TOPICS IN FUNCTIONAL DATA ANALYSIS WITH BIOLOGICAL APPLICATIONS

A Dissertation

by

YEHUA LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2006

Major Subject: Statistics

TOPICS IN FUNCTIONAL DATA ANALYSIS WITH BIOLOGICAL APPLICATIONS

A Dissertation

by

YEHUA LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Raymond J. Carroll
	Tailen Hsing
Committee Members,	Naisyin Wang
	Joseph D. Ward
Head of Department,	Simon J. Sheather

August 2006

Major Subject: Statistics

ABSTRACT

Topics in Functional Data Analysis with Biological Applications. (August 2006)

Yehua Li, B.S., Tsinghua University;

M.S., Texas A&M University

Co-Chairs of Advisory Committee: Dr. Raymond J. Carroll
Dr. Tailen Hsing

Functional data analysis (FDA) is an active field of statistics, in which the primary subjects in the study are curves. My dissertation consists of two innovative applications of functional data analysis in biology. The data that motivated the research broadened the scope of FDA and demanded new methodology. I develop new nonparametric methods to make various estimations, and I focus on developing large sample theories for the proposed estimators.

The first project is motivated from a colon carcinogenesis study, the goal of which is to study the function of a protein (p27) in colon cancer development. In this study, a number of colonic crypts (units) were sampled from each rat (subject) at random locations along the colon, and then repeated measurements on the protein expression level were made on each cell (subunit) within the selected crypts. In this problem, measurements within each crypt can be viewed as a function, since the measurements can be indexed by the cell locations. The functions from the same subject are spatially correlated along the colon, and my goal is to estimate this correlation function using nonparametric methods. We use this data set as an motivation and propose a kernel estimator of the correlation function in a more general framework. We develop a pointwise asymptotic normal distribution for the proposed estimator when the number of subjects is fixed and the number of units

within each subject goes to infinity. Based on the asymptotic theory, we propose a weighted block bootstrapping method for making inferences about the correlation function, where the weights account for the inhomogeneity of the distribution of the unit locations. Simulation studies are also provided to illustrate the numerical performance of the proposed method.

My second project is on a lipoprotein profile data, where the goal is to use lipoprotein profile curves to predict the cholesterol level in human blood. Again, motivated by the data, we consider a more general problem: the functional linear models (Ramsay and Silverman, 1997) with functional predictor and scalar response. There is literature developing different methods for this model; however, there is little theory to support the methods. Therefore, we focus more on the theoretical properties of this model. There are other contemporary theoretical work on methods based on Principal Component Regression. Our work is different in the sense that we base our method on roughness penalty approach and consider a more realistic scenario that the functional predictor is observed only on discrete points. To reduce the difficulty of the theoretical derivations, we restrict the functions with a periodic boundary condition and develop an asymptotic convergence rate for this problem in Chapter Ⅲ. A more general result based on splines is a future research topic that I give some discussion in Chapter Ⅳ.

To Fei and Benjamin.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Dr. Raymond J. Carroll and Dr. Tailen Hsing, for their guidance, encouragement and support during my Ph.D. study. You opened the door for me to this wonderful statistical world and helped me to build a balanced knowledge in both theory and methodology. The experience to work with both of you is something that I will be proud of and cherish for the rest of my life. Without your help, I would never have accomplished as much as I have achieved.

A special thanks is owed to Dr. Naisyin Wang who served on my committee, although the help she offered was so much more than from a committee member, that I consider her as a third advisor. I would like to thank you for the many helpful discussions and your kindly advice.

My thanks are extended to Dr. Joe Ward who served on both my M.S. committee and Ph.D. committee. Thank you for serving on my committees and being consistently supportive. I would also like to thank Dr. Randy Eubank and Dr. Michael Sherman for many helpful discussions.

I would like to thank my parents for raising me and always believing in me. I will do my best to keep you proud. And finally, to my lovely wife Fei, who is raising our child little Benjamin. You are the most beautiful thing that has happened in my life. I feel so lucky to have you by my side. Honey, all my work is dedicated to you!

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Functional Data Analysis (FDA) is a new area of statistics which combines and extends existing methodologies and theories from nonparametric/semiparametric smoothing, stochastic processes, multivariate analysis and generalized linear models. In contrast to the traditional methods, FDA deals with data sets, in which a data point is a function defined on a fixed compact set, instead of a vector. In other words, the random variables are defined on a functional space instead of a vector space. In this sense, FDA is an extension of multivariate analysis, where the random vectors are of infinite dimension. We need theory of stochastic processes to model the population of these random functions. However, in real life, these functional subjects are always measured on discrete points, and the measurements are usually contaminated with measurement errors. That is why we need nonparametric smoothing methods to recover the functions.

1.1 Functional Data

Ramsay and Silverman (1997) and Ramsay and Silverman (2002) gave a good summary of examples and methods in FDA. In my dissertation, I will work on two projects both on functional data. Both projects are new applications of function data, that extends the scope of FDA described in Ramsay and Silverman (1997). I will give a brief introductions to these data sets, and use them to illustrate how functional data are unique in nature and why we need to develop new methodology to analyze them. The proposed methods, theory and

This thesis follows the style of *Biometrics*.

data analysis are given int the following chapters.

1.1.1 *Colon Carcinogenesis Data*

The biomarker that we are interested in is p27, which is a life cycle protein that affects cell apoptosis, proliferation and differentiation. An important goal of the study is to understand the function of p27 in the early stage of the cancer development process. In the experiment, 12 rats were administered azoxymethane (AOM), which is a colon specific carcinogen. After 24 hours, the rats were terminated and a segment of colon tissue was excised from each rat. About 20 colonic crypts were randomly picked along a linear slice on the colon segment. The physical distances between the crypts were measured. Then, within each crypt, we measured cells at different depths within the crypts, and then the expression level of p27 was measured for each cell within the chosen crypts.

The first plot in Figure 1 shows the colonic crypts. As we can see that the cells line up within the crypt so that we can index the measurements within a crypt by the relative cell depth. If we denote the cell location in the bottom of a crypt to be 0 and top to be 1, it is natural to consider the true p27 expression levels within a crypt to be a continuous function on $[0,1]$. Therefore, the measurements on cells can be considered as discrete observations on the function. The number of cells per crypt is roughly 30, but it varies from crypt to crypt. Consequently, the observation locations are different from function to function.

1.1.2 *Lipoprotein Profile Data*

The goal of this project is to use lipoprotein profile curves to predict cholesterol levels from a patient. The lipoprotein curves were generated by the following protocol. A standard amount ($50\mu l$) of diluted serum sample from each patient was process in the centrifugation machine, such that different types of lipoproteins were separated into layers according to their densities. A chemical stain was added to show the concentrations of the cholesterol, and the profile curve was generated by recording the intensity of the stain at different heights using digital cameras.

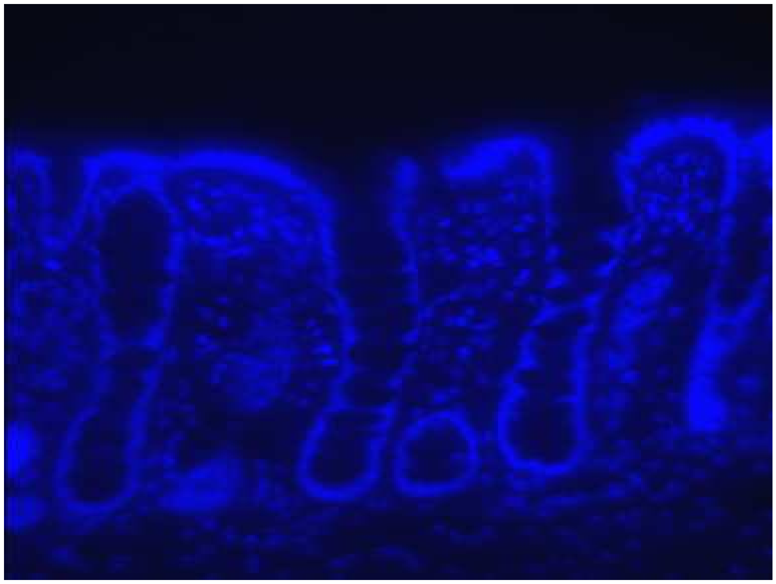
The second plot in Figure 1 shows a typical lipoprotein profile generated in this study. We re-scale the abscissa to make each profile curve defined on $[0, 1]$, with 0 corresponds to the top of the tub and 1 corresponds to the bottom. As one can see, there are usually three peaks in a profile curve, which correspond to the three major types of lipoprotein with different densities. The three types of lipoprotein are Very Low Density Lipoprotein (VLDL), Low Density Lipoprotein (LDL) and High Density Lipoprotein (HDL), from the left to the right of the profile. In this study we have 24 patients, each with a lipoprotein profile curve and a cholesterol level measured separately. Each profile curve consists of over 1,000 equally spaced measurements. The goal is to build a linear model to predict the cholesterol level from these profile curves.

Other typical examples of functional data described in Ramsay and Silverman (1997) are the growth curves, temperature curves and so on.

1.1.3 Why Is Functional Data Special

The compelling reasons for developing new methodology for functional data instead of applying multivariate analysis are the following:

- In many functional data, the dimension of the vector is much higher than the number of subjects. For example, in the lipoprotein profile project, the number of points in each curve depends on the resolution of the camera. In this kind of study, the number of subject is always limited, but as technology advances we can sample more and more points on each curve. Traditional multivariate analysis does not apply in this case, and new methods should be developed to take into account the smoothness of the curves underlying the discrete observations.



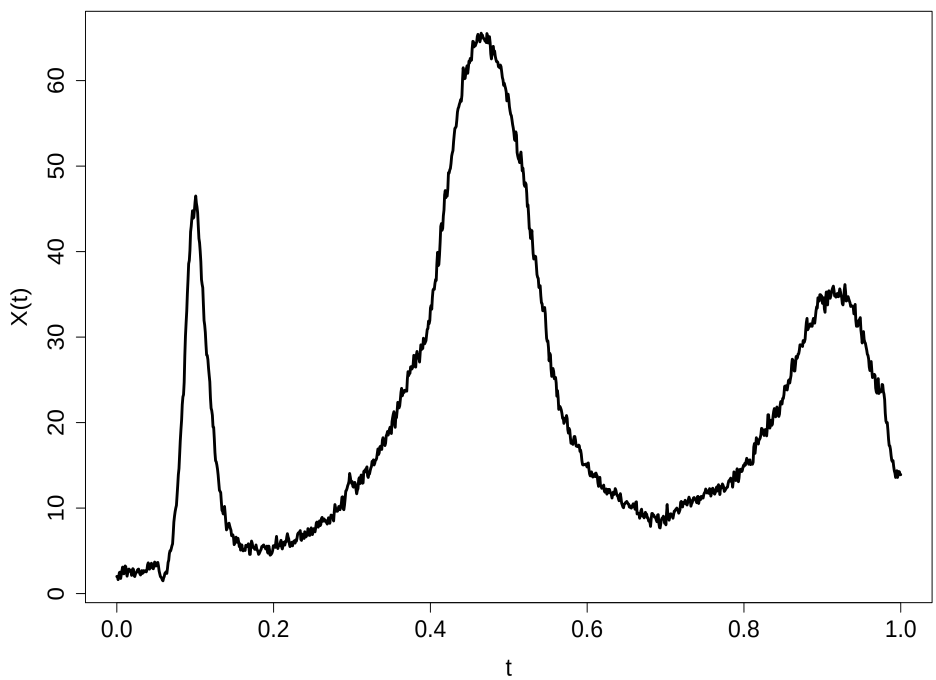


Figure 1: Functional data. Upper panel: a picture of colonic crypts; lower panel: a lipoprotein profile curve.

- The observation locations are different from curve to curve, for example, the colon carcinogenesis data. In this case, the dimension of the vectors could be different from one curve to another, values on the same entries in different vectors could mean different things.

- In some application, the model is concerned with the functional properties of these curves. For example, Ramsay and Silverman (1997) gives an example on growth curves. In this study, the heights of a group of children were measured over time, while the acceleration (second derivative) of these growth curves are of interest.

1.2 General Ideas of Functional Data Analysis

In this section, I will review some general ideas and methods in Functional Data Analysis, which are related to my research topic.

In FDA, the data are a sample of curves. It is natural to model the population of these curves as a stochastic process $X(t)$ defined on the same compact set. Two important ideas are generally used to deal with functional data: one is to do dimension reduction and reduce the problem to multivariate analysis; the other is use a roughness penalty approach that utilize the smooth nature of the curves. Both method will help to reduce the variability of the result but introduce some bias.

1.2.1 Functional Principal Component Analysis

In stochastic process, there is a long-established result that a stochastic process $X(t)$ defined on $[0, 1]$ has the following Karhunen-Leóve expansion (Ash and Gardner, 1975). Suppose $R(s, t) = \text{cov}\{X(s), X(t)\}$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$, with $\phi_1(\cdot), \phi_2(\cdot), \cdots$ being the corresponding eigenfunctions, then

$$X(t) = m(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t),$$

where $m(t) = E\{X(t)\}$, ξ_k are the principal component scores with $E(\xi_k) = 0$ and var$(\xi_k) = \lambda_k$.

1.2.2 Roughness Penalty

The roughness penalty idea is probably first used in nonparametric regressions.

One of the most popular methods in nonparametric regression is the penalized spline method(Ruppert, Wand, and Carroll, 2003), the general framework of which is as the following. Suppose the data we observed are

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

where f is an unknown function. This is a typical regression problem, but f may not be of any parametric form. To increase the flexibility of the model, we can estimate f by $\widehat{f}$ which is spanned by a set of basis functions, $\boldsymbol{B} = \{B_1(x), \cdots, B_K(x)\}^{\mathrm{T}}$. The penalized spline estimator is defined as

$$\widehat{f}(x) = \widehat{\beta}^{\mathrm{T}} \boldsymbol{B}(x), \tag{1.1}$$

where $\widehat{\beta}$ is the minimizer of a penalized least square

$$\sum_{i=1}^{n} \{Y_i - \beta^{\mathrm{T}} \boldsymbol{B}(x_i)\}^2 + \lambda \beta^{\mathrm{T}} D \beta,$$

where D is positive definite matrix and $\lambda > 0$ is a tuning parameter.

In this method, we usually include a relatively large number of basis functions to make the model flexible, and use the roughness penalty $\beta^{\mathrm{T}} D \beta$ to force the estimated curve to be smooth. λ controls the tradeoff between flexibility and variation.

The penalized spline given by (1.1) is quite general, it includes many previous spline variants as special cases. For example, when $\boldsymbol{B}(x)$ are B-spline functions and the penalty is on the divided differences of the coefficients, it is the penalized B-spline method introduced by Eilers and Marx (1996). When $\boldsymbol{B}(x)$ are the natural spline basis with knots on all x_i and the penalty is on $J(f) = \int \{f^{(m)}(x)\}^2 dx$ with

$$D = \int \boldsymbol{B}^m(x)\{\boldsymbol{B}^{(m)}(x)\}^{\mathrm{T}} dx,$$

the penalized spline is equivalent to smoothing spline (Eubank, 1988). Ruppert et al. (2003) propose to use the truncated power series as spline basis,

$$\boldsymbol{B}(x) = \{1, x, x^2, \cdots, x^p, (x-\kappa_1)_+^p, \cdots, (x-\kappa_{K-p-1})_+^p\},$$

and let $D = \mathrm{diag}(\mathbf{0}_{p+1}, \mathbf{1}_{K-p-1})$. Another possibility is let $\boldsymbol{B}(x)$ be the fourier basis, $\{1, \cos(j\pi x), \sin(j\pi x); j = 1, 2, \cdots\}$, and use the same penalty as in smoothing spline, this method is referred as periodic smoothing spline in Eubank (1988).

The roughness penalty approach has a close relation with the mixed effect models, and has been widely used beyond the scope of nonparametric regression discussed above. For example, the roughness penalty idea has been extended to penalized likelihood and penalized quasilikelihood methods for the generalized additive models. See Ruppert et al. (2003) for an overview.

Ramsay and Silverman (1997) also applied the roughness penalty idea in Functional Data Analysis. For example, for the functional linear model that we will discuss in Chapter III,

$$Y_i = \mu + \int_0^1 X_i(t) f(t) dt + \epsilon_i,$$

where they span $X_i(\cdot)$ on a set of basis functions and estimate the unknown coefficient function $f(\cdot)$ by minimizing a penalized least square, with the penalty on the second derivative of f.

1.3 Overview Structure

The following is the general structure of my dissertation.

In Chapter II, I present the results for the project on colon carcinogenesis data. In this data set, the measurements within a crypt are discrete observations on a function, but these functions are correlated within the same subject (rat). I study the nonparametric kernel methods to estimate the spatial correlation between these functions. Asymptotic normal distributions are developed for the proposed estimators. I will also discuss other issues in data analysis, for example bandwidth selection and inference procedure. Simulation studies are also provided to check the performance of the proposed methods.

In Chapter III, I will focus on the theoretical properties for the functional linear model with functional predictor and scalar response. Although I believe the best method for such models should be based splines due to the various desirable properties of spline functions, I will restrict my theoretical derivation to methods based on periodic spline simply because they are much more tractable mathematically. I derive an asymptotic convergence rate for the functional linear model under some periodic boundary conditions, but the result can be inferred to more general spline methods.

In Chapter IV, I will discuss some possible extensions of my work. I will talk about spline methods for functional linear models, apply them to the lipoprotein profile data and compare the results to those of the periodic spline methods discussed in the previous chapter.

All theoretical derivations are given in the appendix.

CHAPTER II

NONPARAMETRIC CORRELATION ESTIMATION FOR THE COLON CARCINOGENESIS DATA

2.1 Introduction

This project concerns kernel-based nonparametric estimation of covariance and correlation functions. Our methods and theory are applicable to longitudinal and spatial data as well as time series data, where observations within the same subject at different time points or locations have strong correlations, which are stationary in time or distance lags. The structure for the observation at a particular time or location within one subject can be very general, for example a vector or even a function.

Our study arises from a colon carcinogenesis experiment. The biomarker that we are interested in is p27, which is a life cycle protein that affects cell apoptosis, proliferation and differentiation. An important goal of the study is to understand the function of p27 in the early stage of the cancer development process. In the experiment, 12 rats were administered azoxymethane (AOM), which is a colon specific carcinogen. After 24 hours, the rats were terminated and a segment of colon tissue was excised from each rat. About 20 colonic crypts were randomly picked along a linear slice on the colon segment. The physical distances between the crypts were measured. Then, within each crypt, we measured cells at different depths within the crypts, and then the expression level of p27 was measured for each cell within the chosen crypts. In this data set, crypts are naturally functional data (Ramsay and Silverman 1997), that the responses within a crypt are coordinated by cell depths. There is a literature about similar data, for example Morris et al. (2001).

However, in this project, we will be focused on a very different perspective. In this application, the spatial correlation between crypts is of biological interest, because it helps answer the question: if we observe a crypt with high p27 expression, how likely are the neighboring crypts to have high p27 expression? We will phrase much of our discussion in terms of this example, but as seen later sections, we have a quite general structure that includes time series as a special case. In that context, the asymptotic theory is as the number of ”time series locations”, i.e., crypts, increases to infinity.

Although motivated by a very specific problem, nonparametric covariance/correlation estimators worth being investigated in their own right. They can be used in a statistical analysis as: (a) an exploratory device to help formulate a parametric model; (b) an intermediate tool to do spatial prediction (kriging); (c) a diagnostic for parametric model; (d) a robust tool to test correlation. Understanding the theoretical properties of the nonparametric estimator is important under any of these situation. A limiting distribution theory would be especially valuable for purpose (d).

There is previous work on the subject of nonparametric covariance estimation. Hall et al. (1994) developed an asymptotic convergence rate of a kernel covariance estimator in a time series setting. They required not only an increasing time domain, but increasingly denser observations. Diggle and Verbyla (1998) suggested a kernel weighted local linear regression estimator for estimating the non-stationary variogram in longitudinal data, without developing asymptotic theory. Guan, Sherman and Calvin (2004) used a kernel variogram estimator when assessing isotropy in geostatistics data. They proved asymptotic normality for their kernel variogram estimator in a geostatistics setting, where they required the spatial locations to be sampled from the field according to a two dimensional homogeneous Poisson process.

As we will show below and as implied by the result from Guan et al. (2004) if the observation locations (or times) in the design are random, Hall’s assumption, namely that the number of observation on a unit domain goes to infinity, is too restrictive and not necessary. However, in the setting of Guan et al., given the sample size, spatial locations are uniformly distributed within the field, which does not fit our problem, where crypt locations within a rat are, in fact, not even close to uniformly distributed.

Our work differs from the previous work on the kernel covariance estimators in the following ways. First, our approach accommodates more complex data structure at each location or time. Secondly, we allow the spatial locations to be sampled in an inhomogeneous way, and as we will show below that this inhomogeneity will affect the asymptotic results and inference procedures. In doing so, we generalize the setting of Guan et al. (2004), and link it to the setting of Hall et al. (1994). Also, Guan et al. (2004) is mainly concerned with comparing variograms on a few pre-selected distance lags, we, on the other hand, are more interested in the correlation as a function. Thirdly, we propose an inference procedure, thus filling a gap in the previous literature.

This chapter is organized as follows. Section 2.2 introduces our model assumptions and estimators, while asymptotic results are given in Section 2.3. A brief analysis of the data motivating this work is given in Section 2.4, where we also discuss bandwidth selection and a procedure to estimate the standard deviation of the correlation estimator. Section 2.5 describes a simulation study, and final comments are given in Section 2.6. All proofs are given in the appendix.

2.2 Model Assumptions and Estimators

The data considered here have the following structure:

- There are are $r = 1, ..., R$ independent subjects, which in our example are rats. We allow $R = 1$.

- The data for each subject have two levels. The first level has an increasing domain, as in time series or spatial statistics, and are the crypts in our example. We label this first level as a ”unit”, and it is these units that have time series or spatial structure in their locations. Within each subject, there are $i = 1, ..., N_r$ such units.

- The second level of the data consists of observations within each of the primary units. In our case, these are the cells within the primary units, the colonic crypts. We will label this secondary level as the ”sub-units”, which are labelled with locations. The locations with the sub-units are on the interval $[0, 1]$. For simplicity, we will assume there are exactly m sub-units (cells) within each unit (crypt), with the j^{th} sub-unit having location (relative cell depth) $x = (j-1)/(m-1)$. However, all theories and methods in our paper will go through if the sub-units take the form of an arbitrary finite set.

- In the time series setting of Hall et al. (1994) or the spatial setting of Guan et al. (2004), $m = 1$.

Let $\Theta(s, x)$ be a random field on $\mathcal{T} \times \mathcal{X}$, where s is the unit (crypt) location and and x is the sub-unit(cell) location, so that $\mathcal{T} = [0, \infty)$, $\mathcal{X} = \{(j-1)/(m-1), j = 1, \cdots, m\}$. Assume that $\Theta_r(\cdot, \cdot)$, $r = 1, \cdots, R$, are independent realizations of $\Theta(\cdot, \cdot)$. We use the short-hand notation $\Theta_{ri}(x) = \Theta_r(S_{ri}, x)$, where S_{ri} is the location of the i^{th} unit (crypt) within the r^{th} subject (rat). Our model for the observed data is that

$$Y_{rij} = \Theta_{ri}(x_j) + \epsilon_{rij}, \qquad (2.1)$$

where Y is the response (logarithm of p27 level), ϵ_{rij} are zero-mean uncorrelated measurement errors with variance σ^2_ϵ, $r = 1, \cdots, R$, $i = 1, \cdots, N_r$ and $j = 1, \cdots, m$ are the indices for subjects (rats), units (crypts) and sub-units (cells). Define $\Theta_r(\cdot) = E_r\{\Theta_{ri}(\cdot)\}$ to be the subject-level mean, and the notation “E_r” refers to expectation conditional on the subject. Another way to understand $\Theta_r(\cdot)$ is to decompose the random field $\Theta_r(\cdot, \cdot)$ into the following random effect model, $\Theta_{ri}(x) = \Theta_r(x) + \Lambda_{ri}(x)$, where $\Theta_r(\cdot)$ is the subject (rat) effect, Λ_{ri} are the zero-mean, spatially correlated unit (crypt) effects.

Within each subject, we assume that the correlation of the mean unit (crypt)-level functions is stationary over the distances between the units. In addition, the covariance between unit locations (s_1, s_2) at sub-unit (cell) locations (x_1, x_2) is assumed to have the following form:

$$\mathcal{V}\{x_1, x_2, \Delta\} \;=\; E[\{\Theta_r(s_1, x_1) - \Theta_r(x_1)\}\{\Theta_r(s_2, x_2) - \Theta_r(x_2)\}], \tag{2.2}$$

where $\Delta = s_1 - s_2$. While we develop general results for model (2.2), in many cases it is reasonable to assume that the covariance function is separable, i.e.,

$$\mathcal{V}(x_1, x_2, \Delta) = G(x_1, x_2)\rho(\Delta). \tag{2.3}$$

When the covariance function is separable, the correlation function at the unit-level, $\rho(\cdot)$, is of interest in itself. In our application, $\rho(\cdot)$ is the correlation between crypts. We provide an estimator of $\rho(\cdot)$ as well as an asymptotic theory for that estimator.

A first estimator for the covariance function has the following form:

$$\begin{aligned}\widehat{\mathcal{V}}(x_j, x_l, \Delta) \;=\; & [\sum_r \sum_i \sum_{k \neq i} K_h\{\Delta_r(i,k) - \Delta\}(Y_{rij} - \overline{Y}_{r\cdot j})(Y_{rkl} - \overline{Y}_{r\cdot l})] \\ & \times [\sum_r \sum_i \sum_{k \neq i} K_h\{\Delta_r(i,k) - \Delta\}]^{-1},\end{aligned} \tag{2.4}$$

where $\overline{Y}_{r\cdot j} = N_r^{-1} \sum_{i=1}^{N_r} Y_{rij}$, $\Delta_r(i,k) = S_{ri} - S_{rk}$, $K_h(\cdot) = h^{-1}K(\cdot/h)$ with K being a kernel function satisfying the conditions in Section 3.

It is usually reasonable to assume that $\mathcal{V}(x_1, x_2, \Delta)$ has some symmetry property, that it is an even function in Δ and $\mathcal{V}(x_1, x_2, \Delta) = \mathcal{V}(x_2, x_1, \Delta)$. However, the estimator defined in (2.4) does not enjoy this property. To see this, we observe that, for $x_j \neq x_l$, although $(Y_{rij} - \overline{Y}_{r\cdot j})(Y_{rkl} - \overline{Y}_{r\cdot l})$ and $(Y_{ril} - \overline{Y}_{r\cdot l})(Y_{rkj} - \overline{Y}_{r\cdot j})$ estimate the same thing, they only contribute to $\widehat{\mathcal{V}}(x_j, x_l, \Delta)$ and $\widehat{\mathcal{V}}(x_j, x_l, -\Delta)$, respectively. We also observe that $\widehat{\mathcal{V}}(x_1, x_2, \Delta) = \widehat{\mathcal{V}}(x_2, x_1, -\Delta)$.

To correct the asymmetry of the covariance estimator, for $\Delta \geq 0$, define

$$\widetilde{\mathcal{V}}(x_j, x_l, \Delta) = [\sum_r \sum_i \sum_{k \neq i} K_h\{|\Delta_r(i,k)| - \Delta\}(Y_{rij} - \overline{Y}_{r\cdot j})(Y_{rkl} - \overline{Y}_{r\cdot l})] \times [\sum_r \sum_i \sum_{k \neq i} K_h\{|\Delta_r(i,k)| - \Delta\}]^{-1}, \quad (2.5)$$

and let $\widetilde{\mathcal{V}}(x_j, x_l, \Delta) = \widetilde{\mathcal{V}}(x_j, x_l, -\Delta)$ for $\Delta < 0$. As shown in the proof of Theorem II.2, for a fixed $\Delta \neq 0$, $\widetilde{\mathcal{V}}(x_1, x_2, \Delta)$ is asymptotically equivalent to $\{\widehat{\mathcal{V}}(x_1, x_2, \Delta) + \widehat{\mathcal{V}}(x_1, x_2, -\Delta)\}/2$.

In addition, when the separable structure (2.3) is assumed, define estimators

$$\widehat{G}(x_1, x_2) = \widetilde{\mathcal{V}}(x_1, x_2, 0), \quad (2.6)$$

and

$$\widehat{\rho}(\Delta) = \{\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \leq x_1} \widetilde{\mathcal{V}}(x_1, x_2, \Delta)\}/\{\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \leq x_1} \widehat{G}(x_1, x_2)\}. \quad (2.7)$$

2.3 Asymptotic Results

The following are our model assumptions. Each subject (rat) is of length L, where in our example L is the length of the segment of tissue from each rat. The units (crypts) are located on the interval $[0, L]$, and in our asymptotics we let $L \to \infty$, so that we have an increasing domain. Suppose that the positions of the units (crypts) within the r^{th} subject (rat) are $S_{r1}, \cdots, S_{rN_r}$, where the S_{ri}'s are points from an inhomogeneous Poisson process on $[0, L]$. Then $\Delta_{r,ik} = S_{ri} - S_{rk}$. The definition of an inhomogeneous Poisson process is adopted from Cressie (1993). We assume the inhomogeneous Poisson process has a local intensity $\nu g^*(s)$, where ν is a positive constant and $g^*(s) = g(s/L)$ for a continuous density function $g(\cdot)$ on $[0, 1]$.

A special case of our setting is that $g(\cdot)$ is a uniform density function and the units (crypts) are sampled according to a homogeneous Poisson process. This is the setting investigated in Guan et al. (2004). Our setting resembles that of Hall et al. (1994) in the sense that we also model the unit locations as random variables with the same distribution: in our setting, the number of units within a subject (rat) is $N_r \sim \mathrm{Poisson}(\nu L)$; given N_r, $S_{r1}/L, \cdots, S_{r,N_r}/L$ are independent and identically distributed with density $g(\cdot)$. By properties of Poisson processes, $N_r/L = O(\nu)$ almost surely, as $L \to \infty$, that is, the number of units (crypts) on a unit length tends to a constant. It is worth noting that Hall et al. (1994) required this ratio to go to infinity. We require less samples on the domain than do Hall et al. (1994).

In what follows, we provide a list of definitions and conditions needed to present our theoretical findings.

1. We assume that $g(\cdot)$ is continuous and $c_1 \geq g(t) \geq c_2 > 0$ for all $t \in [0, 1]$. Suppose t_i, $i = 1, 2, 3, 4$, are independent random variables with density $g(\cdot)$, define f_1, f_2, f_3 to be the density for $t_1 - t_2$, $(t_1 - t_2, t_3 - t_2)$, $(t_1 - t_2, t_3 - t_4, t_2 - t_4)$, respectively. Since $g(\cdot)$ is bounded, one can easily derive that $f_1(0)$, $f_2(0, 0)$ and $f_3(0, 0, 0)$ are positive. We also assume that f_2 is Lipschitsz continuous in the neighborhood of $\mathbf{0}$, i.e. $|f_2(u, v) - f_2(0, 0)| \leq \lambda_1|u| + \lambda_2|v|$, for $\forall u, v$ and some fixed constants $\lambda_1, \lambda_2 > 0$.

2. Assume $\mathcal{V}(x_1, x_2, \Delta)$ has two bounded continuous partial derivatives in Δ, and that $\sup_{x_1,x_2} \int |\mathcal{V}(x_1, x_2, \Delta)|d\Delta < \infty$.

3. Let

$$\begin{aligned}
&\mathcal{M}(x_1, x_2, x_3, x_4, u, v, w) \\
&\quad = E_r\Big[\{\Theta_{ri_1}(x_1) - \Theta_r(x_1)\}\{\Theta_{ri_2}(x_2) - \Theta_r(x_2)\}\{\Theta_{ri_3}(x_3) - \Theta_r(x_3)\} \\
&\qquad \{\Theta_{ri_4}(x_4) - \Theta_r(x_4)\}|\Delta_r(i_1, i_2) = u, \Delta_r(i_3, i_4) = v, \\
&\qquad \Delta_r(i_2, i_4) = w\Big] - \mathcal{V}(x_1, x_2, u)\mathcal{V}(x_3, x_4, v).
\end{aligned}$$

We assume $\mathcal{M}$ has bounded partial derivatives in u, v and w, and

$$\sup_{x_1,x_2,x_3,x_4,u,v} \int |\mathcal{M}(x_1,x_2,x_3,x_4,u,v,w)|dw < \infty. \tag{2.8}$$

4. Denote $b_r(x_1,x_2,\Delta) = L^{-1}\sum_i \sum_{k\neq i} K_h\{\Delta - \Delta_r(i,k)\}\{Y_r(S_{ri},x_1) - \Theta_r(x_1)\}$ $\times\{Y_r(S_{rk},x_2) - \Theta_r(x_2)\}$. We assume that, for any fixed Δ,

$$\sup_{L,x_1,x_2} E(|\text{var}^{-1/2}\{b_r(x_1,x_2,\Delta)\}[b_r(x_1,x_2,\Delta) - E\{b_r(x_1,x_2,\Delta)\}]|^{2+\eta}) \\ \leq C_\eta < \infty \tag{2.9}$$

   for some $\eta > 0$.

5. Let $\mathcal{F}(T)$ be the σ-algebra generated by $\{\Theta(s,x), s \in T, x \in \mathcal{X}\}$, for any Borel set $T \subset \mathcal{T}$. Assume that the random field satisfies the following mixing condition

$$\begin{aligned} \alpha(\tau) &= \sup_t[|P(A_1 \cap A_2) - P(A_1)P(A_2)| : \\ &\qquad\qquad A_1 \in \mathcal{F}\{(-\infty,t]\}, A_2 \in \mathcal{F}\{[t+\tau,\infty)\}] \\ &= O(\tau^{-\delta}) \text{ for some } \delta > 0. \end{aligned} \tag{2.10}$$

6. The kernel function K is a symmetric, continuous probability density function, supported on $[-1,1]$. Define $\sigma_K^2 = \int u^2K(u)du$ and $R_K = \int K^2(v)dv$.

7. Assume that m and R are fixed numbers, $L \to \infty$, $h \to 0$, $Lh \to \infty$, and $Lh^5 = O(1)$.

In assumption 1, we are imposing some regularity conditions on g and f_i. In fact, when g is differentiable f_i are piecewise differentiable, but usually not differentiable at $\mathbf{0}$. However, the Lipschitz condition on f_2 is easily satisfied when, for example, g is Lipschitz. Since f_1 is a marginal density of f_2's, this condition means f_1 is also Lipschitz at 0.

Since we are estimating the covariance function, which is the second moment function, we need a regularity condition on the 4^{th} moment function as in (2.8). Condition (2.9) may seem a little too strong at first sight, but it really is just a condition to bound the tail probability of our statistics. In fact, for example, if we have an assumption analogous to (2.8) for the 8^{th} moment of $\Theta_r(s,x)$, one can use arguments as in Lemma A.3 to show that $E([b_r(x_1,x_2,\Delta) - E\{b_r(x_1,x_2,\Delta)\}]^4) = O(L^{-3}h^{-3})$, therefore condition (2.9) is satisfied for $\eta = 2$.

Denote $\mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta) = \partial^2 \mathcal{V}(x_1,x_2,\Delta)/\partial\Delta^2$. Let $\mathcal{V}(\Delta)$, $\widehat{\mathcal{V}}(\Delta)$ and $\widetilde{\mathcal{V}}(\Delta)$ denote the vectors collecting $\mathcal{V}(x_1,x_2,\Delta)$, $\widehat{\mathcal{V}}(x_1,x_2,\Delta)$ and $\widetilde{\mathcal{V}}(x_1,x_2,\Delta)$ respectively, for all distinct pairs of (x_1,x_2). The following are our main results for the asymptotic theories, all proofs are provided in the appendix. Note that Theorem II.1 refers to $\widehat{\mathcal{V}}(\cdot)$ in (2.4), while Theorem II.2 refers to $\widetilde{\mathcal{V}}(\cdot)$ in (2.5).

<u>Theorem</u> **II.1** Under assumptions 1-7, for $\Delta \neq \Delta'$, we have

$$(RLh)^{1/2} \begin{bmatrix} \widehat{\mathcal{V}}(\Delta) - \mathcal{V}(\Delta) - \text{bias}\{\widehat{\mathcal{V}}(\Delta)\} \\ \widehat{\mathcal{V}}(\Delta') - \mathcal{V}(\Delta') - \text{bias}\{\widehat{\mathcal{V}}(\Delta')\} \end{bmatrix}$$
$$\Rightarrow \text{Normal}\left[0, \{\nu^2 f_1(0)\}^{-1} \begin{pmatrix} \Sigma(\Delta) & C(\Delta,\Delta') \\ C^T(\Delta,\Delta') & \Sigma(\Delta') \end{pmatrix}\right],$$

where the asymptotic bias bias$\{\widehat{\mathcal{V}}(\Delta)\}$ is a vector having entries bias$\{\widehat{\mathcal{V}}(x_1,x_2,\Delta)\} = \sigma_K^2 \mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta)h^2/2$, $\Sigma(\Delta)$ is the covariance matrix with the entry corresponding to cov$\{\widehat{\mathcal{V}}(x_1,\, x_2,\Delta), \widehat{\mathcal{V}}(x_3,x_4,\Delta)\}$ equal to $R_K\{\mathcal{M}(x_1,x_2,x_3,x_4,\Delta,\Delta,0) + I(x_2 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_1,x_3,0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2,x_4,0) + (x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\} + I(\Delta = 0)R_K \{\mathcal{M}(x_1,x_2,x_3,x_4,0,0,0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) + I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1,x_4,0) + I(x_1 = x_4, x_2 = x_3)\sigma_\epsilon^4\}$; $C(\Delta,\Delta')$ is the matrix with the entry corresponding to cov$\{\widehat{\mathcal{V}}(x_1,x_2,\Delta), \widehat{\mathcal{V}}(x_3,x_4,\Delta')\}$ equal to $I(\Delta' = -\Delta)\{\mathcal{M}(x_1,x_2,x_3,x_4,\Delta,-\Delta,-\Delta) + I(x_2 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_1,x_4,0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) + I(x_1 = x_4, x_2 = x_3)\sigma_\epsilon^4\}$.

<u>Theorem</u> **II.2** Under assumptions 1-7, for $\Delta \neq \pm\Delta'$, we have

$$(RLh)^{1/2}\begin{bmatrix} \widetilde{\mathcal{V}}(\Delta) - \mathcal{V}(\Delta) - \text{bias}\{\widetilde{\mathcal{V}}(\Delta)\} \\ \widetilde{\mathcal{V}}(\Delta') - \mathcal{V}(\Delta') - \text{bias}\{\widetilde{\mathcal{V}}(\Delta')\} \end{bmatrix}$$

$$\Rightarrow \text{Normal}\left[0, \{\nu^2 f_1(0)\}^{-1}\begin{pmatrix} \Omega(\Delta) & 0 \\ 0 & \Omega(\Delta') \end{pmatrix}\right],$$

where bias$\{\widetilde{\mathcal{V}}(\Delta)\}$ is a vector with entries bias$\{\widetilde{\mathcal{V}}(x_1, x_2, \Delta)\} = \sigma_K^2 \mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta)h^2/2$, $\Omega(\Delta)$ is the covariance matrix with the entry corresponding to cov$\{\widetilde{\mathcal{V}}(x_1, x_2, \Delta), \widetilde{\mathcal{V}}(x_3, x_4, \Delta)\}$ equal to $(1/2)R_K\{\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta, 0) + \mathcal{M}(x_1, x_2, x_3, x_4, \Delta, -\Delta, -\Delta) + I(x_2 = x_4)\sigma_\epsilon^2\ \mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, 0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4 + I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2, x_3, 0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\} + I(\Delta = 0)(1/2)R_K\{2\mathcal{M}(x_1, x_2, x_3, x_4, 0, 0, 0) + I(x_2 = x_4)\sigma_\epsilon^2\mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, 0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4 + I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2, x_3, 0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}$.

<u>Corollary</u> **II.1** Suppose the covariance function has the separable structure in (2.3) with $\sum_{x_1}\sum_{x_2 \le x_1} G(x_1, x_2) \neq 0$, and $\widehat{\rho}(\Delta)$ is defined in (2.7). Then for $\Delta \neq 0$, we have

$$(RLh)^{1/2}[\widehat{\rho}(\Delta) - \rho(\Delta) - \text{bias}\{\widehat{\rho}(\Delta)\}] \Rightarrow \text{Normal}[0, \{\nu^2 f_1(0)\}^{-1}\sigma_\rho^2(\Delta)],$$

where bias$\{\widehat{\rho}(\Delta)\} = \{\rho^{(2)}(\Delta) - \rho(\Delta)\rho^{(2)}(0)\}\sigma_K^2 h^2/2$ is the asymptotic bias of $\widehat{\rho}(\Delta)$, $\sigma_\rho^2(\Delta) = \{\sum_{x_1}\sum_{x_2 \le x_1} G(x_1, x_2)\}^{-2}\{\mathbf{1}^T\Omega(\Delta)\mathbf{1} + \rho^2(\Delta)\mathbf{1}^T\Omega(0)\mathbf{1}\}$.

We have the following remarks on our theoretical results:

1. The measurement errors in (2.1) affect the covariance estimator mainly though the nugget effect (Cressie, 1993). In our covariance estimators (2.4) and (2.5), we get rid of the nugget effect by excluding the $k = i$ terms in the summation. As a result, the measurement errors do not introduce bias to our covariance estimators. However,

they do affect the variation of the covariance estimators and hence the correlation estimator, because σ_ϵ^2 is in the variance expressions for all our estimators.

2. The result in Theorem II.2 suggest that the covariance estimators at different distance lag are asymptotically independent. This result may seem counterintuitive. It is caused by the kernel smoothing: we choose the bandwidth to make this happen. This result holds for two fixed values, Δ and Δ', when h goes to 0.

2.4 Data Analysis

In this section we apply our methods to study the between-crypt dependence in the carcinogenesis experiment. Recall that the main subjects are rats, the units of interest are colonic crypts and the sub-units within a unit are cells, at which we observe the logarithms of p27 in a cell. The sub-unit locations that we work with in this illustration are at $x = 0, 0.1, 0.2, \cdots, 1.0$. We discuss three key issues in our analysis, namely bandwidth selection, standard error estimation and positive semi-definite adjustment in the following three subsections.

2.4.1 Bandwidth Selection

2.4.1.1 Global Bandwidth

Diggle and Verbyla (1998) suggested a cross-validation procedure to choose the bandwidth for a kernel variogram estimator. We modify their procedure into the following two types of 'leave-one-subject-out' cross-validation criteria. The first is based on prediction error without assuming any specific covariance structure, and is given as

$$CV_1(h) = \sum_r \sum_{|\Delta_r(i,k)|<\Delta_0} \sum_{j=1}^m \sum_{l=1}^m [v_{r,ik}(x_j, x_l) - \tilde{\mathcal{V}}_{(-r)}\{x_j, x_l, \Delta_r(i,k)\}]^2, \qquad (2.11)$$

where $v_{r,ik}(x_j, x_l) = (Y_{rij} - \overline{Y}_{r\cdot j})(Y_{rkl} - \overline{Y}_{r\cdot l})$, $\tilde{\mathcal{V}}_{(-r)}(x_1, x_2, \Delta)$ is the kernel covariance estimator using bandwidth h, as defined in (2.5), with all information on the r^{th} subject (rat) left out. Here we focus on the range $|\Delta_r(i,k)| < \Delta_0$, where Δ_0 is a pre-chosen cut-off point. The criterion $CV_1(h)$ thus evaluates the prediction error for different h within the range of $|\Delta_r(i,k)| < \Delta_0$.

Cross-validation criterion (2.11) assumes no specific covariance structure, while our second cross-validation criterion takes into account the separable structure in (2.3), and is given as

$$CV_2(h) = \sum_r \sum_{|\Delta_r(i,k)|<\Delta_0} \sum_{j=1}^{m} \sum_{l=1}^{m} [v_{r,ik}(x_j, x_l) - \widehat{G}_{(-r)}(x_j, x_l)\widehat{\rho}_{(-r)}\{\Delta_r(i,k)\}]^2 \quad (2.12)$$

where $\widehat{G}_{(-r)}(x_1, x_2)$ and $\widehat{\rho}_{(-r)}(\Delta)$ are the estimators of G and ρ defined in (2.6) and (2.7), with the r^{th} subject (rat) left out.

We evaluated both criteria to estimate the bandwidth h. We choose $\Delta_0 = 500$ microns. The first two columns of Table 1 gives the minimum points and minimum values of the two cross-validation criterions.

By observing Table 1, we find the two criteria gave almost identical minimum values. Since the cross-validation scores are estimates of the prediction errors, the two cross-validation criteria represent prediction errors with or without the separable structure (2.3). The phenomenon, that $CV_1(\cdot)$ and $CV_2(\cdot)$ have almost the same minimum values, suggests that the separability assumption (2.3) fits the data well.

2.4.1.2 Two Bandwidths

The independent variables in the kernel estimator are $|\Delta_r(i,k)|$ for all pairs of crypts within one subject. As shown in Figure 2, the distribution of $|\Delta_r(i,k)|$ that are less than 1000 microns, even more than the target range of interest, is locally somewhat akin to a uniform distribution.

As a robustness check on the global bandwidth, we repeated our analysis, except we used one bandwidth for $|\Delta| \leq 200$ microns, and we used a second bandwidth for $|\Delta| >$

200, and then repeated the cross-validation calculations in (2.11) and (2.12). The minimum values of the two cross-validation criterions are reported in the 3^{rd} column of Table 1.

Comparing the results in columns 2 and 3 in Table 1, we find the minimum values of the cross-validation functions did not change much, i.e. an extra smoothing parameter did not substantially reduce the prediction error for the domain $|\Delta| \leq 500$ microns. In other words, it appears sufficient to use a global bandwidth to estimate $\rho(\Delta)$ for $|\Delta| \leq 500$. For the following analysis, we use the bandwidth $h = 122$ microns, as suggested by CV_2.

	optimal h	min CV score	min score, 2 par
CV_1	124.2334	6.5073	6.4867
CV_2	122.7202	6.4955	6.4788

Table 1: Outcomes of two cross-validation procedures on the carcinogenesis p27 data. The data used in the validation are those with Δ values less than $\Delta_0 = 500$ microns. The first column gives the optimal global bandwidth, the second column gives the value of the cross-validation function at the optimal global bandwidth; the third column gives the minimum value of cross-validation functions using two different smoothing parameters.

2.4.2 Standard Error Estimation

Our primary goal in this section is to construct an estimate of the standard error for $\widehat{\rho}(\Delta)$.

The asymptotic variance of $\widehat{\rho}(\Delta)$ has a very complicated form, which involves the 4^{th} moment function of the random field, $\mathcal{M}(x_1, x_2, x_3, x_4, u, v, w)$. With so many estimates of higher order moments involved, a plug-in method, while feasible, is not desirable. We instead use a bootstrap method to estimate the variance directly.

In our model assumptions, the number of subjects (rats) R is fixed, which means that bootstrapping solely on the subject level will not give a consistent estimator of the variance. Consequently, we decided to sub-sample within each subject. When the data

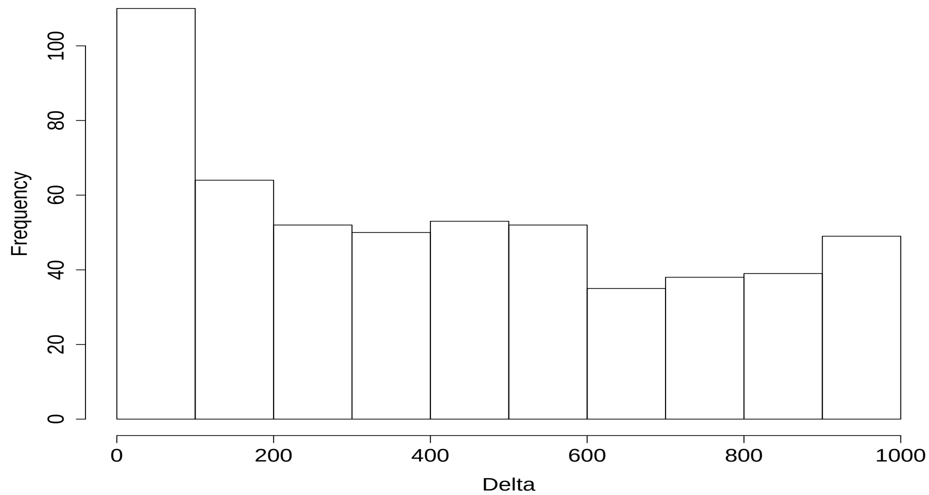


Figure 2: Histogram of $|\Delta_r(i,k)|$ in the carcinogenesis p27 data. $|\Delta|$ less than 1000 microns are considered.

are dependent, block bootstrap methods have been investigated and used, see Shao and Tu (1995). Politis and Sherman (2001) also justified using a block sub-sampling method to estimate the variance of a statistic when the data are from a marked point process. Our data can be viewed as a marked inhomogeneous Poisson process. However, the inhomogeneity does require a modification of their procedure: if we sub-sample a block from each subject and compute the statistic $\widehat{\rho}(\Delta)$ by combining these blocks, then the variance of the statistic depends on the locations of these blocks.

By letting $R = 1$ in Corollary II.1, our theory implies that if the number of units goes to infinity, each subject will provide a consistent estimator of $\rho(\Delta)$. Now, suppose the Poisson process for each subject has a different local intensity, $\nu_r g_r^*(s)$, $r = 1, \cdots, R$. With a slight modification of our theoretical derivations, one can show that,

$$\{\sum_{r=1}^{R} \nu_r^2 f_{r,1}(0) Lh\}^{1/2} [\widehat{\rho}(\Delta) - \rho(\Delta) - \text{bias}\{\widehat{\rho}(\Delta)\}] \Rightarrow \text{Normal}\{0, \sigma_\rho^2(\Delta)\}$$

where $f_{r,1}(t) = \int g_r(t+u) g_r(u) du$, $r = 1, \cdots, R$, are the counterparts of $f_1(t)$ used in Theorem II.1, II.2 and Corollary II.1.

Define $A(\Delta) = \sum_r \sum_i \sum_{k \neq i} K_h\{\Delta_r(i,k) - \Delta\}$, then by Lemma A.2,

$$A(\Delta)/\{\sum_{r=1}^{R} \nu_r^2 f_{r,1}(0)L\} \to 1, \text{ in } L^2.$$

Now we propose our weighted bootstrap procedure:

1. Re-sample R subjects (rats) with replacement from the original collection of subjects.

2. Within each re-sampled subject, randomly sub-sample a block with length L^*.

3. Combine the R blocks as our re-sampled data, compute $\widehat{\rho}(\Delta)$ and $A(\Delta)$ using the re-sampled data, with the same bandwidth h as for the kernel estimator (2.7).

4. Repeat steps 1-3 B times, denoting the results from the b^{th} iteration as $\widehat{\rho}_b^*(\Delta)$ and $A_b^*(\Delta)$.

5. Obtain the estimator of the standard deviation as

$$\widehat{\text{sd}}\{\widehat{\rho}(\Delta)\} = [A^{-1}(\Delta)B^{-1}\sum_{b=1}^{B} A_b^*(\Delta)\{\widehat{\rho}_b^*(\Delta) - \widehat{\rho}_{\cdot}^*(\Delta)\}^2]^{1/2},$$

   where $\widehat{\rho}_{\cdot}^*(\Delta) = B^{-1}\sum_{b=1}^{B} \widehat{\rho}_b^*(\Delta)$.

The block length L^* should increase slowly with L. Politis and Sherman (2001) suggested taking $L^* = L^c$, for some $0 < c < 1$, but choosing a good block length under a finite sample size is still a challenging problem. One operational idea in our context is to choose L^* such that the correlation dies out outside the block but still keep a relatively large numbers of blocks. In our analysis, we took $L^* = 1$ cm (=10,000 microns). We used the same choice of L^* in our simulation study and got quite successful results.

Figure 3 shows the kernel estimator $\widehat{\rho}(\cdot)$ and $\widehat{\rho} \pm 1$ standard deviation. The plot implies that the correlation is practically zero when the crypt distance is larger than about 500 microns.

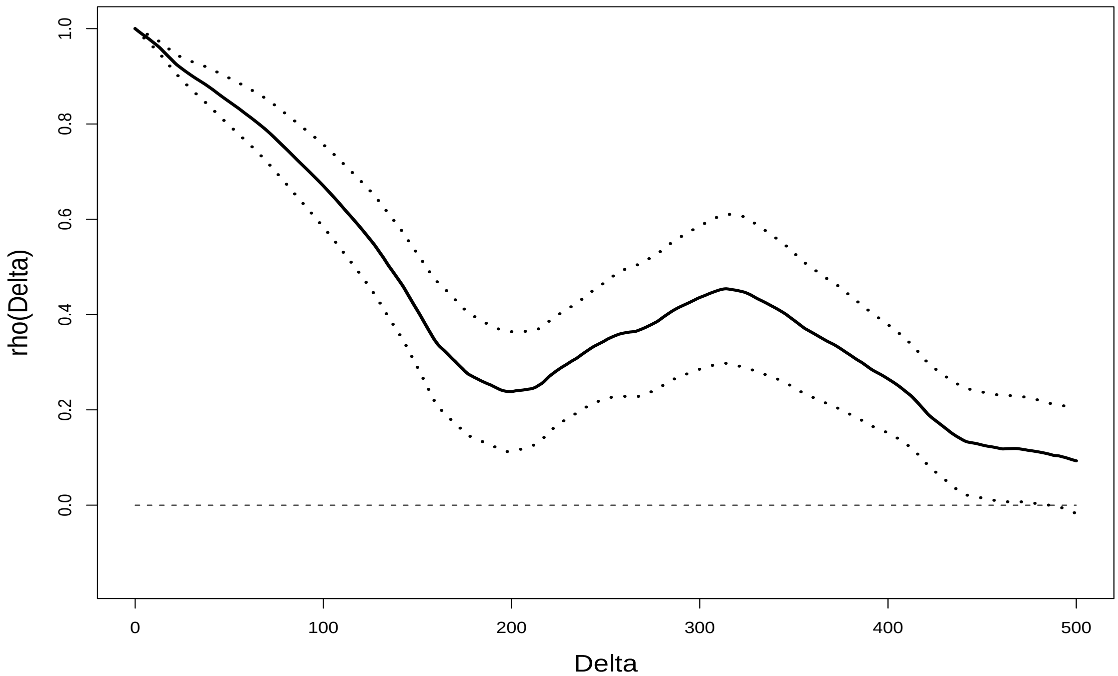


Figure 3: The estimate and the standard deviation band of ρ for the carcinogenesis p27 data. The solid curve is $\widehat{\rho}(\Delta)$ with bandwidth $h = 122$ microns. The dotted curves are $\widehat{\rho}(\Delta) \pm \widehat{SD}\{\widehat{\rho}(\Delta)\}$.

2.4.3 *Positive Semi-Definite Adjustment*

By definition, $\rho(\Delta)$ is a stationary correlation function, therefore is positive semi-definite, i.e. $\int\int \rho(\Delta_1 - \Delta_2)\omega(\Delta_1)\omega(\Delta_2)d\Delta_1 d\Delta_2 \geq 0$ for all integrable functions $\omega(\cdot)$. By Bochner's theorem, the positive semi-definiteness is equivalent to nonnegativity of the Fourier transformation of ρ, i.e. $\rho^+(\theta) \geq 0$ for all θ, where $\rho^+(\theta) = \int_{-\infty}^{\infty} \rho(\Delta)\exp(i\theta\Delta)d\Delta$ $= 2\int_0^{\infty} \rho(\Delta)\cos(\theta\Delta)d\Delta$.

To make $\widehat{\rho}$ a valid correlation function, we apply an adjustment procedure suggested by Hall (1994). First, we compute the Fourier transformation of $\widehat{\rho}(\cdot)$,

$$\widehat{\rho}^+(\theta) = 2\int_0^{\infty} \widehat{\rho}(\Delta)\cos(\theta\Delta)d\Delta.$$

In practice, we can not accurately estimate $\rho(\Delta)$ for a large Δ because of data constraints.

So, what we should do is to multiply $\widehat{\rho}$ by a weight function $w(\Delta) \le 1$, and let

$$\widehat{\rho}^{+}(\theta) = 2\int_0^\infty \widehat{\rho}(\Delta) w(\Delta) \cos(\theta\Delta) d\Delta.$$

Possible choices of $w(\cdot)$, suggested by Hall et al. (1994), are $w_1(\Delta) = I(|\Delta| \le D)$ for some threshold value $D > 0$; and $w_2(\Delta) = 1$ if $|\Delta| < D_1$, $(D_2 - |\Delta|)/(D_2 - D_1)$ if $D_1 \le |\Delta| \le D_2$, 0 if $|\Delta| > D_2$.

Now, let $\theta_0 = \inf\{\theta : \widehat{\rho}^{+}(\theta) < 0, \theta \ge 0\}$, then the adjusted estimator is defined by

$$\widetilde{\rho}(\Delta) = (2\pi)^{-1} \int_{-\theta_0}^{\theta_0} \widehat{\rho}^{+}(\theta) \cos(\theta\Delta) d\theta.$$

Figure 4 shows $\widehat{\rho}(\cdot)$ and $\widetilde{\rho}(\cdot)$ for the colon carcinogenesis data. The size of the correlation even at 200-300 microns is surprising. We have done other, parametric analysis that will be reported elsewhere with a Matérn correlation structure, and this parametric analysis yields correlation estimates at 200-300 microns that are very similar to those seen in Figure 4.

2.5 Simulation Studies

We present two simulation studies to illustrate the numerical performance of the kernel correlation estimation under different settings.

2.5.1 Simulation 1

Our first simulation study is to mimic the colon carcinogenesis data, so that the result could be inferred to evaluate the performance of our estimators in the data analysis and to justify our choice of tuning parameters.

The simulated data arise from the model

$$Y_r^*(s_{ri}, x_j) = \Theta_r^*(s_{ri}, x_j) + \epsilon_{rij}^*,$$

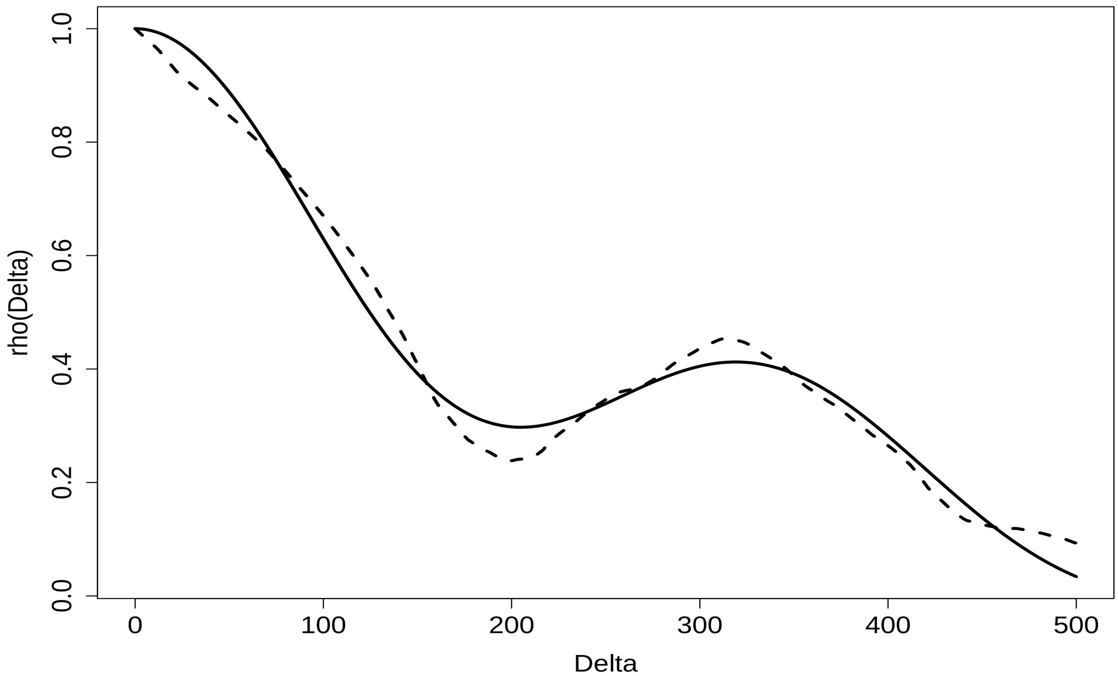


Figure 4: Positive semi-definite adjusted estimate of $\rho(\Delta)$ for the carcinogenesis p27 data. The dashed curve is the unadjusted correlation estimate $\widehat{\rho}(\Delta)$, while the solid curve is the adjusted estimate $\widetilde{\rho}(\Delta)$.

where $\Theta_r^*(s,x)$ is the r^{th} replicate of a zero-mean Gaussian random field $\Theta^*(s,x)$, $r = 1,\cdots,12$. As in our data analysis, x takes values in $\{0.0, 0.1, \cdots, 0.9, 1.0\}$. We used the actual unit (crypt) locations from the data as the sample locations s_{ri} in the simulated data. In addition, $\Theta^*(s,x)$ has covariance structure (2.2) and (2.3), with

$$G^*(x_1,x_2) = (\sum_{r=1}^{12} N_r)^{-1} \sum_{r=1}^{12} \sum_{i=1}^{N_r} \{Y_{ri}(x_1) - \overline{Y}_{r\cdot}(x_1)\}\{Y_{ri}(x_2) - \overline{Y}_{r\cdot}(x_2)\}, \quad (2.13)$$

which is computed from the data, and $\rho^*(\Delta)$ chosen from the Matérn correlation family $\rho^*(\Delta;\phi,\kappa) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(\Delta/\phi)^{\kappa}K_{\kappa}(\Delta/\phi)$, where $K_{\kappa}(\cdot)$ is the modified Bessel function, see Stein (1999). In our simulation, we chose $\kappa = 1.5$ and $\phi = 120$ microns. In addition, the ϵ_{rij}^* are independent identically distributed with Normal$(0,\sigma_{\epsilon^*}^2)$. For $\sigma_{\epsilon^*}^2$, we use an estimate of σ_{ϵ}^2 from the data: $\sigma_{\epsilon^*}^2 = \frac{1}{11}\sum_{j=1}^{11}\{G^*(x_j,x_j) - \widehat{G}(x_j,x_j)\}$, where

$x_j = (j-1)/10$, $j = 1, \cdots, 11$, G^* and $\widehat{G}$ are defined in (2.13) and (2.6), respectively.

For each simulated data set, we computed $\widehat{\rho}(\Delta)$ and the standard deviation estimator $\widehat{SD}\{\widehat{\rho}(\Delta)\}$ that we proposed in Section 2.4.2, for bandwidth $h = 120$ and 200 microns. When doing the bootstrap, we used block size $L^* = 1$ cm, as we did in the p27 data analysis. We repeated the simulation 200 times.

Figure 5 shows the means, 5% and 95% pointwise percentiles of $\widehat{\rho}$ for the two bandwidths, and compares them to the truth ρ^*. Obviously, as expected from the theory, the larger bandwidth incurs the bigger bias. By the plots, it seems that when $h = 120$ the kernel estimator $\widehat{\rho}$ behaves quite well. We compare the true bias from the simulation study to the asymptotic bias computed with the true correlation function ρ^*, under bandwidth $h = 120$. We find the difference between the two are less than 0.04. This means the bias shown in Figure 5 is explainable by our asymptotic theory.

In Fig. 6, we show the pointwise standard deviation of $\widehat{\rho}$ from the simulation and the mean of the bootstrap standard deviation estimates. The closeness of the two curves implies that our bootstrap procedure in Section 2.4.2 gives an approximately unbiased estimator of the true standard deviation, which also implies that our choice of block length, $L^* = 1$ cm, is reasonable. In our simulation, we also tried other block sizes, and the results are almost the same.

2.5.2 Simulation 2

We also provide another simulation study to justify our theoretical assertion that when the locations or times are from an inhomogeneous Poisson process, we have a consistent estimator for ρ as $L \to \infty$. Also, we intend to show the usefulness of a nonparametric correlation estimator in a situation that an 'off-the-shelf' parametric model fails to fit the data.

We found that when the spectrum density is a multi-mode mixture density function

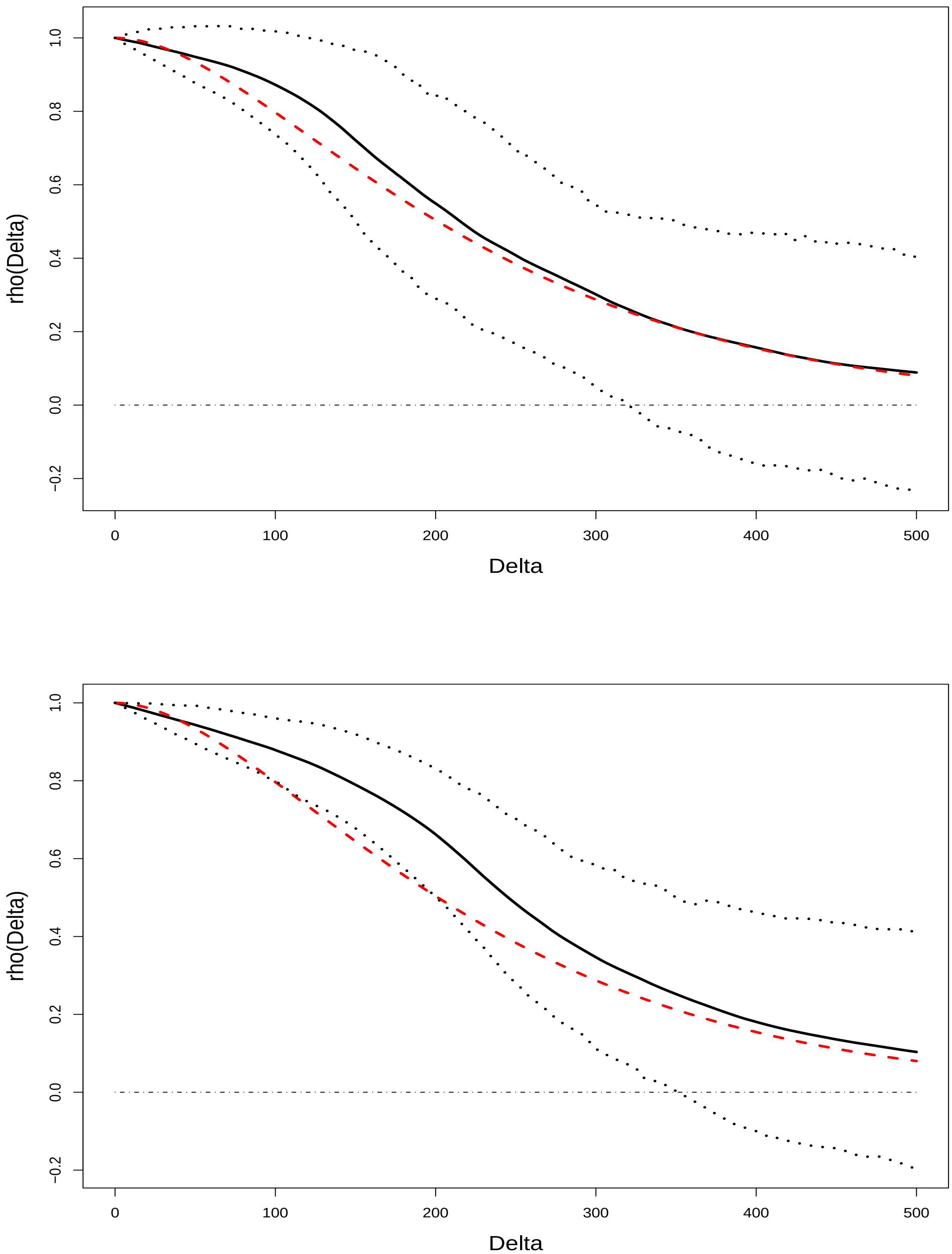


Figure 5: Plots of $\widehat{\rho}(\Delta)$ in the simulation study. Upper panel: $h = 120$; lower panel: $h = 200$. In each plot, the solid curve is the mean of $\widehat{\rho}(\cdot)$, the dashed curve is the true correlation function $\rho(\cdot)$, and the dotted curves are the 5% and 95% pointwise percentiles of $\widehat{\rho}$, respectively.

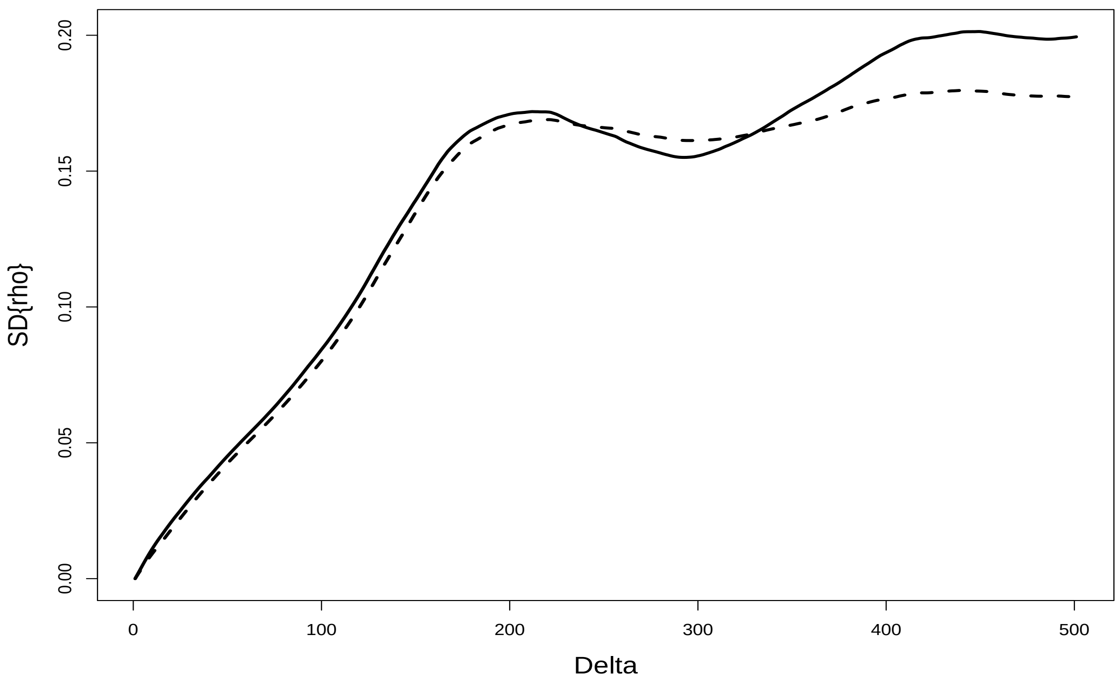


Figure 6: Standard deviation of $\widehat{\rho}$. The solid curve is the pointwise standard deviation of $\widehat{\rho}$ from the simulation, and the dashed curve is the mean of the 200 bootstrap standard deviation estimates. The bandwidth $h = 120$ was used.

like in the upper panel of Fig. 7, the correlation function will have bumpy shape as the dashed curve in the second plot in Fig. 7, which is also the target correlation function in our second simulation study. We simulate only one time series with correlation function given in Fig. 7, and we observe the process on a prolonged time domain $[0, L]$. For simplicity, the observation at each time point is a single value. The observation times are sampled from an inhomogeneous Poisson process with local intensity function $\nu g(\cdot/L)$, where we take $g(\cdot)$ to be a truncated normal density function on $[0, 1]$. The expected number of time points is set to be 500. We also impose some measurement errors to our observations.

We simulated the marked Poisson process described above for 200 times, and computed our kernel correlation estimator for each simulated data set. In the second plot of Fig.

7, the mean of our kernel correlation estimator is given by the solid curve, while the dotted curve is the best approximation to the true correlation function from the Matérn family. We also make a comparison for mean of our bootstrap standard deviation estimator with the true pointwise standard deviation curve in Fig. 7.

As one can see, our nonparametric method can consistently estimate a non-monotone correlation function as we chose in this simulation study, while many parametric models would not be consistent even with large sample size, simply because of their restricted shapes.

2.6 Discussion

We have proposed an estimator of stationary correlation functions for longitudinal or spatial data, where within-subject observations have a complex data structure. The application we presented has a functional data flavor, in that each unit (crypt) in a ”time series” has sub-units (cells) the values from which can be viewed as a function. However, in this paper, we have focused on estimating the spatial correlation between the units.

We established an asymptotic normal limit distribution for the proposed estimator. The techniques used in our theoretical derivation were significantly different from the standard kernel regression literature. In our theoretical framework, as long as we have an increasing number of observations within a subject, each subject yields a consistent estimate of the correlation function. Our method and theory are especially useful to the cases that the number of subject is limited but we have a relatively large number of repeated measurements within each subject. Since having more subjects will just further reduce the variation of the estimator, our main theorems hold when R goes to infinity as well. In that case, we need to replace the condition that $Lh^5 = O(1)$ in assumption 7 in Section 2.3 with $RLh^5 = O(1)$. In fact, when the number of subject $R \to \infty$, we can consistently estimate the within-subject covariance without a large number of units within each subject. For example, Yao,

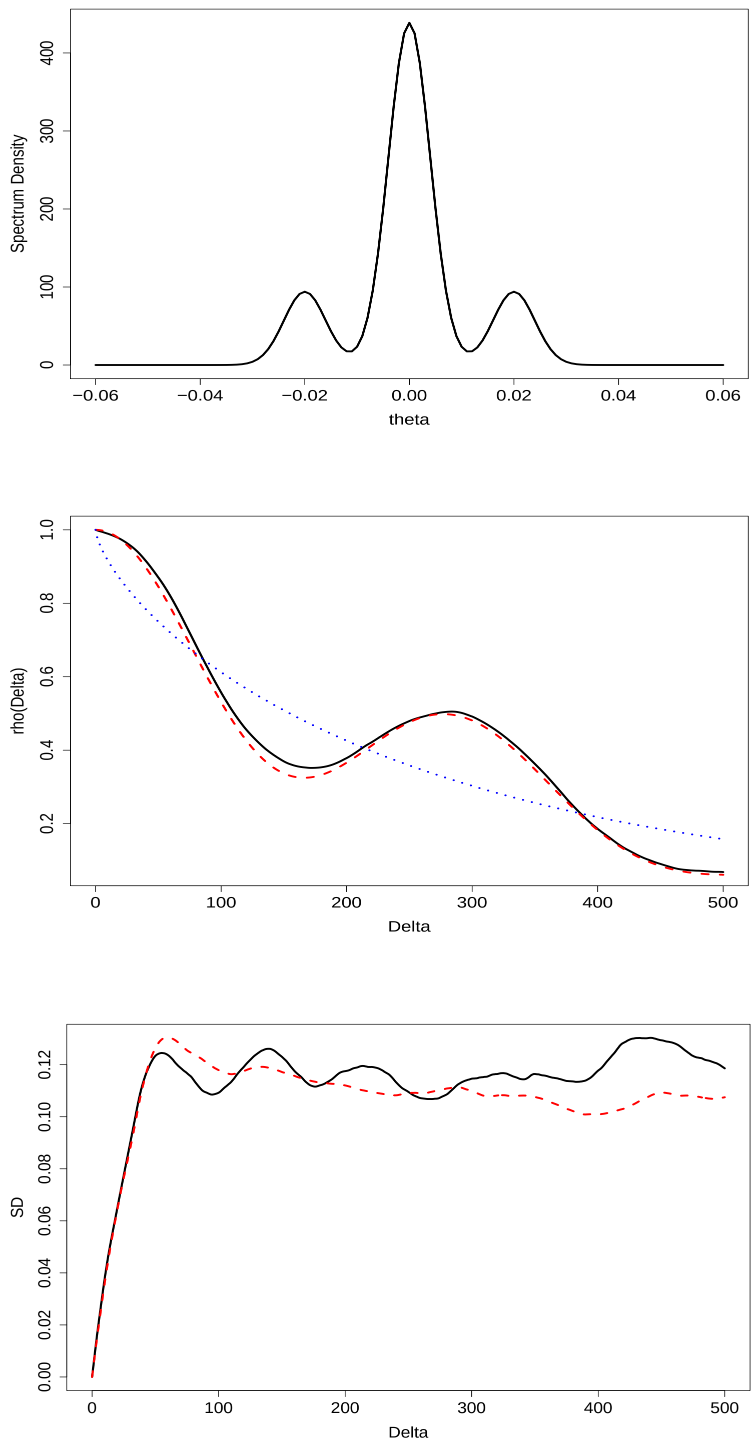


Figure 7: Simulation 2. Upper panel: the spectrum density of the correlation used in the simulation; middle panel: the dashed curve is the true correlation function, the solid curve is the mean of the kernel correlation estimator and the dotted curve is the best Matérn approximation to the true correlation; lower panel: the solid curve is the true pointwise standard deviation for the kernel correlation estimator, the dashed curve is the mean for the bootstrap standard deviation estimator.

Müller, and Wang (2005) proposed using smoothing methods to estimate within subject covariance for sparse longitudinal data, where the covariance is not necessary stationary.

In spatial statistics, many authors prefer intrinsic stationary to second-order stationary, for example Besag, York, and Mollie (1991) and Besag and Higdon (1999), because it is a slightly weaker assumption. In our case, each unit within a subject has further structure, we can define cross-variogram (Cressie 1993) instead of covariance function $\mathcal{V}(x_1, x_2, \Delta)$, and similar limiting distribution theorems can be proved as in Theorem II.1 and II.2. However, when it comes to Spatio-Temporal modelling, many authors, Cressie and Huang (1999) and Stein (2005), would come back to covariance because it is a more natural way to introduce the separable structure (2.3). In our data analysis, we provided some practical ideas to justify the separable structure in our data, where we compare the cross-validation scores with or without the separable assumption.

Lemma A.2 in the proofs implies that the denominator of estimator (2.4) gives the order of the asymptotic distribution of $\widehat{\mathcal{V}}$, $\widetilde{\mathcal{V}}$ and $\widehat{\rho}$. Based on this fact, we proposed a weighted Bootstrap method to estimate the standard deviation of the correlation estimator $\widehat{\rho}$. Our simulation shows that our correlation estimator and the bootstrap standard deviation estimator work well.

The analysis of the colon carcinogenesis p27 data suggests that the correlation of the crypts diminishes to 0 at about $\Delta = 500$ microns. The estimator and the standard deviation band also suggests the shape of the correlation function.

CHAPTER III

FUNCTIONAL LINEAR MODEL

3.1 Introduction

As a byproduct of modern science and technology, a lot of the data that are observed or collected in many fields nowadays are exceptionally high-dimensional. One such example is functional data, where each observation in a sample can be viewed as a function, as opposed to a scalar or a vector. In reality, for one reason or another if not simply human limitation, instead of observing such functions in their entirety, one only observes the values of the functions at a finite set of points. Nevertheless, the number of values observed per function may be quite large, sometimes much larger than the total number of functions, so that traditional multivariate analytical theory and methodology are not directly applicable. Indeed, the analysis of functional data has been steadily gaining attention among statisticians and practitioners, and there has been much progress on the methodology front in trying to understand how to deal with such data. The books Ramsay and Silverman (1997, 2002, 2005) and their website "http://ego.psych.mcgill.ca/misc/fda" contain a substantial amount of information in that regard. However, there has been much less progress on the theory front. This is not a surprise in view of the nature of the difficulties, as a theoretical result in this regard invariably involves the theory of functional analysis, multivariate analysis, optimization, and nonparametric function estimation.

One relatively simple problem, the linear regression, did receive a considerable attention theory-wise. Consider the model

$$Y_i = \mu + \int_a^b X_i(t) f(t) dt + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{3.1}$$

where the response Y_i, the fixed intercept μ and error ε_i are scalar, and the predictor X_i and regression weight function f are functions on $[a, b]$. For now, assume that we observe the X_i, Y_i and we are interested in the inference of μ, f and the variance of ε. In this regard, we mention the papers by Cai and Hall (2006), Cardot, Ferraty, and Sarda (1999, 2003), Cardot and Sarda (2005), Hall and Horowitz (2004) and Müller and Stadtmüller (2005). These papers contain highly sophisticated analysis and results which will be useful for many other situations in functional data analysis. However, all of the papers assumed that the functional predictors X_i are completely observed. This assumption is crucial for their approaches, but is seldom met in practice. To make matters worse, in reality there may be measurement error in observing X_i. The goal of the present paper is to address the linear regression problem under these practical situations.

This problem is ill-posed in the sense that a minute change in the data may lead to a huge changes in the resulting estimates, see Tykohonov and Arsenin (1977). One of the greatest challenges here (and elsewhere in functional data analysis) is how to interface the finite-dimensional space where the data reside and the infinite-dimensional space where the truth resides. Ramsay and Silverman (1997) proposed the following practical solution. First represent both the X_i and g, any candidate estimate of f, in terms of a set of pre-selected basis functions $\phi_1, \ldots, \phi_K$, say, so that

$$\tilde{X}_{i,\rho} = \sum_{k=1}^{K} b_{i,k}\phi_k \quad \text{and} \quad g = \sum_{k=1}^{K} c_k\phi_k,$$

where the "$\tilde{}$" in $\tilde{X}_{i,\rho}$ signifies the fact that this is a function that approximates the true function X_i based on the finitely observed values of X_i. Then estimate μ and f by the minimizer of the following penalized least squares criterion function

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\Big[Y_i - \nu - \int_0^1 \tilde{X}_{i,\rho} g\Big]^2 + \lambda\int_0^1 [g^{(m)}]^2 \\ = \; & n^{-1}\sum_{i=1}^{n}\Big[Y_i - \nu - \sum_{k=1}^{K}\sum_{\ell=1}^{K} b_{i,k}c_\ell\int_0^1 \phi_k\phi_\ell\Big]^2 + \lambda\, \sum_{k=1}^{K}\sum_{\ell=1}^{K} c_k c_\ell\int_0^1 \phi_k^{(m)}\phi_\ell^{(m)}, \end{aligned}$$

where λ is a smoothing parameter. They noted that the choice of the basis functions depends on the nature of the problem and the data. This is appealing since now we have a finite-dimensional optimization problem to cope with. Furthermore, if the basis functions are such that the matrix $\{\int_0^1 \phi_k^{(m)} \phi_\ell^{(m)}\}_{k,\ell=1}^K$ has a band structure, the computations will be even more straightforward; examples of such basis functions include Fourier, B-splines, and natural splines. It is also worth mentioning that variations of the basis-function approaches are adopted by other authors in studying the linear regression model; they include Cardot et al. (2003) who studied penalized B-splines, and James (2002) who considered a parametric approach.

The splines are general and flexible approximating functions which have a lot of desirable properties for this problem. However, theoretically they are more difficult to deal with, and we will address that problem in a forthcoming paper. In the present paper, we will focus on Fourier basis. The Fourier functions are a ideal basis if the data are smooth and exhibit periodicity; an example of that is the Canadian weather data in Ramsay and Silverman (1997). They are certainly the most convenient basis functions to work with in terms of of proving theory, since they are orthogonal, their m-th derivatives are orthogonal, and the orthogonality even carries over to the discretized basis vectors when the set of discrete points are equally-spaced. Our goal of this paper is to study the rate of convergence of the penalized least squares estimation using the Fourier basis. We will show that the rate of convergence is similar to that obtained in nonparametric regression function estimation.

The nature of the topic makes it necessary to employ some standard functional analysis terminology and results. They are quite basic and will not go beyond the first course in functional analysis. The reader is referred to Conway (1990) for details. This paper is structured as follows. Section 3.2 describes the assumptions and main results. All proofs and lemmas are collected in Section 3.3.

3.2 Main Results

Assume that the functional predictor X is a real-valued zero-mean, second-order stochastic process on $[0,1]$. Further, for some positive integer m, assume that with probability one X belongs to the periodic Soblev Space

$$\begin{aligned} W_{2,\text{per}}^m &= \{g \in L^2[0,1] : g \text{ is } m\text{-times differentiable where } g^{(m)} \in L^2[0,1] \text{ and } g^{(\nu)} \text{ is} \\ &\quad \text{absolutely continuous with } g^{(\nu)}(0) = g^{(\nu)}(1), 0 \le \nu \le m-1\}. \end{aligned}$$

It is well known that $W_{2,\text{per}}^m$ is dense in $L^2[0,1]$, therefore our methodology below based on this assumption applies to even situations for which this assumption is not met. However, relaxing the smoothness and boundary conditions do affect the convergence rate of our estimator. Denote by $R(s,t)$ the covariance function

$$R(s,t) = \mathrm{E}[X(s)X(t)],\ s,t \in [0,1],$$

and T the corresponding covariance operator

$$T : g \to \int_{s=0}^{1} R(s,\cdot)g(s)ds, g \in L^2[0,1],$$

For convenience, we will assume throughout without further mention that $\mathrm{E}\|X\|_{L^2}^4 < \infty$. This implies, among other things, that

$$\begin{aligned} \int_0^1 \int_0^1 R^2(s,t)dsdt &= \int_0^1 \int_0^1 \mathrm{E}^2[X(s)X(t)]dsdt \\ &\le \int_0^1 \int_0^1 \mathrm{E}[X^2(s)X^2(t)]dsdt = \mathrm{E}(\|X\|_{L^2}^4) < \infty, \end{aligned}$$

which shows that T is a Hilbert-Schmidt operator.

Let $X_i, 1 \le i \le n$, be n independent realizations of X. Below we consider the linear regression model (3.1) with $\mu = 0$. This simplification is minor for our results, but entails a considerable saving in term of notation. Let $t_j = (2j-1)/(2J), 1 \le j \le J$, be the locations where we observe the X_i; assume that the data that are observed are $Y_i, 1 \le i \le n$, and

$$\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,J})^T = (X_i(t_1) + \varsigma_{i,1}, \ldots, X_i(t_J) + \varsigma_{i,J})^T,\ 1 \le i \le n,$$

where $\varsigma_{i,j}$ is measurement error for $X_i(t_j)$. The $\varepsilon_i, \varsigma_i$ are assumed to be mutually uncorrelated, and independent of the $X_i(t_j)$, with mean zero and

$$\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 \quad \text{and} \quad \text{var}(\varsigma_{i,j}) = \sigma_\varsigma^2.$$

The Fourier basis functions that we use here are the complex Fourier functions $\phi_k(t) = e^{2\pi ikt}$, $k = 0, \pm 1, \pm 2, \cdots$. It is well known that each $g \in W_{2,\text{per}}^m$ can be uniquely represented as $g = \sum_{k=-\infty}^{\infty} c_k \phi_k$ in the L^2 sense with $\bar{c}_k = c_{-k}$. Let

$$K = [(J-1)/2] = \max\{k \le (J-1)/2; k \text{ is an integer}\},$$

and put

$$\phi(t) = (\phi_{-K}(t), \cdots, \phi_0(t), \cdots, \phi_K(t))^T,$$

$$\Phi = \{\phi_k(t_j)\}_{j=1,\cdots,J;k=-K,\cdots,K},$$

and

$$W = \{\langle \phi_i^{(m)}, \phi_j^{(m)} \rangle_{L^2}\}_{i,j=-K}^{K}.$$

Clearly,

$$W = \text{diag}\{(2\pi K)^{2m}, \cdots, (2\pi)^{2m}, 0, (2\pi)^{2m}, \ldots, (2\pi K)^{2m}\}, \tag{3.2}$$

and, since the t_j are equally spaced,

$$J^{-1}\bar{\Phi}^T\Phi = I.$$

First, for some smoothing parameter ρ, approximate each $\boldsymbol{Z}_i$ by $\tilde{X}_{i,\rho}$, the minimizer $u = \sum_{k=-K}^{K} b_k \phi_k$ of the penalized least squares criterion function

$$J^{-1} \sum_{j=1}^{J} [Z_{i,j} - u(t_j)]^2 + \rho \int_0^1 |u^{(m)}|^2,$$

namely,

$$\tilde{X}_{i,\rho}(t) = \phi^T(t) J^{-1} (I + \rho W)^{-1} \bar{\Phi}^T \boldsymbol{Z}_i = \phi^T(t) P_\rho \boldsymbol{Z}_i. \tag{3.3}$$

Since the $\boldsymbol{Z}_i$ are real, it is easily seen that the $\tilde{X}_{i,\rho}$ are real. Note that this smoother is an approximation to the periodic smoothing spline which uses infinitely many Fourier basis functions. In Theorem III.2 below, we will show that, under certain assumptions, the convergence rate of the smoother in (3.3) is comparable to that of periodic smoothing spline given by Rice and Rosenblatt (1981). See Eubank (1988), Section 6.3.1 for more details on periodic splines. This justifies the usage of roughly the same number of Fourier basis functions as the number of points. Using a finite number of basis functions is, of course, crucial for the computations that have to be performed in this problem.

Now, in addition to the smoothing parameter ρ that we used for obtaining the $\tilde{X}_{i,\rho}$ let λ be a second smoothing parameter, and $\hat{f}_{\lambda,\rho}$ be the minimizer $g \in W_{2,\text{per}}^m$ of the following criterion function:

$$n^{-1} \sum_{i=1}^{n} |Y_i - \langle \tilde{X}_{i,\rho}, g \rangle|^2 + \lambda \int_0^1 |g^{(m)}|^2, \tag{3.4}$$

where $\langle g, h \rangle$ is defined as $\int_0^1 g\bar{h}$ for complex-valued functions g, h.

Our main results below address the rate of convergence of $\hat{f}_{\lambda,\rho}$ as functions of ρ, λ, as well as the sample sizes n, J. Denote by $\mathcal{E}$ the space spanned by the eigenfunctions of the covariance operator T. It is clear that if f is not in $\mathcal{E}$ then it is not possible to estimate f consistently since the information that we have on f comes ultimately from $\langle X_i, f \rangle$. More generally, if the eigenspaces that correspond to small eigenvalues are estimated poorly due to a small or biased sample or having significant measurement error in the sample, then it is also unrealistic to expect good estimates of the projections of f on those subspaces. To "standardize" the estimation error of $\hat{f}_{\lambda,\rho}$ relative to the amount of information available, a reasonable measure of distance between $\hat{f}_{\lambda,\rho}$ and f is

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}), \tag{3.5}$$

where $\boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n)^T$, $\tilde{T}_\rho$ is the covariance operator of the sample covariance function

$$\tilde{R}_\rho(s,t) = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_{i,\rho}(s)\tilde{X}_{i,\rho}(t), \tag{3.6}$$

and

$$\langle g, h\rangle_{\tilde{T}_\rho} = \langle g, \tilde{T}_\rho h\rangle \quad \text{and} \quad \|g\|^2_{\tilde{T}_\rho} = \langle g, g\rangle_{\tilde{T}_\rho}.$$

This consideration is not new. For example, Cardot et al. (2003) considered $\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_T)$; also, since

$$\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} = n^{-1}\sum_{i=1}^{n} |\langle \tilde{X}_{i,\rho}, f\rangle - \langle \tilde{X}_{i,\rho}, \hat{f}_{\lambda,\rho}\rangle|^2,$$

the distance measure in (3.5) is similar to that in Cai and Hall (2004).

Now we state the main results, all proofs are given in Appendix B.

<u>Theorem</u> **III.1** *There exists some finite constant C that depends only on f such that*

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) \le C\left(\lambda + n^{-1}\lambda^{-1/(2m)}\nu^2_\rho + \frac{1}{n}\sum_{i=1}^{n} E(\|\tilde{X}_{i,\rho} - X_i\|^2_{L^2}|\boldsymbol{Z}_i)\right) \tag{3.7}$$

for all n, λ, and ρ, where ν_ρ is the largest eigenvalue of $\tilde{T}_\rho$.

Note that the first and second terms on the right of (3.7) describe the square bias and variance, respectively, of the procedure; the third term there essentially reflects the error of approximating X_i by $\tilde{X}_{i,\rho}$.

To obtain a concrete rate of convergence for $\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\boldsymbol{Z})$, the following Theorem III.2 is crucial.

<u>THEOREM</u> **III.2** *Assume that* $X \in W_{2,\mathrm{per}}^m$ *a.s. and* $\mathrm{E}\|X^{(m)}\|_{L^2}^2 < \infty$*. Then for all* $J \to \infty$, $\rho \to 0$, *with* $J^{2m}\rho \to \infty$, *there exists a finite constant* C *which doesn't depend on* J *or* ρ *such that*

$$\mathrm{E}(\|\tilde{X}_{1,\rho} - X_1\|_{L^2}^2) \leq C(\rho + J^{-1}\rho^{-1/(2m)}). \tag{3.8}$$

If, in addition, $X \in W_{2,\mathrm{per}}^{2m}$ *a.s. and* $\mathrm{E}\|X^{(2m)}\|_{L^2}^2 < \infty$*, then for* J, ρ *as above, there exists a finite constant* C *which doesn't depend on* J *or* ρ *such that*

$$\mathrm{E}(\|\tilde{X}_{1,\rho} - X_1\|_{L^2}^2) \leq C(\rho^2 + J^{-1}\rho^{-1/(2m)}). \tag{3.9}$$

Theorem III.2 is similar in spirit to Theorem 2 of Rice and Rosenblatt (1981), which studies the rate of convergence of the periodic smoothing spline estimator in nonparametric regression. As we mentioned before, even though $\tilde{X}_{1,\rho}$ is estimated with a finite number of Fourier basis functions, the rate of convergence is comparable to that of the periodic smoothing spline estimator using an infinite number of basis functions. The result (3.9) shows that with the extra conditions $X_1 \in W_{2,\mathrm{per}}^{2m}$ a.s. and $\mathrm{E}(\|X_1^{(m)}\|_{L^2}^2) < \infty$ in place but not specifically taken into account in the estimation procedure, the rate of convergence will nevertheless improve. This also parallels Rice and Rosenblatt's treatment of the periodic smoothing spline.

For the case where $\sigma_\varsigma = 0$, i.e. the $X_i(t_j)$ are observed without measurement error, we have

$$\mathrm{E}(\|\tilde{X}_{1,0} - X_1\|_{L^2}^2) \leq CJ^{-(2m-1)},$$

under $\mathrm{E}\|X_1^{(m)}\|_{L^2}^2 < \infty$. The proof of this result follows in a straightforward manner from the derivations in the proof of Theorem III.2 and is omitted.

The term $n^{-1}\sum_{i=1}^n \mathrm{E}(\|\tilde{X}_{i,\rho} - X_i\|_{L^2}^2|\boldsymbol{Z}_i)$ in (3.7) clearly converges in L^1 to 0 at the rates described by (3.8) and (3.9) under the respective settings there. Further, since T is bounded, it is natural to expect that $\tilde{T}_\rho$ is also bounded so that $\nu_\rho = O_p(1)$ under appropriate conditions. Thus, we state the following result.

<u>THEOREM</u> **III.3** *Suppose that* $X \in W^m_{2,\mathrm{per}}$ *a.s., and* $\mathrm{E}(\|X^{(m)}\|^2_{L^2}) < \infty$. *Then*

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) = O_p(\lambda + \rho + n^{-1}\lambda^{-1/(2m)} + J^{-1}\rho^{-1/(2m)}), \tag{3.10}$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\lambda \to \infty$. *If, in addition,* $X \in W^{2m}_{2,\mathrm{per}}$ *a.s., and* $\mathrm{E}(\|X^{(2m)}\|^2_{L^2}) < \infty$, *then we have*

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) = O_p(\lambda + \rho^2 + n^{-1}\lambda^{-1/(2m)} + J^{-1}\rho^{-1/(2m)}) \tag{3.11}$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\lambda \to \infty$.

It follows from (3.10) that the optimal rate of convergence of $\hat{f}_{\lambda,\rho}$ in $\tilde{T}_\rho$-norm is

$$n^{-2m/(2m+1)} + J^{-2m/(2m+1)}$$

under the general assumptions of Theorem III.3; the rate can be improved to

$$n^{-2m/(2m+1)} + J^{-4m/(4m+1)}$$

under the additional assumptions $X \in W^{2m}_{2,\mathrm{per}}$ a.s. and $\mathrm{E}(\|X^{(2m)}\|^2_{L^2}) < \infty$, as described by (3.11).

In the following, we consider rates of convergence if $\tilde{T}_\rho$ is replaced by T. To do that we need to quantify the distance between $\tilde{T}_\rho$ and T, for which the Hilbert-Schmidt norm seems ideal. For any self-adjoint operator A on $L^2[0,1]$, let $\|A\|_{\mathcal{H}}$ be the Hilbert-Schmidt norm of the operator.

<u>THEOREM</u> **III.4** *Suppose that, for some* $m \geq 2$, $X \in W^m_{2,\mathrm{per}}$ *a.s.,* $\sup_t \mathrm{E}\{[X^{(m)}(t)]^2\} < \infty$, *and* $\mathrm{E}(\|X^{(m)}\|^4_{L^2}) < \infty$. *Also assume that* $\mathrm{E}(\varsigma^4) < \infty$. *Then*

$$\mathrm{E}(\|\tilde{T}_\rho - T\|^2_{\mathcal{H}}) = O(n^{-1} + \rho + J^{-2}\rho^{-1/(2m)})$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\lambda \to \infty$. *If, in addition,* $X \in W_{2,\text{per}}^{2m}$ *a.s.,* $\sup_t \mathrm{E}\{[X^{(2m)}(t)]^2\} < \infty$, *and* $\mathrm{E}(\|X^{(2m)}\|_{L^2}^4) < \infty$, *then*

$$\mathrm{E}(\|\tilde{T}_\rho - T\|_{\mathcal{H}}^2) = O(n^{-1} + \rho^2 + J^{-2}\rho^{-1/(2m)})$$

for $\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty$, *and* $J^{2m}\lambda \to \infty$.

Note that Theorem III.4 should be compared with the results in Dauxois, Pousse, and Romain (1982) which were proved under the assumption that the X_i are completely and precisely observed.

The following result gives the rates of convergence of $\hat{f}_{\lambda,\rho}$ in T-norm.

<u>Theorem</u> **III.5** *Suppose that for some* $m \geq 2$, $X \in W_{2,\text{per}}^{m}$ *a.s.,* $\sup_t \mathrm{E}\{[X^{(m)}(t)]^2\} < \infty$, *and* $\mathrm{E}(\|X^{(m)}\|_{L^2}^4) < \infty$. *Also assume that* $\mathrm{E}(\varsigma^4) < \infty$. *Then*

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_T^2 | \boldsymbol{Z}) = O_p(n^{-1/2} + \lambda + n^{-1}\lambda^{-1/(2m)} + \rho^{1/2} + J^{-1}\rho^{-1/2m})$$

for n, J, λ, ρ *with*

$$\lambda \to 0, \rho \to 0, n^{2m}\lambda \to \infty, J^{2m}\lambda \to \infty, \text{ and } (\rho + J^{-1}\rho^{-1/(2m)})/\lambda = O(1), \quad (3.12)$$

If, in addition, $X \in W_{2,\text{per}}^{2m}$ *a.s.,* $\sup_t \mathrm{E}\{[X^{(2m)}(t)]^2\} < \infty$, *and* $\mathrm{E}(\|X^{(2m)}\|_{L^2}^4) < \infty$, *then*

$$\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_T^2 | \boldsymbol{Z}) = O_p(n^{-1/2} + \lambda + n^{-1}\lambda^{-1/(2m)} + \rho + J^{-1}\rho^{-1/2m})$$

for n, J, λ, ρ *satisfying (3.12).*

CHAPTER IV

CONCLUSIONS, EXTENSIONS AND FUTURE WORK

4.1 Correlation Estimation for Functional Data

In Chapter II, we proposed a very general framework for the colon carcinogenesis data, that is, each crypt is a function and we want to estimate the spatial correlation between functions. However, to make the mathematics in the derivation of the asymptotic theory tractable, we did some dimension reduction in the crypts. In other words, we reduced the functions to vectors.

As we argued in Section 1.1, the same entry in different vectors do not have the measurements at the same cell location. Therefore, recovering each unit into a curve and do estimation in the functional way may bring more accuracy. Therefore, in the next step, we will consider the case that the number of cells per crypt m also goes to infinity, and we will do smoothing within each crypt.

Another possible extension is apply local linear regression in the Δ direction. This extension will not incur stronger assumption on the underlying random field θ, but will increase the efficiency a lot. Applying higher order local polynomial regression to the Δ direction is possible, but it will generally require higher order partial derivatives of the covariance function $\mathcal{V}(x_1, x_2, \Delta)$, which may not be an appropriate assumption.

In terms of methodology, we also need to develop a spatial adaptive bandwidth selection procedure. When estimating the covariance function, we generally have less observation at larger distance lags. Therefore, in principle, we should use a bigger bandwidth at a larger value of Δ.

4.2 Spline Methods for Functional Linear Model

In Chapter III, we explore the convergence rate for the functional linear model using a two-stage roughness penalty approach, where we smooth each curve by a periodic smoothing spline in the first step and estimate the unknown coefficient function f by minimizing a penalized least square in the second step. We apply the methodology to the lipoprotein profile data that we introduced in Chapter I, and use the profile curves to predict the total cholesterol level in the patients. The first plot in Figure 8 shows the estimated coefficient function $\widehat{f}$ using this method.

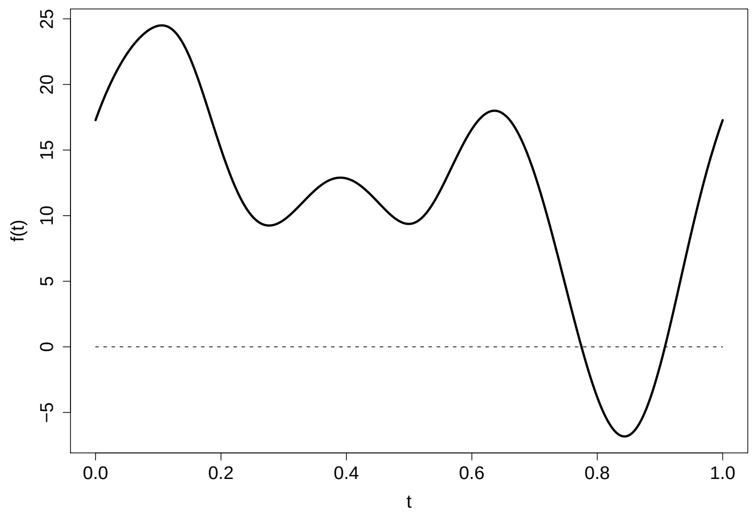
In general, periodic splines require the period boundary condition given in Chapter III, which is too restrictive for statistical practice. The penalized spline methods given in Section 3 in Chapter I are more flexible and more popular in data analyses.

The method that we are considering for the future work is the following. We first choose a set of spline basis function $\boldsymbol{B}(t)$, and smooth each curve by the P-Spline given by (1.1). And then minimize a penalized least square criterion similar to (3.4),

$$n^{-1}\sum_{i=1}^{n}\{Y_i - \mu - \langle \tilde{X}_{i,\rho}, g\rangle\}^2 + \lambda J(g),$$

where we restrict $g = \beta^{\mathrm{T}}\boldsymbol{B}$ to be in the functional subspace spanned by $\boldsymbol{B}(\cdot)$, $J(g) = \beta^{\mathrm{T}}D\beta$ is the roughness penalty and D is a positive semi-definite matrix. All smoothing parameters can be selected by generalized cross-validation (GCV).

We apply the spline method to the lipoprotein profile data, and the result for $\widehat{f}$ is shown in the second plot of Figure 8. As we can see that the methods based fourier basis and spline basis give almost the same results. At the first sight, the result looks strange, since the estimated coefficient function sees to down weight the region corresponding to HDL. In Figure 9, we show the P-Spline fit for the first 5 profile curves. As we can see that everybody seems to have a similar HDL component, but there are much more variations



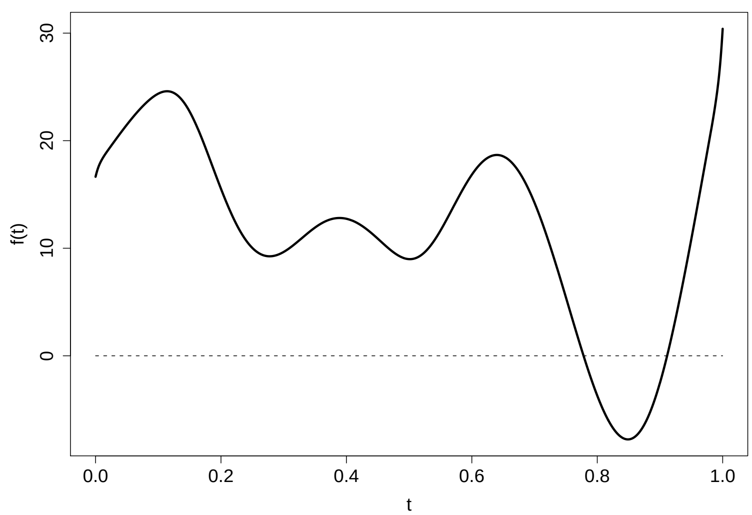


Figure 8: Functional linear regression applied to the lipoprotein profile data. The first plot is the estimated coefficient function $\widehat{f}$ using periodic spline method; the second plot is $\widehat{f}$ by the P-spline method.

in the VLDL and LDL components. So, the right way to interpret $\widehat{f}$ is that the VLDL and LDL parts have some positive effects on the total cholesterol level from the intercept μ.

One of my future research goal is to develop some asymptotic theory for the P-Spline method. However, it is going to be a very difficult problem, since a general asymptotic theory for smoothing one curve via P-Spline is still missing.

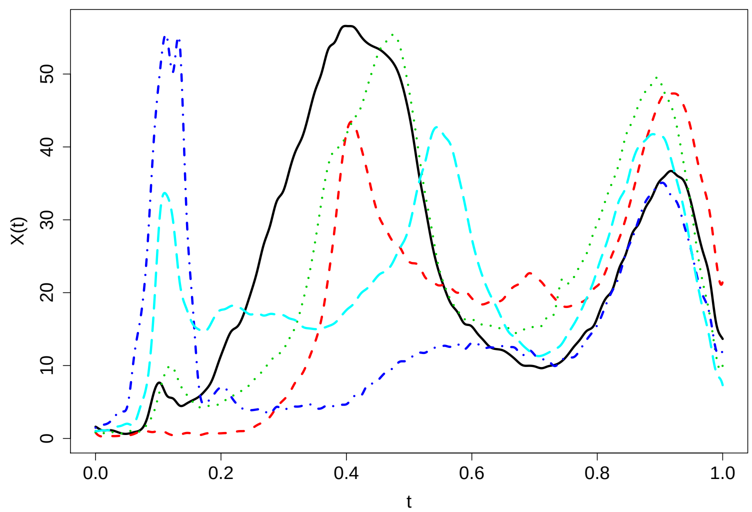


Figure 9: P-Spline fit for the first 5 lipoprotein profiles.

REFERENCES

Ash, R. and Gardner, M. (1975). *Topics in Stochastic Processes*. New York, NY: Academic Press.

Besag, J. and Higdon, D. (1999). Baysian analysis of agricultural field experiments. *Journal of the Royal Statistical Society, Series B,* **61**, 691–746.

Besag, J., York, J., and Mollie, A. (1991). Baysian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics* **43**, to appear.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*, **45**, 11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica,* **13**, 571–591.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis,* **92**, 24–41.

Conway, J. B. (1990). *A course in Functional Analysis*. New York, NY: Springer-Verlag.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik,* **31**, 377–403.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York, NY: Wiley.

Cressie, N. A. C. and Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association,* **94**, 1330–1340.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis,* **12**, 136–154.

Diggle, P. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics,* **54**, 401–415.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penaltyies (with discussion). *Statistical Science,* **11**, 89–121.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. New York, NY: Marcel Dekker.

Guan, Y., Sherman, M., and Calvin, J. A. (2004). A nonparametric test for spatial isotropy using subsampling. *Journal of the American Statistical Association,* **99**, 810–821.

Hall, P., Fisher, N. I., and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics* **22**, 2115–2134.

Hall, P. and Horowitz, J. (2004). Methodology and convergence rates for functional linear regression. *Unpublished manuscript*, http://www.faculty.econ.northwestern.edu/faculty/horowitz/papers/hhor–final.pdf.

James, G. (2002). Generalized linear models with functional predictor variables. *Journal of the Royal Statistical Society, Series B,* **64**, 411–432.

Merikoski, J. and Kumar, R. (2004). Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes,* **4**, 150–159.

Morris, J. S., Wang, N., R., L. J., Chapkin, R. S., Turner, N. D., Hong, M. Y., and Carroll, R. J. (2001). Parametric and nonparametric methods for understanding the relationship between carcinogen-induced dna adduct levels in distal and proximal regions of the colon. *Journal of the American Statistical Association*, **96**, 816–826.

Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, **33**, 774–805.

Politis, D. N. and Sherman, M. (2001). Moment estimation for statistics from marked point processes. *Journal of the Royal Statistical Society, Series B*, **63**, 261–275.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York, NY: Springer-Verlag.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York, NY: Springer-Verlag.

Rice, J. and Rosenblatt, M. (1981). Integrated mean square error of a smoothing spline. *Journal of Approximation Theory*, **33**, 353–369.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. New York, NY: Cambridge University Press.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York, NY: Wiley.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York, NY: Springer-Verlag.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY: Springer-Verlag.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, **100**, 310–321.

Tykohonov, A. and Arsenin, V. (1977). *Solution to Ill-Posed Problems*. New York, NY: Wiley.

Yao, F., Müller, H. G., and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.

APPENDIX A

PROOFS FOR CHAPTER II

The proofs are organized in the following way: in the first section, we provide lemmas regarding asymptotic properties of the covariance estimators when there is only one subject; in the second section, we provide lemmas on the estimators with multiple subjects, and the proofs of Theorems II.1, II.2 and Corollary II.1 are given in the end.

Estimation Within One Subject

We will first discuss a case that there is only one subject and the number of units goes to infinity. Let $N(\cdot)$ be the inhomogeneous Poisson process on $[0, L]$ with local intensity $\nu g^*(s)$. As in Karr (1986), denote $N_2(ds_1, ds_2) = N(ds_1)N(ds_2)I(s_1 \neq s_2)$. Let $\Theta(s, \cdot)$ denote the unit-level mean at unit location s, and $\Theta(\cdot)$ denote the subject-level mean. Define

$$\begin{aligned} a(\Delta) &= L^{-1}\sum_i \sum_{k\neq i} K_h(\Delta - \Delta_{ik}) = L^{-1}\int_0^L \int_0^L K_h\{\Delta - (s_1 - s_2)\}N_2(ds_1, ds_2); \\ b(x_1, x_2, \Delta) &= L^{-1}\sum_i \sum_{k\neq i} K_h(\Delta - \Delta_{ik})\{Y(S_i, x_1) - \Theta(x_1)\}\{Y(S_k, x_2) - \Theta(x_2)\} \\ &= L^{-1}\int_0^L \int_0^L K_h\{\Delta - (s_1 - s_2)\}\{Y(s_1, x_1) - \Theta(x_1)\} \\ &\qquad \times \{Y(s_2, x_2) - \Theta(x_2)\}N_2(ds_1, ds_2). \end{aligned}$$

<u>Lemma</u> **A.1** Let that X_1 and X_2 be real valued random variables measurable with respect to $\mathcal{F}\{[0, t]\}$ and $\mathcal{F}\{[t + \tau, \infty)\}$ respectively, such that $|X_i| < C_i$, $i = 1, 2$. Then $|\text{cov}(X_1, X_2)| \leq 4C_1C_2\alpha(\tau)$. If X_1 and X_2 are complex random variables, this inequality holds with the constant 4 replaced by 16.

<u>Proof</u>: The proof is analogous to that of Theorem 17.2.1 in Ibragimov and Linnik (1971).

Denote $T_1 = [0,t]$, $T_2 = [t+\tau, \infty)$, then we have

$$\begin{aligned} & |E(X_1X_2) - E(X_1)E(X_2)| = |E[E\{X_1X_2|\mathcal{F}(T_1)\}] - E(X_1)E(X_2)| \\ = & |E(X_1[E\{X_2|\mathcal{F}(T_1)\} - E(X_2)])| \leq C_1 E|E\{X_2|\mathcal{F}(T_1)\} - E(X_2)| \\ = & C_1 E(u_1[E\{X_2|\mathcal{F}(T_1)\} - E(X_2)]) \end{aligned}$$

where $u_1 = \text{sign}[E\{X_2|\mathcal{F}(T_1)\} - E(X_2)]$. It is easy to see that u_1 is measurable with respect to $\mathcal{F}(T_1)$, therefore $|E(X_1X_2) - E(X_1)E(X_2)| \leq C_1|E(u_1X_2) - E(u_1)E(X_2)|$. By the same argument, we have $|E(u_1X_2) - E(u_1)E(X_2)| \leq C_2|E(u_1u_2) - E(u_1)E(u_2)|$, where $u_2 = \text{sign}[E\{u_1|\mathcal{F}(T_2)\} - E(u_1)]$. Now, we have $|E(X_1X_2) - E(X_1)E(X_2)| \leq C_1C_2|E(u_1u_2) - E(u_1)E(u_2)|$. Define the events $A_1 = \{u_1 = 1\} \in \mathcal{F}(T_1)$, $\overline{A}_1 = \{u_1 = -1\} \in \mathcal{F}(T_1)$, $A_2 = \{u_2 = 1\} \in \mathcal{F}(T_2)$ and $\overline{A}_2 = \{u_2 = -1\} \in \mathcal{F}(T_2)$. Then,

$$\begin{aligned} |E(u_1u_2) - E(u_1)E(u_2)| &= |P(A_1A_2) - P(A_1\overline{A}_2) - P(\overline{A}_1A_2) + P(\overline{A}_1\overline{A}_2) \\ &\quad -P(A_1)P(A_2) + P(A_1)P(\overline{A}_2) + P(\overline{A}_1)P(A_2) - P(\overline{A}_1)P(\overline{A}_2)| \\ &\leq |P(A_1A_2) - P(A_1)P(A_2)| + |P(A_1\overline{A}_2) - P(A_1)P(\overline{A}_2)| \\ &\quad +|P(\overline{A}_1A_2) - P(\overline{A}_1)P(A_2)| + |P(\overline{A}_1\overline{A}_2) - P(\overline{A}_1)P(\overline{A}_2)| \\ &\leq 4\alpha(\tau). \end{aligned}$$

Thus, the proof is completed for the real random variable case. If X_1 and X_2 are complex, we can apply the same arguments to the real and imaginary parts separately.

<u>LEMMA</u> **A.2** With the assumptions stated in Section 2.3, for any fixed Δ, we have $a(\Delta) \to \nu^2 f_1(0)$ in L^2 sense, as $L \to \infty$.

<u>Proof</u>: Recall that by definition of $f_1(\cdot)$, if X_1 and X_2 are independent and identically distributed with density $g(\cdot)$, then $f_1(u) = \int g(t+u)g(t)dt$ is the density of $X_1 - X_2$. Thus, for fixed Δ,

$$E\{a(\Delta)\} = \nu^2 L^{-1} \int\int_{s_1 \neq s_2} K_h\{\Delta - (s_1 - s_2)\} g(s_1/L) g(s_2/L) ds_1 ds_2$$

$$\begin{aligned}
&= \nu^2 L \int_0^1 \int_0^1 K_h\{\Delta - L(t_1 - t_2)\} g(t_1) g(t_2) dt_1 dt_2 \\
&= \nu^2 L \int \int K_h(\Delta - Lu) g(t_2 + u) g(t_2) du dt_2 \\
&= \nu^2 L \int K_h(\Delta - Lu) f_1(u) du = \nu^2 \int K(v) f_1\{(\Delta - hv)/L\} dv \\
&= \nu^2 \int K(v)\{f_1(0) + O(L^{-1})\} dv = \nu^2 f_1(0) + O(L^{-1}).
\end{aligned}$$

Next,

$$\begin{aligned}
E\{a^2(\Delta)\} &= L^{-2} \int_0^L \int_0^L \int_0^L \int_0^L K_h\{\Delta - (s_1 - s_2)\} K_h\{\Delta - (s_3 - s_4)\} \\
&\qquad \times E\{N_2(ds_1, ds_2) N_2(ds_3, ds_4)\}.
\end{aligned}$$

Calculations as in Guan et al. (2004) show that

$$\begin{aligned}
&E\{N_2(ds_1, ds_2) N_2(ds_3, ds_4)\} = \nu^4 g^*(s_1) g^*(s_2) g^*(s_3) g^*(s_4) ds_1 ds_2 ds_3 ds_4 \\
&\quad + \nu^3 g^*(s_1) g^*(s_2) g^*(s_4) \epsilon_{s_1}(ds_3) ds_1 ds_2 ds_4 + \nu^3 g^*(s_1) g^*(s_2) g^*(s_3) \epsilon_{s_1}(ds_4) ds_1 ds_2 ds_3 \\
&\quad + \nu^3 g^*(s_1) g^*(s_2) g^*(s_4) \epsilon_{s_2}(ds_3) ds_1 ds_2 ds_4 + \nu^3 g^*(s_1) g^*(s_2) g^*(s_3) \epsilon_{s_2}(ds_4) ds_1 ds_2 ds_3 \\
&\quad + \nu^2 g^*(s_1) g^*(s_2) \epsilon_{s_1}(ds_3) \epsilon_{s_2}(ds_4) ds_1 ds_2 + \nu^2 g^*(s_1) g^*(s_2) \epsilon_{s_1}(ds_4) \epsilon_{s_2}(ds_3) ds_1 ds_2,
\end{aligned}$$

where $\epsilon_x(\cdot)$ is a point measure defined in Karr (1986), such that $\epsilon_x(dy) = 1$ if $x \in dy$, 0 otherwise. Here dy is defined to be a small disc centered at y. There are 7 terms in the expression above, so the expression for $E\{a^2(\Delta)\}$ can be decomposed into 7 integrals: denote them as A_{11}-A_{17}. Similar to the calculations of $E\{a(\Delta)\}$, we have

$$\begin{aligned}
A_{11} &= \nu^4 L^{-2} \int \int_{s_1 \neq s_2} \int \int_{s_3 \neq s_4} K_h\{\Delta - (s_1 - s_2)\} K_h\{\Delta - (s_3 - s_4)\} \\
&\qquad \times g(s_1/L) g(s_2/L) g(s_3/L) g(s_4/L) ds_1 ds_2 ds_3 ds_4 \\
&= \nu^4 f_1^2(0) + o(1).
\end{aligned}$$

$$A_{12} = \nu^3 L^{-2} \int_{s_1 \neq s_2, s_4} K_h\{\Delta - (s_1 - s_2)\} K_h\{\Delta - (s_1 - s_4)\}$$

$$
\begin{aligned}
& \times g(s_1/L)g(s_2/L)g(s_4/L)ds_1 ds_2 ds_4 \\
&= \nu^3 L \int\int K_h(\Delta - Lu_1)K_h\{\Delta - L(u_1-u_2)\}f_2(u_1,u_2)du_1 du_2 \\
& \qquad \text{(by definition of } f_2) \\
&= \nu^3 L^{-1} \int\int K(v_1)K(v_2)f_2\{(\Delta - hv_1)/L, (v_2-v_1)h/L\}dv_1 dv_2 \\
&= \nu^3 L^{-1} f_2(0,0) + O(L^{-2}).
\end{aligned}
$$

Similarly, $A_{13} - A_{15}$ are of order $O(L^{-1})$. Next,

$$
\begin{aligned}
A_{16} &= \nu^2 L^{-2} \int_{s_1 \neq s_2} K_h^2\{\Delta - (s_1 - s_2)\}g(s_1/L)g(s_2/L)ds_1 ds_2 \\
&= \nu^2 \int K_h^2(\Delta - Lu)f_1(u)du \\
&= \nu^2 L^{-1}h^{-1} \int K^2(v)f_1\{(\Delta - hv)/L\}dv \\
&= \nu^2 L^{-1}h^{-1} f_1(0)R_K + o(Lh^{-1}).
\end{aligned}
$$

Similarly, we can show that A_{17} is of the same order as A_{16}. This means that A_{11} is the leading term in $E\{a^2(\Delta)\}$. Hence, $E\{a(\Delta) - \nu^2 f_1(0)\}^2 \to 0$, completing the proof.

LEMMA **A.3** For any fixed Δ, define $\beta(x_1, x_2, \Delta) = b(x_1, x_2, \Delta) - a(\Delta)\mathcal{V}(x_1, x_2, \Delta)$. Then

$$
\begin{aligned}
& E\{\beta(x_1,x_2,\Delta)\} = \nu^2 f_1(0)\{\mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta)\sigma_K^2 h^2/2 + o(h^2)\}, \\
& \text{cov}\{\beta(x_1,x_2,\Delta), \beta(x_3,x_4,\Delta')\} \\
& \quad = \nu^2 L^{-1}h^{-1}R_K f_1(0)[I(\Delta = \Delta')\{\mathcal{M}(x_1,x_2,x_3,x_4,\Delta,\Delta,0) \\
& \quad + I(x_2 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_1,x_3,0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2,x_4,0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\} \\
& \quad + I(\Delta = -\Delta')\{\mathcal{M}(x_1,x_2,x_3,x_4,\Delta,-\Delta,-\Delta) + I(x_2 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_1,x_4,0) \\
& \quad + I(x_1 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_2,x_3,0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}] + o(L^{-1}h^{-1}),
\end{aligned}
$$

where $\mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta) = \partial^2 \mathcal{V}(x_1,x_2,\Delta)/\partial\Delta^2$.

Proof: Rewrite

$$\beta(x_1,x_2,\Delta) = L^{-1}\int\int K_h\{\Delta-(s_1-s_2)\}[\{Y(s_1,x_1)-\Theta(x_1)\} \times\{Y(s_2,x_2)-\Theta(x_2)\}-\mathcal{V}(x_1,x_2,\Delta)]N_2(ds_1,ds_2),$$

it follows that

$$\begin{aligned}E\{\beta(x_1,x_2,\Delta)\} &= \nu^2L^{-1}\int\int_{s_1\neq s_2}K_h\{\Delta-(s_1-s_2)\} \\ &\quad\times\{\mathcal{V}(x_1,x_2,s_1-s_2)-\mathcal{V}(x_1,x_2,\Delta)\}g(s_1/L)g(s_2/L)ds_1ds_2 \\ &= \nu^2L\int K_h(\Delta-Lu)\{\mathcal{V}(x_1,x_2,Lu)-\mathcal{V}(x_1,x_2,\Delta)\}f_1(u)du \\ &= \nu^2\int K(v)\{-\mathcal{V}^{(0,0,1)}(x_1,x_2,\Delta)hv+\mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta)h^2v^2/2+o(h^2)\} \\ &\quad\times\{f_1(0)+f_1'(0)(\Delta-hv)/L+o(L^{-1})\}dv \\ &= \nu^2\{f_1(0)\mathcal{V}^{(0,0,2)}(x_1,x_2,\Delta)\sigma_K^2h^2/2+o(h^2)\}.\end{aligned}$$

In addition,

$$\begin{aligned}&\text{cov}\{\beta(x_1,x_2,\Delta),\beta(x_3,x_4,\Delta')\} \\ &= L^{-2}\int\int\int\int K_h\{\Delta-(s_1-s_2)\}K_h\{\Delta'-(s_3-s_4)\} \\ &\quad[\mathcal{V}(x_1,x_2,\Delta)\mathcal{V}(x_3,x_4,\Delta')-\mathcal{V}(x_1,x_2,s_1-s_2)\mathcal{V}(x_3,x_4,\Delta') \\ &\quad-\mathcal{V}(x_1,x_2,\Delta)\mathcal{V}(x_3,x_4,s_3-s_4)+\mathcal{V}(x_1,x_2,s_1-s_2)\mathcal{V}(x_3,x_4,s_3-s_4) \\ &\quad+\mathcal{M}\{x_1,x_2,x_3,x_4,(s_1-s_2),(s_3-s_4),(s_2-s_4)\} \\ &\quad+I(s_1=s_3)I(s_2\neq s_4)I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}\{x_2,x_4,(s_2-s_4)\} \\ &\quad+I(s_1=s_4)I(s_2\neq s_3)I(x_1=x_4)\sigma_\epsilon^2\mathcal{V}\{x_2,x_3,(s_2-s_3)\} \\ &\quad+I(s_2=s_3)I(s_1\neq s_4)I(x_2=x_3)\sigma_\epsilon^2\mathcal{V}\{x_1,x_4,(s_1-s_4)\} \\ &\quad+I(s_2=s_4)I(s_1\neq s_3)I(x_2=x_4)\sigma_\epsilon^2\mathcal{V}\{x_1,x_3,(s_1-s_3)\} \\ &\quad+I(s_1=s_3,s_2=s_4)\{I(x_2=x_4)\sigma_\epsilon^2\mathcal{V}(x_1,x_3,0)+I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}(x_2,x_4,0) \\ &\quad+I(x_1=x_3,x_2=x_4)\sigma_\epsilon^4\}\end{aligned}$$

$$+I(s_1 = s_4, s_2 = s_3)\{I(x_2 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_2, x_3, 0)$$
$$+I(x_1 = x_4, x_2 = x_3)\sigma_\epsilon^4\}]E\{N_2(ds_1, ds_2)N_2(ds_3, ds_4)\}$$
$$-\nu^4 L^{-2} \int\int\int\int K_h\{\Delta - (s_1 - s_2)\}K_h\{\Delta' - (s_3 - s_4)\}$$
$$\times\{\mathcal{V}(x_1, x_2, s_1 - s_2) - \mathcal{V}(x_1, x_2, \Delta)\}\{\mathcal{V}(x_3, x_4, s_3 - s_4) - \mathcal{V}(x_3, x_4, \Delta')\}$$
$$\times g(s_1/L)g(s_2/L)g(s_3/L)g(s_4/L)ds_1 ds_2 ds_3 ds_4.$$

As in Lemma A.2, according to the expression for $E\{N_2(ds_1, ds_2)N_2(ds_3, ds_4)\}$, we can summarize this covariance expression as the sum of 7 terms, denoted as A_{21}-A_{27}.

$$\begin{aligned}
A_{21} &= \nu^4 L^{-2} \int_0^L \int_0^L \int_0^L \int_0^L K_h\{\Delta - (s_1 - s_2)\}K_h\{\Delta' - (s_3 - s_4)\} \\
&\qquad \times\mathcal{M}\{x_1, x_2, x_1, x_2, (s_1 - s_2), (s_3 - s_4), (s_2 - s_4)\} \\
&\qquad \times g(s_1/L)g(s_2/L)g(s_3/L)g(s_4/L)ds_1 ds_2 ds_3 ds_4 \\
&= \nu^4 L^2 \int_0^1 \int_0^1 \int_0^1 \int_0^1 K_h\{\Delta - L(t_1 - t_2)\}K_h\{\Delta' - L(t_3 - t_4)\} \\
&\qquad \times\mathcal{M}\{x_1, x_2, x_1, x_2, L(t_1 - t_2), L(t_3 - t_4), L(t_2 - t_4)\} \\
&\qquad \times g(t_1)g(t_2)g(t_3)g(t_4)dt_1 dt_2 dt_3 dt_4 \\
&= \nu^4 L^2 \int\int\int K_h(\Delta - Lu_1)K_h(\Delta' - Lu_2)\mathcal{M}(x_1, x_2, x_1, x_2, Lu_1, Lu_2, Lu_3) \\
&\qquad f_3(u_1, u_2, u_3)du_1 du_2 du_3 \\
&= \nu^4 L^{-1} \int\int\int K(v_1)K(v_2)\mathcal{M}(x_1, x_2, x_1, x_2, \Delta - hv_1, \Delta' - hv_2, v_3) \\
&\qquad f_3\{(\Delta - hv_1)/L, (\Delta' - hv_2)/L, v_3/L\}dv_1 dv_2 dv_3 \\
&\leq \nu^4 L^{-1} C \int \mathcal{M}(x_1, x_2, x_1, x_2, \Delta, \Delta', v)dv + o(L^{-1}),
\end{aligned}$$

where C is the upper bound for the density function $f_3(u, v, w)$ on $[-1, 1]^3$. By assumption 1 in Section 2.3 that $g(\cdot)$ is bounded, one can easily derive that C is a finite constant.

$$\begin{aligned}
A_{22} &= \nu^3 L^{-2} \int\int\int K_h\{\Delta - (s_1 - s_2)\}K_h\{\Delta' - (s_1 - s_4)\} \\
&\qquad \times([\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}\{x_1, x_2, (s_1 - s_2)\}][\mathcal{V}(x_3, x_4, \Delta') - \mathcal{V}\{x_3, x_4, (s_1 - s_4)\}]
\end{aligned}$$

$$
\begin{aligned}
&+\mathcal{M}\{x_1, x_2, x_3, x_4, (s_1-s_2), (s_1-s_4), (s_2-s_4)\} \\
&+I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}\{x_2, x_4, (s_2-s_4)\})g(s_1/L)g(s_2/L)g(s_4/L)ds_1ds_2ds_4 \\
= &\ \nu^3 L\int\int\int K_h\{\Delta - L(t_1-t_2)\}K_h\{\Delta' - L(t_1-t_4)\} \\
&\times([\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}\{x_1, x_2, L(t_1-t_2)\}][\mathcal{V}(x_3, x_4, \Delta') - \mathcal{V}\{x_3, x_4, L(t_1-t_4)\}] \\
&+\mathcal{M}\{x_1, x_2, x_3, x_4, L(t_1-t_2), L(t_1-t_4), L(t_2-t_4)\} \\
&+I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}\{x_2, x_4, L(t_2-t_4)\})g(t_1)g(t_2)g(t_4)dt_1dt_2dt_4 \\
= &\ \nu^3 L\int\int K_h(\Delta + Lu_1)K_h(\Delta' + Lu_2) \\
&\times[\{\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}(x_1, x_2, -Lu_1)\}\{\mathcal{V}(x_3, x_4, \Delta') - \mathcal{V}(x_3, x_4, -Lu_2)\} \\
&+\mathcal{M}\{x_1, x_2, x_3, x_4, -Lu_1, -Lu_2, L(u_1-u_2)\} \\
&+I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}\{x_2, x_4, L(u_1-u_2)\}]f_2(u_1, u_2)du_1du_2 \\
= &\ \nu^3 L^{-1}\int\int K(v_1)K(v_2)[I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}\{x_2, x_4, (v_1-v_2)h + \Delta' - \Delta\} \\
&+\{\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}(x_1, x_2, \Delta - hv_1)\}\{\mathcal{V}(x_3, x_4, \Delta) - \mathcal{V}(x_3, x_4, \Delta' - hv_2)\} \\
&+\mathcal{M}\{x_1, x_2, x_3, x_4, \Delta - hv_1, \Delta' - hv_2, (v_1-v_2)h + \Delta' - \Delta\}] \\
&\times f_2\{(-\Delta + hv_1)/L, (-\Delta + v_1 h)/L\}dv_1dv_2 \\
= &\ \nu^3 L^{-1}f_2(0,0)\{\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta', \Delta' - \Delta) \\
&+I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, \Delta' - \Delta)\} + o(L^{-1}).
\end{aligned}
$$

It is easy to see that A_{23}-A_{25} have the same order as A_{22}. Further, we have

$$
\begin{aligned}
A_{26} = &\ \nu^2 L^{-2}\int\int K_h\{\Delta - (s_1-s_2)\}K_h\{\Delta' - (s_1-s_2)\} \\
&\qquad\times(\mathcal{M}\{x_1, x_2, x_3, x_4, (s_1-s_2), (s_1-s_2), 0\} \\
&+[\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}\{x_1, x_2, (s_1-s_2)\}][\mathcal{V}(x_3, x_4, \Delta') - \mathcal{V}\{x_3, x_4, (s_1-s_2)\}] \\
&+\{I(x_2=x_4)\sigma_\epsilon^2\mathcal{V}(x_1, x_3, 0) + I(x_1=x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, 0) \\
&+I(x_1=x_3, x_2=x_4)\sigma_\epsilon^4\}) \times g(s_1/L)g(s_2/L)ds_1ds_2
\end{aligned}
$$

$$
\begin{aligned}
&= I(\Delta = \Delta')\nu^2 \int\int K_h^2\{\Delta - L(t_1 - t_2)\} \\
&\quad \times(\mathcal{M}\{x_1, x_2, x_3, x_4, L(t_1 - t_2), L(t_1 - t_2), 0\} \\
&\quad +[\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}\{x_1, x_2, L(t_1 - t_2)\}][\mathcal{V}(x_3, x_4, \Delta) - \mathcal{V}\{x_3, x_4, L(t_1 - t_2)\}] \\
&\quad +\{I(x_2 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2, x_4, 0) \\
&\quad +I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}) \times g(t_1)g(t_2)dt_1dt_2 \\
&= I(\Delta = \Delta')\nu^2 \int K_h^2(\Delta - Lu)[\mathcal{M}(x_1, x_2, x_3, x_4, Lu, Lu, 0) \\
&\quad +\{\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}(x_1, x_2, Lu)\}\{\mathcal{V}(x_3, x_4, \Delta) - \mathcal{V}(x_3, x_4, Lu)\} \\
&\quad +\{I(x_2 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2, x_4, 0) \\
&\quad +I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}] \times f_1(u)du \\
&= I(\Delta = \Delta')\nu^2 L^{-1}h^{-1} \int K^2(v)[\mathcal{M}(x_1, x_2, x_3, x_4, \Delta - hv, \Delta - hv, 0) \\
&\quad +\{\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}(x_1, x_2, \Delta - hv)\}\{\mathcal{V}(x_3, x_4, \Delta) - \mathcal{V}(x_3, x_4, \Delta - hv)\} \\
&\quad +\{I(x_2 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2, x_4, 0) \\
&\quad +I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}] \times f_1\{(\Delta - hv)/L\}dv \\
&= I(\Delta = \Delta')\nu^2 L^{-1}h^{-1} R_K f_1(0)[\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta, 0) + \{I(x_2 = x_4)\sigma_\epsilon^2 \\
&\quad \times\mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_2, x_4, 0) + I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\} + o(1)].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
A_{27} &= \nu^2 L^{-2} \int\int K_h\{\Delta - (s_1 - s_2)\}K_h\{\Delta' - (s_2 - s_1)\} \\
&\quad \times(\mathcal{M}\{x_1, x_2, x_3, x_4, (s_1 - s_2), (s_2 - s_1), (s_2 - s_1)\} \\
&\quad +[\mathcal{V}(x_1, x_2, \Delta) - \mathcal{V}\{x_1, x_2, (s_1 - s_2)\}][\mathcal{V}(x_3, x_4, \Delta') - \mathcal{V}\{x_3, x_4, (s_2 - s_1)\}] \\
&\quad +\{I(x_2 = x_3)\sigma_\epsilon^2 \mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4)\sigma_\epsilon^2 \mathcal{V}(x_2, x_3, 0) \\
&\quad +I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}) \times g(s_1/L)g(s_2/L)ds_1ds_2 \\
&= I(\Delta = -\Delta')\nu^2 \int\int K_h^2\{\Delta - L(t_1 - t_2)\}
\end{aligned}
$$

$$
\begin{aligned}
& \times(\mathcal{M}\{x_1,x_2,x_3,x_4,L(t_1-t_2),L(t_2-t_1),L(t_2-t_1)\} \\
& +[\mathcal{V}(x_1,x_2,\Delta)-\mathcal{V}\{x_1,x_2,L(t_1-t_2)\}][\mathcal{V}(x_3,x_4,-\Delta)-\mathcal{V}\{x_3,x_4,L(t_2-t_1)\}] \\
& +\{I(x_2=x_3)\sigma_\epsilon^2\mathcal{V}(x_1,x_4,0)+I(x_1=x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) \\
& +I(x_1=x_3,x_2=x_4)\sigma_\epsilon^4\})\times g(t_1)g(t_2)dt_1dt_2 \\
= \; & I(\Delta=-\Delta')\nu^2\int K_h^2(\Delta-Lu)[\mathcal{M}(x_1,x_2,x_3,x_4,Lu,-Lu,-Lu) \\
& +\{\mathcal{V}(x_1,x_2,\Delta)-\mathcal{V}(x_1,x_2,Lu)\}\{\mathcal{V}(x_3,x_4,\Delta)-\mathcal{V}(x_3,x_4,Lu)\} \\
& +\{I(x_2=x_3)\sigma_\epsilon^2\mathcal{V}(x_1,x_4,0)+I(x_1=x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) \\
& +I(x_1=x_3,x_2=x_4)\sigma_\epsilon^4\}]\times f_1(u)du \\
= \; & I(\Delta=-\Delta')\nu^2L^{-1}h^{-1}\int K^2(v)[\mathcal{M}(x_1,x_2,x_3,x_4,\Delta-hv,-\Delta+hv,-\Delta+hv) \\
& +\{\mathcal{V}(x_1,x_2,\Delta)-\mathcal{V}(x_1,x_2,\Delta-hv)\}\{\mathcal{V}(x_3,x_4,\Delta)-\mathcal{V}(x_3,x_4,\Delta-hv)\} \\
& +\{I(x_2=x_3)\sigma_\epsilon^2\mathcal{V}(x_1,x_4,0)+I(x_1=x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) \\
& +I(x_1=x_3,x_2=x_4)\sigma_\epsilon^4\}]\times f_1\{(\Delta-hv)/L\}dv \\
= \; & I(\Delta=-\Delta')\nu^2L^{-1}h^{-1}R_Kf_1(0)[\mathcal{M}(x_1,x_2,x_3,x_4,\Delta,-\Delta,-\Delta) \\
& +\{I(x_2=x_3)\sigma_\epsilon^2\mathcal{V}(x_1,x_4,0)+I(x_1=x_4)\sigma_\epsilon^2\mathcal{V}(x_2,x_3,0) \\
& +I(x_1=x_3,x_2=x_4)\sigma_\epsilon^4\}+o(1)].
\end{aligned}
$$

Both A_{26} and A_{27} are of order $O\{(Lh)^{-1}\}$, while the rest terms are of order $O(L^{-1})$. The proof is completed by summarizing the contribution of each term to cov$\{\beta(x_1,x_2,\Delta),$ $\beta(x_3,x_4,\Delta')\}$.

LEMMA **A.4** With $\beta(x_1,x_2,\Delta)$ defined as in Lemma A.3, and with all assumptions in Section 2.3, we have

$$(Lh)^{1/2}[\beta(x_1,x_2,\Delta)-E\{\beta(x_1,x_2,\Delta)\}]\Rightarrow \text{Normal}\{0,\nu^2f_1(0)\sigma^2(x_1,x_2,\Delta)\},$$

where $\sigma^2(x_1,x_2,\Delta)=R_K\{\mathcal{M}(x_1,x_2,x_1,x_2,\Delta,\Delta,0)+\sigma_\epsilon^2\mathcal{V}(x_1,x_1,0)+\sigma_\epsilon^2\mathcal{V}(x_2,x_2,0)+\sigma_\epsilon^4\}+I(\Delta=0)R_K[\{\mathcal{M}(x_1,x_2,x_1,x_2,0,0,0)+I(x_1=x_2)\{2\sigma_\epsilon^2\mathcal{V}(x_1,x_1,0)+\sigma_\epsilon^4\}]$.

Proof: The proof shares the similar structure to that of Theorem 2 in Guan et al. (2004). Define $a_1 = 0$, $b_1 = L^p - L^q$, $a_i = a_{i-1} + L^p$, $b_i = a_i + L^p - L^q$, $i = 2, \cdots, k_L$, for some $1/(1+\delta) < q < p < 1$ (δ is defined in (2.10)). We thus have divided the interval $[0, L]$ into $k_L \approx L/L^p$ disjoint subintervals each having length $L^p - L^q$ and at least L^q apart. Define $I_i = [a_i, b_i]$, $I = \cup_{i=1}^{k_L} I_i$, $I_i' = [a_i/L, b_i/L]$, $I' = \cup_{i=1}^{k_L} I_i'$, and

$$\begin{aligned}\beta_i(x_1, x_2, \Delta) &= L^{-1} \int\int_{I_i \times I_i} K_h\{\Delta - (s_1 - s_2)\}[\{Y(s_1, x_1) - \Theta(x_1)\} \\ &\qquad \times \{Y(s_2, x_2) - \Theta(x_2)\} - \mathcal{V}(x_1, x_2, \Delta)]N_2(ds_1, ds_2), \\ \widetilde{\beta}(x_1, x_2, \Delta) &= \sum_{i=1}^{k_L} \beta_i(x_1, x_2, \Delta).\end{aligned}$$

Define independent random variables $\gamma_i(x_1, x_2, \Delta)$ on a different probability space, such that they have the same distributions as $\beta_i(x_1, x_2, \Delta)$, and define

$$\gamma(x_1, x_2, \Delta) = \sum_{i=1}^{k_L} \gamma_i(x_1, x_2, \Delta).$$

Let $\phi(\xi)$ and $\psi(\xi)$ be the characteristic functions of $(Lh)^{1/2}[\widetilde{\beta}(x_1, x_2, \Delta) - E\{\widetilde{\beta}(x_1, x_2, \Delta)\}]$ and $(Lh)^{1/2}[\gamma(x_1, x_2, \Delta) - E\{\gamma(x_1, x_2, \Delta)\}]$, respectively.

We finish the proof in the following 3 steps:

(i) $(Lh)^{1/2}([\{\beta(x_1, x_2, \Delta) - E\{\beta(x_1, x_2, \Delta)\}] - \{\widetilde{\beta}(x_1, x_2, \Delta) - E\{\widetilde{\beta}(x_1, x_2, \Delta)\}]) \xrightarrow{p} 0$;

(ii) $\psi(\xi) - \phi(\xi) \to 0$;

(iii) $(Lh)^{1/2}[\gamma(x_1, x_2, \Delta) - E\{\gamma(x_1, x_2, \Delta)\}] \Rightarrow \text{Normal}\{0, \nu^2 f_1(0)\sigma^2(x_1, x_2, \Delta)\}$.

To show (i), notice that, with $|I_i| \to \infty$, calculations as in Lemma A.3 show that

$$\sum_{i=1}^{k_L} \text{var}\{\beta_i(x_1, x_2, \Delta)\} = \sum_{i=1}^{k_L} \nu^2 L^{-1} h^{-1} R_K f_{i,1}(0)\{\sigma^2(x_1, x_2, \Delta) + o(1)\}, \quad \text{(A.1)}$$

where $f_{i,1}(u)=\int g_i(u+t)g_i(t)dt$ is the counterpart of $f_1(u)$, with $g_i(t)=g(t)I(t\in I_i')$. Since $g(\cdot)$ is bounded away from both 0 and ∞, $f_{i,1}(0)=\int_{I_i'}g^2(t)=O(|I_i'|)=O(L^{p-1})$, and var$\{\beta_i(x_1,x_2,\Delta)\}=O(L^{p-2}h^{-1})$.

Observe that $|I'|=\sum_{i=1}^{k_L}|I_i'|=k_L\times(L^p-L^q)/L\approx L/L^p\times(L^p-L^q)/L=1-L^{q-p}\to 1$, and

$$\sum_{i=1}^{k_L}f_{i,1}(0)=\sum_{i=1}^{k_L}\int_{I_i'}g(t)^2dt=\int_{I'}g(t)^2dt\to\int_0^1 g(t)^2dt=f_1(0). \qquad \text{(A.2)}$$

Therefore, $\sum_{i=1}^{k_L}$ var$\{\beta_i(x_1,x_2,\Delta)\}=$ var$\{\beta(x_1,x_2,\Delta)\}+o(L^{-1}h^{-1})$. Further but equivalent derivations show that $\sum_{i\neq j}$ cov$\{\beta_i(x_1,x_2,\Delta),\beta_j(x_1,x_2,\Delta)\}=O(L^{-1})$. The calculations here are similar to those in Lemma A.3, except that the $i\neq j$ condition excluded terms like A_{22} through A_{27}. Now we have

$$\begin{aligned}\text{var}\{\widetilde{\beta}(x_1,x_2,\Delta)\} &= \sum_{i=1}^{k_L}\text{var}\{\beta_i(x_1,x_2,\Delta)\}+\sum_{i\neq j}\text{cov}\{\beta_i(x_1,x_2,\Delta),\beta_j(x_1,x_2,\Delta)\}\\ &= \text{var}\{\beta(x_1,x_2,\Delta)\}+o(L^{-1}h^{-1}).\end{aligned}$$

Similarly, one can show that

$$\text{cov}\{\widetilde{\beta}(x_1,x_2,\Delta),\beta(x_1,x_2,\Delta)\}=\text{var}\{\beta(x_1,x_2,\Delta)\}+o(L^{-1}h^{-1}).$$

Therefore, (Lh)var$[\{\beta(x_1,x_2,\Delta)-\{\widetilde{\beta}(x_1,x_2,\Delta)\}]\to 0$, and step (i) is established.

To show (ii), we follow similar arguments that prove Theorem 2 (S2) in Guan et al. (2004). Denote $U_i=\exp(Ix(Lh)^{1/2}[\beta_i(x_1,x_2,\Delta)-E\{\beta_i(x_1,x_2,\Delta)\}])$, where I is the unit imaginary number. Then by definitions, $\phi(x)=E(\prod_{i=1}^{k_L}U_i)$, $\psi(x)=\prod_{i=1}^{k_L}E(U_i)$.

Observing $|E(U_i)|\leq 1$ for all U_i, we have

$$\begin{aligned}|\phi(x)-\psi(x)| &\leq |E(\prod_{i=1}^{k_L}U_i)-E(\prod_{i=1}^{k_L-1}U_i)E(U_{k_L})|+|E(\prod_{i=1}^{k_L-1}U_i)E(U_{k_L})-\prod_{i=1}^{k_L}E(U_i)|\\ &\leq |E(\prod_{i=1}^{k_L}U_i)-E(\prod_{i=1}^{k_L-1}U_i)E(U_{k_L})|+|E(\prod_{i=1}^{k_L-1}U_i)-\prod_{i=1}^{k_L-1}E(U_i)||E(U_{k_L})|\end{aligned}$$

$$\leq \quad |E(\prod_{i=1}^{k_L} U_i) - E(\prod_{i=1}^{k_L-1} U_i)E(U_{k_L})| + |E(\prod_{i=1}^{k_L-1} U_i) - \prod_{i=1}^{k_L-1} E(U_i)|.$$

By induction,

$$|\phi(x) - \psi(x)| \quad \leq \quad \sum_{j=1}^{k_L-1} |E(\prod_{i=1}^{j+1} U_i) - E(\prod_{i=1}^{j} U_i)E(U_{j+1})| = \sum_{j=1}^{k_L-1} |\text{cov}(\prod_{i=1}^{j} U_i, U_{j+1})|.$$

Observe that $\prod_{i=1}^{j} U_i$ and U_{j+1} are $\mathcal{F}([0, b_j])$ and $\mathcal{F}([a_{j+1}, b_{j+1}])$ measurable respectively, with $|\prod_{i=1}^{j} U_i| \leq 1$ and $|U_{j+1}| \leq 1$, and the index sets are at least L^q away. By Lemma A.1,

$$|\phi(x) - \psi(x)| \quad \leq \quad \sum_{j=1}^{k_L-1} 16\alpha(L^q) \leq 16L^{1-p} \times L^{-q\delta}.$$

By our choice of p and q, it is easy to check $1-p-q\delta < 0$, and therefore $|\phi(x)-\psi(x)| \to 0$.

(iii) can be proved by applying Lyapounov's central limit theorem and by the fact that

$$(Lh) \sum_{i=1}^{k_L} \text{var}\{\gamma_i(x_1, x_2, \Delta)\} \to \nu^2 f_1(0)\sigma^2(x_1, x_2, \Delta),$$

which has been shown in (A.1) and (A.2).

It remains to check the Lyapounov's condition. By condition (2.9),

$$\begin{aligned}
\sum_{i=1}^{k_L} \frac{E(|\gamma_i(x_1, x_2, \Delta) - E\{\gamma_i(x_1, x_2, \Delta)\}|^{2+\eta})}{[\text{var}\{\gamma(x_1, x_2, \Delta)\}]^{(2+\eta)/2}} &= L^{1-p} \times \frac{O\{(L^{p-2}h^{-1})^{(2+\eta)/2}\}}{O\{(L^{-1}h^{-1})^{(2+\eta)/2}\}} \\
&= O(L^{-(1-p)\eta/2}) \to 0.
\end{aligned}$$

The proof is thus complete.

<u>LEMMA</u> **A.5** Let $\vec{\beta}(\Delta)$ be the vector collecting all $\beta(x_1, x_2, \Delta)$ for distinct pairs of (x_1, x_2). Then, with all assumptions above, for $\Delta' \neq \Delta$,

$$(Lh)^{1/2} \begin{bmatrix} \vec{\beta}(\Delta) - E\{\vec{\beta}(\Delta)\} \\ \vec{\beta}(\Delta') - E\{\vec{\beta}(\Delta')\} \end{bmatrix} \Rightarrow \text{Normal} \left\{ 0, \nu^2 f_1(0) \begin{pmatrix} \Sigma(\Delta) & C(\Delta, \Delta') \\ C^T(\Delta, \Delta') & \Sigma(\Delta') \end{pmatrix} \right\},$$

where $\Sigma(\Delta)$ is the covariance matrix with the entry corresponding to cov$\{\beta(x_1, x_2, \Delta), \beta(x_3, x_4, \Delta)\}$ equal to $R_K\{\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta, 0) + I(x_2 = x_4)\sigma_\epsilon^2\mathcal{V}(x_1, x_3, 0) + I(x_1 = x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, 0) + (x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\} + I(\Delta = 0)R_K\{\mathcal{M}(x_1, x_2, x_3, x_4, 0, 0, 0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2, x_3, 0) + I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4, x_2 = x_3)\sigma_\epsilon^4\}$; $C(\Delta, \Delta')$ is the matrix with the entry corresponding to cov$\{\beta(x_1, x_2, \Delta), \beta(x_3, x_4, \Delta')\}$ equal to $I(\Delta' = -\Delta)\{\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, -\Delta, -\Delta) + I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1, x_4, 0) + I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2, x_3, 0) + I(x_1 = x_4, x_2 = x_3)\sigma_\epsilon^4\}$.

Proof: Using similar proofs as for Lemma A.3 and A.4, we can show that any linear combination $\sum_{i=1}^k c_i\beta(x_{i1}, x_{i2}, \Delta) + \sum_{i=1}^{k'} c_i'\beta(x_{i1}, x_{i2}, \Delta')$ is asymptotically normal. By the Crámer-Wold device (Serfling 1980), the joint normality is established.

Note: If $\Delta' = -\Delta$, the limiting distribution on the right hand side is a degenerate multivariate normal distribution, because $\beta(x_1, x_1, \Delta) = \beta(x_1, x_1, -\Delta)$ for all x_1.

Estimation With Multiple Subjects

Now suppose we have R subjects, and R is a fixed number. Define

$$
\begin{aligned}
&Y_{r,ik}(x_j, x_l) = \{Y_{rij} - \Theta_r(x_j)\}\{Y_{rkl} - \Theta_r(x_l)\},\\
&a_r(\Delta) = L^{-1}\sum_i\sum_{k\neq i} K_h(\Delta - \Delta_{r,ik}),\\
&b_r(x_j, x_l, \Delta) = L^{-1}\sum_i\sum_{k\neq i} Y_{r,ik}(x_j, x_l)K_h\{\Delta - \Delta_r(i,k)\},\\
&\beta_r(x_j, x_l, \Delta) = b_r(x_j, x_l, \Delta) - a_r(\Delta)\mathcal{V}(x_j, x_l, \Delta),\\
&c_r(x_j, \Delta) = L^{-1}\sum_i\sum_{k\neq i}\{Y_{rij} - \Theta_r(x_j)\}K_h\{\Delta - \Delta_r(i,k)\}.
\end{aligned}
$$

Further, define

$$
\begin{aligned}
&a(\Delta) = \sum_r a_r(\Delta), \quad b(x_j, x_l, \Delta) = \sum_r b_r(x_j, x_l, \Delta),\\
&\beta(x_j, x_l, \Delta) = \sum_r \beta_r(x_j, x_l, \Delta),\\
&\widehat{\mathcal{V}}_0(x_1, x_2, \Delta) = b(x_1, x_2, \Delta)/a(\Delta).
\end{aligned}
$$

Let $\widehat{\mathcal{V}}_0(\Delta)$ and $\mathcal{V}(\Delta))$ be the vectors collecting all $\widehat{\mathcal{V}}_0(x_1, x_2, \Delta)$ and $\mathcal{V}(x_1, x_2, \Delta))$ for all distinct pairs of (x_1, x_2), respectively.

<u>Lemma</u> **A.6** With the assumptions in Section 2.3, for $\Delta' \neq \Delta$,

$$(RLh)^{1/2}\left\{\begin{array}{c} \widehat{\mathcal{V}}_0(\Delta) - \mathcal{V}(\Delta) - \sigma_K^2\mathcal{V}^{(2)}(\Delta)h^2/2 \\ \widehat{\mathcal{V}}_0(\Delta') - \mathcal{V}(\Delta') - \sigma_K^2\mathcal{V}^{(2)}(\Delta')h^2/2 \end{array}\right\}$$

$$\Rightarrow \text{Normal}\left[0, \{\nu^2 f_1(0)\}^{-1}\left(\begin{array}{cc} \Sigma(\Delta) & C(\Delta,\Delta') \\ C^T(\Delta,\Delta') & \Sigma(\Delta') \end{array}\right)\right],$$

where $\mathcal{V}^{(2)}(\Delta)$ is the vector collecting $\mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta)$ for all distinct pairs of (x_1, x_2).

<u>Proof</u>: Notice that

$$\begin{aligned} \widehat{\mathcal{V}}_0(x_1, x_2, \Delta) - \mathcal{V}(x_1, x_2, \Delta) &= [\sum_{r=1}^{R}\{b_r(x_1, x_2, \Delta) - a_r(\Delta)\mathcal{V}(x_1, x_2, \Delta)\}]/\{\sum_{r=1}^{R} a_r(\Delta)\} \\ &= \beta(x_1, x_2, \Delta)/a(\Delta) \end{aligned}$$

Since subjects are independent, by Lemma A.2, $a(\Delta)/\{\nu^2 R f_1(0)\} \xrightarrow{p} 1$. Also, by Lemma A.5, $(R^{-1}Lh)^{1/2}\{\vec{\beta}(\Delta)^T, \vec{\beta}(\Delta')^T\}^T$ are asymptotically jointly normal with the covariance matrix given in Lemma A.5. Thus, by Slutsky's theorem (Serfling, 1980),

$$(RLh)^{1/2}\left[\begin{array}{c} \beta(\Delta)/a(\Delta) - E\{\beta(\Delta)\}/a(\Delta) \\ \beta(\Delta')/a(\Delta') - E\{\beta(\Delta')\}/a(\Delta') \end{array}\right]$$

$$\Rightarrow \text{Normal}\left[0, \{\nu^2 f_1(0)\}^{-1}\left(\begin{array}{cc} \Sigma(\Delta) & C(\Delta,\Delta') \\ C^T(\Delta,\Delta') & \Sigma(\Delta') \end{array}\right)\right].$$

Finally, by Lemma A.3, $E\{\beta(x_1, x_2, \Delta)\} = R\nu^2 f_1(0)\{\mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta)\sigma_K^2 h^2/2 + o(h^2)\}$, so that we have $E\{\beta(x_1, x_2, \Delta)\}/a(\Delta) = \sigma_K^2\mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta)h^2/2 + o_p(h^2)$. The $o_p(h^2)$ term is eliminated by the assumption that $Lh^5 = O(1)$.

<u>Lemma</u> **A.7** With all the assumptions above, we have that

$$\widehat{\mathcal{V}}(x_1, x_2, \Delta) = \widehat{\mathcal{V}}_0(x_1, x_2, \Delta) + O_p\{L^{-1}h^{-1/2}\}$$

Proof: Notice that

$$
\begin{aligned}
\widehat{\mathcal{V}}(x_j, x_l, \Delta) &= \widehat{\mathcal{V}}_0(x_j, x_l, \Delta) + \Big[\sum_r \{\overline{Y}_{r\cdot j} - \Theta_r(x_j)\} c_r(x_l, \Delta) \\
&\quad + \sum_r \{\overline{Y}_{r\cdot j} - \Theta_r(x_j)\}\{\overline{Y}_{r\cdot l} - \Theta_r(x_l)\} a_r(\Delta) \\
&\quad + \sum_r \{\overline{Y}_{r\cdot l} - \Theta_r(x_l)\} c_r(x_j, \Delta) \Big] \times a(\Delta)^{-1}, \qquad \text{(A.3)} \\
c_r(x_1, \Delta) &= L^{-1} \int\int \{Y(s_1, x_1) - \Theta_r(x_1)\} K_h\{\Delta - (s_1 - s_2)\} N_2(ds_1, ds_2).
\end{aligned}
$$

Using the expression above, it is easy to see that $E\{c_r(x_1, \Delta)\} = 0$, and calculations as in Lemma A.3 show that

$$
\begin{aligned}
\text{var}\{c_r(x_1, \Delta)\} &= L^{-2} \int\int\int\int K_h\{\Delta - (s_1 - s_2)\} K_h\{\Delta - (s_3 - s_4)\} \\
&\quad \times [\mathcal{V}\{x_1, x_1, (s_1 - s_3)\} + I(s_1 = s_3)\sigma_\epsilon^2] \\
&\quad \times E\{N_2(ds_1, ds_2) N_2(ds_3, ds_4)\} \\
&= O\{\nu^2 L^{-1} h^{-1}\}.
\end{aligned}
$$

On the other hand, $\overline{Y}_{r\cdot j} - \Theta_r(x_j) = \frac{1}{N_r} \int \{Y_r(s, x_j) - \Theta_r(s, x_j)\} N(ds)$. It is easy to see that $E\{\overline{Y}_{r\cdot j} - \Theta_r(x_j)\} = 0$, and that

$$
\begin{aligned}
\text{var}[N_r\{\overline{Y}_{r\cdot j} - \Theta_r(x_j)\}] &= \int\int [\mathcal{V}\{x_j, x_j, (s_1 - s_2)\} + I(s_1 = s_2)\sigma_\epsilon^2] \\
&\quad \times \{\nu^2 g(s_1/L) g(s_2/L) ds_1 ds_2 + \nu g(s_1/L) \epsilon_{s_1}(ds_2) ds_1\} \\
&= \nu^2 L^2 \int \mathcal{V}(x_j, x_j, Lu) f_1(u) du \\
&\quad + \nu L \int \{\mathcal{V}(x_j, x_j, 0) + \sigma_\epsilon^2\} g(s_1) ds_1 \\
&= \nu^2 L f_1(0) \int \mathcal{V}(x_j, x_j, u) du + \nu L \{\mathcal{V}(x_j, x_j, 0) + \sigma_\epsilon^2\} + o(L).
\end{aligned}
$$

By properties of Poisson processes, we have $N_r/(\nu L) \to 1$ a.s.. Therefore, we have $\overline{Y}_{r\cdot j} - \Theta_r(x_j) = O_p\{L^{-1/2}\}$, $c_r(x_1, \Delta) = O_p\{L^{-1/2} h^{-1/2}\}$. By Lemma A.2, $a_r(\Delta) = O_p(1)$. Therefore, $\widehat{\mathcal{V}}(x_1, x_2, \Delta) - \widehat{\mathcal{V}}_0(x_1, x_2, \Delta) = O_p\{L^{-1} h^{-1/2}\}$, completing the proof.

Proof of Theorem II.1: This is a direct result from Lemma A.6 and A.7.

Proof of Theorem II.2: For a fixed $\Delta \neq 0$, when $h \leq |\Delta|$, we have

$$\widetilde{\mathcal{V}}(x_1, x_2, \Delta) = \{\widehat{\mathcal{V}}(x_1, x_2, \Delta) + \widehat{\mathcal{V}}(x_1, x_2, -\Delta)\}/2.$$

This equation is true automatically for $\Delta = 0$. Therefore, asymptotic distribution of $\widetilde{\mathcal{V}}(\Delta)$ is the same as that of $\{\widehat{\mathcal{V}}(\Delta) + \widehat{\mathcal{V}}(-\Delta)\}/2$, for any fixed Δ.

For $\Delta_1 \neq \pm\Delta_2$, by Theorem II.1, $\{\widehat{\mathcal{V}}(\Delta_1), \widehat{\mathcal{V}}(-\Delta_1)\}^T$ and $\{\widehat{\mathcal{V}}(\Delta_2), \widehat{\mathcal{V}}(-\Delta_2)\}^T$ are asymptotically independent, and the joint asymptotic normality of the four vectors can be established. Therefore $\widetilde{\mathcal{V}}(\Delta_1)$ and $\widetilde{\mathcal{V}}(\Delta_2)$ are jointly asymptotic normal and asymptotically independent. It suffices to show that $\Omega(\Delta)$ is the asymptotic covariance matrix of $\widetilde{\mathcal{V}}(\Delta)$.

For $\Delta \neq 0$, apply the delta method to the joint asymptotic distribution of $\widehat{\mathcal{V}}(\Delta)$ and $\widehat{\mathcal{V}}(-\Delta)$, the following gives the asymptotic covariance between $\widetilde{\mathcal{V}}(x_1, x_2, \Delta)$ and $\widetilde{\mathcal{V}}(x_3, x_4, \Delta)$:

$$\begin{aligned}(1/4)(RLh)^{-1}\{\nu^2 f_1(0)\}^{-1}R_K \times \{&\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta, 0)\\ &+\mathcal{M}(x_1, x_2, x_3, x_4, -\Delta, -\Delta, 0)\\ &+2\mathcal{M}(x_1, x_2, x_3, x_4, \Delta, -\Delta, -\Delta) + 2I(x_2 = x_4)\sigma_\epsilon^2\mathcal{V}(x_1, x_3, 0)\\ &+2I(x_1 = x_3)\sigma_\epsilon^2\mathcal{V}(x_2, x_4, 0) + 2I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\\ &+2I(x_2 = x_3)\sigma_\epsilon^2\mathcal{V}(x_1, x_4, 0) + 2I(x_1 = x_4)\sigma_\epsilon^2\mathcal{V}(x_2, x_3, 0)\\ &+2I(x_1 = x_3, x_2 = x_4)\sigma_\epsilon^4\}\end{aligned}$$

Note that $\mathcal{M}(x_1, x_2, x_3, x_4, -\Delta, -\Delta, 0) = \mathcal{M}(x_1, x_2, x_3, x_4, \Delta, \Delta, 0)$ by the symmetry in the definition of $\mathcal{M}(x_1, x_2, x_3, x_4, u, v, w)$. Next, for $\Delta = 0$, we have $\widetilde{\mathcal{V}}(x_1, x_2, 0) = \widehat{\mathcal{V}}(x_1, x_2, 0)$, the asymptotic covariance between $\widetilde{\mathcal{V}}(x_1, x_2, 0)$ and $\widetilde{\mathcal{V}}(x_3, x_4, 0)$ is given in Theorem II.1. The proof is completed.

Proof of Corollary II.1: The result follows from Theorem II.2 and the Delta-method. To see this, note that, with the separable structure in (2.3), we have $\mathcal{V}(x_1, x_2, \Delta) = G(x_1, x_2)\rho(\Delta)$

and $\mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta) = G(x_1, x_2)\rho^{(2)}(\Delta)$. By the Delta-method, the asymptotic mean of $\widehat{\rho}(\Delta)$ is

$$\begin{aligned} & \frac{\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \leq x_1} \{\mathcal{V}(x_1, x_2, \Delta) + \sigma_K^2 \mathcal{V}^{(0,0,2)}(x_1, x_2, \Delta) h^2/2 + o_p(h^2)\}}{\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \leq x_1} \{G(x_1, x_2) + \sigma_K^2 G(x_1, x_2) \rho^{(2)}(0) h^2/2 + o_p(h^2)\}} \\ = \;& \{\rho(\Delta) + \sigma_K^2 \rho^{(2)}(\Delta) h^2/2 + o_p(h^2)\}/\{1 + \sigma_K^2 \rho^{(2)}(0) h^2/2 + o_p(h^2)\} \\ = \;& \{\rho(\Delta) + \sigma_K^2 \rho^{(2)}(\Delta) h^2/2 + o_p(h^2)\} * \{1 - \sigma_K^2 \rho^{(2)}(0) h^2/2 + o_p(h^2)\} \\ = \;& \rho(\Delta) + \{\rho^{(2)}(\Delta) - \rho(\Delta)\rho^{(2)}(0)\}\sigma_K^2 h^2/2 + o_p(h^2). \end{aligned}$$

The asymptotic variance of $\widehat{\rho}(\Delta)$ also follows from the Delta-method.

APPENDIX B

PROOFS FOR CHAPTER III

Throughout this appendix chapter, we adopt the following notations: for any complex-valued vector b, $\|b\|^2 = b^T\bar{b}$; for any zero-mean, complex-valued, random variables X and Y, cov$(X,Y) = \mathrm{E}(X\bar{Y})$.

Proof of Theorem III.1

Since $\tilde{X}_{i,\rho} \in \mathrm{span}\{\phi_i(\cdot), i = -K, \cdots, K\}$, by orthogonality of Fourier basis, $\hat{f}_{\lambda,\rho}$ is spanned by the same set of basis functions. Write $\hat{f}_{\lambda,\rho} = \phi^T\hat{\beta}$. Then $\hat{\beta}$ minimizes

$$n^{-1}\sum_{i=1}^{n}|Y_i - \boldsymbol{Z}_i^T V\bar{b}|^2 + \lambda b^T W\bar{b}$$

among all $2K+1$ dimensional complex vectors, where $V = P_\rho^T\langle\phi,\phi^T\rangle_{L^2} = J^{-1}\bar{\Phi}(I + \rho W)^{-1}$, P_ρ being defined in (3.3). The penalized least square function above is a variant of (3.4). Since Y_i and $\tilde{X}_{i,\rho}$ are real valued, it is clear that $\hat{f}_{\lambda,\rho}$ is real-valued, and $\hat{\beta}$ satisfies $\overline{\hat{\beta}_j} = \hat{\beta}_{-j}$, $j = -K, \cdots, K$.

Define

$$\Omega_{n,J} = n^{-1}\boldsymbol{Z}^T\boldsymbol{Z}, \text{ and } \check{\Omega}_\rho = n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{Z}V, \tag{B.1}$$

and

$$r_i = \langle X_i, f\rangle_{L^2} - \langle\tilde{X}_{i,\rho}, f\rangle_{L^2} \quad \text{and} \quad \boldsymbol{r} = (r_1,\cdots,r_n)^T.$$

It is easy to check that $\check{\Omega}_\rho$ is an Hermite matrix, i.e. $\overline{\check{\Omega}}_\rho^T = \check{\Omega}_\rho$. Suppose that $f = \sum_{j=-\infty}^{\infty}\beta_j\phi_j$, and denote by $\check{f}$ the projection of f onto $\mathrm{span}\{\phi_{-K},\cdots,\phi_K\}$, and $\check{\beta} = (\beta_{-K},\cdots,\beta_K)^T$. By orthogonality of Fourier basis, $\langle\tilde{X}_{i,\rho}, f\rangle_{L^2} = \langle\tilde{X}_{i,\rho}, \check{f}\rangle_{L^2} = \boldsymbol{Z}_i^T V\overline{\check{\beta}}$.

On the other hand, since both $\tilde{X}_{i,\rho}$ and f are real-valued functions, we have $\langle \tilde{X}_{i,\rho}, f\rangle_{L^2} = \boldsymbol{Z}_i^T \bar{V}\check{\beta}$. Writing $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, we have

$$\begin{aligned} \hat{f}_{\lambda,\rho}(t) &= \phi^T(t)(V^T\boldsymbol{Z}^T\boldsymbol{Z}\bar{V} + n\lambda W)^{-1}V^T\boldsymbol{Z}^T\boldsymbol{Y} \\ &= \phi^T(t)(\check{\Omega}_\rho^T + \lambda W)^{-1}n^{-1}V^T\boldsymbol{Z}^T(\boldsymbol{Z}\bar{V}\check{\beta} + \boldsymbol{r} + \boldsymbol{\varepsilon}) \\ &=: \phi^T(t)(\check{\beta}_\lambda + \hat{\beta}_r + \hat{\beta}_\epsilon) \\ &=: \check{f}_{\lambda,\rho}(t) + g_\lambda(t) + h_\lambda(t). \end{aligned} \tag{B.2}$$

By (3.6),

$$\tilde{R}_\rho(s,t) = \frac{1}{n}\sum_{i=1}^n \tilde{X}_{i,\rho}(s)\tilde{X}_{i,\rho}(t) = \phi(s)^T P_\rho \Omega_{n,J} P_\rho^T \phi(t).$$

For any $g(t) = \phi(t)^T b \in \mathrm{span}\{\phi_{-K}, \ldots, \phi_K\}$, we have

$$(\tilde{T}_\rho g)(t) = \int \tilde{R}_\rho(s,t)g(s)ds = \phi^T(t)P_\rho\Omega_{n,J}\bar{P}_\rho^T\{\int \bar{\phi}(s)\phi^T(s)ds\}b = \phi^T(t)P_\rho\Omega_{n,J}\bar{V}b$$

and hence by (B.1),

$$\|g\|_{\tilde{T}_\rho}^2 = \langle g, \tilde{T}_\rho g\rangle_{L^2} = b^T\bar{V}^T\Omega_{n,J}V\bar{b} = b^T\check{\Omega}_\rho\bar{b} = n^{-1}\|\boldsymbol{Z}\bar{V}b\|^2. \tag{B.3}$$

It follows from (B.2) that

$$\|\hat{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2 \le 3\|\check{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2 + 3\|g_\lambda\|_{\tilde{T}_\rho}^2 + 3\|h_\lambda\|_{\tilde{T}_\rho}^2. \tag{B.4}$$

First, consider $\|\check{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2$ which is equal to $\|\check{f}_{\lambda,\rho} - \check{f}\|_{\tilde{T}_\rho}^2$. By (B.3),

$$\|\check{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2 = \|\check{f}_{\lambda,\rho} - \check{f}\|_{\tilde{T}_\rho}^2 = n^{-1}\|\boldsymbol{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2$$

Since $\check{f}_{\lambda,\rho}$ is the solution of the following problem

$$\min_{g=\phi^T b}\{n^{-1}\|\check{\boldsymbol{Z}}\bar{V}(b - \check{\beta})\|^2 + \lambda b^T W\bar{b}\},$$

we conclude that

$$\begin{aligned}\|\check{f}_{\lambda,\rho} - f\|^2_{\tilde{T}_\rho} &= n^{-1}\|\boldsymbol{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2 \\ &\leq n^{-1}\|\boldsymbol{Z}\bar{V}(\check{\beta}_\lambda - \check{\beta})\|^2 + \lambda\check{\beta}_\lambda^T W\overline{\check{\beta}_\lambda} \\ &\leq \lambda\check{\beta}^T W\overline{\check{\beta}} = \lambda\|\check{f}^{(m)}\|^2_{L^2} \leq \lambda\|f^{(m)}\|^2_{L^2}. \end{aligned} \tag{B.5}$$

Note that this is an approach for handling the bias introduced by Craven and Wahba (1979). Next, by (B.3),

$$\begin{aligned}\mathrm{E}(\|h_\lambda\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) &= \mathrm{E}\{\boldsymbol{\varepsilon}^T n^{-1}\boldsymbol{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\overline{\check{\Omega}}_\rho^T + \lambda W)^{-1}n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{\varepsilon}|\boldsymbol{Z}\} \\ &= n^{-1}\sigma_\varepsilon^2 \mathrm{tr}\{(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho + \lambda W)^{-1}\check{\Omega}_\rho\}.\end{aligned}$$

Let ω_i, ν_i, ρ_i, and η_i be the i-th smallest eigenvalue of $W, \check{\Omega}_{n,J}$, $\check{\Omega}_{n,J} + \lambda W$, and $(\check{\Omega}_{n,J} + \lambda W)^{-1}\check{\Omega}_{n,J}$, respectively. Observe that for any complex-valued function $g = \phi^T b$, we have

$$\frac{\langle g, \tilde{T}_\rho g\rangle_{L^2}}{\|g\|^2_{L^2}} = \frac{b\check{\Omega}_\rho\bar{b}}{b^T\bar{b}}, \tag{B.6}$$

and hence the eigenvalues of $\check{\Omega}_\rho$ are the same as those of $\tilde{T}_\rho$, then $\check{\Omega}_\rho$ is positive semi-definite and $\nu_i \geq 0$ for all i. Note that ν_{2K+1} is denoted as ν_ρ in the statement of the theorem to emphasize its dependence on ρ. By (3.2), $\omega_i = ([i/2]2\pi)^{2m}$, $i = 1, \cdots, 2K+1$. We clearly also have

$$\eta_i \leq 1 \quad \text{and} \quad \rho_i \geq \lambda([i/2]2\pi)^{2m} \geq \lambda((i-1)\pi)^{2m} \quad \text{for all } i.$$

It follows from Theorem 7 of Merikoski and Kumar (2004) that

$$\eta_{(2K+1)-i+1} \leq \rho_i^{-1}\nu_\rho \quad \text{for all } i.$$

Thus,

$$\eta_{(2K-1)-i+1} \leq \min(1,\ \nu_\rho\pi^{-2m}\lambda^{-1}(i-1)^{-2m}) \quad \text{for all } i,$$

and therefore

$$\mathrm{tr}\{(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_{n,J}\}$$
$$=\sum_{i=1}^{2K+1}\eta_i^2 \le \sum_{i=1}^{[\lambda^{-1/(2m)}]} 1+\frac{\nu_\rho^2}{\pi^{4m}\lambda^2}\sum_{i=[\lambda^{-1/(2m)}]+1}^{2K+1}(i-1)^{-4m} \le C\lambda^{-1/(2m)}\nu_\rho^2.$$

Hence,

$$\mathrm{E}(\|h_\lambda\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) \le Cn^{-1}\lambda^{-1/(2m)}\nu_\rho^2. \tag{B.7}$$

Next,

$$\mathrm{E}(\|g_\lambda\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) = \mathrm{E}\{\boldsymbol{r}^T n^{-1}\boldsymbol{Z}V(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho+\lambda W)^{-1}n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{r}|\boldsymbol{Z}\}.$$

Since the eigenvalues of $\boldsymbol{Z}V(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho+\lambda W)^{-1}n^{-1}\bar{V}^T\boldsymbol{Z}^T$ are the same as those of

$$(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_\rho(\check{\Omega}_\rho+\lambda W)^{-1}n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{Z}V = [(\check{\Omega}_\rho+\lambda W)^{-1}\check{\Omega}_\rho]^2,$$

which are bounded by 1, we conclude that

$$\mathrm{E}(\|g_\lambda\|^2_{\tilde{T}_\rho}|\boldsymbol{Z}) \le \frac{1}{n}\mathrm{E}(\boldsymbol{r}^T\boldsymbol{r}|\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(r_i^2|\boldsymbol{Z}_i) \le \frac{\|f\|^2_{L^2}}{n}\sum_{i=1}^n \mathrm{E}(\|\tilde{X}_{i,\rho}-X_i\|^2_{L^2}|\boldsymbol{Z}_i), \tag{B.8}$$

by the Cauchy-Schwarz inequality. The result follows from (B.4), (B.5), (B.7), and (B.8).

Proof of Theorem III.2

LEMMA **B.1** *Suppose* $t_j=(2j-1)/(2J)$, $j=1,\cdots,J$, *then*

$$\sum_{j=1}^J \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \begin{cases} (-1)^s J, & \textit{if } k_2=k_1+sJ; \\ 0 & \textit{otherwise}. \end{cases}$$

PROOF. If k_2-k_1 is not a multiple of J, we have

$$\sum_{j=1}^J \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \sum_{j=1}^J e^{(k_2-k_1)2\pi i t_j} = \frac{e^{(k_2-k_1)2\pi i(2J+1)/(2J)}-e^{(k_2-k_1)2\pi i/(2J)}}{1-e^{(k_2-k_1)2\pi i/J}} = 0.$$

Next, suppose $k_2 = k_1 + sJ$ for some integer s, we have

$$\sum_{j=1}^{J} \bar{\phi}_{k_1}(t_j)\phi_{k_2}(t_j) = \sum_{j=1}^{J} e^{s\pi i(2j-1)} = \sum_{j=1}^{J}(-1)^s = J(-1)^s.$$

<u>LEMMA</u> **B.2** *Let* $g \in W_{2,\text{per}}^m$ *have the Fourier basis representation* $g = \sum_{j=-\infty}^{\infty} c_j\phi_j(t)$ *in* $L^2[0,1]$*. Then the Fourier basis representation for* $g^{(m)}$ *is*

$$\sum_{j=-\infty}^{\infty} c_j(2\pi ji)^m\phi_j(t),$$

and we have

$$\|g^{(m)}\|_{L^2}^2 = \sum_{j=-\infty}^{\infty} (2\pi j)^{2m}|c_j|^2.$$

PROOF. Let $g_k = \sum_{j=-k}^{k} c_j\phi_j$ and consider $g^{(m)} - g_k^{(m)}$. Note that the assumption implies that $g^{(\nu)}(0) = g^{(\nu)}(1), 0 \leq \nu \leq m-1$, and $\phi_j^{(s)}(0) = \phi_j^{(s)}(1)$ for all s. Integrating by parts repeatedly,

$$\langle g^{(m)} - g_k^{(m)}, \phi_j\rangle_{L^2} = (-1)^m\langle g - g_k, (2\pi ji)^m\phi_j\rangle_{L^2} = 0, \text{ for all } j = -k, \cdots, k.$$

This means that $g_k^{(m)} = \sum_{i=-k}^{k}(2\pi ji)^m c_j\phi_j$ is the L^2 projection of $g^{(m)}$ on $\text{span}\{\phi_j, j = -k, \cdots, k\}$. Since $g^{(m)} \in L^2[0,1]$, and the Fourier basis is complete, we conclude that $g_k^{(m)} \to g^{(m)}$ in $L^2[0,1]$ and the result follows.

PROOF OF THEOREM III.2. The proof is similar to those in Rice and Rosenblatt (1981). As before,

$$\boldsymbol{X} = (X(t_1), \cdots, X(t_J))^T,\ \varsigma = (\varsigma_1, \cdots, \varsigma_J), \text{ and } \boldsymbol{Z} = \boldsymbol{X} + \varsigma.$$

Let $X(t) = \sum_{j=-\infty}^{\infty} a_j\phi_j(t)$. By Lemma B.1,

$$\tilde{a}_j \ := \ J^{-1}\sum_{l=1}^{J}\phi_j(t_l)X(t_l) = \sum_{s=-\infty}^{\infty}(-1)^s a_{j+sJ},\ j = -K, \cdots, K$$

Thus,

$$\begin{aligned}\tilde{X}_\rho(t) &= \phi^T(t)(I+\rho W)^{-1}J^{-1}\Phi^T\boldsymbol{Z}\\ &= \sum_{j=-K}^{K}(1+(2\pi)^{2m}\rho j^{2m})^{-1}(\tilde{a}_j+\tilde{\varsigma}_j)\phi_j(t),\end{aligned}$$

where $\tilde{\varsigma}_j = J^{-1}\sum_{k=1}^{J}\phi_j(t_k)\varsigma_k$. Next,

$$\begin{aligned}&\mathrm{E}(\|\tilde{X}_\rho - X\|^2)\\ =\;& \mathrm{E}\Big[\int|\sum_{j=-K}^{K}\{a_j-(1+(2\pi)^{2m}\rho j^{2m})^{-1}(\tilde{a}_j+\tilde{\varsigma}_j)\}\phi_j(t)|^2dt\Big]\\ &\qquad\qquad +\mathrm{E}\Big[\int|\sum_{|j|>K}a_j\phi_j(t)|^2dt\Big]\\ =\;& \mathrm{E}\Big[\sum_{j=-K}^{K}\frac{|\rho(2\pi)^{2m}j^{2m}a_j-(\tilde{a}_j-a_j)-\tilde{\varsigma}_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\Big]+\mathrm{E}\Big[\sum_{|j|>K}|a_j|^2\Big]\\ \le\;& 2\mathrm{E}\Big[\sum_{j=-K}^{K}\frac{\rho^2(2\pi)^{4m}j^{4m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\Big]+2\mathrm{E}\Big[\sum_{j=-K}^{K}\frac{|\tilde{a}_j-a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\Big]\\ &\qquad +\sum_{j=-K}^{K}\frac{J^{-1}\sigma_\varsigma^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}+\mathrm{E}\Big[\sum_{|j|>K}|a_j|^2\Big].\end{aligned}$$

Note that, by Lemma B.2,

$$\|X^{(m)}\|_{L^2}^2=\sum_{j=-\infty}^{\infty}(2j\pi)^{2m}|a_j|^2 \quad \text{a.s.}$$

and therefore, with probability one,

$$\sum_{j=-K}^{K}\frac{\rho^2(2\pi)^{4m}j^{4m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2}\le\sum_{j=-K}^{K}\frac{\rho(2\pi)^{2m}j^{2m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})}\le\rho\sum_{j=-K}^{K}(2\pi)^{2m}j^{2m}|a_j|^2\le\rho\|X^{(m)}\|_{L^2}^2.$$

and

$$\sum_{|j|>K}|a_j|^2\le(2\pi)^{-2m}K^{-2m}\sum_{|j|>K}(2\pi)^{2m}j^{2m}|a_j|^2\le K^{-2m}(2\pi)^{-2m}\|X^{(m)}\|_{L^2}^2.$$

By integral approximation,

$$\sum_{j=-K}^{K}(1+\rho(2\pi j)^{2m})^{-r}\sim(2\pi)^{-1}\rho^{-1/(2m)}\int_{-\infty}^{\infty}(1+x^{2m})^{-r}dx,\ r\ge 1. \tag{B.9}$$

On the other hand, with probability one,

$$\begin{aligned}|\tilde{a}_j - a_j|^2 &= |\sum_{s\neq 0}(-1)^s a_{j+sJ}|^2 \\ &\leq \{\sum_{s\neq 0}(j+sJ)^{-2m}\}\{\sum_{s\neq 0}(j+sJ)^{2m}|a_{j+sJ}|^2\} \\ &= O(J^{-2m})\|X^{(m)}\|_{L^2}^2,\end{aligned}$$

so that

$$\sum_{j=-K}^{K}\frac{|\tilde{a}_j - a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2} = O(J^{-2m}\rho^{-1/2m})\|X^{(m)}\|_{L^2}^2$$

Combining these and applying the assumptions $K \sim J/2$, and $\mathrm{E}(\|X^{(m)}\|_{L^2}^2) < \infty$, we obtain (3.8).

Next, if $X \in W_{2,\mathrm{per}}^{2m}$, we have $\sum_{j=-\infty}^{\infty}(2\pi j)^{4m}|a_j|^2 = \|X^{(2m)}\|_{L^2}^2$. The proof of (3.9) is the same as that for (3.8), except that now we have, with probability one,

$$\sum_{|j|>K}|a_j|^2 \leq (2\pi)^{-4m}K^{-4m}\sum_{|j|>K}(2\pi)^{4m}j^{4m}|a_j|^2 \leq K^{-4m}(2\pi)^{-4m}\|X^{(2m)}\|_{L^2}^2;$$

$$|\tilde{a}_j - a_j|^2 \leq \{\sum_{s\neq 0}(j+sJ)^{-4m}\}\{\sum_{s\neq 0}(j+sJ)^{4m}|a_{j+sJ}|^2\} \leq O(J^{-4m})\|X^{(2m)}\|_{L^2}^2;$$

$$\sum_{j=-K}^{K}\frac{\rho^2(2\pi)^{4m}j^{4m}|a_j|^2}{(1+\rho(2\pi)^{2m}j^{2m})^2} \leq \rho^2\sum_{j=-K}^{K}(2\pi)^{4m}j^{4m}|a_j|^2 \leq \rho^2\|X^{(2m)}\|_{L^2}^2.$$

Therefore, the term $O(\rho)$ in (3.8) is replaced by $O(\rho^2)$ and (3.9) follows.

Proof of Theorem III.3

In view of the discussions prior to the statement of the theorem, it suffices to show that $\nu_\rho = O_p(1)$ where ν_ρ is the largest eigenvalue of $\tilde{T}_\rho$. By (B.6), $\tilde{T}_\rho$ and $\check{\Omega}_\rho$ have the same eigenvalues. By (B.1), the eigenvalues of $\check{\Omega}_\rho$ are bounded by those of $J^{-1}\Omega_{n,J} = J^{-1}n^{-1}\boldsymbol{Z}^T\boldsymbol{Z}$. Hence it suffices to show that

$$\sup_{n,J} J^{-2}\mathrm{E}[\mathrm{tr}(\Omega_{n,J}^2)] < \infty.$$

Straightforward computations show that

$$\frac{1}{J^2}\mathrm{E}[\mathrm{tr}(\Omega_{n,J}^2)] = \frac{1}{nJ^2}\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{E}[(X_1(t_j)+\varsigma_{1,j})^2(X_1(t_k)+\varsigma_{1,k})^2] + \frac{n-1}{nJ^2}\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{E}^2[(X_1(t_j)+\varsigma_{1,j})(X_1(t_k)+\varsigma_{1,k})].$$

By the Cauchy-Schwarz inequality, it is sufficient to deal with the first expression on the right. In view of independence,

$$\frac{1}{J^2}\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{E}[(X_1(t_j)+\varsigma_{1,j})^2(X_1(t_k)+\varsigma_{1,k})^2] = \frac{1}{J^2}\sum_{j=1}^{J}\sum_{k=1}^{J}\{\mathrm{E}[X_1^2(t_j)X_1^2(t_k)] + \sigma_\varsigma^2(\mathrm{E}[X_1^2(t_j)] + \mathrm{E}[X_1^2(t_k)]) + \sigma_\varsigma^4\},$$

which will be bounded under the assumption that $\mathrm{E}(\|X_1\|_{L^2}^4) < \infty$.

Proof of Theorem III.4

Let $X_{\{\ell\}} = \sum_{j=-\ell}^{\ell} a_j\phi_j$, namely the projection of X on span$\{\phi_k, -\ell \le k \le \ell\}$. Then

$$R(s,t) = \mathrm{E}[X(s)\bar{X}(t)] = \lim_{\ell_1,\ell_2\to\infty}\mathrm{E}[X_{\{\ell_1\}}(s)\bar{X}_{\{\ell_2\}}(t)] = \sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\mathrm{E}(a_{j_1}\bar{a}_{j_2})\phi_{j_1}(s)\bar{\phi}_{j_2}(t).$$

For convenience, write $a_{j_1,j_2} = \mathrm{E}(a_{j_1}\bar{a}_{j_2})$. By an argument similar to that used in Lemma B.2, using the assumption $\mathrm{E}\{[X^{(m)}(s)]^2\} < \infty$ for all s, we have

$$R^{(m,m)}(s,t) = \sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty} a_{j_1,j_2}(2\pi i j_1)^m(-2\pi i j_2)^m\phi_{j_1}(s)\bar{\phi}_{j_2}(t) = \lim_{\ell_1,\ell_2\to\infty}\mathrm{E}[X_{\{\ell_1\}}^{(m)}(s)X_{\{\ell_1\}}^{(m)}(t)] = \mathrm{E}[X^{(m)}(s)X^{(m)}(t)].$$

Consequently, we have

$$\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}(2\pi j_1)^{2m}(2\pi j_2)^{2m}|a_{j_1,j_2}|^2 = \int_0^1\int_0^1[R^{(m,m)}(s,t)]^2 ds dt = \int_0^1\int_0^1 \mathrm{E}^2[X^{(m)}(s)X^{(m)}(t)] \le \mathrm{E}\|X^{(m)}\|_{L^2}^4 < \infty. \tag{B.10}$$

Similarly, under the additional conditions $X \in W_{2,\mathrm{per}}^{2m}$ a.s., $\mathrm{E}\{[X^{(2m)}(s)]^2\} < \infty$ for all s, and $\mathrm{E}\|X^{(2m)}\|_{L^2}^2 < \infty$, we can show

$$\begin{aligned}\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}(2\pi j_1)^{4m}|a_{j_1,j_2}|^2 &= \int_0^1\int_0^1 |R^{(2m,0)}(s,t)|^2 = \int_0^1\int_0^1 \mathrm{E}^2[X^{(2m)}(s)X(t)] \\ &\le \left(\mathrm{E}\|X^{(2m)}\|_{L^2}^4 \mathrm{E}\|X\|_{L^2}^4\right)^{1/2} < \infty. \qquad \text{(B.11)}\end{aligned}$$

Define $R_\rho = \mathrm{E}[\tilde{R}_\rho(s,t)]$ and let T_ρ be the corresponding covariance operator. The following calculations are similar to those in Lemma III.2. Let $\Sigma = E(\boldsymbol{X}_1\boldsymbol{X}_1^T) = (R(t_l,t_k))_{l,k=1}^J$, it follows that

$$\begin{aligned}R_\rho(s,t) &= \phi^T(s)P_\rho E(\boldsymbol{Z}_1\boldsymbol{Z}_1^T)\bar{P}_\rho^T\bar{\phi}(t) \\ &= J^{-2}\phi^T(s)(I+\rho W)^{-1}\bar{\Phi}^T(\Sigma+\sigma_\varsigma^2 I)\Phi(I+\rho W)^{-1}\bar{\phi}(t) \\ &= \sum_{j=-K}^{K}\sum_{k=-K}^{K}\tilde{a}_{jk}\{1+\rho(2\pi j)^{2m}\}^{-1}\{1+\rho(2\pi k)^{2m}\}^{-1}\phi_j(s)\bar{\phi}_k(t) \\ &\quad +\sigma_\varsigma^2 J^{-1}\sum_{j=-K}^{K}\{1+\rho(2\pi j)^{2m}\}^{-2}\phi_j(s)\bar{\phi}_j(t),\end{aligned}$$

where

$$\begin{aligned}\tilde{a}_{j_1,j_2} &= \sum_{l=1}^{J}\sum_{k=1}^{J}R(t_l,t_k)\bar{\phi}_{j_1}(t_l)\phi_{j_2}(t_k) \\ &= \sum_{s_1=-\infty}^{\infty}\sum_{s_2=-\infty}^{\infty}(-1)^{s_1+s_2}a_{j_1+s_1J,j_1+s_2J},\ -K \le j_1,j_2 \le K.\end{aligned}$$

Then

$$\begin{aligned}\|T_\rho - T\|_{\mathcal{H}}^2 &= \int_0^1\int_0^1[R_\rho(s,t)-R(s,t)]^2 dsdt \\ &\le 2\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}|\tilde{a}_{j_1,j_2}\{1+\rho(2\pi j_1)^{2m}\}^{-1}\{1+\rho(2\pi j_2)^{2m}\}^{-1} - a_{j_1,j_2}|^2 \\ &\quad +2\sigma_\varsigma^4 J^{-2}\sum_{j=-K}^{K}\{1+\rho(2\pi j)^{2m}\}^{-4} + \sum_{|j_1|>K}\sum_{|j_2|>K}|a_{j_1,j_2}|^2.\end{aligned}$$

Suppose first that $X \in W_{2,\mathrm{per}}^m$ and $\mathrm{E}\|X^{(m)}\|_{L^2}^4 < \infty$. By (B.10),

$$\sum_{|j_1|>K}\sum_{|j_2|>K}|a_{j_1,j_2}|^2 \le (2\pi K)^{-4m}\sum_{|j_1|>K}\sum_{|j_2|>K}(2\pi j_1)^{2m}(2\pi j_2)^{2m}|a_{j_1,j_2}|^2 = O_p(J^{-4m}),$$

and by (B.9) with $r = 4$,

$$J^{-2}\sum_{j=-K}^{K}\{1+\rho(2\pi j)^{2m}\}^{-4} = O(J^{-2}\rho^{-1/(2m)}).$$

Now,

$$\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}|\tilde{a}_{j_1,j_2}\{1+\rho(2\pi j_1)^{2m}\}^{-1}\{1+\rho(2\pi j_2)^{2m}\}^{-1} - a_{j_1,j_2}|^2 \le 2(A+B)$$

where

$$\begin{aligned} A &= \sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}|\tilde{a}_{j_1,j_2} - a_{j_1,j_2}|^2\{1+\rho(2\pi j_1)^{2m}\}^{-2}\{1+\rho(2\pi j_2)^{2m}\}^{-2} \\ B &= \sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}|a_{j_1,j_2}\{1+\rho(2\pi j_1)^{2m}\}^{-1}\{1+\rho(2\pi j_2)^{2m}\}^{-1} - a_{j_1,j_2}|^2. \end{aligned}$$

It follows that

$$\begin{aligned} |\tilde{a}_{j_1,j_2} - a_{j_1,j_2}|^2 &\le \Big(\sum_{(s_1,s_2)\neq(0,0)}|a_{j_1+s_1J,j_1+s_2J}|\Big)^2 \\ &\le \Big\{\sum_{(s_1,s_2)\neq(0,0)}(j_1+s_1J)^{-2m}(j_2+s_2J)^{-2m}\Big\} \\ &\qquad \times\Big\{\sum_{(s_1,s_2)\neq(0,0)}(j_1+s_1J)^{2m}(j_2+s_2J)^{2m}|a_{j_1,j_2}|^2\Big\} \\ &\le \Big[\Big\{\sum_{s_1\neq 0}(j_1+s_1J)^{-2m}\Big\}\Big\{\sum_{s_2=-\infty}^{\infty}(j_2+s_2J)^{-2m}\Big\} \\ &\qquad +\Big\{\sum_{s_1=-\infty}^{\infty}(j_1+s_1J)^{-2m}\Big\}\Big\{\sum_{s_2\neq 0}(j_2+s_2J)^{-2m}\Big\}\Big] \\ &\qquad\qquad \times\int_0^1\int_0^1|R^{(m,m)}(s,t)|^2\,ds\,dt \\ &= O(J^{-2m}). \end{aligned} \tag{B.12}$$

By (B.9) with $r = 2$ and (B.12),

$$A = O(J^{-2m}\rho^{-1/m}).$$

Next,

$$\begin{aligned} B &\leq 3\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}\frac{\rho^2(2\pi j_1)^{4m}|a_{j_1,j_2}|^2+\rho^2(2\pi j_2)^{4m}|a_{j_1,j_2}|^2}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2} \\ &\quad +3\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}\frac{\rho^4(2\pi j_1)^{4m}(2\pi j_2)^{4m}|a_{j_1,j_2}|^2}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2}. \end{aligned}$$

Clearly,

$$\frac{\rho^2(2\pi j_1)^{4m}+\rho^2(2\pi j_2)^{4m}}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2}\leq \rho(2\pi j_1)^{2m}+\rho(2\pi j_2)^{2m},$$

and

$$\frac{\rho^4(2\pi j_1)^{4m}(2\pi j_2)^{4m}}{\{1+\rho(2\pi j_1)^{2m}\}^2\{1+\rho(2\pi j_2)^{2m}\}^2}\leq \rho^2(2\pi j_1)^{2m}(2\pi j_2)^{2m}.$$

We thus have

$$B \leq 3\sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K}\{\rho(2\pi j_1)^{2m}+\rho(2\pi j_2)^{2m}+\rho^2(2\pi j_1)^{2m}(2\pi j_2)^{2m}\}|a_{j_1,j_2}|^2 = O(\rho).$$

Combining the various computations, using the fact that if $m \geq 2$ and $J^{-1}\rho^{-1/(2m)} \to 0$ then $J^{-2m}\rho^{-1/m} = o(J^{-2}\rho^{-1/(2m)})$, we conclude

$$\mathrm{E}(\|T_\rho - T\|^2_{\mathcal{H}}) = O(\rho) + O(J^{-2}\rho^{-1/(2m)}) \text{ if } X \in W^m_{2,\text{per}} \text{ and } \mathrm{E}\|X^{(m)}\|^4_{L^2} < \infty. \quad \text{(B.13)}$$

Now if $X \in W^{2m}_{2,\text{per}}$ and $\mathrm{E}\|X^{(2m)}\|^2_{L^2} < \infty$, the same approach shows that $B = O(\rho^2)$. Thus,

$$\mathrm{E}(\|T_\rho - T\|^2_{\mathcal{H}}) = O(\rho^2) + O(J^{-2}\rho^{-1/(2m)}) \text{ if } X \in W^{2m}_{2,\text{per}} \text{ and } \mathrm{E}\|X^{(2m)}\|^4_{L^2} < \infty. \quad \text{(B.14)}$$

Next,

$$\begin{aligned} \mathrm{E}(\|\tilde{T}_\rho - T_\rho\|^2_{\mathcal{H}}) &= \mathrm{E}\int_0^1\int_0^1[\tilde{R}_\rho(s,t)-R_\rho(s,t)]^2dsdt \\ &= n^{-1}\int_0^1\int_0^1 \mathrm{var}\{\tilde{X}_\rho(s)\tilde{X}_\rho(t)\}dsdt \leq n^{-1}\mathrm{E}(\|\tilde{X}_\rho\|^4_{L^2}). \end{aligned}$$

If follows that

$$\mathrm{E}(\|\tilde{X}_\rho\|_{L^2}^4) \leq 8\mathrm{E}(\|X\|_{L^2}^4) + 8\mathrm{E}(\|\tilde{X}_\rho - X\|_{L^2}^4),$$

where $\mathrm{E}(\|X\|_{L^2}^4) < \infty$ by assumption. By calculations in Lemma III.2, we have

$$\|\tilde{X}_\rho - X\|_{L^2}^2 \leq \{\rho + O(J^{-2m})\}\|X\|_{\mathbf{S}}^2 + 2\sum_{j=1}^{J} \frac{\tilde{\varsigma}_j^2}{(1+\rho\pi^{2m}j^{2m})^2}.$$

Some tedious but straightforward calculations show that $\mathrm{E}(\tilde{\varsigma}_j^2\tilde{\varsigma}_k^2) = O(J^{-2})$, and we obtain

$$\mathrm{E}(\|\tilde{X}_\rho - X\|_{L^2}^4) = O(\rho^2) + O(J^{-2}\rho^{-1/m}).$$

We have shown

$$\mathrm{E}(\|\tilde{T}_\rho - T_\rho\|_{\mathcal{H}}^2) = O(n^{-1}). \tag{B.15}$$

The results in theorem follow from (B.13)-(B.15).

Proof of Theorem III.5

Define bilinear forms

$$L(g) = \langle g, T_\rho g\rangle_{L^2} + \langle g^{(m)}, g^{(m)}\rangle_{L^2} \quad \text{and} \quad \tilde{L}(g) = \langle g, \tilde{T}_\rho g\rangle_{L^2} + \langle g^{(m)}, g^{(m)}\rangle_{L^2}.$$

<u>Lemma</u> **B.3** *Assume that the conditions of Theorem III.4 hold. Let* n, J, ρ *be such that* $n^{-1} + \rho + J^{-1}\rho^{-1/(2m)} \to 0$. *The following can be shown:*

1. $\liminf_{J,\rho} \inf_{g=\phi^T b, \|g\|_{L^2}=1} L(g) > 0$, *and*
2. $\lim_{n,J,\rho} P(\inf_{g=\phi^T b, \|g\|_{L^2}=1} \tilde{L}(g) > c) = 1$.

Proof. For convenience let n, J, ρ be indexed by k and $n_k^{-1} + \rho_k + J_k^{-1}\rho_k^{-1/(2m)} \to 0$ as $k \to \infty$.

(i) Notice that the null space of $\langle g^{(m)}, g^{(m)}\rangle_{L^2}$ is spanned by ϕ_0. On the other hand, by our assumption, $\langle \phi_0, T\phi_0\rangle_{L^2} = c_0 > 0$. Fix $0 < \epsilon < c_0$ and pick a large enough k so that we have $\|T_\rho - T\| < \epsilon$ and $\langle \phi_0, T_\rho\phi_0\rangle_{L^2} \geq c_0 - \epsilon$ by Theorem III.4. Note that in this proof, the operator norm can be the usual sup-norm, which is dominated by the Hilbert-Schmidt norm. For any $g = \phi^T b$ with $b^T\bar{b} = 1$, let $h = g - b_0\phi_0$, then $\|h\|^2 = 1 - |b_0|^2$. Thus,

$$\begin{aligned} L(g) &\geq c_1|b_0|^2 - 2|b_0||\langle \phi_0, T_\rho h\rangle_{L^2}| + \langle h^{(m)}, h^{(m)}\rangle_{L^2} \\ &\geq c_1|b_0|^2 - 2c_2\|h\| + c_3\|h\|^2 \text{ for all large } k; \end{aligned}$$

on the other hand, $L(g) \geq \langle g^{(m)}, g^{(m)}\rangle_{L^2} = \langle h^{(m)}, h^{(m)}\rangle_{L^2} \geq c_3\|h\|^2$, therefore

$$L(g) \geq (c_1|b_0|^2 - 2c_2\|h\|)_+ + c_3\|h\|^2 \text{ for all large } k.$$

Note that the minimum of this lower bound does not depend on J or ρ.

(ii) Let c be the $\liminf$ in part (i). For any $0 < \epsilon < c$,

$$\begin{aligned} P(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) \geq c - \epsilon) &= P(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) \geq c - \epsilon, \|\tilde{T}_\rho - T_\rho\| \leq \epsilon) \\ &\quad + P(\inf_{g=\phi^T b, b^T\bar{b}=1} \tilde{L}(g) > c - \epsilon, \|\tilde{T}_\rho - T_\rho\| > \epsilon). \end{aligned}$$

When k is large enough, the first term is equal to $P(\|\tilde{T}_\rho - T_\rho\| \leq \epsilon)$ which tends to 1 as $k \to \infty$, and the second term tends to 0.

Proof of Theorem III.5: We will start by showing that $E(\|\hat{f}_{\lambda,\rho}\|^2|\boldsymbol{Z}) = O_p(1)$. First,

$$\|\check{f}_{\lambda,\rho}\|^2_{L^2} = \|(\check{\Omega}^T_\rho + \lambda W)^{-1}\check{\Omega}^T_\rho\check{\beta}\|^2 \leq \|\check{\beta}\|^2 = \|\check{f}\|^2_{L^2} < \|f\|^2_{L^2}.$$

Second, let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ be the functions that return the smallest and largest eigenvalues of a matrix. For any function $g(t) = \phi^T(t)b$, we have $\tilde{L}(g) = b^T(\check{\Omega}_\rho + W)\bar{b}$, then by Lemma B.3, $\lambda_{\min}(\check{\Omega}_\rho + W) = O_p(1)$. Hence, $\lambda_{\min}(\check{\Omega}_\rho + \lambda W) = O_p(\lambda)$, and $\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} = O_p(\lambda^{-1})$. Also, as stated before, the eigenvalues of $\check{\Omega}_\rho$ are the same as those of $\tilde{T}_\rho$, and hence $\check{\Omega}_\rho$ is positive semi-definite. It then follows that

$$\mathrm{E}(\|h_\lambda\|^2_{L^2}|\boldsymbol{Z}) = E\{\boldsymbol{\varepsilon} n^{-1}\boldsymbol{Z}V(\check{\Omega}_\rho + \lambda W)^{-2}n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{\varepsilon}|\boldsymbol{Z}\}$$

$$
\begin{aligned}
&= n^{-1}\sigma_\epsilon^2 \mathrm{tr}\{(\check{\Omega}_\rho + \lambda W)^{-2}\check{\Omega}_\rho\} \\
&\leq n^{-1}\sigma_\epsilon^2 \mathrm{tr}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} \\
&\leq n^{-1}\sigma_\epsilon^2 [\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\} + 2\lambda^{-1}\sum_{j=1}^{K}(2\pi j)^{-2m}] \\
&= O_p(n^{-1}\lambda^{-1}).
\end{aligned}
$$

Thirdly, by assumption that $(\rho + J^{-1}\rho^{-1/(2m)})/\lambda$ is bounded,

$$
\begin{aligned}
\mathrm{E}(\|g_\lambda\|_{L^2}^2|\boldsymbol{Z}) &= E\{\boldsymbol{r}n^{-1}\boldsymbol{Z}V(\check{\Omega}_\rho + \lambda W)^{-2}n^{-1}\bar{V}^T\boldsymbol{Z}^T\boldsymbol{r}|\boldsymbol{Z}\} \\
&\leq n^{-1}\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\}E\{\boldsymbol{r}n^{-1}\boldsymbol{Z}V(\check{\Omega}_\rho + \lambda W)^{-1}\bar{V}\boldsymbol{Z}^T\boldsymbol{r}|\boldsymbol{Z}\} \\
&\leq n^{-1}\lambda_{\max}\{(\check{\Omega}_\rho + \lambda W)^{-1}\}\mathrm{E}(\boldsymbol{r}^T\boldsymbol{r}|\boldsymbol{Z}) \\
&= O_p(\lambda^{-1})\{O_p(\rho) + O_p(J^{-1}\rho^{-1/(2m)})\} = O_p(1).
\end{aligned}
$$

Now, by (B.2), we have $\mathrm{E}(\|\hat{f}_{\lambda,\rho}\|_{L^2}^2|\boldsymbol{Z}) = O_p(1)$, and therefore $\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_{L^2}^2|\boldsymbol{Z}) = O_p(1)$. Finally, notice that

$$
\begin{aligned}
\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_T^2|\boldsymbol{Z}) &= \mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2|\boldsymbol{Z}) + \mathrm{E}\{\langle\hat{f}_{\lambda,\rho} - f, (T - \tilde{T}_\rho)(\hat{f}_{\lambda,\rho} - f)\rangle_{L^2}|\boldsymbol{Z}\} \\
&\leq \mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_{\tilde{T}_\rho}^2|\boldsymbol{Z}) + \|T - \tilde{T}_\rho\|_{\mathcal{H}}\mathrm{E}(\|\hat{f}_{\lambda,\rho} - f\|_{L^2}^2|\boldsymbol{Z}).
\end{aligned}
$$

The result follows from Theorems III.3 and III.4.

VITA

Yehua Li was born in Changchun, China. He received his Bachelor of Science degree in Applied Mathematics in 2000 from Tsinghua University. He was admitted to the graduate program in the Department of Statistics at Texas A&M University in year 2001. He received his Master's degree in August 2003, and Ph.D. in August 2006. His permanent address is:

268- 3 Dongsandao Street, Apartment 507

Nanguan District, Changchun, Jilin 130000

P.R. China