Statistical Analysis of Symmetric Positive-definite Matrices

Ying Yuan

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2011

Approved by:

Hongtu Zhu, Advisor

J. S. Marron, Advisor

Joseph. G. Ibrahim, Committee Member

Weili Lin, Committee Member

Haipeng Shen, Committee Member

Abstract

YING YUAN: Statistical Analysis of Symmetric Positive-definite Matrices.
(Under the direction of Hongtu Zhu and J. S. Marron.)

This dissertation is motivated by addressing the statistical analysis of symmetric positive definite (SPD) matrix valued data, which arise in many applications. Due to the nonlinear structure of such data, it is challenging to apply well-established statistical methods to them. Our goal is to develop statistical models and perform statistical inferences on the Riemannian manifold of the space of SPD matrices. This dissertation has three major parts.

In the first part, we develop a local polynomial regression model for the analysis of data with SPD matrix responses on the Riemannian manifold. The independent variable of this model is from Euclidean space. We examine two commonly used metrics including the affine invariant metric and the Log-Euclidean metric on the space of SPD matrices. Under each metric, we develop an associated cross-validation bandwidth selection method, and derive the asymptotic bias, variance, and normality of the intrinsic local constant and local linear estimators and compare their asymptotic mean square errors. Simulation studies are further used to compare the estimators under the two metrics and examine their finite sample performance.

In the second part, we develop a functional data analysis framework to model diffusion tensors along fiber bundles as functional responses with a set of covariates of interest, such as age, diagnostic status and gender, in real applications. We propose a statistical model with varying coefficient functions to characterize the dynamic association between functional SPD matrix-valued responses and covariates. We calculate a weighted least squares estimation of the varying coefficient functions under the Log-Euclidean metric in the space of SPD matrices. We also develop a global test statistic to test specific hypotheses about these coefficient functions and construct their simultaneous confidence bands. Simulated data are further used to examine the finite sample performance of the estimated varying coefficient functions.

The third part is to develop a varying coefficient model framework under the affine invariant metric. This framework is very similar to that in the second part. However, this metric is more complex than the Log-Euclidean metric, which makes the subsequent estimation of the varying coefficient functions and the theoretical derivations very challenging. Since there is no explicit form formula for the estimators, we developed an optimization method for calculating it. We also derive the asymptotic properties for the estimated coefficient functions, which are important for constructing the simultaneous confidence band and the global test statistic. Moreover, comparisons of the statistical powers of the varying coefficient models under the affine invariant and Log-Euclidean metrics are made by using simulated data.

Keywords: Symmetric positive definite matrix, Log-Euclidean metric, affine invariant metric, Local polynomial regression, Functional data analysis, Varying coefficient model, Global test statistic, Simultaneous confidence band.

Acknowledgments

This is a great opportunity to extend my deepest appreciation to all who were involved in the completion of this dissertation.

First of all, I am sincerely grateful to my advisors, Professor Hongtu Zhu and Professor J. S. Marron for their encouragement, direction and support. Throughout my dissertation work, they continually stimulated my analytical thinking through discussions and greatly assisted me with scientific writing. I improved myself a lot in formulating scientific problems and in developing methods to solve these problems.

I would like to thank my committee members, Professor Joseph G. Ibrahim, Professor Weili Lin and Professor Haipeng Shen for their insightful suggestions on this dissertation that led me to greatly improve my work.

Special thanks to Professor Martin Styner and Professor John H. Gilmore from Department of Psychiatry at UNC-Chapel Hill for kindly providing the data that led to interesting applications and Professor Pew-Thian Yap from Department of Radiology for helping with visulization.

Last, but not the least, I would like to express my gratitude to my family, who has been a constant source of love, concern, support and strength all these years. To my parents for their unconditional love and support. To my wonderful husband for listening to me, understanding me, caring for me and relieving me of the daily worries. To my parents-in-law for helping me take care of my twin baby girls. To my lovely baby girls for bringing me happiness and laughter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Symmetric positive-definite (SPD) matrix-valued data occur in a wide variety of important applications. For instance,

(1) An approach to computational anatomy has a representing shape in terms of deformation, where a SPD deformation vector $(JJ^T)^{1/2}$ is computed to capture the directional information of shape change encoded in the Jacobian matrices J at each location in an image (Grenander and Miller, 2007).

(2) An different type of computational anatomy is done in diffusion tensor imaging (DTI) (Basser et al., 1994b), where a 3×3 SPD diffusion tensor (DT), which tracks the effective diffusion of water molecules, is estimated at each voxel (a 3 dimensional ($3D$) pixel) of an imaging space.

(3) Brain function is studied in functional magnetic resonance imaging (fMRI), where a SPD covariance matrix is calculated to delineate functional connectivity between different neural assemblies when the subject is performing a complex cognitive task or perceptual process (Fingelkurts et al., 2005; Gao et al., 2009).

(4) In classical multivariate statistics, a fundamental task is to model and estimate SPD covariance matrices for multivariate measurements, longitudinal data, time series data, among many others (Pourahmadi, 1999, 2000; Anderson, 2003).

In this dissertation, we consider the situation where the data at hand are random SPD matrix-valued samples instead of parameters like in the fourth application. Specifically, we focus on two types of data: (i) the SPD matrix-valued data measured along one curve or (ii) a bunch of curves.

One example application is in DTI, where we obtained: (1) DT's measured along a fiber tract as a function of locations; (2) a collection of DT's along a fiber tract measured as a function of age, diagnostic status, and gender, while controlling for other clinical variables.

Figure 1.1 (a) displays one fiber tract in the brain, called the splenium tract projected onto a slice of the FA image from a DTI scan. Along this tract, we have DT's for each of different subjects with some specific attributes, such as age, gender and diagnostic status. Each DT is geometrically represented by an ellipsoid (Figure 1.1 (b)-(e)). In this representation, the lengths of the semiaxes of the ellipsoid equal the square root of the eigenvalues of a DT, while the eigenvectors define the direction of the three axes.

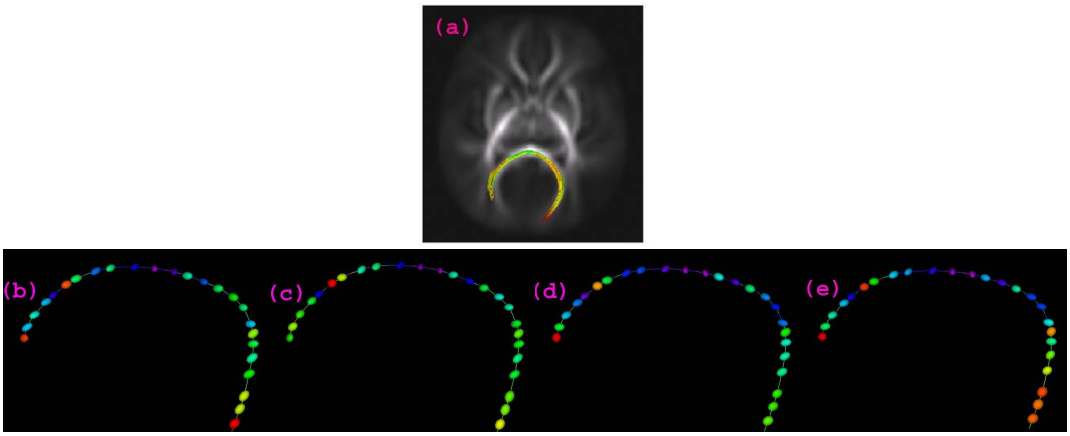

Figure 1.1: (a) shows the splenium tract extracted from the tensor atlas projected onto one axial slice of the FA image from a DTI scan, with color representing FA value. (b)-(e) shows the ellipsoidal representations of DT's along the tract for four selected subjects, with some attributes such as age, gender and diagnostic status, etc..

Four issues arise when we deal with SPD matrix-valued data statistically.

(1) Should the analysis procedure be based on the derived scalar quantities or be based on the whole SPD matrices?

(2) Should the estimation be parametric or nonparametric?

(3) which metric should be used, Euclidean or non-Euclidean metrics?

(4) Which statistical model should be developed for a collection of SPD matrix-valued curves, functional data analysis or others?

In current practice, many statistical analyses of SPD matrix valued data are based on the derived scalar quantities or carried out for each individual element of SPD matrices (Smith et al. (2006); O'Donnell et al. (2009); Gao et al. (2009)). For instance, in DTI, classical statistical methods are applied to fractional anisotropy values (FA is one derived quantity of a diffusion tensor) along the fiber tracts in order to investigate the change of FA's. And, in fMRI, they are applied to inter-regional correlation data in order to explore the development curve of each inter-regional correlation. This is relatively simple and straightforward because many classical statistical methods developed for data on Euclidean space can be directly applied in this situation. However, this method of analysis is not adequate for investigating SPD matrix valued data or their derived scalar quantities because they only use of part of the information from SPD matrices and so may not reveal some important characteristics of the data. Moreover, the derived scalar quantities are linear or nonlinear functions of the estimated eigenvalues of the SPD matrices and thus contain inherent bias. Hence, estimation and inference based on them can be substantially misleading. Our simulation results will make this point clearly. Improvements will be made by developing statistical methods that model the whole SPD matrix as a multivariate unit in this dissertation.

For Data type (i), we have SPD matrix-valued data measured at different spatial locations (or different time points). We wish to denoise the data and reconstruct the underlying SPD matrix-valued function. As is known (Fan and Gijbels (1996)), when some assumed underlying functional forms and distributions are correct, a parametric method is better than a nonparametric method in terms of computational speed and convergence rate of the estimator. However, the reality is that the correct model is rarely available. In this situation, the performance of parametric methods can be very poor and inference based on incorrect assumptions about functional forms and distributions can be highly doubtful. Nonparametric statistical methods can reduce the reliance on the assumptions required for estimation and inference, thereby reducing the opportunities for obtaining misleading results. One contribution of this dissertation is developing a nonparametric regression method for the analysis of data with SPD matrix valued responses. Specifically, we develop the local polynomial regression method for SPD matrix valued data, which can be used to reconstruct the average change of SPD matrices as a function of some covariate, e.g. spatial location, time point or $3D$ coordinate. The estimation process of the mean function involves the choice of metric, which determines the geometry of the space of SPD matrices.

Here we mention three metrics. One is the Euclidean metric and the other two are non-Euclidean metrics: the Log-Euclidean and affine invariant metrics (Arsigny (2006) and Schwartzman (2006)). They correspond to three different geometries: the Euclidean, Log-Euclidean and affine invariant geometries. If we analyze SPD matrix valued data using the Euclidean metric, the usual estimation methods widely studied on Euclidean spaces can be directly applied to them. However, this direct application cannot guarantee the positive-definiteness of the estimates since the estimation highly depends on the noise structure and the structure of the SPD matrices. It is possible to find the closest points on the space of SPD matrices to the non-positive estimators. This falls into the category of extrinsic methods. However, in general, extrinsic methods do not perform well, especially for the statistical inference (Fletcher (2004)). In contrast, estimations using the Log-Euclidean or affine invariant metrics do not suffer from this problem since the operations are carried out by considering the space as a non-Euclidean space, which are intrinsic methods. Hence, in this work, we develop statistical methods with respect to the above two non-Euclidean metrics. Our simulation results also show that in some scenarios, for example at moderate noise levels or for non-gaussian noises, the estimators using the Euclidean metric fail to be SPD matrices and thus cannot outperform those using the Log-Euclidean or affine invariant metrics. Moreover, even if using the Euclidean metric can retain the positive-definiteness in some cases, estimators using this metric suffer from the swelling effect, i.e. the determinant of the Euclidean estimators can be larger than the original determinants. As said in Fletcher (2004), in DTI, diffusion tensors are assumed to be covariance matrices of the local Brownian motion of water molecules. Introducing more dispersion in computations amounts to introducing more diffusion, which is physically unacceptable.

For Data type (ii), we have a collection of SPD matrix-valued data measured along a curve (a temporal or spatial curve) as a function of age, diagnostic status, and gender, while controlling for other clinical variables. These data are in essence functional data. Moreover, one of the main characteristics of these data is the combination of longitudinal or spatial information (time series or spatial data) with cross-sectional information (attribute data). Data can contain trends that vary in longitudinal or spatial aspects, that vary across different groups of subjects or objects. Take the study of fiber tract in DTI as an example. There, each subject is characterized by a set of DT's along one of the fiber tracts in his/her brain coupled with its additional attributes such as age, diagnostic status and gender. The natural approach to dealing with this type of data is functional

data analysis (FDA). Compared with the classical statistics where the interest centers around a set of data vectors, FDA can capture trend, processes and dynamic information which are inherent in the data and thus increase the statistical power in detecting interesting features and in exploring variability. While there is extensive interest in developing FDA methods, however, most research focuses are on functional data with the responses in Euclidean space. The functional data in this dissertation have SPD matrix-valued responses, which lie in a non-Euclidean space. Those methods for the data on Euclidean space cannot be directly applied.

Another major contribution of this dissertation is the development of varying coefficient model frameworks for the analysis of SPD matrix valued functions and their association with a set of covariates of interest, such as age, diagnostic status and gender, in real applications. Our modeling will be formulated under the two affine invariant metrics mentioned above. The varying coefficient model framework consists of four integrated components:

(i) a varying coefficient model for characterizing the association between SPD matrix valued functions with a set of covariates of interest,

(ii) the local polynomial regression method for estimating the coeffient functions in the model,

(iii) global and local test statistics for testing hypotheses of interest

(iv) a resampling or χ^2 approximation method for approximating the p-value of the global test statistic.

The remainder of this dissertation is organized as follows. In Chapter 2, we develop an intrinsic local polynomial regression model for SPD matrix valued data under two Riemannian metrics: affine invariant and Log-Euclidean metrics. Under each metric, we develop an associated cross-validation method for nonparametric analysis of random SPD matrix-valued data and investigate the asymptotic properties of the estimators. Simulation studies are performed to examine the finite sample performance of the estimator. In Chapter 3, we establish the *varying coefficient* model framework under the *Log-Euclidean* metric, called VCLE, for analysis of the association between fiber bundle diffusion tensors and a set of covariates of interest. Under this metric, a varying coefficient model is formulated and correspondingly, an estimation procedure based on the local polynomial regression method is proposed to estimate the parameters in the model. We also

develop both local and global test statistics to test hypotheses on the varying coefficient functions and a resampling method is used to approximate the p-value of the test statistics. We construct a simultaneous confidence band to quantify the uncertainty in the estimated coefficient functions and propose a resampling method to approximate the critical point. We examine the finite sample performance of VCLE via simulation studies. In Chapter 4, we formulate the *varying coefficient* model framework developed in Chapter 3 under another commonly used metric for the space of SPD matrices, the *affine invariant* metric, called VCAI. Specification of the model under this metric involves the matrix exponential transformation and thus brings many more theoretical and computational difficulties than that in Chapter 3. We develop an annealing evolutionary stochastic approximation Monte Carlo algorithm for computing the coefficient functions (Liang (2010)). We also derived the asymptotic properties of the coefficient functions, based on which a global test statistic is developed to test hypotheses on the varying coefficient functions and a simultaneous confidence band is constructed.

Chapter 2

Local Polynomial Regression for SPD Matrices

2.1 Introduction

Symmetric positive-definite (SPD) matrix-valued data occur in many applications. This has motivated the recent development of several methods for statistical analysis of SPD matrices as response variables in a Riemannian manifold. Schwartzman et al. (2008) has proposed several parametric models for SPD matrices and derived the distributions of several test statistics for comparing differences between the means of the two (or multiple) groups of SPD matrices. Kim and Richards (2010) have developed a nonparametric estimator for the common density function of a random sample of positive definite matrices. Zhu et al. (2009) develop a semi-parametric regression model with SPD matrices as responses in a Riemannian manifold and covariates in a Euclidean space. This model extends the two-group models studied by Schwartzman (2006) and Schwartzman et al. (2008) to allow for general covariates and generalizes the parametric modeling in Schwartzman (2006) by assuming only that certain appropriately defined residuals have mean zero. Barmpoutis et al. (2007) and Davis et al. (2010) have proposed tensor splines and local constant regressions for interpolating DTI tensor fields based on the affine invariant metric, but these two papers do not address several important issues of analyzing random SPD matrices including the asymptotic properties of the nonparametric estimate proposed. Recently, Dryden et al. (2009) compare the various choices of metrics of the space of SPD matrices and their properties.

To the best of our knowledge, this is the very first paper for developing an intrinsic local polynomial regression (ILPR) model for estimating an intrinsic *conditional expectation* of a SPD matrix response, S, given a covariate vector x from a set of observations $(x_1, S_1), \cdots, (x_n, S_n)$, where the x_i can be either univariate or multivariate. In practice, x can be the arc-length of a specific fiber

tract (e.g., right internal capsule tract), the coordinates in the 3D imaging space, and demographic variables such as age. Important applications of ILPR include smoothing diffusion tensors along fiber tracts and diffusion tensor fields, quantifying the change of diffusion and deformation tensors across groups and along time, and the evolution of inter-regional functional connectivity matrix with time.

Compared with the existing literature, we make several contributions in this chapter. Since the space of SPD matrices is a curved space, the standard local polynomial regression method is not adequate and can lead to an unpleasant effect in image processing (Chefd'hotel et al., 2004). To account for the curved nature of the SPD space, we consider the affine invariant metric and the Log-Euclidean metric for the space of SPD matrices in order to examine the effect of different metrics on carrying out statistical inference in the Riemannian manifold. Under each metric, we develop the ILPR method for estimating the intrinsic conditional expectation of random SPD responses given the covariate and derive an approximation to its associated cross-validation method for bandwidth selection. We establish the asymptotic properties of the ILPR estimators. We show the superiority of the intrinsic local linear estimator over the intrinsic local constant estimator by examining their asymptotic mean square errors. As shown in Figure 2.7 in Section 2.5, substantial improvements can be made by treating SPD matrices in the Riemannian manifold.

The rest of this chapter is organized as follows. In Section 2.2, we develop the ILPR method and its associated cross-validation method for nonparametric analysis of random SPD matrix-valued data. We investigate the asymptotic properties of the estimators proposed under the affine invariant metric in Section 2.3 and the Log-Euclidean metric in Section 2.4, respectively. We examine the finite sample performance of the estimator via simulation studies in Section 2.5. Finally, we will analyze a real data set to illustrate a real-world application of the proposed ILPR method in Section 2.6 before offering some concluding remarks in Section 2.7.

2.2 Intrinsic Local Polynomial Regression for SPD matrices

Let $\mathrm{Sym}^+(m)$ and $\mathrm{Sym}(m)$ be, respectively, the set of $m \times m$ SPD matrices and the set of $m \times m$ symmetric matrices with real entries. Suppose that $(x_i, S_i), i = 1, \cdots, n$ is an independent and identically distributed random sample, where $S_i \in \mathrm{Sym}^+(m)$. For notational simplicity, we

focus on a univariate covariate throughout the paper. The question of interest is to estimate an 'intrinsic conditional expectation' of S at each x, denoted by $D(x)$, in $\mathrm{Sym}^+(m)$.

Since $\mathrm{Sym}^+(m)$ is a curved space, one cannot directly define $D(x) = E(S|X = x)$ for a random SPD matrix $S \in \mathrm{Sym}^+(m)$. Instead, we introduce an $m \times m$ *residual matrix* of S at $D(x)$, denoted by $\mathcal{E}_D(x)$, in $\mathrm{Sym}(m)$. The space $\mathrm{Sym}(m)$ is a Euclidean space with the Euclidean inner product given by $< A_1, A_2 >= \mathrm{tr}(A_1 A_2)$ for any $A_1, A_2 \in \mathrm{Sym}(m)$. Thus, since $\mathcal{E}_D(x)$ is in a Euclidean space, we can directly compute its conditional expectation, which leads to the definition of 'intrinsic conditional expectation' of S at each x as follows:

$$E\{\mathcal{E}_D(X)|X = x\} = O_m, \tag{2.2.1}$$

where O_m is the $m \times m$ matrix with all elements zero. Heuristically, consider S and $D(x)$ as lying in Euclidean space, then $\mathcal{E}_D(x)$ could be defined as $S - D(x)$ and (2.2.1) would reduce to $D(x) = E(S|X = x)$.

To rigorously define $\mathcal{E}_D(X)$, we review some basic facts about the geometrical structure of $\mathrm{Sym}^+(m)$ in order to introduce *addition* and *subtraction* operations in $\mathrm{Sym}^+(m)$ (Zhu et al., 2009; Schwartzman, 2006; Dryden et al., 2009). We first introduce the tangent vector and tangent space at $D(x)$ in $\mathrm{Sym}^+(m)$. For a small scalar $\delta > 0$, let $C(t)$ be a differentiable map from $(-\delta, \delta)$ to $\mathrm{Sym}^+(m)$ passing through $C(0) = D(x)$. A tangent vector at $D(x)$ is defined as the derivative of the smooth curve $C(t)$ with respect to t evaluated at $t = 0$. The set of all tangent vectors at $D(x)$ forms the tangent space of $\mathrm{Sym}^+(m)$ at $D(x)$, denoted as $T_{D(x)}\mathrm{Sym}^+(m)$. We need to introduce a specific metric $<< \cdot, \cdot >>_{D(x)}$ as an inner product for tangent vectors on $T_{D(x)}\mathrm{Sym}^+(m)$. For the given metric, we can calculate $<< U_{D(x)}, V_{D(x)} >>_{D(x)}$ for any $U_{D(x)}$ and $V_{D(x)}$ on $T_{D(x)}\mathrm{Sym}^+(m)$ and then measure the length of the curve in $\mathrm{Sym}^+(m)$. Furthermore, one can compute the shortest path, called the geodesic, between points in $\mathrm{Sym}^+(m)$. Let $\gamma_{D(x)}(t, U_{D(x)})$ be the geodesic as a function of t passing through $\gamma_{D(x)}(0, U_{D(x)}) = D(x)$ in the direction of $U_{D(x)} \in T_{D(x)}\mathrm{Sym}^+(m)$.

We introduce the Riemannian exponential and logarithmic maps on $\mathrm{Sym}^+(m)$ as the addition and subtraction operations. The Riemannian exponential mapping $\mathrm{Exp}_{D(x)} : T_{D(x)}\mathrm{Sym}^+(m) \to \mathrm{Sym}^+(m)$ is defined as $\mathrm{Exp}_{D(x)}(U_{D(x)}) = \gamma_{D(x)}(1; U_{D(x)})$. Intuitively, the Riemannian exponential map starts at $D(x) \in \mathrm{Sym}^+(m)$ and then moves on $\mathrm{Sym}^+(m)$ in the direction $U_{D(x)} \in$

$T_{D(x)}\text{Sym}^+(m)$. We define an open sphere of radius r centered at the 'origin' $D(x)$ in $T_{D(x)}\text{Sym}^+(m)$ as

$$B_{D(x)}(r) = \{U_{D(x)} \in T_{D(x)}\text{Sym}^+(m) :<< U_{D(x)}, U_{D(x)} >>_{D(x)} < r^2\}. \tag{2.2.2}$$

Since $\gamma_{D(x)}(t; sU_{D(x)}) = \gamma_{D(x)}(st; U_{D(x)})$ and $T_{D(x)}\text{Sym}^+(m)$ is a Euclidean space, there is an $r > 0$ such that the Riemannian exponential map is a diffeomorphism. Inversely, the inverse of the Riemannian exponential map $\text{Log}_{D(x)}(\cdot) = \text{Exp}^{-1}_{D(x)}(\cdot)$ is called the Riemannian logarithmic map from $\text{Sym}^+(m)$ to a vector in $T_{D(x)}\text{Sym}^+(m)$. For small t, we have $\text{Log}_{D(x)}(\gamma_{D(x)}(t; U_{D(x)})) = tU_{D(x)}$. For instance, we consider any two symmetric matrices A_1 and A_2 in the Euclidean space $\text{Sym}(m)$ with the usual Euclidean metric. Thus, we have $\text{Log}_{A_1}(A_2) = A_2 - A_1 \in T_{A_1}\text{Sym}(m)$ and $\text{Exp}_{A_1}(A_2 - A_1) = A_1 + A_2 - A_1 = A_2 \in \text{Sym}(m)$.

We define $\mathcal{E}_D(X)$ to be $\text{Log}_{D(X)}(S)$ in $T_{D(X)}\text{Sym}^+(m)$. Thus, the intrinsic conditional expectation of S at $X = x$ is defined as $D(x) \in \text{Sym}^+(m)$ such that

$$E\{\text{Log}_{D(X)}(S)|X = x\} = O_m. \tag{2.2.3}$$

The next question is to estimate $D(X)$ at each $X = x$ using the observed data $\{(x_i, S_i), i = 1, \cdots, n\}$.

Since $\text{Log}_{D(x)}(S_i)$ is in the Euclidean space $\text{Sym}(m)$, we consider estimating $D(X)$ at $X = x_0$ by minimizing a weighted intrinsic least square criterion given by

$$G_n(D(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0) << \text{Log}_{D(x_0)}(S_i), \text{Log}_{D(x_0)}(S_i) >>_{D(x_0)}, \tag{2.2.4}$$

where $K_h(u) = K(u/h)h^{-1}$, in which h is a positive scalar, and $K(\cdot)$ is a kernel function such as the Epanechnikov kernel (see Fan and Gijbels (1996); Fan and Yao (1998); Wand and Jones (1995) for additional discussion). Since $<< \text{Log}_{D(x_0)}(S), \text{Log}_{D(x_0)}(S) >>_{D(x_0)}$ equals the square of the geodesic distance between S and $D(x_0)$, denoted by $g(D(x_0), S)$, $G_n(D(x_0))$ can be rewritten as

$$G_n(D(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0) g(D(x_0), S_i)^2. \tag{2.2.5}$$

Directly minimizing $G_n(D(x_0))$ with respect to $D(x_0)$ leads to a weighted intrinsic mean of $S_1, \cdots, S_n \in$

$\text{Sym}^+(m)$ at x_0, denoted by $\hat{D}_I(x_0)$, which can be regarded as a generalization of the intrinsic mean of $S_1, \cdots, S_n \in \text{Sym}^+(m)$ (Bhattacharya and Patrangenaru, 2005). This is exactly the intrinsic local constant estimator of $D(x_0)$ considered in Davis et al. (2010).

We propose the intrinsic local polynomial regression for estimating $D(X)$ at $X = x_0$ as follows. Since $D(x)$ is in the curved space, we cannot directly expand $D(x)$ at x_0 by using a Taylor's series expansion. Instead, we consider the logarithmic map of $D(x)$ at $D(x_0)$ in $T_{D(x_0)}\text{Sym}^+(m)$. Since $\text{Log}_{D(x_0)}(D(x))$ for different x_0 are in different tangent spaces, we may rotate them back to the same tangent space $T_{I_m}\text{Sym}^+(m)$ through a rotation mapping (or parallel transport) $\phi_{D(x_0)} : T_{D(x_0)}\text{Sym}^+(m) \to T_{I_m}\text{Sym}^+(m)$, where I_m is an $m \times m$ identity matrix. That is, $Y(x) = \phi_{D(x_0)}(\text{Log}_{D(x_0)}(D(x))) \in T_{I_m}\text{Sym}^+(m)$ and $\text{Log}_{D(x_0)}(D(x)) = \phi^{-1}_{D(x_0)}(Y(x))$, where $\phi^{-1}_{D(x_0)}(\cdot)$ is the inverse map of $\phi_{D(x_0)}(\cdot)$. Moreover, since $Y(x_0) = \phi_{D(x_0)}(O_m) = O_m$ and $Y(x)$ is in the same space $T_{I_m}\text{Sym}^+(m)$, we expand $Y(x)$ at x_0 by using the Taylor's series expansion as follows:

$$\text{Log}_{D(x_0)}(D(x)) = \phi^{-1}_{D(x_0)}(Y(x)) \approx \phi^{-1}_{D(x_0)}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k), \tag{2.2.6}$$

where k_0 is an integer and $Y^{(k)}(x)$ is the $k-$th derivative of $Y(x)$ with respect to x divided by $k!$. Equivalently, $D(x)$ can be approximated by

$$\begin{aligned} D(x) &= \text{Exp}_{D(x_0)}(\phi^{-1}_{D(x_0)}(Y(x))) \\ &\approx \text{Exp}_{D(x_0)}(\phi^{-1}_{D(x_0)}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k)) = D(x, \alpha(x_0), k_0), \end{aligned} \tag{2.2.7}$$

where $\alpha(x_0)$ contains all unknown parameters in $\{D(x_0), Y^{(1)}(x_0), \cdots, Y^{(k_0)}(x_0)\}$.

To estimate $\alpha(x_0)$, we calculate an intrinsic weighted least square estimator of $\alpha(x_0)$ defined by

$$\hat{\alpha}_I(x_0; h) = \text{argmin}_{\alpha(x_0)} G_n(\alpha(x_0)), \tag{2.2.8}$$

where $G_n(\alpha(x_0))$ is given by

$$G_n(\alpha(x_0)) = \sum_{i=1}^{n} K_h(x_i - x_0) g(\text{Exp}_{D(x_0)}(\phi^{-1}_{D(x_0)}(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k)), S_i)^2. \tag{2.2.9}$$

Then we can calculate $D(x, \hat{\alpha}_I(x_0; h), k_0)$, denoted by $\hat{D}_I(x, h)$, as an intrinsic local polynomial regression estimator (ILPRE) of $D(x)$.

We propose to use a leave-one-out cross validation method for bandwidth selection due to its conceptual simplicity. Larger bandwidths may gain on the variance side, but lose on the bias side due to oversmoothing. Smaller bandwidths may gain on the bias side, but loses on the variance side due to undersmoothing. Let $\hat{D}_I^{(-i)}(x_i; h)$ be the estimate of $D(x_i)$ obtained by minimizing $G_n(\alpha(x_i))$ with (x_i, S_i) deleted for a given bandwidth h and all i. The cross-validation score is defined as follows:

$$\mathrm{CV}(h) = n^{-1} \sum_{i=1}^{n} g(S_i, \hat{D}_I^{(-i)}(x_i; h))^2. \tag{2.2.10}$$

The optimal h, denoted by $\hat{h}$, can be obtained by minimizing $\mathrm{CV}(h)$. However, since computing $\hat{D}_I^{(-i)}(x_i; h)$ for all i can be computationally prohibitive, we suggest to use the first-order approximation of $\mathrm{CV}(h)$, whose details will be given below under each specific metric. Although it is possible to develop other bandwidth selection methods, such as plug-in and bootstrap methods (Rice, 1984; Park and Marron, 1990; Hall et al., 1992; Härdle et al., 1992), we must deal with additional computational and theoretical challenges, which will be left for future research.

2.3 Affine Invariant Metric

As discussed in Dryden et al. (2009), various metrics can be defined for tangent vectors on $T_{D(x)}\mathrm{Sym}^+(m)$. To assess the effect of different metrics on ILPREs, we consider two commonly used metrics including the affine invariant and Log-Euclidean metrics and compare the asymptotic properties of ILPREs, such as asymptotic normality, under these two metrics. Furthermore, we systematically compare the intrinsic local constant and linear estimators under each metric.

2.3.1 ILPR under the Affine Invariant Metric

We give the explicit form of $\mathrm{Log}_{D(x)}(S)$ for the affine invariant metric. We add the notation of 'A' into all geometric quantities under the affine invariant metric. Let $\exp(\cdot)$ and $\log(\cdot)$ be the common matrix exponential and logarithm, respectively. A scaled Frobenius inner product on

$T_{D(x)}\mathrm{Sym}^+(m)$ based on the affine invariant metric is defined as

$$<< U_{D(x)}, V_{D(x)} >>_{D(x),A} = \mathrm{tr}(U_{D(x)} D(x)^{-1} V_{D(x)} D(x)^{-1}). \tag{2.3.1}$$

The geodesic $\gamma_{D(x),A}(t; U_{D(x)})$ is given by $G(x)\exp(tG(x)^{-1}U_{D(x)}G(x)^{-T})G(x)^T$ for any t, where $G(x)$ is any square root of $D(x)$ such that $D(x) = G(x)G(x)^T$. Without of loss of generality, $G(x)$ is assumed to be a lower triangular matrix with strictly positive diagonal terms due to the Cholesky decomposition.

The Riemannian exponential and logarithm maps are, respectively, given by

$$\begin{aligned}
&\mathrm{Exp}_{D(x),A}(U_{D(x)}) = \gamma_{D(x),A}(1; U_{D(x)}) = G(x)\exp(G(x)^{-1}U_{D(x)}G(x)^{-T})G(x)^T, \\
&\mathrm{Log}_{D(x),A}(S) = G(x)\log(G(x)^{-1}SG(x)^{-T})G(x)^T.
\end{aligned} \tag{2.3.2}$$

The geodesic distance between $D(x)$ and S, denoted by $g_A(D(x), S)$, is given by

$$\sqrt{\mathrm{tr}\{\log^2(G(x)^{-1}SG(x)^{-T})\}} = \sqrt{\mathrm{tr}\{\log^2(S^{-1/2}D(x)S^{-T/2})\}}, \tag{2.3.3}$$

where $S^{1/2}$ is any square root of S.

We consider two SPD matrices $D(x)$ and $D(x_0) = G(x_0)G(x_0)^T$. For any $U_{D(x_0)} \in T_{D(x_0)}\mathrm{Sym}^+(m)$, the rotation map $\phi_{D(x_0),A}$ is defined by

$$\phi_{D(x_0),A}(U_{D(x_0)}) = G(x_0)^{-1}U_{D(x_0)}G(x_0)^{-T} \in T_{I_m}\mathrm{Sym}^+(m). \tag{2.3.4}$$

Since $\mathrm{Log}_{D(x_0)}(D(x)) \in T_{D(x_0)}\mathrm{Sym}^+(m)$, combining (2.3.2) and (2.3.4) yields that

$$\begin{aligned}
&Y(x) = \phi_{D(x_0),A}(\mathrm{Log}_{D(x_0)}(D(x))) = \log(G(x_0)^{-1}D(x)G(x_0)^{-T}), \\
&D(x) = G(x_0)\exp(Y(x))G(x_0)^T.
\end{aligned} \tag{2.3.5}$$

In this case, $\mathcal{E}_D(X)$ can be defined to be $\log(G(X)^{-1}SG(X)^{-T})$ such that

$$E\{\log(G(X)^{-1}SG(X)^{-T})|X = x\} = O_m. \tag{2.3.6}$$

To compute the ILPR estimator, we use the Taylor's series expansion to expand $Y(x)$ at x_0 as follows:

$$D(x) \approx G(x_0) \exp(\sum_{k=1}^{k_0} Y^{(k)}(x_0)(x - x_0)^k/k!) G(x_0)^T = D_A(x, \alpha_A(x_0), k_0), \tag{2.3.7}$$

where $\alpha_A(x_0)$ contains all unknown parameters in $G(x_0)$ and $Y^{(k)}(x_0)$ for $k = 1, \cdots, k_0$. Thus, we can compute $\hat{\alpha}_{IA}(x_0; h)$ by minimizing $G_n(\alpha_A(x_0))$. Minimizing $G_n(\alpha_A(x_0))$ represents a computational challenge for the affine invariant metric. According to our experience, the standard gradient methods do not perform well for optimizing $G_n(\alpha_A(x_0))$ when $k_0 > 0$. Hence, we develop an annealing evolutionary stochastic approximation Monte Carlo algorithm (see Liang (2010) for good discussion) for computing $\hat{\alpha}_{IA}(x_0; h)$, whose details can be found in the supplementary document.

To simplify the computation of the cross-validation score $\mathrm{CV}_A(h)$, we suggest the first-order approximation to $\mathrm{CV}_A(h)$ as follows:

$$\mathrm{CV}_A(h) \approx n^{-1} \sum_{i=1}^{n} g_A(S_i, \hat{D}_{IA}(x_i; h, k_0))^2 + 2p_n(h), \tag{2.3.8}$$

where $\hat{D}_{IA}(x; h, k_0) = D_A(x, \hat{\alpha}_{IA}(x_0; h), k_0)$. The $\mathrm{CV}_A(h)$ is close to Akaike's information criterion (AIC) (Sakamoto et al., 1999) and $p_n(h)$ can be regarded as degree of freedom. The explicit form of $p_n(h)$ will be presented in Appendix A.

2.3.2 Asymptotic Properties

To understand the statistical properties of $\hat{D}_{IA}(x_0; h)$, we establish the consistency and asymptotic normality of the local polynomial estimators $\hat{\alpha}_{IA}(x_0; h)$.

We need to introduce some notation for discussion. Let $\mathbf{u} = (u_1, \cdots, u_{k_0})^T$ and $\mathbf{v} = (v_1, \cdots, v_{k_0})^T$ be $k_0 \times 1$ vectors and $\mathcal{U}_2 = (u_{i+j})$ and $\mathcal{V}_2 = (v_{i+j})$ for $1 \le i, j \le k_0$ be two $k_0 \times k_0$ matrices with $u_k = \int x^k K(x) dx$ and $v_k = \int x^k K(x)^2 dx$ for $k \ge 0$. Let $f_X(x)$ and $f_X^{(1)}(x)$ be the marginal density function of X and its first-order derivative with respect to x, respectively. Consider a function

$$\psi(S, G, Y) = g_A(S, G \exp(Y) G^T)^2, \tag{2.3.9}$$

where G is an $m \times m$ lower triangle matrix, $S \in \text{Sym}^+(m)$, and $Y \in \text{Sym}(m)$. Let $\text{vecs}(C) = (c_{11}, c_{21}, c_{22}, \cdots, c_{m1}, \cdots, c_{mm})^T$ for any $m \times m$ matrix $C = (c_{ij})$ and $\text{vec}(A) = (a_{11}, ..., a_{1m}, a_{21}, ..., a_{2m}, \cdots, a_{m1}, \cdots, a_{mm})^T$ be the vectorization of an $m \times m$ matrix $A = (a_{ij})$. Let $\alpha = (\alpha_G, \alpha_Y)$, in which $\alpha_G = \text{vecs}(G)$, and $\alpha_Y = \text{vecs}(Y)$. Let $\partial_\alpha \psi(S, G, Y)$ and $\partial^2_\alpha \psi(S, G, Y)$ be the first and second order derivatives of $\psi(S, G, Y)$ with respect to α, respectively. By substituting $Y(X)$ into $\partial_\alpha \psi(S, G, Y)$ and $\partial^2_\alpha \psi(S, G, Y)$ and using the decomposition of $\alpha = (\alpha_G, \alpha_Y)$, we define

$$\begin{pmatrix} \Psi_1(x) & \Psi_2(x) \\ \Psi_2(x)^T & \Psi_3(x) \end{pmatrix} = E\{\partial^2_\alpha \psi(S, G, Y(X))|X = x\},$$
$$\begin{pmatrix} \Psi_{11}(x) & \Psi_{12}(x) \\ \Psi_{12}(x)^T & \Psi_{22}(x) \end{pmatrix} = E[\{\partial_\alpha \psi(S, G, Y(X))\}^{\otimes 2}|X = x],$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector or matrix $\mathbf{a}$ and the expectation is taken with respect to S given $X = x$. Let $\mathbf{1}_{k_0 \times 1}$ be a $k_0 \times 1$ column vector with all elements ones. Finally, we define $\aleph(x_0; h) = (w_1(x_0; h)^T \Psi_2(x_0), \mathbf{w}(x_0; h)\{\mathbf{1}_{k_0 \times 1} \otimes \Psi_3(x_0)\})^T$ and $\mathbf{w}(x_0; h) = (w_2(x_0; h)^T, \cdots, w_{k_0+1}(x_0; h)^T)$, in which for $0 < k \leq k_0 + 1$,

$$w_k(x_0; h) = \begin{cases} u_{k_0+k} \text{vecs}(Y^{(k_0+1)}(x_0)) & \text{if } k_0 + k \text{ is even;} \\ h u_{k_0+k+1} \text{vecs}(Y^{(k_0+1)}(x_0) f_X^{(1)}(x_0)/f_X(x_0) + Y^{(k_0+2)}(x_0)/(k_0 + 2)) & \text{if } k_0 + k \text{ is odd.} \end{cases}$$

We have the following results, whose proof can be found in Appendix C.

Theorem 2.3.1. *Suppose that x_0 is an interior point of $f_X(\cdot)$. Let $\mathcal{H} = diag(1, h, \cdots, h^{k_0}) \otimes I_q$ and $\alpha_A(x) = (vecs(G(x))^T, vecs(Y^{(1)}(x))^T, \cdots, vecs(Y^{(k_0)}(x)/k_0!)^T)^T$, in which I_q is an identity matrix of size $q = m(m+1)/2$.*

(i) Under (2.3.6) and conditions (C1)-(C8) in Appendix C, there exist solutions $\hat{\alpha}_{IA}(x_0; h)$ to equation $\partial G_n(\alpha_A(x_0))/\partial \alpha_A(x_0) = \mathbf{0}$ such that $\mathcal{H}\{\hat{\alpha}_{IA}(x_0; h) - \alpha_A(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

(ii) For $k_0 = 0$, under (2.3.6), conditions (C1)-(C8) in Appendix C and that $f_X^{(1)}(x)$ is continuous

in a neighborhood of x_0*, conditioning on* $\mathbf{x} = \{x_1, \cdots, x_n\}$*, we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IA}(x_0; h) - \alpha_A(x_0)\} - h^2 u_2 vecs\{G^{(1)}(x_0)\frac{f_X^{(1)}(x_0)}{f_X(x_0)} + 0.5G^{(2)}(x_0)\}] \rightarrow^L N(\mathbf{0}, \Omega_0(x_0)) \quad (2.3.10)$$

where $\Omega_0(x_0) = u_0^{-2} f_X^{-1}(x_0) v_0 \Psi_1(x_0)^{-1} \Psi_{11}(x_0) \Psi_1(x_0)^{-1}$ *and* $\rightarrow^L$ *denotes convergence in distribution.*

(iii) For $k_0 > 0$*, under the conditions of Theorem 2.3.1 (ii) and condition (C9), conditioning on* $\mathbf{x}$*, we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IA}(x_0; h) - \alpha_A(x_0)\} - \frac{h^{k_0+1}}{(k_0+1)!}\mathcal{N}(x_0)^{-1}\aleph(x_0; h)] \rightarrow^L N(\mathbf{0}, \Omega(x_0)), \quad (2.3.11)$$

where $\Omega(x_0) = f_X^{-1}(x_0)\mathcal{N}(x_0)^{-1}\mathcal{N}^*(x_0)\mathcal{N}(x_0)^{-1}$ *and* $\mathcal{N}(x)$ *and* $\mathcal{N}^*(x)$ *are, respectively, given by*

$$\mathcal{N}(x) = \begin{pmatrix} u_0\Psi_1(x) & \mathbf{u} \otimes \Psi_2(x) \\ \mathbf{u}^T \otimes \Psi_2(x)^T & \mathcal{U}_2 \otimes \Psi_3(x) \end{pmatrix}, \mathcal{N}^*(x) = \begin{pmatrix} v_0\Psi_{11}(x) & \mathbf{v} \otimes \Psi_{12}(x) \\ \mathbf{v}^T \otimes \Psi_{12}(x)^T & \mathcal{V}_2 \otimes \Psi_{22}(x) \end{pmatrix}.$$

Theorem 2.3.1 delineates the asymptotic bias, covariance, and asymptotic normality of $\hat{\alpha}_{IA}(x_0; h)$ for $k_0 \geq 0$. Based on Theorem 2.3.1, it is straightforward to derive the asymptotic bias, covariance, and asymptotic normality of $\hat{D}_{IA}(x_0; h, k_0)$ for $k_0 \geq 0$. Moreover, to have a direct comparison between the affine invariant and Log-Euclidean metrics, we calculate the asymptotic biases and covariances of $\log(\hat{D}_{IA}(x_0; h, k_0)) \in \text{Sym}(m)$. Subsequently, we calculate the asymptotic mean squared error (AMSE) conditional on $\mathbf{x}$ as

$$\begin{aligned} &\text{AMSE}(\log(\hat{D}_{IA}(x_0; h, k_0))) = E(\text{tr}[\{\log(\hat{D}_{IA}(x_0; h, k_0)) - \log(D(x_0))\}^2]|\mathbf{x}) \\ &= \text{tr}\{\text{bias}(\text{vecs}(\log(\hat{D}_{IA}(x_0; h, k_0))|\mathbf{x})^{\otimes 2}\} + \text{tr}\{\text{Cov}(\text{vecs}(\log(\hat{D}_{IA}(x_0; h, k_0)))|\mathbf{x})\}. \end{aligned}$$

Furthermore, we may consider a constant bandwidth that minimizes the asymptotic mean integrated squared error (AMISE) as

$$\text{AMISE}(\log(\hat{D}_{IA}(.; h, k_0))) = \int \text{AMSE}(\log(\hat{D}_{IA}(x; h, k_0)))w(x)dx$$

for a given weight function $w(x)$.

We are interested in comparing the asymptotic properties of the intrinsic local constant $\hat{D}_{IA}(x_0; h, 0)$ and the local linear estimator $\hat{D}_{IA}(x_0; h, 1)$. For the intrinsic local constant estimator, it follows from the delta method that AMSE($\log(\hat{D}_{IA}(x_0; h, 0))$) can be approximated as

$$h^4 u_2^2 \mathrm{tr}([G_D(x_0)^T \mathrm{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2}) + \{nh\}^{-1} \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}, \tag{2.3.12}$$

where $G_D(x_0) = \{\partial \mathrm{vec}(\log(G(x_0)^{\otimes 2}))/\partial \mathrm{vecs}(G(x_0))^T\}^T$. The asymptotic bias and variance of $\hat{D}_{IA}(x_0; h, 0)$ are similar to those of the Nadaraya-Watson estimator when both response and covariate are in Euclidean space (Fan, 1992). Minimizing AMSE($\log(\hat{D}_{IA}(x_0; h, 0))$) leads to the asymptotically optimal local bandwidth which is given by

$$h_{opt,A}(x_0; 0) = \left[\frac{n^{-1} \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}}{4 u_2^2 \mathrm{tr}([G_D(x_0)^T \mathrm{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2})} \right]^{1/5}. \tag{2.3.13}$$

With some calculations, the optimal bandwidth for minimizing AMISE($\log(\hat{D}_{IA}(x_0; h, 0))$), denoted by $h_{opt,A}(0)$, is given by

$$h_{opt,A}(0) = \left[\frac{n^{-1} \int \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\} w(x) dx}{4 u_2^2 \int \mathrm{tr}([G_D(x_0)^T \mathrm{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2}) w(x) dx} \right]^{1/5}. \tag{2.3.14}$$

For the intrinsic local linear estimator, AMSE($\log(\hat{D}_{IA}(x_0; h, 1))$) is given by

$$0.25 h^4 u_2^2 \mathrm{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \mathrm{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}] + (nh)^{-1} \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}. \tag{2.3.15}$$

Minimization of AMSE($\log(\hat{D}_{IA}(x_0; h, 1))$) leads to the asymptotically optimal local bandwidth, denoted by $h_{opt,A}(x_0; 1)$, which is given by

$$h_{opt,A}(x_0; 1) = \left(\frac{n^{-1} \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}}{u_2^2 \mathrm{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \mathrm{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}]} \right)^{1/5}. \tag{2.3.16}$$

Furthermore, the optimal bandwidth for minimizing AMISE($\log(\hat{D}_{IA}(x_0;h,1))$) is

$$h_{opt,A}(1) = \left(\frac{n^{-1} \int \mathrm{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x)\} w(x) dx}{u_2^2 \int \mathrm{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \mathrm{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}] w(x) dx} \right)^{1/5}. \tag{2.3.17}$$

For interior points, both the intrinsic local constant and linear estimators have the same covariance, but different biases. Particularly, the intrinsic local constant estimator has two leading terms with one associated with the marginal density $f_X(.)$. This result is similar to the well-known results on the superiority of the local linear over the Nadaraya-Watson estimator when both response and covariate are in Euclidean space (see Fan and Gijbels (1996) for good discussion). However, for the intrinsic local constant estimator, its bias depends on the first and second order derivatives of $G(x)$, whereas the bias of the intrinsic local linear estimator is only associated with the second-order derivative of $Y(x)$.

We consider ILPRE near the edge of the support of $f_X(x)$. Without loss of generality, we assume that the design density $f_X(.)$ has a bounded support [0, 1] and consider the left-boundary point $x_0 = dh$ for some positive constant d. The asymptotic consistency and normality of ILPRE are valid for the boundary points after slight modifications on the definition of u_k and v_k. Denote $u_{k,d} = \int_{-d}^{\infty} x^k K(x) dx$ and $v_{k,d} = \int_{-d}^{\infty} x^k K^2(x) dx$. Correspondingly, $\mathbf{u}$, $\mathcal{U}_2$, $\mathcal{V}_2$, $\mathcal{U}_0$ and $\mathcal{V}_0$ are replaced by $\mathbf{u}_d$, $\mathcal{U}_{2,d}$, $\mathcal{V}_{2,d}$, $\mathcal{U}_{0,d}$ and $\mathcal{V}_{0,d}$, respectively. Let $\mathbf{c}_{k_0+2,d} = (u_{k_0+2,d}, \cdots, u_{2k_0+1,d})^T$ and $\aleph_d(0+) = (u_{k_0+1,d}\Psi_2(0+), \mathbf{c}_{k_0+2,d} \otimes \Psi_3(0+))^T \mathrm{vecs}(Y^{(k_0+1)}(0+))$. For the boundary points, we have the following asymptotic results under the affine invariant metric.

Theorem 2.3.2. *Suppose that $x_0 = dh$ is a left boundary point of $f_X(.)$.*
(i) Under conditions (C1)-(C8) in Appendix C, there exist solutions, denoted by $\hat{\alpha}_{IA}(x_0;h)$, to the equation $\partial G_n(\alpha_A(x_0))/\partial \alpha_A(x_0) = \mathbf{0}$ such that $\mathcal{H}\{\hat{\alpha}_{IA}(x_0,h) - \alpha_A(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.
(ii) For $k_0 = 0$ and under conditions (C1)-(C8) in Appendix C, conditioning on $\mathbf{x} = \{x_1, \cdots, x_n\}$, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IA}(0+;h) - \alpha_A(0+)\} - h u_{0,d}^{-1} u_{1,d} G^{(1)}(0+)] \to^L N(\mathbf{0}, \Omega_{0,d}(0+)) \tag{2.3.18}$$

where $\Omega_{0,d}(0+) = f_X^{-1}(0+) u_{0,d}^{-2} v_{0,d} \Psi_1(0+)^{-1} \Psi_{11}(0+) \Psi_1(0+)^{-1}$.

(iii) For $k_0 > 0$ and under the same conditions as in Theorem 2.3.2 (ii) and condition (C9), conditioning on $\mathbf{x} = \{x_1, \cdots, x_n\}$, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IA}(0+;h) - \alpha_A(0+)\} - \frac{h^{k_0+1}}{(k_0+1)!}\mathcal{N}_d(0+)^{-1}\aleph_d(0+)] \to^L N(\mathbf{0}, \Omega_d(0+)). \tag{2.3.19}$$

where $\Omega_d(0+) = f_X^{-1}(0+)\mathcal{N}_d(0+)^{-1}\mathcal{N}_d^*(0+)\mathcal{N}_d(0+)^{-1}$ and $\mathcal{N}_d(0+)$ and $\mathcal{N}_d^*(0+)$ are, respectively, given by

$$\mathcal{N}_d(0+) = \begin{pmatrix} u_{0,d}\Psi_1(0+) & \mathbf{u}_d \otimes \Psi_2(0+) \\ \mathbf{u}_d{}^T \otimes \Psi_2(0+)^T & \mathcal{U}_{2,d} \otimes \Psi_3(0+) \end{pmatrix} \quad \textit{and}$$

$$\mathcal{N}_d^*(0+) = \begin{pmatrix} v_{0,d}\Psi_{11}(0+) & \mathbf{v}_d{}^T \otimes \Psi_{12}(0+) \\ \mathbf{v}_d \otimes \Psi_{12}^T(0+) & \mathcal{V}_{2,d} \otimes \Psi_{22}(0+) \end{pmatrix}.$$

It follows from Theorem 2.3.2 (ii) and (iii) that when x_0 is at the boundary, the asymptotic average mean squared errors of intrinsic local constant and linear estimators are, respectively, $\text{AMSE}(\log(\hat{D}_{IA}(0+;h,0))) = O_p(h^2 + (nh)^{-1})$ and $\text{AMSE}(\log(\hat{D}_{IA}(0+;h,1))) = O_p(h^4 + (nh)^{-1})$. Thus, the rate of convergence for intrinsic local constant estimator at boundary points is slower than that at points in the interior. Thus, the intrinsic local constant estimator suffers from the well-known boundary effects. However, the intrinsic local linear estimator adapts automatically at the boundary points and its rate of convergence is not influenced by the location of points. Thus, the intrinsic local linear (or polynomial) estimators share the same property of automatic adaptation to the boundary as local polynomial estimators in Euclidean space (Fan, 1992).

2.4 Log-Euclidean Metric

2.4.1 ILPR under Log-Euclidean Metric

We introduce an 'L' into all geometric quantities under the Log-Euclidean metric (see Arsigny (2006) for details of the geometry of the space of SPD matrices under this metric). Let $\partial_{D(x)}\log(.)$ and $\partial_{\log(D(x))}\exp(.)$ be, respectively, the directional derivatives in the direction of $D(x) \in \text{Sym}^+(m)$ of the log map and in the direction of $\log(D(x)) \in \text{Sym}(m)$ of the exp map. The map $\partial_{D(x)}\log(.)$ is a linear mapping from $T_{D(x)}\text{Sym}^+(m)$ to $T_I\text{Sym}^+(m)$. A scaled Frobenius inner product on

$T_{D(x)}\mathrm{Sym}^+(m)$ based on the Log-Euclidean metric is defined as

$$<< U_{D(x)}, V_{D(x)} >>_{D(x),L} =< \partial_{D(x)} \log(U_{D(x)}), \partial_{D(x)} \log(V_{D(x)}) >, \tag{2.4.1}$$

where $U_{D(x)}$ and $V_{D(x)}$ are in $T_{D(x)}Sym^+(m)$. The geodesic $\gamma_{D(x),L}(t, U_{D(x)})$ is given by $\exp(\log(D(x)) + t\partial_{D(x)} \log(U_{D(x)}))$ for any t. Under the Log-Euclidean metric, the Riemannian exponential and logarithm maps are, respectively, given by

$$\begin{aligned} &\mathrm{Exp}_{D(x),L}(U_{D(x)}) = \exp(\log(D(x)) + \partial_{D(x)} \log(U_{D(x)})), \\ &\mathrm{Log}_{D(x),L}(S) = \partial_{\log(D(x))} \exp(\log(S) - \log(D(x))). \end{aligned} \tag{2.4.2}$$

The geodesic distance between $D(x)$ and S is uniquely given by

$$d_L(D(x), S) = \sqrt{tr[\{\log(D(x)) - \log(S)\}^{\otimes 2}]}. \tag{2.4.3}$$

We consider two SPD matrices $D(x)$ and $D(x_0)$. For any $U_{D(x_0)} \in T_{D(x_0)}Sym^+(m)$, the rotation map $\phi_{D(x_0),L} : T_{D(x_0)}Sym^+(m) \to T_{I_m}\mathrm{Sym}^+(m)$ is defined by

$$\phi_{D(x_0),L}(U_{D(x_0)}) = \partial_{D(x_0)} \log(U_{D(x_0)}) \in T_{I_m}\mathrm{Sym}^+(m). \tag{2.4.4}$$

Recall that $\mathrm{Log}_{D(x_0),L}(D(x))$ is in $T_{D(x_0)}\mathrm{Sym}^+(m)$. Combining (2.4.2) and (2.4.4) yields that

$$\begin{aligned} &Y(x) = \phi_{D(x_0),L}(\mathrm{Log}_{D(x_0)}(D(x))) = \log(D(x)) - \log(D(x_0)), \\ &D(x) = \exp(\log(D(x_0)) + Y(x)). \end{aligned} \tag{2.4.5}$$

In this case, $\mathcal{E}_D(X)$ can be defined to be $\log(S) - \log(D(X))$ such that

$$E\{\log(S) - \log(D(X))|X = x\} = O_m. \tag{2.4.6}$$

To compute the ILPR estimator, we use the Taylor's series expansion to expand $\log(D(x_0))$ at

x_0 as follows:

$$D(x) \approx \exp(\sum_{k=0}^{k_0} (\log(D(x_0)))^{(k)}(x_0)(x - x_0)^k/k!) = D_L(x, \alpha_L(x_0), k_0), \tag{2.4.7}$$

where $\alpha_L(x_0) = (\text{vecs}(\log(D(x_0)))(x_0))^T, \cdots, \text{vecs}(\log(D(x_0)))^{(k_0)}(x_0))^T)^T$. Thus, we can compute $\hat{\alpha}_{IL}(x_0; h)$ by minimizing $G_n(\alpha(x_0))$. For the Log-Euclidean metric, $\hat{\alpha}_{IL}(x_0; h)$ has the explicit expression as

$$\hat{\alpha}_{IL}(x_0; h) = \text{vec}(\{\sum_{i=1}^{n} K_h(x_i - x_0)\mathcal{X}_i(x_0)^{\otimes 2}\}^{-1} \sum_{i=1}^{n} K_h(x_i - x_0)\mathcal{X}_i(x_0)\text{vecs}(\log(S_i))^T),$$

where $\mathcal{X}_i(x) = (1, (x_i - x), \cdots, (x_i - x)^{k_0})^T$. Furthermore, substituting $\hat{\alpha}_{IL}(x_0; h)$ into $D_L(x, \alpha_L(x_0), k_0)$, we have $\hat{D}_{IL}(x; h, k_0) = D_L(x, \hat{\alpha}_{IL}(x_0; h), k_0)$.

Let $e_{k_0+1,i}$ be the $(k_0 + 1)$ unit vector having 1 in the i-th entry and 0 elsewhere. Let $e_{k_0+1,i}^T\{\sum_{j=1}^{n} K_h(x_j - x)\mathcal{X}_j(x)^{\otimes 2}\}^{-1} K_h(x_i - x)\mathcal{X}_i(x) = a_i(x)$. The cross-validation score $\text{CV}(h)$ can be simplified as follows ,

$$\text{CV}(h) = n^{-1} \sum_{i=1}^{n} g_L(S_i, \hat{D}_{IL}(x_i; h))^2/\{1 - a_i(x_i)\}^2. \tag{2.4.8}$$

Replacing $a_i(x_i)$ in equation (2.4.8) by the average of $a_1(x_1), \cdots, a_n(x_n)$, we can get the following generalized cross-validation (GCV) score

$$\text{GCV}(h) = n^{-1} \sum_{i=1}^{n} g_L(S_i, \hat{D}_{IL}(x_i; h))^2/\{1 - \sum_{i=1}^{n} a_i(x_i)/n\}^2. \tag{2.4.9}$$

Without special saying, for the Log-Euclidean metric, we use generalized cross-validation score $\text{GCV}(h)$ to select the bandwidth throughout this paper.

2.4.2 Asymptotic Properties

Under the Log-Euclidean metric, we examine the consistency and asymptotic normality of ILPRE for both of the interior and boundary points. We need some additional notation. Let $\mathcal{U}_0 = (u_{i+j})$ and $\mathcal{V}_0 = (v_{i+j})$ for $0 \le i, j \le k_0$ be two $(k_0 + 1) \times (k_0 + 1)$ matrices. Let $\Sigma_{\mathcal{E}_D}(x)$ be $\text{Cov}(\text{vecs}(\log(S) - \log(D(x)))|X = x)$. We define $\mathcal{M}(x_0; h) = (\mathcal{M}_1(x_0; h)^T, \cdots, \mathcal{M}_{k_0+1}(x_0; h)^T)^T$,

in which for $0 < k \leq k_0 + 1$,

$$\mathcal{M}_k(x_0; h) = \begin{cases} u_{k_0+k}\text{vecs}((\log(D(x_0)))^{(k_0+1)}) & \text{if } k_0 + k \text{ is even;} \\ hu_{k_0+k+1}\text{vecs}((\log(D(x_0)))^{(k_0+1)} f_X^{(1)}(x_0)/f_X(x_0) & \text{if } k_0 + k \text{ is odd.} \\ +(\log(D(x_0)))^{(k_0+2)}/(k_0+2)) & \end{cases}$$

Theorem 2.4.1. *Suppose that x_0 is an interior point of $f_X(.)$.*

(i) Under conditions (C1)-(C4) in Appendix C, we have $\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\}$ converges to $\mathbf{0}$ in probability as $n \to \infty$.

(ii) For $k_0 = 0$, under (2.4.6) and conditions (C1)-(C4) and (C10) in Appendix C and that $f_X^{(1)}(x)$ is continuous in a neighborhood of x_0, conditioning on $\mathbf{x} = \{x_1, \cdots, x_n\}$, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\} - h^2 u_2 vecs(0.5(\log(D(x_0)))^{(2)} + \frac{f_X^{(1)}(x_0)}{f_X(x_0)}(\log D(x_0))^{(1)})] \\ \to^L N(\mathbf{0}, \Sigma_0(x_0)), \quad (2.4.10)$$

where $\Sigma_0(x_0) = f_X^{-1}(x_0) v_0 \Sigma_{\mathcal{E}_D}(x_0)$.

(iii) For $k_0 > 0$, under the conditions of Theorem 2.4.1 (ii), conditioning on $\mathbf{x}$, we have

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\} - \frac{h^{k_0+1}}{(k_0+1)!}(\mathcal{U}_0^{-1} \otimes I_q)\mathcal{M}(x_0; h)] \to^L N(\mathbf{0}, \Sigma(x_0)), \quad (2.4.11)$$

where $\Sigma(x_0) = f_X^{-1}(x_0)(\mathcal{U}_0^{-1}\mathcal{V}_0\mathcal{U}_0^{-1}) \otimes \Sigma_{\mathcal{E}_D}(x_0)$.

Theorem 2.4.1 delineates the asymptotic properties of $\hat{\alpha}_{IL}(x_0; h, k_0)$. Since $\text{vecs}(\log(\hat{D}_{IL}(x_0; h, k_0)))$ is a subvector of $\hat{\alpha}_{IL}(x_0; h, k_0)$, Theorem 3 covers the asymptotic properties of the intrinsic local constant and linear estimators of $D(x_0)$ for $k_0 = 0, 1$. In particular, the asymptotic bias and variance of $\hat{D}_{IL}(x_0; h, 0)$ are closely related to those of the Nadaraya-Watson estimator when both response and covariate are in Euclidean space (Fan, 1992). For the intrinsic local constant estimator, by Theorem 2.4.1 (iii), $\text{AMSE}(\log(\hat{D}_{IL}(x_0; h, 0)))$ equals

$$h^4 u_2^2 \text{tr}([\text{vecs}\{0.5(\log(D(x_0)))^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}(\log(D(x_0)))^{(1)}\}]^{\otimes 2}) \\ + v_0(nhf_X(x_0))^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}. \quad (2.4.12)$$

Minimizing AMSE($\log(\hat{D}_{IL}(x_0;h,0))$) leads to the asymptotically optimal local bandwidth which is given by

$$h_{opt,L}(x_0;0)=\left[\frac{v_0\{nf_X(x_0)\}^{-1}\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}{4u_2^2\mathrm{tr}([\mathrm{vecs}\{0.5(\log(D(x_0)))^{(2)}+f_X^{(1)}(x_0)f_X(x_0)^{-1}(\log(D(x_0)))^{(1)}\}]^{\otimes 2})}\right]^{1/5}. \tag{2.4.13}$$

With some calculations, the optimal bandwidth for minimizing AMISE($\log(\hat{D}_{IL}(x_0;h,0))$), denoted by $h_{opt,L}(0)$, is given by

$$\left[\frac{v_0\{nf_X(x_0)\}^{-1}\int\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}w(x)dx}{4u_2^2\int\mathrm{tr}([\mathrm{vecs}\{0.5(\log(D(x_0)))^{(2)}+f_X^{(1)}(x_0)f_X(x_0)^{-1}(\log(D(x_0)))^{(1)}\}]^{\otimes 2})w(x)dx}\right]^{1/5}. \tag{2.4.14}$$

For the intrinsic local linear estimator, AMSE($\log(\hat{D}_{IL}(x_0;h,1))$) is given by

$$0.25h^4u_2^2\mathrm{tr}[\{\mathrm{vecs}(\log(D(x_0))^{(2)})\}^{\otimes 2}]+v_0\{nhf_X(x_0)\}^{-1}\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}. \tag{2.4.15}$$

Intrinsic local constant and linear estimators have the same covariance, their differences are concerned only with their biases. The local constant estimator has one more term $h^2u_2f_X^{(1)}(x_0)f_X(x_0)^{-1}$ $\mathrm{vecs}(\log(D(x_0))^{(1)})$, which depends on the marginal density $f_X(.)$. Minimization of AMSE($\log(\hat{D}_{IL}(x_0;h,1))$) leads to the asymptotically optimal local bandwidth, denoted by $h_{opt,L}(x_0;1)$, which is given by

$$h_{opt,L}(x_0;1)=\left[\frac{\{nf_X(x_0)\}^{-1}v_0\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}{u_2^2\mathrm{tr}[\{\mathrm{vecs}(\log(D(x_0))^{(2)})\}^{\otimes 2}]}\right]^{1/5}. \tag{2.4.16}$$

Furthermore, the optimal bandwidth for minimizing AMISE($\log(\hat{D}_{IL}(x_0;h,0))$) is

$$h_{opt,L}(1)=\left[\frac{n^{-1}v_0\int\mathrm{tr}\{\Sigma_{\mathcal{E}_D}(x)\}\{f_X(x)\}^{-1}w(x)dx}{u_2^2\int\mathrm{tr}[\{\mathrm{vecs}(\log(D(x))^{(2)})\}^{\otimes 2}]w(x)dx}\right]^{1/5}. \tag{2.4.17}$$

We present the asymptotic properties of the intrinsic estimators under the Log-Euclidean metric at boundary points below.

Theorem 2.4.2. *Suppose that $x_0=dh$ is a left boundary point of $f_X(.)$. Let $d_{k_0,d}=(u_{k_0+1,d},\cdots,u_{2k_0+1,d})^T$.*

(i)Under conditions (C1)-(C4) in Appendix C, we have $\mathcal{H}\{\hat{\alpha}_{IL}(x_0;h)-\alpha(x_0)\}$ converges to $\mathbf{0}$ in

probability as $n \to \infty$.

(ii) For $k_0 = 0$, *under conditions (C1)-(C4) and (C10) in Appendix C, conditioning on* $\mathbf{x} = \{x_1, \cdots, x_n\}$, *we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha(x_0)\} - hu_{0,d}^{-1}u_{1,d}vecs((\log(D(0+)))^{(1)})] \to^L N(\mathbf{0}, \Sigma_{0,d}(0+)), \tag{2.4.18}$$

where $\Sigma_{0,d}(0+) = f_X^{-1}(0+)u_{0,d}^{-2}v_{0,d}\Sigma_{\mathcal{E}_D}(0+)$.

(iii) For $k_0 > 0$, *under the conditions of Theorem 2.4.2 (ii), conditioning on* $\mathbf{x} = \{x_1, \cdots, x_n\}$, *we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha(x_0)\} - \frac{h^{k_0+1}}{(k_0+1)!}(\mathcal{U}_{0,d}^{-1} \otimes I_q)(d_{k_0,d} \otimes vecs((\log(D(0+)))^{(k_0+1)}))] \to^L N(\mathbf{0}, \Sigma_d(0+)), \tag{2.4.19}$$

where $\Sigma_d(0+) = f_X^{-1}(0+)(\mathcal{U}_{0,d}^{-1}\mathcal{V}_{0,d}\mathcal{U}_{0,d}^{-1}) \otimes \Sigma_{\mathcal{E}_D}(0+)$.

2.4.3 Comparisons

We study the asymptotic relative efficiency of the intrinsic local constant and linear estimators in an interior point x_0 under the affine invariant and Log-Euclidean metrics by comparing their AMSEs/AMISEs. We use AMSE_{opt} and AMISE_{opt} to denote the AMSE and AMISE evaluated at their optimal bandwidth. We first compare AMSE_{opt} for the intrinsic local constant estimators under the affine invariant and Log-Euclidean metrics. Specifically, with some calculations, we see that as n approaches ∞, the ratio of $\text{AMSE}_{opt}(\log(\hat{D}_{IA}(x_0; h, 0)))$ over $\text{AMSE}_{opt}(\log(\hat{D}_{IL}(x_0; h, 0)))$ converges to

$$\begin{aligned}\text{rMSE}(A, L; 0) = &\left[\frac{\text{tr}\{G_D(x_0)^{\otimes 2}\Psi_1(x_0)^{-1}\Psi_{11}(x_0)\Psi_1(x_0)^{-1}\}}{\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}\right]^{4/5} \\ &\times \left\{\frac{\text{tr}([G_D(x_0)^T\text{vecs}\{f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0) + 0.5G^{(2)}(x_0)\}]^{\otimes 2})}{\text{tr}[\text{vecs}\{f_X^{(1)}(x_0)f_X(x_0)^{-1}\log(D(x_0))^{(1)} + 0.5\log(D(x_0))^{(2)}\}^{\otimes 2}]}\right\}^{1/5},\end{aligned} \tag{2.4.20}$$

which is the product of two terms. The first term is associated with the ratio of the covariances of the intrinsic local constant estimators of $\log(D(x_0))$ under the two metrics, while the second term is associated with their biases. Consider the simplest scenario with $m = 1$ such that $D(x_0) = G(x_0)^2$

and $G(x_0) > 0$. By simple calculations, we can show that the first term equals one and the second term equals

$$\left\{ \frac{f_X^{(1)}(x_0) f_X(x_0)^{-1} G^{(1)}(x_0) + 0.5 G^{(2)}(x_0)}{f_X^{(1)}(x_0) f_X(x_0)^{-1} G^{(1)}(x_0) + 0.5 G^{(2)}(x_0) - 0.5 (G^{(1)}(x_0))^2 / G(x_0)} \right\}^{2/5},$$

which yields that $\text{rMSE}(A, L; 0) > 1$ if and only if

$$0.25 (G^{(1)}(x_0))^2 / G(x_0) < f_X^{(1)}(x_0) f_X(x_0)^{-1} G^{(1)}(x_0) + 0.5 G^{(2)}(x_0).$$

Thus, whether $\text{rMSE}(A, L; 0)$ is greater than 1 depends on both the design density and $D(x)$ itself as a function of x.

Similarly, we define $\text{rMISE}(A, L; 0)$ as the ratio of $\text{AMISE}_{opt}(\log(\hat{D}_{IA}(x_0; h, 0)))$ over $\text{AMISE}_{opt}(\log(\hat{D}_{IL}(x_0; h, 0)))$. Following similar arguments to $\text{rMSE}(A, L; 0)$, when $m = 1$, we have that $\text{rMISE}(A, L; 0) > 1$ if and only if

$$0.25 \int \{G^{(1)}(x)\}^2 G(x)^{-1} w(x) dx < \int \{f_X^{(1)}(x) f_X(x)^{-1} G^{(1)}(x) + 0.5 G^{(2)}(x)\} w(x) dx.$$

Therefore, in terms of AMSE_{opt} and AMISE_{opt}, the affine invariant metric is not uniformly superior to or worse than the Log-Euclidean metric for reconstructing all $D(x)$.

We compare AMSE_{opt} for the intrinsic local linear estimators under the affine invariant and Log-Euclidean metrics. As n approaches ∞, the ratio of $\text{AMSE}_{opt}(\log(\hat{D}_{IA}(x_0; h, 1)))$ over $\text{AMSE}_{opt}(\log(\hat{D}_{IL}(x_0; h, 1)))$ converges to

$$\begin{aligned} \text{rMSE}(A, L; 1) &= \left[\frac{\text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}}{\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}} \right]^{4/5} \\ &\quad \times \left\{ \frac{\text{tr}(\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \text{vecs}(Y^{(2)}(x_0))\}^{\otimes 2})}{\text{tr}[\text{vecs}\{\log(D(x_0))^{(2)}\}^{\otimes 2}]} \right\}^{1/5}. \end{aligned} \quad (2.4.21)$$

We also consider the simplest scenario with $m = 1$ such that $D(x) = G(x_0)^2 \exp(Y(x))$ and $G(x_0) > 0$. With some calculations, we can show that $\text{rMSE}(A, L; 1)$ equals one when $m = 1$. Thus, the two metrics are actually the same for one dimensional case. However, when $m > 1$, it is unclear whether $\text{rMSE}(A, L; 1)$ equals to 1 or not.

2.5 Simulation

We conducted three sets of Monte Carlo simulations to examine the finite sample performance of ILPREs for SPD matrices. The first set of simulations was to compare our intrinsic local constant and linear estimators under the affine invariant and Log-Euclidean metrics with other smoothing methods including local constant and linear estimators under the Euclidean metric and tensor spline estimator in Barmpoutis et al. (2007). The second set was to compare the performance of the intrinsic local linear estimators under the two metrics: the affine invariant and Log-Euclidean metrics. The third set was to evaluate the importance and effect of directly smoothing SPDs on the SPD-derived scalar measures. We set $m = 3$ and assume that the underlying SPD matrix function has the form: $D(x) = \exp(M(x))$, where $M(x) = \begin{pmatrix} -0.1(x+0.1) & 0.2(x+0.1) & \sin(0.75x) \\ 0.2(x+0.1) & 0.6(x+0.1) & -0.4(x+0.1) \\ \sin(0.75x) & -0.4(x+0.1) & 0.5(x+0.1) \end{pmatrix}$. We generated n design points $x_i, i = 1, \cdots, n$ independently from a $N(0, 0.25)$ distribution. In the following we generated the simulated data using three different noise models:

1. Riemannian log normal noise model: it assume that $C(x_i)^{-1} \log_{D(x_i)}(S_i) C(x_i)^{-T}$ has a symmetric matrix variate normal distribution $N(0, \Sigma)$ (see Schwartzman (2006) for details), where $D(x_i) = C(x_i)^{\otimes 2}$ and S_i is the simulated SPD matrix. Thus we first simulated $\epsilon_i \in \text{Sym}(3)$ from $N(0, \Sigma)$ and then calculated $S_i = C(x_i) \exp(\epsilon_i) C(x_i)^T$. We chose covariance matrice Σ as follows:

$$\Sigma_1 = 2 \begin{pmatrix} 0.3 & 0.049 & 0.052 \\ 0.049 & 0.2 & 0.0424 \\ 0.052 & 0.0424 & 0.1 \end{pmatrix}$$

and $\Sigma_2 = 4\Sigma_1$. The simulated SPD matrix data is shown in Figure 2.1 (a).

2. Log normal noise model: it assume that $\log(S_i)$ has a symmetric matrix variate normal distribution $N(0, \Sigma)$ (see Schwartzman (2006)). Thus we simulated ϵ_i from $N(0, \Sigma)$ and then calculated $S_i = \exp(\log(D(x_i)) + \epsilon_i)$. The simulated SPD matrix data is shown in Figure 2.1 (b).

3. Rician noise model: this noise model is commonly assumed in DT-MRI studies. The diffusion-weighted MR signal was simulated for 31 gradient directions $r_k, k = 1, \cdots, 31$ with b-factor $b_k = 1000 s/mm$ representing the magnitude of the diffusion gradients and four baseline images are

considered. We set the baseline signal intensity w_0 at 1500 and $\sigma_0 = 75$ to obtain a signal-to-noise ratio 25. For a given diffusion tensor $D(x_i)$, we add complex Gaussian noise having a standard deviation σ_0 to the simulated real signal $w_0 \exp(-b_k r_k D(x_i) r_k)$ at the k-th acquisition. i.e. we calculated

$$w_{i,k} = \sqrt{(w_0 \exp(-b_k r_k D(x_i) r_k) + \epsilon_{1,k})^2 + \epsilon_{2,k}^2}$$

where $\epsilon_{1,k}$ and $\epsilon_{2,k}$ were generated from a Gaussian random generator with mean zero and standard deviation σ_0 as the resulting diffusion weighted MR signal at the k-th acquisition. Then the diffusion tensor S_i was estimated from the 35 diffusion-weighted images by applying a least squares technique to the log-transformed signal intensities. The simulated DT (SPD matrix) data is shown in Figure 2.1 (c).

We set $n = 50$ and then simulated 100 data sets for each case. For each simulated dataset $\{(x_i, S_i) : i = 1, \cdots, n\}$, we apply our methods and also the local constant and linear methods under the Euclidean metric to estimate the SPD matrix function and the bandwidth was chosen separately for each estimator using its corresponding cross-validation method.

Figure 2.2 displays the estimated SPD using local linear regression under the affine invariant, Log-Euclidean and Euclidean metrics for the three different noise models. Note that three smoothing methods did an excellent job of recovering ground truth for the Rician noise model. For the other two noise models, our intrinsic local linear regression methods outperform that using the Euclidean metric. There is a clear swelling effect when using Euclidean smoothing method.

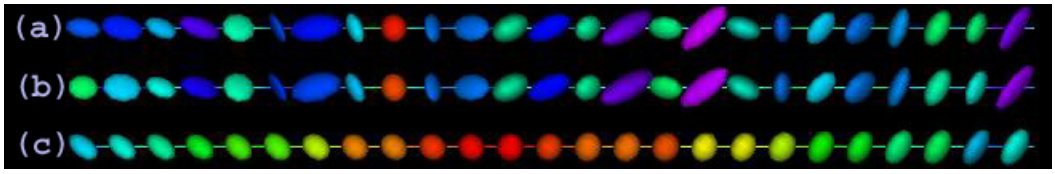

Figure 2.1: Ellipsoidal representations of the simulated SPD matrix data along the design points under the three different noise models: (a) Riemannian log normal, (b) log normal and (c) Rician noise models, colored with FA values defined in (2.5.1).

2.5.1 Simulation 1

We compared our intrinsic local constant and linear estimators for SPD matrices under both the affine invariant and Log-Euclidean metrics with local constant and linear estimators under the

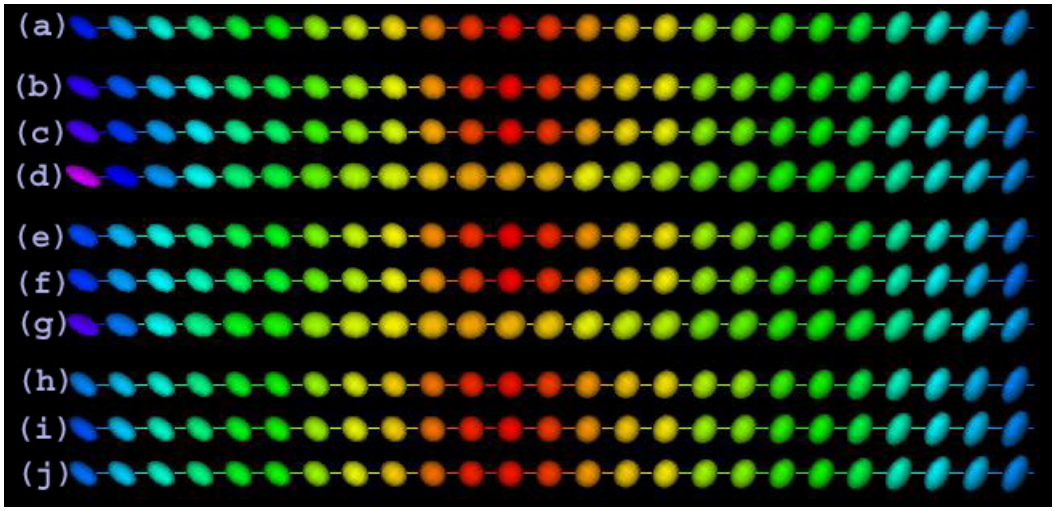

Figure 2.2: Ellipsoidal representations of the true (a) and estimated SPD matrix data along the design points using three smoothing methods: ((b), (e) and (h)) ILPR under the affine invariant metric, ((c), (f) and (i)) ILPR under the Log-Euclidean metric and ((d), (g) and (j)) LPR under the Euclidean metric and under the three different noise models: (b)-(d) Riemannian log normal, (e)-(g) log normal and (h)-(j) Rician noise models, colored with FA values defined in (2.5.1).

Euclidean metric and tensor spline method in Barmpoutis et al. (2007), respectively. For each simulated dataset, an Average (over the design points, x_i, for $i = 1, \cdots, n$) Geodesic Distance (AGD) across all pairs of the estimated and true SPD matrices $\hat{D}(x_i)$ and $D(x_i)$ was calculated as

$$\text{AGD}_j = n^{-1} \sum_{i=1}^{n} g(10^3 \hat{D}_j(x_i), 10^3 D(x_i)).$$

At each sample point x_i, a Local Average (over simulated realizations, $j = 1, \cdots, N$) Geodesic Distance (LAGD) was calculated as

$$\text{LAGD}(x_i) = N^{-1} \sum_{j=1}^{N} g(10^3 \hat{D}_j(x_i), 10^3 D(x_i)),$$

where $\hat{D}_j(x_i)$ are the estimated SPD matrices at the sample point x_i at the j-th replication. We take three different metrics for calculating the geodesic distance: affine invariant, Log-Euclidean and Euclidean metrics. Results for the three different noise models are summarized in Figure 2.3 and Figure 2.4.

First, we observe that the comparison measurements based on three different metrics reveal

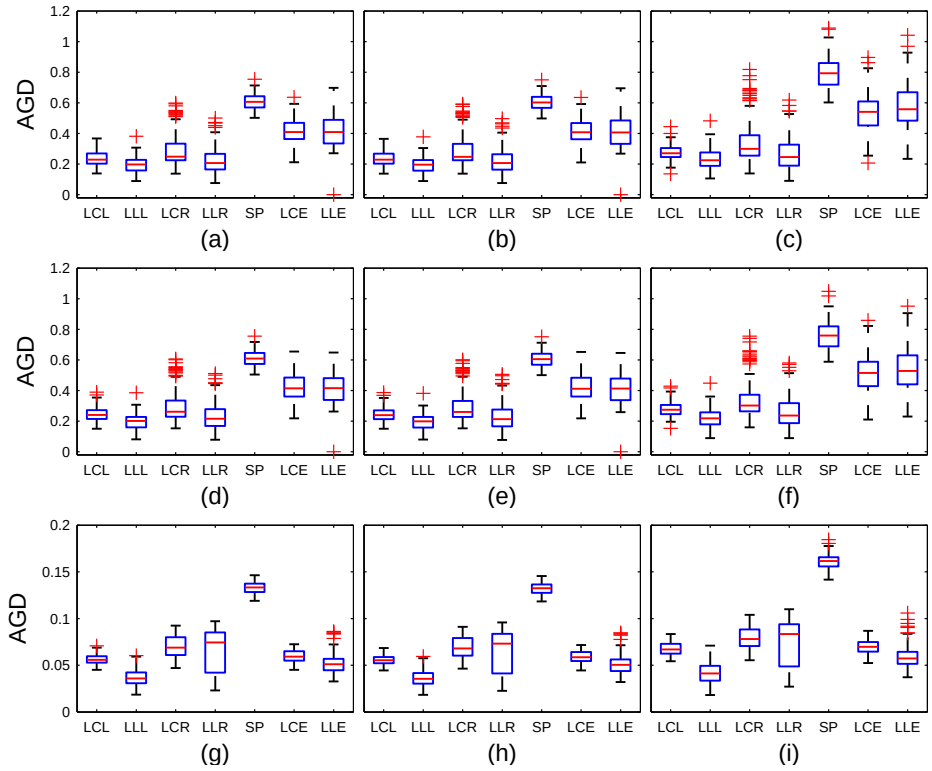

Figure 2.3: Boxplots of the AGD using the intrinsic local constant and linear estimators under the Log-Euclidean, affine invariant and Euclidean metrics, and tensor spline smoothing method, based on 100 replications for the three noise models: (a)-(c) Riemannian log normal, (d)-(f) log normal, and (g)-(i) Rician at sample size 50. The first, second and third columns correspond to comparisons based on geodesic distances under the affine invariant, Log-Euclidean and Euclidean metrics. LCL, LLL, LCR, LLR, LCE, LLE and SP represent the intrinsic local constant and linear estimators under the Log-Euclidean, affine invariant and Euclidean metrics and tensor spline smoothing method, respectively.

similar results. In addition, as expected, under each of three metrics considered, the local linear estimator is superior to the local constant estimator. Also, our intrinsic estimators perform better than the local constant and linear estimators under the Euclidean metric and the tensor spline method for the first two noise models. For the third noise model, our intrinsic local estimator under the Log-Euclidean metric works best. The second one is the estimator under the Euclidean metric. The third one is that under the affine invariant metric. It appears that the tensor spline method cannot compete with the other six methods. Moreover, it is observed that the variances of the AGD's based on the affine invariant metric are larger than those based on the Log-Euclidean metric. This suggests that the Log-Euclidean metric may be more useful. More can be observed from Figure 2.4, which displays the LAGD's at each sample point . We observed that, at interior

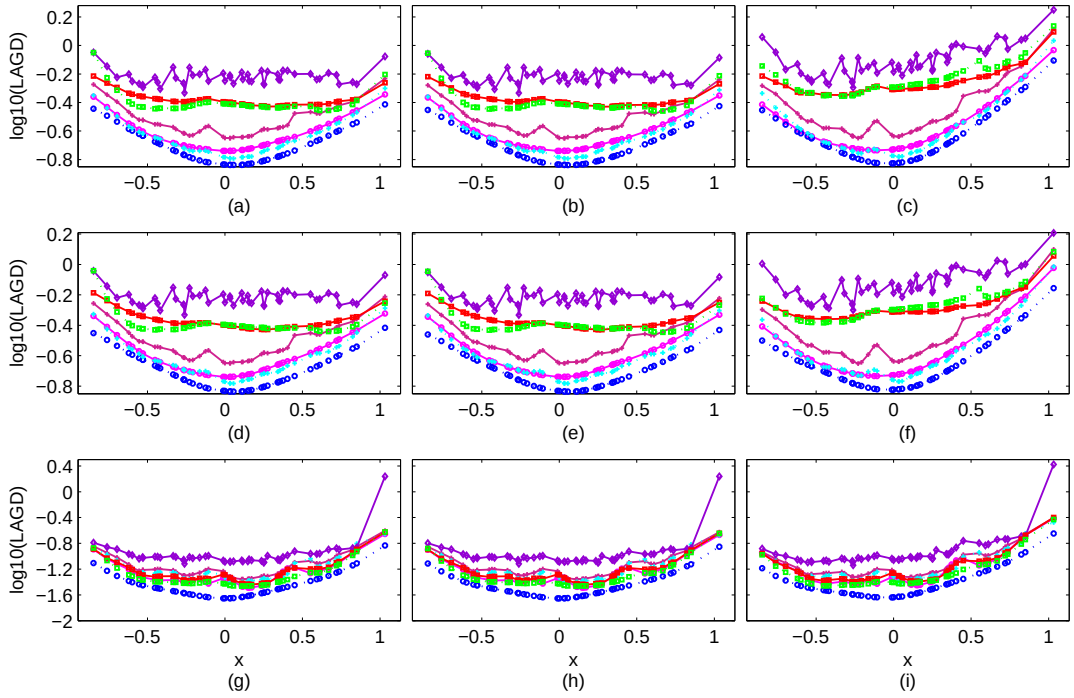

Figure 2.4: The $\log_{10}$(LAGD) curves at each sample point using the intrinsic local constant (solid line) and linear (dotted line) estimators under the Log-Euclidean(lines with circles), affine invariant (lines with +'s) and Euclidean metrics (lines with squares), and tensor spline smoothing method (lines with diamonds), based on 100 replications for the three noise models: (a)-(c) Riemannian log normal, (d)-(f) log normal, and (g)-(i) Rician at sample sizes 50. The first, second and third columns correspond to comparisons based on geodesic distances under affine invariant, Log-Euclidean and Euclidean metrics.

points, the differences are generally of smaller magnitude than those near the boundaries. In addition, note that the LAGD curves using the intrinsic local method under the affine invariant metric are rather rough. This is because the AIC-like bandwidth selection method selects a small bandwidth in order to adapt to the poor performances of the estimators at the edge.

2.5.2 Simulation 2

The second set of simulation studies, which was the same as the first, except for the relatively high noise level, Σ_2, compared the finite sample performance of the intrinsic local linear estimators under the affine invariant and Log-Euclidean metrics. Figure 2.3 did not reveal the superiority of the affine invariant metric over the Log-Euclidean metric. Figure 2.5 and 2.6 suggests that, when the noise level is high, the intrinsic local linear estimator under the affine invariant metric outperforms that under the Log-Euclidean metric. The results are similar by using comparison

measurements under three different metrics.

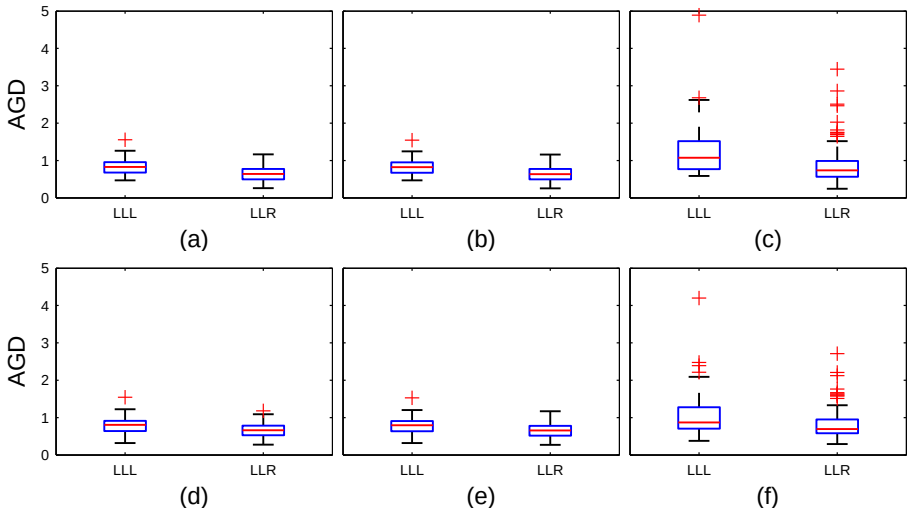

Figure 2.5: Boxplots of the AGD using the intrinsic local linear estimators under the Log-Euclidean, Riemannian and metrics, based on 100 replications for the first two noise models: (a)-(c) Riemannian log normal and (d)-(f) log normal at sample size 50. The first, second and third columns correspond to comparisons based on geodesic distances under the affine invariant, Log-Euclidean and Euclidean metrics. LLL and LLR represent the intrinsic local linear estimators under the Log-Euclidean and affine invariant metrics, respectively.

2.5.3 Simulation 3

The third set of simulation studies was to examine the importance and effect of directly smoothing SPDs on the more commonly considered SPD-derived scalar summary measures. We considered the usual scalar measure derived from a 3×3 SPD matrix, called fractional anisotropy (FA), which is a measure of the dominance of the largest eigenvalue of the SPD matrix. FA is a scalar value between zero (all eigenvalues are the same) and one (two eigenvlues equal 0) used in DTI, defined as

$$FA = \sqrt{\frac{3\{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2\}}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{2.5.1}$$

with eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and their average $\bar{\lambda}$, that describes the variation of the three eigenvalues of a 3×3 SPD matrix. We compared two different methods for smoothing FA's. The first smoothing method was to first calculate the FA's from all SPD matrices and then use the classic local linear regression in Euclidean space to smooth the FA's. The second and third smoothing methods were to first apply the intrinsic local linear estimator under the Log-Euclidean and affine invariant

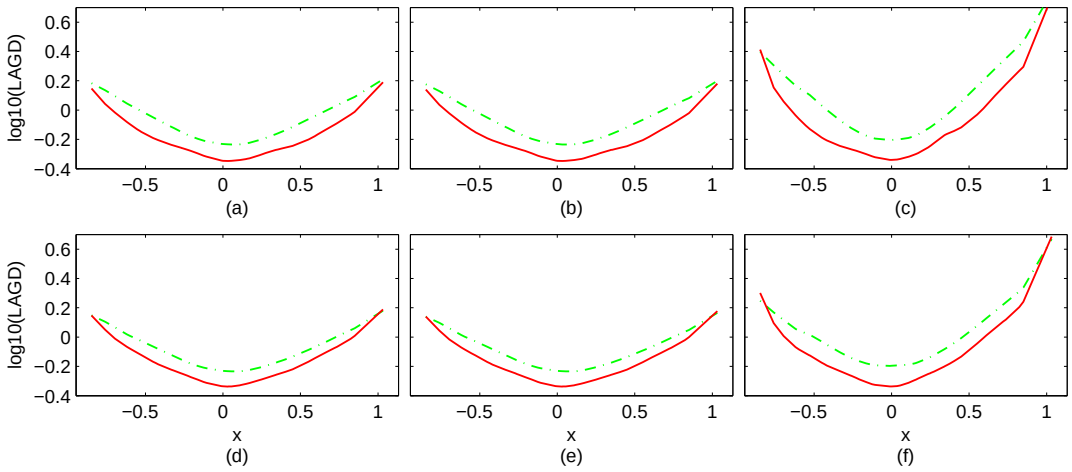

Figure 2.6: The $\log_{10}$(LAGD) curves at each sample point using the intrinsic local linear estimators under the Log-Euclidean (solid line) and affine invariant (dotted line) metrics based on 100 replications for the first two noise models: (a)-(c) Riemannian log normal and (d)-(f) log normal at sample sizes 50. The first, second and third columns correspond to comparisons based on geodesic distances under the affine invariant, Log-Euclidean and Euclidean metrics.

metrics, respectively, to smooth SPD matrices and then calculate smoothed FA curves based on the smoothed SPD matrices. For each method, we assess its performance via the Mean Absolute Deviation Error (MADE) defined by MADE $= n^{-1}\sum_{i=1}^{n}|FA(x_i) - \widehat{FA}(x_i)|$, where $FA(x_i)$ and $\widehat{FA}(x_i)$ are, respectively, the true and estimated FA values at the sample point x_i for $i = 1, 2, \cdots, n$.

Figure 2.7 reveals that the more intrinsic smoothing method outperforms the first smoothing method for the first two noise models. For the Rician noise model, three methods can be comparable but the method under the Log-Euclidean metric is slightly better. This also indicates the possible major improvements of directly smoothing manifold data over the post smoothing method done by the first method. We also calculated the median of MADEs based on the 100 replications (see Figure 2.7(d)-(f)). The first method cannot reconstruct the trend of the FA curve for the first two noise models, whereas the second and third intrinsic methods can accurately estimate the FA curve and reveal its critical features such as the valley.

2.6 HIV Imaging Data

We assess the integrity of white matter in human immunodeficiency virus (HIV) by using DTI and our IPLRE. We consider 46 subjects with 28 HIV+ subjects (20 males and 8 females whose mean age is 40.0 with SD 5.6 years) and 18 healthy controls (9 males and 9 females whose mean

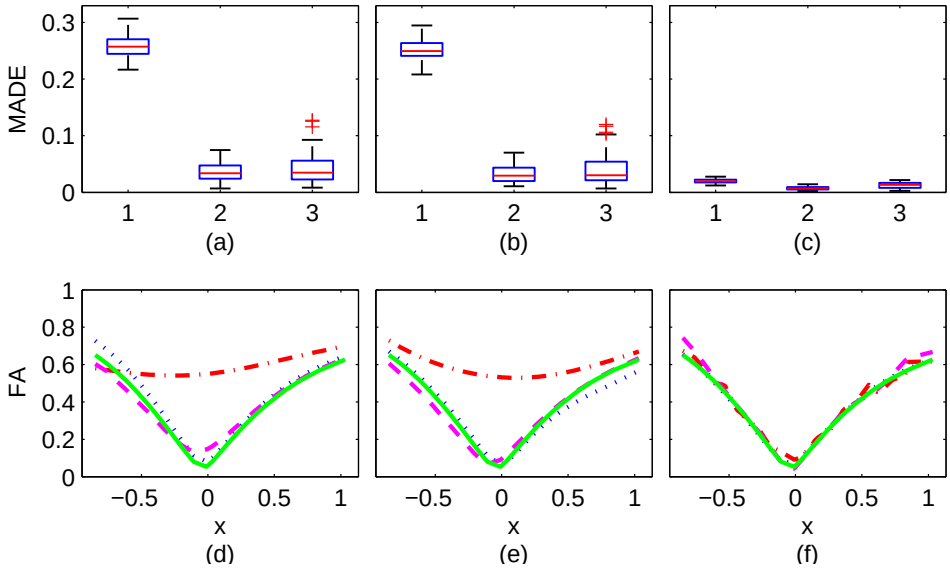

Figure 2.7: Boxplot of the MADE's using the three smoothing methods (1 (2 or 3) represents the first (second or third) method) based on 100 replications for three noise models: (a) Riemannian log normal and (b) log normal with covariance matrix Σ_1 and (c) Rician at sample size 50. Smoothed FA curves for the realizations with median MADE for three noise models: (d) Riemannian log normal and (e) log normal with covariance matrix Σ_1 and (f) Rician. The true FA curve (the solid line), the estimated FA curve using the first method (the dash-dotted line), using the second method (the dotted line) and using the third method (dashed line). This shows that the more intrinsic approach is much better for the first two models and three methods can be comparable but the method under the Log-Euclidean metric is slightly better.

age is 41.2 with SD 7.4 years). Diffusion-weighted images and T1 weighted images were acquired for each subject. The diffusion tensor acquisition scheme includes 18 repeated measures of six non-collinear directions, (1,0,1), (-1,0,1), (0,1,1), (0,1,-1), (1,1,0), and (-1,1,0) at a b-value of 1000 s/mm^2 and a $b = 0$ reference scan. Forty-six contiguous slices with a slice thickness of 2 mm covered a field of view (FOV) of 256 mm^2 with an isotropic voxel size of $2 \times 2 \times 2$ mm^3. High resolution T1 weighted (T1W) images were acquired using a 3D MP-RAGE sequence. A weighted least square estimation method was used to construct the diffusion tensors (Zhu et al., 2007b). Since in the previous DTI findings, the diffusion tensors in the splenium of the corpus callosum were found significantly different between the HIV+ and control groups, we examine the finite sample performance of our method by using this fiber tract. The tensors along the tract were extracted by using the method described in Zhu et al. (2010b). Figure 2.8 displays the splenium of the corpus callosum and the ellipsoidal representation of full tensors on that tract from one selected subject. This involves three steps: (i) registration and atlas construction, (ii) fiber tracking on the

atlas and (iii) collection of tensor data on the atlas fiber tracts.

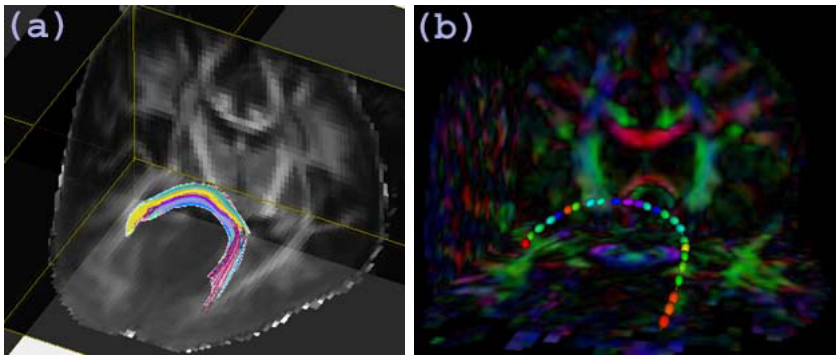

Figure 2.8: (a)The splenium of the corpus callosum in the analysis of HIV DTI data. (b)The ellipsoidal representation of full tensors on the fiber tract from a selected subject.

We calculated the intrinsic local linear estimator of the SPD matrices along this selected tract for each subject under each metric. See Figure 2.9 for the raw and estimated sequences of tensors along the fiber tracts from one subject. It is observed from the ellipsoidal representation of diffusion tensor data (Figure 2.9 (a)) that the data are noisy. It is shown from Figure 2.9 (b) and (c) that the tensors are more spherical at the beginning with low FA values and more anisotropic in the middle part with high FA values. The methods under both metrics reveal very similar trend of diffusion tensors changing along the fiber tract, especially at the first and last third parts. Some differences appear on the right side of the middle part. The estimated tensors in this part are more anisotropic using the method under the Log-Euclidean metric.

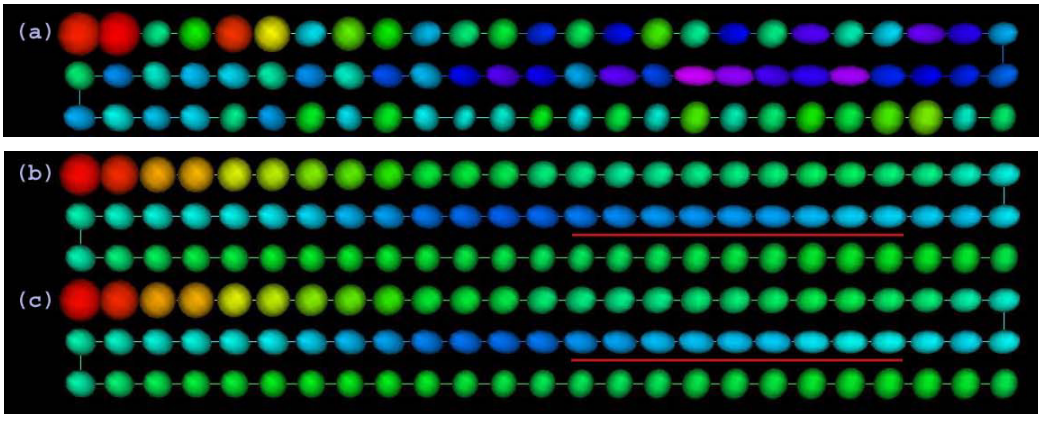

Figure 2.9: (a) Ellipsoidal representations of the diffusion tensor data and estimated tensors using the intrinsic local linear regression under the (b)Log-Euclidean and (c) affine invariant metrics along the fiber tract f1 colored with FA values. The estimated tensors in the middle right part (highlighted in the red line) are more anisotropic using the method under the Log-Euclidean metric.

In many applications, tensor-derived diffusion measures (such as FA, MD (the trace of a diffusion tensor) and PE (the largest eigenvalue of a diffusion tensor)) are derived from noisy diffusion tensor data and standard statistical methods for data in Euclidean space are applied for statistical analysis. However, these scalar measures do not capture all information in the full diffusion tensor and thus can decrease the sensitivity of detecting subtle changes of the white matter structure. For comparison purposes, we also applied the standard local linear regression to the FA, MD and PE data as we did in Section 3.3.3, where we observed differences between the two different smoothing methods for the scalar measures in the simulation. Figure 2.10 shows that there is not much difference between the intrinsic local linear estimators under the two metrics. However, the two smoothing methods are quite different in smoothing FA and PE curves, especially in the middle part around 10 to 25. The smoothed curve is more inclined to the raw curve using the naive method of directly smoothing FA and MD data. The reason is that this naive method is based on only part of information from the diffusion tensor data.

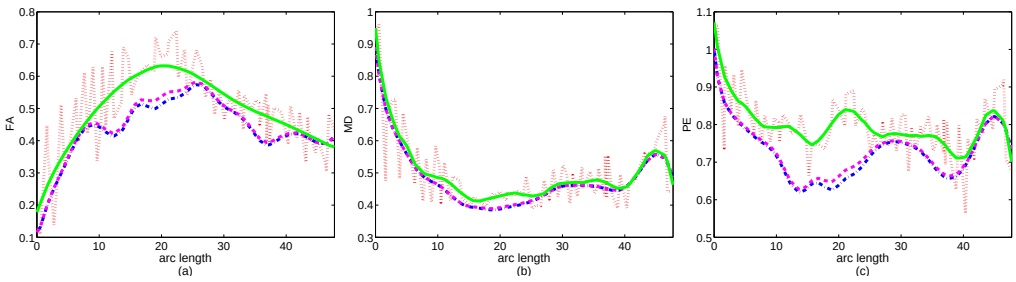

Figure 2.10: (a) FA's , (b) MD's and (c) PE's derived from the raw tensor data (dot line) and estimated tensors using the intrinsic local linear regression under the affine invariant (dash-dot line) and Log-Euclidean (dash line) metrics as the function of arc-length along the tract f1. Estimated FA function along the fiber tract f1 by using the standard local linear regression for scalars (solid line).

Finally, we estimated the mean diffusion tensor curve for each of the two groups: HIV and control groups. In order to detect meaningful group differences, registration is crucial. The 46 HIV DTI data used in our studies, included the splenium tracts and tensors on them are the registered data, which are in the same atlas space. Figure 2.11 displays the estimated mean diffusion tensors along the fiber tract for the two groups using the intrinsic local linear regression for SPD matrices under both the Log-Euclidean and affine invariant metrics. We can observe some obvious changes of diffusion tensors of HIV subjects along the splenium corpus callosum compared with those in

the control group. We also calculated the differences of FA values derived from the the estimated mean diffusion tensors, which corresponds to the color differences in Figure 2.11, and the geodesic distances between estimated mean tensors at each point along the tract. The results are shown in Figure 2.12. This result agrees with previous DTI findings that the splenium of the corpus callosum has been detected as abnormal for the HIV group (Filippi et al. (2001) and Chen et al. (2009)).

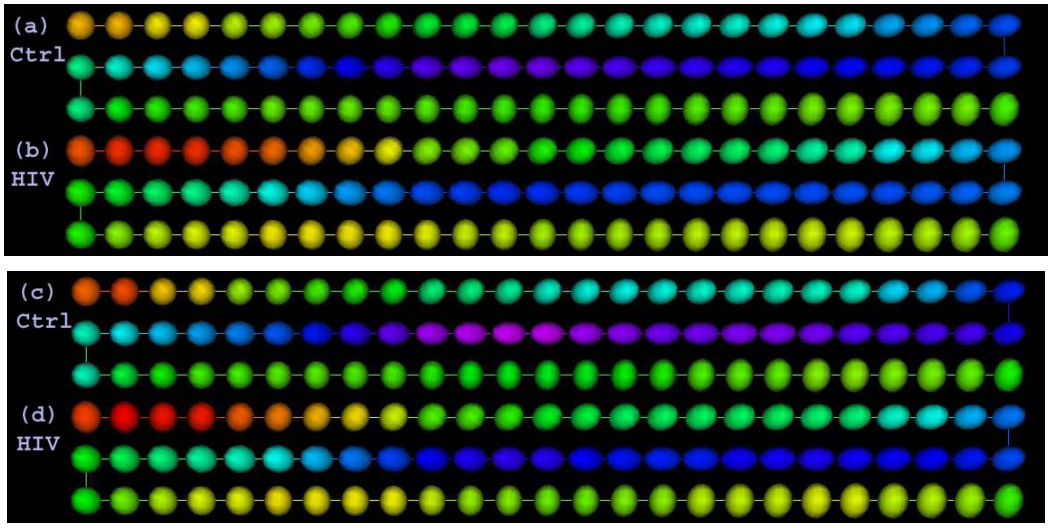

Figure 2.11: Ellipsoidal representations of estimated mean tensors along the fiber tract f1 for the control and HIV groups using the intrinsic local linear regression under the Log-Euclidean ((a) and (b)) and affine invariant ((c) and (d)) metrics colored with FA values.

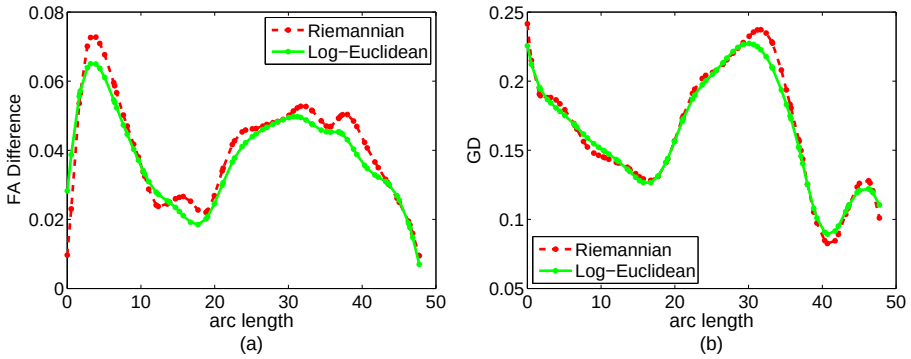

Figure 2.12: (a) FA differences and (b) geodesic distances between pairs of mean diffusion tensors of HIV and control groups along the fiber f1 under the Log-Euclidean (the solid line) and affine invariant (the dashed line) metrics.

2.7 Conclusion and Discussion

We have systematically investigated the intrinsic local polynomial regression methods under two metrics on the space of SPD matrices. Moreover, we have derived the asymptotic bias and variance of ILPRE. We also have developed a relatively straightforward cross-validation bandwidth selection method under each metric. Our theoretical and numerical results have shown that the intrinsic local linear estimator outperforms the intrinsic local constant estimator under both metrics. The proposed cross validation bandwidth selector is straightforward and relatively simple to derive and implement for SPD matrix variate data. However, the relatively high variance of the cross validation bandwidth selector is regarded widely as an impedient to its good performance (Jones et al., 1996; Härdle et al., 1992). In the future, we plan to develop better bandwidth selection methods to reduce the variability of cross validation.

Moreover, we have applied our ILPRE to HIV diffusion tensor curve data in Section 2.6. From the average diffusion tensor curves for the HIV and control groups, we can observe some obvious changes of diffusion tensors of HIV subjects along the spleninium corpus callosum compared with the control group. However, it is important to compare the two groups of tensor curves by using some formal statistical hypothesis testing method. For curve data with responses in Euclidean space, Fan and Lin (1998) developed an adaptive Neyman test. Zhang and Chen (2007) proposed a global L_2-norm-based test statistic to test a general hypothesis testing problem about the group differences. However, no literature is on hypothesis testing for curve data with responses on the space of SPD matrices. In the future, it will be interesting to propose tests for comparing the differences across multiple groups of SPD curves.

Chapter 3

Varying Coefficient Models for Modeling Diffusion Tensors Along White Matter Bundles

3.1 Introduction

Diffusion Tensor Imaging (DTI), which can track the effective diffusion of water in the human brain *in vivo*, has been widely used to map the microstructure and organization of fiber tracts and to assess the integrity of anatomical connectivity in white matter (Basser et al., 1994a,b). In DTI, the degree of diffusivity and the directional dependence of water diffusion in each voxel can be quantified by a 3×3 symmetric positive definite (SPD) matrix, called a diffusion tensor (DT), and its three eigenvalue-eigenvector pairs $\{(\lambda_k, \mathbf{v}_k) : k = 1, 2, 3\}$ with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Fiber tracts in white matter can be constructed by consecutively connecting the principal directions ($\mathbf{v}_1$) of DTs in adjacent voxels (Basser et al., 2000). Therefore, DTs and tensor-derived quantities (e.g., fractional anisotropy (FA)) are distributed along these white matter fiber tracts for each subject. As an illustration, Figure 3.1 (a) presents the right internal capsule tract and Figure 3.1 (b) presents DTs along this tract obtained from 10 subject's, in which each DT is geometrically represented by an ellipsoid. In this representation, the lengths of the semiaxes of the ellipsoid equal the square root of the eigenvalues of a DT, while the eigenvectors define the direction of the three axes. Mathematically, these diffusion tensors along the fiber tract are functionals of SPD matrices. Our research of interest is to statistically model SPD functionals as responses with covariates of interest, such as age and gender, across multiple subjects.

Statistical approaches have been developed for the statistical analysis of tensor-derived quantities along fiber tracts. A tract-based spatial statistics framework was developed to construct local diffusion properties along a white matter skeleton and then perform pointwise hypothesis tests

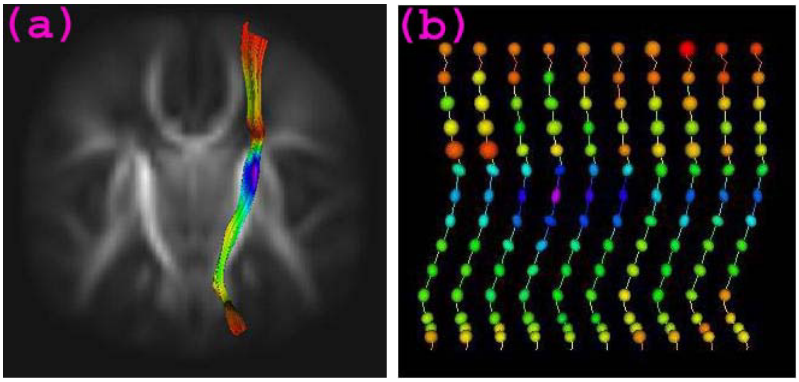

Figure 3.1: (a) The right internal capsule tract. (b) The ellipsoidal representation of full tensors on the fiber tract from 10 selected subjects, colored with FA values.

at each grid point of the skeleton (Smith et al., 2006). A model-based framework was developed to construct the medial manifolds of fiber tracts and then to test pointwise hypotheses based on diffusion properties along the medial manifolds (Yushkevich et al., 2008). However, since these two methods ignore the functional nature of diffusion properties along fiber tracts, they can suffer from low statistical power in detecting interesting features and in exploring variability in tract-based diffusion properties. A functional data analysis framework was used to compare a univariate diffusion property along fiber tracts across two (or more) populations for a single hypothesis test per tract by using functional principal component analysis and the Hotelling T^2 statistic (Goodlett et al., 2009). Their method has two major limitations including only consideration of a univariate diffusion property and the lack of control for other covariates of interest, such as age. To address these two limitations, a functional regression framework was proposed to analyze multiple diffusion properties along fiber tracts as functional responses with a set of covariates of interest, such as age, diagnostic status and gender (Zhu et al., 2010b). An alternative approach, called the generalized functional linear model, was developed with a scalar outcome (e.g., diagnostic group) as responses and fiber bundle diffusion properties as varying covariate functions (or functional predictors) (Goldsmith et al., 2010).

The calculated diffusion properties, which are nonlinear and linear functions of the estimated three eigenvalues of DT containing inherent bias, may be substantially different from the true diffusion properties (Pierpaoli and Basser, 1996; Zhu et al., 2007b; Anderson, 2001). Numerical simulations have shown that estimates of the largest eigenvalue in a DT usually overestimate the true value of λ_1 and that estimates of the smallest eigenvalue usually underestimate λ_3 (Pierpaoli

and Basser, 1996; Zhu et al., 2007b; Anderson, 2001). These differences between the estimated and true eigenvalues, referred to as "sorting bias", subsequently bias the estimation of diffusion properties that are calculated from the values of these estimated eigenvalues (Pierpaoli and Basser, 1996; Zhu et al., 2007b; Anderson, 2001). The sorting bias is pronounced in three types of degenerate DT including isotropic ($\lambda_1 = \lambda_2 = \lambda_3$), oblate ($\lambda_1 = \lambda_2 > \lambda_3$), or prolate ($\lambda_1 > \lambda_2 = \lambda_3$). Previous studies have shown that a major portion of DTs along fiber tracts are prolate tensors (Zhu et al., 2006), and thus directly comparing these biased diffusion properties along fiber tracts can create 'statistical artifacts' including biased parameter estimates and incorrect test statistics and $p-$values for hypotheses of interest as shown in Section 3.3.

To avoid these statistical artifacts, it is important to directly analyze estimated DTs along fiber tracts. There are several advantages of comparing the estimated DTs along fiber tracts with covariates. The first one is that the standard weighted least squared estimates of true DTs are almost unbiased (Zhu et al., 2007b). Moreover, as shown in Section 3.3.1, directly modelling DTs along fiber tracts as a smooth SPD process allows us to incorporate smoothness constraint to further reduce noise in the estimated DTs along fiber tracts, which subsequently leads to reduced noise in estimated diffusion properties along the fiber tracts. Furthermore, the use of scalar diffusion properties ignores the direction information of DT, and thus it can lose the statistical power in detecting the differences in DT oriented in different directions.

There is a growing interest in the DTI literature in developing statistical methods for direct analysis of DTs in the space of SPD matrices. Schwartzman et al. (2008) proposed several parametric models for SPD matrices and derived the distributions of several test statistics for comparing differences between the means of the two (or multiple) groups of SPD matrices. Kim and Richards (2010) developed a nonparametric estimator for the common density function of a random sample of positive definite matrices. Zhu et al. (2009) developed a semi-parametric regression model with SPD matrices as responses in a Riemannian manifold and covariates in a Euclidean space. Barmpoutis et al. (2007) and Davis et al. (2010) proposed tensor splines and local constant regressions for interpolating DTI tensor fields based on the Riemannian metric.

In this chapter, we propose a varying coefficient model (VCLE) to use varying coefficient functions to characterize the association between fiber bundle diffusion tensors and a set of covariates. Since the space of SPD matrices is a Riemanian manifold, to the best of our knowledge, our VCLE

is the first paper for developing a functional data analysis framework for modeling functional manifold-valued responses with covariates in Euclidean space. To account for the curved nature of the SPD space, we employ the Log-Euclidean metric in Arsigny et al. (2006) and then use a weighted least squares estimation method based on the geodesic distance under the Log-Euclidean metric to estimate the varying coefficient functions. Furthermore, we develop global test statistics to test hypotheses on the varying coefficient functions and use a resampling method for approximating its $p-$value. Finally, we construct a simultaneous confidence band to quantify the uncertainty in the estimated coefficient functions and propose a resampling method to approximate the critical point.

The rest of this chapter is organized as follows. Section 3.2 presents VCLE and related statistical inference. Section 3.4 illustrates an application of VCLE in a clinical study of neurodevelopment. Section 3.3 examines the finite sample performance of VCLE via simulation studies. Section 3.5 presents concluding remarks.

3.2 Methodologies

In this section, we present our VCLE for the statistical analysis of DTs along fiber tracts as functional responses with a set of covariates. To compare DTs in populations of DTIs, we use the DTI atlas building followed by atlas fiber tractography and fiber parametrization as described in Goodlett et al. (2009) to extract DTI fibers and establish DTI fiber correspondence across all DTI fiber correspondence across all DTI datasets from different subjects. For the sake of simplicity, we do not include these image processing steps here, which have been discussed in details in Goodlett et al. (2009); Zhu et al. (2010b).

3.2.1 Varying Coefficient Model for Functional SPD data

We need to introduce some notation. Let $\mathrm{Sym}^+(3)$ and $\mathrm{Sym}(3)$ be, respectively, the set of 3×3 SPD matrices and the set of 3×3 symmetric matrices with real entries. Let $\mathrm{vecs}(C) = (c_{1,1}, c_{2,1}, c_{2,2} \cdots, c_{m_1,1}, \cdots, c_{m_1,m_1})^T$ for any $m_1 \times m_1$ symmetric matrix $C = (c_{k,l})$. Let $\mathrm{Ivecs}(\cdot)$ be the inverse operator of $\mathrm{vecs}(\cdot)$ and (a_l) be a $q \times 1$ vector with the l-th element a_l. Let $\mathrm{vec}(C) = (c_{1,1}, \cdots, c_{1,m_2}, \cdots, c_{m_1,1}, \cdots, c_{m_1,m_2})^T$ for any $m_1 \times m_2$ matrix $C = (c_{k,l})$ and $C \otimes D$ denote the

Kronecker product of two matrices C and D.

Let $x \in [0, L_0]$ be the arc length of any point on a specific fiber bundle relative to a fixed end point of the fiber bundle, where L_0 is the longest arc length on the fiber bundle. For the i-th subject, we measure a diffusion tensor, denoted by $S_i(x_j) \in \text{Sym}^+(3)$, at the arc length $x_j \in [0, L_0]$ for the j-th location grid point on the fiber bundle for $j = 1, \cdots, n_G$ and $i = 1, \cdots, n$, where n_G and n denote the numbers of grid points and subjects, respectively. We consider a varying coefficient model given as follows:

$$\log(S_i(x)) \quad = \quad \text{Ivecs}((\mathbf{z}_i^T \beta_l(x))) + \mathcal{U}_i(x) + \mathcal{E}_i(x) \quad \text{for} \quad i = 1, \cdots, n, \tag{3.2.1}$$

where $\log(\cdot)$ denotes the matrix logarithm, $\mathcal{E}_i(x) \in \text{Sym}(3)$ is a 3×3 symmetric matrix of measurement errors, and $\mathcal{U}_i(x) \in \text{Sym}(3)$ characterizes both individual matrix variations from $\text{Ivecs}((\mathbf{z}_i^T \beta_l(x)))$ and the correlation structure between $\log(S_i(x))$ and $\log(S_i(x'))$ for different x and x'. Moreover, $\mathbf{z}_i$ and $\beta_l(x) = (\beta_{1l}(x), \cdots, \beta_{rl}(x))^T$ are, respectively, a $r \times 1$ vector of covariates of interest with $z_{i,1} = 1$ and its associated vector of varying coefficient functions of x for $l = 1, \cdots, 6$. Model (3.2.1) can be regarded as a generalization of varying coefficient models, which have been widely studied and developed for longitudinal, time series, and functional data (Fan and Zhang, 1999; Wu and Chiang, 2000; Fan et al., 2003; Fan and Zhang, 2008; Wang et al., 2008).

Let $\text{SP}(\mu, \Sigma)$ denote a stochastic process with mean $\mu(x)$ and covariance matrix function $\Sigma(x, x')$ for any $x, x' \in [0, L_0]$. It is also assumed that $\text{vecs}(\mathcal{E}_i(x))$ and $\text{vecs}(\mathcal{U}_i(x))$ are independent and respectively, independent and identical copies of $SP(\mathbf{0}, \Sigma_\mathcal{E})$ and $SP(\mathbf{0}, \Sigma_\mathcal{U})$. Moreover, $\text{vecs}(\mathcal{E}_i(x))$ and $\text{vecs}(\mathcal{E}_i(x'))$ for $x \neq x'$ are assumed to be independent and thus $\Sigma_\mathcal{E}(x, x')$ takes the form of $\Sigma_\mathcal{E}(x)\mathbf{1}_{x=x'}$. Finally, the covariance structure of $\text{vecs}(\log(S_i(x_j)))$, denoted by $\Sigma_S(x, x')$, is given by

$$\Sigma_S(x, x') = \text{Cov}(\text{vecs}(\log(S_i(x))), \text{vecs}(\log(S_i(x')))) = \Sigma_\mathcal{U}(x, x') + \Sigma_\mathcal{E}(x, x)\mathbf{1}_{x=x'}. \tag{3.2.2}$$

Since $E(\text{vecs}(\mathcal{E}_i(x))) = E(\text{vecs}(\mathcal{U}_i(x))) = \mathbf{0}$, we are interested in making statistical inference on $\beta(x)$ and reconstruct

$$D(\mathbf{z}, \beta(x)) = \exp(E(\log(S(x)))) = \exp(\text{Ivecs}((\mathbf{z}^T \beta_l(x)))), \tag{3.2.3}$$

where exp(·) denotes the matrix exponential.

As an illustration, in our clinical study on early brain development, we are interested in studying the evolution of the diffusion tensor along one selected fiber tract in 96 healthy pediatric subjects (Figure 3.1). We consider a multivariate varying coefficient model of diffusion tensor along a specific tract as follows:

$$\log(S_i(x)) = \text{Ivecs}((\beta_{1l}(x) + \beta_{2l}(x) \times \text{G}_i + \beta_{3l}(x) \times \text{Gage}_i)) + \mathcal{U}_i(x) + \mathcal{E}_i(x) \tag{3.2.4}$$

for $i = 1, \cdots, 96$, where Gage_i and G_i, respectively, denote the gestational age at the scan time and the gender of the i-th infant. In this case, $\beta(x)$ is an 18×1 vector of coefficient functions with $\beta_l(x) = (\beta_{1l}(x), \beta_{2l}(x), \beta_{3l}(x))^T$ and $\mathbf{z}_i = (1, \text{G}_i, \text{Gage}_i)^T$.

3.2.2 Weighted Least Squares Estimation

Before estimating the coefficient functions, we first introduce the Log-Euclidean metric and the geometry of the space of SPDs under this metric (see Arsigny (2006) for details). Let $\partial_{D(x)} \log(.)$ and $\partial_{\log(D(x))} \exp(.)$ be, respectively, the directional derivatives in the direction of $D(x) \in \text{Sym}^+(m)$ of the log map and in the direction of $\log(D(x)) \in \text{Sym}(m)$ of the exp map. The map $\partial_{D(x)} \log(.)$ is a linear mapping from $T_{D(x)}\text{Sym}^+(m)$ to $T_I\text{Sym}^+(m)$. A scaled Frobenius inner product on $T_{D(x)}\text{Sym}^+(m)$ based on the Log-Euclidean metric is defined as

$$<< U_{D(x)}, V_{D(x)} >>_{D(x)} = < \partial_{D(x)} \log(U_{D(x)}), \partial_{D(x)} \log(V_{D(x)}) >, \tag{3.2.5}$$

where $U_{D(x)}$ and $V_{D(x)}$ are in $T_{D(x)}Sym^+(m)$. The geodesic $\gamma_{D(x)}(t, U_{D(x)})$ is given by $\exp(\log(D(x)) + t\partial_{D(x)} \log(U_{D(x)}))$ for any t. Under the Log-Euclidean metric, the geodesic distance between $D(x)$ and S is uniquely given by

$$d(D(x), S) = \sqrt{\text{tr}[\{\log(D(x)) - \log(S)\}^{\otimes 2}]}. \tag{3.2.6}$$

To estimate the coefficient functions in $\beta(x) = (\beta_1^T(x), \cdots, \beta_q^T(x))^T$, we develop a weighted least squares estimation method based on an adaptive local polynomial kernel (LPK) smoothing technique (Fan and Gijbels, 1996; Wand and Jones, 1995; Wu and Zhang, 2006; Ramsay and

Silverman, 2005; Welsh and Yee, 2006; Zhang and Chen, 2007) under the Log-Euclidean metric. Let $h^{(1)}$ be a given bandwidth. Specifically, using Taylor's expansion, we can expand $\beta_l(x_j)$ at x to obtain

$$\beta_l(x_j) \approx \beta_l(x) + \dot{\beta}_l(x)(x_j - x) = A_l(x)\mathbf{y}_{h^{(1)}}(x_j - x), \tag{3.2.7}$$

where $\mathbf{y}_h(x_j - x) = (1, (x_j - x)/h)^T$ and $A_l(x) = (\beta_l(x), h^{(1)}\dot{\beta}_l(x))$ is an $r \times 2$ matrix, in which $\dot{\beta}_l(x) = (\dot{\beta}_{1l}(x), \cdots, \dot{\beta}_{rl}(x))^T$ is an $r \times 1$ vector and $\dot{\beta}_{kl}(s) = d\beta_{kl}(s)/ds$ for $k = 1, \cdots, r$. We calculate a weighted least squares estimate of $A_l(x)$ as follows.

Let $K(\cdot)$ be a kernel function, such as the Gaussian and uniform kernels (Fan and Gijbels, 1996; Wand and Jones, 1995). For a fixed bandwidth $h^{(1)}$, we calculate a weighted least squares estimate of $A_l(x)$, denoted by $\hat{A}_l(x)$, which minimizes an objective function given by

$$\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathrm{tr}\{[\log(S_i(x_j)) - \mathrm{Ivecs}((\mathbf{z}_i^T A_l(x)\mathbf{y}_{h^{(1)}}(x_j - x)))]^{\otimes 2}\}, \tag{3.2.8}$$

where $K_{h^{(1)}}(\cdot) = K(\cdot/h^{(1)})/h^{(1)}$ is a rescaled kernel function and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector or any matrix $\mathbf{a}$. It can be shown that $\mathrm{vec}(\hat{A}_l(x)) = (\hat{\beta}_{1l}(x), h^{(1)}\hat{\dot{b}}_{1l}(x), \cdots, \hat{\beta}_{rl}(x), h^{(1)}\hat{\dot{b}}_{rl}(x))^T$ is given by

$$\Sigma(h^{(1)}, x)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i \otimes \mathbf{y}_{h^{(1)}}(x_j - x)](\log(S_i(x_j)))_l, \tag{3.2.9}$$

where $\Sigma(h^{(1)}, x) = \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i^{\otimes 2} \otimes \mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2}]$. Thus, we have

$$\hat{\beta}_l(x) = (\hat{\beta}_{1k}(x), \cdots, \hat{\beta}_{rl}(x))^T = [I_r \otimes (1, 0)]\mathrm{vec}(\hat{A}_l(x)), \tag{3.2.10}$$

where I_r is an $r \times r$ identity matrix.

We pool the data from all n subjects and select an estimated bandwidth $h^{(1)}$, denoted by $\hat{h}_e^{(1)}$ by minimizing the cross-validation score given by

$$CV_1(h^{(1)}) = (nn_G)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\mathrm{tr}\{[\log(S_i(x_j)) - \mathrm{Ivecs}((\mathbf{z}_i^T\hat{\beta}_l(x_j, h^{(1)})^{(-i)}))]^{\otimes 2}\}, \tag{3.2.11}$$

where $\hat{\beta}_l(x, h^{(1)})^{(-i)}$ is the weighted least squares estimator of $\beta_l(x)$ for the bandwidth $h^{(1)}$ based on observed data with the observations from the i-th subject excluded. Finally, by substituting

$\hat{h}_e^{(1)}$ into Equation (3.2.10), we can obtain an estimate of $\beta_l(x)$, denoted by $\hat{\beta}_{l,e}(x)$. Combining all $\hat{\beta}_{l,e}(x)$ leads to $\hat{\beta}_e(x) = [\hat{\beta}_{1,e}(x), \cdots, \hat{\beta}_{q,e}(x)]$.

3.2.3 Smoothing Individual Functions and Estimating Covariance Matrices

To simultaneously construct the individual function $\mathcal{U}_i(x)$, we also employ the local polynomial kernel smoothing technique. Specifically, using Taylor's expansion, we can expand $\mathcal{U}_i(x_j)$ at x to obtain

$$\mathcal{U}_i(x_j) \approx \mathcal{U}_i(x) + \dot{\mathcal{U}}_i(x)(x_j - x), \tag{3.2.12}$$

which is equivalent to

$$\text{vecs}(\mathcal{U}_i(x_j)) \approx \text{vecs}(\mathcal{U}_i(x)) + \text{vecs}(\dot{\mathcal{U}}_i(x))(x_j - x) = \mathcal{D}_i(x)^T \mathbf{y}_{h^{(2)}}(x_j - x), \tag{3.2.13}$$

where $\mathcal{D}_i(x) = (\text{vecs}(\mathcal{U}_i(x)), h^{(2)}\text{vecs}(\dot{\mathcal{U}}_i(x)))^T$ is a 2×6 matrix. We develop an algorithm to estimate $\mathcal{D}_i(x)$ as follows. For each fixed x and each bandwidth $h^{(2)}$, the weighted least square estimator of $\mathcal{D}_i(x)$ can be calculated by minimizing an objective function given by

$$\sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\text{tr}\{[\log(S_i(x_j)) - \text{Ivecs}((\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j)) - (\mathcal{D}_i(x)^T\mathbf{y}_{h^{(2)}}(x_j - x)))]^{\otimes 2}\}.$$

With some calculation, it can be shown that the weighted least square estimate of $\mathcal{D}_i(x)$, denoted as $\hat{\mathcal{D}}_i(x)$, has the following explicit expression:

$$\Sigma_1(h^{(2)}, x)^{-1}\sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)(\text{vecs}(\log(S_i(x_j)) - (\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j))))^T, \tag{3.2.14}$$

where $\Sigma_1(h^{(2)}, x) = \sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)^{\otimes 2}$. Let $\mathbf{e}_{2,1} = (1, 0)^T$. Then, $\text{vecs}(\mathcal{U}_i(x))$ can be estimated by

$$\text{vecs}(\hat{\mathcal{U}}_i(x))^T = \mathbf{e}_{1,2}^T\hat{\mathcal{D}}_i(x) = \sum_{j=1}^{n_G}\tilde{K}^0_{h^{(2)}}(x_j - x, x)(\text{vecs}(\log(S_i(x_j))) - (\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j)))^T, \tag{3.2.15}$$

where $\tilde{K}^0_{h^{(2)}}(\cdot,\cdot)$ is the empirical equivalent kernel (Fan and Gijbels, 1996). Finally, let R_i be a matrix with the j-th row $\text{vecs}(\log(S_i(x_j))) - (\mathbf{z}_i^T\hat{\beta}_{l,e}(x_j))^T$ and $\mathcal{S}$ be a $n_G \times n_G$ smoothing matrix

with the (i,j)-th element $\tilde{K}^0_{h^{(2)}}(x_j - x_i, x_i)$. We can obtain

$$(\text{vecs}(\hat{\mathcal{U}}_i(x_1)), \cdots, \text{vecs}(\hat{\mathcal{U}}_i(x_{n_G})))^T = \mathcal{S}R_i. \tag{3.2.16}$$

Let $(C)_j$ represent the j-th column vector in a $m_1 \times m_2$ matrix C and $\mathcal{S}_{jj}$ be the (j,j) element of the matrix $\mathcal{S}$. Let Λ be a 6×6 diagonal matrix with q diagonal entries $1, 2, 1, 2, 2, 1$. We pool the data from all n subjects and select an estimated bandwidth of $h^{(2)}$, denoted as $\hat{h}_e^{(2)}$, by minimizing the cross-validation score given by

$$CV_2(h^{(2)}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_G} \frac{\|\Lambda^{1/2}((R_i)_j - (\mathcal{S}R_i)_j)\|^2}{(1 - \mathcal{S}_{jj})^2}. \tag{3.2.17}$$

Replacing $\mathcal{S}_{jj}$ in Equation (3.2.17) by the average of $\mathcal{S}_{11}, \cdots, \mathcal{S}_{n_G n_G}$, we can get the following generalized cross-validation (GCV) score

$$GCV_2(h^{(2)}) = n^{-1} \frac{\sum_{i=1}^{n} \text{tr}\{(R_i - \mathcal{S}R_i)^{\otimes 2}\}}{(1 - n^{-1}\text{tr}(\mathcal{S}))^2}. \tag{3.2.18}$$

Without special saying, for this part, we use generalized cross-validation score $GCV_2(h^{(2)})$ to select the bandwidth throughout this paper. Based on $\hat{h}_e^{(2)}$, we can use Equation (3.2.15) to estimate $\text{vecs}(\mathcal{U}_i(x))$ and $\mathcal{U}_i(x)$, denoted by $\text{vecs}(\hat{\mathcal{U}}_{i,e}(x))$ and $\hat{\mathcal{U}}_{i,e}(x)$, respectively, for all i.

After obtaining $\hat{\mathcal{U}}_{i,e}(x)$, we can estimate the mean function $\mathcal{U}(x)$ and the covariance function $\Sigma_{\mathcal{U}}(x, x')$. Specifically, we estimate $\mathcal{U}(x)$ and $\Sigma_{\mathcal{U}}(x, x')$ by using their empirical counterparts based on the estimated $\hat{\mathcal{U}}_{i,e}(x)$:

$$\hat{\mathcal{U}}_e(x) = n^{-1} \sum_{i=1}^{n} \hat{\mathcal{U}}_{i,e}(x) \quad \text{and}$$
$$\hat{\Sigma}_{\mathcal{U}}(x, x') = (n - q)^{-1} \sum_{i=1}^{n} \text{vecs}(\hat{\mathcal{U}}_{i,e}(x)) \text{vecs}(\hat{\mathcal{U}}_{i,e}(x'))^T.$$

We construct a nonparametric estimator of the covariance matrix $\Sigma_{\mathcal{E}}(x, x)$ as follows. Let $\hat{\mathcal{E}}_i(x_j) = \log(S_i(x_j)) - \text{Ivecs}((\mathbf{z}_i^T \hat{\beta}_{l,e}(x_j))) - \hat{\mathcal{U}}_{i,e}(x_j)$ be estimated residuals for $i = 1, \cdots, n$ and

$j = 1, \cdots, n_G$. We consider the kernel estimate of $\Sigma_{\mathcal{E}}(x, x)$ given by

$$\hat{\Sigma}_{\mathcal{E}}(x, x) = (n - q)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_G} \frac{K_{h^{(3)}}(x_j - x) \text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}}{\sum_{j=1}^{n_G} K_{h^{(3)}}(x_j - x)}. \tag{3.2.19}$$

Let $\tilde{\Sigma}_{\mathcal{E}}(x_j, x_j) = (n-q)^{-1} \sum_{i=1}^{n} \text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}$. To select an estimated bandwidth $h^{(3)}$, denoted by $\hat{h}_e^{(3)}$, we minimize the cross-validation score given by

$$CV_3(h^{(3)}) = (nn_G)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_G} \text{tr}\{[\text{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2} - \hat{\Sigma}_{\mathcal{E}}(x_j, x_j, h^{(3)})^{(-i)}]^{\otimes 2} \tilde{\Sigma}_{\mathcal{E}}(x_j, x_j)^{-1}]^{\otimes 2}\}, \tag{3.2.20}$$

where $\hat{\Sigma}_{\mathcal{E}}(x, x, h^{(3)})^{(-i)}$ is the weighted least squares estimator of $\hat{\Sigma}_{\mathcal{E}}(x, x)$ based on observed data with the observations from the i-th subject excluded. Based on $\hat{h}_e^{(3)}$, we can use Equation (3.2.19) to estimate $\Sigma_{\mathcal{E}}(x, x)$, denoted by $\hat{\Sigma}_{\mathcal{E},e}(x, x)$.

3.2.4 Asymptotic Properties

We will use the following theorems to construct the global test statistics and simultaneous confidence bands for coefficient functions, whose detailed assumptions can be found in Appendix D and whose proofs are similar to those in Zhu et al. (2010a).

Theorem 3.2.1. *If assumptions (C1)-(C6) in Appendix D are true, then* $\sqrt{n}\{vec(\hat{\beta}(x) - \beta(x) - 0.5u_2\ddot{\beta}(x)h^{(1)2}[1 + o_p(1)]) : x \in [0, L_0]\}$ *converges weakly to a centered Gaussian process* $G(\cdot)$ *with covariance matrix* $\Sigma_{\mathcal{U}}(x, x') \otimes \Omega_Z^{-1}$.

Theorem 3.2.1 establishes weak convergence of $\hat{\beta}(x)$ as a stochastic process indexed by $x \in [0, L_0]$ and forms the foundation for constructing global test statistic and simultaneous confidence bands for $\beta(\cdot)$.

Theorem 3.2.2. *If assumptions (C1)-(C7) in Appendix D are true, then*

$$\sup_{(x,t)\in[0,L_0]^2} |\hat{\Sigma}_{\mathcal{U}}(x, t) - \Sigma_{\mathcal{U}}(x, t)| = O_p(n^{-1/2} + (n_G h^{(2)})^{-1} + \hat{h}_e^{(1)2} + h^{(2)2} + (\log n/n)^{1/2}).$$

Theorem 3.2.2 shows the uniform convergence of $\hat{\Sigma}_{\mathcal{U}}(x, t)$. This result is useful for constructing the global and local test statistics for testing the covariate effects.

3.2.5 Hypothesis Test

In neuroimaging studies, some scientific questions require the comparison of fiber bundle diffusion tensors along fiber bundles across two (or more) diagnostic groups and the assessment of the development of fiber bundle diffusion properties along time. Such questions can often be formulated as linear hypotheses of $\beta(x)$ as follows:

$$H_0 : \mathbf{R}\beta(x) = \mathbf{b}_0(x) \text{ for all } x \text{ vs. } H_1 : \mathbf{R}\beta(x) \neq \mathbf{b}_0(x), \tag{3.2.21}$$

where $\mathbf{R}$ is a $t \times 6r$ matrix of full row rank and $\mathbf{b}_0(x)$ is a given $t \times 1$ vector of functions .

We propose both local and global test statistics. The local test statistic can identify the exact location of significant grid point on a specific tract. At a given grid point x_j on a specific tract, we test the local null hypothesis $H_0(x_j) : \mathbf{R}\beta(x_j) = \mathbf{b}_0(x_j)$ against $H_1(x_j) : \mathbf{R}\beta(x_j) \neq \mathbf{b}_0(x_j)$. We use a local test statistic $T_n(x_j)$ defined by

$$T_n(x_j) = n\mathbf{d}(x_j)^T\{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x_j, x_j) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1}\mathbf{d}(x_j), \tag{3.2.22}$$

where $\hat{\Omega}_Z = n^{-1}\sum_{i=1}^n \mathbf{z}_i^{\otimes 2}$ and $\mathbf{d}(x) = \mathbf{R}(\hat{\beta}_o(x)^T - \text{bias}(\hat{\beta}_o(x)^T)) - \mathbf{b}_0(x)$. Following Fan and Zhang (2000), a smaller bandwidth leads to a small value of $\text{bias}(\hat{\beta}_o(x))$. Moreover, according to our simulation studies below, we have found that the effect of dropping $\text{bias}(\hat{\beta}_o(x))$ is negligible and therefore, we drop it from now on.

We test the null hypothesis $H_0 : \mathbf{R}\beta(x) = \mathbf{b}_0(x)$ for all x using a global test statistic T_n defined by

$$T_n = n\int_0^{L_0} \mathbf{d}(x)^T[\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x, x) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T]^{-1}\mathbf{d}(x)dx. \tag{3.2.23}$$

Let $X_{\mathbf{R}}(\cdot)$ be a Gaussian process with zero mean and covariance structure $\Sigma_{\mathbf{R}}(x, x')$, which is given by

$$\Sigma_{\mathbf{R}}(x, x') = \{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x, x) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1/2}\{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x, x') \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}\{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x', x') \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-T/2}. \tag{3.2.24}$$

It follows from Theorem 3.2.1 that $\sqrt{n}\{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x, x) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1}\mathbf{d}(x)$ converges weakly to $X_{\mathbf{R}}(x)$. Therefore, let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes, it follows from the

continuous mapping theorem that as both n and n_G converges to infinity, we have

$$T_n \Rightarrow \int_0^{L_0} X_{\mathbf{R}}(x)^T X_{\mathbf{R}}(x) dx.$$

Based on that result, we develop a wild bootstrap method to approximate the p-value of T_n.

Step (i): Fit model (3.2.1) under the null hypothesis H_0, which yields $\hat{\beta}_e^*(x_j) = (\hat{\beta}_{l,e}^*(x_j))$, $\hat{\mathcal{U}}_{i,e}^*(x_j)$ and $\hat{\mathcal{E}}_{i,e}^*(x_j)$ for $i = 1, \cdots, n$ and $j = 1, \cdots, n_G$.

Step (ii): Generate a random sample $\tau_i^{(g)}$ and $\tau_i(x_j)^{(g)}$ from a $N(0,1)$ generator for $i = 1, \cdots, n$ and $j = 1, \cdots, n_G$ and then construct

$$\hat{S}_i(x_j)^{(g)} = \exp(\log(D(\mathbf{z}_i, \hat{\beta}_{l,e}^*(x_j))) + \tau_i^{(g)} \hat{\mathcal{U}}_i^*(x_j) + \tau_i(x_j)^{(g)} \hat{\mathcal{E}}_i^*(x_j)).$$

Then, based on $\hat{S}_i(x_j)^{(g)}$, we recalculate $\tilde{h}_e^{(1)}$, $\hat{\beta}_e(x)^{(g)}$, and $\mathbf{d}(x)^{(g)} = \mathbf{R}\hat{\beta}_e(x)^{(g)} - \mathbf{b}_0(x)$. We compute

$$\begin{aligned} T_n^{(g)} &= n \int_0^{L_0} \mathbf{d}(x)^{(g)T} \{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x,x) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1} \mathbf{d}(x)^{(g)} dx, \\ T_n(x_j)^{(g)} &= n\mathbf{d}(x_j)^{(g)T} \{\mathbf{R}(\hat{\Sigma}_{\mathcal{U}}(x_j,x_j) \otimes \hat{\Omega}_Z^{-1})\mathbf{R}^T\}^{-1} \mathbf{d}(x_j)^{(g)} \quad \text{for} \quad j = 1, \cdots, n_G. \end{aligned}$$

Step (iii): Aggregate the results of Step (2) over $g = 1, \cdots, G$ to obtain $\{T_{n,\max}^{(g)} = \max_{1 \le j \le n_G} T_n(x_j)^{(g)} : g = 1, \cdots, G\}$ and calculate $p(x_j) = G^{-1} \sum_{g=1}^{G} 1(T_{n,\max}^{(g)} \ge T_n(x_j))$ for each x_j. The $p(x_j)$ is the corrected p-value at the location x_j.

Step (iv): Aggregate the results of Step (2) over $g = 1, \cdots, G$ to obtain $\{T_n^{(g)} : g = 1, \cdots, G\}$ and calculate $p = G^{-1} \sum_{g=1}^{G} 1(T_n^{(g)} \ge T_n)$. If p is smaller than a pre-specified significance level α, say 0.05, then we reject the null hypothesis H_0.

3.2.6 Confidence Band

We construct a confidence band for $D(\mathbf{z}, \beta(x)) \in \mathrm{Sym}^+(3)$ over $x \in [0, L_0]$ given a fixed $\mathbf{z}$. Specifically, for a given significance level α, we construct a simultaneous confidence region in the space of SPD matrices for each $\mathbf{z}$, based on the critical value $C_z(\alpha)$ such that

$$P(d(D(\mathbf{z}, \beta(x)), D(\mathbf{z}, \hat{\beta}(x))) \le C_z(\alpha) \quad \text{for all} \quad x \in [0, L_0]) = 1 - \alpha. \tag{3.2.25}$$

Note that $d(D(\mathbf{z},\beta(x)),D(\mathbf{z},\hat{\beta}(x))) = \sqrt{\text{tr}[\{\text{Ivecs}(\mathbf{z}^T\{\beta_l(x)-\hat{\beta}_l(x)\})\}^2]}$. By the Theorem 3.2.1, $\sup_{x\in[0,L_0]}\sqrt{n}\{\hat{\beta}_{l,e}(x)-\beta_l(x)\}$ converges in distribution to $G_l(x)$, where $G_l(\cdot)$ is a centered Gaussian process indexed by $x\in[0,L_0]$, we have that as $n\to\infty$,

$$\sqrt{n}d(D(\mathbf{z},\beta(x)),D(\mathbf{z},\hat{\beta}(x))) \Rightarrow \sqrt{\text{tr}[\{\text{Ivecs}((\mathbf{z}^T G_l(x)))\}^2]}.$$

To determine $C_z(\alpha)$, we develop an efficient resampling method (Zhu et al., 2007a; Kosorok, 2003) to approximate it. Let $\mathbf{e}_{r,l}$ be the $r\times 1$ vector with the l-th element 1 and all others 0, and $\hat{\mathbf{r}}_{i,l}(x_j) = (\text{vecs}(\log(S_i(x_j))))_l - \mathbf{z}_i^T\hat{\beta}_{l,e}(x_j)$. For $g=1,\cdots,G$, we independently simulate $\{\tau_i^{(g)} : i=1,\cdots,n\}$ from $N(0,1)$ for all l, and then we calculate a stochastic process $G_l(x)^{(g)}$, which is defined as follows:

$$\sqrt{n}\mathbf{e}_l^T[I_r\otimes(1,0)]\text{vec}(\Sigma(h^{(1)},x)^{-1}\sum_{i=1}^{n}\tau_i^{(g)}\sum_{j=1}^{n_G}K_{h^{(1)}}(x_j-x)[\mathbf{z}_i\otimes\mathbf{y}_{h^{(1)}}(x_j-x)]\hat{\mathbf{r}}_{i,l}(x_j)). \quad (3.2.26)$$

Finally, we calculate $\sup_{x\in[0,L_0]}\sqrt{\text{tr}[\{\text{Ivecs}((\mathbf{z}^T G_l(x)^{(g)}))\}^2]/n}$ for all $g=1,\cdots,G$ and use the resulting $1-\alpha$ empirical percentile to estimate $C_z(\alpha)$.

Moreover, we can construct confidence bands for the functional coefficients $\beta(x)=(\beta_1(x)^T,\cdots,\beta_q(x)^T)^T$ with $\beta_l(x)=(\beta_{1l},\cdots,\beta_{rl})^T$ in model (3.2.1). Specifically, for a given significance level α, we construct a confidence band for each $\beta_{kl}(x),\ \ k=1,\cdots,r$ such that

$$P(\hat{\beta}_{kl}^{L,\alpha}(x)<\beta_{kl}(x)<\hat{\beta}_{kl}^{U,\alpha}(x)\ \text{ for all }\ x\in[0,L_0])=1-\alpha, \quad (3.2.27)$$

where $\hat{\beta}_{kl}^{L,\alpha}(x)$ and $\hat{\beta}_{kl}^{U,\alpha}(x)$ are the lower and upper limits of the confidence band. It follows from Theorem 3.2.1 that $\sup_{x\in[0,L_0]}|\sqrt{n}[\hat{\beta}_{kl,e}(x)-\beta_{kl}(x)]|$ converges in distribution to $\sup_{x\in[0,L_0]}|G_{kl}(x)|$. We define the critical point $C_{kl}(\alpha)$ such that $P(\sup_{x\in[0,L_0]}|G_{kl}(x)|\le C_{kl}(\alpha))=1-\alpha$. Thus, a $1-\alpha$ simultaneous confidence band for $\beta_{kl}(x)$ is given as follows:

$$\left(\hat{\beta}_{kl,e}(x)-\frac{C_{kl}(\alpha)}{\sqrt{n}},\quad \hat{\beta}_{kl,e}(x)+\frac{C_{kl}(\alpha)}{\sqrt{n}}\right). \quad (3.2.28)$$

Let $\mathbf{e}_{kl}$ be an $6r\times 1$ vector with the $(l-1)r+k$-th element 1 and all others 0. The critical point $C_{kl}(\alpha)$ can be calculated as $1-\alpha$ empirical percentile of $\sup_{x\in[0,L_0]}|\mathbf{e}_{kl}G_l(x)^{(g)}|$ for all $g=1,\cdots,G$.

3.3 Simulation Studies

We conducted three sets of Monte Carlo simulations. The first set of simulations was to evaluate the Type I and II error rates of the global test statistic T_n. The second set was to compare the power in detecting the group effect using either the whole diffusion tensor or the diffusion properties. The third set was to evaluate the coverage probabilities of the simultaneous confidence bands of the functional coefficients $\beta(x)$ and $D(\mathbf{z}, \beta(x))$ for each $\mathbf{z}$. We simulate diffusion tensors along the right internal capsule tract (Figure 3.1 (a)) as follows:

$$S_i(x) = \exp(\text{Ivecs}((\beta_{1l}(x) + \beta_{2l}(x) \times \text{G}_i + \beta_{3l}(x) \times \text{Gage}_i)) + \mathcal{U}_i(x) + \mathcal{E}_i(x)) \tag{3.3.1}$$

where Gage_i and G_i, respectively, denote the gestational age at the scan time and gender of the i-th infant, $\text{vecs}(\mathcal{U}_i(x)) = ((\mathcal{U}_i(x))_1, \cdots, (\mathcal{U}_i(x))_6)^T$ is a Gaussian process with zero mean and covariance matrix $\Sigma_{\mathcal{U}}(x, x')$ and $\text{vecs}(\mathcal{E}_i(x)) = ((\mathcal{E}_i(x))_1, \cdots, (\mathcal{E}_i(x))_6)^T$ is a Gaussian random vector with zero mean and covariance matrix $\Sigma_{\mathcal{E}}(x, x)\mathbf{1}(x = x')$. To mimic imaging data, we used the diffusion tensors along the right internal capsule tract from all 96 infants in our clinical data to estimate $\beta(x)$ by $\hat{\beta}(x)$ via equation (3.2.10), $\mathcal{U}(x)$ by $\hat{\mathcal{U}}(x)$ via equation (3.2.15), and $\mathcal{E}(x)$ by $\hat{\mathcal{E}}(x)$ via $\hat{\mathcal{E}}(x) = \log(S_i(x)) - \text{Ivecs}((\hat{\beta}_l(x)^T\mathbf{z})) - \hat{\mathcal{U}}(x)$. We fixed all the parameters at their values obtained from our clinical data, except that we assumed $(\beta_{31}(x), \cdots, \beta_{36}(x)) = c(\hat{\beta}_{31}(x), \cdots, \hat{\beta}_{36}(x))$, where c is a scalar which takes a different value as specified below to study the power and $(\hat{\beta}_{31}(x), \cdots, \hat{\beta}_{36}(x))$ were estimators obtained from our clinical data. Figure 3.2 displays the estimated diffusion tensors using the VCLE method when $c = 1$ by using two covariate (Figure 3.2 (c)) and univariate (Figure 3.2 (d)) modeling. Note that the multivariate modeling did a better job of recovering ground truth than univariate modeling. It is also observed from Figure 3.3 that the multivariate modeling obtains a smaller geodesic distance between simulated and estimated diffusion tensors the the univariate modeling does.

3.3.1 Simulation 1

In neuroimaging studies, a lot of scientific questions require the assessment of the development of fiber bundle diffusion tensors across time. In this simulation study, the questions were formulated

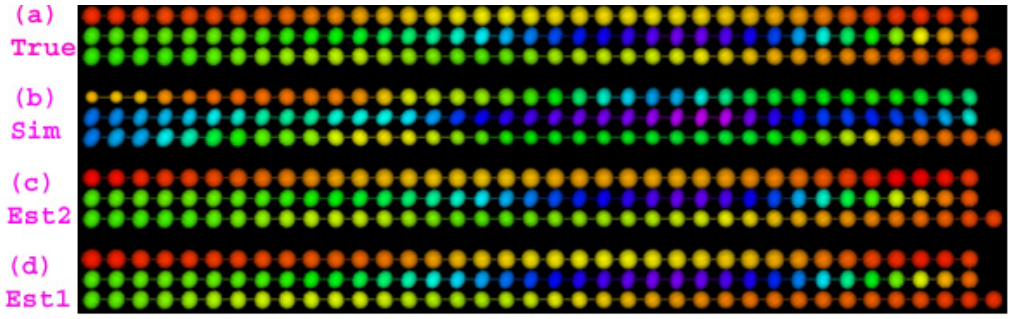

Figure 3.2: Ellipsoidal representations of the true (a), simulated (b) and estimated (c) (based on two covariates) and (d) (based on univariate) diffusion tensors along the the right internal capsule tract, colored with FA values.

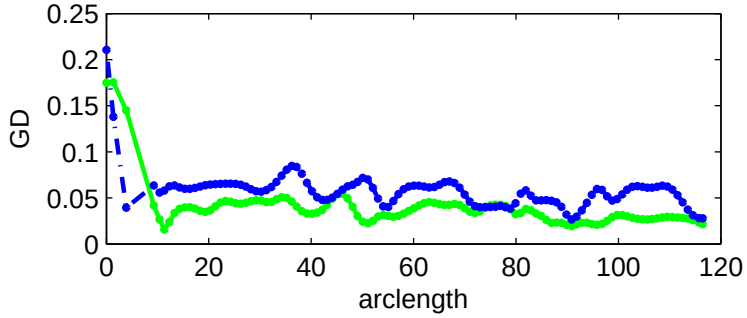

Figure 3.3: Geodesic distances between simulated and estimated diffusion tensors (solid lines (based on two covariates) and dash-dotted (based on univariate)) along the the right internal capsule tract.

as the hypotheses test $H_0 : \beta_{31}(x) = \cdots = \beta_{36}(x) = 0$ for all x along the right internal capsule tract against $H_1 : \beta_{3l}(x) \neq 0$ for at least one x on the tract for some $l = 1, \cdots, 6$. We first assumed $c = 0$ to assess the Type I error rates (size)for the global test statistic T_n, and then we assumed $c = .2, .4, .6$, and .8 to examine the Type II error rates (power) for T_n at different effect sizes. In order to evaluate the Type I and II error rates at different sample sizes, we also consider sample size $n = 48$ except $n = 96$. For $n = 96$, the values of gender and gestational age were taken those of the 96 infants in our clinical study. For $n = 48$, we randomly chose 18 males from the 36 male infants and 30 females from the 60 female infants. Then for all simulations, we used their values of gender and gestational age to simulate the diffusion tensors along the right internal capsule tract. Note that the number of grid points on the right internal capsule equals $n_G = 112$ for both cases.

We applied the VCLE procedure to the simulated diffusion tensors. Particularly, we approximated the p-value of T_n using the wild bootstrap method described in the hypothesis test discussed in Section 3.2.5. For each simulation, the significance levels were set at $\alpha = .05$ and .01, and 100 replications were used to estimate the rejection rates. For a fixed α, if the Type I rejection rate is smaller than α, then the test is conservative, whereas if the Type I rejection rate is greater than α, then the test is anticonservative, or liberal.

Figure 3.4(a) displays the rejection rates for T_n based on the resampling method for both sample sizes ($n = 48$ or 96) and all effect sizes ($c = 0$, .2, .4, .6, or .8) at both significance levels ($\alpha = .01$ or .05) using full diffusion tensors. The statistical power for rejecting the null hypothesis increases with the sample size, the effect size and the significance level, which is consistent with our expectation.

3.3.2 Simulation 2

For each simulated diffusion tensor generated in 3.3.1 , we calculated its three eigenvalues, FA and MD values. Then we applied the VCLE procedure for multiple measures to the simulated values of FA, MD, joint FA and MD, each of the eigenvalues, and the three eigenvalues together, respectively and then tested the significance of the gestational age effect. Figure 3.4(b)-(h) displays the power for detecting the gestational age effect in each of these cases. First, as in Figure 3.4 (a), it is also observed that the statistical power for rejecting the null hypothesis increases with the sample size, the effect size and the significance level. Second, it is observed that the statistical power is much higher when we use the whole tensor instead of FA, MD, λ_1, λ_2 or λ_3 alone. Third, the power

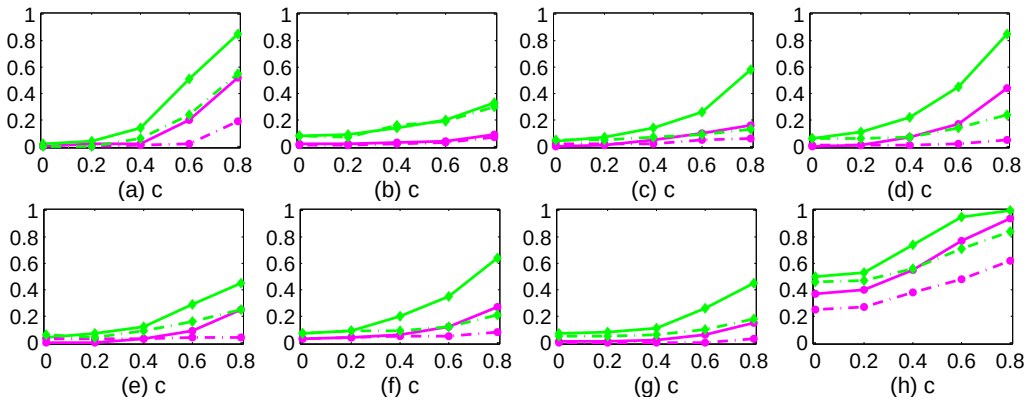

Figure 3.4: Simulation study: Type I and Type II error rates. Rejection rates of T_n based on the resampling method are calculated at five different values of the effect size c for sample sizes of 48 (dash-dotted lines), 96(solid lines) subjects at the .05 (lines with diamond markers) and .01(lines with circle markers) significance levels using (a) diffusion tensor; (b) FA values; (c) MD values; (d) joint values of FA and MD; (e) λ_1; (f)λ_2; (g)λ_3; and (h) joint values of λ_1, λ_2 and λ_3.

is a little higher at small effect size when we use the joint FA and MD values instead of full diffusion tensors while at higher effect size, they almost have the same power (Fig. 3.4(a) and (d)). Fourth, the power is higher using three eigenvalues together than using the full diffusion tensors. Last, it is noted that the Type I error is greater than the .05 significance level when we use the joint FA and MD values (Fig. 3.4(d)) or use the three eigenvalues together (Fig. 3.4(h)). It means that the test is liberal. As mentioned in Section 3.1, the calculated diffusion properties, which are nonlinear and linear functions of the estimated three eigenvalues of DT containing inherent bias (sorting bias), may be substantially different from the true diffusion properties. Thus directly comparing these biased diffusion properties along fiber tracts can create incorrect p-values for hypotheses of interest.

3.3.3 Simulation 3

The third set of Monte Carlo simulations was carried out to evaluate the coverage probabilities of the simultaneous confidence bands for regression coefficients $\beta(x)$ and for $D(\mathbf{z}, \beta(x))$ for a fixed $\mathbf{z}$. The simulation setup is the same as in the first set of simulations except $c = 1$ and the sample size is 96. The 95% and 99% simultaneous confidence bands were considered. As suggested by Fan and Zhang (2000), an appropriate smaller bandwidth would improve the accuracy of the confidence bands. In our simulations, we used a shrinkage factor .6. For simplicity, we do not consider estimating the bias of $\hat{\beta}(x)$.

Table 3.1: Simulated coverage probabilities for coefficient functions based on 500 simulations for VCLE methods

	$\alpha = .05$			$\alpha = .01$		
	intercept	gender	Gage	intercept	gender	Gage
$\beta_1(x)$	.942	.938	.946	.988	.988	.988
$\beta_2(x)$	.936	.952	.956	.982	.992	.988
$\beta_3(x)$	.940	.936	.948	.990	.992	.988
$\beta_4(x)$	.946	.960	.938	.990	.998	.994
$\beta_5(x)$	.932	.950	.942	.984	.988	.990
$\beta_6(x)$	.942	.940	.946	.990	.988	.988

Table 3.1 summarizes the empirical coverage probabilities based on 500 simulations for $\alpha = .01$ and and .05. The coverage probabilities are quite close to the claimed confidence levels. Figure 3.5 depicts typical 95% simultaneous confidence bands for vectors of coefficient functions $\beta_l(x)$ for $l = 1, \cdots, 6$. For each l, $\beta_l(x)$ contains three coefficient functions which correspond the intercept, age and gestational age. It is observed that the length of the simultaneous confidence bands for the second and third functions in $\beta_1(x)$, $\beta_3(x)$ and $\beta_6(x)$ are large. This is because of large variances of corresponding parts in data.

Figure 3.6 summarizes the empirical coverage probabilities for $D(\mathbf{z}, \beta(x))$ based on 500 simulations for $\alpha = .01$ and and .05. The coverage probabilities are quite close to the claimed confidence levels.

3.4 A Real Example

We investigate early brain development by using DTI and our VCLE. We consider 96 healthy infants (36 males and 60 females) whose mean gestational age was 245.6 days with SD: 18.5 days (range: 192-270 days). A 3T Allegra head only MR system was used to acquire all the images. The system was equipped with a maximal gradient strength of 40 mT/m and a maximal slew rate

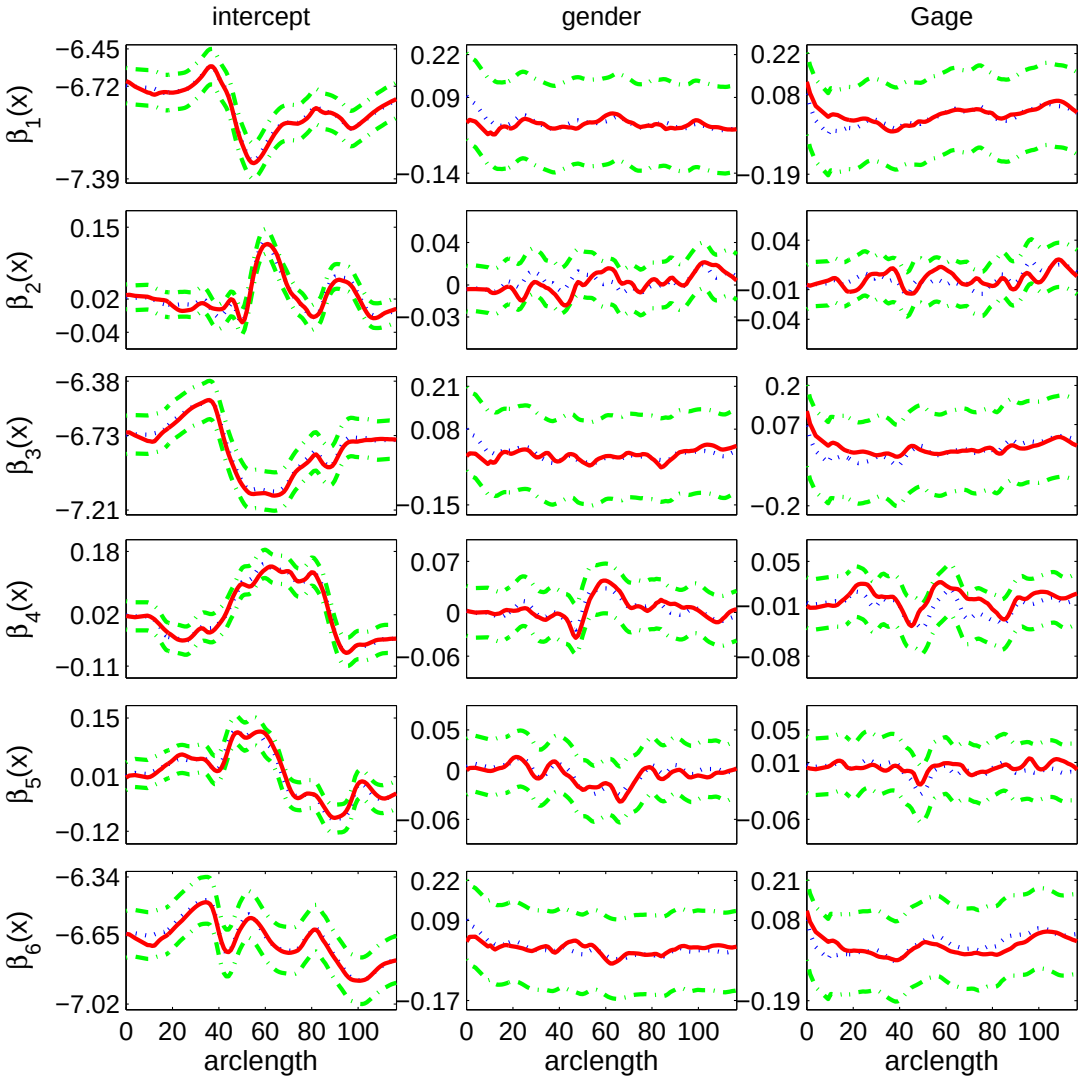

Figure 3.5: Typical 95% simultaneous confidence bands for vectors of coefficient functions $\beta_l(x)$ for $l = 1, \cdots, 6$. The solid, dotted, dash-dotted curves are the true curves, the estimated coefficient functions and the 95% confidence bands, respectively.

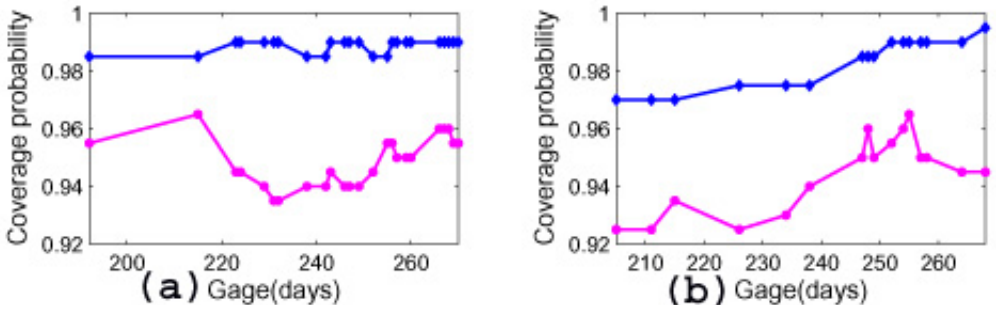

Figure 3.6: Simulated coverage probabilities for $D(\mathbf{z}, \beta(x))$ based on 500 simulations for $\alpha = .05$ (solid lines with diamond markers) and $\alpha = .01$ (solid lines with circle markers), (a)for female (b) for male at different gestational ages, respectively.

of 400 mT/(m·msec). The DTI images were obtained by using a single shot EPI DTI sequence (TR/TE=5400/73 msec) with eddy current compensation. The six non-collinear directions at the b-value of 1000 s/mm^2 with a reference scan (b=0) was applied. The voxel resolution was isotropic 2 mm, and the in-plane field of view was set as 256 mm in both directions. To improve the signal-to-noise ratio of the images, a total of five scans were acquired and averaged. A weighted least square estimation method (Zhu et al., 2007b; Basser et al., 1994a) was used to construct the diffusion tensors. Then DTI atlas building followed by atlas based tractography procedure was employed to process all 96 DTI datasets. We chose the right internal capsule tract to illustrate the applicability of our method. Diffusion tensors were extracted along this fiber tract for all the 96 infants (Goodlett et al., 2009).

In this study, we have two specific aims. The first one is to compare diffusion tensors along the selected fiber bundle across the male and female groups and thus illuminate the gender effect on the development of these fiber bundle diffusion tensors. The second one is to delineate the development of fiber bundle diffusion tensor across the gestational age effect. To statistically test the effects, we applied our VCLE to diffusion tensors along the fiber tract. For the the selected tract, we fitted the VCLE model (3.2.1) to the diffusion tensors from all 96 subjects, in which $\mathbf{z} = (1, \text{gender}, \text{Gage})^T$. Then, we used equation (3.2.10) to estimate the functional coefficients $\beta(x)$. For the hypothesis testing, we constructed the global test statistic T_n via equation (3.2.23) to test the gender and age effects for the diffusion tensors. The p value of T_n was approximated using the resampling method with $G = 2,000$ replications. Finally, we constructed the 95% simultaneous confidence bands for

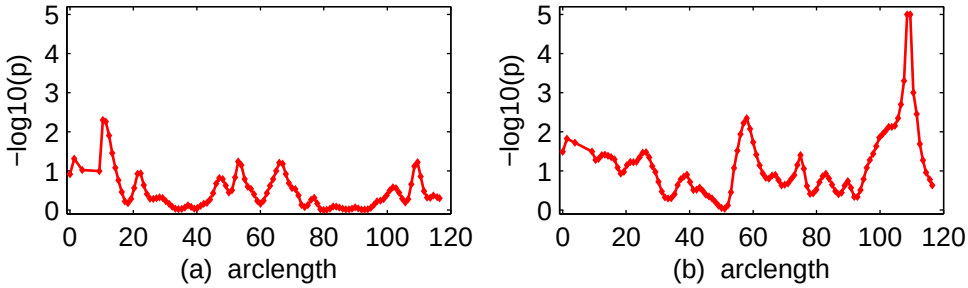

Figure 3.7: The $-\log_{10}(p)$ values of test statistics $T_n(x_j)$ for testing gender or gestational age effect of diffusion tensors on the right internal capsule tract in panel (a) showing no significant gender effect, in panel (b) showing significant gestational age effect.

the functional coefficients $\beta(x)$.

We statistically test the effects of gender and gestational age on the diffusion tensors along the right internal capsule tract. To test the gender effect, we calculated the local test statistics $T_n(x_j)$ and their corresponding p values across all grid points on the right internal capsule tract. It is observed from Figure 3.7 (a) that all grid points have not significant $-\log_{10}(p)$ values, which are less than 0.8. Then, we also computed the global test statistic $T_n = 792.70$ and its associated p-value $p = 0.335$ indicating no gender effect. It is observed from Figure 3.7 (b) that the $-\log_{10}(p)$ values of $T_n(x_j)$ for testing the gestational age effect at some grid points of the right tail are greater than 1.4 while a very high significant gestational age effect was found with $T_n = 1361.16$ and its $p-$value $p = 0.01$. This indicates that diffusion tensors along the right internal capsule tract do not differ significantly between male and female groups but are significantly associated with the gestational age. We picked a grid point with the significant p value of $T_n(x)$ and observed that the diffusion tensors become more spherical with the gestational age (Figure 3.8). This indicates the increasing pattern of diffusion.

The estimated coefficient functions along with the 95% simultaneous confidence bands are depicted in Figure 3.9. From Table 2, one can see that the gender effect, is not significant at the 5% level. In fact, from Figure 3.9, the values of coefficient functions for this variable are close to zero over a lot of parts of the whole interval. The coefficients for the gestational age are statistically significant.

Finally, the 95% critical values for $D(\mathbf{z}, \beta(x))$ and the estimated $D(\mathbf{z}, \beta(x))$ along the right internal capsule tract over gestational ages for female and male groups, respectively, are depicted

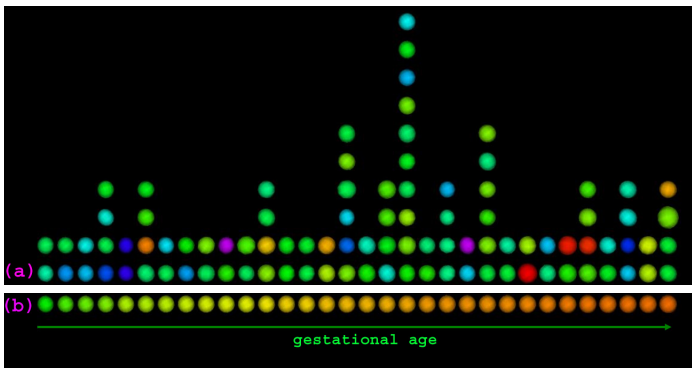

Figure 3.8: The ellipsoidal representations of (a) raw and (b) smoothed diffusion tensors changing with the gestational age at one point on the right internal capsule tract with significant gestational age effect.

in Figure 3.10. From this figure, we can see that the uncertainty of estimation of $D(\mathbf{z}, \beta(x))$ is larger on the two sides (especially on the left side) and smaller in the middle.

3.5 Discussion

We have developed the VCLE method for diffusion tensors along fiber tracts in the Riemannian manifold of SPD matrices under the Log-Euclidean metric. From the application end, VCLE is demonstrated in a clinical study of neurodevelopment for revealing the complex inhomogeneous spatiotemporal maturation patterns as the apparent changes in fiber bundle diffusion tensors.

Another commonly used metric on the Riemannian manifold of SPD matrices is the Riemannian metric. In contrast, some operations, e.g. average or interpolation of a set of tensors under the Log-Euclidean and Riemannian metrics, are theoretically and practically very similar (Arsigny (2006)). Moreover, some statistical methods based on the two metrics have very similar results. For instance, in Chapter 2, no big differences are found in using local polynomial smoothing methods based on them. However, the Riemannian metric is affine invariant. Affine invariance is a desirable feature for imaging processing, e.g. segmentation. In this scenario, as shown in Barmpoutis et al. (2007), the method using the Riemannian metric outperformed the method using the Log-Euclidean metric. So it is interesting to develop the varying coefficient method under the Riemannian metric and then compare the statistical powers of detecting group differences under these two different metrics.

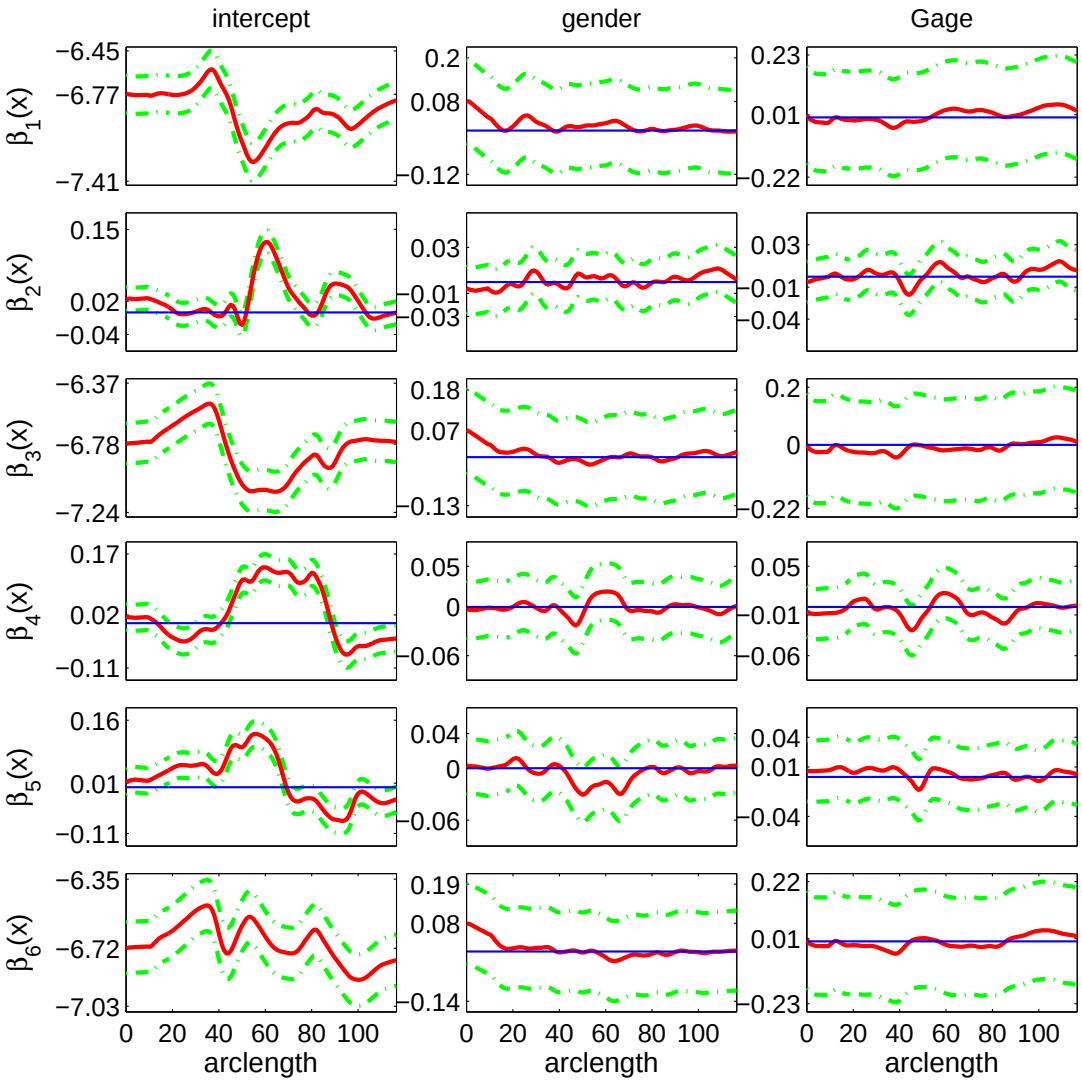

Figure 3.9: 95% simultaneous confidence bands for coefficient functions. The solid curves are the estimated coefficient functions, and the dashed curves are the 95% confidence bands. The thin horizontal line is the line crossing $(0, 0)$.

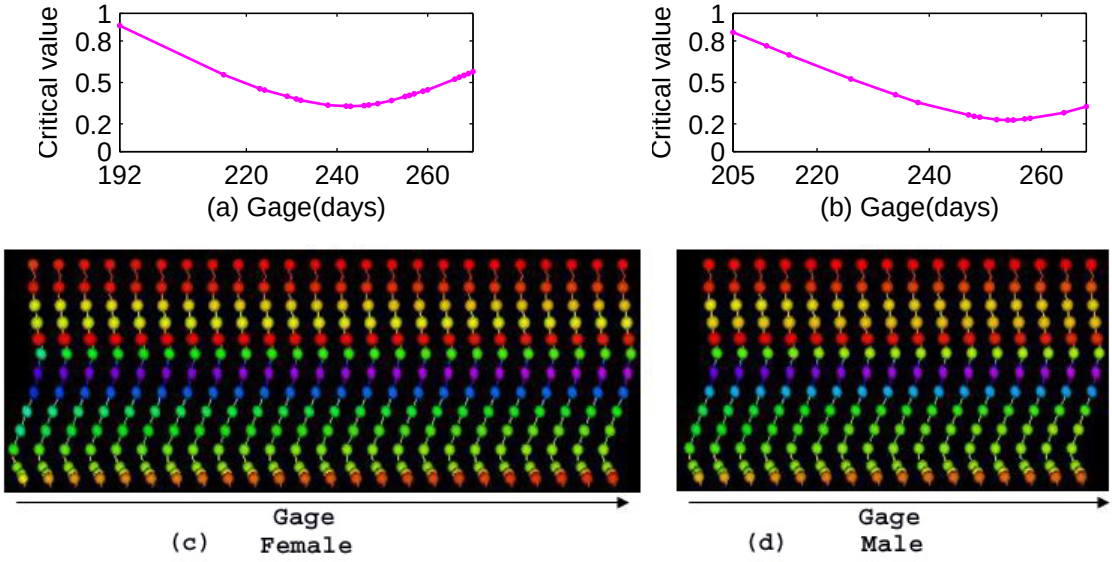

Figure 3.10: The 95% critical values for $D(\mathbf{z}, \beta(x))$ over gestational ages for female (a) and male (b) groups, respectively. The ellipsoidal representation of the estimated $D(\mathbf{z}, \beta(x))$ along the right internal capsule tract over gestational ages for female (c) and male (d) groups, respectively.

Chapter 4

Varying Coefficient Model for SPD Matrix Valued Functional Data under the Affine Invariant Metric

4.1 Introduction

SPD-matrix valued functional data appear in many applications. One of the most important applications is in medical imaging. As studied in Chapter 3, in DTI, a collection of DT's are measured along a specific white matter fiber tract (e.g. the splenium or right internal capsule tract) for each subject. Mathematically, these DT's along the fiber tract are functionals of SPD matrices. The natural approach to dealing with this type of data is functional data analysis (FDA).

The simple and straightforward way is to perform FDA based on the derived scalar quantities of the SPD matrices, e.g. eigenvalues of the matrix or FA values calculated based on the eigenvalues. However, as shown in the simulation studies of our previous chapters, this way of analysis is not appropriate because it can only account for a portion of variability in the data and risk loss of information and low statistical power in detecting group differences. At this point, a statistical analysis framework is needed to characterize the association between SPD matrix valued functions and a set of covariates, and permit to perform group comparisons and statistical inferences based on the whole SPD matrix.

The literature of recent years reflects a lot of interests in developing statistical methods in the space of SPD matrices (Schwartzman et al. (2008); Kim and Richards (2010); Zhu et al. (2009);Barmpoutis et al. (2007); Davis et al. (2010); Whitcher et al. (2007); Commowick et al. (2008)). This involves a basic but important question: defining the distance between SPD matrices since the measurement of the distance between SPD matrices is the foundation on which the subsequent statistical analysis, such as model formulation, estimation or statistical inference is

based. To define the distance, a metric is needed to given at first. So far, several metrics have been proposed for statistical analysis of SPD matrix valued data, including the Eulcidean and non-Euclidean (e.g. affine invariant and Log-Euclidean) metrics (Whitcher et al. (2007); Arsigny et al. (2006); Pennec et al. (2006)). Since more than one metric is admissible for the space of SPD matrices, selecting among them and determining which one would best characterize the distance between SPD matrices and would lead to high statistical power for hypothesis tests are challenging issues. Whitcher et al. (2007) investigated the effect of Euclidean and Log-Euclidean metrics on the tensor statistical test and their results were in favor of the Euclidean metric. The results in Barmpoutis et al. (2007) show that the affine invariant and Log-Euclidean metrics are better than the Euclidean metric in smoothing diffusion tensors using tensor spline methods. Pasternak et al. (2010) pointed out that the affine invariant metric may be the appropriate metric for quantities that are log-normally distributed and the Euclidean metric appropriate for quantities that are normally distributed. Dryden et al. (2009) made comparisons of various choices of metrics in different noise models. They observed that in some cases, the Log-Euclidean metric outperformed the Euclidean metric while in some cases, the affine invariant metric perform the best compared with the Log-Euclidean and Euclidean metrics.

We have developed a functional data analysis framework for SPD matrix data under the Log-Euclidean metric in Chapter 3. To investigate the effect of the metric on statistical analysis of SPD matrix valued functional data, we formulate the varying coefficient model under the affine invariant metric, called VCAI. Similar to Chapter 3, we use varying coefficient functions to characterize the association between fiber bundles diffusion tensor and a set of covariates. We use a weighted least squares estimation method based on the geodesic distance to estimate the varying coefficient functions. However, the geodesic distance is defined under the affine invariant metric. Compared with the Log-Euclidean metric, this metric is more complex in terms of computational and theoretical difficulties. There is no closed form formula for the estimators under the affine invariant metric. Nonlinear optimization algorithm has to be used for computing the estimators. We also establish the asymptotic properties of the estimators, which are the foundation of constructing the global test statistic and simultaneous confidence band. Furthermore, we develop a global test statistic to test hypotheses on the varying coefficient functions and use a resampling method for approximating its $p-$value. Finally, we construct a simultaneous confidence band to quantify the uncertainty in

the estimated coefficient functions and propose a resampling method to approximate the critical point.

The rest of the paper is organized as follows. Section 4.2 presents VCAI and related statistical inference. Section 4.3 examines the finite sample performance of VCAI via simulation studies. Section 4.4 illustrates an application of VCAI in a clinical study of neurodevelopment. Section 4.5 presents concluding remarks.

4.2 Methodologies

In this section, we present our VCAI for the statistical analysis of functional data with SPD matrix valued responses with a set of covariates under the affine invariant metric. This framework is similar to that in Chapter 3. The challenges in computation and theoretical derivations come from the complexity of the affine invariant metric. The details are given below.

4.2.1 Varying Coefficient Model for SPD Matrix Valued Functional Data

Consider a varying coefficient model to characterize the relationship between $p \times p$ SPD matrix-valued functional responses $S_i(x)$ under the common design and a $r \times 1$ vector of covariates of interest, denoted by $\mathbf{z}_i$, observed from n independent subjects. The model is given as follows:

$$\log[D(\mathbf{z}_i, \beta(x_j))^{-1/2} S_i(x_j) D(\mathbf{z}_i, \beta(x_j))^{-T/2}] \;=\; \mathcal{U}_i(x_j) + \mathcal{E}_i(x_j), i = 1, 2, \cdots, n, \quad (4.2.1)$$

where $\beta(x)$ is a $qr \times 1$ vector of functions of x with $q = p(p+1)/2$, $\mathcal{E}_i(x) \in \mathrm{Sym}(p)$ are measurement errors and $\mathcal{U}_i(x) \in \mathrm{Sym}(p)$ characterize individual curve variations from $D(\mathbf{z}_i, \beta(x))$.

For the sake of comparison with model (3.2.1) in Chapter 3, we use a matrix logarithm model for $D(\mathbf{z}_i, \beta(x))$. This model assumes that the matrix logarithm of $D(\mathbf{z}_i, \beta(x))$ can be denoted as

$$\log(D(\mathbf{z}_i, \beta(x))) = \log(C(\mathbf{z}_i, \beta(x))^{\otimes 2}) = Y(\mathbf{z}_i, \beta(x)). \quad (4.2.2)$$

where $Y(\mathbf{z}_i, \beta(x))$ is modeled as $\mathrm{vecs}(Y(\mathbf{z}_i, \beta(x))) = (\mathbf{z}_i^T \beta_1(x), \cdots, \mathbf{z}_i^T \beta_q(x))^T$ with $\beta_l(x) = (\beta_{1l}(x), \cdots, \beta_{rl}(x))^T$ and $C(\mathbf{z}_i, \beta(x)) = \exp(Y(\mathbf{z}_i, \beta(x))/2)$.

For notational simplicity, functional responses $S_i(x)$ are assumed to be measured at the same

n_G location points $x_1 = 0, x_2, \cdots, x_{n_G} = L_0$ for $i = 1, \cdots, n$. Let SP(μ, Σ) denote a stochastic process with mean $\mu(x)$ and covariance matrix function $\Sigma(x, y)$. We assume the stochastic process has finite support, i.e. $x \in [0, L_0], 0 < L_0 < \infty$. We also assume that vecs$(\mathcal{E}_i(x))$ and vecs$(\mathcal{U}_i(x))$ are independent with vecs$(\mathcal{E}_i(x)) \sim SP(\mathbf{0}, \Sigma_{\mathcal{E}}), \Sigma_{\mathcal{E}}(x, y) = \Sigma_{\mathcal{E}}(x)\mathbf{1}_{x=y}$ and vecs$(\mathcal{U}_i(x)) \sim SP(\mathbf{0}, \Sigma_{\mathcal{U}})$. Moreover, vecs$(\mathcal{E}_i(x))$ and vecs$(\mathcal{E}_i(x'))$ for $x \neq x'$ are assumed to be independent and thus $\Sigma_{\mathcal{E}}(x, x')$ takes the form of $\Sigma_{\mathcal{E}}(x)\mathbf{1}_{x=x'}$.

Like model (3.2.1) in Chapter 3, model (4.2.1) can also be regarded as a generalization of varying coefficient models, which have been widely studied and developed for longitudinal, time series, and functional data (Fan and Zhang, 1999; Wu and Chiang, 2000; Fan et al., 2003; Fan and Zhang, 2008; Wang et al., 2008). Models (3.2.1) and (4.2.1) are very similar. Both use a matrix logarithm model for $D(\mathbf{z}, \beta(x))$. Both models assume the same error distributions. The big difference is that the residual matrix of the functional response $S(x)$ at the population mean function $D(\mathbf{z}, \beta(x))$ in model (4.2.1) is defined based on the Riemannian logarithm map under the affine invariant metric while in model (3.2.1), it is based on the Riemannian logarithm map under the Log-Euclidean metric (see Section 2.2, 2.3 and 2.4 in Chapter 2 for the formulation of a residual matrix of $S(x)$ at $D(\mathbf{z}, \beta(x))$) .

Like the classical varying coefficient model for functional data with responses on the Euclidean space (see Fan and Zhang (2008) for details), for model (4.2.1), we focus on carrying out statistical inferences on $\beta(x)$, including constructing simultaneous confidence bands for $\beta(x)$ and developing global test statistics for general hypothesis testing problem on $\beta(x)$.

4.2.2 Weighted Least Squares Estimation

To estimate the coefficient function in $\beta(x)$, we have developed a weighted least squares estimation method based on an adaptive local polynomial kernel (LPK) smoothing technique under the affine invariant metric. Specifically, using Taylor's expansion, we can expand $\beta(x)$ at x to obtain

$$\beta(x_j) \approx \beta(x) + \dot{\beta}(x)(x_j - x) = A(x)\mathbf{y}_{h^{(1)}}(x_j - x), \tag{4.2.3}$$

where $\mathbf{y}_{h^{(1)}}(x_j - x) = (1, (x_j - x)/h^{(1)})^T$ and $A(x) = (\beta(x), h^{(1)}\dot{\beta}(x))$ is an $rq \times 2$ matrix, in which $\dot{\beta}(x)$ are the derivatives of the vector $\beta(x)$ with respect to x, respectively. For each fixed x and

each bandwidth $h^{(1)}$, a weighted least square estimator of $A(x)$, can be calculated by minimizing an objective function given by

$$\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathrm{tr}[\{\log(C(\mathbf{z}_i, A(x)\mathbf{y}_{h^{(1)}}(x_j - x))^{-1} S_i(x_j) C(\mathbf{z}_i, A(x)\mathbf{y}_{h^{(1)}}(x_j - x))^{-T})\}^{\otimes 2}] \quad (4.2.4)$$

where $K(.)$ be a kernel function and $K_h(.) = K(./h)/h$ is a rescaled kernel function. Since the distance in the objective function (4.2.4) is the geodesic distance under the affine invariant metric, the weight least square estimate of $A(x)$, denoted as $\hat{A}(x)$, does not appear to have the explicit expression. This minimization problem is computationally challenging since the standard gradient methods do not perform well in this scenario. Hence, we will develop an annealing evolutionary stochastic approximation Monte Carlo algorithm (Liang (2010)) for calculating the estimate. After we have the estimate of $\mathrm{vec}(\hat{A}(x)) = (\hat{\beta}_{11}, h^{(1)}\hat{\dot{\beta}}_{11}, \cdots, \hat{\beta}_{r1}, h^{(1)}\hat{\dot{\beta}}_{r1}, \cdots, \hat{\beta}_{1q}, h^{(1)}\hat{\dot{\beta}}_{1q}, \cdots, \hat{\beta}_{rq}, h^{(1)}\hat{\dot{\beta}}_{rq})^T$, we can obtain the estimate of the coefficient function $\beta(x)$

$$\hat{\beta}(x) = \{I_{qr} \otimes (1, 0)\}\mathrm{vec}(\hat{A}(x)). \quad (4.2.5)$$

We now investigate the asymptotic properties of the coefficient functions $\hat{\beta}(x)$, whose assumptions and detailed proofs can be found in the Appendix E. To establish the asymptotic properties of $\hat{\beta}(x)$. We need the following notations. Let $\psi(S, C) = \mathrm{tr}[\{\log(C^{-1}SC^{-T})\}^{\otimes 2}]$, be a function of $C = \exp(Y/2)$ and $S \in \mathrm{Sym}^+(m)$, $\partial_Y \psi(S, C)$ and $\partial_Y^2 \psi(S, C)$ be the first and second derivatives of $\psi(S, C)$ with respect to Y, respectively. Let Λ be a $q \times q$ diagonal matrix with q diagonal entries $1, 2, 1, 2, 2, 1 \cdots, 2, \cdots, 2, 1$ and Y_k denote the k-th element of $\mathrm{vecs}(Y)$. Let $\mathcal{C}^{(k)}(x, \mathbf{z}) = -C(\mathbf{z}, \beta(x))^{-1}\dfrac{\partial CC^T}{\partial Y_k}C(\mathbf{z}, \beta(x))^{-T}$ and $M^{(k,k')}(v, u, \mathbf{z}) = \Lambda\mathrm{vecs}(\mathcal{C}^{(k)}(v, \mathbf{z}))(\Lambda\mathrm{vecs}(\mathcal{C}^{(k')}(u, \mathbf{z})))^T$. Let $M_{\mathcal{E}}(v, u, \mathbf{z})$ and $M_{\mathcal{U}}(v, u, \mathbf{z})$ be $q \times q$ matrices with the (k, k') entry $\mathrm{tr}\{\Sigma_{\mathcal{E}}(v, u)M^{(k,k')}(v, u), \mathbf{z})\}$ and $\mathrm{tr}\{\Sigma_{\mathcal{U}}(v, u)M^{(k,k')}(v, u, \mathbf{z})\}$, respectively. Let $\Omega_{\mathcal{E}}(v, u) = E\{M_{\mathcal{E}}(v, u, \mathbf{z}) \otimes \mathbf{z}^{\otimes 2}\}$ and $\Omega_{\mathcal{U}}(v, u) = E\{M_{\mathcal{U}}(v, u, \mathbf{z}) \otimes \mathbf{z}^{\otimes 2}\}$. Let $\mathcal{X}$ be denoted as $\{x_1, \cdots, x_{n_G}\}$.

Theorem 4.2.1. *If the assumptions (M1)-(M9) are true in Appendix E, then the following results hold:*

(i) when the assumption (M10) in Appendix E is true, $\sqrt{n}\{\hat{\beta}(x) - \beta(x) - 0.5u_2\ddot{\beta}(x)h^{(1)2}\{1 + o(1)\}) : x \in [0, L_0]\}$ *converges weakly to a centered Gaussian process with covariance matrix* $\Sigma_\beta(x, x') =$

$4[E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}]^{-1}\Omega_{\mathcal{U}}(x,x')\left[E\{\partial_Y^2\psi(S(x'),C(\mathbf{z},\beta(x')))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-T}$.

(ii) the asymptotic bias and covariance of $\hat{\beta}(x)$ *given* $\mathcal{X}=\{x_j, j=1,\cdots,n_G\}$ *are, respectively,* $0.5u_2\ddot{\beta}(x)h^{(1)2}[1+o(1)]$ *and*

$$\begin{aligned}
&\Sigma_{\beta,\mathcal{X}}(x,h) = 4\left[E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-1}(n^{-1}e_n(x)+\\
&(nn_Gh^{(1)})^{-1}\pi(x)^{-1}v_0\{\Omega_{\mathcal{U}}(x,x)+\Omega_{\mathcal{E}}(x,x)\}\{1+O(h^{(1)})\}+\\
&n^{-1}[\Omega_{\mathcal{U}}(x,x)+h^{(1)2}u_2\{\Omega_{\mathcal{U}}^{(2,0)}(x,x)\pi(x)+2\Omega_{\mathcal{U}}^{(1,0)}(x,x)\dot{\pi}(x)+\Omega_{\mathcal{U}}(x,x)\ddot{\pi}(x)\}\pi(x)^{-1}\\
&+O_p(n_G^{-1})+o_p(h^{(1)2})]\left[E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-T}, \qquad (4.2.6)
\end{aligned}$$

where $e_n(x)=O_p((n_Gh^{(1)})^{-1/2})$ *is a random matrix of* $\mathcal{X}$ *and* $\Omega_{\mathcal{U}}(x,x)$ *with* $E\{e_n(x)\}=0$ *and is defined in the appendix.*

Theorem 4.2.1 establishes weak convergence of $\hat{\beta}(x)$ as a stochastic process indexed by $x\in[0,L_0]$. Based on Theorem 4.2.1, it is straightforward to derive the asymptotic bias and variance of $\log(D(\mathbf{z},\beta(x)))$, which are combined together to give the asymptotic mean square error (AMSE) of $\log(D(\mathbf{z},\beta(x)))$ as

$$\begin{aligned}
\text{AMSE}(\log(D(\mathbf{z},\hat{\beta}(x)))) &= E[\text{tr}\{\text{Ivecs}(Y(\mathbf{z},\hat{\beta}(x))-Y(\mathbf{z},\beta(x)))^{\otimes 2}\}]\\
&= 0.25u_2^2h^{(1)4}\text{tr}[\{(\Lambda^{1/2}\otimes\mathbf{z})\ddot{\beta}(x)\}^{\otimes 2}]+\text{tr}\{(\Lambda^{1/2}\otimes\mathbf{z})\Sigma_{\beta,\mathcal{X}}(x,h^{(1)})(\Lambda^{1/2}\otimes\mathbf{z})^T\}.
\end{aligned}$$

Note that the bandwidth that minimizes AMSE$(\log(D(\mathbf{z},\beta(x))))$ is of the order $(nn_G)^{-1/5}$ if the terms $n^{-1}h^{(1)2}$ and $n^{-1}e_n$ in Equation (4.2.6) are ignored. Theorem 4.2.1 also forms the foundation for constructing global test statistic and simultaneous confidence bands for $\beta(\cdot)$.

A straightforward bandwidth selection method is cross-validation. The idea is to pool the data from all n subjects and select an estimated bandwidth of $h^{(1)}$, by minimizing the cross-validation score given by

$$CV(h^{(1)})=\frac{1}{nn_G}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\text{tr}[\{\log(C(\mathbf{z}_i,\hat{\beta}^{(-i)}(x_j,h^{(1)}))^{-1}S_i(x_j)C(\mathbf{z}_i,\hat{\beta}^{(-i)}(x_j,h^{(1)}))^{-T})\}^{\otimes 2}] \quad (4.2.7)$$

where $\hat{\beta}^{(-i)}(x,h^{(1)})$ is the weighted least squares estimator of $\beta(x)$ for the bandwidth $h^{(1)}$ based on observed data with the observations from the i-th subject excluded. However, since computing

$\hat{\beta}^{(-i)}(x_j, h^{(1)})$ for all i can be computationally prohibitive, we use plug-in method (Park and Marron, 1990) which is based on the result in Theorem 4.2.1. For a given weight function $w(x)$, the asymptotic mean integrated squared error (AMISE) of $D(\mathbf{z}, \beta(x))$ is

$$\begin{aligned}
\text{AMISE}(\log(D(\mathbf{z}, \hat{\beta}(x)))) &= \int \text{AMSE}(\log(D(\mathbf{z}, \hat{\beta}(x))))w(x)dx \\
&= 0.25u_2^2 h^{(1)4} \int \text{tr}[\{(\Lambda^{1/2} \otimes \mathbf{z})\ddot{\beta}(x)\}^{\otimes 2}]w(x)dx \\
&\quad + \int \text{tr}\{(\Lambda^{1/2} \otimes \mathbf{z})\Sigma_{\beta,\mathcal{X}}(x, h^{(1)})(\Lambda^{1/2} \otimes \mathbf{z})^T\}w(x)dx. \quad (4.2.8)
\end{aligned}$$

So with respect to this criterion, the optimal bandwidth is the one minimizing AMISE$(\log(D(\mathbf{z}, \hat{\beta}(x))))$, denoted by $\hat{h}_e^{(1)}$. If the terms $n^{-1}h^{(1)2}$ and $(n)^{-1}e_n$ are ignored, then we have

$$\hat{h}_e^{(1)} = \left(\frac{(nn_G)^{-1}v_0 \int \text{tr}\{(\Lambda^{1/2} \otimes \mathbf{z})\Sigma_{\beta,0}(x,x)(\Lambda^{1/2} \otimes \mathbf{z})^T\}\pi(x)^{-1}w(x)dx}{0.25u_2^2 \int \text{tr}[\{(\Lambda^{1/2} \otimes \mathbf{z})\ddot{\beta}(x)\}^{\otimes 2}]w(x)dx} \right)^{1/5}, \quad (4.2.9)$$

where $\Sigma_{\beta,0}(x,x) = \left(\left[E\{\partial_Y^2 \psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]^{-1} \{\Omega_{\mathcal{U}}(x,x) + \Omega_{\mathcal{E}}(x,x)\}^{1/2}\right)^{\otimes 2}$. The plug-in bandwidth selection idea is to replace the unknown quantities in Equation 4.2.9 by their estimators. For unknown quantity $\beta(x)$ and $\ddot{\beta}(x)$, we replace them by a parametric polynomial fit of degree 4. We take $w(x) = \pi(x)w_0(x)$, where $w_0(x)$ is a given weight function. In this way the problem of $\pi(x)$ sorts itself out. Based on $\hat{h}_e^{(1)}$, we can obtain an estimate of $\beta(x)$, denoted by $\hat{\beta}_e(x)$. This bandwidth selection method may tend to oversmoothing since we ignore two terms in Equation (E.0.36), which are introduced by the within-curve dependence. So in practice, we select the bandwidth slightly smaller than this optimal bandwidth.

4.2.3 Smoothing Individual Functions and Estimating Covariance Matrices

As we did in Chapter 3, we also employ the local polynomial kernel smoothing technique to simultaneously construct the individual function $\mathcal{U}_i(x)$ and then construct a nonparametric estimator of the covariance matrix $\Sigma_{\mathcal{E}}(x,x)$. The computational procedures are similar to that in Section 3.2.3 of Chapter 3. The only difference comes from modeling the residual matrix of the functional response $S(x)$ at the population mean function $D(\mathbf{z}, \beta(x))$. In the following, we will give

detailed derivations. Using Taylor's expansion, we can expand $\mathcal{U}_i(x_j)$ at x to obtain

$$\mathcal{U}_i(x_j) \approx \mathcal{U}_i(x) + \dot{\mathcal{U}}_i(x)(x_j - x), \tag{4.2.10}$$

which is equivalent to

$$\text{vecs}(\mathcal{U}_i(x_j)) \approx \text{vecs}(\mathcal{U}_i(x)) + \text{vecs}(\dot{\mathcal{U}}_i(x))(x_j - x) = \mathcal{D}_i(x)^T \mathbf{y}_{h^{(2)}}(x_j - x), \tag{4.2.11}$$

where $\mathbf{y}_{h^{(2)}}(x_j - x) = (1, (x_j - x)/h^{(2)})^T$ and $\mathcal{D}_i(x) = (\text{vec}(\mathcal{U}_i(x)), h^{(2)}\text{vec}(\dot{\mathcal{U}}_i(x)))^T$ is a $2 \times q$ matrix. For each fixed x and each bandwidth $h^{(2)}$, the weighted least square estimator of $\mathcal{D}_i(x)$ can be calculated by minimizing an objective function given by

$$\sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\text{tr}[\{\log(C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-T}) - \text{Ivecs}(\mathcal{D}_i(x)^T \mathbf{y}_{h^{(2)}}(x_j - x))\}^{\otimes 2}].$$

With some calculation, it can be shown that the weighted least square estimate of $\mathcal{D}_i(x)$, denoted as $\hat{\mathcal{D}}_i(x)$, has the following explicit expression:

$$\Sigma_1(h^{(2)}, x)^{-1} \sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)\{\text{vecs}(\log(C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-T}))\}^T, \tag{4.2.12}$$

where $\Sigma_1(h^{(2)}, x) = \sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)\mathbf{y}_{h^{(2)}}(x_j - x)^{\otimes 2}$. Then, $\text{vecs}(\mathcal{U}_i(x))$ can be estimated by

$$\text{vecs}(\hat{\mathcal{U}}_i(x)) = \mathbf{e}_{1,2}^T \hat{\mathcal{D}}_i(x) = \sum_{j=1}^{n_G} \tilde{K}^0_{n_G, h^{(2)}}(x_j - x, x)\{\text{vecs}(\log(C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-T}))\}, \tag{4.2.13}$$

where $\tilde{K}^0_{n_G, h^{(2)}}(.,.)$ is the empirical equivalent kernel. Finally, let R_i be a $n_G \times q$ matrix with the j-th row $\text{vecs}(\log(C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-T}))^T$ and $\mathcal{S}$ be a $n_G \times n_G$ smoothing matrix with the (i, j)-th element $\tilde{K}^0_{n_G, h^{(2)}}(x_j - x_i, x_i)$. We can obtain

$$(\text{vecs}(\hat{\mathcal{U}}_i(x_1)), \cdots, \text{vecs}(\hat{\mathcal{U}}_i(x_{n_G})))^T = \mathcal{S} R_i. \tag{4.2.14}$$

We pool the data from all n subjects and select an estimated bandwidth of $h^{(2)}$, denoted as

$\hat{h}_e^{(2)}$, by minimizing the cross-validation score given by

$$CV_2(h^{(2)}) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\frac{\|\Lambda^{1/2}((R_i)_j - (\mathcal{S}R_i)_j)\|^2}{(1-\mathcal{S}_{jj})^2}. \tag{4.2.15}$$

Replacing $\mathcal{S}_{jj}$ in equation (4.2.15) by the average of $\mathcal{S}_{11},\cdots,\mathcal{S}_{n_Gn_G}$, we can get the following generalized cross-validation (GCV) score

$$GCV_2(h^{(2)}) = n^{-1}\frac{\sum_{i=1}^{n}\mathrm{tr}[\{\Lambda^{1/2}(R_i - \mathcal{S}R_i)\}^{\otimes 2}]}{(1-n^{-1}\mathrm{tr}(\mathcal{S}))^2}. \tag{4.2.16}$$

Without special saying, for this part, we use generalized cross-validation score $GCV_2(h^{(2)})$ to select the bandwidth throughout this chapter.

After obtaining $\hat{\mathcal{U}}_{i,e}(x)$, we can estimate the mean function $\mathcal{U}(x)$ and the covariance function $\Sigma_{\mathcal{U}}(x,y)$ by using their empirical counterparts based on the estimated $\hat{\mathcal{U}}_{i,e}(x)$:

$$\begin{aligned}\hat{\mathcal{U}}_e(x) &= n^{-1}\sum_{i=1}^{n}\hat{\mathcal{U}}_{i,e}(x) \quad \text{and} \\ \hat{\Sigma}_{\mathcal{U}}(x,y) &= (n-q)^{-1}\sum_{i=1}^{n}\mathrm{vecs}(\hat{\mathcal{U}}_{i,e}(x))\mathrm{vecs}(\hat{\mathcal{U}}_{i,e}(y)).\end{aligned} \tag{4.2.17}$$

We construct a nonparametric estimator of the covariance matrix $\Sigma_{\mathcal{E}}(x,x)$ as follows. Let $\hat{\beta}_e(x) = [\hat{\beta}_{1,e}(x),\cdots,\hat{\beta}_{q,e}(x)]$ and $\hat{\mathcal{E}}_i(x_j) = \log(C(\mathbf{z}_i,\hat{\beta}_e(x_j))^{-1}S_i(x_j)C(\mathbf{z}_i,\hat{\beta}_e(x_j))^{-T}) - \hat{\mathcal{U}}_{i,e}(x_j)$ be estimated residuals for $i=1,\cdots,n$ and $j=1,\cdots,n_G$. We consider the kernel estimate of $\Sigma_{\mathcal{E}}(x,x)$ given by

$$\hat{\Sigma}_{\mathcal{E}}(x,x) = (n-q)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\frac{K_{h^{(3)}}(x_j-x)\mathrm{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}}{\sum_{j=1}^{n_G}K_{h^{(3)}}(x_j-x)}. \tag{4.2.18}$$

Let $\tilde{\Sigma}_{\mathcal{E}}(x_j,x_j) = (n-q)^{-1}\sum_{i=1}^{n}\mathrm{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2}$. To select an estimated bandwidth $h^{(3)}$, denoted by $\hat{h}_e^{(3)}$, we minimize the cross-validation score given by

$$CV_3(h^{(3)}) = (nn_G)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\mathrm{tr}\{[\mathrm{vecs}(\hat{\mathcal{E}}_i(x_j))^{\otimes 2} - \hat{\Sigma}_{\mathcal{E}}(x_j,x_j,h^{(3)})^{(-i)}]^{\otimes 2}\tilde{\Sigma}_{\mathcal{E}}(x_j,x_j)^{-1}]^{\otimes 2}\}, \tag{4.2.19}$$

where $\hat{\Sigma}_{\mathcal{E}}(x,x,h^{(3)})^{(-i)}$ is the weighted least squares estimator of $\hat{\Sigma}_{\mathcal{E}}(x,x)$ based on observed data with the observations from the i-th subject excluded. Based on $\hat{h}_e^{(3)}$, we can use (4.2.18) to estimate

$\Sigma_{\mathcal{E}}(x,x)$, denoted by $\hat{\Sigma}_{\mathcal{E},e}(x,x)$.

We will use the following theorems to construct the global test statistics and simultaneous confidence bands for coefficient functions, whose detailed proofs can be found in the Appendix E.

Theorem 4.2.2. *We have the following results.*

(i) If assumptions (M1)-(M6) and (M10) in Appendix E are true, then

$$\sup_{(x,t)\in[0,L_0]^2} |\hat{\Sigma}_{\mathcal{U}}(x,t) - \Sigma_{\mathcal{U}}(x,t)| = O_p(n^{-1/2} + (n_G h^{(2)})^{-1} + \hat{h}_e^{(1)2} + h^{(2)2} + (\log n/n)^{1/2}). \quad (4.2.20)$$

(ii) If assumptions (M1)-(M6) and (M10)-(M11) in Appendix E are true, then $\sup_{x\in[0,L_0]} |\hat{\Sigma}_{\mathcal{E}}(x,x) - \Sigma_{\mathcal{E}}(x,x)| = o_p(1)$.

Theorem 4.2.2 (i) shows the uniform convergence rate of $\hat{\Sigma}_{\mathcal{U}}(x,t)$ which is useful for constructing the global and local test statistics for testing the covariate effects. Theorem 4.2.2 (ii) do not present the uniform convergence rate of $\hat{\Sigma}_{\mathcal{E}}(x,x)$ because one focus of this Chapter is to derive the asymptotic distribution of a global test statistic, which does not involve $\hat{\Sigma}_{\mathcal{E}}(x,x)$.

4.2.4 Hypothesis Test

Like the classical varying coefficient models, it is interesting to investigate if an estimated coefficient function is significantly away from zero or if the estimated coefficient function is really varying (Fan and Zhang, 2000, 2008). In general, consider the linear hypotheses of $\beta(x)$ as follows:

$$H_0 : \mathbf{R}\beta(x) = \mathbf{b}_0(x) \ \text{ for all } x \ \text{ vs. } \quad H_1 : \mathbf{R}\beta(x) \neq \mathbf{b}_0(x), \quad (4.2.21)$$

where $\mathbf{R}$ is a $t \times rq$ matrix of full row rank and $\mathbf{b}_0(x)$ is a given $t \times 1$ vector of functions. In order to test the null hypothesis $H_0 : \mathbf{R}(\beta(x)) = \mathbf{b}_0(x)$ for all x, we need to construct a global test statistic T_n. We can construct it in the same manner as we did in Section 3.2.5 of Chapter 3. However, the two test statistics are not the same since they depend on the asymptotic covariance of the estimated coefficient functions. Under the affine invariant metric, the global test statistic is defined by

$$\mathbf{T}_n = n \int_0^{L_0} \mathbf{d}(x)^T \{R\Sigma_\beta(x,x)R^T\}^{-1}\mathbf{d}(x). \quad (4.2.22)$$

where $\mathbf{d}(x) = \mathbf{R}\text{vec}(\hat{\beta}_e(x) - \text{bias}(\hat{\beta}_e(x))) - \mathbf{b}_0(x)$. Using the same arguments as those in Section 3.2.5 of Chapter 3, we can show that

$$T_n \Rightarrow \int_0^{L_0} X_{\mathbf{R}}(x)^T X_{\mathbf{R}}(x) dx,$$

where $X_{\mathbf{R}}(\cdot)$ is a Gaussian process with zero mean and covariance structure $\Sigma_{\mathbf{R}}(x, x^{'})$, which is given by

$$\Sigma_{\mathbf{R}}(x, x^{'}) = \{\mathbf{R}\hat{\Sigma}_\beta(x,x)\mathbf{R}^T\}^{-1/2}\{\mathbf{R}\hat{\Sigma}_\beta(x,x^{'})\mathbf{R}^T\}\{\mathbf{R}\hat{\Sigma}_\beta(x^{'},x^{'})\mathbf{R}^T\}^{-T/2}. \tag{4.2.23}$$

Based on this result, we develop a wild bootstrap method to approximate the p-value of T_n.

Step (1): Fit model (4.2.1) under the null hypothesis H_0, which yields $\hat{\beta}_e^*(x_j)$, $\hat{\mathcal{U}}_{i,e}^*(x_j)$ and $\hat{\mathcal{E}}_{i,e}^*(x_j)$ for $i = 1, \cdots, n$ and $j = 1, \cdots, n_G$.

Step (2): For $g = 1, \cdots, G$, generate a random sample $\tau_i^{(g)}$ and $\tau_i(x_j)^{(g)}$ from a $N(0,1)$ generator for $i = 1, \cdots, n$ and $j = 1, \cdots, n_G$ and then construct

$$\hat{S}_i(x_j)^{(g)} = C(\mathbf{z}_i, \hat{\beta}_e^*(x_j)) \exp(\tau_i^{(g)}\hat{\mathcal{U}}_i^*(x_j) + \tau_i(s_j)^{(g)}\hat{\mathcal{E}}_i^*(x_j)) C(\mathbf{z}_i, \hat{\beta}_e^*(x_j))^T.$$

Then, based on $\hat{S}_i(x_j)^{(g)}$, we recalculate $\tilde{h}_e^{(1)}$, $\hat{\beta}_e(x)^{(g)}$, $\text{bias}(\hat{\beta}_e(x)^{(g)})$, and $\mathbf{d}(x)^{(g)} = \mathbf{R}(\hat{\beta}_e(x)^{(g)} - \text{bias}(\hat{\beta}_e(x)^{(g)})) - \mathbf{b}_0(x)$. We also note that $\mathbf{R}(\hat{\beta}_e(x)^{(g)}) \approx \mathbf{b}_0$ and $\mathbf{R}(\text{bias}(\hat{\beta}_e(x)^{(g)})) \approx \mathbf{0}$. Thus, we can drop the term $\text{bias}(\hat{\beta}_e(x)^{(g)})$ in $\mathbf{d}(x)^{(g)}$ for computational efficiency. Subsequently, we compute

$$\begin{aligned} \mathbf{T}_n^{(g)} &= n\int_0^{L_0} \mathbf{d}(x)^{(g)T}\{R\Sigma_\beta(x,x)\mathbf{R}^T\}^{-1}\mathbf{d}(x)^{(g)}dx, \\ \mathbf{T}_n(x_j)^{(g)} &= n\mathbf{d}(x_j)^{(g)T}\{\mathbf{R}\Sigma_\beta(x,x)\mathbf{R}^T\}^{-1}\mathbf{d}(s_j)^{(g)} \quad \text{for} \quad j = 1, \cdots, n_G. \end{aligned} \tag{4.2.24}$$

Step (3): Aggregate the results of Step (2) to obtain $\{T_{n,\max}^{(g)} = \max_{1\le j\le n_G} T_n(x_j)^{(g)} : g = 1, \cdots, G\}$ and calculate $p(x_j) = G^{-1}\sum_{g=1}^{G} 1(T_{n,\max}^{(g)} \ge T_n(x_j))$ for each x_j. The $p(x_j)$ is the corrected p-value at the location x_j.

Step (4):Aggregate the results of Step (2) to obtain $\{T_n^{(g)} : g = 1, \cdots, G\}$ and calculate $p = G^{-1}\sum_{g=1}^{G} 1(T_n^{(g)} \ge T_n)$. If p is smaller than a pre-specified significance level α, say 0.05, then we reject the null hypothesis H_0.

4.2.5 Confidence Band

We construct confidence bands for the functional coefficients $\beta(x) = (\beta_1(x)^T, \cdots, \beta_q(x)^T)^T$ with $\beta_l(x) = (\beta_{1l}, \cdots, \beta_{rl})^T$ in model (4.3.1). Specifically, for a given significance level α, we construct a confidence band for each $\beta_{kl}(x)$ such that

$$P(\hat{\beta}_{kl}^{L,\alpha}(x) < \beta_{kl}(x) < \hat{\beta}_{kl}^{U,\alpha}(x) \text{ for all } x \in [0, L_0]) = 1 - \alpha, \tag{4.2.25}$$

where $\hat{\beta}_{kl}^{L,\alpha}(x)$ and $\hat{\beta}_{kl}^{U,\alpha}(x)$ are the lower and upper limits of the confidence band. It follows from Theorem 4.2.1 that $\sup_{x\in[0,L_0]} |\sqrt{n}[\hat{\beta}_{kl,e}(x) - \beta_{kl}(x)]|$ converges in distribution to $\sup_{x\in[0,L_0]} |G_{kl}(x)|$. We define the critical point $C_{kl}(\alpha)$ such that $P(\sup_{x\in[0,L_0]} |G_{kl}(x)| \leq C_{kl}(\alpha)) = 1 - \alpha$. Thus, a $1 - \alpha$ simultaneous confidence band for $\beta_{kl}(x)$ is given as follows:

$$\left(\hat{\beta}_{kl,e}(x) - \frac{C_{kl}(\alpha)}{\sqrt{n}}, \quad \hat{\beta}_{kl,e}(x) + \frac{C_{kl}(\alpha)}{\sqrt{n}} \right). \tag{4.2.26}$$

Recall that $\mathbf{e}_{kl}$ be an $qr \times 1$ vector with the $(l-1)r + k$-th element 1 and 0 otherwise. For $g = 1, \cdots, G$, we independently simulate $\{\tau_i^{(g)} : i = 1, \cdots, n\}$ from $N(0, 1)$ for all l, and then we calculate a stochastic process $G_l(x)^{(g)}$, which is defined as follows:

$$\begin{aligned} &\sqrt{n}\{(1,0) \otimes I_{qr}\} \left[\sum_{i=1}^{n} \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) \left\{ \mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2} \otimes \partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \hat{\beta}(x_j)))^T \otimes \mathbf{z}_i^{\otimes 2} \right\} \right]^{-1} \\ &\sum_{i=1}^{n} \tau_i^g \sum_{j=1}^{n_G} H_{h^{(1)}}(x_j - x) \otimes \{\partial_Y \psi(S_i(x_j), C(\mathbf{z}_i, \hat{\beta}(x_j))) \otimes \mathbf{z}_i\}, \end{aligned} \tag{4.2.27}$$

where $H_{h^{(1)}}(u - x) = K_{h^{(1)}}(u - x)\mathbf{y}_{h^{(1)}}(u - x)$. The critical point $C_{kl}(\alpha)$ can be calculated as $1 - \alpha$ empirical percentile of $\sup_{x\in[0,L_0]} |\mathbf{e}_{kl} G_l(x)^{(g)}|$ for all g.

Theorem 4.2.3. *If assumptions (M1)-(M6) and (M9)-(M11) in Appendix E are true, then $G_l^{(g)}(\cdot)$ converges weakly to $G_l(\cdot)$ conditional on the data.*

4.3 Simulation Studies

We conducted three sets of Monte Carlo simulations. The first set was to examine the bias and variance of the estimated coefficient functions using varying coefficient model under the affine invariant and Log-Euclidean metrics. The second set was to compare the Type I and II error rates of the global test statistic T_n using varying coefficient model under the affine invariant and Log-Euclidean metrics. The third set was to evaluate the coverage probabilities of the simultaneous confidence bands of the functional coefficients $\beta(x)$ under the affine invariant metric.

We simulate 3×3 SPD matrices (DT's) along the right internal capsule tract (Figure 3.1 (a)) as follows:

$$\begin{aligned} S_i(x) &= \exp(\text{Ivecs}((\beta_{1l}(x) + \beta_{2l}(x) \times \text{G}_i + \beta_{3l}(x) \times \text{Gage}_i)/2)) \exp(\mathcal{U}_i(x) + \mathcal{E}_i(x)) \\ &\quad \exp(\text{Ivecs}((\beta_{1l}(x) + \beta_{2l}(x) \times \text{G}_i + \beta_{3l}(x) \times \text{Gage}_i)/2)) \end{aligned} \tag{4.3.1}$$

where Gage_i, G_i, $\text{vecs}(\mathcal{U}_i(x))$ and $\text{vecs}(\mathcal{E}_i(x))$ are the same as those in Section 3.3 of Chapter 3. To mimic imaging data, we also used the diffusion tensors along the right internal capsule tract from all 96 infants in our clinical data to estimate $\beta(x)$ by $\hat{\beta}(x)$ via solving the nonlinear optimization problem in Equation (4.2.4), $\mathcal{U}(x)$ by $\hat{\mathcal{U}}(x)$ via Equation (4.2.14), and $\mathcal{E}(x)$ by $\hat{\mathcal{E}}(x)$ via $\hat{\mathcal{E}}(x) = \log(\exp(-\text{Ivecs}((\hat{\beta}_l(x)^T\mathbf{z}/2)))S(x)\exp(-\text{Ivecs}((\hat{\beta}_l(x)^T\mathbf{z}/2))) - \hat{\mathcal{U}}(x)$. We fixed all the parameters at their values obtained from our clinical data, except that we assumed $(\beta_{31}(x), \cdots, \beta_{36}(x)) = c(\hat{\beta}_{31}(x), \cdots, \hat{\beta}_{36}(x))$, where c is a scalar which takes a different value as specified below to study the power and $(\hat{\beta}_{31}(x), \cdots, \hat{\beta}_{36}(x))$ were estimators obtained from our clinical data.

4.3.1 Simulation 1

The first set of Monte Carlo simulations was carried out to compare the varying coefficient models under both the Log-Euclidean and affine invariant metrics in estimating the coefficient functions. Recall that VCAI (VCLE) is denoted as the varying coefficient model under the affine invariant (Log-Euclidean) metric. For each simulated data, we use the VCAI and VCLE to estimate the coefficient functions, respectively. Then we calculate the bias and standard error of $\beta(x)$. The results are presented in Table 4.1. All estimators have smaller biases than the standard deviations.

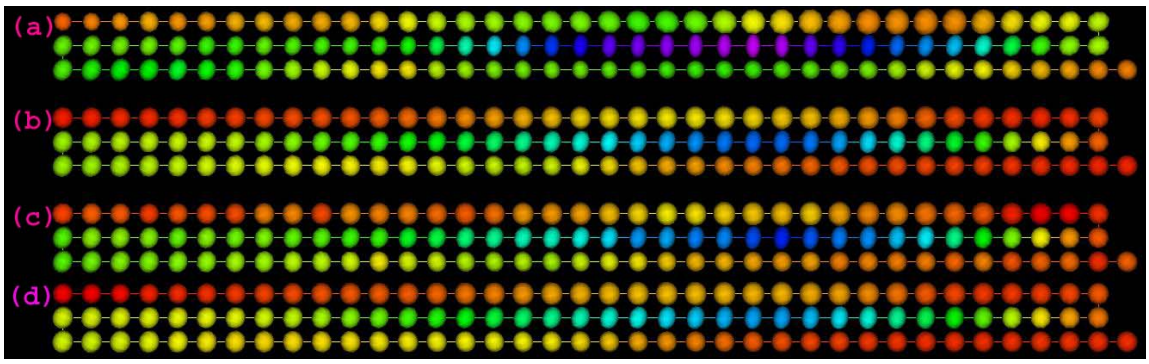

Figure 4.1: Ellipsoidal representations of the simulated (a), true (b) and estimated diffusion tensors along the the right internal capsule tract using VCAI (c) and VCLE (d), respectively, for one selected subject from one simulation, colored with FA values.

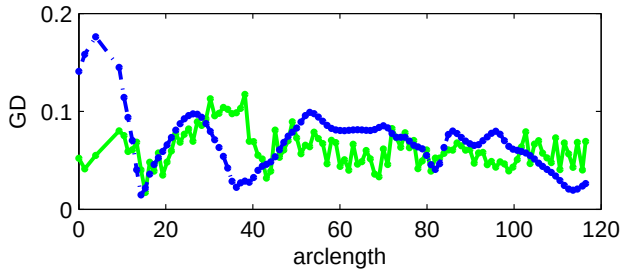

Figure 4.2: Average geodesic distances between simulated and estimated diffusion tensors along the the right internal capsule tract using VCAI (solid line) and VCLE (dash-dotted line), respectively.

Moreover, the biases using VCAI are smaller than those using VCLE while the standard deviations using VCLE are smaller than those using VCAI. One possible reason is that the estimators in VCAI do not have explicit form and is obtained by using a nonlinear optimization method. This may make the results more variable. However, both estimators using VCAI and VCLE perform reasonably well in reconstructing the individual SPD matrix. Figure 4.1 displays the raw, true and estimated SPD matrices using the VCAI and VCLE methods when $c = 1$ for one selected subject from one simulation. Note that both VCAI and VCLE did a comparably good job in recovering ground truth. We also calculate the average geodesic distances between the true and estimated SPD matrices. It is observed from Figure 4.2 that the geodesic distance curve obtained by using VCAI is more rough compared with that using VCLE. This may also be due to the nonlinear estimation methods.

Table 4.1: Bias and standard deviation (SD) for coefficient functions based on 200 simulations

		VCAI			VCLE		
		intercept	gender	Gage	intercept	gender	Gage
Bias ×100	$\beta_1(x)$	0.2633	0.0965	0.1416	0.3098	0.2342	0.1659
	$\beta_2(x)$	0.0957	0.0579	0.0516	0.1298	0.0538	0.0522
	$\beta_3(x)$	0.2667	0.0795	0.1830	0.2733	0.1401	0.1392
	$\beta_4(x)$	0.1122	0.0660	0.0787	0.1407	0.0599	0.0864
	$\beta_5(x)$	0.1258	0.0595	0.0637	0.1330	0.0808	0.0589
	$\beta_6(x)$	0.2339	0.0904	0.1520	0.3144	0.1813	0.1456
SD×100	$\beta_1(x)$	1.9963	1.9658	1.9171	1.8672	1.8102	2.0228
	$\beta_2(x)$	0.6975	0.6761	0.6916	0.5224	0.5247	0.5032
	$\beta_3(x)$	1.8477	1.8600	1.7693	1.6346	1.6529	1.8108
	$\beta_4(x)$	0.8943	0.8742	0.8692	0.7348	0.7379	0.7158
	$\beta_5(x)$	0.8267	0.8191	0.8186	0.6686	0.6686	0.6531
	$\beta_6(x)$	1.7861	1.8263	1.7439	1.6529	1.6318	1.7623

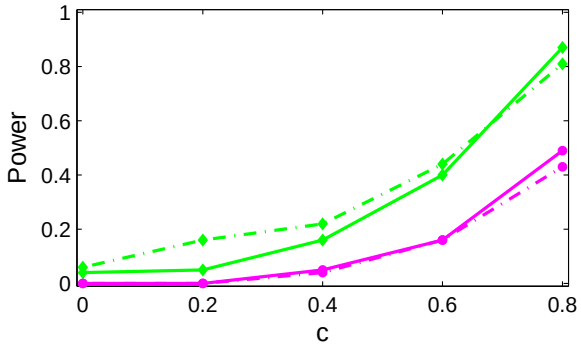

Figure 4.3: Simulation study: Type I and Type II error rates. Rejection rates of T_n based on the resampling method are calculated at five different values of the effect size c using varying coefficient models under the affine invariant (dash-dotted lines) and Log-Euclidean (solid lines) metrics at sample sizes of 96 subjects at the .05 (lines with diamond markers) and .01(lines with circle markers) significance levels.

4.3.2 Simulation 2

In neuroimaging studies, some scientific questions require the assessment of the development of fiber bundle diffusion tensors across time. In this simulation study, the questions were formulated as the hypotheses test $H_0 : \beta_{31}(x) = \cdots = \beta_{36}(x) = 0$ for all x along the right internal capsule tract against $H_1 : \beta_{3l}(x) \neq 0$ for at least one x on the tract for some $l = 1, \cdots, 6$. We first assumed $c = 0$ to assess the Type I error rates (size) for the global test statistic T_n, and then we assumed $c = .2, .4, .6$, and .8 to examine the Type II error rates (power) for T_n at different effect sizes. For all simulations, we used the values of gender and gestational age from our clinical data to simulate the 3×3 SPD matrices (DT's) along the right internal capsule tract. For each simulated SPD matrices, we applied VCAI and VCLE to test the significance of the gestational age effect, respectively. The significance levels were set at $\alpha = .05$ and .01 and 100 replications were used to estimate the rejection rates. Figure 4.3 displays the power for detecting the gestational age effect in each of these cases. It is also observed that the statistical power for rejecting the null hypothesis increases with the effect size and the significance level. In addition, the statistical power using the varying coefficient models under the Log-Euclidean and affine invariant metrics are comparable for large effect sizes. For the small effect sizes, the statistical power under the affine invariant metric is higher. However, the type I error rate is greater than the .05 significance level under this metric.

Table 4.2: Simulated coverage probabilities for coefficient functions based on 200 simulations for VCAI methods

	$\alpha = .05$			$\alpha = .01$		
	intercept	gender	Gage	intercept	gender	Gage
$\beta_1(x)$	.960	.930	.925	.995	.990	.985
$\beta_2(x)$	.915	.925	.945	.990	.995	1.0
$\beta_3(x)$	.970	.930	.935	.990	.985	.985
$\beta_4(x)$	.960	.955	.935	.995	.995	.975
$\beta_5(x)$	.940	.965	.920	.995	.980	.970
$\beta_6(x)$	.965	.930	.930	.995	.985	.985

4.3.3 Simulation 3

The third set of Monte Carlo simulations was carried out to evaluate the coverage probabilities of the simultaneous confidence bands for regression coefficients $\beta(x)$ using VCAI. The simulation setup is the same as in the first set of simulations except $c = 1$ and the sample size is 96. The 95% and 99% simultaneous confidence bands were considered. As suggested by Fan and Zhang (2000), an appropriate smaller bandwidth would improve the accuracy of the confidence bands. In our simulations, we used a shrinkage factor .6. For simplicity, we do not consider estimating the bias of $\hat{\beta}(x)$.

Table 4.2 summarizes the empirical coverage probabilities based on 200 simulations for $\alpha = .01$ and and .05. The coverage probabilities are quite close to the claimed confidence levels. Figure 4.4 depicts typical 95% simultaneous confidence bands for vectors of coefficient functions $\beta_l(x)$ for $l = 1, \cdots, 6$. For each l, $\beta_l(x)$ contains three coefficient functions which correspond the intercept, age and gestational age. It is observed that the length of the simultaneous confidence bands for the second and third functions in $\beta_1(x)$, $\beta_3(x)$ and $\beta_6(x)$ are large. This is because of large variances of corresponding parts in data.

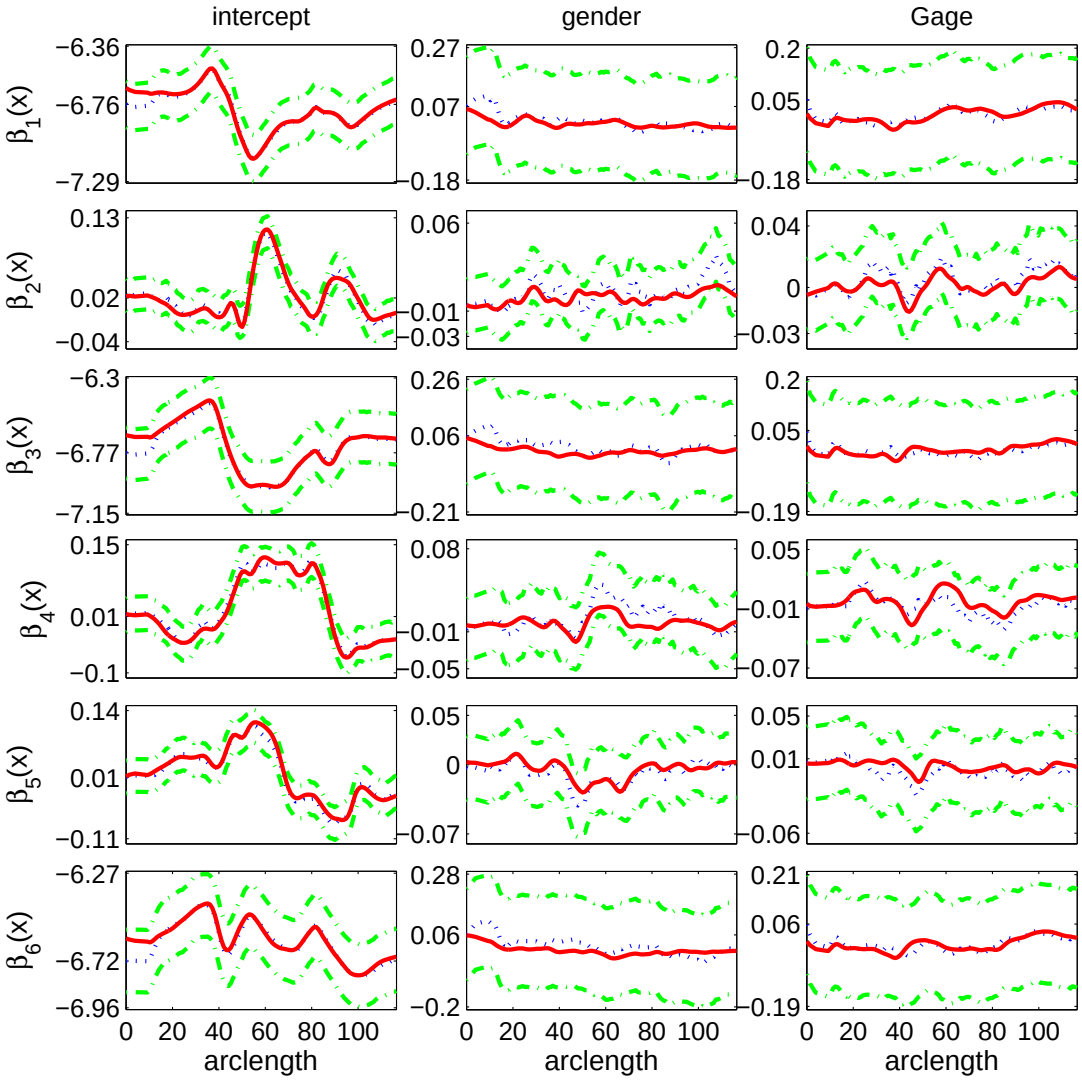

Figure 4.4: Typical 95% simultaneous confidence bands for vectors of coefficient functions $\beta_l(x)$ for $l = 1, \cdots, 6$. The solid, dotted, dash-dotted curves are the true curves, the estimated coefficient functions and the 95% confidence bands, respectively.

4.4 A Real Example

In this section, we apply our VCAI procedure to the same DT data in Section 3.4 of Chapter 3 in order to investigate early brain development. In this study, we also have the same two specific aims. The first one is to compare diffusion tensors along the selected fiber bundle across the male and female groups and thus illuminate the gender effect on the development of these fiber bundle diffusion tensors. The second one is to delineate the development of fiber bundle diffusion tensor across the gestational age effect. To statistically test the effects, we applied our VCAI to diffusion tensors along the fiber tract. For the the selected tract, we fitted the VCAI model (4.3.1) to the diffusion tensors from all 96 subjects, in which $\mathbf{z} = (1, \text{gender}, \text{Gage})^T$. Then, we used the annealing evolutionary stochastic approximation Monte Carlo algorithm to estimate the functional coefficients $\beta(x)$. For the hypothesis testing, we constructed the global test statistic T_n via equation (4.2.22) to test the gender and age effects for the diffusion tensors. The p value of T_n was approximated using the resampling method with $G = 100$ replications. Finally, we constructed the 95% simultaneous confidence bands for the functional coefficients $\beta(x)$.

We statistically test the effects of gender and gestational age on the diffusion tensors along the right internal capsule tract. To test the gender effect, we calculated the local test statistics $T_n(x_j)$ and their corresponding p values across all grid points on the right internal capsule tract. It is observed from Figure 4.5 (a) that most grid points do not have significant $-\log_{10}(p)$ values, which are less than 1.0. Then, we also computed the global test statistic $T_n = 850.40$ and its associated p-value $p = 0.37$ indicating no gender effect. It is observed from Figure 4.5 (b) that the $-\log_{10}(p)$ values of $T_n(x_j)$ for testing the gestational age effect at some grid points of the right tail are greater than 2.0 while a very high significant gestational age effect was found with $T_n = 1397.82$ and its $p-$value $p = 0.01$. This indicates that diffusion tensors along the right internal capsule tract do not differ significantly between male and female groups but are significantly associated with the gestational age. Moreover, the model under the affine invariant metric can detect more significant grid points than that under the Log-Euclidean metric.

The estimated coefficient functions along with the 95% simultaneous confidence bands are depicted in Figure 4.6.

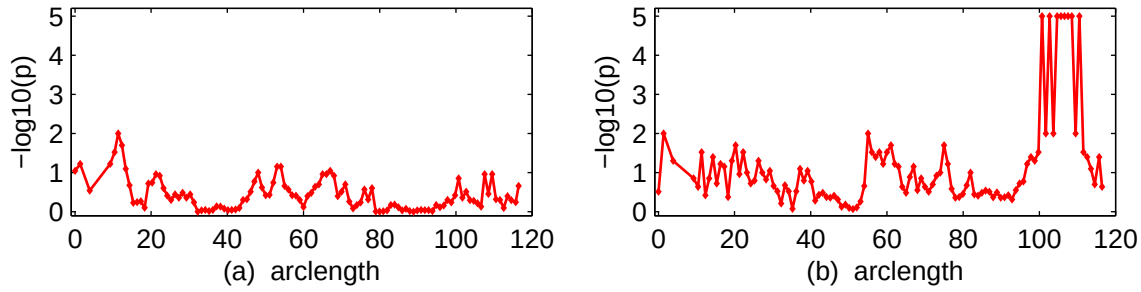

Figure 4.5: The $-\log_{10}(p)$ values of test statistics $T_n(x_j)$ for testing gender or gestational age effect of diffusion tensors on the right internal capsule tract in panel (a) showing no significant gender effect, in panel (b) showing significant gestational age effect.

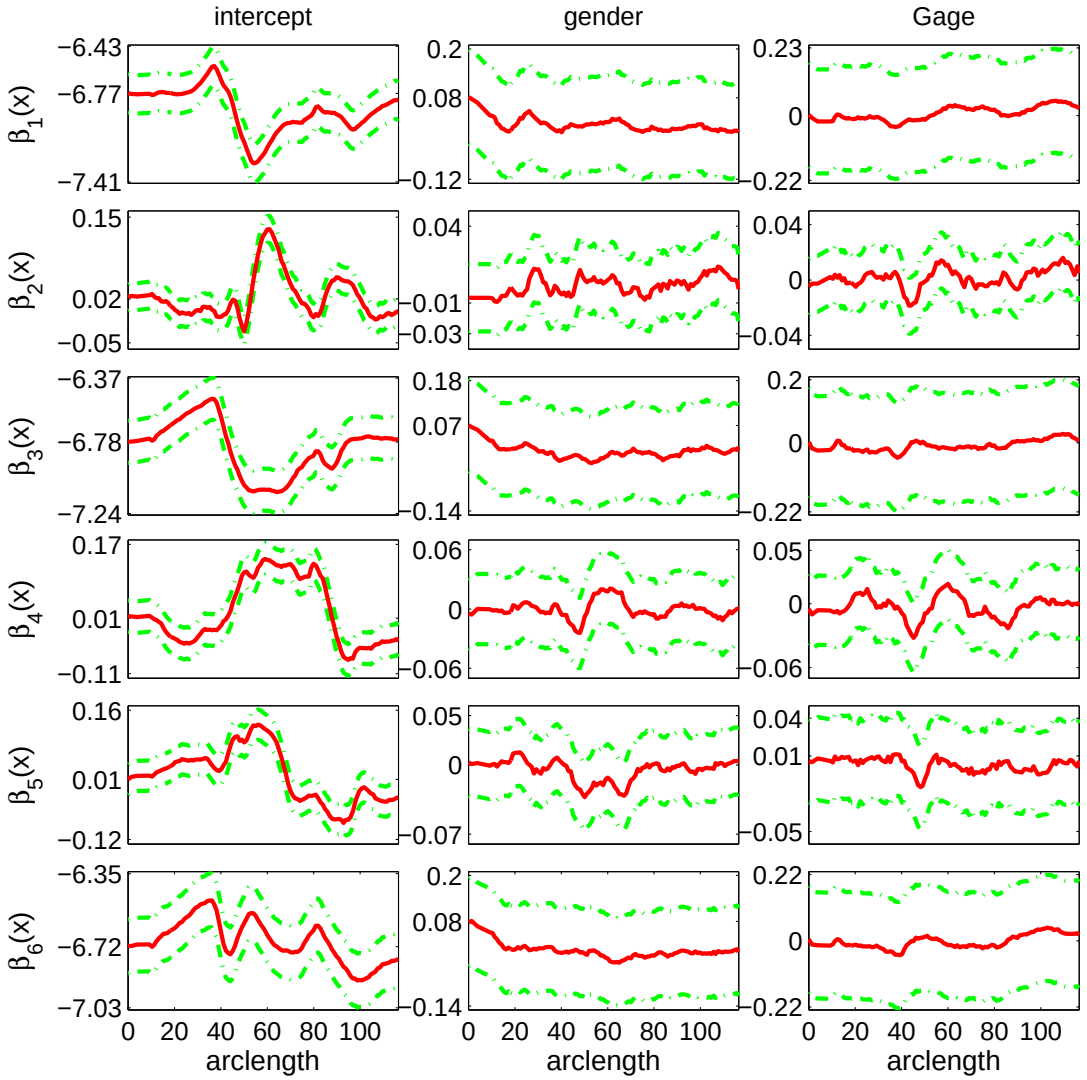

Figure 4.6: 95% simultaneous confidence bands for coefficient functions. The solid curves are the estimated coefficient functions, and the dashed curves are the 95% confidence bands.

4.5 Summary

In this chapter, we have developed a varying coefficient model for SPD matrix valued functional data under the affine invariant metric in order to investigate the effect of different metrics for the space of SPD matrices on statistical inferences. Our simulation results suggest that the estimators based on the affine invariant metric have smaller biases and larger variances compared with those based on the Log-Euclidean metric. Both estimators give reasonable good approximations to the ground truth. In addition, the statistical power using the invariant metric can be slightly higher in detecting the group differences when the effect size is small. However, using VCAI involves more complex computations and in general is time-consuming. It will be the case for the statistical analysis of general manifold data, for which there is no alternative simple metric like the Log-Euclidean metric for the space of SPD matrices.

In addition, our model can be generalized to statistical analysis of any general manifold valued functional data given their Riemannian geometry. For instance, a new imaging technique called high angular resolution diffusion imaging (HARDI) (Tuch (2002)) has been used to imaging complex oriented structures in the brain. Unlike DTI, which models the diffusion with a single tensor and can only reveal a single prevalent fiber direction, HARDI measures water diffusion along many uniformly distributed directions on the sphere and can characterize more complex fiber orientations. Based on HARDI measures, the orientation distribution function (ODF) can be reconstructed at each voxel (Frank (2002); Tuch (2004); Rathi et al. (2009)), which can be used to characterize diffusion. This provides a new type of random samples, which are manifold valued data. The question arises naturally: how to perform statistical analysis on this manifold valued data? By considering the ODF data on a Riemannian manifold, it is possible to extend our methods for DT's to the analysis of ODF data.

Appendix A

Cross Validation Bandwidth Selection (2.3.8)

In this section, we will derive the first-order approximation to the cross validation score $\mathrm{CV}_A(h)$ in (2.3.8). We need some notation as follows:

$$
\begin{aligned}
&M(S,D) = \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\},\\
&G_{n,A}^{(-i)}(\alpha_A(x)) = \sum_{j\neq i} g_A(S_j, D_A(x_j,\alpha_A(x),k_0))^2,\\
&\hat{\alpha}_A^{(-i)}(x,h) = \mathrm{argmin}_{\alpha_A(x)} G_{n,A}^{(-i)}(\alpha_A(x)),\\
&\hat{\alpha}_A(x,h) = \mathrm{argmin}_{\alpha_A(x)} G_{n,A}(\alpha_A(x)), \quad \Delta\hat{\alpha}_A^{(-i)}(x,h) = \hat{\alpha}_A^{(-i)}(x,h) - \hat{\alpha}_A(x,h),\\
&\partial_D M(S,D) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\}}{\partial \mathrm{vecs}(D)},\\
&\partial_G M(S,GG^T) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}GG^TS^{-T/2})^2\}}{\partial \mathrm{vecs}(G)},\\
&\partial_D^2 M(S,D) = \frac{\partial^2 \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\}}{\partial \mathrm{vecs}(D)\partial \mathrm{vecs}(D)^T},\\
&E(S,D,x,x_0) = D_A(x,\alpha_A(x_0),k_0)^{-1/2} S D_A(x,\alpha_A(x_0),k_0)^{-T/2},\\
&M(S,x-x_0,\alpha_A(x_0)) = \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_A(x_0),k_0)S^{-T/2})^2\},\\
&\partial_{\alpha_A(x_0)} M(S,x-x_0,\alpha_A(x_0)) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_A(x_0),k_0)S^{-T/2})^2\}}{\partial \alpha_A(x_0)},\\
&\partial^2_{\alpha_A(x_0)} M(S,x-x_0,\alpha_A(x_0)) = \frac{\partial^2 \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_A(x_0),k_0)S^{-T/2})^2\}}{\partial \alpha_A(x_0)\partial \alpha_A(x_0)^T},\\
&H^{(-i)}(x,\alpha_A(x),h) = \sum_{j\neq i} K_h(x_j-x)\partial^2_{\alpha_A(x)} M(S_j, x_j-x, \alpha_A(x)).
\end{aligned}
$$

Let $\hat{G}_A^{(-i)}(x_i,h)$ and $\hat{G}_A(x_i,h)$ be, respectively, the subcomponents of $\hat{\alpha}_A^{(-i)}(x,h)$ and $\hat{\alpha}_A(x,h)$ corresponding to $G(x)$. Then, by using the Taylor's series expansion, we can approximate the cross validation at bandwidth h by

$$
\mathrm{CV}_A(h) \;=\; n^{-1}\sum_{i=1}^n g_A(S_i, \hat{D}_A^{(-i)}(x_i,h))^2 = n^{-1}\sum_{i=1}^n \mathrm{tr}\{\log(S_i^{-1/2}\hat{D}_A^{(-i)}(x_i,h)S_i^{-T/2})^2\}
$$

$$\approx \quad n^{-1}\sum_{i=1}^{n} g_A(S_i, \hat{D}_A(x_i,h))^2 + 2p_n(h),$$

where $p_n(h)$ can be regarded as the degree of freedom for ILPR and is given by

$$p_n(h) = (2n)^{-1}\sum_{i=1}^{n}\{\partial_G M(S_i, \hat{D}_A(x_i,h))^T \mathrm{vecs}(\hat{G}_A^{(-i)}(x_i,h) - \hat{G}_A(x_i,h))\}. \tag{A.0.1}$$

Since $\hat{\alpha}_A^{(-i)}(x_i,h)$ and $\hat{\alpha}_A(x_i,h)$ minimize $G_{n,A}^{(-i)}(\alpha_A(x_i))$ and $G_{n,A}(\alpha_A(x_i))$, respectively, we have

$$\begin{aligned}
0 &= \sum_{j\neq i}\{K_h(x_j - x_i)\partial_{\alpha(x_i)} M(S_j, x_j - x_i, \hat{\alpha}_A^{(-i)}(x_i,h))\} \qquad (A.0.2)\\
&\approx \sum_{j\neq i}\{K_h(x_j - x_i)\partial_{\alpha(x_i)} M(S_j, x_j - x_i, \hat{\alpha}_A(x_i,h))\}\\
&\quad + \sum_{j\neq i} K_h(x_j - x_i)\partial^2_{\alpha(x_i)} M(S_j, x_j - x_i, \hat{\alpha}_A(x_i,h))\Delta\hat{\alpha}_A^{(-i)}(x_i,h)\\
&= -K_h(0)\partial_{\alpha(x_i)} M(S_i, 0, \hat{\alpha}_A(x_i,h))\\
&\quad + \sum_{j\neq i} K_h(x_j - x_i)\partial^2_{\alpha(x_i)} M(S_j, x_j - x_i, \hat{\alpha}_A(x_i,h))\Delta\hat{\alpha}_A^{(-i)}(x_i,h).
\end{aligned}$$

This yields that

$$\Delta\hat{\alpha}_A^{(-i)}(x_i,h) = K_h(0)H^{(-i)}(x_i, \hat{\alpha}_A(x_i,h), h)^{-1}\partial_{\alpha(x_i)} M(S_i, 0, \hat{\alpha}_A(x_i,h)). \tag{A.0.3}$$

Furthermore, at a given x_i, we consider $\alpha_A(x_i) = (\alpha_{(1)}(x_i)^T, \alpha_{(2)}(x_i)^T)^T$ and $\alpha_A^{(-i)}(x_i,h) = (\alpha_{(1),IA}^{(-i)}(x_i,h)^T,$ $\alpha_{(2),IA}^{(-i)}(x_i,h)^T)^T$, in which $\alpha_{(1)}(x_i)$ and $\alpha_{(1),IA}^{(-i)}(x_i,h)$, respectively, correspond to the unknown parameters in $G(x_i)$. Suppose that $H^{(-i)}(x_i, \hat{\alpha}_A(x_i,h), h)$ can be decomposed according to the decomposition $\alpha(x_i) = (\alpha_{(1)}(x_i)^T, \alpha_{(2)}(x_i)^T)^T$ as follows:

$$H^{(-i)}(x_i, \hat{\alpha}_A(x_i,h), h) = \begin{pmatrix} H_{11,IA}^{(-i)}(x_i,h) & H_{12,IA}^{(-i)}(x_i,h) \\ H_{21,IA}^{(-i)}(x_i,h) & H_{22,IA}^{(-i)}(x_i,h) \end{pmatrix}.$$

It can be shown that all elements in $\partial_{\alpha_{(2)}(x_i)} M(S_i, 0, \hat{\alpha}_A(x_i,h))$ equal zero. Thus, by using the nullity theorem, we have

$$\Delta\hat{\alpha}_{(1),IA}^{(-i)}(x_i,h) = K_h(0)H_{11\cdot 2,IA}^{(-i)}(x_i,h)^{-1}\partial_{\alpha_{(1)}(x_i)} M(S_i, 0, \hat{\alpha}_A(x_i,h)), \tag{A.0.4}$$

where $H^{(-i)}_{11\cdot 2,IA}(x_i,h) = H^{(-i)}_{11,IA}(x_i,h) - H^{(-i)}_{12,IA}(x_i,h)H^{(-i)}_{22,IA}(x_i,h)^{-1}H^{(-i)}_{21,IA}(x_i,h)$. By substituting (A.0.4) into (A.0.1), we have

$$p_n(h) = (2n)^{-1}K_h(0)\sum_{i=1}^{n}\{\partial_G M(S_i,\hat{D}_A(x_i,h))^T H^{(-i)}_{11\cdot 2,IA}(x_i,h)^{-1}\partial_G M(S_i,\hat{D}_A(x_i,h))\}, \quad \text{(A.0.5)}$$

which finishes the proof of (2.3.8).

In order to calculate $\mathrm{CV}_A(h)$, we need the first order derivatives of matrix logarithm and exponential, $\partial_G M(S,GG^T)$, and the first-order and second-order derivatives of $M(S,x-x_0,\alpha_A(x_0))$ with respect to $\alpha_A(x_0)$ as follows.

Lemma A.0.1. *(i) The first-order derivative of $M(S,GG^T)$ with respect to vecs(G) is given by*

$$\frac{\partial M(S,GG^T)}{\partial G_j} = 2tr\{\log(G^TS^{-1}G)G^{-1}\frac{\partial(GG^T)}{\partial G_j}G^{-T}\}, \quad \text{(A.0.6)}$$

where G_j is the $j-th$ unknown element in vecs(G).

(ii) Suppose that $D(\alpha) \in Sym^+(m)$ and $M(\alpha) \in Sym(m)$ are differentiable functions of α, the first order derivatives of $\log(D(\alpha))$ and $\exp(M(\alpha))$ with respect to the $j-th$ component of α_j are, respectively, given by

$$\frac{\partial \exp(M(\alpha))}{\partial \alpha_j} = \int_0^1 \exp((1-s)M(\alpha))\frac{\partial M(\alpha)}{\partial \alpha_j}\exp(sM(\alpha))ds, \quad \text{(A.0.7)}$$

$$\frac{\partial \log(D(\alpha))}{\partial \alpha_j} = \int_0^1 [\{D(\alpha)-I_m\}s+I_m]^{-1}\frac{\partial D(\alpha)}{\partial \alpha_j}[\{D(\alpha)-I_m\}s+I_m]^{-1}ds. \quad \text{(A.0.8)}$$

Proof of Lemma A.0.1. Since $\mathrm{tr}\{\log(S^{-1/2}GG^TS^{-T/2})^2\} = \mathrm{tr}\{\log(G^{-1}SG^{-T})^2\}$, It follows from Proposition 2.1 in Maher (2005) that

$$\frac{\partial M(S,GG^T)}{\partial G_j} = 2\mathrm{tr}\{\log(G^{-1}SG^{-T})G^TS^{-1}G\partial_{G_j}(G^{-1}SG^{-T})\}.$$

Because $\log(G^{-1}SG^{-T})G^TS^{-1}G = G^TS^{-1}G\log(G^{-1}SG^{-T})$ and

$$\partial_{G_j}(G^{-1}SG^{-T}) = -G^{-1}(\partial_{G_j}G)G^{-1}SG^{-T} - G^{-1}SG^{-T}(\partial_{G_j}G)G^{-T},$$

we have

$$\frac{\partial M(S, GG^T)}{\partial G_j} = -2\sum_{i=1}^{n} \mathrm{tr}\{\log(G^{-1}SG^{-T})(G^{-1}\frac{\partial G}{\partial G_j} + \frac{\partial G^T}{\partial G_j}G^{-T})\},$$

which yields (A.0.6). The proof of (A.0.7) and (A.0.8) can be found in Higham (2008).

Lemma A.0.2. *Let $\alpha_j(x_0)$ be the jth element of $\alpha_A(x_0)$. The jth element of the vector $\partial_{\alpha_A(x_0)} M(S, x - x_0, \alpha_A(x_0))$ is given by*

$$\partial_{\alpha_j(x_0)} M(S, x - x_0, \alpha_A(x_0)) = -2tr[\log(E(S, D, x, x_0))D(x, \alpha_A(x_0), k_0)^{-1/2} \times \{\partial_{\alpha_j(x_0)} D(x, \alpha_A(x_0), k_0)\}D(x, \alpha_A(x_0), k_0)^{-T/2}].$$

The $(j,k)-$th element of $\partial^2_{\alpha_A(x_0)} M(S, x - x_0, \alpha_A(x_0))$ is given by

$$\frac{\partial^2 M(S, x - x_0, \alpha_A(x_0))}{\partial \alpha_j(x_0) \partial \alpha_k(x_0)} = -2tr[\frac{\partial \log(E(S, D, x, x_0))}{\partial \alpha_k(x_0)} D(x, \alpha_A(x_0), k_0)^{-1/2} \times \{\partial_{\alpha_j(x_0)} D(x, \alpha_A(x_0), k_0)\}D(x, \alpha_A(x_0), k_0)^{-T/2}] - 2tr(\log(E(S, D, x, x_0)) \frac{\partial [D(x, \alpha_A(x_0), k_0)^{-1/2} \{\partial_{\alpha_j(x_0)} D(x, \alpha_A(x_0), k_0)\} D(x, \alpha_A(x_0), k_0)^{-T/2}]}{\partial \alpha_k(x_0)}).$$

Proof of Lemma A.0.2. By using Lemma A.0.1 and matrix differentiation, we can easily prove Lemma A.0.2.

Appendix B

Annealing Evolutionary Stochastic Approximation Monte Carlo

We now develop an annealing evolutionary stochastic approximation Monte Carlo algorithm for computing $\hat{\alpha}_{IA}(x_0; h)$. Quite recently, the stochastic approximation Monte Carlo algorithm (Liang et al., 2007) has been proposed in the literature as a general simulation technique, which possesses a nice feature in that the moves are self-adjustable and thus not likely to get trapped by local energy minima. The annealing evolutionary SAMC algorithm (Liang, 2010) represents a further improvement of stochastic approximation Monte Carlo for optimization problems by incorporating some features of simulated annealing (Kirkpatrick et al., 1983) and the genetic algorithm (Goldberg, 1989) into its search process.

Like the genetic algorithm, annealing evolutionary stochastic approximation Monte Carlo works on a population of samples. Let $\alpha^l = (\alpha_{A,[1]}, \ldots, \alpha_{A,[l]})$ denote the population, where l is the population size, and $\alpha_{A,[i]} = (\alpha_{i1}, \ldots, \alpha_{iq(k_0+1)})$ is a $q(k_0+1)$-dimensional vector called an individual or chromosome in terms of genetic algorithms. Thus, the minimum of the objective function $G_n(\alpha_A(x_0))$, $\alpha_A(x_0) \in \mathcal{B}$, can be obtained by minimizing the function $U(\alpha^l) = \sum_{i=1}^{l} G_n(\alpha_{A,[i]})$. An unnormalized Boltzmann density can be defined for the population as follows,

$$\psi(\alpha^l) = \exp\Big\{-U(\alpha^l)/\tau\Big\}, \quad \alpha^l \in \mathcal{B}^l, \tag{B.0.1}$$

where $\tau = 1$ is called the temperature, and $\mathcal{B}^l = \mathcal{B} \times \cdots \times \mathcal{B}$ is a product sample space. The sample space can be partitioned according to the function $U(\alpha^l)$ into b subregions: $\mathbb{E}_1 = \{\alpha^l : U(\alpha^l) \le \delta_1\}$, $\mathbb{E}_2 = \{\alpha^l : \delta_1 < U(\alpha^l) \le \delta_2\}$, $\cdots$, $\mathbb{E}_{b-1} = \{\alpha^l : \delta_{b-2} < U(\alpha^l) \le \delta_{b-1}\}$, and $\mathbb{E}_b = \{\alpha^l : U(\alpha^l) > \delta_{b-1}\}$, where $\delta_1 < \delta_2 < \ldots < \delta_{b-1}$ are $b-1$ known real numbers. We note that here the sample space is not necessarily partitioned according to the function $U(\alpha^l)$, for example, the function $\lambda(\alpha^l) = \min\{G_n(\alpha_{A,[1]}), \ldots, G_n(\alpha_{A,[l]})\}$ also works.

Let $\varpi(\delta)$ denote the index of the subregion that a sample with energy u belongs to. For example,

if $\alpha^l \in E_j$, then $\varpi(U(\alpha^l)) = j$. Let $\mathcal{B}^{(t)}$ denote the sample space at iteration t. The algorithm initiates its search in the entire sample space $\mathcal{B}_0 = \bigcup_{i=1}^{b} E_i$, and then iteratively searches in the set

$$\mathcal{B}_t = \bigcup_{i=1}^{\varpi(U_{\min}^{(t)}+\aleph)} E_i, \quad t = 1, 2, \ldots, \tag{B.0.2}$$

where $U_{\min}^{(t)}$ is the best function value obtained until iteration t, and $\aleph > 0$ is a user specified parameter which determines the broadness of the sample space at each iteration. Note that in this method, the sample space shrinks iteration by iteration. To ensure the convergence of the algorithm to the set of global minima, the moves at each iteration are required to admit the following distribution as the invariant distribution,

$$f_{w^{(t)}}(\alpha^l) \propto \sum_{i=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \frac{\psi(\alpha^l)}{e^{w_i^{(t)}}} I(\alpha^l \in \mathbb{E}_i), \quad x \in \mathcal{B}_t^l, \tag{B.0.3}$$

where $w_i^{(t)}$ are the working parameters which will be updated from iteration to iteration as described in the algorithm below.

The annealing evolutionary stochastic approximation Monte Carlo includes five types of moves, the MH-Gibbs mutation, K-point mutation, K-point crossover, snooker crossover, and linear crossover operators. See Liang (2010) for the details of the moves. Let $\rho_1, \ldots, \rho_5$, $0 < \rho_i < 1$ and $\sum_{i=1}^{5} \rho_i = 1$, denote the respective working probabilities of the five types of moves. The algorithm can be summarized as follows.

The algorithm:

(a) (Initialization) Partition the sample space $\mathcal{B}^l$ into b disjoint subregions $\mathbb{E}_1, \ldots, \mathbb{E}_b$; choose the threshold value $\aleph$ and the working probabilities $\rho_1, \ldots, \rho_5$; initialize a population $\alpha^{l(0)}$ at random; and set $w^{(0)} = (w_1^{(0)}, \ldots, w_b^{(0)}) = (0, 0, \ldots, 0)$, $\mathcal{B}_0^l = \bigcup_{i=1}^{b} \mathbb{E}_i$, $U_{\min}^{(0)} = U(\alpha^{l(0)})$ and $t = 0$. Let Θ be a compact set in R^m.

(b) (Sampling) Update the current population $\alpha^{l(t)}$ using the MH-Gibbs mutation, K-point mutation, K-point crossover, snooker crossover, and linear crossover operators according to the respective working probabilities.

(c) (Working weight updating) Update the working weight $w^{(t)}$ by setting

$$w_i^* = w_i^{(t)} + \gamma_{t+1} H_i(w^{(t)}, \alpha^{l(t+1)}), \quad i = 1, \ldots, \varpi(U_{\min}^{(t)} + \aleph),$$

where $H_i(w^{(t)}, \alpha^{l(t+1)}) = I(\alpha^{l(t+1)} \in \mathbb{E}_i)$ for the crossover operators, $H_i(w^{(t)}, \alpha^{l(t+1)}) = \sum_{j=1}^{l} I(\alpha^{l(t+1,j)} \in \mathbb{E}_i)/l$ for the mutation operators, and γ_{t+1} is called the gain factor. If $w^* \in \Theta$, set $w^{(t+1)} = w^*$; otherwise, set $w^{(t+1)} = w^* + c^*$, where $c^* = (c^*, \ldots, c^*)$ and c^* is chosen such that $w^* + c^* \in \Theta$.

(d) (Termination Checking) Check the termination condition, e.g., whether a fixed number of iterations has been reached. Otherwise, set $t \to t + 1$ and go to step (b).

In this article, we follow Liang (2010) to set $\rho_1 = \rho_2 = 0.05$, $\rho_3 = \rho_4 = \rho_5 = 0.3$, and the gain factor sequence

$$\gamma_t = \frac{t_0}{\max(t_0, t)}, \quad t = 0, 1, 2, \ldots, \tag{B.0.4}$$

with $t_0 = 5000$. In general, a large value of t_0 will allow the sampler to reach all the subregions very quickly even for a large system. As shown in Liang (2010), it can converge weakly toward a neighboring set of global minima of $U(\alpha^l)$ in the space of energy. More precisely, the sample $\alpha^{l(t)}$ converges in distribution to a random population with the density function

$$f_w(\alpha^l) \propto \sum_{i=1}^{\varpi(U_{\min}+\aleph)} \frac{\psi(\alpha^l)}{\int_{\mathbb{E}_i} \psi(\alpha^l) d\alpha^l} I(x \in \mathbb{E}_i), \tag{B.0.5}$$

where $U_{\min}$ is the global minimum value of $U(\alpha)$,

Regarding the setting of other parameters, we have the following suggestions. In the algorithm, the moves are reduced to the Metropolis-Hastings moves (Metropolis et al., 1953; Hastings, 1970) within the same subregions. Hence, the sample space should be partitioned such that the MH moves within the same subregion have a reasonable acceptance rate. In this article, we set $u_{i+1} - u_i \equiv 0.2$ for $i = 1, \ldots, b - 1$.

The crossover operator has been modified to serve as a proposal for the moves, and it is no longer as critical as to the genetic algorithm. Hence, the population size l is usually set to a moderate number, ranging from 10 to 100. Since $\aleph$ determines the size of the neighboring set toward which

the method converges, ℵ should be chosen carefully for efficiency of the algorithm. If ℵ is too small, it may take a long time for the algorithm to locate the global minima. In this case, the sample space may contain a lot of separated regions, and most of the proposed transitions will be rejected if the proposal distribution is not spread out enough. If ℵ is too large, it may also take a long time for the algorithm to locate the global energy minimum due to the broadness of the sample space. In practice, the values of l and ℵ can be determined through a trial and error process based on the diagnosis for the convergence of the algorithm. If it fails to converge, the parameters should be tuned to larger values. As suggested by Liang (2010), the convergence of the method can be diagnosed by examining the difference of the patterns of the working weights obtained in multiple runs. In this article, we set $l = 50$ and $\aleph = 50$.

Appendix C

Assumptions and Proof of Theorems 2.3.1, 2.3.2, 2.4.1 and 2.4.2

Assumptions

The following assumptions are needed to facilitate development of our methods, although they are not the weakest possible conditions. We need some notation. Recall that $\psi(S, G, Y) = g_A(S, G\exp(Y)G^T)^2$ and $\alpha = (\alpha_G, \alpha_Y)$, where G is an $m \times m$ lower triangle matrix, $S \in \mathrm{Sym}^+(m)$, $Y \in \mathrm{Sym}(m)$, $\alpha_G = \mathrm{vecs}(G)$, and $\alpha_Y = \mathrm{vecs}(Y)$. We define

$$\begin{pmatrix} \partial_{\alpha_G}\psi(S,G,Y) \\ \partial_{\alpha_Y}\psi(S,G,Y) \end{pmatrix} = \begin{pmatrix} \psi_G(S,G,Y) \\ \psi_Y(S,G,Y) \end{pmatrix},$$
$$\begin{pmatrix} \partial^2_{\alpha_G}\psi(S,G,Y) & \partial^2_{\alpha_G\alpha_Y}\psi(S,G,Y) \\ \partial^2_{\alpha_Y\alpha_G}\psi(S,G,Y) & \partial^2_{\alpha_Y}\psi(S,G,Y) \end{pmatrix} = \begin{pmatrix} \psi_{GG}(S,G,Y) & \psi_{GY}(S,G,Y) \\ \psi_{YG}(S,G,Y) & \psi_{YY}(S,G,Y) \end{pmatrix}.$$

(C1) The kernel function $K(.)$ is a continuous symmetric probability density function with bounded support, say $[-1, 1]$.

(C2) The regression function $D(x) \in \mathrm{Sym}^+(m)$ has a continuous $(k_0 + 1)$-th order derivative in a neighborhood of x_0.

(C3) The bandwidth h tends to zero and $nh \to \infty$.

(C4) The design density $f_X(.)$ is continuous in a neighborhood of x_0 and $f_X(x_0) > 0$.

(C5) The conditional density $f(S|X = x)$ is continuous in a neighborhood of x_0.

(C6) $E\{\partial^2_\alpha\psi(S, G, Y(X))|X = x\}$ and $E[\{\partial_\alpha\psi(S, G, Y(X))\}^{\otimes 2}|X = x]$ are continuous in a neighborhood of x_0.

(C7) The matrices $\mathcal{N}(x) = \begin{pmatrix} u_0\Psi_1(x) & \mathbf{u}\otimes\Psi_2(x) \\ \mathbf{u}^T\otimes\Psi_2(x)^T & \mathcal{U}_2\otimes\Psi_3(x) \end{pmatrix}$ is positive definite in a neighborhood of x_0.

(C8) Let $\|.\|$ be the L_2 norm of a matrix, η_0 be a lower triangle matrix, $\eta_1 \in \mathrm{Sym}(m)$, and $U_\delta = \{(\eta_0,\eta_1) : \|\eta_0\|^2 + \|\eta_1\|^2 \le \delta^2\}$. As $\delta \to 0$,

$$
\begin{aligned}
&E\{\sup_{U_\delta}\|\partial_\alpha^2\psi(S,G(x_0)+\eta_0,Y(X)+\eta_1) - \partial_\alpha^2\psi(S,G(x_0),Y(X))\||X=x\} = o(1),\\
&E\{\sup_{U_\delta}\|\partial_\alpha\psi(S,G(x_0)+\eta_0,Y(X)+\eta_1) - \partial_\alpha\psi(S,G(x_0),Y(X))\\
&-\partial_\alpha^2\psi(S,G(x_0),Y(X))(\mathrm{vecs}(\eta_0)^T,\mathrm{vecs}(\eta_1)^T)^T\||X=x\} = o(\delta),
\end{aligned}
$$

are uniformly in x in a neighborhood of x_0.

(C9) There exists a $b>0$ such that $E\{\|\partial_\alpha\psi(S,G(x_0),Y(X))\|^{b+2}|X=x\}$ is bounded in a neighborhood of x_0.

(C10) Suppose that $\Sigma_{\mathcal{E}_D}(x) = \mathrm{Cov}\{\mathcal{E}_D(X)|X=x\}$ is continuous in a neighborhood of x_0 and there exists a $b>0$ such that $E\{\|\mathcal{E}_D(X)\|^{b+2}|X=x\}$ is bounded in a neighborhood of x_0.

Remark: Assumptions (C1)-(C10) are standard conditions for ensuring the asymptotic properties of local polynomial estimators when x_0 is an interior point of $f_X(\cdot)$ (Fan and Gijbels, 1996; Fan and Yao, 1998; Wand and Jones, 1995). Some conditions can be released with additional technicalities of proofs. For instance, the bounded support restriction on $K(\cdot)$ in (C1) is not essential and can be removed if we put restriction on the tail of $K(\cdot)$. Condition (C2) ensures that $Y(x) = \log(G(x_0)^{-1}D(x)G(x_0)^{-T})$, $G(x)$, and $\log(D(x))$ have a continuous (k_0+1)-th order derivative in a neighborhood of x_0. Moreover, assume that $f_X(.)$ has a bounded support $[0,1]$. All assumptions can be easily modified when x_0 is a boundary point, say left boundary point $x_0 = dh$ or right boundary point $x_0 = 1-dh$ for some $d>0$. For instance, we require that conditions (C2)-(C10) hold in the left neighborhood of 0 or the right neighborhood of 1. For condition (C2), we also need to introduce $f_X(0+)$ as x_0 is the left boundary point and $f_X(1-)$ as x_0 is the right boundary point. For condition (C7), $\mathcal{N}(x)$ is also needed to make some modifications. For simplicity, we omit these details.

Proof of Theorems 2.3.1, 2.3.2, 2.4.1 and 2.4.2

In this section, we will give the detailed proof of Theorems 2.3.1-2.4.2. Since the proof of Theorem 2.3.2 involves more details, we only prove Theorem 2.3.2 and the proof of Theorem 2.3.1, 2.4.1 and 2.4.2 follows the same lines of arguments. In addition, for notational simplicity, we only prove these theorems for the local linear regression for the $k_0 > 0$ case. We also prove theorems for the local constant regression because the proof requires some different treatments. The following lemmas are needed for our technical proofs.

Lemma C.0.3. *Let $R(X) = Y(X) - Y^{(1)}(x_0)(X - x_0)$ and assume that conditions (C1)-(C5), (C7) and (C8) hold. For any random $m \times m$ lower triangle matrix sequence η_{i0} and any random symmetric matrix sequence $\eta_{i1} \in Sym(m)$, for $i = 1, \cdots, n$, if $\max_{1\le i\le n} \|\eta_{i0}\| = o_p(1)$ and $\max_{1\le i\le n} \|\eta_{i1}\| = o_p(1)$, then we have the following results:*

$$\begin{aligned}
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) \\
= \;& nhf_X(0+)u_{0,d}\Psi_1(0+)\{1 + o_p(1)\}, && \text{(C.0.1)} \\
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l \\
= \;& nh^{l+1}f_X(0+)u_{l,d}\Psi_2(0+)\{1 + o_p(1)\}, && \text{(C.0.2)} \\
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{YY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l \\
= \;& nh^{l+1}f_X(0+)u_{l,d}\Psi_3(0+)\{1 + o_p(1)\}, && \text{(C.0.3)} \\
&\sum_{i=1}^{n} hK(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})^T vecs(R(x_i))(x_i - x_0)^l \\
= \;& \frac{1}{2}nh^{l+3}f_X(0+)u_{l+2,d}\Psi_2(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\}, && \text{(C.0.4)} \\
&\sum_{i=1}^{n} hK(x_i - x_0)\psi_{YY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})^T vecs(R(x_i))(x_i - x_0)^l \\
= \;& \frac{1}{2}nh^{l+3}f_X(0+)u_{l+2,d}\Psi_3(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\}. && \text{(C.0.5)}
\end{aligned}$$

Proof of Lemma C.0.3. We only prove (C.0.2), while the remainings can be shown using the same arguments. It is easy to see that

$$\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0), Y(x_i))(x_i - x_0)^l + \\
&\quad \sum_{i=1}^{n} hK_h(x_i - x_0)\{\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\}(x_i - x_0)^l \\
&= T_{n1} + T_{n2}.
\end{aligned}
$$

Let $Z_{j,k} = hK_h(X - x_0)(\psi_{GY})_{j,k}(X - x_0)^l$, where $(\psi_{GY})_{j,k}$ is the (j,k)-th element of the matrix ψ_{GY}. For the (j,k)-th element $(T_{n1})_{j,k}$ in the matrix T_{n1}, we have

$$
(T_{n1})_{j,k} = nE(Z_{j,k}) + O_p\left(\sqrt{nE(Z_{j,k}^2)}\right). \tag{C.0.6}
$$

We calculate the first two moments of $Z_{j,k}$ below. Note that

$$
\begin{aligned}
E(Z_{j,k}) &= E\{hK_h(X - x_0)(\psi_{GY})_{j,k}(X - x_0)^l\} \\
&= \int_0^1 hK_h(y - x_0)(y - x_0)^l(\Psi_2(y))_{j,k} f_X(y)dy \\
&= h^{l+1}\int_{-\min\{d,1\}}^{1} K(z)z^l(\Psi_2(zh + x_0))_{j,k} f_X(zh + x_0)dz,
\end{aligned} \tag{C.0.7}
$$

which can be approximated by $h^{l+1}\mathcal{A}$ with $\mathcal{A} = \int_{-\min\{d,1\}}^{1} K(z)z^l(\Psi_2(0+))_{j,k} f_X(x_0)dz$. Specifically, we consider the difference given by

$$
\mathcal{I}_{n,1} \equiv |\int_{-\min\{d,1\}}^{1} K(z)z^l(\Psi_2(zh + x_0))_{j,k} f_X(zh + x_0)dz - \mathcal{A}|.
$$

Applying the dominated convergence theorem together with the boundedness and continuity assumptions on $f_X(.)$ and $\Psi_2(.)$, we get $\lim_{n\to\infty} \mathcal{I}_{n,1} = 0$. Thus,

$$
E(Z_{j,k}) = h^{l+1} f_X(0+)(\Psi_2(0+))_{j,k} u_{l,d}\{1 + o(1)\}. \tag{C.0.8}
$$

Since $f_X(x)$ and $E\{(\psi_{GY})_{j,k}^2 | X = x\}$ are bounded, there exists a $d_3 > 0$ such that $|f_X(x)E\{(\psi_{GY})_{j,k}^2 | X = x\}| < d_3$ for all x. So we have

$$
E(Z_{j,k}^2) = E\{h^2 K_h^2(X - x_0)(\psi_{GY})_{j,k}^2(X - x_0)^{2l}\}
$$

$$
\begin{aligned}
&= \int_0^1 h^2 K_h^2(y-x_0)(y-x_0)^{2l} f_X(y) E\{(\psi_{GY})_{j,k}^2 | X=y\} dy \\
&= h^{2l+1} \int_{-\min\{d,1\}}^1 K^2(z) z^{2l} f_X(zh+x_0) E\{(\psi_{GY})_{j,k}^2 | X = zh+x_0\} dz \\
&\leq d_3 h^{2l+1} \int_{-\min\{d,1\}}^1 K^2(z) z^{2l} dz.
\end{aligned}
$$

By the continuity of $K(.)$, we have $\int_{-\min\{d,1\}}^1 K^2(z) z^{2l} dz < d_4$ for a given $d_4 > 0$ and

$$
E(Z_{j,k}^2) \leq d_3 d_4 h^{2l+1}. \tag{C.0.9}
$$

Combining with (C.0.6), (C.0.8) and (C.0.9), we have

$$
\begin{aligned}
(T_{n1})_{j,k} &= nh^{l+1}(f_X(x_0)\Psi_2(x_0) u_{l,d}\{1+o(1)\} + O_p(1/\sqrt{nh}))_{j,k} \\
&= nh^{l+1} f_X(0+)(\Psi_2(0+))_{j,k} u_{l,d}\{1+o_p(1)\}.
\end{aligned}
$$

That is, $T_{n1} = nh^{l+1} f_X(0+) u_{l,d} \Psi_2(0+)\{1+o_p(1)\}$.

To prove (C.0.2), it suffices to show that $T_{n2} = o_p(nh^{l+1})$. Let $\Delta_{n0} = \{\eta_{10}, \cdots, \eta_{n0}\}$ and $\Delta_{n1} = \{\eta_{11}, \cdots, \eta_{n1}\}$, where η_{i0} is lower triangle matrix and $\eta_{i1} \in \mathrm{Sym}(m)$ for $i = 1, \cdots, n$. For any given $\delta > 0$, denote $D_\delta = \{\Delta_n = (\Delta_{n0}, \Delta_{n1}) : \|\eta_{i0}\|^2 + \|\eta_{i1}\|^2 \leq \delta^2, \forall i \leq n\}$. Define

$$
\begin{aligned}
V(\Delta_n) &= \frac{1}{nh^{l+1}} \sum_{i=1}^n hK_h(x_i - x_0)(x_i - x_0)^l \{\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) \\
&\quad - \psi_{GY}(S_i, G(x_0), Y(x_i))\}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\sup_{D_\delta} \|V(\Delta_n)\| &\leq \frac{1}{nh^{l+1}} \sum_{i=1}^n hK_h(x_i - x_0)(x_i - x_0)^l \\
&\quad \sup_{D_\delta} \|\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\|.
\end{aligned}
$$

By using condition (C8), as $\delta \to 0$, we have

$$
\epsilon_\delta = E\{\sup_{D_\delta} \|\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\| | x_i = x\} = o(1).
$$

Therefore, as $\delta \to 0$, we have

$$E\left\{\sup_{D_\delta}\|V(\Delta_n)\|\right\} \le \epsilon_\delta \frac{1}{nh^{l+1}} E\left\{\sum_{i=1}^n hK_h(x_i - x_0)(x_i - x_0)^l\right\} \to 0.$$

Since $\max_{1\le i\le n}\|\eta_{i0}\| = o_p(1)$ and $\max_{1\le i\le n}\|\eta_{i1}\| = o_p(1)$, we have $V(\hat{\Delta}_n) = o_p(1)$ for $\hat{\Delta}_n = (\hat{\Delta}_{n0}, \hat{\Delta}_{n1})$ with $\hat{\Delta}_{n0} = (\eta_{10}, \cdots, \eta_{n0})^T$ and $\hat{\Delta}_{n1} = (\eta_{11}, \cdots, \eta_{n1})^T$. Thus, $T_{n2} = nh^{l+1}V(\hat{\Delta}_n) = o_p(nh^{l+1})$, which finishes the proof of (C.0.2).

Lemma C.0.4. *Assume that conditions (C1)-(C8) hold. Then we have*

$$\begin{aligned}
&\sum_{i=1}^n hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y^{(1)}(x_0)(x_i - x_0)) \\
=\ & \frac{1}{2}nh^3 f_X(0+)u_{2,d}\Psi_2(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\} + \\
&\sum_{i=1}^n hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y(x_i)), \qquad \text{(C.0.10)} \\
&\sum_{i=1}^n hK_h(x_i - x_0)\psi_Y(S_i, G(x_0), Y^{(1)}(x_0)(x_i - x_0))(x_i - x_0)^l \\
=\ & \frac{1}{2}nh^{l+3} f_X(0+)u_{l+2,d}\Psi_3(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\} + \\
&\sum_{i=1}^n hK_h(x_i - x_0)\psi_Y(S_i, G(x_0), Y(x_i))(x_i - x_0)^l. \qquad \text{(C.0.11)}
\end{aligned}$$

Proof of Lemma C.0.4. We just prove (C.0.10), while the second one can be similarly shown. We consider

$$\begin{aligned}
J_n \ \equiv\ & \sum_{i=1}^n hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y(x_i) - R(x_i)) \\
=\ & \sum_{i=1}^n hK_h(x_i - x_0)[\psi_G(S_i, G(x_0), Y(x_i)) + \psi_{GY}(S_i, G(x_0), Y(x_i))^T \mathrm{vecs}(-R(x_i)) \\
& +\{\psi_G(S_i, G(x_0), Y(x_i) - R(x_i)) - \psi_G(S_i, G(x_0), Y(x_i)) \\
& -\psi_{GY}(S_i, G(x_0), Y(x_i))^T \mathrm{vecs}(-R(x_i))\}] \\
=\ & J_{n1} + J_{n2} + J_{n3}. \qquad \text{(C.0.12)}
\end{aligned}$$

We need to consider J_{n1}, J_{n2}, and J_{n3} as follows. By using condition (C2) and the Taylor's

series expansion, we have

$$\max_{1\le i\le n}\{\|R(x_i)\|\mathbf{1}(|x_i-x_0|\le h)\}\le\frac{1}{2}\sup_{|\xi-x_0|\le h}\|Y^{(2)}(\xi)\|h^2=O_p(h^2).$$

Let $D_\delta=\{\Delta_n=(\eta_1,\cdots,\eta_n):\|\eta_j\|\le\delta,\eta_j\in\mathrm{Sym}(m),\forall j\le n\}$ for any $\delta>0$. Define

$$\begin{aligned}V(\Delta_n) &= \frac{1}{nh}\sum_{i=1}^n hK_h(x_i-x_0)\{\psi_G(S_i,G(x_0),Y(x_i)+\eta_i)-\psi_G(S_i,G(x_0),Y(x_i))\\ &\quad -\psi_{GY}(S_i,G(x_0),Y(x_i))^T\mathrm{vecs}(\eta_i)\},\end{aligned}$$

By using condition (C8), as $\delta\to 0$, we have $\epsilon_\delta=E\{\sup_{D_\delta}\|\psi_G(S_i,G(x_0),Y(x_i)+\eta_i)-\psi_G(S_i,G(x_0),Y(x_i))-\psi_{GY}(S_i,G(x_0),Y(x_i))^T\mathrm{vecs}(\eta_i)\||x_i=x\}=o(\delta)$ uniformly in a neighborhood of x_0. Therefore, for all $|x_i-x_0|\le h$, we have, as $\delta\to 0$,

$$E\{\sup_{D_\delta}\|V(\Delta_n)\|\}\le\epsilon_\delta\frac{1}{nh}E\left\{\sum_{i=1}^n hK_h(x_i-x_0)\right\}=o(1).$$

Since $\max_{1\le i\le n}\{\|R(x_i)\|\mathbf{1}(|x_i-x_0|\le h)\}=O_p(h^2)=o_p(1)$, we have $V(\hat\Delta_n)=o_p(h^2)$, where $\hat\Delta_n=(R(x_1),\cdots,R(x_n))$. This leads to $J_{n3}=nhV(\hat\Delta_n)=o_p(nh^3)$. Applying equation (C.0.4) in Lemma C.0.3 to the second term J_{n2} in (C.0.12), we get

$$J_{n2}=\frac{1}{2}nh^3f_X(0+)u_{2,d}\Psi_2(0+)^T\mathrm{vecs}(Y^{(2)}(0+))\{1+o_p(1)\},$$

which yields (C.0.10).

Lemma C.0.5. *Assume that conditions (C1)-(C9) hold. Let*

$$\mathcal{T}_n\equiv\begin{pmatrix}\sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i,G(x_0),Y(x_i))\\ \sum_{i=1}^n hK_h(x_i-x_0)(x_i-x_0)\psi_Y(S_i,G(x_0),Y(x_i))/h\end{pmatrix}. \tag{C.0.13}$$

Then $\mathcal{T}_n/\sqrt{nh}$ is asymptotically normal with mean zero and covariance matrix

$$\Sigma_{\mathcal{T}}=f_X(0+)\begin{pmatrix}v_{0,d}\Psi_{11}(0+) & v_{1,d}\Psi_{12}(0+)\\ v_{1,d}\Psi_{12}(0+)^T & v_{2,d}\Psi_{22}(0+)\end{pmatrix}\{1+o(1)\}, \tag{C.0.14}$$

where $v_{j,d}$ for $j = 0, 1, 2$ and $\Psi_{11}(x)$, $\Psi_{12}(x)$ and $\Psi_{22}(x)$ are defined in Section 2.3.2.

Proof of Lemma C.0.5. Let $\mathcal{T}_G^{(ij)}$, and $\mathcal{T}_Y^{(ij)}$ denote the jth elements of $\psi_G(S_i, G(x_0), Y(x_i))$ and $\psi_Y(S_i, G(x_0), Y(x_i))$, respectively. Let

$$\mathcal{T}_{ni} = \begin{pmatrix} hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y(x_i)) \\ hK_h(x_i - x_0)(x_i - x_0)\psi_Y(S_i, G(x_0), Y(x_i))/h \end{pmatrix}.$$

Note that $\mathcal{T}_{ni}$ are independent and $\mathcal{T}_n = \sum_{i=1}^n \mathcal{T}_{ni}$.

It follows from (2.3.6) and Lemma A.0.2 that $E(\mathcal{T}_{ni}) = E(\mathcal{T}_n) = \mathbf{0}$ and the covariance matrix of $\mathcal{T}_n/\sqrt{nh}$ is

$$\frac{n}{nh} E\left\{ h^2 K_h^2(X - x_0) \begin{pmatrix} \psi_G\psi_G^T, & \psi_G\psi_Y^T(X - x_0)/h \\ \psi_Y\psi_G^T(X - x_0)/h, & \psi_Y\psi_Y^T(X - x_0)^2/h^2 \end{pmatrix} \right\}.$$

Using the same arguments as in Lemma C.0.3, we can obtain the asymptotic expression for the covariance matrix of $\mathcal{T}_n$, which equals Σ in (C.0.14). Finally, we will show that the sequence $\mathcal{T}_{ni}/\sqrt{nh}$ satisfies the Linderberg-Feller condition:

$$\sum_{i=1}^n E[\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\|^2 1\left\{\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\| > \epsilon\right\}] \to 0 \quad \text{for any} \ \ \epsilon > 0. \tag{C.0.15}$$

Let $b > 0$ and $\|\mathcal{T}'_{ni}\|^2 = (\mathcal{T}_G^{(i1)})^2 + \cdots + (\mathcal{T}_G^{(iq)})^2 + h^{-2}\{(\mathcal{T}_Y^{(i1)})^2 + \cdots + (\mathcal{T}_Y^{(iq)})^2\}(x_i - x_0)^2$.

$$\begin{aligned} & E\left[\left\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\right\|^2 1\left\{\left\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\right\| > \epsilon\right\}\right] \le \mathcal{K}_i \equiv E\left\{\frac{\left(K(h^{-1}(x_i - x_0))\right)^{b+2}\|\mathcal{T}'_{ni}\|^{(b+2)}}{(\sqrt{nh})^{b+2}(\epsilon)^b}\right\} \\ &= \int_{-\min\{d,1\}}^1 \frac{K(z)^{b+2}E\{\|\mathcal{T}'_{ni}\|^{(b+2)}|x_i = zh + x_0\}}{(\sqrt{nh})^{b+2}h^{-1}\epsilon^b} f_X(zh + x_0)dz \\ &\le \int_{-\min\{d,1\}}^1 \frac{E\{\|\psi_\alpha(S_i, G(x_0), Y(X))\|^{b+2}|X = zh + x_0\}}{\{K(z)\}^{-b-2}(\sqrt{nh})^{b+2}f_X(zh + x_0)^{-1}h^{-1}\epsilon^b} dz. \end{aligned}$$

Combining conditions (C4) and (C9) yields that there is a constant $d_5 > 0$ such that $\sum_{i=1}^n \mathcal{K}_i \le d_5 nh/(\sqrt{nh})^{b+2} \to 0$. Thus, it follows from the Linderberg-Feller theorem that $\mathcal{T}_n/\sqrt{nh}$ is asymptotically normal with mean 0 and covariance Σ.

Proof of Theorem 2.3.2 (i). Let $\tilde{Y}^{(1)}(x_0) = hY^{(1)}(x_0)$, $\gamma = (\text{vecs}(G(x))^T, \text{vecs}(\tilde{Y}^{(1)}(x))^T)^T$ and

$\gamma_0 = (\text{vecs}(G(x_0))^T, \text{vecs}(hY^{(1)}(x_0))^T)^T$. Thus, $G_n(\gamma)$ can be written as

$$\sum_{i=1}^{n} hK_h(x_i - x_0)\psi(S_i, G(x_0), \tilde{Y}^{(1)}(x_0)(x_i - x_0)/h).$$

Let $U_\delta(\gamma_0) = \{\gamma : ||\gamma - \gamma_0|| < \delta\}$ for $\delta > 0$. We will show that for any small $\delta > 0$,

$$\lim_{n\to\infty} P\{\inf_{\gamma \in U_\delta(\gamma_0)} G_n(\gamma) > G_n(\gamma_0)\} = 1. \tag{C.0.16}$$

By a Taylor's series expansion, we have

$$G_n(\gamma) - G_n(\gamma_0) = G_n^{(1)}(\gamma_0)^T(\gamma - \gamma_0) + \frac{1}{2}(\gamma - \gamma_0)^T G_n^{(2)}(\gamma^*)(\gamma - \gamma_0), \tag{C.0.17}$$

where $\gamma \in U_\delta(\gamma_0)$ and $\gamma^* = (\text{vecs}(G(x_*))^T, \text{vecs}(\tilde{Y}^{(1)}(x_*))^T)^T$ lies in between $\gamma(x_0)$ and γ.

It follows from lemmas C.0.4 and C.0.5 that

$$\begin{aligned}
& (nh)^{-1}G_n^{(1)}(\gamma_0) \\
= \; & \frac{1}{nh}\sum_{i=1}^{n} hK_h(x_i - x_0) \begin{pmatrix} \psi_G(S, G(x_0), Y^{(1)}(x_0)(x_i - x_0)) \\ \psi_Y(S, G(x_0), Y^{(1)}(x_0)(x_i - x_0))(x_i - x_0)/h \end{pmatrix} \\
= \; & \frac{1}{2}h^2 f_X(0+)\text{vecs}(Y^{(2)}(0+)) \begin{pmatrix} u_{2,d}\Psi_2(0+) \\ u_{3,d}\Psi_3(0+) \end{pmatrix} \{1 + o_p(1)\} \\
& + \frac{1}{nh}\sum_{i=1}^{n} hK_h(x_i - x_0) \begin{pmatrix} \psi_G(S_i, G(x_0), Y(x_i))^T \\ h^{-1}(x_i - x_0)\psi_Y(S_i, G(x_0), Y(x_i))^T \end{pmatrix} \\
= \; & \frac{1}{2}h^2 f_X(0+)(u_{2,d}\Psi_2(0+)^T, u_{3,d}\Psi_3(0+)^T)^T \text{vecs}(Y^{(2)}(0+))\{1 + o_p(1)\} + \frac{1}{nh}\mathcal{T}_n \\
= \; & o_p(1).
\end{aligned}$$

We define $\Delta(S, G, Y, X)$ as

$$\begin{pmatrix} \psi_{GG}(S, G, Y) & (X - x_0)h^{-1}\psi_{YG}(S, G, Y) \\ (X - x_0)h^{-1}\psi_{GY}(S, G, Y) & (X - x_0)^2 h^{-2}\psi_{YY}(S, G, Y) \end{pmatrix}. \tag{C.0.18}$$

With some calculation, we have

$$\frac{1}{nh}G_n^{(2)}(\gamma^*) = \frac{1}{nh}\sum_{i=1}^{n} hK_h(x_i - x_0)[\{\Delta(S_i, G(x_*), \tilde{Y}^{(1)}(x_*)(x_i - x_0)h^{-1}, x_i)$$
$$-\Delta(S_i, G(x_0), Y(x_i), x_i)\} + \Delta(S_i, G(x_0), Y(x_i), x_i)]$$
$$\equiv M_{n1} + M_{n2}.$$

It follows from (C.0.1), (C.0.2) and (C.0.3) in Lemma C.0.3 that

$$M_{n2} = f_X(x_0)\begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T \\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+) \end{pmatrix}\{1 + o_p(1)\},$$

which is positive definite in probability. Note that $\|G(x_*) - G(x_0)\| \le \delta$ and for all $|x_i - x_0| < h$, we have $\max_i \|\tilde{Y}^{(1)}(x_*)(x_i - x_0)/h - Y(x_i)\| \le \max_i \|R(x_i)\| + \delta$. Then it follows from Lemma C.0.3 that $\limsup_{\delta\to 0}\limsup_{n\to\infty}\|M_{n1}\| = 0$ in probability. Thus, for any $\gamma \in U_\delta(\gamma_0)$, we have for sufficiently small $\delta > 0$

$$\lim_{n\to\infty} P[\inf_{\gamma\in U_\delta(\gamma_0)} \frac{1}{nh}(\gamma - \gamma_0)^T G_n^{(2)}(\gamma^*)(\gamma - \gamma_0) > 0.5af_X(x_0)\delta^2] = 1,$$

which yields that $G_n(\gamma)$ has a local minimum $\hat{\gamma} = (\mathrm{vecs}(\hat{G})^T, \mathrm{vecs}(h\hat{Y}^{(1)})^T)^T$ in the interior of $U_\delta(\gamma_0)$. Then, we have $\lim_{n\to\infty} P\{\|\hat{\gamma}_n - \gamma_0\| \le \delta\} = 1$. This implies Theorem 2.3.2 (i).

Proof of Theorem 2.3.2(iii). Let $\hat{\eta}_0 = \hat{G} - G(x_0)$ and $\hat{\eta}_{1i} = \{\hat{Y}^{(1)} - Y^{(1)}(x_0)\}(x_i - x_0) - R(x_i)$. The estimator $\hat{\gamma}$ satisfies the following local estimating equations:

$$\sum_{i=1}^{n} K_h(x_i - x_0)\partial_\gamma\psi(S_i, \hat{G}, \hat{Y}^{(1)}(x_i - x_0)) = \mathbf{0}, \tag{C.0.19}$$

where $\partial_\gamma\psi(S, G, Y) \equiv \left(\psi_G(S, G, Y)^T, (X - x_0)\psi_Y(S, G, Y)^T\right)^T$. It follows from (C.0.19) that

$$\sum_{i=1}^{n} hK_h(x_i - x_0)[\partial_\gamma\psi(S_i, G(x_0), Y(x_i)) + \partial^2_{\alpha\gamma}\psi(S, G(x_0), Y(x_i))^T(\mathrm{vecs}(\hat{\eta}_0)^T, \mathrm{vecs}(\hat{\eta}_{1i})^T)^T$$
$$+ \{\partial_\gamma\psi(S, G(x_0) + \hat{\eta}_0, Y(x_i) + \hat{\eta}_{1i}) - \partial_\gamma\psi(S, G(x_0), Y(x_i))$$
$$- \partial^2_{\alpha\gamma}\psi(S, G(x_0), Y(x_i))^T(\mathrm{vecs}(\hat{\eta}_0)^T, \mathrm{vecss}(\hat{\eta}_{1i})^T)^T\}] = 0. \tag{C.0.20}$$

Note that the second term on the left hand side of (C.0.20) equals the sum of L_{n1} and L_{n2}, which

can, respectively, be approximated by

$$\begin{aligned}
L_{n1} &= \sum_{i=1}^{n} hK_h(x_i - x_0)\Delta(S_i, G(x_0), Y(x_i), x_i) \begin{pmatrix} \text{vecs}(\hat{\eta}_0) \\ h \cdot \text{vecs}(\hat{Y}^{(1)} - Y^{(1)}(x_0)) \end{pmatrix} \\
&= f_X(0+)nh \begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+) \\ u_{1,d}\Psi_2(0+)^T & u_{2,d}\Psi_3(0+) \end{pmatrix} \begin{pmatrix} \text{vecs}(\hat{\eta}_0) \\ h \cdot \text{vecs}(\hat{Y}^{(1)} - Y^{(1)}(x_0)) \end{pmatrix} \times \\
&\quad \{1 + o_p(1)\}, \\
L_{n2} &= \sum_{i=1}^{n} hK_h(x_i - x_0)\Delta(S_i, G(x_0), Y(x_i), x_i) \begin{pmatrix} \text{vecs}(O_m) \\ h \cdot \text{vecs}(R(x_i)) \end{pmatrix} \\
&= \frac{1}{2} f_X(0+)nh^3 \begin{pmatrix} u_{2,d}\Psi_2(0+) \\ u_{3,d}\Psi_3(0+) \end{pmatrix} \text{vecs}(Y^{(2)}(0+))\{1 + o_p(1)\}.
\end{aligned}$$

By the consistency of $\hat{\gamma} = (\text{vecs}(\hat{G})^T, \text{vecs}(h\hat{Y}^{(1)})^T)^T$, we have $\|\hat{\eta}_0\| = o_p(1) = O_p(\|\hat{G} - G(x_0)\|)$ and

$$\begin{aligned}
&\sup_{\{i:|x_i - x_0| \le h\}} \|\hat{\eta}_{i1}\| \le \sup_{\{i:|x_i - x_0| \le h\}} \{\|R(x_i)\| + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|\} \\
= \; & O_p(h^2 + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|) = o_p(1).
\end{aligned}$$

Thus, it follows from (C8) and Lemma C.0.4 that the third term on the left hand side of (C.0.20) is at the order of $o_p(nh)\{h^2 + \|\hat{G} - G(x_0)\| + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|\}$. Let

$$B_n = \frac{h^2}{2} \begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T \\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+) \end{pmatrix}^{-T} \begin{pmatrix} u_{2,d}\Psi_2(0+)^T \\ u_{3,d}\Psi_3(0+)^T \end{pmatrix} \text{vecs}(Y^{(2)}(0+))\{1 + o_p(1)\}.$$

Then from (C.0.20), we obtain that

$$\hat{\gamma} - \gamma_0 = B_n + \{nhf_X(0+)\}^{-1} \begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T \\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+) \end{pmatrix}^{-T} \mathcal{T}_n\{1 + o_p(1)\}.$$

Finally, Theorem 2.3.2 (iii) follows from Lemma C.0.5 and the Slutsky's theorem.

The above derivation holds for any $k_0 > 0$. When $k_0 = 0$ and $K(.)$ is symmetric, the following

modifications need to be made.

Lemma C.0.6. *Let $\eta_0(X) = G(X) - G(x_0)$. Assume that conditions (C1)-(C5) and (C7) hold. If x_0 is an interior point of $f_X(\cdot)$, then we have*

$$\begin{aligned} & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))vecs(\eta_0(x_i)) \\ = \; & nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1 + o_p(1)\}. \end{aligned} \tag{C.0.21}$$

If x_0 is a boundary point of $f_X(\cdot)$, then we have

$$\begin{aligned} & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))vecs(\eta_0(x_i)) \\ & \qquad = nh^2 f_X(0+)\Psi_1(0+)G^{(1)}(0+)u_{1,d}\{1 + o_p(1)\}. \end{aligned} \tag{C.0.22}$$

Proof of Lemma C.0.6. We only prove equation (C.0.21). Let

$$T_{n1} = \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))\text{vecs}(\eta_0(x_i)).$$

For the (j)-th element $(T_{n1})_j$ of the vector T_{n1}, we have

$$(T_{n1})_j = nE(Z_j) + O_p(\sqrt{nE(Z_j^2)}). \tag{C.0.23}$$

We calculate the first two moments of Z_j below. Note that

$$\begin{aligned} E(Z_j) &= E[hK_h(X - x_0)(\psi_{GG}(S_i, G(X))\text{vecs}(\eta_0(X)))_j] \\ &= \int_{-\infty}^{\infty} hK_h(y - x_0)(E\{\psi_{GG}(S_i, G(X))|X = y\}\text{vecs}(\eta_0(y)))_j f_X(y)dy \\ &= h\int_{-1}^{1} K(z)(E\{\psi_{GG}(S_i, G(X))|X = zh + x_0\}\text{vecs}(\eta_0(zh + x_0)))_j f_X(zh + x_0)dz, \end{aligned}$$

By a Taylor's series expansion, we have $\eta_0(zh + x_0) = G^{(1)}(x_0)zh + G^{(2)}(x_0)(zh)^2/2$. Applying the dominated convergence theorem together with the continuity assumptions on $f_X^{(1)}(\cdot)$, $G(.)^{(2)}$ and

$\Psi_1(.)$ yields

$$E(Z_j) = h^3 u_2(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j\{1+o(1)\}. \tag{C.0.24}$$

By the continuity $f_X^{(1)}(x)$ and $G(x)^{(2)}$ together with condition (C7), we have

$$\begin{aligned} E(Z_j^2) &= E\{h^2K_h^2(X-x_0)(\psi_{GG}(S_i,G(x_i))\text{vecs}(\eta_0(x_i)))_j^2\} \\ &= \int_{-\infty}^{\infty} h^2K_h^2(y-x_0)f_X(y)(E\{\psi_{GG}(S,G(X))|X=y\}\text{vecs}(\eta_0(y)))_j^2dy \\ &= h\int_{-1}^{1} K^2(z)z^{2l}f_X(zh+x_0)(E\{\psi_{GG}(S,G(X))|X=zh+x_0\} \\ &\quad \text{vecs}(\eta_0(zh+x_0)))_j^2dz \\ &= h^3\{1+O(h)\} \end{aligned} \tag{C.0.25}$$

Combining with (C.0.23),(C.0.24) and (C.0.25), we have

$$\begin{aligned} (T_{n1})_j &= nh^3[(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j u_2\{1+o(1)\} \\ &\quad +O_p(1/\sqrt{nh^3})] \\ &= nh^3(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j u_2\{1+o_p(1)\}. \end{aligned}$$

That is, $T_{n1} = nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1+o_p(1)\}$.

Lemma C.0.7. *Assume that conditions (C1)-(C7) hold. If x_0 is an interior point of $f_X(\cdot)$, then we have*

$$\begin{aligned} &\sum_{i=1}^{n} hK_h(x_i-x_0)\psi_G(S_i,G(x_i)) \\ =\ & nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1+o_p(1)\} + \\ &\sum_{i=1}^{n} hK_h(x_i-x_0)\psi_G(S_i,G(x_0)). \end{aligned} \tag{C.0.26}$$

If x_0 is the left boundary point of $f_X(\cdot)$,

$$\sum_{i=1}^{n} hK_h(x_i-x_0)\psi_G(S_i,G(x_i)) \tag{C.0.27}$$

$$= \quad nh^2 f_X(0+)\Psi_1(0+)G^{(1)}(0+)u_{1,d}\{1+o_p(1)\} + \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_G(S_i, G(x_0)).$$

Lemma C.0.8. *Assume that conditions (C1)-(C9) hold. Let*

$$\mathcal{T}_n \equiv \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_G(S_i, G(x_i)). \tag{C.0.28}$$

Then $\mathcal{T}_n/\sqrt{nh}$ *is asymptotically normal with mean zero and covariance matrices*

$$\Sigma = f_X(x_0)v_0\Psi_{11}(x_0)\{1+o(1)\} \text{ if } x_0 \text{ is an interior point of } f_X(x), \tag{C.0.29}$$

$$\Sigma = f_X(0+)v_0\Psi_{11}(0+)\{1+o(1)\} \text{ if } x_0 = dh \text{ is a boundary point of } f_X(x). \tag{C.0.30}$$

The proof is similar to the proof of Lemma C.0.5.

Proof of Theorem 2.3.2(ii). We use Lemmas 6-8 and follow the same lines of Theorem 2.3.2 (iii) to prove Theorem 2.3.2 (ii).

Appendix D

Assumptions in Theorems 3.2.1 and 3.2.2

Assumption C1. $\mathcal{E}_i(x)$ and $\mathcal{U}_i(x)$ are identical and independent copies of $\mathrm{SP}(0, \Sigma_{\mathcal{E}})$ and $\mathrm{SP}(0, \Sigma_{\mathcal{U}})$, respectively, and $\mathcal{E}_i(x)$ and $\mathcal{U}_i(t)$ are independent for $x \neq t$. Moreover, with probability one, the sample path of $\mathcal{U}_i(x)$ has continuous second-order derivatives on $[0, L_0]$ and $E[\sup_{x\in[0,L_0]} ||\mathcal{U}(x)||_2^{r_1}] < \infty$ and $E\{\sup_{x\in[0,L_0]}[||\dot{\mathcal{U}}(x)||_2 + ||\ddot{\mathcal{U}}(x)||_2]^{r_2}\} < \infty$ for all $r_1, r_2 \in (2, \infty)$, where $||\cdot||_2$ is the Euclidean norm.

Assumption C2. All components of $\beta(x)$ and $\Sigma_{\mathcal{E}}(x, x)$ have continuous second derivatives on $[0, L_0]$. The fourth moments of $\mathcal{E}_i(x)$ are continuous on $[0, L_0]$. All components of $\Sigma_{\mathcal{U}}(x, t)$ have continuous second-order partial derivatives with respect to $(x, t) \in [0, L_0]^2$. Moreover, $\Sigma_{\mathcal{E}}(x, x)$ and $\Sigma_{\mathcal{U}}(x, x)$ are positive for all $x \in [0, L_0]$.

Assumption C3. The grid points $\mathcal{X} = \{x_j, j = 1, \cdots, n_G\}$ are independently and identically distributed with density function $\pi(x)$, which has the bounded support $[0, L_0]$. For some constants π_L and $\pi_U \in (0, \infty)$ and any $x \in [0, L_0]$, $\pi_L \leq \pi(x) \leq \pi_U$ and $\pi(x)$ has continuous second-order derivative.

Assumption C4. The kernel function $K(t)$ is a symmetric density function with a compact support [-1, 1] and Lipschitz continuous.

Assumption C5. The covariate vectors $\mathbf{z}_i$ are independently and identically distributed with $E\mathbf{z}_i = \mu_z$ and $E[||\mathbf{z}_i||_2^4] < \infty$ and that $E[\mathbf{z}_i^{\otimes 2}] = \Omega_Z$ is invertible.

Assumption C6. Both n and n_G converge to ∞, $h^{(1)} = o(1)$, $n_G h^{(1)} \to \infty$, $h^{(1)-1}|\log h^{(1)}|^{1-2/q_1} \leq n_G^{1-2/q_1}$, where $q_1 \in (2, 4)$.

Assumption C7. $E[|\mathcal{E}_i(x_j)|^{q_2}] < \infty$ for some $q_2 \in (4, \infty)$; $h^{(2)} = o(1)$, $n_G h^{(2)} \to \infty$, and $(h^{(2)})^{-4}(\log n/n)^{1-2/q_2} = o(1)$.

Appendix E

Proof of Theorems 4.2.1, 4.2.2 and 4.2.3

Assumptions

The following assumptions are needed to facilitate the technical details, although they are not the weakest possible conditions.

Assumption M1. vecs($\mathcal{E}_i(x)$) and vecs($\mathcal{U}_i(x)$) are identical and independent copies of SP$(0, \Sigma_{\mathcal{E}})$ and SP$(0, \Sigma_{\mathcal{U}})$, respectively, and $\mathcal{E}_i(x)$ and $\mathcal{E}_i(x')$ are independent for $x \neq x'$. Moreover, with probability one, the sample path of $\mathcal{U}_i(x)$ has continuous second-order derivatives on $[0, L_0]$ and $E[\sup_{x\in[0,L_0]} ||\mathcal{U}(x)||_2^{r_1}] < \infty$ and $E\{\sup_{x\in[0,L_0]}[||\dot{\mathcal{U}}(x)||_2 + ||\ddot{\mathcal{U}}(x)||_2]^{r_2}\} < \infty$ for $r_1, r_2 \in (2, \infty)$, where $||\cdot||_2$ is the Euclidean norm.

Assumption M2. All components of $\beta(x)$ and $\Sigma_{\mathcal{E}}(x, x)$ have continuous second derivatives on $[0, L_0]$. The fourth moments of $\mathcal{E}_i(x)$ are continuous on $[0, L_0]$. All components of $\Sigma_{\mathcal{U}}(x, x')$ have continuous second-order partial derivatives with respect to $(x, x') \in [0, L_0]^2$. Moreover, $\Sigma_{\mathcal{E}}(x, x)$ and $\Sigma_{\mathcal{U}}(x, x)$ are positive for all $x \in [0, L_0]$.

Assumption M3. The grid points $\mathcal{X} = \{x_j, j = 1, \cdots, n_G\}$ are independently and identically distributed with density function $\pi(x)$, which has the bounded support $[0, L_0]$. For some constants π_L and $\pi_U \in (0, \infty)$ and any $x \in [0, L_0]$, $\pi_L \leq \pi(x) \leq \pi_U$ and $\pi(x)$ has continuous second-order derivative.

Assumption M4. The kernel function $K(x)$ is a symmetric density function with a compact support [-1, 1] and Lipschitz continuous.

Assumption M5. The covariate vectors $\mathbf{z}_i$ are independently and identically distributed with $E\mathbf{z}_i = \mu_z$ and $E[||\mathbf{z}_i||_2^4] < \infty$ and that $E[\mathbf{z}_i^{\otimes 2}] = \Omega_Z$ is invertible.

Assumption M6. Both n and n_G converge to ∞, $h^{(1)} = o(1)$, $n_G h^{(1)} \to \infty$, $h^{(1)-1}|\log h^{(1)}|^{1-2/q_1} \leq n_G^{1-2/q_1}$, where $q_1 \in (2, 4)$.

Assumption M7. $E\{\partial_Y^2 \psi(S(X), C(\mathbf{z}, \beta(X)))|X = x, \mathbf{Z}\}$ is bounded, where $\mathbf{Z} = \{\mathbf{z}_1, \cdots, \mathbf{z}_n\}$.

Assumption M8. Let $\|.\|$ be the L_2 norm of a matrix, η be a lower triangle matrix, $U_\delta = \{\eta : \|\eta\| \leq$

$\delta\}$. As $\delta \to 0$, $E\{\sup_{U_\delta} \|\partial_Y^2 \psi(S(X), C(\mathbf{z}, \beta(X)+\eta)) - \partial_Y^2 \psi(S(X), C(\mathbf{z}, \beta(X)))\| | X = x, \mathbf{Z}\} = o(1)$ are uniformly in x in a neighborhood of x.

Assumption M9. There exists a $b > 0$ such that $E\left\{\|\partial_Y \psi(S(X), C(\mathbf{z}, \beta(X)))\|^{2+b} | X = x, \mathbf{Z}\right\}$ is bounded in a neighborhood of x.

Assumption M10. $E[|\mathcal{E}_{i,k}(x_j)|^{q_2}] < \infty$ for some $q_2 \in (4, \infty)$ and all k; $h^{(2)} = o(1)$, $n_G h^{(2)} \to \infty$, and $h^{(2)-4}(\log n/n)^{1-2/q_2} = o(1)$.

Assumption M11. $E[|\mathcal{E}_{i,k}(x_j)|^{q_3}] < \infty$ for some $q_3 \in (4, \infty)$ and all k; $h^{(3)} = o(1)$, $n_G h^{(3)} \to \infty$, and $h^{(3)-2}(\log n/n)^{1-2/q_3} = o(1)$.

Proof of Theorem 4.2.1

We need some notations. Recall that Λ is a $q \times q$ diagonal matrix with q diagonal entries $1, 2, 1, 2, 2, 1 \cdots, 2, \cdots, 2, 1$. Let

$$
\begin{aligned}
&X_n(x) = \sqrt{n}\{\hat{\beta}(x) - \beta(x) - 0.5u_2 h^{(1)2} \ddot{\beta}(x)\} + o_p(h^{(1)2}) \\
&\mathcal{C}^{(k)}(x_j, \mathbf{z}_i) = -C(\mathbf{z}_i, \beta(x_j))^{-1} \frac{\partial C C^T}{\partial Y_k} C(\mathbf{z}_i, \beta(x_j))^{-T} \\
&\mathcal{C}(x_j, \mathbf{z}_i) = (\Lambda \text{vecs}(\mathcal{C}^{(1)}(x_j, \mathbf{z}_i)), \cdots, \Lambda \text{vecs}(\mathcal{C}^{(q)}(x_j, \mathbf{z}_i)))^T \\
&\Delta_{n_G}(x; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) = n_G^{-1} \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) \mathbf{y}_{h^{(1)}}(x_j - x) \otimes \{\mathcal{C}(x_j, \mathbf{z}_i) \text{vecs}(\mathcal{U}_i(x_j))\} - \\
&\qquad \int K_{h^{(1)}}(u - x) \mathbf{y}_{h^{(1)}}(u - x) \otimes \{\mathcal{C}(u, \mathbf{z}_i) \text{vecs}(\mathcal{U}_i(u))\} \pi(u) du, \\
&\mathcal{T}_{\mathcal{E}}(h^{(1)}, x) = 2 \sum_{i=1}^{n} \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) \mathbf{y}_{h^{(1)}}(x_j - x) \otimes [\{\mathcal{C}(x_j, \mathbf{z}_i) \text{vecs}(\mathcal{E}_i(x_j))\} \otimes \mathbf{z}_i] \\
&\mathcal{T}_{\mathcal{U}}(h^{(1)}, x) = 2 \sum_{i=1}^{n} \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) \mathbf{y}_{h^{(1)}}(x_j - x) \otimes [\{\mathcal{C}(x_j, \mathbf{z}_i) \text{vecs}(\mathcal{U}_i(x_j))\} \otimes \mathbf{z}_i] \\
&H_{h^{(1)}}(u - x) = K_{h^{(1)}}(u - x) \mathbf{y}_{h^{(1)}}(u - x).
\end{aligned}
$$

Recall that

$$
\psi(S, C) = \text{tr}[\{\log(C^{-1} S C^{-T})\}^{\otimes 2}], \tag{E.0.1}
$$

where $C = \exp(Y/2)$ and $S \in \text{Sym}^+(m)$ and $\partial_Y \psi(S, C)$ and $\partial_Y^2 \psi(S, C)$ are the first and second derivatives of $\psi(S, C)$ with respect to Y, respectively.

By Equation (4.2.4), we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x) \otimes \{\partial_Y \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x))) \otimes \mathbf{z}_i\} = 0. \tag{E.0.2}$$

Let $R(X) = \beta(X) - \beta(x) - \dot{\beta}(x)(X - x)$.

Lemma E.0.9. *The first-order derivative of $\psi(S, C)$ with respect to vecs(Y) is given by*

$$\frac{\partial \psi(S, C)}{\partial Y_j} = -2tr\{\log(C^{-1}SC^{-T})C^{-1}\frac{\partial CC^T}{\partial Y_j}C^{-T}\}, \tag{E.0.3}$$

where Y_j is the j-th element in vecs(Y).

The proof is similar to Lemma A.0.1. It follows from Lemma E.0.9 that

$$\partial_Y \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) = 2\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j) + \mathcal{U}_i(x_j)). \tag{E.0.4}$$

Lemma E.0.10. *Assume that assumptions (M2)-(M8) hold. For any random vector $\eta_j \in R^{rq}$, for $j = 1, \cdots, n_G$, if $\max_{1\le j\le n_G} \|\eta_j\| = o_p(1)$, then we have*

$$\begin{aligned}
&h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) \otimes \mathbf{z}_i^{\otimes 2}\}(x_j - x)^l \\
&= nh^{(1)l+1}\pi(x)u_l E\{\partial_Y^2 \psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}(1 + o_p(1)), \qquad \text{(E.0.5)} \\
&h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) \otimes \mathbf{z}_i^{\otimes 2}\}R(x_j)(x_j - x)^l \\
&= \frac{1}{2}nh^{(1)l+3}\pi(x)u_{l+2} E\{\partial_Y^2 \psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x)(1 + o_p(1)). \qquad \text{(E.0.6)}
\end{aligned}$$

Proof of Lemma E.0.10. We only prove (E.0.6) while the other one can be shown using the same arguments. It is easy to see that

$$\begin{aligned}
&h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) \otimes \mathbf{z}_i^{\otimes 2}\}R(x_j)(x_j - x)^l \\
&= h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) + \{\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) \\
&-\partial_Y^2 \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\}] \otimes \mathbf{z}_i^{\otimes 2}\}R(x_j)(x_j - x)^l
\end{aligned}$$

$= T_{n1} + T_{n2}$.

Let M_k denote the k-th element of $h^{(1)}K_{h^{(1)}}(X-x)\{(\partial_Y^2\psi(S(X), C(\mathbf{z}, \beta(X))) \otimes \mathbf{z}^{\otimes 2}\}R(X)(X-x)^l$. For the k-th element $(T_{n1})_k$ in the vector T_{n1}, we have

$$(T_{n1})_k = nn_G E(M_k) + O_p\left(\sqrt{nn_G E(M_k^2)}\right). \tag{E.0.7}$$

We calculate the first two moments of M_k. Applying the dominated convergence theorem and the boundedness and continuity assumptions on $\pi(.)$, $\ddot{\beta}(x)$ and $E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}$, we have

$$\begin{aligned}
&E(M_k) = E[h^{(1)}K_{h^{(1)}}(X-x)(\{\partial_Y^2\psi(S(X), C(\mathbf{z}, \beta(X))) \otimes \mathbf{z}^{\otimes 2}\}R(X))_k(X-x)^l] \\
&= \frac{1}{2}h^{(1)l+3}\int_{-1}^{1} h^{(1)}K(t)t^{l+2}E[(\{\partial_Y^2\psi(S(th^{(1)}+x), C(\mathbf{z}, \beta(th^{(1)}+x))) \otimes \mathbf{z}^{\otimes 2}\} \\
&\ddot{\beta}(\xi h^{(1)}+x))_k]\pi(th^{(1)}+x)dt \\
&= \frac{1}{2}h^{(1)l+3}u_{l+2}\pi(x)(E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x))_k\{1+o(1)\}
\end{aligned} \tag{E.0.8}$$

Since $E\{\partial_Y^2\psi(S(X), C(\mathbf{z}, \beta(X)))|X=x, Z=\mathbf{z}\}$ is bounded, there exists $d_1, d_2 > 0$ such that

$$\begin{aligned}
E(M_k^2) &= E[h^{(1)2}K_{h^{(1)}}^2(X-x)(\{\partial_Y^2\psi(S(X), C(\mathbf{z}, \beta(X))) \otimes \mathbf{z}^{\otimes 2}\}R(X))_k^2(X-x)^{2l}] \\
&= \frac{1}{4}h^{(1)2l+5}\int_{-1}^{1} h^{(1)}K^2(t)t^{2l+4}(E\{\partial_Y^2\psi(S(th^{(1)}+x), C(\mathbf{z}, \beta(th^{(1)}+x))) \otimes \mathbf{z}^{\otimes 2}\} \\
&\quad \ddot{\beta}(\xi h^{(1)}+x))_k^2\pi(th^{(1)}+x)dt \\
&\leq d_1 h^{(1)2l+5}\int_{-1}^{1} K^2(t)t^{2l+4}dt \\
&\leq d_1 d_2 h^{(1)2l+5}
\end{aligned} \tag{E.0.9}$$

Combining (E.0.7), (E.0.8) and (E.0.9), we have

$$\begin{aligned}
&(T_{n1})_k = \frac{1}{2}nn_G h^{(1)l+3}[u_{l+2}\pi(x)(E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x))_k\{1+o(1)\} \\
&+O_p(1/\sqrt{nn_G h^{(1)}})] \\
&= nn_G h^{(1)l+3}u_{l+2}\pi(x)(E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x))_k\{1+o_p(1)\}
\end{aligned}$$

That is, $T_{n1} = nn_G h^{(1)l+3} u_{l+2}\pi(x)E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x)\{1 + o_p(1)\}/2$. To prove (E.0.6), it suffices to show that $T_{n2} = o_p(nn_G h^{(1)l+3})$. Let $\triangle_{n_G} = \{\eta_1, \cdots, \eta_{n_G}\}$ where $\eta_j \in R^{rq}$ for $j = 1, \cdots, n_G$. For any given $\delta > 0$, denote $D_\delta = \{\triangle_{n_G} : \|\eta_j\| \leq \delta, \forall j \leq n_G\}$. Define

$$V(\triangle_{n_G}) = \frac{1}{nn_G h^{(1)l+3}} h^{(1)} \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j))$$
$$-\partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\} \otimes \mathbf{z}_i^{\otimes 2} R(x_j)(x_j - x)^l$$

By assumptions (M5) and (M8), as $\delta \to 0$, we have

$$\epsilon_\delta = E\{\sup_{D_\delta}\|\{\partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) - \partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\} \otimes \mathbf{z}_i\|\} = o(1).$$

Therefore, as $\delta \to 0$, we have

$$E\{\sup_{D_\delta}\|V(\triangle_{n_G})\|\} \leq \epsilon_\delta E\{\sum_{i=1}^{n}\sum_{j=1}^{n_G} h^{(1)} K_{h^{(1)}}(x_j - x)\|R(x_j)\|(x_j - x)^l \to 0$$

Since $\max_{1\leq j\leq n_G}\|\eta_j\| = o_p(1)$, we have $V(\hat{\triangle}_{n_G}) = o_p(1)$ for $\hat{\triangle}_{n_G} = \{\eta_1, \cdots, \eta_{n_G}\}$ Thus, $T_{n2} = nn_G h^{(1)l+3} V(\hat{\triangle}_{n_G}) = o_p(nn_G h^{(1)l+3})$, which completes the proof.

Lemma E.0.11. *Assume that assumptions (M2)-(M8) hold. Then we have*

$$h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) - R(x_j))) \otimes \mathbf{z}_i\}(x_j - x)^l$$
$$= \frac{1}{2} nh^{(1)l+3}\pi(x)u_{l+2}E\{\partial_Y^2\psi(S(x), C(Z, \beta(x))) \otimes Z^{\otimes 2}\}\ddot{\beta}(x)(1 + o_p(1)) +$$
$$h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) \otimes \mathbf{z}_i\}(x_j - x)^l. \tag{E.0.10}$$

Proof of Lemma E.0.11. We consider

$$J_n \equiv h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) - R(x_j))) \otimes \mathbf{z}_i\}(x_j - x)^l$$
$$= h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)(x_j - x)^l[\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) +$$
$$\partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))(-R(x_j)) +$$

$$\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) - R(x_j))) - \partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) -$$
$$\partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))(-R(x_j))\}] \otimes \mathbf{z}_i$$
$$= J_{n1} + J_{n2} + J_{n3}.$$

We will consider J_{ni} $i = 1, 2, 3$. By using the assumption (M2) and Taylor's series expansion, we have

$$\max\|R(x_j)\|\mathbf{1}(x_j - x) \leq h^{(1)} \leq \frac{1}{2}\sup_{|\xi - x| \leq h}\|\ddot{\beta}(\xi)\|h^{(1)2} = O_p(h^{(1)2}).$$

Let $D_\delta = \{\triangle_{n_G} = (\eta_1, \cdots, \eta_{n_G}) : \eta_j \in R^{rq}, \forall j \leq n_G\}$ for any $\delta > 0$. Define

$$V(\triangle_{n_G}) = \frac{1}{nn_G h^{(1)}} h^{(1)} \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)(x_j - x)^l\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j) + \eta_j)) -$$
$$\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) - \partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\eta_j\} \otimes \mathbf{z}_i.$$

By Assumption M8, as $\delta \to 0$, We have $\epsilon_\delta = E[\|\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)+\eta_j)) - \partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) - \partial_Y^2\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\text{vecs}(\eta_j)\|x_j = x] = o(\delta)$ uniformly in a neighborhood of x. Therefore,

$$E\{\sup_{D_\delta}\|V(\triangle_{n_G})\|\} \leq \epsilon_\delta \frac{1}{nn_G h^{(1)}} E\left\{h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)(x_j - x)^l\right\} = o(1).$$

Since $\max\|R(x_j)\|\mathbf{1}(x_j - x) \leq h^{(1)} \leq O_p(h^{(1)2})$, we have $V(\hat{\triangle}_{n_G}) = o_p(h^{(1)2})$, where $\hat{\triangle} = (R(x_1), \cdots, R(x_{n_G}))$. This leads to $J_{n3} = nh^{(1)}V(\hat{\triangle}_{n_G}) = o_p(nh^{(1)3})$. Applying Equation (E.0.6) in Lemma E.0.10 to J_{n2}, we get

$$J_{n2} = \frac{1}{2}nh^{(1)l+3}\pi(x)u_{l+2}E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\ddot{\beta}(x)(1 + o_p(1)),$$

which yields (E.0.10).

Lemma E.0.12. *Let*

$$\mathcal{T}_n \equiv \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x) \otimes \{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) \otimes \mathbf{z}_i\}$$

Then (i) if assumptions (M1)-(M6) and (M9) are true, then conditioning on $\mathcal{X}$, the covariance

matrix of $\{(1,0)\otimes I_{qr}\}\mathcal{T}_n/(\sqrt{n}n_G)$ *is*

$$Cov(\{(1,0)\otimes I_{qr}\}\mathcal{T}_n/(\sqrt{n}n_G)|\mathcal{X}) = 4(n_G h^{(1)})^{-1}\pi(x)v_0[\Omega_{\mathcal{E}}(x,x)+\Omega_{\mathcal{U}}(x,x)+$$
$$O_p\{(n_G h^{(1)})^{-1/2}+h^{(1)}\}] + e_n(x) + \pi(x)^2\Omega_{\mathcal{U}}(x,x) + h^{(1)2}u_2\{\Omega^{(2,0)}_{\mathcal{U}}(x,x)\pi(x)+$$
$$2\Omega^{(1,0)}_{\mathcal{U}}(x,x)\dot{\pi}_X(x)+\Omega_{\mathcal{U}}(x,x)\ddot{\pi}_X(x)\}\pi(x)+o(h^{(1)2}), \tag{E.0.11}$$

where $e_n(x) = O_p\{(n_G h^{(1)})^{-1/2}\}$ *with* $E\{e_n(x)\} = \mathbf{0}$.

(ii) if assumptions (M1)-(M6) and (M9) are true, then conditioning on $\mathcal{X}$, $\{(1,0)\otimes I_{qr}\}\mathcal{T}_n/(\sqrt{n}n_G)$ *is asymptotically normal with mean zero and covariance matrix* $\Sigma_{\mathcal{T}} = 4\pi(x)^2\Omega_{\mathcal{U}}(x,x)$.

Proof of Lemma E.0.12. Let

$$\mathcal{T}_{ni} \equiv \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x)\otimes\{\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i\}.$$

Note that conditioning on $\mathcal{X}$, $\mathcal{T}_{ni}$ are independent and $\mathcal{T}_n = \sum_{i=1}^n \mathcal{T}_{ni}$. It follows from Equation (4.2.1) and Lemma E.0.9 that $E(\mathcal{T}_{ni}|\mathcal{X}) = \mathbf{0}$. Conditioning on $\mathcal{X}$, the covariance matrix of $Cov(\mathcal{T}_n|\mathcal{X}) = \sum_{i=1}^n Cov(\mathcal{T}_{ni}|\mathcal{X}) = \sum_{i=1}^n E(\mathcal{T}_{ni}^{\otimes 2}|\mathcal{X})$. Note that $E(\mathcal{T}_{ni}^{\otimes 2}|\mathcal{X}) = E\{E(\mathcal{T}_{ni}^{\otimes 2}|\mathcal{X},\mathbf{Z})\}$. Let $H_{h^{(1)}}(u-x) = K_{h^{(1)}}(u-x)\mathbf{y}_{h^{(1)}}(u-x)$ with $\mathbf{y}_{h^{(1)}}(u-x) = (1,(u-x)/h^{(1)})^T$. With some calculations, we have

$$E(\mathcal{T}_{ni}^{\otimes 2}|\mathcal{X},\mathbf{Z}) = \sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j-x)H_{h^{(1)}}(x_{j'}-x)^T\otimes$$
$$E\{\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\partial_Y\psi(S_i(x_{j'}),C(\mathbf{z}_i,\beta(x_{j'})))^T|\mathcal{X},\mathbf{Z}\}\otimes\mathbf{z}_i^{\otimes 2}.$$

Recall that $M^{(k,k')}(v,u,\mathbf{z}) = \Lambda\text{vecs}(\mathcal{C}^{(k)}(v,\mathbf{z}))(\Lambda\text{vecs}(\mathcal{C}^{(k')}(u,\mathbf{z})))^T$ and $M_{\mathcal{E}}(v,u,\mathbf{z})$ and $M_{\mathcal{U}}(v,u,\mathbf{z})$ are $q\times q$ matrix with the (k,k') entries $\text{tr}\{\Sigma_{\mathcal{E}}(v,u)M^{(k,k')}(v,u),\mathbf{z})\}$ and $\text{tr}\{\Sigma_{\mathcal{U}}(v,u)M^{(k,k')}(v,u,\mathbf{z})\}$, respectively. Let $\mathbf{c}(x_j,\mathbf{z}_i)$ be a $q\times 1$ vector with the k-th element $\text{tr}[\{\mathcal{E}_i(x_j)+\mathcal{U}_i(x_j)\}\mathcal{C}^{(k)}(x_j,\mathbf{z}_i)]$. By Lemma E.0.9

$$E\{\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\partial_Y\psi(S_i(x_{j'}),C(\mathbf{z}_i,\beta(x_{j'})))^T|\mathcal{X},\mathbf{Z}\}$$
$$= 4E\{\mathbf{c}(x_j,\mathbf{z}_i)\mathbf{c}(x_{j'},\mathbf{z}_i)^T|\mathcal{X},\mathbf{Z}\}$$
$$= 4(M_{\mathcal{E}}(x_j,x_{j'},\mathbf{z}_i)+M_{\mathcal{U}}(x_j,x_{j'},\mathbf{z}_i))$$

It follows that

$$E(\mathcal{T}_n^{\otimes 2}|\mathcal{X},\mathbf{Z}) = 4n\sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_{j'} - x)^T \otimes E\{(M_{\mathcal{E}}(x_j, x_{j'}, \mathbf{z}) + M_{\mathcal{U}}(x_j, x_{j'}, \mathbf{z})) \otimes \mathbf{z}^{\otimes 2}\}$$
$$\equiv 4n(A_{\mathcal{E}}(x) + A_{\mathcal{U}}(x)) \quad \text{(E.0.12)}$$

We approximate $A_{\mathcal{E}}(x)$ and $A_{\mathcal{U}}(x)$ as follows. Since $A_{\mathcal{E}}(x) = E\{A_{\mathcal{E}}(x)\} + O_p(\sqrt{\text{Var}(A_{\mathcal{E}}(x))})$, we have

$$A_{\mathcal{E}}(x) = n_G h^{(1)-1}[\pi(x)\begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes \Omega_{\mathcal{E}}(x,x) + O_p\{(n_G h^{(1)})^{-1/2} + h^{(1)}\}], \quad \text{(E.0.13)}$$

where $\Omega_{\mathcal{E}}(v,u) = E(M_{\mathcal{E}}(v,u,\mathbf{z}) \otimes \mathbf{z}^{\otimes 2})$.

Moreover, $A_{\mathcal{U}}(x)$ can be written as the sum of $A_{\mathcal{U}}^{(1)}(x) = \sum_{j=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_j - x)^T \otimes E(M_{\mathcal{U}}(x_j, x_j, \mathbf{z}) \otimes \mathbf{z}^{\otimes 2})$ and $A_{\mathcal{U}}^{(2)}(x) = \sum_{j \neq j'}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_{j'} - x)^T \otimes E(M_{\mathcal{U}}(x_j, x_{j'}, \mathbf{z}) \otimes \mathbf{z}^{\otimes 2})$. Particularly, $A_{\mathcal{U}}^{(2)}(x)/\{n_G(n_G - 1)\}$ is a U-statistic (DasGupta, 2008). Similar to $A_{\mathcal{E}}(x)$, it can be shown that

$$A_{\mathcal{U}}^{(1)}(x) = n_G h^{(1)-1}[\pi(x)\begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes E(M_{\mathcal{U}}(x,x,\mathbf{z}) \otimes \mathbf{z}^{\otimes 2})O_p\{(n_G h^{(1)})^{-1/2} + h^{(1)}\}].$$

For $A_{\mathcal{U}}^{(2)}(x)$, we define three $2qr \times 2qr$ matrices: $U_{\mathcal{U}}(x) = (U_{\mathcal{U},ml}(x)) = A_{\mathcal{U}}^{(2)}(x)/\{n_G(n_G - 1)\}$, $\theta(x) = (\theta_{ml}(x)) = E(U_{\mathcal{U}}(x))$ and $P_{\mathcal{U}}(v) = (P_{\mathcal{U},ml}(v)) = \int_0^{L_0} H_{h^{(1)}}(v - x)H_{h^{(1)}}(u - x)^T \otimes E(M_{\mathcal{U}}(v,u,\mathbf{z}) \otimes \mathbf{z}^{\otimes 2})\pi(u)du$, where $m = (m' - 1)qr + (k-1)r + m_1$ and $l = (l' - 1)qr + (k' - 1)r + l_1$ for $m', l' = 1, 2$ and $m_1, l_1 = 1, \cdots, r$. By using the Hajek projection, we have

$$U_{\mathcal{U},ml}(x) = \theta_{ml}(x) + \frac{2}{n_G}\sum_{j=1}^{n_G}(P_{\mathcal{U},ml}(x) - \theta_{ml}(x)) + \tilde{E}_{n,ml}(x) \text{ for } m, l = 1, \cdots, 2qr,$$

where $2\sum_{j=1}^{n_G}(P_{\mathcal{U},ml}(x) - \theta_{ml}(x))/n_G$ is the projection of $U_{\mathcal{U},ml}(x) - \theta_{ml}(x)$ onto the set of all statistics the form $\sum_{j=1}^{n_G} f_j(x)$. Thus, with some calculations, we have

$$\text{Var}(\tilde{E}_{n,ml}(x)) = \text{Var}(U_{\mathcal{U},ml}(x) - \theta_{ml}(x)) - \text{Var}\{\frac{2}{n_G}\sum_{j=1}^{n_G}(P_{\mathcal{U},ml}(x) - \theta_{ml}(x))\} = O((n_G h^{(1)})^{-2}).$$

Let $\Omega_{\mathcal{U}}(v,u) = (\Omega_{\mathcal{U},kk',m_1l_1}(v,u))$ with $\Omega_{\mathcal{U},kk',m_1l_1}(v,u) = E[\mathrm{tr}(\Sigma_{\mathcal{U}}(v,u)M^{(k,k')}(v,u,\mathbf{z}))(\mathbf{z}^{\otimes 2})_{m_1,l_1}]$. As $h^{(1)} \to 0$, it follows from Taylor's expansion that

$$
\begin{aligned}
&\theta_{ml}(x) + o(h^{(1)2}) = \pi(x)^2 u_{m'-1}u_{l'-1}\Omega_{\mathcal{U},kk',m_1l_1}(x,x) + \\
&h^{(1)}\pi(x)^2\Omega^{(1,0)}_{\mathcal{U},kk',m_1l_1}(x,x)(u_{m'}u_{l'-1} + u_{m'-1}u_{l'}) + \\
&h^{(1)}\Omega_{\mathcal{U},kk',m_1l_1}(x,x)\dot{\pi}_X(x)\pi(x)(u_{m'}u_{l'-1} + u_{m'-1}u_{l'}) + \\
&0.5h^{(1)2}u_{m'+1}u_{l'-1}\{\Omega^{(2,0)}_{\mathcal{U},kk',m_1l_1}(x,x)\pi(x) + 2\Omega^{(1,0)}_{\mathcal{U},kk',m_1l_1}(x,x)\dot{\pi}_X(x) + \\
&\Omega_{\mathcal{U},kk',m_1l_1}(x,x)\ddot{\pi}_X(x)\}\pi(x) + h^{(1)2}u_{m'}u_{l'}\{\Omega^{(1,1)}_{\mathcal{U},kk',m_1l_1}(x,x)\pi(x)^2 + \\
&2\Omega^{(1,0)}_{\mathcal{U},kk',m_1l_1}(x,x)\dot{\pi}_X(x)\pi(x) + \Omega_{\mathcal{U},kk',m_1l_1}(x,x)\dot{\pi}^2_X(x)\} + \\
&0.5h^{(1)2}u_{m'-1}u_{l'+1}\{\Omega^{(2,0)}_{\mathcal{U},kk',m_1l_1}(x,x)\pi(x) + 2\Omega^{(1,0)}_{\mathcal{U},kk',m_1l_1}(x,x)\dot{\pi}_X(x) + \\
&\Omega_{\mathcal{U},kk',m_1l_1}(x,x)\ddot{\pi}_X(x)\}\pi(x)
\end{aligned}
$$

Let $m_0 = (k-1)r + m_1$ and $l_0 = (k^{'}-1)r + l_1$ and $e_n = (e_{n,m_0l_0}(x))$ with $e_{n,m_0l_0}(x) = 2\sum_{j=1}^{n_G}\{P_{\mathcal{U},m_0l_0}(x) - \theta_{m_0l_0}(x)\}/n_G + \tilde{E}_{n,m_0l_0}(x)$. We have $e_{n,m_0l_0}(x) = O_p((n_Gh^{(1)})^{-1/2})$ and thus

$$
\begin{aligned}
&((1,0)\otimes I_{qr})A_{\mathcal{U}}(x)((1,0)\otimes I_{qr})^T = n_G^2e_n(x) + n_G^2\pi(x)^2\Omega_{\mathcal{U}}(x,x) + n_G^2o_p(h^{(1)2}) \\
&n_Gh^{(1)-1}\{\pi(x)v_0\Omega_{\mathcal{U}}(x,x) + O_p((n_Gh^{(1)})^{-1/2}) + O_p(h^{(1)})\} + \\
&n_G^2h^{(1)2}u_2\{\Omega^{(2,0)}_{\mathcal{U}}(x,x)\pi(x) + 2\Omega^{(1,0)}_{\mathcal{U}}(x,x)\dot{\pi}_X(x) + \Omega_{\mathcal{U}}(x,x)\ddot{\pi}_X(x)\}\pi(x) \qquad \text{(E.0.14)}
\end{aligned}
$$

Substituting (E.0.13) and (E.0.14) into (E.0.12), we can obtain (E.0.11).

Finally, we will show that the sequence $\mathcal{T}_n/(\sqrt{n}n_G)$ satisfies the Linderberg-Feller condition:

$$
\sum_{i=1}^{n} E\left\{\left\|\frac{\mathcal{T}_{ni}}{\sqrt{n}n_G}\right\|^2 \mathbf{1}\left(\left\|\frac{\mathcal{T}_{ni}}{\sqrt{n}n_G}\right\| > \epsilon\right) \mid \mathcal{X}\right\} \to 0 \text{ for any } \epsilon > 0.
$$

Let $\mathcal{T}_{nij} \equiv K_{h^{(1)}}(x_j - x)\begin{pmatrix} 1 \\ \frac{x_j - x}{h^{(1)}} \end{pmatrix} \otimes (\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) \otimes \mathbf{z}_i)$ and $\|\mathcal{T}^{'}_{nij}\|^2 = \|\partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))\|^2\{1 + (x_j - x)^2/h^{(1)2}\}\|\mathbf{z}_i\|^2$. Then by partial sums of moment inequality, we have

$$
E\left\{\left\|\frac{\mathcal{T}_{ni}}{\sqrt{n}n_G}\right\|^2 \mathbf{1}\left(\left\|\frac{\mathcal{T}_{ni}}{\sqrt{n}n_G}\right\| > \epsilon\right) |\mathcal{X}, \mathbf{Z}\right\} \le \mathcal{K}_i \equiv \frac{E\left\{\|\sum_{j=1}^{n_G}\mathcal{T}_{nij}\|^{2+b}|\mathcal{X}, \mathbf{Z}\right\}}{(nn_G^2)^{1+b/2}\epsilon^b}
$$

$$
\begin{aligned}
&\leq \frac{n_G^{1+b}\sum_{j=1}^{n_G} E\left\{\|\mathcal{T}_{nij}\|^{2+b}|\mathcal{X},\mathbf{Z}\right\}}{(nn_G^2)^{1+b/2}\epsilon^b} \\
&= \frac{n_G^{1+b}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j-x)^{2+b}E\left\{\|\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\|^{2+b}|\mathcal{X},\mathbf{Z}\right\}}{(nn_G^2)^{1+b/2}\epsilon^b\|\mathbf{z}_i\|^{2+b}\{1+(x_j-x)^2/h^{(1)2}\}^{-1-b/2}} \\
&= \frac{(v_0+v_2)^{1+b/2}E\left\{\|\partial_Y\psi(S(X),C(\mathbf{z}_i,\beta(X)))\|^{2+b}|X=x,\mathbf{Z}\right\}}{n^{1+b/2}\epsilon^b\|\mathbf{z}_i\|^{2+b}}
\end{aligned}
$$

Combining assumptions (M5) and (M9)yields that there is a constant $d>0$ such that $\sum_{i=1}^{n}\sum_{j=1}^{n_G}\mathcal{K}_i \leq dn/n^{1+b/2}\to 0$. Thus, it follows from the Linderberg-Feller theorem that $\mathcal{T}_n/\sqrt{n}n_G$ is asymptotically normal with mean $\mathbf{0}$ and covariance $\Sigma_{\mathcal{T}}$.

Lemma E.0.13. *If assumptions (M1)-(M9) are true, then conditioning on $\mathcal{X}$, $X_n(x)$ is asymptotically normal with mean zero and covariance matrix* $\Sigma_\beta(x,x^{'}) = 4\left[E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-1}$ $\Omega_{\mathcal{U}}(x,x^{'})\left[E\{\partial_Y^2\psi(S(x^{'}),C(\mathbf{z},\beta(x^{'})))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-T}$.

Proof of Lemma E.0.13. Let $-\hat{\eta}(x_j) = R(x_j)-(\hat{\beta}(x)-\beta(x))-(\hat{\dot{\beta}}(x)-\dot{\beta}(x))(x_j-x)$. By Equation (4.2.4), we have

$$
\begin{aligned}
&h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j-x)\mathbf{y}_{h^{(1)}}(x_j-x)\otimes(\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i \\
&+\{\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^T\otimes\mathbf{z}_i\}\hat{\eta}(x_j)+[\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)+\hat{\eta}(x_j)))\otimes\mathbf{z}_i \\
&-\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i-\{\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^T\otimes\mathbf{z}_i\}\hat{\eta}(x_j)]) \\
&=0. \qquad\qquad (E.0.15)
\end{aligned}
$$

Note that the second term on the left hand side of (E.0.15) is

$$
\begin{aligned}
&-h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j-x)\left\{\mathbf{y}_{h^{(1)}}(x_j-x)\otimes\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^{\otimes 2}\right\}R(x_j) \\
&+h^{(1)}\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j-x)\left\{\mathbf{y}_{h^{(1)}}(x_j-x)^{\otimes 2}\otimes\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^{\otimes 2}\right\} \\
&\begin{pmatrix}\hat{\beta}(x)-\beta(x)\\ h^{(1)}\{\hat{\dot{\beta}}(x)-\dot{\beta}(x)\}\end{pmatrix}\equiv L_1+L_2 \qquad\qquad (E.0.16)
\end{aligned}
$$

By Lemma E.0.10, L_1 and L_2 can, respectively, be approximated by

$$\begin{aligned} L_1 &= -\frac{1}{2}\pi(x)nn_G h^{(1)3}\left[\begin{pmatrix} u_2 \\ 0 \end{pmatrix} \otimes E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{Z}^{\otimes 2}\}\right]\ddot{\beta}(x)(1+o_p(1)), \\ L_2 &= \pi(x)nn_G h^{(1)}\left[\begin{pmatrix} 1 & 0 \\ 0 & u_2 \end{pmatrix} E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{Z}^{\otimes 2}\}\right]\begin{pmatrix} \hat{\beta}(x) - \beta(x) \\ h^{(1)}\{\dot{\hat{\beta}}(x) - \dot{\beta}(x)\} \end{pmatrix} \\ & \quad (1+o_p(1)). \end{aligned}$$

By the same arguments as those in Lemma E.0.11 and Assumption (M8), we have that the third term on the left hand side of Equation (E.0.15) is given by $o_p(nn_G h^{(1)})\{\|h^{(1)2}\| + \|\hat{\beta}(x) - \beta(x)\| + h^{(1)}\|\dot{\hat{\beta}}(x) - \dot{\beta}(x)\|\}$, which converges to zero in probability faster than the second term on the left hand side of Equation (E.0.15).

Let

$$\begin{aligned} B &= \frac{h^{(1)2}}{2}\left[\begin{pmatrix} 1 & 0 \\ 0 & u_2 \end{pmatrix} \otimes E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]^{-1} \\ & \quad \left[\begin{pmatrix} u_2 \\ 0 \end{pmatrix} \otimes E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]\ddot{\beta}(x)(1+o_p(1)) \end{aligned}$$

Then from Equation (E.0.15), we have

$$\begin{aligned} \begin{pmatrix} \hat{\beta}(x) - \beta(x) \\ h^{(1)}\{\dot{\hat{\beta}}(x) - \dot{\beta}(x)\} \end{pmatrix} &= B - \{nn_G h^{(1)}\pi(x)\}^{-1}\left[\begin{pmatrix} 1 & 0 \\ 0 & u_2 \end{pmatrix} \otimes E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]^{-1} \\ & \quad \mathcal{T}_n(1+o_p(1)). \end{aligned}$$

So, we have

$$\begin{aligned} \sqrt{n}\{\hat{\beta}(x) - \beta(x) - 0.5u_2 h^{(1)2}\ddot{\beta}(x)\} &= -\sqrt{n}\{nn_G h^{(1)}\pi(x)\}^{-1}\left[E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]^{-1} \\ & \quad \{(1,0) \otimes I_{qr}\}\mathcal{T}_n(1+o_p(1)). \end{aligned} \tag{E.0.17}$$

Finally, it follows from Equation E.0.17, Lemma E.0.12 and the Slutsky's theorem that

$$\sqrt{n}\{\hat{\beta}(x) - \beta(x) - 0.5u_2 h^{(1)2}\ddot{\beta}(x)\} \to^L N(\mathbf{0}, \Sigma_\beta(x, x^{'})).$$

Lemma E.0.14. *If assumptions (M1)-(M6) are true, then we have*

$$\sup_{x\in[0,L_0]} n^{-1/2}h^{(1)}|T_{\mathcal{E}}(h^{(1)}, x)| = O_p(\sqrt{n_G h^{(1)}|\log h^{(1)}|}) = o_p(n_G h^{(1)}). \tag{E.0.18}$$

Proof of Lemma E.0.14. We define $F_n(x_j) = 2n^{-1/2}\sum_{i=1}^n[\{\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j))\} \otimes \mathbf{z}_i]$,

$$\begin{aligned}
n^{-1/2}h^{(1)}T_{\mathcal{E}}(h^{(1)}, x) &= h^{(1)}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x) \otimes F_n(x_j), \\
n^{-1/2}h^{(1)}T'_{\mathcal{E}}(h^{(1)}, x) &= h^{(1)}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x) \otimes F_n(x_j)\mathbf{1}(||F_n(x_j)||_2 \le \gamma_{n_G}), \\
n^{-1/2}h^{(1)}\tilde{T}_{\mathcal{E}}(h^{(1)}, x) &= n^{-1/2}h^{(1)}\{T'_{\mathcal{E}}(h^{(1)}, x) - E[T'_{\mathcal{E}}(h^{(1)}, x)|\mathcal{X}, \mathbf{Z}]\},
\end{aligned}$$

where γ_{n_G} is a positive scalar. The proof of Lemma E.0.14 consists of three steps. In Step 1, we show that

$$\sup_{x\in[0,L_0]} n^{-1/2}h^{(1)}||T_{\mathcal{E}}(h^{(1)}, x) - \tilde{T}_{\mathcal{E}}(h^{(1)}, x)||_2 = o_p(\sqrt{n_G h^{(1)}|\log h^{(1)}|}). \tag{E.0.19}$$

In Step 2, we define an equally-spaced grid $\tilde{\mathcal{X}} = \{\tilde{x}_l = lh^{(1)} : l = 0, \cdots, L_0 h^{(1)-1}\}$ and then show that

$$\max_l h^{(1)}||n^{-1/2}\tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)||_2 = O_p(\sqrt{n_G h^{(1)}|\log h^{(1)}|}). \tag{E.0.20}$$

In Step 3, we show that

$$\max_l \sup_{x\in[\tilde{x}_{l-1}, \tilde{x}_l]} n^{-1/2}h^{(1)}||\tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_{l-1}) - \tilde{T}_{\mathcal{E}}(h^{(1)}, x)||_2 = O(\sqrt{n_G h^{(1)}|\log h^{(1)}|}). \tag{E.0.21}$$

Combing equations (E.0.19)-(E.0.21), we can finish the proof of Lemma E.0.14.

In Step 1, to show (E.0.19), we note that $n^{-1/2}h^{(1)}||T_{\mathcal{E}}(h^{(1)}, x) - \tilde{T}_{\mathcal{E}}(h^{(1)}, x)||_2$ is upper bounded

by

$$d\sum_{j=1}^{n_G}||F_n(x_j)||_2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G}) + d\sum_{j=1}^{n_G}E[||F_n(x_j)||_2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}],$$

where d is a positive scalar. Let $\gamma_{n_G} = \delta(n_G/|\log h^{(1)}|)^{1/q_1}$. It follows from the partial sums moment inequality (DasGupta, 2008) that as $\gamma_{n_G} \to \infty$, we have

$$\max_j E[||F_n(x_j)||_2^{q_1}\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}] = o_p(1).$$

Let $c_k(x_j, \mathbf{z}_i)$ be the k-th element of the $qr \times 1$ vector $\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j))\} \otimes \mathbf{z}_i$ for $k = 1, \cdots, qr$. For any $c > 0$ with $q_1 + c < 4$, we have

$$\begin{aligned}
&\max_j E[||F_n(x_j)||_2^{q_1}\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}] \\
&\leq \max_j E[\|n^{-1/2}\sum_{i=1}^{n}\{\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j))\} \otimes \mathbf{z}_i\|^{q_1+c}|\mathcal{X}, \mathbf{Z}]/\gamma_{n_G}^c \\
&\leq \max_j n^{-(q_1+c)/2}c(q_1)n^{(q_1+c)/2-1}(qr)^{(q_1+c)/2-1}\sum_{i=1}^{n}\sum_{k=1}^{qr}E[|c_k(x_j, \mathbf{z}_i)|^{q_1+c}]/\gamma_{n_G}^c = o(1).
\end{aligned}$$

Therefore, we can show that

$$\begin{aligned}
&\sum_{j=1}^{n_G}E[||F_n(x_j)||_2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}] \\
&\leq \sum_{j=1}^{n_G}E[||F_n(x_j)||_2^{q_1}\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}]/\gamma_{n_G}^{q_1-1} \\
&\leq o(1)n_G^{1/q_1}|\log h^{(1)}|^{1-1/q_1} \leq o(\sqrt{n_G h^{(1)}|\log h^{(1)}|}).
\end{aligned} \tag{E.0.22}$$

Furthermore, we have

$$\begin{aligned}
&\text{Var}(\sum_{j=1}^{n_G}||F_n(x_j)||_2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})) \\
&\leq E_Z E_{\mathcal{E}}\{[\sum_{j=1}^{n_G}||F_n(x_j)||_2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})]^2|\mathcal{X}, \mathbf{Z}\} \\
&\leq E_Z E_{\mathcal{E}}\{[\sum_{j=1}^{n_G}d_2||F_n(x_j)||_2^2\mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})]|\mathcal{X}, \mathbf{Z}\}
\end{aligned}$$

$$
\begin{aligned}
&\leq E_Z E_{\mathcal{E}}[\sum_{j=1}^{n_G} d_2 ||F_n(x_j)||_2^{q_1} \mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}]/\gamma_{n_G}^{q_1-2} \\
&\leq \max_j E[||F_n(x_j)||_2^{q_1} \mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})] n_G/\gamma_{n_G}^{q_1-2} \\
&\leq \max_j E[||F_n(x_j)||_2^{q_1} \mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G})] n_G h^{(1)} = o(n_G h^{(1)}|\log h^{(1)}|), \qquad \text{(E.0.23)}
\end{aligned}
$$

where d_2 is a finite universal constant (DasGupta, 2008).

Therefore, combing equations (E.0.22) and (E.0.23), we have

$$
\sum_{j=1}^{n_G} ||F_n(x_j)||_2 \mathbf{1}(||F_n(x_j)||_2 \geq \gamma_{n_G}) = o_p(\sqrt{n_G h^{(1)}|\log h^{(1)}|}),
$$

which yields (E.0.19).

In Step 2, to prove (E.0.20), we note that

$$
\begin{aligned}
&h^{(1)}||K_{h^{(1)}}(x_j - x)\{F_n(x_j)\mathbf{1}(||F_n(x_j)||_2 \leq \gamma_{n_G}) - \qquad \text{(E.0.24)} \\
&E[\mathbf{y}_{h^{(1)}}(x_j - s) \otimes F_n(x_j)\mathbf{1}(||F_n(x_j)||_2 \leq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}]\}||_2 \\
&\leq d_3(n_G/|\log h^{(1)}|)^{1/q_1} \leq d_3\sqrt{n_G h^{(1)}/|\log h^{(1)}|}, \qquad \text{(E.0.25)}
\end{aligned}
$$

where d_3 is a positive scalar.

Furthermore, let $E_{\mathbf{X}}$ denote the expectation to x_j, we have

$$
\begin{aligned}
&\text{Var}(\sum_{j=1}^{n_G} h^{(1)} K_{h^{(1)}}(x_j - x)\mathbf{y}_{h^{(1)}}(x_j - x) \otimes F_n(x_j)\mathbf{1}(||F_n(x_j)||_2 \leq \gamma_{n_G})|\mathbf{Z}) \\
&\leq \sum_{j=1}^{n_G} E_{\mathcal{X}}\{h^{(1)2} K_{h^{(1)}}(x_j - x)^2 E[\mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2} \otimes F_n(x_j)^{\otimes 2}\mathbf{1}(||F_n(x_j)||_2 \leq \gamma_{n_G})|\mathcal{X}, \mathbf{Z}]\} \\
&\leq \sum_{j=1}^{n_G} h^{(1)2} E_{\mathcal{X}}[K_{h^{(1)}}(x_j - x)^2 n^{-1} \sum_{i=1}^{n} \mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2} \otimes \{\mathcal{C}(x_j, \mathbf{z}_i)\Sigma_{\mathcal{E}}(x_j, x_j)\mathcal{C}(x_j, \mathbf{z}_i)^T \otimes \mathbf{z}_i^{\otimes 2}\}] \\
&= O_p(n_G h^{(1)}).
\end{aligned}
$$

Therefore, by applying Bernstein inequality to each component of $h^{(1)} n^{-1/2} \tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)$ (van der Vaar and Wellner, 1996), we can prove (E.0.20). For instance, let $\mathbf{e}_1$ be a $\dim(\tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)) \times 1$

vector with the first element 1 and zero otherwise, we have

$$P(\max_l |\mathbf{e}_1 h^{(1)} n^{-1/2} \tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)| > t) \leq E[\exp(-\frac{1}{2}\frac{t^2}{v(\mathbf{Z}) + t d_3 \sqrt{n_G h^{(1)}/|\log h^{(1)}|}/3})|\mathbf{Z}], \quad \text{(E.0.26)}$$

where t is a positive scalar, and $v(\mathbf{Z}) \geq \text{Var}(\mathbf{e}_1 h^{(1)} n^{-1/2} \tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)|\mathbf{Z})$. By setting $t = C'\sqrt{n_G h^{(1)}|\log h^{(1)}|}$ for large $C' > 0$, we can show that the right hand side of (E.0.26) is at the order of $h^{(1)C''}$, where C'' is a positive scalar. Thus,

$$P(\max_l |\mathbf{e}_1 h^{(1)} n^{-1/2} \tilde{T}_{\mathcal{E}}(h^{(1)}, \tilde{x}_l)| > C'\sqrt{n_G h^{(1)}|\log h^{(1)}|}) \to 0 \quad \text{as} \quad h^{(1)} \to 0.$$

In Step 3, we focus on the first component of $\mathbf{y}_{h^{(1)}}(x_j - x)$. We first consider a function class, denoted by $\mathcal{F}_l$, which is given by

$$\mathcal{F}_l = \{w_l(X; x) = h^{(1)}[K_{h^{(1)}}(X - \tilde{x}_l) - K_{h^{(1)}}(X - x)]F_n(X)\mathbf{1}(||F_n(X)||_2 \leq \gamma_{n_G}) : x \in [\tilde{x}_{l-1}, \tilde{x}_l]\}.$$

It follows from Assumption M4 and γ_{n_G} that $\mathcal{F}_j$ is a pointwise measurable class of functions and $\sup_{x\in[0,L_0]} |w_l(X; x)| \leq d_4\gamma_{n_G} \leq d_4'\sqrt{n_G h^{(1)}/|\log h^{(1)}|}$. Let $||\phi||_B = \sup_{z\in B} |\phi(z)|$ for any real valued function ϕ defined on a set B and $\tau_1, \cdots, \tau_{n_G}$ be a sequence of independent Rademacher random variables independent of observed data. It follows from an inequality of Talagrand (Talagrand, 1994; Einmahl and Mason, 2000) that conditional on $\mathbf{Z}$, we have for suitable finite constants $A_1, A_2 > 0$

$$\begin{aligned} &P\{||\sum_{j=1}^{n_G}[w_l(x_j; x) - E[w_l(x_j; x)|\mathbf{Z}]||_{\mathcal{F}_l} \geq A_1(E[||\sum_{j=1}^{n_G}\tau_j w_l(x_j; x)||_{\mathcal{F}_l}|\mathbf{Z}] + t)|\mathbf{Z}\} \\ &\leq 2[\exp(-A_2 t^2/(n_G V_{\mathcal{F}_l}(\mathbf{Z}))) + \exp(-A_2 t/(d_4'\sqrt{n_G h^{(1)}/|\log h^{(1)}|}))], \end{aligned} \quad \text{(E.0.27)}$$

where $V_{\mathcal{F}_l}(\mathbf{Z}) = \sup_{x\in[\tilde{x}_{l-1},\tilde{x}_l]} \text{Var}(w_l(X; x)|\mathbf{Z})$. It can be shown that

$$\begin{aligned} V_{\mathcal{F}_l}(\mathbf{Z}) &\leq \sup_{x\in[\tilde{x}_{l-1},\tilde{x}_l]} E_X\{h^{(1)2}[K_{h^{(1)}}(X - \tilde{x}_l) - K_{h^{(1)}}(X - x)]^2 E[F_n(X)^{\otimes 2}|\mathbf{X}, \mathbf{Z}]\} \\ &\leq d_5 h^{(1)} n^{-1} \sum_{i=1}^{n} E_X\{\mathcal{C}(X, \mathbf{z}_i)\Sigma_{\mathcal{E}}(X, X)\mathcal{C}(X, \mathbf{z}_i)^T\} \otimes \mathbf{z}_i^{\otimes 2}. \end{aligned}$$

where d_5 is a positive scalar. By setting $t = d_6\sqrt{n_G h^{(1)}|\log h^{(1)}|}$ for large d_6, we can show that $A_2 t^2/(n_G V_{\mathcal{F}_{j'}}(\mathbf{Z})) = d_6'|\log h^{(1)}|$ and $A_2 t/(d_4'\sqrt{n_G h^{(1)}/|\log h^{(1)}|}) = d_6''|\log h^{(1)}|$. Moreover, it follows from assumption M4 that $\mathcal{F}_{j'}$ is a pointwise measurable Vapnik and Cervonenkis (VC) class (van der Vaar and Wellner, 1996). By using Proposition A.1 of (Einmahl and Mason, 2000), we can show that $\max_l E[||\sum_{j=1}^{n_G}\tau_j w_l(x_j;x)||_{\mathcal{F}_l}|\mathbf{Z}] \leq O(\sqrt{n_G h^{(1)}|\log h^{(1)}|})$. This yields (E.0.21).

Let $\Pi_{n_G}(\cdot)$ be the sampling distribution function based on $\mathcal{X}$ and $\Pi(\cdot)$ be the distribution function of x_j.

Lemma E.0.15. *Let $(\mathcal{C}(u,\mathbf{z}_i)vecs(\mathcal{E}_i(u))_{(k)}$ be the k-th element of the $q\times 1$ vector $\mathcal{C}(u,\mathbf{z}_i)vecs(\mathcal{E}_i(u)$ If assumptions (M1)-(M6) are true, then for any $r \geq 0$ and i, we have*

$$\sup_{x\in[0,L_0]}\left|\int K_{h^{(1)}}(u-x)\frac{(u-x)^r}{h^{(1)r}}d[\Pi_{n_G}(u)-\Pi(u)]\right| = O_p(n_G^{-1/2}h^{(1)-1}), \tag{E.0.28}$$

$$\sup_{x\in[0,L_0]}\left|\int K_{h^{(1)}}(u-x)\frac{(u-x)^r}{h^{(1)r}}(\mathcal{C}(u,\mathbf{z}_i)vecs(\mathcal{E}_i(u)))_{(k)}d\Pi_{n_G}(u)\right| = O_p(\sqrt{\frac{|\log(h^{(1)})|}{n_G h^{(1)}}}). \tag{E.0.29}$$

Proof of Lemma E.0.15. Equation (E.0.28) follows from the integration by parts, while (E.0.29) can be proved by using the same arguments of Lemma E.0.14.

Lemma E.0.16. *If assumptions (M1)-(M6) are true, then we have*

$$\sup_{x\in[0,L_0]}\|n^{-1/2}\sum_{i=1}^{n}\Delta_{n_G}(x;\mathcal{U}_i,\mathbf{z}_i,h^{(1)})\otimes\mathbf{z}_i\| = O_p((n_G h^{(1)})^{-1/2}+h^{(1)-1}n_G^{-1/2}). \tag{E.0.30}$$

Proof of Lemma E.0.16. Let $\tau_1,\cdots,\tau_n$ be a sequence of independent Rademacher random variables independent of observed data. It follows from the symmetrization inequality (van der Vaar and Wellner, 1996) that

$$E||n^{-1/2}\sum_{i=1}^{n}\Delta_{n_G}(x;\mathcal{U}_i,\mathbf{z}_i,h^{(1)})\otimes\mathbf{z}_i||_{[0,L_0]} \leq 2E\{E_\tau||n^{-1/2}\sum_{i=1}^{n}\Delta_{n_G}(x;\mathcal{U}_i,\mathbf{z}_i,h^{(1)})\otimes\tau_i\mathbf{z}_i||_{[0,L_0]}]\}.$$

We consider a function class, denoted by $\mathcal{F}_{\mathcal{U}}$, which is given by

$$\mathcal{F}_{\mathcal{U}} = \{f(\mathbf{z},\mathcal{U};x) = \Delta_{n_G}(x;\mathcal{U},\mathbf{z},,h^{(1)})\otimes\mathbf{z} : x\in[0,L_0]\}.$$

Let $N(\epsilon, \mathcal{F}_{\mathcal{U}}, d_Q)$ be the minimal number of d_Q-balls with radius ϵ needed to cover $\mathcal{F}_{\mathcal{U}}$. It follows from that for any fixed $\{(\mathbf{z}_1, \mathcal{U}_1), \cdots, (\mathbf{z}_n, \mathcal{U}_n)\}$, we have

$$
\begin{aligned}
&E_\tau || n^{-1/2} \sum_{i=1}^{n} \Delta_{n_G}(x; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \tau_i \mathbf{z}_i ||_{[0, L_0]} \\
&\leq E_\tau \| n^{-1/2} \sum_{i=1}^{n} \Delta_{n_G}(x_0; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \tau_i \mathbf{z}_i \|_2 + c_0 \int_0^{D_n} \sqrt{\log N(\epsilon, \mathcal{F}_{\mathcal{U}}, d_{2;z,\mathcal{U}})} d\epsilon, \quad \text{(E.0.31)}
\end{aligned}
$$

where x_0 is any point in $[0, L_0]$ and and $c_0 > 0$ is a positive constant. Moreover, we define

$$
\begin{aligned}
D_n^2 &= \sup_{x \in [0, L_0]} | n^{-1} \sum_{i=1}^{n} \mathrm{tr}\{ [\Delta_{n_G}(x; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \mathbf{z}_i]^{\otimes 2} \} |, \quad \text{(E.0.32)} \\
d_{2;z,\mathcal{U}}(x_1, x_2)^2 &= n^{-1} \sum_{i=1}^{n} \mathrm{tr}\{ \mathbf{z}_i^{\otimes 2} \} || \Delta_{n_G}(x_1; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) - \Delta_{n_G}(x_2; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) ||_2^2.
\end{aligned}
$$

The proof of (E.0.31) consists of two steps. First, we note that

$$
[E_\tau \| n^{-1/2} \sum_{i=1}^{n} \Delta_{n_G}(x_0; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \tau_i \mathbf{z}_i \|_2]^2 \leq n^{-1} \sum_{i=1}^{n} \| \mathbf{z}_i \|_2^2 || \Delta_{n_G}(x_0; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) ||_2^2. \quad \text{(E.0.33)}
$$

With some calculations, we have

$$
\begin{aligned}
&E[|| \Delta_{n_G}(x_0; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) ||_2^2] = n_G^{-1} E[\int_0^{L_0} K_{h^{(1)}}(u - x_0)^2 \mathrm{tr}(\mathbf{y}_{h^{(1)}}(u - x_0)^{\otimes 2}) \\
&\| \mathcal{C}(u, \mathbf{z}_i) \mathrm{vecs}(\mathcal{U}_i(u)) \|_2^2 du - \int_0^{L_0} K_{h^{(1)}}(u - x_0) K_{h^{(1)}}(v - x_0) \mathbf{y}_{h^{(1)}}(u - x_0)^T \mathbf{y}_{h^{(1)}}(v - x_0) \\
&\{ \mathcal{C}(u, \mathbf{z}_i) \mathrm{vecs}(\mathcal{U}_i(u)) \}^T \{ \mathcal{C}(v, \mathbf{z}_i) \mathrm{vecs}(\mathcal{U}_i(v)) \} du dv] \\
&= O(n_G^{-1} + n_G^{-1} h^{(1)-1}). \quad \text{(E.0.34)}
\end{aligned}
$$

Thus, we have $[E_\tau | n^{-1/2} \sum_{i=1}^{n} \Delta_{n_G}(x_0; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \tau_i \mathbf{z}_i |]^2 = O(n_G^{-1} + n_G^{-1} h^{(1)-1})$.

Second, it is noted that

$$
\Delta_{n_G}(x; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) = \int_0^{L_0} K_{h_{(1)}}(u - x) \mathbf{y}_{h^{(1)}}(u - x) \otimes \mathcal{C}(u, \mathbf{z}_i) \mathrm{vecs}(\mathcal{U}_i(u)) d[\Pi_{n_G}(u) - \Pi(u)].
$$

It follows from an integration by parts that $||\Delta_{n_G}(x;\mathcal{U}_i,\mathbf{z}_i,h^{(1)})||_2$ is upper bounded by

$$c_1 h^{(1)-1} \sup_{x\in[0,L_0]} |\Pi_{n_G}(x)-\Pi(x)| \sup_{x\in[0,L_0]} [\|\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2 + \|\partial_x\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2],$$

where c_1 is a positive constant. Therefore, we have

$$D_n \le A_n = c_0 h^{(1)-1} \sup_{x\in[0,L_0]} |\Pi_{n_G}(x)-\Pi(x)|$$
$$\sqrt{n^{-1}\sum_{i=1}^{n}[\|\mathbf{z}_i\|_2^2\{\sup_{x\in[0,L_0]}\|\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2 + \|\partial_x\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2\}^2]}.$$

It follows from the integration by parts that there is a positive scalar $\tilde{c}_1$ such that

$$\begin{aligned}
&\|\Delta_{n_G}(x_1;\mathcal{U}_i,\mathbf{z}_i,h^{(1)}) - \Delta_{n_G}(x_2;\mathcal{U}_i,\mathbf{z}_i,h^{(1)})\|_2 \\
&= \|\int_0^{L_0}[H_{h^{(1)}}(u-x_1)-H_{h^{(1)}}(u-x_2)]\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))d[\Pi_{n_G}(u)-\Pi(u)]\| \\
&\le |x_1-x_2|\tilde{c}_1 h^{(1)-1} \sup_{x\in[0,L_0]} |\Pi_{n_G}(x)-\Pi(x)| \sup_{x\in[0,L_0]} [\|\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2 + \\
&\|\partial_x\mathcal{C}(x,\mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))\|_2]. \qquad \text{(E.0.35)}
\end{aligned}$$

For any given $x_1, x_2 \in [0, L_0]$, we have $d_{2;z,\mathcal{U}}(x_1,x_2) \le c_2|x_1-x_2|A_n$, where c_2 is positive scalar. Therefore, we have $N(\epsilon A_n, \mathcal{F}_\mathcal{U}, d_{2;z,\mathcal{U}}) = O(\epsilon^{-1})$ and

$$\begin{aligned}
&\int_0^{D_n}\sqrt{\log N(\epsilon,\mathcal{F}_\mathcal{U},d_{2;z,\mathcal{U}})}d\epsilon \le \int_0^1\sqrt{\log N(\epsilon A_n,\mathcal{F}_\mathcal{U},d_{2;z,\mathcal{U}})}d\epsilon A_n \\
&= O(h^{(1)-1}n_G^{-1/2}\log\log(n_G)) = O_p(h^{(1)-1}n_G^{-1/2}).
\end{aligned}$$

Proof of Theorem 4.2.1 . The proof of Theorem 4.2.1 (i) consists of two steps. The first step is to show the finite convergence of $\{X_n(x) : x \in [0, L_0]\}$. The second step is to check the asymptotic continuity of $X_n(x)$ as x varies in $[0, L_0]$. Moreover, Theorem 4.2.1 (ii) is a consequence of Theorem 4.2.1 (i).

In the first step, it follows from Lemma E.0.13 that at a single point x, $X_n(x)$ converges weakly to $N(0,\Sigma_\beta(x))$. The finite convergence can be directly verified by generalizing the asymptotic

distribution of $X_n(x)$ at one point to any finite number of points using the Cramer-Wold theorem (DasGupta, 2008).

In the second step, we will show the stochastic continuity of $X_n(x)$. It follows from equations (E.0.4) and (E.0.17) that

$$\sqrt{n}\{\hat{\beta}(x) - \beta(x) - 0.5u_2 h^{(1)2}\ddot{\beta}(x)\} = -\sqrt{n}\{nn_G h^{(1)}\pi(x)\}^{-1}$$
$$\left[E\{\partial_Y^2\psi(S(x), C(\mathbf{z}, \beta(x))) \otimes \mathbf{z}^{\otimes 2}\}\right]^{-1}\{\mathcal{T}_{\mathcal{E}}(h^{(1)}, x) + \mathcal{T}_{\mathcal{U}}(h^{(1)}, x)\}\{1 + o_p(1)\}.$$

It follows from lemmas E.0.14 and E.0.15 that

$$n^{-1/2}n_G^{-1}T_{\mathcal{E}}(h^{(1)}, x) = O_p((n_G h^{(1)})^{-1/2}\sqrt{|\log h^{(1)}|})O_p(|\log(h_{n_G,k})|(n_G h_{n_G,k})^{-1})$$

hold uniformly for all $x \in [0, L_0]$. Thus, it follows from the Slutsky's lemma and the continuous mapping theorem that we only need to show the asymptotic tightness of the sequence $\{\tilde{X}_n(x) : x \in [0, L_0]\}$ defined by $\tilde{X}_n(x) = n^{-1/2}n_G^{-1}T_{\mathcal{U}}(h^{(1)}, x)$. Note that $\tilde{X}_n(x)$ can be written as the sum of $n^{-1/2}\sum_{i=1}^n \Delta_{n_G}(x; \mathcal{U}_i, \mathbf{z}_i, h^{(1)}) \otimes \mathbf{z}_i$ and $\hat{X}_n(x) = n^{-1/2}\sum_{i=1}^n \int_0^{L_0} H_{h^{(1)}}(u - x) \otimes \{\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x_j))\}\pi(u)du \otimes \mathbf{z}_i$. It follows from Lemma E.0.16 that we only need to show that $\hat{X}_n(x)$ is asymptotic tight.

The asymptotic tightness of $\hat{X}_n(x)$ can be proved by verifying a classical, simple sufficient condition (van der Vaar and Wellner, 1996). The $\hat{X}_n(x)$ can be written as

$$\hat{X}_n(x) = n^{-1/2}\sum_{i=1}^n \int_0^{L_0} H_{h^{(1)}}(u - x) \otimes [\mathcal{C}(u, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(u)) - \mathcal{C}(x, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))]\pi(u)du$$
$$\otimes \mathbf{z}_i + n^{-1/2}\sum_{i=1}^n \int_0^{L_0} H_{h^{(1)}}(u - x)\pi(u)du \otimes \{\mathcal{C}(x, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x)) \otimes \mathbf{z}_i\} \quad \text{(E.0.36)}$$

Let $\mathbf{e}_m$ be a $2qr \times 1$ vector with the m-th element 1 and zero otherwise. Note that the first term on the right hand side of (E.0.36) equals

$$n^{-1/2}\sum_{i=1}^n \int_{[0,L_0]\cap[x-h^{(1)}, x+h^{(1)}]} H_1(u)[\mathcal{C}(x + h^{(1)}u, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x + h^{(1)}u)) -$$
$$\mathcal{C}(x, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x))]\pi(x + h^{(1)}u)du \otimes \mathbf{z}_i$$

Therefore, it follows from assumptions M1-M3 that

$$E\{\mathbf{e}_m[\hat{X}_n(x) - \hat{X}_n(x_2)]\}^2 = (x_1 - x_2)^2 O(1). \tag{E.0.37}$$

Therefore, we can finish the proof of Theorem 4.2.1.

Proof of Theorem 4.2.2

We define

$$\begin{aligned}
\overline{\mathcal{E}}_i(x) &= \sum_{j=1}^{n_G} \tilde{K}_{n_G,h^{(2)}}(x_j - x)\text{vecs}(\mathcal{E}_i(x_j)), \\
\Delta\mathcal{U}_i(x) &= \sum_{j=1}^{n_G} \tilde{K}_{n_G,h^{(2)}}(x_j - x)\text{vecs}(\mathcal{U}_i(x_j) - \mathcal{U}_i(x)), \\
\Delta\beta(x, \mathbf{z}_i) &= \sum_{j=1}^{n_G} \tilde{K}_{n_G,h^{(2)}}(x_j - x)\text{vecs}(\log(C(\mathbf{z}_i, \beta(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \beta(x_j))^{-T}) \\
&- \log(C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-1} S_i(x_j) C(\mathbf{z}_i, \hat{\beta}_e(x_j))^{-T})), \\
\Delta_i(x) &= \overline{\mathcal{E}}_i(x) + \Delta\mathcal{U}_i(x) + \Delta\beta(x, \mathbf{z}_i).
\end{aligned} \tag{E.0.38}$$

Let $\overline{\mathcal{E}}_{i,k}(x)$, $\Delta\mathcal{U}_{i,k}(x)$, $\Delta\beta_k(x, \mathbf{z}_i)$, $\hat{\mathcal{U}}_{i,k}(x)$ and $\mathcal{U}_{i,k}(x)$ denote the k-th element of $\overline{\mathcal{E}}_i(x)$, $\Delta\mathcal{U}_i(x)$, $\Delta\beta(x, \mathbf{z}_i)$, $\hat{\mathcal{U}}_i(x)$ and $\mathcal{U}_{i,k}(x)$ for $k = 1, \cdots, q$, respectively. We need the following two lemmas, whose detailed proofs can be found in Zhu et al. (2010a).

Lemma E.0.17. *If assumptions (M1), (M3), (M4) and M(10) are true, then we have*

$$\sup_{(x,t)} n^{-1} |\sum_{i=1}^{n} \overline{\mathcal{E}}_{i,k}(x)\mathcal{U}_{i,k'}(t)| = O_p(n^{-1/2}(\log n)^{1/2}). \tag{E.0.39}$$

Lemma E.0.18. *If assumptions (M1), (M3), (M4) and (M10) are true, then we have*

$$\sup_{(x,t)} n^{-1} |\sum_{i=1}^{n} \overline{\mathcal{E}}_{i,k}(x)\overline{\mathcal{E}}_{i,k'}(t)| = O((n_G h^{(2)})^{-1} + (\log n/n)^{1/2}) = o_p(1). \tag{E.0.40}$$

Proof of Theorem 4.2.2 Recall that $\Delta_{i,k}(x) = \text{vecs}(\hat{\mathcal{U}}_{i,k}(x) - \mathcal{U}_{i,k}(x))$ and

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\hat{\mathcal{U}}_{i,k}(x)\hat{\mathcal{U}}_{i,k'}(t) &= n^{-1}\sum_{i=1}^{n}\Delta_{i,k}(x)\Delta_{i,k'}(t) + n^{-1}\sum_{i=1}^{n}\mathcal{U}_{i,k}(x)\Delta_{i,k'}(t) \\ &\quad + n^{-1}\sum_{i=1}^{n}\Delta_{i,k}(x)\mathcal{U}_{i,k'}(t) + n^{-1}\sum_{i=1}^{n}\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t). \end{aligned} \tag{E.0.41}$$

The proof consists of two steps. The first step is to show that the first three terms on the right hand side of (E.0.41) converge to zero uniformly for all $(x,t) \in [0, L_0] \times [0, L_0]$ in probability. The second step is to show the uniform convergence of $n^{-1}\sum_{i=1}^{n}\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t)^T$ to $\Sigma_{\mathcal{U},kk'}(x,t)$ over $(x,t) \in [0, L_0] \times [0, L_0]$ in probability.

We first show that

$$\sup_{(x,t)} n^{-1}|\sum_{i=1}^{n}\Delta_{i,k}(x)\mathcal{U}_{i,k'}(t)| = O_p(n^{-1/2} + \hat{h}_e^{(1)2} + h^{(2)2} + (\log n/n)^{1/2}). \tag{E.0.42}$$

Since

$$\begin{aligned} &|\sum_{i=1}^{n}\Delta_{i,k}(x)\mathcal{U}_{i,k'}(t)| \le n^{-1}\{|\sum_{i=1}^{n}\overline{\mathcal{E}}_{i,k}(x)\mathcal{U}_{i,k'}(t)| + |\sum_{i=1}^{n}\Delta\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t)| + \\ &|\sum_{i=1}^{n}\Delta\beta_k(x,\mathbf{z}_i)\mathcal{U}_{i,k'}(t)|\}, \end{aligned} \tag{E.0.43}$$

it is sufficient to focus on the three terms on the right-hand side of (E.0.43).

Since $\sqrt{n}\{\hat{\beta}_e(\cdot) - \beta(\cdot) - 0.5u_2\hat{h}_e^{(1)2}\ddot{\beta}(\cdot)(1+o_p(1))\}$ weakly converges to a Gaussian process in $\ell^\infty([0,L_0])$ as $n \to \infty$, it is asymptotically tight. Thus, by Tayler's expansion, we have

$$\begin{aligned} &\Delta\beta(x,\mathbf{z}_i) = -\sum_{j=1}^{n_G}\tilde{K}_{n_G,h^{(2)}}(x_j - x)\partial_Y \text{vecs}(\log(C(\mathbf{z}_i,\beta(x_j))^{-1}S_i(x_j)C(\mathbf{z}_i,\beta(x_j))^{-T})) \otimes \mathbf{z}_i^T 0.5u_2 \\ &\hat{h}_e^{(1)2}\ddot{\beta}(x_j)(1+o_p(1)) + \sum_{j=1}^{n_G}\tilde{K}_{n_G,h^{(2)}}(x_j - x)\partial_Y \text{vecs}(\log(C(\mathbf{z}_i,\beta(x_j))^{-1}S_i(x_j)C(\mathbf{z}_i,\beta(x_j))^{-T})) \otimes \mathbf{z}_i^T \\ &\{0.5u_2\hat{h}_e^{(1)2}\ddot{\beta}(x_j)(1+o_p(1)) + \beta(x_j) - \hat{\beta}_e(x_j)\} \end{aligned}$$

Thus,

$$\|\Delta\beta_k(x,\mathbf{z}_i)\| \leq (O_p(h_e^{(1)2}) + O_p(n^{-1/2}))$$
$$\sup_x \|E[\partial_Y \text{vecs}(\log(C(\mathbf{z}_i,\beta(x))^{-1}S_i(x)C(\mathbf{z}_i,\beta(x))^{-T}))] \otimes \mathbf{z}_i^T\|_2. \tag{E.0.44}$$

So, we have

$$n^{-1}\sum_{i=1}^n \|\Delta\beta_k(x,\mathbf{z}_i)\mathcal{U}_{i,k'}(t)\| \leq (O_p(h_e^{(1)2}) + O_p(n^{-1/2}))$$
$$\sup_{(x,t)} n^{-1}\sum_{i=1}^n \|E[\partial_Y \text{vecs}(\log(C(\mathbf{z}_i,\beta(x))^{-1}S_i(x)C(\mathbf{z}_i,\beta(x))^{-T}))] \otimes \mathbf{z}_i^T\|_2 |\mathcal{U}_{i,k'}(t)|$$
$$= O_p(h_e^{(1)2}) + O_p(n^{-1/2}).$$

Similarly, we have

$$n^{-1}|\sum_{i=1}^n \Delta\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t)| \leq n^{-1}\sum_{i=1}^n \sup_{x,t\in[0,L_0]} |\Delta\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t)| = O_p(h^{(2)2}) = o_p(1).$$

It follows from Lemma E.0.17 that $\sup_{(x,t)} n^{-1}\{|\sum_{i=1}^n \overline{\mathcal{E}}_{i,k}(x)\mathcal{U}_{i,k'}(t)| = O((\log n/n)^{1/2})$. Similarly, we can show that $\sup_{(x,t)} n^{-1}|\sum_{i=1}^n \Delta_{i,k}(t)\mathcal{U}_{i,k'}(x)| = O_p(n^{-1/2} + \hat{h}_e^{(1)2} + h^{(2)2} + (\log n/n)^{1/2})$.

We can show that

$$\sup_{(x,t)} \|n^{-1}\sum_{i=1}^n [\mathcal{U}_{i,k}(x)\mathcal{U}_{i,k'}(t) - \Sigma_{\mathcal{U},kk'}(x,t)]\| = O_p(n^{-1/2}). \tag{E.0.45}$$

Since $|\mathcal{U}_{i,k}(x_1)\mathcal{U}_{i,k'}(t_1) - \mathcal{U}_{i,k}(x_2)\mathcal{U}_{i,k'}(t_2)| \leq 2(|x_1-x_2|+|t_1-t_2|) \sup_{x\in[0,L_0]} |\dot{\mathcal{U}}_{i,k}(x)| \sup_{x\in[0,L_0]} |\mathcal{U}_{i,k'}(x)|$ holds for any (x_1,t_1) and (x_2,t_2), the functional class $\{\mathcal{U}_{i,k}(u)\mathcal{U}_{i,k'}(v) : (u,v) \in [0,L_0]^2\}$ is a Vapnik and Cervonenkis (VC) class (van der Vaar and Wellner, 1996; Kosorok, 2008). Thus, it yields that (E.0.45) is true.

Finally, we can show that

$$\sup_{(x,t)} n^{-1}|\sum_{i=1}^n \Delta_{i,k}(x)^T\Delta_{i,k'}(t)| = O_p((n_G h^{(2)})^{-1} + (\log n/n)^{1/2} + \hat{h}_e^{(1)4} + h^{(2)4}). \tag{E.0.46}$$

It follows from the Cauchy-Schwartz inequality that

$$|\sum_{i=1}^{n}\Delta_{i,k}(x)\Delta_{i,k'}(t)| \leq O(1)\sup_{(x,t)}[|\sum_{i=1}^{n}\overline{\mathcal{E}}_{i,k}(x)\overline{\mathcal{E}}_{i,k'}(t)| + |\sum_{i=1}^{n}\Delta\mathcal{U}_{i,k}(x)\Delta\mathcal{U}_{i,k'}(t)| + |\sum_{i=1}^{n}\Delta\beta_k(x,\mathbf{z}_i)\Delta\beta'_k(t,\mathbf{z}_i)|].$$

It follows from Lemma E.0.18 that $\sup_{(x,t)} n^{-1}|\sum_{i=1}^{n}\overline{\mathcal{E}}_{i,k}(x)\overline{\mathcal{E}}_{i,k'}(t)| = O((n_G h^{(2)})^{-1}+(\log n/n)^{1/2})$. Since $\sup_{x\in[0,L_0]}|\Delta\mathcal{U}_{i,k}(x)| = O(1)\sup_{x\in[0,L_0]}|\ddot{\mathcal{U}}_{i,k}(x)|h^{(2)2}$, we have $\sup_{(x,t)} n^{-1}|\sum_{i=1}^{n}\Delta\mathcal{U}_{i,k}(x)\Delta\mathcal{U}_{i,k'}(t)| = O(1)h^{(2)4}$. Furthermore, by (E.0.44), we have $|\sum_{i=1}^{n}\Delta\beta_k(x,\mathbf{z}_i)\Delta\beta_{k'}(t,\mathbf{z}_i)| = O_p(n^{-1}+\hat{h}_e^{(1)4})$.

Combining (E.0.41)-(E.0.45) leads to $\sup_{(x,t)}|\hat{\Sigma}_{\mathcal{U},kk'}(x,t)-\Sigma_{\mathcal{U},kk'}(x,t)| = O_p(n^{-1/2}+(n_G h^{(2)})^{-1}+\hat{h}_e^{(1)2}+h^{(2)2}+(\log n/n)^{1/2})$. So $\sup_{(x,t)}\|\hat{\Sigma}_{\mathcal{U}}(x,t)-\Sigma_{\mathcal{U}}(x,t)\| = O_p(n^{-1/2}+(n_G h^{(2)})^{-1}+\hat{h}_e^{(1)2}+h^{(2)2}+(\log n/n)^{1/2})$. Similar to the arguments in (E.0.41)-(E.0.45), we can show that $\sup_x|\hat{\Sigma}_{\mathcal{E}}(x,x)-\Sigma_{\mathcal{E}}(x,x)| = o_p(1)$.

Proof of Theorem 4.2.3

Proof of Theorem 4.2.3. We define $\tilde{G}(x)^{(g)}$ as follows:

$$\sqrt{n}\{(1,0)\otimes I_{qr}\}\left[\sum_{i=1}^{n}\sum_{j=1}^{n_G}K_{h^{(1)}}(x_j-x)\left\{\mathbf{y}_{h^{(1)}}(x_j-x)^{\otimes 2}\otimes\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^{\otimes 2}\right\}\right]^{-1}$$
$$\sum_{i=1}^{n}\tau_i^g\sum_{j=1}^{n_G}H_{h^{(1)}}(x_j-x)\otimes\{\partial_Y\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i\}. \quad \text{(E.0.47)}$$

Following the arguments in (Kosorok, 2003; Zhu and Zhang, 2006; Zhu et al., 2010a), we will prove Theorem 4.2.3 in three steps. In Step 1, we will prove the unconditional weak convergence of $\tilde{G}(x)^{(g)}$. In Step 2, we will prove the weak convergence of $\tilde{G}(x)^{(g)}$ conditional on the data. In Step 3, we will prove the weak convergence of $G(x)^{(g)}$ conditional on the data by showing that $\tilde{G}(x)^{(g)}$ and $G(x)^{(g)}$ are asymptotically equivalent as $n\to\infty$.

In Step 1, it follows from Equation (E.0.4) that

$$\tilde{G}(x)^{(g)} = 2\sqrt{n}\{(1,0)\otimes I_{qr}\}$$
$$\left[\sum_{i=1}^{n}\sum_{j=1}^{n_G}K_{h^{(1)}}(x_j-x)\left\{\mathbf{y}_{h^{(1)}}(x_j-x)^{\otimes 2}\otimes\partial_Y^2\psi(S_i(x_j),C(\mathbf{z}_i,\beta(x_j)))\otimes\mathbf{z}_i^{\otimes 2}\right\}\right]^{-1}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n_G} H_{h^{(1)}}(x_j - x) \otimes \{\mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j) + \mathcal{U}_i(x_j)) \otimes \tau_i^g \mathbf{z}_i\}.$$

Therefore, by treating $\tau_i^{(g)}\mathbf{z}_i$ as the new 'covariate' vector, we can apply the same arguments in the proof of Lemma (E.0.13) to prove that $\tilde{G}_k^{(g)}$ converges to G_k in distribution; that is, $\tilde{G}_k^{(g)}$ is asymptotically measurable.

In step 2, we define

$$\begin{aligned}
S(x,t) &= n^{-1}n_G^{-2}\sum_{i=1}^{n}\sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_{j'} - t)^T \otimes \partial_Y \psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j))) \\
&\quad \partial_Y \psi(S_i(x_{j'}), C(\mathbf{z}_i, \beta(x_{j'})))^T \otimes \mathbf{z}_i^{\otimes 2}, \\
S_{\mathcal{U}\mathcal{U}}(x,t) &= 4n^{-1}n_G^{-2}\sum_{i=1}^{n}\sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h_{n_G,k}}(x_{j'} - t)^T \otimes \mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x_j)) \\
&\quad \text{vecs}(\mathcal{U}_i(x_{j'}))^T\mathcal{C}(x_{j'}, \mathbf{z}_i)^T \otimes \mathbf{z}_i^{\otimes 2}, \\
S_{\mathcal{U}\mathcal{E}}(x,t) &= 4n^{-1}n_G^{-2}\sum_{i=1}^{n}\sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_{j'} - t)^T \otimes \mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{U}_i(x_j)) \\
&\quad \text{vecs}(\mathcal{E}_i(x_{j'}))^T\mathcal{C}(x_{j'}, \mathbf{z}_i)^T \otimes \mathbf{z}_i^{\otimes 2}, \\
S_{\mathcal{E}\mathcal{E}}(x,t) &= 4n^{-1}n_G^{-2}\sum_{i=1}^{n}\sum_{j,j'=1}^{n_G} H_{h^{(1)}}(x_j - x)H_{h^{(1)}}(x_{j'} - t)^T \otimes \mathcal{C}(x_j, \mathbf{z}_i)\text{vecs}(\mathcal{E}_i(x_j)) \\
&\quad \text{vecs}(\mathcal{E}_i(x_{j'}))^T\mathcal{C}(x_{j'}, \mathbf{z}_i)^T \otimes \mathbf{z}_i^{\otimes 2}.
\end{aligned}$$

By Lemma E.0.10, conditional on data, $\tilde{G}(x)^{(g)}$ is a normal random vector with zero mean and covariance given by $(nn_G\pi(x))^{-2}\{(1,0)\otimes I_{qr}\}\left[\text{diag}(1,u_2)\otimes E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-1}$ $S(x,t)\left[\text{diag}(1,u_2)\otimes E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-1}\{(1,0)^T\otimes I_{qr}\}$. It is easy to see that

$$S(x,t) = S_{\mathcal{U}\mathcal{U}}(x,t) + S_{\mathcal{U}\mathcal{E}}(x,t) + S_{\mathcal{U}\mathcal{E}}(t,x) + S_{\mathcal{E}\mathcal{E}}(x,t). \tag{E.0.48}$$

Following the arguments of lemmas E.0.17 and E.0.18, we can show that $S_{\mathcal{U}\mathcal{E}}(x,t) + S_{\mathcal{U}\mathcal{E}}(t,x) + S_{\mathcal{E}\mathcal{E}}(x,t) = o(1)$. Furthermore, it can be shown that $E[S_{\mathcal{U}\mathcal{U}}(x,t)] = 4\text{diag}(1,0)\otimes\Omega_{\mathcal{U}}(x,x) + O(\hat{h}_e^{(1)})$ and $\text{Cov}[S_{\mathcal{U}\mathcal{U}}(x,t)] = O(n^{-1})$. Let Cov_τ be denoted as the covariance with respect to $\tau_i^{(g)}$ conditional on the data. Therefore, we have $\text{Cov}_\tau[\tilde{G}(x)^{(g)}, \tilde{G}(t)^{(g)}]$ converges to $4\{E(\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))$ $\otimes\mathbf{z}^{\otimes 2})\}^{-1}\Omega_{\mathcal{U}}(x,x)\left[E\{\partial_Y^2\psi(S(x),C(\mathbf{z},\beta(x)))\otimes\mathbf{z}^{\otimes 2}\}\right]^{-1}$ in probability. We can obtain the marginal convergence of $\tilde{G}(x)^{(g)}$ in the conditional central limit theorem by using the Cramer-Wald method.

For each $\delta > 0$, let $\tilde{\mathbf{X}}_\delta = \{l\delta : l = 0, \cdots, L_0\delta^{-1}\}$ be an equally $\delta-$spaced grid and $[0, L_0]_\delta(x)$ assign to each $x \in [0, L_0]$ a closest element of $\tilde{\mathbf{X}}_\delta$. The finite convergence results yield $\sup_{f\in\mathcal{B}L_1(\ell^\infty([0,L_0])}$ $|E_\tau f(\tilde{G}^{(g)}([0, L_0]_\delta)) - Ef(G([0, L_0]_\delta))| \to 0$ in probability, as $g \to \infty$. Due to the continuity of $G(x)$, we have $G([0, L_0]_\delta(x)) \to G(x)$ almost surely as $\delta \to 0$, that is, $\lim_{\delta\to 0} \sup_{f\in\mathcal{B}L_1(\ell^\infty([0,L_0])} |E_\tau f(\tilde{G}([0, L_0]_\delta))$ $-Ef(G([0, L_0]))| = 0$. Finally, we have

$$\sup_{f\in\mathcal{B}L_1(\ell^\infty([0,L_0])} |E_\tau f(\tilde{G}^{(g)}([0, L_0]_\delta(\cdot))) - E_\tau f(\tilde{G}^{(g)})| \leq E_\tau(\sup_{|x-x'|_2<\delta} |\tilde{G}^{(g)}(x) - \tilde{G}^{(g)}(x')|).$$

Thus, the expectation on the left side of the above equation is smaller than $E(\sup_{|x-x'|_2<\delta} |\tilde{G}^{(g)}(x) - \tilde{G}^{(g)}(x')|)$, which was established by the unconditional weak convergence of $\tilde{G}^{(g)}(\cdot)$ in Step 1. This finishes the proof of Step 2.

Let $\Delta_{\partial_Y\psi}(\mathbf{z}_i, x_j) = \partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \hat{\beta}(x_j))) - \partial_Y\psi(S_i(x_j), C(\mathbf{z}_i, \beta(x_j)))$. In Step 3, following the arguments in Theorem 4.2.3 of Kosorok (2003), we only need to prove that

$$\Delta_{n,\beta} = \sup_{x\in[0,L_0]} n^{-1} \sum_{i=1}^{n} \mathrm{tr}[\{n_G^{-1} \sum_{j=1}^{n_G} H_{h^{(1)}}(x_j - x) \otimes \Delta_{\partial_Y\psi}(\mathbf{z}_i, x_j)\}^{\otimes 2} \otimes \mathbf{z}_i^{\otimes 2}] = o_p(1).$$

It follows from the proof of Theorem 4.2.2 that $\Delta_{n,\beta} = O_p(n^{-1} + (\hat{h}_e^{(1)})^4)$, which converges to zero in probability. This finishes the proof of Theorem 4.2.3.

Bibliography

Anderson, A. (2001), "Theoretical analysis of the effects of noise on diffusion tensor imaging," *Magnetic Resonance in Medicine*, 46, 1174–1188.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis (3rd ed.)*, Wiley Series in Probability and Statistics.

Arsigny, V. (2006), "Processing Data in Lie Groups: An Algebraic Approach. Application to Non-linear Registration and Diffusion Tensor MRI," Ph.D. dissertation.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006), "Log-Euclidean metrics for fast and simple caculus on diffusion tensors," *Magnetic Resonance in Medicine*, 56, 411–421.

Barmpoutis, A., Vemuri, B. C., Shepherd, T. M., and Forder, J. R. (2007), "Tensor splines for interpolation and approximation of DT-MRI With applications to segmentation of isolated rat hippocampi," *IEEE Transactions on Medical Imaging*, 26, 1537–1546.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994a), "Estimation of the effective self- diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance Ser. B*, 103, 247–254.

— (1994b), "MR diffusion tensor spectroscopy and imaging," *Biophysical Journal*, 66, 259–267.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000), "In vivo fiber tractography using DT-MRI data," *Magnetic Resonance in Medicine*, 44, 625–632.

Bhattacharya, R. and Patrangenaru, V. (2005), "Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds-II," *Annals of Statistics*, 33, 1225–1259.

Chefd'hotel, C., Tschumperlé, D., Deriche, R., and Faugeras, O. (2004), "Regularizing flows for constrained matrix-valued images," *Journal of Mathematical Imaging and Vision*, 20, 147–162.

Chen, Y. S., An, H. Y., Zhu, H. T., Stone, T., Smith, J. K., Hall, C., Bullitt, E., Shen, D. G., and Lin, W. L. (2009), "White matter abnormalities revealed by diffusion tensor imaging in non-demented and demented HIV+ patients," *NeuroImage*, 47, 1154–1162.

Commowick, O., Fillard, P., Clatz, O., and Warfield, S. (2008), "Detection of DTI white matter bbnormalities in multiple sclerosis patients," *Med Image Comput Comput Assist Interv*, 11, 975 – 982.

DasGupta, A. (2008), *Asymptotic Theory of Statistics and Probability*, Springer Verlag, New York.

Davis, B. C., Bullitt, E., Fletcher, P. T., and Joshi, S. (2010), "Population Shape Regression From Random Design Data," *International Journal of Computer Vision*, 90, 255–266.

Dryden, I. L., Koloydenko, A., and Zhou, D. (2009), "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *Annals of Applied Statistics*, 3, 1102–1123.

Einmahl, U. and Mason, D. M. (2000), "An empirical process approach to the uniform consistency of kernel-type function estimators," *Journal of Theoretical Probability*, 13, 1–37.

Fan, J. (1992), "Design-adaptive nonparametric regression," *Journal of the American Statistical Association*, 87, 998–1004.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.

Fan, J. and Lin, S.-K. (1998), "Test of significance when data are curves," *J. Amer. Statist. Assoc.*, 93, 1007–1021.

Fan, J. and Yao, Q. (1998), "Efficient estimation of conditional variance functions in stochastic regression," *Biometrika*, 85, 645–660.

Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive varying-coefficient linear models," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65, 57–80.

Fan, J. and Zhang, W. (1999), "Statistical estimation in varying coefficient models," *Ann. Statist.*, 27, 1491–1518.

— (2000), "Simultaneous confidence bands and hypothesis testing in varying-coefficient models," *Scandinavian Journal of Statistics*, 27, 715–731.

— (2008), "Statistical methods with varying coefficient models," *Statistics and Its Interface*, 1, 179–195.

Filippi, C., Ulug, A., Ryan, E., Ferrando, S., and van Gorp, W. (2001), "Diffusion tensor imaging of patients with HIV and normal-appearing white matter on MR images of the brain," *Ajnr: Am. J. Neuroradiol*, 22, 277–283.

Fingelkurts, A. A., Fingelkurts, A. A., and Kahkonen., S. (2005), "Functional connectivity in the brain-is it an elusive concepts?" *Neuroscience and Biobehavioral Reviews*, 28, 827–836.

Fletcher, P. T. (2004), "Statistical Variability in Nonlinear Spaces:Application to Shape Analysis and DT-MRI," Ph.D. thesis, university of North Carolina at Chapel Hill.

Frank, L. R. (2002), "Characterization of anisotropy in high angular resolution diffusionweighted MRI," *Magnetic Resonance in Medicine*, 47, 1083–1099.

Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., and Lin, W. (2009), "Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects." *Proceedings of the National Academy of Sciences*, 106, 6790–6795.

Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, Massachusetts: Addison Wesley.

Goldsmith, J., Feder, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2010), "Penalized functional regression," Tech. rep., Working Paper 2010-204, Johns Hopkins University Dept. of Biostatistics.

Goodlett, C. B., Fletcher, P. T., Gilmore, J. H., and Gerig, G. (2009), "Group analysis of DTI fiber tract statistics with application to neurodevelopment," *NeuroImage*, 45, S133–S142.

Grenander, U. and Miller, M. I. (2007), *Pattern Theory From Representation to Inference*, Oxford University Press.

Hall, P., Marron, J. S., and Park, B. U. (1992), "Smoothed cross-validation," *Probability Theory and Related Fields*, 92, 1–20.

Härdle, W., Hall, P., and Marron, J. S. (1992), "Regression smoothing parameters that are not far from their optimum," *Journal of American Statistical Association*, 87, 227–233.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

Higham, N. J. (2008), *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, U.S.

Jones, M., Marron, J., and Sheather, S. (1996), "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, 91, 401–407.

Kim, P. T. and Richards, D. S. (2010), *Deconvolution density estimation on spaces of positive definite symmetric matrices*, A Festschrift of Tom Hettmansperger, IMS Lecture Notes Monograph Series.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), "Optimization by simulated annealing," *Science*, 220, 671–680.

Kosorok, M. R. (2003), "Bootstraps of sums of independent but not identically distributed stochastic processes," *J. Multivariate Anal.*, 84, 299–318.

— (2008), *Introduction to Empirical Processes and Semiparametric Inference.*, Springer: New York.

Liang, F. (2010), "Evolutionary Stochastic Approximation Monte Carlo for Global Optimization," *Statistics and Computing*, In press.

Liang, F., Liu, C., and Carroll, R. (2007), "Stochastic approximation in Monte Carlo computation," *J. Am. Statist. Assoc.*, 102, 305–320.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of state calculations by fast computing machines," *J. Chemical Physics*, 21, 1087–1091.

O'Donnell, L., Westin, C.-F., and Golby, A. (2009), "Tract-based morphometry for white matter group analysis," *Neuroimage*, 45, 832–844.

Park, B. U. and Marron, J. S. (1990), "Comparison of data-driven bandwidth selectors," *Journal of American Statistical Association*, 85, 66–72.

Pasternak, O., Sochen, N., and Basser, P. J. (2010), "The effect of metric selection on the analysis of diffusion tensor MRI data," *NeuroImage*, 49, 2190 – 2204.

Pennec, X., Fillard, P., , and Ayache, N. (2006), "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, 66, 41–66.

Pierpaoli, C. and Basser, P. J. (1996), "Toward a quantitative assessment of diffusion anisotropy," *Magn Reson Med*, 36, 893–906.

Pourahmadi, M. (1999), "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," *Biometrika*, 86, 677–690.

— (2000), "Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix," *Biometrika*, 87, 425–435.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer-Verlag.

Rathi, Y., Michailovich, O., and Shenton, M. E.and Bouix, S. (2009), "Directional functions for orientation distribution estimation," *Medical Image Analysis*, 13, 432 – 444.

Rice, J. (1984), "Bandwidth choice for nonparametric regression," *Annals of Statistics*, 12, 1215–1230.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1999), *Akaike Information Criterion Statistics (Mathematics and its Applications)*, Springer.

Schwartzman, A. (2006), "Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data," Ph.D. thesis, stanford University.

Schwartzman, A., Mascarenhas, W., and Taylor, J. E. (2008), "Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices," *Ann. Statist.*, 36, 1423–1431.

Smith, S. M., M., J., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M., Matthews, P., and Behrens, T. E. (2006), "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *NeuroImage*, 31, 1487 – 1505.

Talagrand, M. (1994), "Sharper bounds for Gaussian and empirical processes,," *Ann. Probab*, 22, 28–76.

Tuch, D. S. (2002), "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, 48, 577–582.

— (2004), "Q-ball imaging," *Magnetic Resonance in Medicine*, 52, 1358–1372.

van der Vaar, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag Inc.

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.

Wang, L., Li, H., and Huang, J. Z. (2008), "Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements," *J. Amer. Statist. Assoc.*, 103, 1556–1569.

Welsh, A. H. and Yee, T. W. (2006), "Local regression for vector responses," *Journal of Statistical Planning and Inference*, 136, 3007–3031.

Whitcher, B., Wisco, J., Hadjikhani, N., and Tuch, D. (2007), "Statistical group comparison of diffusion tensors via multivariate hypothesis testing," *Magnetic Resonance in Medicine*, 57, 1065 – 1074.

Wu, C. O. and Chiang, C.-T. (2000), "Kernel smoothing on varying coefficient models with longitudinal dependent variable," *Statist. Sinica*, 10, 433–456.

Wu, H. L. and Zhang, J. T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, Hoboken, New Jersey.: John Wiley & Sons, Inc.

Yushkevich, P. A., Zhang, H., Simon, T., and Gee, J. C. (2008), "Structure-specific statistical mapping ofwhite matter tracts," *Neuroimage*, 41, 448–461.

Zhang, J. and Chen, J. (2007), "Statistical inference for functional data," *Annals of Statistics*, 35, 1052–1079.

Zhu, H. and Zhang, H. P. (2006), "Generalized score test of homogeneity for mixed effects models." *Annals of Statistics*, 34, 1545–1569.

Zhu, H. T., Chen, Y. S., Ibrahim, J. G., Li, Y. M., and Lin, W. L. (2009), "Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging," *Journal of the American Statistical Association*, 104, 1203–1212.

Zhu, H. T., Ibrahim, J. G., Tang, N., Rowe, D., Hao, X., Bansal, R., and Peterson, B. S. (2007a), "A statistical analysis of brain morphology using wild bootstrapping," *IEEE Trans Med Imaging*, 26, 954–966.

Zhu, H. T., Li, R. Z., and Kong, L. L. (2010a), "Multivariate varying coefficient models for functional responses," Tech. rep., University of North Carolina at Chapel Hill.

Zhu, H. T., Styner, M., Tang, N. S., Liu, Z. X., Lin, W. L., and Gilmore, J. (2010b), "FRATS: Functional regression analysis of dti tract statistics," *IEEE Transactions on Medical Imaging*, 29, 1039–1049.

Zhu, H. T., Xu, D., Amir, R., Hao, X., Zhang, H., Alayar, K., Ravi, B., and Peterson, B. (2006), "A statistical framework for the classification of tensor morphologies in diffusion tensor images," *Magnetic Resonance Imaging*, 24, 569–582.

Zhu, H. T., Zhang, H. P., Ibrahim, J. G., and Peterson, B. G. (2007b), "Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion)," *Journal of the American Statistical Association*, 102, 1085–1102.