat{\boldsymbol{\beta}}_k^{B,1}(t,z_k) \\
&+ \sum_{l \neq k}^{d} \int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathbf{w}_{ijk}(t,z_k) \mathbf{w}_{ijl}(t,z_l)^\top X_{ijk} X_{ijl} \\
&\qquad \times K_{h_0,h_k,h_l}(t,z_k,z_l;T_{ij},Z_{ijk},Z_{ijl}) \hat{\boldsymbol{\beta}}_l^{B,1}(t,z_l) dz_l \\
= &\boldsymbol{\Psi}_k(t,z_k) \hat{\boldsymbol{\beta}}_k^{B,1}(t,z_k) + \sum_{l \neq k}^{d} \int \boldsymbol{\Psi}_{kl}(t,z_k,z_l) \hat{\boldsymbol{\beta}}_l^{B,1}(t,z_l) dz_l + o_p(n^{-1/3})
\end{aligned}
$$

uniformly for $t \in I_0$ and $z_k \in I_k$. Comparing the two systems of equations, we conclude that

$$
\hat{\boldsymbol{\beta}}_k^{B,1} = \frac{1}{2} \left(\int u^2 K(u) du \right) \mathbf{1}_1 \left[\frac{\partial^2 \beta_k(t,z_k)}{\partial t^2} h_0^2 + \frac{\partial^2 \beta_k(t,z_k)}{\partial z_k^2} h_k^2 \right] + o_p(n^{-1/3})
$$

uniformly for $t \in I_0$ and $z_k \in I_k$. This completes the proof of Lemma 2. □

Proof of Theorem 2

By the standard techniques of kernel smoothing, the stochastic term $\tilde{\boldsymbol{\beta}}_k^A(t, z_k)$ has mean zero and is asymptotically normal. Since

$$h_0 h_k \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \mathrm{var}\left[\mathbf{w}_{ijk}(t, z_k) K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk}) X_{ijk} Y_{ij} | T_{ij}, \mathbf{Z}_{ij}, \mathbf{X}_{ij}\right]$$

$$= \mathbf{D}_2 E\left[X_k^2 \sigma^2(T, \mathbf{Z}, \mathbf{X}) | T = t, Z_k(T) = z_k\right] p_k(t, z_k) + o_p(1),$$

and $\hat{\boldsymbol{\Psi}}_k(t, z_k) = \boldsymbol{\Psi}_k(t, z_k) + o_p(1)$, we find that the asymptotic variance of $\tilde{\boldsymbol{\beta}}_k^A(t, z_k)$ equals

$$(nEN)^{-1} h_0^{-1} h_k^{-1} (\mathbf{D}^{-1} \mathbf{D}_2 \mathbf{D}^{-1}) \frac{E\left[X_k^2 \sigma^2(T, \mathbf{Z}, \mathbf{X}) | T = t, Z_k(T) = z_k\right]}{\left[E\left(X_k^2 | T = t, Z_k(T) = z_k\right)\right]^2 p_k(t, z_k)},$$

where

$$\mathbf{D}_2 = \begin{pmatrix} \mu_0(K^2)^2 & \mu_1(K^2)\mu_0(K^2) & \mu_1(K^2)\mu_0(K^2) \\ \mu_1(K^2)\mu_0(K^2) & \mu_2(K^2)\mu_0(K^2) & \mu_1(K^2)\mu_1(K^2) \\ \mu_1(K^2)\mu_0(K^2) & \mu_1(K^2)\mu_1(K^2) & \mu_2(K^2)\mu_0(K^2) \end{pmatrix},$$

$\mathbf{D} = \mathrm{diag}\Big((1, \mu_2(K), \mu_2(K))\Big)$ and $\mu_p(K^q) = \int u^p K^q(u) du$. This completes the proof of Theorem 2. $\square$

List of Notations

$\beta_k(t, z_k)$: two-dimensional coefficient function of t and z_k

$\beta_{k,1}/h_0$: partial derivatives of β_k with respect to the first argument

$\beta_{k,2}/h_k$: partial derivatives of β_k with respect to the second argument

$\boldsymbol{\beta}_k(t, z_k)$ $= (\beta_k(t, z_k), h_0\partial\beta_k(t, z_k)/\partial t, h_k\partial\beta_k(t, z_k)/\partial z_k)^\top$

$\hat{\boldsymbol{\beta}}(t, \mathbf{z})$ $= (\hat{\boldsymbol{\beta}}_1(t, z_1)^\top, \cdots, \hat{\boldsymbol{\beta}}_d(t, z_d)^\top)^\top$,

proposed estimator, minimizer of $L_t(\mathbf{f})$ over $\mathbf{f} \in \mathcal{H}(\hat{\mathbf{M}}_t)$

$\tilde{\boldsymbol{\beta}}(t, \mathbf{z})$: full-dimensional estimator, minimizer of $L_t(\mathbf{f})$ over $\mathbf{f} \in L_2(\hat{\mathbf{M}}_t)$

$\tilde{\boldsymbol{\beta}}_k(t, z_k)$ $= \hat{\boldsymbol{\Psi}}_k(t, z_k)^{-1}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_{in}}\mathbf{w}_{ijk}(t, z_k)K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk})X_{ijk}Y_{ij}$

δ_{ij} : stochastic variation of the true longitudinal process ($= \delta_i(T_{ij})$)

e_{ij} : measurement error

ε_{ij} $= \delta_{ij} + e_{ij}$

$\|\mathbf{f}\|^2_{\hat{\mathbf{M}}_t}$ $= \int \mathbf{f}(\mathbf{z})^\top\hat{\mathbf{M}}(t, \mathbf{z})\mathbf{f}(\mathbf{z})d\mathbf{z}$

h_0 : bandwidth of t

h_k : bandwidth of z_k

$\mathbf{h}$: bandwidth vector ($= (h_0, h_1, \cdots, h_d)$)

$\mathcal{H}(\hat{\mathbf{M}}_t) \quad = \{\mathbf{f} \in L_2(\hat{\mathbf{M}}_t) : \mathbf{f}_k(\mathbf{z}) = (f_k(z_k), f_{k,1}(z_k), f_{k,2}(z_k))^\top\}$

$\mathcal{H}_k(\hat{\mathbf{M}}_t) \quad = \{\mathbf{f} \in L_2(\hat{\mathbf{M}}_t) : \mathbf{f}_k(\mathbf{z}) = (f_k(z_k), f_{k,1}(z_k), f_{k,2}(z_k))^\top$ and $\mathbf{f}_l \equiv 0$ for $l \neq k\}$

K : baseline kernel function

K_h : boundary corrected kernel that $\int K_h(u; v)du = 1$

$K_{\mathbf{h}}$: product kernel for a bandwidth vector $\mathbf{h}$

$L_t(\mathbf{f})$: objective function

$L_2(\hat{\mathbf{M}}_t) \quad = \{\mathbf{f}(\mathbf{z}) = (\mathbf{f}_1(\mathbf{z})^\top, \cdots, \mathbf{f}_d(\mathbf{z})^\top)^\top :$ $\mathbf{f}_k(\mathbf{z}) = (f_k(\mathbf{z}), f_{k,1}(\mathbf{z}), f_{k,2}(\mathbf{z}))^\top, \|\mathbf{f}\|^2_{\hat{\mathbf{M}}_t} < \infty\}$

$\Lambda \quad = \sum_{k=1}^d \int \left[|\tilde{\boldsymbol{\beta}}_k(t, z_k)|^2 + |\hat{\boldsymbol{\beta}}_k^{[0]}(t, z_k)|^2\right] p_k(t, z_k)dz_k dt$

$\Lambda(t) \quad = \sum_{k=1}^d \int \left[|\tilde{\boldsymbol{\beta}}_k(t, z_k)|^2 + |\hat{\boldsymbol{\beta}}_k^{[0]}(t, z_k)|^2\right] p_k(t, z_k)dz_k$

$\hat{\mathbf{M}}(t, \mathbf{z}) \quad = \mathcal{N}_s^{-1} \sum_{i=1}^n \sum_{j=1}^{N_{in}} \mathbf{v}(t, \mathbf{z}; \mathbf{X}_{ij})\mathbf{v}(t, \mathbf{z}; \mathbf{X}_{ij})^\top K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})$

$\mathbf{M}(t, \mathbf{z}) \quad = p(t, \mathbf{z})\Big[E(\mathbf{X}\mathbf{X}^\top | T = t, \mathbf{Z}(T) = \mathbf{z}) \otimes \mathbf{D}_0$ $+ \text{diag}\left(E(X_k^2 | T = t, \mathbf{Z}(T) = \mathbf{z})\right) \otimes (\mathbf{D} - \mathbf{D}_0)\Big]$

N_{in} : number of time points of ith object

$\mathcal{N}_s \quad = \sum_{i=1}^n N_{ij}$

$\sigma^2(T, \mathbf{Z}, \mathbf{X}) \quad = \text{var}(Y | T, \mathbf{Z}(T), \mathbf{X}(T))$

T_{ij} : jth time point of ith object, $i = 1, \cdots, n$, $j = 1, \cdots, N_{ij}$

$\theta_k(t, z_k) \quad = \frac{1}{2}\left(\int u^2 K(u)du\right)\left[\frac{\partial^2 \beta_k(t,z_k)}{\partial t^2} h_0^2 + \frac{\partial^2 \beta_k(t,z_k)}{\partial z_k^2} h_k^2\right]$

$\boldsymbol{\theta}_k(t, z_k) \quad = \theta_k(t, z_k)\mathbf{1}_1$

$(\Theta(t, \mathbf{z}))_k \quad = \frac{1}{2}\left(\int u^2 K(u)du\right)\left[\frac{\partial^2 \beta_k(t,z_k)}{\partial t^2} c_0^2 + \frac{\partial^2 \beta_k(t,z_k)}{\partial z_k^2} c_k^2\right]$

$$
\begin{aligned}
\mathbf{v}_{ijk}(t, z_k; x_k) &= \mathbf{w}_{ijk}(t, z_k)x_k \\
\mathbf{v}(t, \mathbf{z}; \mathbf{x})^\top &= (\mathbf{w}_{ij1}(t, z_1)^\top x_1, \cdots, \mathbf{w}_{ijd}(t, z_d)^\top x_d) \\
\mathbf{V}(t, \mathbf{z}) &= \operatorname{diag}\left(\left(\int K^2(u)du\right)^2 \frac{E[X_k^2\sigma^2(T,\mathbf{Z},\mathbf{X})|T=t, Z_k(T)=z_k]}{[E(X_k^2|T=t, Z_k(T)=z_k)]^2 p_k(t,z_k)} c_0^{-1} c_k^{-1}\right) \\
\mathbf{w}_{ijk}(t, z_k) &= \left(1, \frac{T_{ij}-t}{h_0}, \frac{Z_{ijk}-z_k}{h_k}\right)^\top \\
\mathbb{Z} &= \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_d \text{ for compact sets } \mathbb{Z}_k \\
\mathbf{1}_1 &= (1, 0, 0)^\top
\end{aligned}
$$

References

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore, MD.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, **17** 453–510.

CAI, Z., FAN, J. and LI, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, **95** 888–902.

CARROLL, R. J., MAITY, A., MAMMEN, E. and YU, K. (2009). Nonparametric additive regression for repeatedly measured data. *Biometrika*, **96** 383–398.

FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, **27** 1491–1518. URL `http://dx.doi.org/10.1214/aos/1017939139`.

Habeebullah, T. M., Munir, S., Awad, A. H. A. A., Morsy, E. A., Seroji, A. R. and Mohammed, A. M. F. (2015). The interaction between air quality and meteorological factors in an arid environment of makkah, saudi arabia. *International Journal of Environmental Science and Development*, **6** 576–580.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** pp. 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85** 809–822.

Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, **14** 763–788.

Kim, S., Hong, K.-H., Jun, H., Park, Y.-J., Park, M. and Sunwoo, Y. (2014). Effect of precipitation on air pollutant concentration in seoul, korea. *Asian Journal of Atmospheric Environment*, **8** 202–211.

KMA (2010-2014). *Annual Report 2014*. Korea Meteorological Administration.

Lee, Y. K., Mammen, E. and Park, B. U. (2012a). Flexible generalized varying coefficient regression models. *The Annals of Statistics*, **40** 1906–1933.

Lee, Y. K., Mammen, E. and Park, B. U. (2012b). Projection-type estimation for varying coefficient regression models. *Bernoulli*, **18** 177–205.

Lee, Y. K., Mammen, E. and Park, B. U. (2014). Backfitting and smooth backfitting in varying coefficient quantile regression. *The Econometrics Journal*, **17** S20–S38.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95** pp. 520–534.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **61** 381–400.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82** 93–100.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, **27** 1443–1490.

Marković, D., Marković, D., Jovanović, A., Lazić, L. and Mijić, Z. (2008). Determination of o3, no2, so2, co and pm10 measured in belgrade urban area. *Environmental Monitoring and Assessment*, **145** 349–359.

MOE (2010-2014). *Annual Report of Ambient Air Quality in Korea*. Ministry of Environment, National Institute of Environmental Research.

Munir, S., Habeebullah, T. M., Seroji, A. R., Morsy, E. A., Mohammed, A. M., Saud, W. A., Abdou, A. E. and Awad, A. H. (2013). Modeling particulate matter concentrations in makkah, applying a statistical modeling approach. *Aerosol and Air Quality Research*, **13** 901–910.

Noh, H. S. and Park, B. U. (2010). Sparse varying coefficient models for longitudinal data. *Statistica Sinica*, **20** 1183–1202.

Park, B. U., Mammen, E., Lee, Y. K. and Lee, E. R. (2015). Varying coefficient regression models: A review and new developments. *International Statistical Review*, **83** 36–64.

Roca-Pardiñas, J. and Sperlich, S. (2009). Feasible estimation in generalized structured models. *Statistics and Computing*, **20** 367–379.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, **13** 689–705.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, **93** 1388–1402.

Wu, C. O. and Yu, K. F. (2002). Nonparametric varying-coefficient models for the analysis of longitudinal data. *International Statistical Review*, **70** 373–393.

Yang, L., Park, B. U., Xue, L. and Härdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, **101** 1212–1227.

Yoo, J.-M., Jeong, M.-J., Kim, D., Stockwell, W. R., Yang, J.-H., Shin, H.-W., Lee, M.-I., Song, C.-K. and Lee, S.-D. (2015). Spatiotemporal variations of air pollutants (o_3, no_2, so_2, co, pm_{10}, and vocs) with land-use types. *Atmospheric Chemistry and Physics*, **15** 10857–10885.

You, J. and Zhou, H. (2007). Two-stage efficient estimation of longitudinal nonparametric additive models. *Statistics & Probability Letters*, **77** 1666 – 1675.

Yu, K., Park, B. U. and Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics*, **36** 228–260.

Zhang, R. and Li, G. (2007). Averaged estimation of functional-coefficient regression models with different smoothing variables. *Statistics & Probability Letters*, **77** 455 – 461.

Zhang, X., Park, B. U. and Wang, J.-L. (2013). Time-varying additive models for longitudinal data. *Journal of the American Statistical Association*, **108** 983–998.

국문초록

이 논문에서 우리는 경시적 자료에 적용 가능한 변수계수모형을 제안하였다. 변수계수모형은 반응변수와 공변량의 관계를 평활 변수에 영향을 받는 계수함수 통하여 설명하는 모형이다. 변수계수모형은 구조화된 비모수 모형이고 공변량의 영향을 해석하기 쉽다. 이러한 변수계수모형의 계수함수를 추정하는 방법에는 여러 가지가 있는데, 우리가 관심이 있는 평활역적합 방법을 통해 추정하면 차원의 저주를 피할 수 있는 장점이 있다. 여기서 우리는 평활역적합 방법을 동적 변수계수모형에 적용 가능하도록 2차원으로 확장하였고, 이 추정량를 추정하는 알고리즘을 제안하였다. 이 알고리즘이 지수적 속도로 수렴함을 보였고, 추정량의 점근적 성질를 보였다. 또한 이 결과는 뒷받침할 수 있는 시뮬레이션 결과와 실제 데이터 분석을 진행하였다.

주요어 : 커널평활법, 경시적 자료분석, 평활역적합, 변수계수모형

학　번 : 2009-20250