Robust Simultaneous Inference for Functional Data

by

Italo Raony Costa Lima

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2017

Keywords: Functional Data Analysis, Simultaneous Confidence Band, Robust Statistics, Spline Smoothing

Approved by

Guanqun Cao, Chair, Assistant Professor of Mathematics and Statistics
Nedret Billor, Co-chair, Professor of Mathematics and Statistics
Ash Abebe, Professor of Mathematics and Statistics
Xiaoyu Li, Assistant Professor of Mathematics and Statistics

Abstract

Advancements in modern technology have enabled the collection of complex, high-dimensional data sets, such as curves, 2D or 3D images, and other objects living in a functional space, thus boosting the investigation of function data. This phenomenon affects all the fields involving applied statistics such as Geophysics (Ferraty et al., 2005), Environmetrics (Febrero et al., 2008), Ecology (Embling et al., 2012), Chemometrics (Daszykowski et al., 2007), and others, see also Ferraty and Vieu (2006). Not only revealing the wide variety of possible fields of application of functional data methods, the previous selected set of references also reflect the large scope of statistical problems, such as regression and inference, in which one may have to deal with under the context of functional data. Hence, functional data analysis (FDA) has become one of the most active fields of research in statistics during the last ten years.

In this dissertation our objective is to develop outlying-resistant methods for the estimation of the mean curve of a functional dataset that remain valid even in the presence of a significant proportion of outlier curves. We also propose a method for the calculation of a simultaneous confidence band for the mean curve that is robust to outliers. Our work is based on B-Spline Smoothing, together with LAD-based and M-based estimation techniques. We also extend both methods for the estimation of the difference of the mean functions of two populations, also obtaining a robust test statistics for the difference of the mean functions of two populations. We analyze the asymptotic properties of the proposed estimators, proving weak consistency and, for the M-based estimator, asymptotic normality. We implement an

extensive numerical simulation to evaluate the performance of the proposed methods. We also demonstrate the applicability of the proposed methods using real datasets.

Acknowledgments

Completing this dissertation was not a solitary endeavor, much less an achievement entirely of my own. I take this opportunity to thank everyone who directly or indirectly has made this possible.

I will be forever grateful for the guidance of my advisors, Dr. Guanqun Cao and Dr. Nedret Billor. Your patience and wise words of motivation helped me progress with my research. Watching you work tirelessly, being always there to help me, and all your other students, and always having insightful suggestions has helped grow as a student and researcher. You will always have my deepest gratitude.

To the members of my committee, Dr. Ash Abebe, Dr. Xiaoyu Li and Dr. Mary J. Sandage, thank you! Your comments and advice are an integral part of this work.

To all my Auburn friends, this work is the culmination of all our team projects, classes we had together, and our insightful discussions. Thank you for making my time here so delightful. To all my Brazilian friends, you will always have a special place in my heart. Cicconet, thank you! Prof. Geraldo, without your help, none of this would have been possible. To my friends from the *Projeto Formação do Talento Matemático Sergipano*, I would have traded our studying weekends for nothing in this world. Pietro, thank you! Professor Valdenberg, words will never express my gratitude for you. If I am here today, I owe it to you. To my parents, *Raimunda* and *Élio, muito obrigado pelo suporte incondicional e por me fazer quem eu sou hoje,* and to my sisters, Eline and Mariane. To my beautiful wife, Carolina, thank you, thank you and thank you. Without you, I wouldn't be me.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main topic of this dissertation is *Functional Data Analysis* (FDA), a recently developed area of statistics, that has gained increasing interest in the latest years. FDA considers *smooth functions* as fundamental atoms of data, markedly in the form of functional curves. Unlike multivariate statistics where the data live in a finite-dimensional space, usually $\mathbb{R}^n$, in FDA the data live in spaces of infinite dimension, usually some subset of L^2, the space of square integrable functions. The literature on FDA is vast, but the books of Silverman and Ramsay (2005) and Ferraty and Vieu (2006) provide a thorough overview of the basics of the area.

In this dissertation we study *Robust Statistical Methods*. Basic statistical methods require assumptions on the data, such as distributional assumptions, e.g. normality of data, and homogeneity of data. When the assumptions are not satisfied, the statistical methods usually deteriorate, generating misleading results, such as wrong estimates or overestimated variance. The assumptions of a model might be broken due to the intrinsic distribution of the data, or due to an external contamination, that is, due to the presence of *outliers*.

Outliers can appear in a dataset in various ways. They might be the result of an error, for example an incorrect measurement, or they can be correct observations that do not follow the pattern of the majority of the data. In the former case, identification and removal of wrong data are preferred. In the latter case, the removal of outliers is not recommended, since it is a valid data point. Rather, a robust method should be used for the analysis of such data. Robust statistics has a long history, going

as far back as the work of Fisher (1922) where the sample mean was studied under small deviations from normality. The seminal works of Hampel et al. (2011) and Huber and Ronchetti (2011) established much of the basic theory for modern robust methods.

The results presented in this dissertation are on the intersection of *robust statistical methods* and *functional data analysis*. Robust statistical methods in functional data analysis are challenging due to the rich outlier structure in functional spaces. The inherent complexity of infinite dimensional functional spaces allows for a wide variety of possible outlier behavior. Among the recent literature, Gervini (2008) proposed a robust estimator for the location parameter of contaminated datasets, by extending the notion of median to functional datasets, and also proposed a robust alternative for the functional principal components analysis (FPCA) based on the spherical principal components defined in Locantore et al. (1999). More recently, Bali et al. (2011) and Lee et al. (2013) proposed robust estimators for the functional principal components by adapting the projection pursuit approach and based on MM estimation, respectively. Kraus and Panaretos (2012) uses a different approach, instead estimating the dispersion operator containing influential observations. More recently, Maronna and Yohai (2013) established a robust version of spline-based estimators, replacing the L^2 loss by a bounded loss function of the residuals in functional linear regression models. Shin and Lee (2016) proposed a robust procedure which uses outlier-resistant loss functions including non-convex loss functions in functional linear regression models, computationally based on iteratively reweighted penalized least-squares algorithm and study the theoretical developments of this estimator by using numerical studies with various types of robust loss.

In this dissertation we obtain robust methods for the simultaneous inference for the mean function of functional datasets. The problem of simultaneous inference for

functional data has not been studied extensively not due to the lack of interest of the research community, but rather due to its challenging technical nature. The focus of this dissertation is on developing robust estimation of simultaneous confidence bands for the mean function of functional data, that is, confidence bands that are valid uniformly for all the points of the mean function when the data are contaminated with outliers. Among the recent literature, Degras (2011) and Cao et al. (2012b) calculated simultaneous confidence bands for the mean function using local kernels and B-Spline basis, respectively, when data are homogeneous. Some studies also discussed robust confidence intervals for the location parameter in the finite-dimensional setting. For instance, Fraiman et al. (2001) discussed globally robust confidence intervals, where the confidence level is maintained in a neighborhood of the base distribution and Adrover et al. (2004) defined globally robust confidence intervals and p-values for the location and simple linear regression models, and Haque and Khan (2012) discussed globally robust confidence intervals by taking into account the bias direction of the robust location estimators.

Two methods are presented to obtain a robust estimator and robust simultaneous confidence band for the mean function of functional data. In Chapter 2, we present an estimation of the mean function by using *Least Absolute Deviation* and B-Spline smoothing. Following, we propose a robust estimation of the covariance function by using the spherical principal components for estimating the eigenfunction decomposition of the covariance operator, aided by a robust estimation of the eigenvalues, leading to a robust reconstruction of the covariance function. The robust simultaneous confidence band is then calculated by using a Monte-Carlo procedure to estimate the maximal quantile of appropriately defined Gaussian processes. The same methodology is extended for the robust estimation of the difference of the mean functions of two populations, and for the definition of a robust statistic for testing

the hypothesis of the difference of the mean function of two populations. In Chapter 3, we present an extension of the robust simultaneous inference in Chapter 2. We propose the B-Spline Smoothed M-Estimator for the mean function of functional data. For this estimator, we research the asymptotic theoretical properties, and we present proofs for the asymptotic consistency and the asymptotic normality of the estimator. Using a modified version of the pseudo-data introduced in Cox (1983), we obtain a robust simultaneous confidence band for the mean function. We also propose a robust test for the difference of the mean functions of two populations, based on the robust simultaneous confidence band of the difference of two mean functions. In Chapter 4, we review and make a comparison of the results of the dissertation, and propose some future research problems.

1.1 Some Motivating Examples

We present here two real datasets that have a natural interpretation as functional datasets. These datasets also contain outliers. These examples will serve as a motivation for the application of the methods presented in this dissertation, and are further analyzed in the Sections 2.5 and 3.5.

1.1.1 Octane dataset

This dataset consists of 39 near infrared (NIR) spectra of gasoline samples, obtained from Esbensen et al. (1996). The absorbance rate of each gasoline sample was measured for every other wavelength between $1100hz$ and $1550hz$. It is known, from the data collection phase, that 6 of the samples contain added ethanol, which corresponds to an upward translation of the absorbance rate for the upper wavelength interval of the spectrum, $1390hz$ onward.

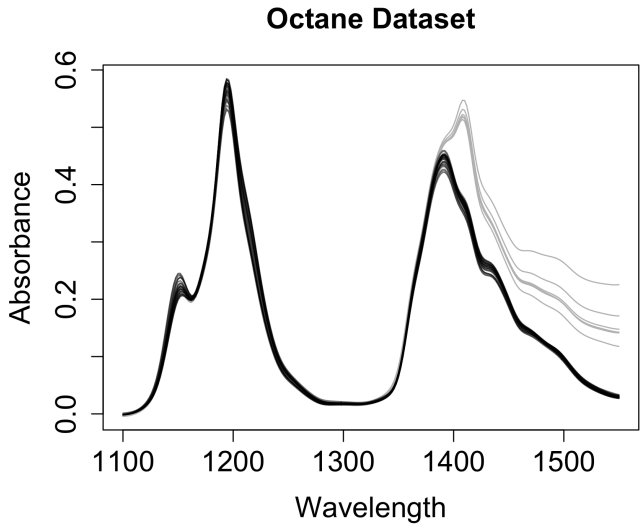

Figure 1.1: Absorbance rate of gasoline for near infrared spectra.

For this dataset, the outlying curves are clear, both visually, see Figure 1.1, and from the description of the data collection. A question of interest about this dataset is to identify the average absorbance rate as a function of the wavelenght in the presence of the outliers. An answer to this question will be given in Sections 2.5 and 3.5.

1.1.2 Ground level Ozone concentration dataset

This dataset consists of hourly average measurements of ground level ozone (O_3) concentrations from a monitoring station in Richmond, BC, Canada, from the years of 2004 to 2012. The presence of Ozone at ground level is highly undesirable, and considered a serious air pollutant. Since the concentration of ground level Ozone typically peaks at summer months, the month of August will usually present a high concentration level. Figure 1.2 represents the daily ground level ozone concentration for the month of August of the years 2005 and 2007.

The same dataset was studied in Boente and Salibian-Barrera (2015), and using S-estimators for the principal components, the presence of outliers was detected in the year of 2005. Based on the existence of outliers, one question of interest is if there is a significant difference between the mean functions of the ground level ozone concentration between the years of 2005 and 2007, in spite of the existence of outliers. An answer to this question will be presented in the Section 2.5.

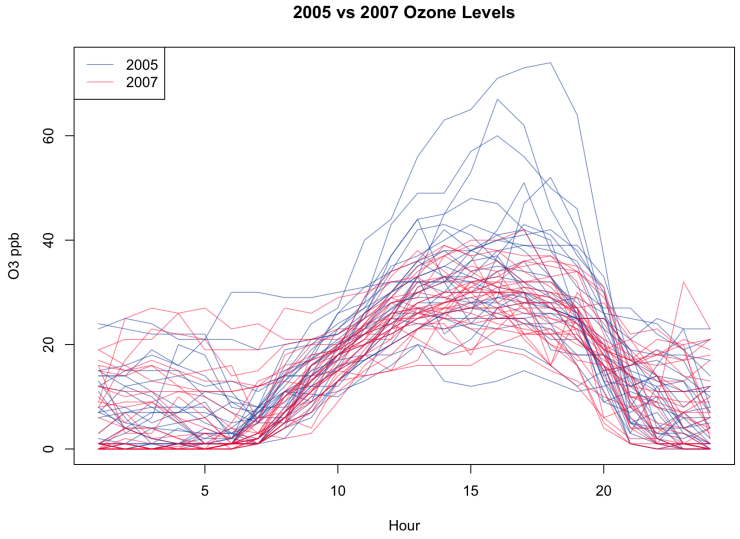

Figure 1.2: Ground Level Ozone Concentration for the month of August for 2005 and 2007.

1.2 Introduction to Functional Data Analysis

In this section we will present some basic definitions for FDA and present some examples of analyses performed on functional datasets.

The recent developments of data management apparatus, including collection, storage and analysis capabilities, made possible the collection of complex, high-dimensional datasets. Examples include the collection of smooth curves, high-definition

images and high-fidelity 3D models, such as datasets originating from fMRI, functional Magnetic Resonance Imaging. These developments have boosted the study of functional data analysis. This phenomenon affects all the fields involving applied statistics such as Geophysics (Ferraty et al., 2005), Environmetrics (Febrero et al., 2008), Ecology (Embling et al., 2012), Chemometrics (Daszykowski et al., 2007), and others, see also Ferraty and Vieu (2006). Not only revealing the wide variety of possible fields of application of functional data methods, the previous selected set of references also reflect the large scope of statistical problems, such as regression and inference, in which one may have to deal with under the context of functional data. Hence, FDA has become one of the most active fields of research in statistics during the last ten years.

The development of FDA was also helped by the seminal book of Silverman and Ramsay (2005), which laid the foundations of functional data analysis, and also presented insightful suggestions for future research. It is still considered the handbook of functional data analysis. The basic definitions presented following are mostly borrowed from that book.

1.2.1 Basic Definitions

The basic data considered in FDA come from experiments in which the measurements can be interpreted as a smooth function. One such example comes from the Berkeley Growth Study (Tuddenham and Snyder, 1954). In this study, the researchers collected the heights of boys and girls at irregular intervals from the birth to eighteen years of age. Although only a discrete set of measurements is presented, the height is best represented as a smooth function observed at discrete points. The assumption of the observations being smooth functions comes with several benefits,

such as the possibility of using the information of derivatives, giving rise to rich measures, such as the acceleration of growth of the kids.

The same assumption of smooth curves also comes with setbacks. For example, *how should the discrete measurement points be interpolated to a smooth function?* Basis representation of functional spaces provide a partial answer to this question. Some examples of such basis expansions are Fourier, B-Splines, and Wavelets.

A formal attempt at defining objects of study of FDA can be made with the help of *stochastic processes*. We consider a *functional dataset* as a collection of i.i.d. random samples, $\{\eta_i(x)\}_{i=1}^n$, from a smooth and square integrable random function $\eta(x) \in L^2$. As mentioned in the previous paragraph, often times the whole curve is not available, but rather a discrete set of measurements for each sample curve, that is, each random curve $\eta_i(\cdot)$ is measured at the points x_{ij}, $1 \leq j \leq N$, $1 \leq i \leq n$, where N goes to infinity together with the sample size n. Then, the j-th observation for the i-th subject can be written as

$$Y_{ij} = \eta_i\left(x_{ij}\right) + \sigma\left(x_{ij}\right)\varepsilon_{ij}, \tag{1.1}$$

where ε_{ij} are measurement errors, and usually assumed to be independent, satisfying $\mathbb{E}(\varepsilon_{ij}) = 0$ and $\mathbb{E}(\varepsilon_{ij}^2) = 1$, and $\sigma(\cdot)$ is the variance of the measurement errors, assumed to be a smooth function.

In order to obtain the full sample from a discrete set of measurements, an approximation basis approach can be used. One example of such approach is the B-Spline smoothing method. In order to illustrate this method, assume that the random curves are defined in a compact interval, assumed here to be $[0, 1]$. Let $t_{1-p} = \cdots = t_0 = 0 < t_1 < \cdots < t_{N_m} < 1 = t_{N_m+1} = \cdots = t_{N_m+p}$ be a set of N_m points on $[0, 1]$, called interior knots. The number of interior knots will control

the amount variation allowed in the interpolation, and will depend on the number of measurement points, N, and the number of sample curves n. Denote by $\mathcal{H}^{(p-2)}$ the p-th order spline space, i.e., $p-2$ times continuously differentiable functions on $[0,1]$ that are polynomials of degree $p-1$ on $[t_J, t_{J+1}]$, $J = 0, \ldots, N_m$. The B-Spline basis of $\mathcal{H}^{(p-2)}$ was defined in De Boor (1978), and consists of an orthogonal basis, containing $N_m + p$ functions $B_{1-p}(\cdot), \cdots, B_{N_m}(\cdot)$. Using the B-Spline basis, we can approximate the whole smooth sample function as

$$\widehat{\eta}_i(x) = \sum_{J=1-p}^{N_m} \hat{\beta}_{i,J} B_J(x),\ i = 1, \ldots, n,$$

where $\{\hat{\beta}_{i,1}, \cdots, \hat{\beta}_{i,N_m+p}\}^{\mathrm{T}}, i = 1, \ldots, n$ is the solution of

$$\left\{\hat{\beta}_{i,1}, \cdots, \hat{\beta}_{i,N_m+p}\right\}^{\mathrm{T}} = \underset{\{\beta_1,\ldots,\beta_{N_m+p}\}\in R^{N_m+p}}{\arg\min} \sum_{j=1}^{N} \left\{ Y_{ij} - \sum_{J=1-p}^{N_m} \beta_J B_J\left(x_{ij}\right) \right\}^2, 1 \le i \le n.$$

Another problem in FDA concerns of the estimation of the covariance function. The covariance function is defined as

$$G(x, x') = \mathbb{E}\left((\eta(x) - m(x))(\eta(x') - m(x'))\right), 0 \le x, x' \le 1,$$

where $m(x) = \mathbb{E}(\eta(x))$, $0 \le x \le 1$. Using the Karhunen-Loève L^2 representation theorem, we can write

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

where ξ_k, $k = 1, 2, \ldots$, are uncorrelated random coefficients with mean zero and variance one and $\phi_k(x) = \sqrt{\lambda_k}\varphi_k(x)$. Also, $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\varphi_k(x)\}_{k=1}^{\infty}$ are the eigenvalues

and the eigenfunctions, respectively, of the covariance function $G(x, x')$ given by

$$G(x, x') = \sum_{k=1}^{\kappa} \lambda_k \cdot \varphi_k(x)\varphi_k(x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k(x').$$

This suggests that the covariance function can be estimated using an extension of the principal components analysis to functional data, the *Functional Principal Component Analysis*.

A presentation of further problems in FDA can be found in Silverman and Ramsay (2005).

Chapter 2

LAD-Based Robust Simultaneous Inference for Functional Data

2.1 Introduction

In this chapter, our objective is to develop outlying-resistant methods that can provide valid statistical inference even in the presence of a significant proportion of outlier curves. In particular, outlier-resistant simultaneous inference for the location function will be studied. Experimenters may have outlier curves in functional data for two reasons. First, they may result in errors in measurements and recording or typing mistakes. These should be identified and corrected if possible, or discarded if not corrected. Second, outlier curves can be correctly observed data curves which are different in the sense that they do not follow the same pattern as that of the majority of the curves. In the presence of outlier curves, the estimate of the functional mean, and therefore simultaneous confidence band (SCB), may be affected badly which would yield misleading statistical conclusions. Further, the presence of outliers is amplified by the inherent complexity of functional spaces. This may lead to arise different types of outlier curves which add further complication to estimating functional mean and constructing SCB for functional mean. It is of particular interest in data analysis to construct SCB for the mean function instead of point-wise confidence intervals and to develop global test statistics for the general hypothesis testing problem on the location functions. For example, Degras (2011), Cao et al. (2012a,b) construct asymptotic SCB for the mean and derivative functions in FDA without considering any outliers. However, all the existing literatures on constructing SCB

for the mean function assume that observed functional data are homogeneous, that is free of outlier curves.

Developing the robust (R) simultaneous inference for functional responses, we encountered many new challenges. First, the greater technical difficulty is to formulate SCB for a mean function and establish their theoretical properties. Second, unlike the scenarios considered in the classical FDA literature, in our settings there is complex outlier structure. Our main contribution is to construct a R simultaneous confidence band (RSCB) based on R estimators of the mean function and covariance function using the least absolute deviation and spline smoothing methods and spherical principal components, respectively. We further extend the simultaneous inference to the two-sample case and evaluate the equality of mean functions from two groups when atypical curves exist. Our Monte Carlo simulation results show that the proposed bands are superior to existing classical methods which do not account for atypical curves. To our best knowledge, this is the first attempt to provide SCB for functional data that are less sensitive to anomalous observations. We note that all currently available methods cannot be immediately used for constructing a R version of SCB for mean functions.

This chapter is organized as follows. We first propose a R estimator for the mean function when the dataset contain outliers, using a proposed robust B-spline smoothing estimation method in the Section 2.2. Then we construct a RSCB for the mean function based on the proposed R estimator for the mean function, and a R covariance estimator obtained from the spherical principal components by using polynomial spline smoothing estimation in the Section 2.3. In addition we extend this method to form a RSCB for the difference of mean functions of two populations in the same section. In the Sections 2.4 and 2.5, the performance of the proposed R methods

and their robustness are demonstrated with an extensive simulation study and real data examples. Finally we conclude the chapter with discussion and conclusion.

2.2 Model

A functional dataset can be defined as a collection of i.i.d. random samples, $\{\eta_i(x)\}_{i=1}^n$, where i is the subject index, from a smooth and square integrable random function $\eta(x) \in L^2$, with unknown mean function, $\mathbb{E}[\eta(x)] = m(x)$, and unknown covariance function $G(x, x') = \text{cov}[\eta(x), \eta(x')]$. For simplicity reasons, the domain of $\eta(\cdot)$ is assumed as $[0, 1]$. In this paper we assume an equally spaced dense design, that is, each random curve $\eta_i(\cdot)$ is measured at the points $x_{ij} = j/N$, $1 \leq j \leq N$, $1 \leq i \leq n$, where N goes to infinity when sample size n goes to infinity. Then, the jth observation for the ith subject can be written as

$$Y_{ij} = \eta_i\left(j/N\right) + \sigma\left(j/N\right)\varepsilon_{ij}, \tag{2.1}$$

where errors ε_{ij}'s are independent and assumed to satisfy $\mathbb{E}(\varepsilon_{ij}) = 0$ and $\mathbb{E}(\varepsilon_{ij}^2) = 1$, and the stochastic process $\mathbb{E}\int_{[0,1]}\eta^2(x)dx < \infty$.

The process $\eta(x)$ can be written, based on Karhunen-Loève L^2 representation, as $\eta(x) = m(x) + \sum_{k=1}^{\infty}\xi_k\phi_k(x)$, where ξ_k's are uncorrelated random coefficients with mean zero and variance one, which we assume to have a symmetrical elliptical distribution, and $\phi_k(x) = \sqrt{\lambda_k}\varphi_k(x)$. Here, $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\varphi_k(x)\}_{k=1}^{\infty}$ are the eigenvalues and eigenfunctions, which form an orthonormal basis of L^2, of the covariance function $G(x, x')$. We assume that $\phi_k(\cdot)$'s are kept in a descending order of λ_k's, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. Assume that $\lambda_k = 0$ for $k > \kappa$, where κ is a positive integer or

∞. This implies that the eigenvalue decomposition of $G(x, x')$ is

$$G(x, x') = \sum_{k=1}^{\kappa} \lambda_k \varphi_k(x) \varphi_k(x') = \sum_{k=1}^{\kappa} \phi_k(x) \phi_k(x'). \tag{2.2}$$

With this representation, we rewrite the model (2.1) as

$$Y_{ij} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k\left(\frac{j}{N}\right) + \sigma\left(\frac{j}{N}\right) \varepsilon_{ij}. \tag{2.3}$$

Although the existence of $\{\lambda_k\}_{k=1}^{\kappa}$ and $\{\phi_k\}_{k=1}^{\kappa}$, and the random coefficients $\{\xi_{ik}\}_{k=1}^{\kappa}$ are guaranteed mathematically, they are unknown and unobservable.

2.3 Proposed Method

2.3.1 The classical estimation of the mean function

We first review an estimation procedure that approximates the mean function by polynomial splines. Let $t_{1-p} = \cdots = t_0 = 0 < t_1 < \cdots < t_{N_m} < 1 = t_{N_m+1} = \cdots = t_{N_m+p}$ be equally-spaced points over $[0, 1]$, called interior knots, in which $t_J = Jh_m$, $0 \leq J \leq N_m$, and $h_m = 1/(N_m + 1)$ is the distance between neighboring knots. Denote by $\mathcal{H}^{(p-2)}$ the p-th order spline space, i.e., $p-2$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $p-1$ on $[t_J, t_{J+1}]$, $J = 0, \ldots, N_m$.

Cao et al. (2012b) proposed to approximate the mean function $m(\cdot)$ by a linear combination of spline basis: $\hat{m}(x) = \sum_{J=1-p}^{N_m} \hat{\beta}_J B_J(x)$, where B_J be the J-th B-spline basis of order p defined in de Boor (2001), and the coefficients $\{\hat{\beta}_{1-p}, \ldots, \hat{\beta}_{N_m}\}^{\mathrm{T}}$ are

the solutions of the following least squares problem

$$\left\{\hat{\beta}_{1-p},\ldots,\hat{\beta}_{N_m}\right\}^{\mathrm{T}} = \underset{\{\beta_{1-p},\ldots,\beta_{N_m}\}\in R^{N_m+p}}{\arg\min} \sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - \sum_{J=1-p}^{N_m}\beta_J B_J\left(j/N\right)\right\}^2.$$

Let $\bar{\mathbf{Y}} = (\bar{Y}_{\cdot 1},\ldots,\bar{Y}_{\cdot N})^{\mathrm{T}}$, where $\bar{Y}_{\cdot j} = 1/n\sum_{i=1} Y_{ij}$, $j = 1,\ldots,N$. One advantage of the least square estimator is the existence of a closed form solution. Indeed, applying elementary algebra, one obtains

$$\hat{m}\left(x\right) = \boldsymbol{B}(x)\left(\mathbb{B}^{\mathrm{T}}\mathbb{B}\right)^{-1}\mathbb{B}^{\mathrm{T}}\bar{\mathbf{Y}}, \tag{2.4}$$

where $\boldsymbol{B}(x) = (B_{1-p,p}(x),\ldots,B_{N_m,p}(x))$ and $\mathbb{B} = (\boldsymbol{B}^{\mathrm{T}}(1/N),\ldots,\boldsymbol{B}^{\mathrm{T}}(N/N))^{\mathrm{T}}$ is the design matrix.

2.3.2 The robust estimation of the mean function

The ordinary least square (OLS) estimation, used in equation (2.4), is susceptible to the presence of outliers. To circumvent this non-robustness, we propose to replace the OLS by the *least absolute deviation* (LAD), that is, we estimate the mean function by a linear combination of spline basis,

$$\tilde{m}(x) = \sum_{J=1-p}^{N_m}\tilde{\beta}_J B_J(x), \tag{2.5}$$

where

$$\left\{\tilde{\beta}_{1-p},\ldots,\tilde{\beta}_{N_m}\right\}^{\mathrm{T}} = \underset{\{\beta_{1-p},\ldots,\beta_{N_m}\}\in R^{N_m+p}}{\arg\min} \sum_{i=1}^{n}\sum_{j=1}^{N}\left|Y_{ij} - \sum_{J=1-p}^{N_m}\beta_J B_J\left(j/N\right)\right|.$$

LAD gives equal emphasis to all observations, in contrast to OLS which, by squaring the residuals, gives more weight to large residuals, that is, outliers in which predicted values are far from actual observations. This is helpful in studies where outliers do not need to be given greater weight than other observations. In this work, we use cubic spline ($p = 4$) basis and the number of interior knots N_m is taken to be $[0.5n^{1/8} \log n]$ and $[a]$ denotes the integer part of a, which is recommended in Cao et al. (2012b). In the following theorem, we show that the proposed R estimator, $\tilde{m}(x)$, is a consistent estimator of the true mean function, $m(x)$, converging at the optimal rate $n^{-1/2}$.

Theorem 1. *Under Assumptions (A1)-(A4) bellow, one has*

$$\|\tilde{m} - m\|_2^2 = o_P(n^{-1}).$$

To give the proof of the Theorem 1, we first need some additional definitions. For any $r \in (0, 1]$, we denote $C^{q,r}[0, 1]$ as the space of Hölder continuous functions on $[0, 1]$, $C^{q,r}[0, 1] = \left\{\phi : \sup_{t \neq s, t, s \in [0,1]} \frac{\left|\phi^{(q)}(t) - \phi^{(q)}(s)\right|}{|t-s|^r} < +\infty\right\}$. Let $f_{ij}(0)$ be the density function of e_{ij}, where $e_{ij} = \sum_{k=1}^{\kappa} \xi_{ik}\phi_k(j/N) + \sigma(j/N)\varepsilon_{ij}$, $1 \leq i \leq n$, $1 \leq j \leq N$. The following technical assumptions are needed.

(A1) The regression function $m \in C^{p_0-1,1}[0, 1]$. The spline order in estimating m satisfies $p \geq p_0$.

(A2) *The standard deviation function* $\sigma(x) \in C^{0,\delta}[0, 1]$ *for some* $\delta \in (0, 1]$ *and for any* $k = 1, 2, \ldots \kappa$, $\phi_k(x) \in C^{p,1}[0, 1]$ *and* $\min_{x \in [0,1]} G(x, x) > 0$;

(A3) The number of knots N_m satisfies $n^{1/(2p)} \ll N_m \ll Nn^{-1}$.

(A4) Uniformly over i and j, $f_{ij}(\cdot)$ is bounded from infinity, and it is bounded away from zero and has a bounded first derivative in the neighborhood of zero.

Assumptions (A1)-(A3) are standard in the spline smoothing and FDA literature; see Cao et al. (2012b), for instance. In particular, (A1) guarantee the orders of the bias terms of the spline smoothers for $m(x)$. Assumption (A2) ensures the covariance function is a uniformly bounded function. Assumption (A3) implies the number of points on each curve N diverges to infinity as $n \to \infty$, which is a well-developed asymptotic scenario for dense functional data. The smoothness of our estimator is controlled by the number of knots, which increases to infinity as specified in (A3). This increasing knots asymptotic framework guarantees the richness of the basis. (A4) is a standard assumption in quantile regression; see Theorem 1 in Wang et al. (2009) for example.

2.3.3 Proof of Theorem 1

Let $\mathbf{K}_i = \text{diag}\{f_{i1}(0), \ldots, f_{iN}(0)\}$, $i = 1, 2, \ldots, n$ and $\mathbf{K} = \text{diag}\{\mathbf{K}_1, \ldots, \mathbf{K}_n\}$. Let $\varsigma = \varsigma(\boldsymbol{\beta}) = N_m^{-1/2}\mathbf{H}_n(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, where $\mathbf{H}_n^2 = N_m\boldsymbol{\Pi}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Pi}$, $\boldsymbol{\Pi} = (\mathbf{B}_1^{\mathrm{T}}, \ldots, \mathbf{B}_n^{\mathrm{T}})$ and $\mathbf{B}_i \equiv \mathbf{B}$, $i = 1, \ldots, n$. We first show that $\|\varsigma(\tilde{\boldsymbol{\beta}})\|_2 = O_p(N_m^{1/2})$. We standardize $\tilde{b}_j = N_m^{1/2}\mathbf{H}_n^{-1}\mathbf{B}_p(j/N)$. Denote $R_{nj} = \mathbf{B}_p(j/N)\boldsymbol{\beta}_0 - m(j/N)$ as the bias from the spline approximation. Thus, we have $\sum_{i=1}^n \sum_{j=1}^N |Y_{ij} - \mathbf{B}_p(j/N)\boldsymbol{\beta}| = \sum_{i=1}^n \sum_{j=1}^N |e_{ij} - \mathbf{B}_p(j/N)\boldsymbol{\beta} - \varsigma^{\mathrm{T}}\tilde{b}_j - R_{nj}|$. By Lemma A.5 of Kim (2007), $\max_j \|\tilde{b}_j\|_2 = O(\sqrt{N_m/n})$. Applying similar arguments used in Theorem 3.1 of Wei and He (2006), for any $\epsilon > 0$, there exits L_ϵ such that $P\big\{ \inf_{\|\varsigma\|_2 > L_\epsilon N_m^{1/2}} \sum_{i=1}^n \sum_{j=1}^N |e_{ij} - \mathbf{B}_p(j/N)\boldsymbol{\beta} - \varsigma^{\mathrm{T}}\tilde{b}_j - R_{nj}| > \sum_{i=1}^n \sum_{j=1}^N |e_{ij} - \mathbf{B}_p(j/N)\boldsymbol{\beta} - R_{nj}| \big\} > 1 - \epsilon$. Since $\sum_{i=1}^n \sum_{j=1}^N |e_{ij} - \mathbf{B}_p(j/N)\boldsymbol{\beta} - \varsigma^{\mathrm{T}}\tilde{b}_j - R_{nj}|$ is minimized at $\varsigma(\tilde{\boldsymbol{\beta}})$ over the space R^{N_m+p}, we have $P(\|\varsigma(\tilde{\boldsymbol{\beta}})\|_2 < L_\epsilon N_m^{1/2}) > 1 - \epsilon$, and $\|\varsigma(\tilde{\boldsymbol{\beta}})\|_2 = O_p(N_m^{1/2})$. According to Lemma A.4 of Cao et al. (2012b), there exists a constant $C > 0$, such that $\|\widetilde{m} - m\|_\infty \leq C \|m\|_\infty J_m^{-p_m}$, where $\|m\|_\infty =$

$\sup_{x\in[0,1]}|m(x)|$. There exists a constant C, such that

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\{\tilde{m}(j/N)-m(j/N)\}^2 &\leq \frac{2}{N}\sum_{j=1}^{N}\{\mathbf{B}_p(j/N)(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}^2+2CN_m^{-2p}\\ &= 2N^{-1}\|\varsigma(\tilde{\boldsymbol{\beta}})\|_2^2+2CN_m^{-2p}=o_p(n^{-1}).\end{aligned}$$

The proof of Theorem 1 is complete.

2.3.4 The robust estimator for the covariance function

In order to construct a R confidence band for the mean function, first we need to obtain a R estimator for the covariance function, $G(x,x')$. The covariance function defined in (2.2) can be recovered if the eigenfunction/eigenvalue decomposition is known. A popular estimator for the eigenfunctions and eigenvalues is derived from eigenvalue decomposition of the empirical or smoothed covariance matrix. Such method has been widely used in functional data analysis (Yao et al., 2005; Cao et al., 2012b). However, this estimator is sensitive to the presence of outliers.

An alternative R estimator, the spherical principal components, is proposed in Locantore et al. (1999). The first step in this method is to normalize each sample curve to mitigate the effect of outliers. That is, the covariance function is replaced by the *normalized covariance function*, i.e.,

$$\rho(x,x')=\mathbb{E}\left\{\frac{[\eta(x)-m(x)][\eta(x')-m(x')]}{\|\eta(\cdot)-m(\cdot)\|_2^2}\right\},$$

where $\|\cdot\|_2$ is the usual L^2-norm. We can then find the eigenfunction/eigenvalue decomposition of this new operator, leading to eigenfunctions φ_k^* and eigenvalues λ_k^*. When $\kappa<\infty$ and $\{\xi_k\}_{k=1}^{\kappa}$ has a symmetric and interchangeable marginal distribution, Gervini (2008) has shown that $\varphi_k^*=\varphi_k$, that is, the eigenfunctions of the

covariance operator are the same as the eigenfunctions of normalized one, and they are in the same order. However the corresponding two types of eigenvalues are not necessarily be the same. These R estimators are generally more robust to outliers than the commonly used sample mean and principal components. Moreover, Boente et al. (2014) extend the above results to the case of $\kappa = \infty$.

Since we do not have the entire stochastic process and the true mean function, we use the discretized sample version of the normalized covariance function,

$$\rho_n(j/N, j'/N) = \frac{1}{n}\sum_{i=1}^{n} \frac{(Y_{ij} - \tilde{m}(j/N))(Y_{ij'} - \tilde{m}(j'/N))}{\|Y_{i\cdot} - \tilde{m}(\cdot)\|_2^2}, 1 \leq j, j' \leq N, \qquad (2.6)$$

where $\|Y_{i\cdot} - \tilde{m}(\cdot)\|_2^2 = 1/N \sum_{j=1}^{N} (Y_{ij} - \tilde{m}(j/N))^2$ and $\tilde{m}(\cdot)$ is the proposed R estimator of the mean function, defined in (2.5). Similarly to the population version, we can find the eigenvector decomposition $\{\varphi_k^*(j/N)\}_{k=1}^{\kappa}$ for the sample version of the normalized covariance function and the eigenfunctions, φ_k^* are called *spherical principal components* (Locantore et al., 1999) since ρ_n is the sample covariance function of the centred curves projected on the unit sphere.

The presence of measurement errors in our model adds a layer of contamination to the calculation of (2.6), but due to the aforementioned robustness of the spherical principal components, if the measurement errors are not excessive, φ_k^* will be close to the true φ_k. In order to reduce the effects of measurement errors further, we smooth each φ_k^* using the B-Splines which yield the smoothed eigenfunctions $\tilde{\varphi}_k(\cdot)$.

To estimate the covariance function in (2.2), we need to recover the eigenvalues λ_k. Notice that the *k-th* eigenvalue is the variance of the projection of the centralized stochastic process on the *k-th* eigenfunction, that is $\lambda_k = \mathrm{Var}\left(\langle \eta(\cdot) - m(\cdot), \varphi_k(\cdot)\rangle\right)$. Adapting this to our model, we can use the proposed R estimator $\tilde{m}$ and the R estimator of the eigenfunctions $\tilde{\varphi}_k$ to obtain $\tilde{\lambda}_k$ as the square of the R estimator of the scale

of $\langle Y_{i\cdot} - \tilde{m}(\cdot), \tilde{\varphi}_k(\cdot)\rangle$. Having estimated the eigenvalues and eigenfunctions, we can recover a R estimator of the covariance function as $\tilde{G}(x, x') = \sum_{k=1}^{\kappa} \tilde{\lambda}_k \tilde{\varphi}_k(x)\tilde{\varphi}_k(x')$. The R estimation of the covariance function can be summarized in the following steps.

Step 1. *Estimate eigenfunction* $\tilde{\varphi}_k(\cdot)$.

Apply the B-spline smoothing $\tilde{\varphi}_k(x) = \boldsymbol{B}(x)\left(\mathbb{B}^{\mathrm{T}}\mathbb{B}\right)^{-1}\mathbb{B}^{\mathrm{T}}\boldsymbol{\varphi}_k^*$, where $\boldsymbol{\varphi}_k^* = (\varphi_k^*(1/N), \ldots, \varphi_k^*(N/N))^{\mathrm{T}}$, $k = 1, 2, \ldots, N$, are the eigenvectors of the sample normalized covariance matrix ρ_n in (2.6).

Step 2. *Estimate eigenvalue* $\tilde{\lambda}_k$.

Using a R scale estimator of the projection of the centralized data onto $\tilde{\varphi}_k$. The chosen R estimator is the Huber's M-estimator (Huber and Ronchetti, 2011), i.e., the solution of the equation

$$H(\sigma) = \frac{1}{n}\sum_{i=1}^{n} \rho_\gamma \left(\frac{\frac{1}{N}\sum_{j=1}^{N}(Y_{ij} - \tilde{m}(j/N))\tilde{\varphi}_k(j/N)}{\sigma} \right) = 0.5,$$

where $\rho_\gamma(u) = \rho(u/\gamma), \gamma > 0$, and ρ is the Tukey's bisquare function. We use $\gamma = 1.5$ in the following numerical studies, which is the suggested value in Huber and Ronchetti (2011). This leads to the estimators $\tilde{\lambda}_k^{1/2}$. Squaring these values results in the estimator $\tilde{\lambda}_k$ of the eigenvalues.

Step 3. *Recover the R covariance estimator* $\tilde{G}(x, x')$.

Define $\tilde{G}(x, x') = \sum_{k=1}^{N} \tilde{\lambda}_k \tilde{\varphi}_k(x)\tilde{\varphi}_k(x')$, $x, x' \in [0, 1]$.

2.3.5 The RSCB for the mean function

In the following, we mimic the construction of SCB procedure in Cao et al. (2012b) to obtain the RSCB for the mean function. Namely, first we obtain an estimator of the $(1-\alpha)100\%$ quantile, $Q_{1-\alpha}$, of the absolute maxima distribution for a standardized Gaussian process $\zeta(x)$, with $\mathbb{E}[\zeta(x)] = 0$, and $\mathbb{E}[\zeta^2(x)] = 1$, and covariance function $\mathbb{E}\left[\zeta(x)\zeta(x')\right] = G(x,x')\left\{G(x,x)G(x',x')\right\}^{-1/2}$. For any $\alpha \in (0,1)$, we denote $Q_{1-\alpha}$ the $100(1-\alpha)$-th percentile of the absolute maxima distribution of $\zeta(x)$, i.e., $\mathbb{P}(\sup_{x\in[0,1]}|\zeta(x)| < Q_{1-\alpha}) = 1-\alpha$, $0 < \alpha < 1$. This can be done using a Monte-Carlo simulation, that is, we first simulate $\zeta_l(x) = \tilde{G}(x,x)^{-1/2}\sum_{k=1}^{\kappa} Z_{k,l}\sqrt{\tilde{\lambda}_k}\tilde{\varphi}_k(x)$, where $Z_{k,l}$ follows i.i.d. standard normal distribution, $1 \le k \le \kappa$, $l = 1, \cdots, L$, and L is a large positive integer, say 1000. Next, one chooses the number κ of eigenfunctions by using the following standard "fraction of variation explained", i.e., selects the number of eigenvalues that can explain, say, 95% of the variation in the data. Then an estimator for $Q_{1-\alpha}$, $\hat{Q}_{1-\alpha}$, is obtained as the empirical $(1-\alpha)$ quantile of the set $\{\sup_{x\in[0,1]}|\zeta_l(x)|, l = 1, \cdots, L\}$.

Implementation of correction factor

In order to make the RSCB have consistent coverage rates for different samples sizes, we perform a simulation to find the correction factor. The detailed discussion is given in the Section 2.4.2. Based on the empirical coverage rates, the $(1-\alpha)100\%$ RSCB for the mean function can then be calculated as $\tilde{m}(x) \pm c(n) \cdot n^{-1/2}\tilde{G}(x,x)^{1/2}\hat{Q}_{1-\alpha}$, where $c(n) = 1.3 \cdot 1.1^{\log_2(n/30)}$ is a correction factor used to improve the consistency of the confidence band. The use of correction factors to improve the consistency of robust confidence intervals is a common procedure in R methodology (Wilcox, 2005).

2.3.6 The RSCB for the difference of two mean functions

The framework proposed here to obtain a RSCB for the mean function can be extended to obtain a RSCB for the difference of the mean functions of two populations. Denote for $d = 1, 2$ the samples coming from each population, satisfying the model defined in (2.3)

$$Y_{dij} = m_d\left(\frac{j}{N}\right) + \sum_{k=1}^{\kappa_d} \xi_{dik}\phi_{dk}\left(\frac{j}{N}\right) + \sigma_d\left(\frac{j}{N}\right)\varepsilon_{dij}, 1 \leq i \leq n_d,\ 1 \leq j \leq N,$$

with covariance function $G_d(x, x') = \sum_{k=1}^{\kappa_d} \phi_{dk}(x)\phi_{dk}(x')$, respectively. Define the ratio of two-sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1\to\infty} \hat{r} = r > 0$. For each group we can obtain the R estimator for the mean function as described in the Section 2.3.2.

Following the procedure in the Section 2.3.4, we can obtain R estimators for the covariance function of each group, $\tilde{G}_d(\cdot, \cdot)$, $d = 1, 2$, then proceed as the Section 2.3.5, by first defining $\zeta_{12}(x)$, $x \in [0, 1]$ the Gaussian process with zero mean, $\mathbb{E}[\zeta_{12}(x)] = 0$, unit variance $\mathbb{E}[\zeta_{12}^2(x)] = 1$, and covariance function

$$\mathbb{E}\left[\zeta_{12}(x)\zeta_{12}(x')\right] = \frac{\tilde{G}_1(x, x') + \hat{r}\tilde{G}_2(x, x')}{\left\{\tilde{G}_1(x, x) + \hat{r}\tilde{G}_2(x, x)\right\}^{1/2}\left\{\tilde{G}_1(x', x') + \hat{r}\tilde{G}_2(x', x')\right\}^{1/2}},$$

where $x, x' \in [0, 1]$. Analogue to the one sample case, the quantile $\hat{Q}_{12,1-\alpha}$ can be estimated using a Monte-Carlo simulation. The RSCB for $m_1(x) - m_2(x)$ is given as

$$(\tilde{m}_1(x) - \tilde{m}_2(x)) \pm c(n_1, n_2) \cdot n_1^{1/2}\left[\tilde{G}_1(x, x) + \hat{r}\tilde{G}_2(x, x)\right]^{1/2}\hat{Q}_{12,1-\alpha},$$

where the correction factor is selected as $c(n_1, n_2) = 1.3 \cdot 1.1^{\log_2(\min\{n_1,n_2\}/30)}$, similarly to the discussion in the Section 2.4.2. The confidence band for the difference of the

mean functions can be used to perform a hypothesis test of the form $H_0 : m_1(x) \equiv m_2(x)$, $\forall x \in [0,1]$ vs. $H_A : m_1(x) \neq m_2(x), \exists\, x \in [0,1]$. The test can be performed by calculating the appropriate $(1-\alpha)\times 100\%$ confidence band. Although the p-value cannot be calculated directly, it can be estimated by finding the largest α when H_0 is rejected.

2.4 Simulation

In this section we perform a simulation study to select the correction factor, and compare the performance of the proposed RSCB with the (non-robust) method proposed by Cao et al. (2012b) for the mean function and the difference of mean functions for two populations. We use empirical confidence band coverage rate as a performance criterion. Since outlier curves often have different types of outlying behaviors in the functional datasetting, we consider several types of outliers in the assessment of the performance of the RSCB.

2.4.1 Simulation setting

We first generate data from the simulation model in Cao et al. (2012b), i.e.,

$$Y_{ij} = m\left(\frac{j}{N}\right) + \sum_{k=1}^{2} \xi_{ik}\phi_k\left(\frac{j}{N}\right) + \sigma\epsilon_{ij},\ 1 \leq j \leq N,\ 1 \leq i \leq n.$$

In this model, ξ_{ik} for $k = 1, 2$ and ϵ_{ij} for $1 \leq j \leq N$, $1 \leq i \leq n$ are generated from $N(0,1)$. The number of subjects is n, the number of observations per curve is taken as $N = \lfloor n^{0.25}\log^2(n) \rfloor$ for functional samples and the number of knots is chosen as $N_m = n^{1/2p}\log(n)$. The mean function, eigenvector functions and the noise level are taken as $m(x) = 10 + \sin\{2\pi(x-1/2)\}$, $\phi_1(x) = -2\cos\{(\pi(x-1/2)\}$, $\phi_2(x) = \sin\{\pi(x-1/2)\}$, and $\sigma = 0.5$. Note that this model implies that $\lambda_1 = 2$,

$\lambda_2 = 0.5$, but this information is not used a priori. For each sample, κ is chosen to contain 95% of the total sum of the eigenvalues.

Under this functional model, we introduce outlier curves (Y_{ij}^o) to the generated functional sample by contaminating a subset, I_O, of the original functional sample. The contamination proportion varies from 0 to 0.5, at 0.05 increment. In order to determine the influence of different types of outliers on constructing SCB for the mean function we consider three different types of outliers, mainly with localized influence, which mimic the types of outlying curves encountered in the real datasets in the Section 2.5.

1. *Peak Outliers.* To simulate an outlier with a punctual influence, each curve was contaminated at a single measurement point, j^*/N, by adding a fixed value s (outliers strength), that is,

$$Y_{ij^*}^o = Y_{ij^*} + s, i \in I_O,\ j^* = \lfloor 0.05N \rfloor.$$

   This produces a peak outlier curve with a *peak* at the point j^*/N.

2. *Bump Outliers.* This type is an extension of the peak outliers and the contamination occurs in an interval, $[b_0, b_1]$, rather than at a single point, that is,

$$Y_{ij^*}^o = Y_{ij^*} + s, i \in I_O, j^*/N \in [b_0, b_1].$$

   This type of outlier is present in the ozone dataset considered in the Section 2.5. In the simulation, the interval is chosen as $[b_0, b_1] = [0.5, 0.53]$.

3. *Step Outliers.* A further extension of the bump outliers is created by contaminating the curve in the interval $[c_i, 1]$, where c_i is randomly chosen from $[0.5, 1]$

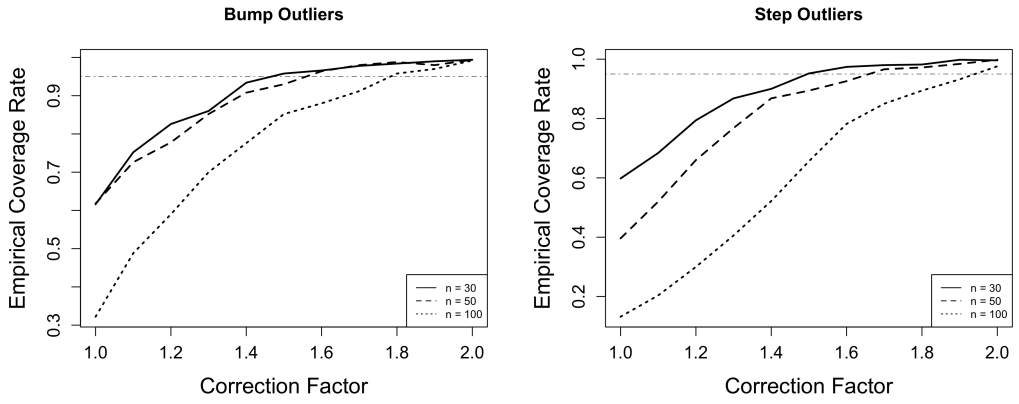

Figure 2.1: Empirical coverage rates of 95% RSCB when the correction factor c varies between 1 and 2.

for each outlying curve, that is,

$$Y^o_{ij^*} = Y_{ij^*} + s, i \in I_O, j^*/N \geq c_i.$$

2.4.2 Selecting correction factor

In this simulation, the contamination ratio of the outliers is 20% and the outliers strength is $s = 5$. Three sample sizes $n = 30, 50$ and 100 are considered and the correction factor c is varied for values between 1 and 2. For each n, the validity of the confidence band is tested at 100 points $\{1/100, \cdots, 99/100, 1\}$, against the value of the true mean function defined as in the simulation setting. The proportion of the valid confidence bands is then evaluated. The simulation is repeated 500 times. Figure 2.1 presents the empirical coverage rates of 95% SCB, i.e., $\tilde{m}(x) \pm c \cdot n^{-1/2}\tilde{G}(x,x)^{1/2}\hat{Q}_{1-\alpha}$, for the cases of bump and step outliers. Given these different empirical coverage rates for various c, we select $c\,(n) = 1.3 \cdot 1.1^{\log_2(n/30)}$ as the correction factor to improve the consistency of the RSCB.

2.4.3 An illustrative example

We first show the performance of the proposed method for one sample SCB construction on an illustrative toy example. Therefore we generate a functional sample of $n = 100$ from the model defined in the Section 2.4.1 and contaminate the data by using all three types of outliers with the contamination proportion 20% and, for this illustrative example, strength of $s = 20$. We first construct the 95% confidence band using the proposed RSCB (black) and non-robust (red) SCB for the mean function for each outlier type. We also construct the non-robust SCB and RSCB for the mean function when the sample does not have outlier curves to assess the consistency of the proposed RSCB. Figure 2.2 depicts the effects of outlier curves on the non-robust SCB and RSCB methods.

The first graph (top left in Figure 2.2) for no outlier case shows that the proposed RSCB behaves the same as the non-robust SCB when there is no outlier curve in the data, therefore it is consistent. For peak and bump outliers (top right and bottom left in Figure 2.2), which are considered as very localized outliers, the width of the non-robust SCB is widened around the outlier location and the estimate of the true mean function is strongly affected, deviating from the true mean function, which results in a confidence band that does not cover the mean function whereas the RSCB is robust to outlier curves even though it is widened slightly around the outlier location, but by a small factor, and the estimated mean is barely affected by the presence of these types of outliers.

For the step outlier (bottom right in Figure 2.2), the mean function estimation and the non-robust SCB are affected dramatically since the estimate of the mean function is deteriorated whereas the RSCB is not affected even though the band is widened a little around the outlier location. This illustrative example provides evidence that when there are outlier curves in a functional dataset, estimate of the

mean function and non-robust SCB are both affected badly while the proposed RSCB based on the R mean estimator performs well for different types of outlier curves. Further the proposed RSCB is consistent since it performs well when there are no outlier curves.

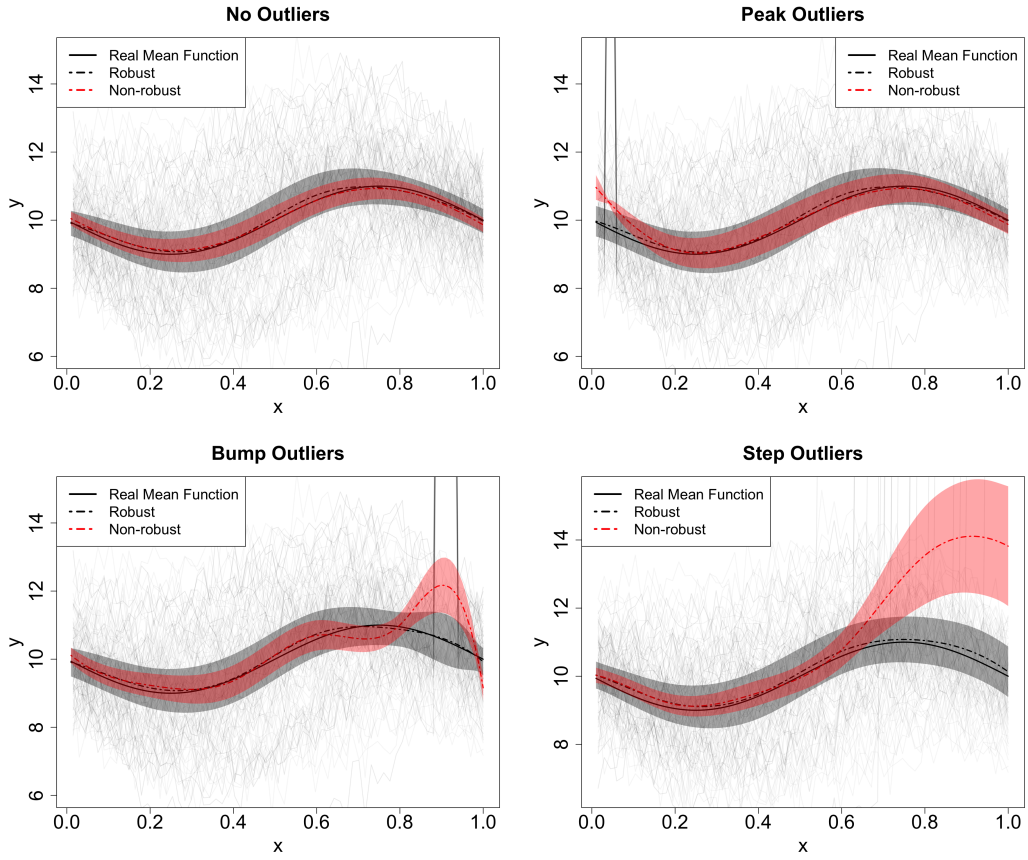

Figure 2.2: Comparison between RSCB (Black) and non-robust SCB (Red) for a simulated dataset.

Another important metric is the relationship between the strength of the outlier curves and the total area of the confidence band. To illustrate this metric, we generate functional samples of size $n = 50$ from the model defined in the Section 2.4.1 and contaminate the data by using peak and step outliers with contamination proportion 20%, and the varying strengths, $s = 5, 10, 20, 30, 40, 50$. The results are presented

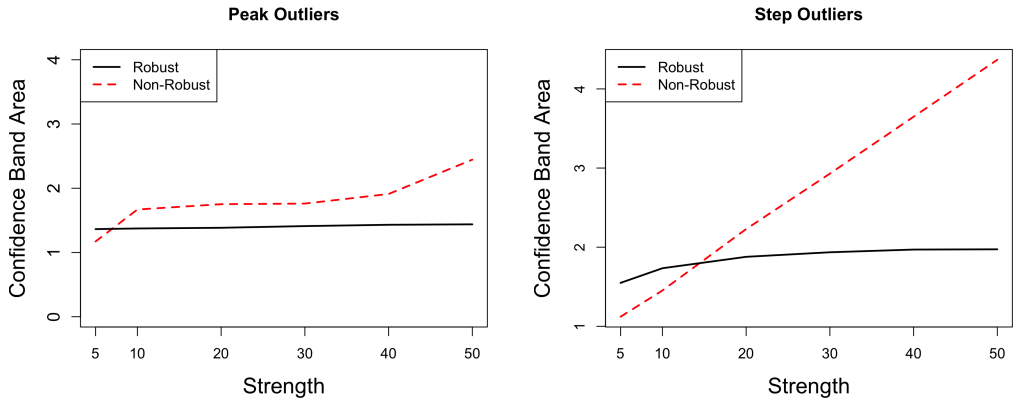

Figure 2.3: Comparison of the area of the RSCB (black) and non-robust SCB (red), for varying strength of outliers.

in Figure 2.3. Note that the area of non-robust SCB is strongly influenced by the strength of the outliers, while the area of the RSCB is maintained at an almost constant level. Similar results are obtained for the bump outliers, which are not shown here.

2.4.4 Simulation for the SCB for the mean function and the difference of two mean functions

Case I: SCB for the mean function, $m(x)$

To evaluate the performance of the proposed RSCB method for the mean function, we calculate the empirical coverage rate. We generate functional samples from the model in the Section 2.4.1 for sample sizes $n = 30, 50$, and 100 with strength is $s = 5$. Each simulation is repeated 500 times.

The empirical coverage rates for contamination proportions varying from 0.05 to 0.50 are presented in Figure 2.4 as well as in Table 2.1. The results for non-contaminated datasets are presented in Table 2.2. For the first two types of outliers (top two graphs in Figure 2.4 and Table 2.1), the proposed RSCB method maintains

Table 2.1: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for four types of outlier curves.

Outlier Type	n	Method	Contamination Proportion *0.05*	*0.10*	*0.15*	*0.20*	*0.25*	*0.30*	*0.35*	*0.40*	*0.45*	*0.50*
Peak	30	R	0.85	0.85	0.86	0.91	0.93	0.92	0.93	0.92	0.91	0.82
		NR	0.64	0.17	0.09	0.01	0.00	0.01	0.00	0.00	0.01	0.00
	50	R	0.91	0.93	0.90	0.88	0.87	0.88	0.86	0.83	0.82	0.78
		NR	0.77	0.62	0.34	0.10	0.05	0.04	0.03	0.02	0.01	0.02
	100	R	0.96	0.99	0.98	0.97	0.96	0.94	0.93	0.94	0.95	0.92
		NR	0.85	0.77	0.62	0.37	0.20	0.08	0.02	0.00	0.00	0.00
	200	R	1.00	0.99	1.00	1.00	0.99	0.98	1.00	0.98	0.98	0.98
		NR	0.90	0.83	0.79	0.68	0.54	0.43	0.28	0.17	0.09	0.00
Bump	30	R	0.84	0.81	0.83	0.76	0.78	0.76	0.79	0.71	0.69	0.66
		NR	0.77	0.72	0.63	0.44	0.34	0.13	0.09	0.03	0.02	0.02
	50	R	0.89	0.90	0.90	0.89	0.89	0.85	0.87	0.82	0.75	0.65
		NR	0.82	0.73	0.65	0.41	0.25	0.07	0.03	0.00	0.00	0.00
	100	R	0.98	0.97	0.96	0.93	0.88	0.91	0.86	0.76	0.68	0.53
		NR	0.81	0.57	0.17	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	200	R	0.99	0.99	0.97	0.91	0.85	0.69	0.62	0.46	0.27	0.13
		NR	0.74	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Step	30	R	0.82	0.85	0.86	0.87	0.88	0.88	0.82	0.76	0.64	0.14
		NR	0.77	0.60	0.49	0.22	0.06	0.00	0.00	0.00	0.00	0.00
	50	R	0.93	0.90	0.88	0.89	0.87	0.83	0.80	0.58	0.34	0.01
		NR	0.79	0.46	0.18	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	100	R	0.97	0.95	0.92	0.86	0.82	0.70	0.46	0.20	0.02	0.00
		NR	0.61	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	200	R	0.97	0.95	0.91	0.74	0.53	0.32	0.10	0.01	0.00	0.00
		NR	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2.2: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for datasets with no outliers.

	Method	n 30	50	100	200
Clean Dataset	R	0.86	0.91	0.98	0.98
	NR	0.86	0.87	0.91	0.90

the nominal level 95%, and has breakdown point close to 50%, while the non-robust SCB breaks down immediately in the presence of outliers, and shows a rapid decrease in the empirical coverage rate. Hence we conclude that proposed RSCB performs superiorly to the non-robust SCB for these two cases.

For the step outlier case, the RSCB method still performs well (bottom graph in Figure 2.4 and Table 2.1). Although the RSCB method does not maintain 50% breakdown point as in the case of peak and bump outliers, but still has reasonably good breakdown point (20% to 30%), whereas the non-robust SCB breaks down immediately in the presence of even small contamination proportions.

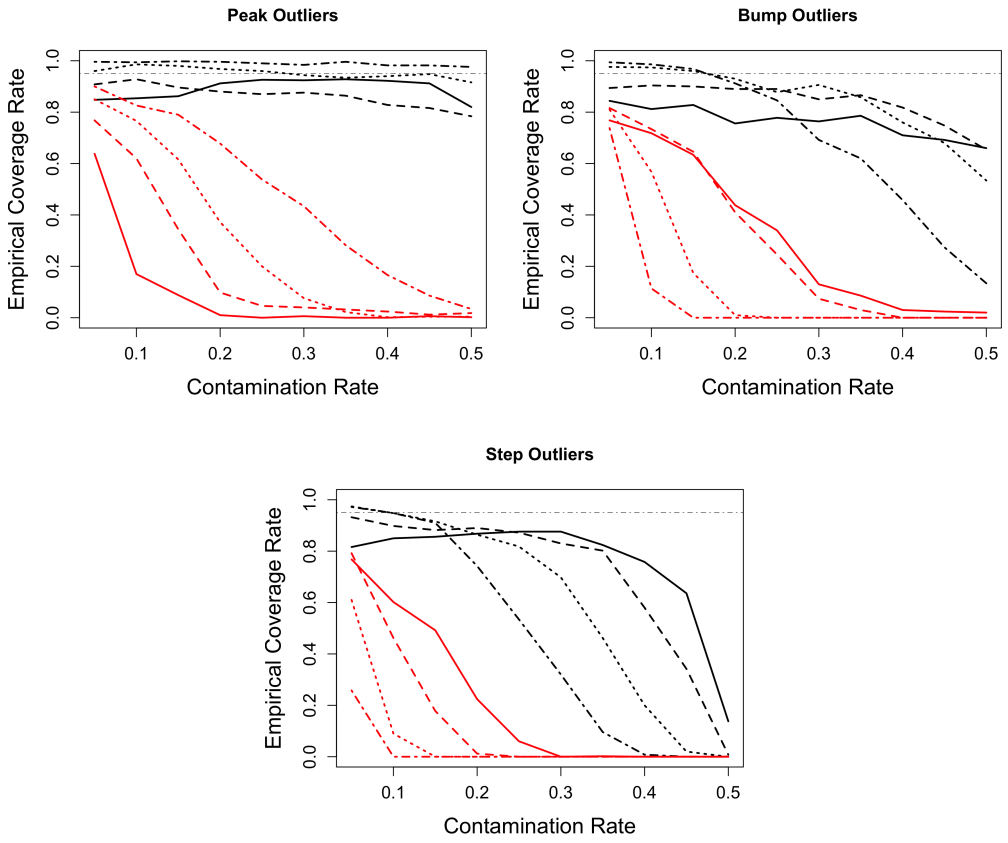

Figure 2.4: Empirical coverage rates of 95% SCB. Non-robust method (Red) vs. Robust method (Black). $n = 30$ (solid), $n = 50$ (dashed), $n = 100$ (dotted), $n = 200$ (dash-dot).

Case II: SCB for the difference of two mean functions, $m_1(x) - m_2(x)$

We also conduct simulation to evaluate the performance of the RSCB method for the difference between two mean functions, by testing the hypotheses described in the Section 2.3.6,

$$H_0 : m_1(x) = m_2(x),\ \forall x \in [0,1] \text{ vs. } H_A : m_1(x) \neq m_2(x), \exists\, x \in [0,1]. \tag{2.7}$$

We employ the same model in the Section 2.4.1 for the one sample case. In this simulation setup, $n_1 = 100$, and $n_2 = 130$ correspond to the sample sizes for the first and the second population, respectively, $N = 100$ are the number of measurement points for both samples, and outlier curves are introduced to the first population.

The results of the simulation are presented in Figure 2.5 for the three types of outliers, using RSCB (red), and using non-robust SCB (black) as proposed in Cao et al. (2012b). For all cases, the type I error is not maintained for the non-robust method, while for the RSCB method, the type I error is kept at, or close to the nominal value. For the step outliers, the type I error is close to the nominal value for small contamination proportions, deteriorating for large contamination proportions, albeit at a much slower rate than the non-robust SCB. This is evidence that our proposed RSCB method has a superior performance compared to the classical one.

2.5 Applications

We illustrate our approach on two datasets: Octane dataset for the one sample case and Ground level Ozone concentration dataset for the two sample case.

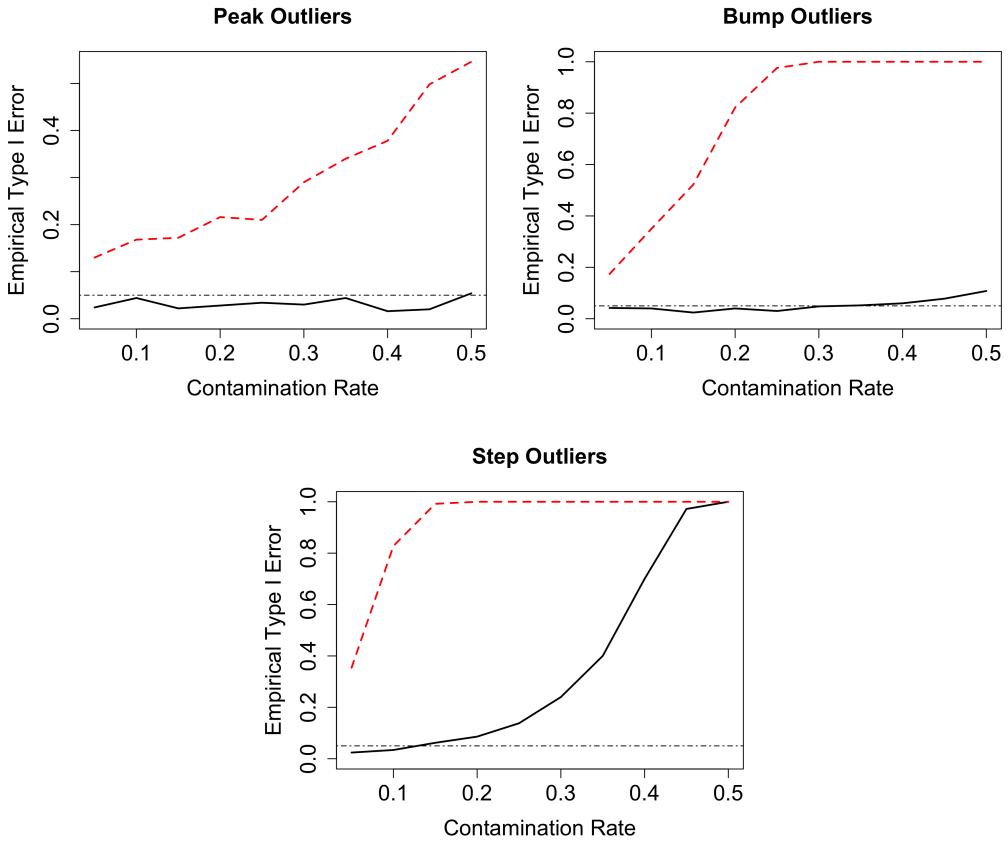

Figure 2.5: Empirical type I error of Hypothesis Test for the difference of means for two populations, with nominal value $\alpha = 0.05$. Non-robust method (Red) vs. Robust method (Black).

2.5.1 Octane dataset

This dataset consists of 39 near infrared (NIR) spectra of gasoline sample, obtained from Esbensen et al. (1996). It is known that 6 of the samples contain added ethanol, which corresponds to an upward translation on the upper wavelength, 1390 onward, interval of the spectrum. This is considered as the step outliers described in the Section 3.4.

The R estimation of the mean and the 95% RSCB are calculated for this dataset, as well as the mean estimator and confidence band using the method in Cao et al.

(2012b). The results are presented in Figure 2.6, showing the full spectrum measure (left panel) and magnified on the second half of the spectrum to display the differences more apparently between the non-robust and robust SCBs (right panel).

We observe that the R mean estimator remains close to the non-outlying curves, while the non-robust estimate of the mean function is heavily influenced by the outliers, resulting in an upward shift. The non-robust SCB is also heavily influenced by the outliers, translating in a very wide band on the second half of the spectrum. However the proposed RSCB maintains a consistent width across the whole spectrum.

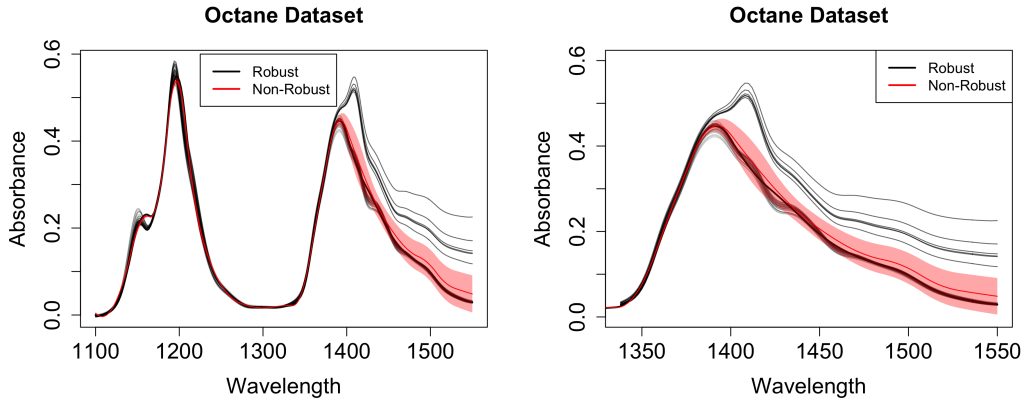

Figure 2.6: 95% SCB comparison for the octane dataset. Non-Robust (Red) vs. Robust (Black) methods. Left: full spectrum. Right: magnified on the second half of the spectrum.

2.5.2 Ground level Ozone concentration dataset

This dataset consists of hourly average measurements of ground level ozone (O_3) concentrations from a monitoring station in Richmond, BC, Canada, from the years of 2004 to 2012. The presence of Ozone at ground level is highly undesirable, and considered a serious air pollutant. Since the concentration of ground level Ozone

typically peaks at summer months, only the month of August is analyzed, resulting in 31 samples, with 24 measurement points for each sample.

The same dataset was studied in Boente and Salibian-Barrera (2015), and using S-estimators for the principal components, the presence of outliers was detected in the year of 2005. For illustrative purposes, we take the ozone levels for the year of 2005 as one sample and the ozone levels for the year of 2007, which we know there are no outlier curves (Boente and Salibian-Barrera, 2015), as the other sample. The plot of the ground level O_3 concentration for years 2005 and 2007 is presented in Figure 2.7, top left panel, with the year of 2005 in gray/black, and the year of 2007 in red. The outliers detected by Boente and Salibian-Barrera (2015) are highlighted.

We set up our hypotheses for testing if there is a difference between the ozone mean functions of the years 2005 and 2007 in Richmond, Canada. The outliers in the dataset are similar to the bump outliers described in the Section 3.4. The ground level O_3 concentration remained the same, except for the aforementioned outliers in 2005 (Figure 2.7, top left panel). The R method does not reject the null hypothesis at a significant level $\alpha = 0.05$, while the non-robust method proposed in Cao et al. (2012b) rejects the null hypothesis, with an empirical p-value calculated as 0.015. The 95% SCB of the difference between the mean functions of the ground level O_3 concentration in years of 2005 and 2007 is presented in the top right panel of Figure 2.7. We also calculate the 95% SCB for the difference between the mean functions with the outliers kept for the RSCB, and excluding the outliers for the non-robust SCB. This is presented in the bottom panel of Figure 2.7. This plot provides a comparison of the SCB between the robust and non-robust methods, highlighting that the former successfully works, despite the presence of outliers.

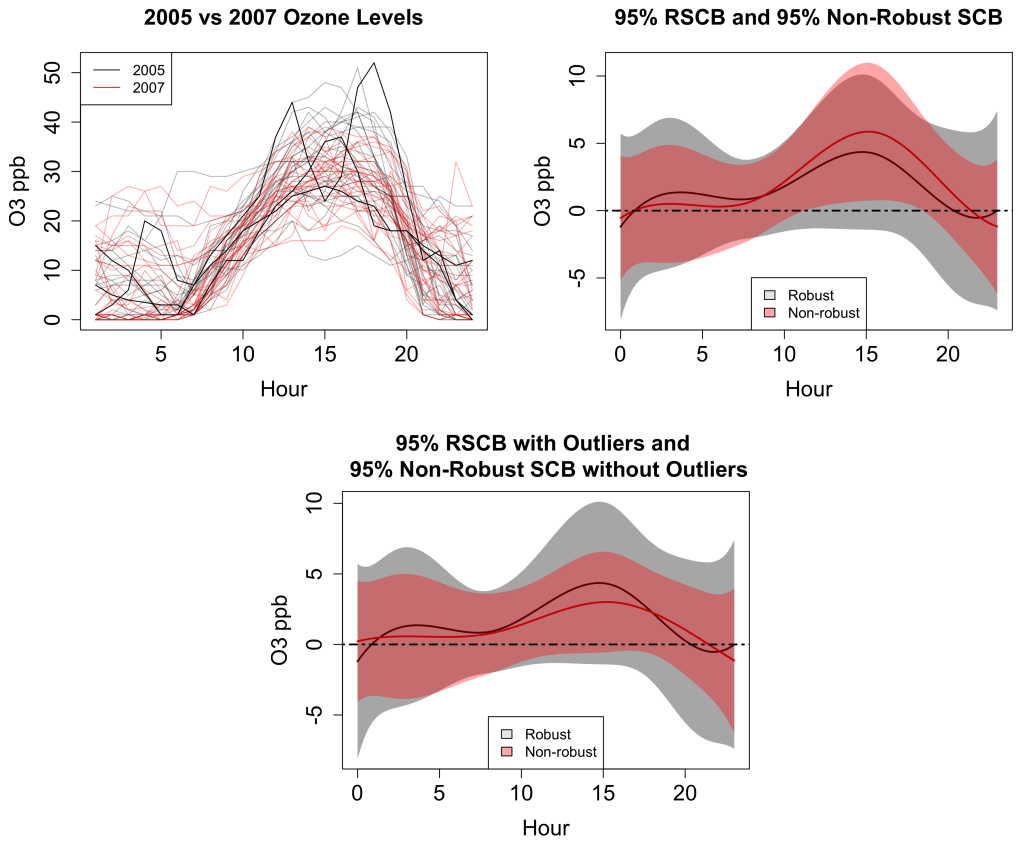

Figure 2.7: Top Left: O_3 Levels in years of 2005 (Gray and black) and 2007 (Red) in Richmond. Black lines are the outliers which are determined in Boente and Salibian-Barrera (2015). Top Right: 95% non-robust SCB (Red) and RSCB (Gray) for the difference between the mean functions of the two years. Bottom: 95% non-robust SCB (Red) and RSCB (Gray) for the difference between the mean functions of the two years, keeping outliers for RSCB, excluding outliers for non-robust SCB.

2.6 Conclusion

In this chapter we presented a novel method of robustly estimating the mean function of the functional data, and a novel method to obtain a RSCB for the mean function. To our best knowledge, this work is the first manuscript investigating the construction of RSCB accounting for the presence of outlying curves. The proposed method performs well under the presence of the three types of local outliers studied, which are typical types of contamination in the R functional data analysis. Although we primarily focus on the computational issue of RSCB, we also proved the consistency of the proposed LAD estimator for the mean function. Careful developing a consistent estimator of covariance function of the LAD estimator may produce a more accurate RSCB, but this procedure is not straightforward. A careful analysis of phase outliers is also necessary due to the use of L^2 norms for distance measurements. We leave these issues for a future work. We also believe that an exploration of RSCB for the functional derivatives of the mean, together with the methods proposed here can improve the understanding of the behavior of the mean function.

Chapter 3

B-Spline Smoothed M-Based Simultaneous Inference for Functional Data

3.1 Introduction

In this chapter we present a complement and extension to the methods presented in the Chapter 2, where we proposed a robust framework for the estimation of the mean function of functional data, and for the simultaneous confidence band for the mean function, and also a robust hypothesis test for the difference of the mean function of two populations. We complement the Chapter 2 by presenting a framework in which the asymptotic theory of the results there presented can be established, and we extend those results by proposing an additional robust estimator, the B-spline smoothed M-Estimator.

Exact theoretical properties of statistical methods for functional data are more often than not out of reach due to the intricate structure of probabilities in functional spaces. This is also added to the greater technical difficulty in formulating SCB for the mean function of an infinite dimensional functional response. In our setting, there is also the complex outlier structure considered, which increases the challenge of a theoretical framework. In this chapter, we propose a robust estimator for the mean function, a robust method to construct a simultaneous confidence band in the functional data setting and a robust hypothesis test statistic for the difference between the mean functions of two populations. The proposed method uses the B-spline smoothed M-Estimation. We also present the asymptotic theoretical properties of the proposed estimators, proving asymptotic consistency and asymptotic normality of the mean function estimator.

This chapter is organized as follows. We first introduce the model used for the functional dataset in the Section 3.2. We then propose a robust estimator for the mean function when the dataset contain outliers, using a B-spline smoothing M-estimation method in the Section 3.3. This is followed by the presentation of the asymptotic properties of the proposed estimator. Then we construct a RSCB for the mean function based on the proposed robust estimator for the mean function. In addition, we extend this method to construct a robust simultaneous confidence band for the difference of the mean functions of two populations. In the Sections 3.4 and 3.5, the performance of the proposed robust methods, and their robustness, are demonstrated with an extensive simulation study and real data examples. Finally we conclude with some discussion.

3.2 Model

Consider the following model for a functional dataset

$$Y_i(x) = m\,(x) + \epsilon_i(x), 1 \leq i \leq n$$

where $m(\cdot)$ is a non-random component and $\{\epsilon_i(\cdot)\}_{i=1}^n$ are independent random noise. Assume also that each sample curve is observed only on a discrete set of measurement points $x_1, \ldots, x_N$, that is,

$$Y_{ij} = Y_i(x_j) = m\,(x_j) + e_{ij}, \tag{3.1}$$

where $e_{ij} = \epsilon_i(x_j)$. For this paper, we assume that the interval of definition of the random functions is the unit interval $[0, 1]$ and the measurements are equally spaced, that is, $x_1 = 1/N, x_2 = 2/N, \ldots, x_N = 1$.

Notice that if $\mathbb{E}(\epsilon_i) = 0$, then $m(\cdot)$ is the mean function in the traditional sense. If we weaken this assumption, only requiring that distribution of $e_{i\cdot}$ is symmetric, $i = 1, \ldots, n$, we cannot guarantee the existence of $\mathbb{E}(Y(\cdot))$, $m(\cdot)$ can be viewed as a center function of the functional data. Although we postpone the specific assumptions required to the Appendix, for simplicity, we will use the term *mean function* when we refer to the function $m(\cdot)$.

3.3 Proposed Methods

In this section we will present the robust estimation method for the mean function. The estimator extends the concept of M-Estimators from classical robust statistics, adding the use of B-Splines basis to improve the quality of estimation and also allow for interpolation of the discrete measurements. We call this the *B-Spline Smoothed M-Estimator for the Mean Function*. We also present the asymptotic properties of estimator, showing that it is (weakly) consistent, with order of convergence approaching the best non-parametric convergence order from Stone (1985).

A method for estimating a *Robust Simultaneous Confidence Band* is also proposed. The method is built on the results obtained for the mean function estimator and a modification of the concept of *pseudo-data*, introduce by Cox (1983). We also extend the robust simultaneous confidence band to obtain a robust statistic for testing the two-sample difference of mean functions.

3.3.1 B-spline Smoothed M-Estimator for Mean Function

Denote by $\mathcal{H}^{(p-2)}$ the *p-th* order spline space, i.e., $p-2$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $p-1$ on $[0, t_1], [t_J, t_{J+1}]$, $J = 1, \ldots, N_m - 2$ and $[t_{N_m}, 1]$. The points $\{t_J\}_{J=1}^{N_m}$ are called interior knots, and for each integer p, introduce left boundary knots $t_{1-p}, \ldots, t_0$ and right boundary knots

$t_{N_m+1}, \ldots, t_{N_m+p}$, satisfying $t_{1-p} = \ldots = t_0 = 0 \leq t_1 < t_2 \cdots < t_{N_m} \leq 1 = t_{N_m+1} = \ldots = t_{N_m+p}$.

Define the *B-spline smoothed M-estimator* of the mean function by

$$\hat{m}(x) = \sum_{J=1-p}^{N_m} \hat{\boldsymbol{\beta}}_J B_j(x),$$

where

$$\left\{\hat{\boldsymbol{\beta}}_{1-p}, \ldots, \hat{\boldsymbol{\beta}}_{N_m}\right\}^{\mathrm{T}} = \mathrm{argmin}_{\{\boldsymbol{\beta}_{1-p}, \ldots, \boldsymbol{\beta}_{N_m}\} \in R^{N_m+p}} \sum_{i=1}^{n} \sum_{j=1}^{N} \rho\left(Y_{ij} - \sum_{J=1-p}^{N_m} \boldsymbol{\beta}_J B_J\left(j/N\right)\right), \tag{3.2}$$

and ρ is a suitably chosen loss function. In this work, we focus on convex loss functions, which guarantees that the equation (3.2) has a unique solution. Different choices of ρ will lead to different estimation properties of the mean function m. For example, if we choose $\rho(x) = x^2$, we obtain the classical OLS estimator which was studied in Cao et al. (2012b). If we consider $\rho(x) = |x|$ we obtain the LAD estimator, which was studied in the Chapter 2.

The robust properties of different choices of ρ functions has been extensively studied in the literature, see Wilcox (2005). One notable example is the Huber loss function defined by

$$\rho_k(x) = \begin{cases} x^2/2, |x| \leq k, \\ k(|x| - k/2), |x| > k, \end{cases}$$

where $k > 0$. Note that ρ_k combines both the OLS and the LAD loss functions for small and large values, respectively. The parameter k controls a trade-off between the resistance to outlying observations and efficiency of the estimator. For example,

a choice of $k = 1.345$ provides 95% asymptotic efficiency on the normal distribution, see Huber et al. (1964).

Define the function $\psi(x) = \rho'(x)$ a.e., then the estimated coefficients in (3.2) can also be obtained by the following system of equations

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\psi\left(Y_{ij} - \sum_{J=1-p}^{N_m}\hat{\boldsymbol{\beta}}_J B_J(j/N)\right)B_k(j/N) = 0,\ 1-p \leq k \leq N_m, \tag{3.3}$$

which can be seen from direct differentiation. A closed form solution to (3.3) usually does not exist. An approximate solution can be obtained using an iteratively re-weighted least squares fitting algorithm. This is performed by first obtaining a crude initial approximation, $m^0(\cdot)$, for the estimator and calculating weights $w_{ij}^0 = \psi(e_{ij}^0)/e_{ij}^0$, where $e_{ij}^0 = Y_{ij} - m^0(j/N)$, $i = 1, \ldots, n$, $j = 1, \ldots, N$. Then, a weighted least square is fit using the weights $w_{ij}^0, i = 1, \ldots, n, j = 1, \ldots, N$, producing a new approximation for the estimator. New weights can be calculated, and the procedure is repeated until convergence. This algorithm is implemented in the function `rlm` from the MASS R-package, Venables and Ripley (2002). Notice that, since we are assuming that ρ is a convex loss function, the uniqueness of the solution is guaranteed.

We perform a simulation analysis comparing the performance of the spline smoothed M-estimator and a non-robust spline smoothed least square method. The results of the analysis are presented in the Section 3.4 and serve to indicated the need for a robust procedure.

3.3.2 Asymptotic Properties of the B-Spline Smoothed M-Estimator

In this section we will explore the asymptotic properties of the proposed B-Spline Smoothed M-Estimator for the mean function of functional data. We prove the asymptotic consistency and the asymptotic normality of the B-spline Smoothed

M-estimator. Before stating the first result, we need to introduce some notations. For $0 < r < 1$, denote by $\mathcal{C}^{p,r}$ the Hölder function space, that is, the space of functions with continuous derivatives up to order p, and with r-Hölder continuous p-derivative. For a real-valued function, f, denote by $\|f\|_2$ the standard L^2 space norm, that is, $\|f\|_2^2 = \int_0^1 |f(x)|^2 dx$. Similarly, for a vector $\boldsymbol{V}^{\mathrm{T}} = (V_i)_{i=1}^k$, let $\|\boldsymbol{V}\|_2^2 = \sum_{i=1}^k |V_i|^2$ and for a matrix $\mathbb{A}$, $\|\mathbb{A}\|_2 = \sup_{\boldsymbol{V}\neq 0} \|\mathbb{A}\boldsymbol{V}\|_2/\|\boldsymbol{V}\|_2$. Let $\lambda_{\max}(\mathbb{A})$ and $\lambda_{\min}(\mathbb{A})$ be the largest and smallest eigenvalue of matrix $\mathbb{A}$, respectively. Note that $\|\mathbb{A}\|_2 = \lambda_{\max}(\mathbb{A})$ and if matrix $\mathbb{A}$ is non-singular, $\|\mathbb{A}^{-1}\|_2 = \lambda_{\min}(\mathbb{A})^{-1}$. Throughout this section, C denotes a uniform positive constant. We need the following assumptions for the asymptotic consistency and the asymptotic normality of the proposed method.

(A1) Let p be the order of the smoothing splines, N_m be the number of B-spline knots, and assume that $(n/N)^{1/3} \ll N_m \ll \min(n^{1/3}, n/N)$ and $N_m \log N_m \ll N$;

(A2) The function $m(x)$ satisfies $m \in \mathcal{C}^{p,r}$;

(A3) Let $\psi(x) = \rho'(x)$, then $\psi(x)$ is continuous, non-decreasing and uniformly bounded, $|\psi(x)| < C$, $\forall x \in \mathbb{R}$. Also, $\rho(\cdot)$ is a *convex function*;

(A4) $\mathbb{E}\psi(e_{ij}) = 0$ and $\mathbb{E}\left[\psi(e_{ij})\right]^2 \leq C$;

(A5) There exists a bounded function $\delta(x)$ satisfying $0 < \inf_{x\in\mathbb{R}} \delta(x) < \sup_{x\in\mathbb{R}} \delta(x) < \infty$, such that

$$\left|\mathbb{E}[\psi(e_{ij}+u)] - \delta\left(\frac{j}{N}\right)\cdot u\right| \leq Cu^2,\ |u| < C;$$

(A6) $\mathbb{E}\left[\psi(e_{ij}+u) - \psi(e_{ij})\right]^2 \leq C|u|$, and $|\psi(u+v) - \psi(v)| < C$, for $|u| < C$, and $v \in \mathbb{R}$;

(A7) Define $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{iN})^{\mathrm{T}}$, $\boldsymbol{\psi}(\boldsymbol{e}_i) = (\psi(e_{i1}), \cdots, \psi(e_{iN}))^{\mathrm{T}}$ and $\mathbb{G}_i = \mathbb{E}\left(\boldsymbol{\psi}(\boldsymbol{e}_i) \cdot \boldsymbol{\psi}(\boldsymbol{e}_i)^{\mathrm{T}}\right)$, $1 \leq i \leq n$. Also, $\min_{1 \leq i \leq n} \lambda_{\min}\left(\mathbb{G}_i\right) \geq \lambda_0$.

Remark 1. Assumption (A2) is standard in B-spline approximation, see for example Huang et al. (2004), Cao et al. (2012b), and allows for arbitrarily good approximations of $m(x)$ by spline functions. Assumption (A3) is a common condition in the theory of M-estimators, and guarantees the existence of the solution of the optimization problem in (3.2). The boundedness of ψ is a technical assumption needed for the proof of the consistency of the estimator. It doesn't pose a large restriction, since most of the ψ functions chosen in practice satisfy this condition, such as the Huber loss function. Assumption (A4) states that the function ψ adds some regularity to the errors e_{ij}. Assumptions (A5) and (A6) are regularity conditions on the function ψ. Assumption (A7) is used to prove the asymptotic normality of the proposed estimator.

Asymptotic Consistency

The first result states that the spline smoothed M-Estimator, $\widehat{m}$, converges in probability to the mean function m at rate approaching the best non-parametric convergence rate established in Stone (1985).

Theorem 2 (Asymptotic Consistency)**.** *Under Assumptions (A1) - (A6) we have*

$$\|\widehat{m} - m\|_2^2 = O_P\left(n^{-1} N_m + N_m^{-2p}\right). \tag{3.4}$$

Remark 2. As a consequence from Theorem 2, the L^2 error between $\widehat{m}$ and m has an order of magnitude bounded by the maximum of $n^{-1}N_m$ and N_m^{-2p}. Choosing $N_m = O(n^{1/(2p+1)})$ produces an optimal convergence rate equal to $o_P(n^{(2p-2)/(2p+1)})$.

Proof of Theorem 2

To prove Theorem 2, we first need to introduce some notations. Let $\widehat{m}$ be the proposed spline-smoothed M-estimator of the mean function, i.e., $\widehat{m}(\cdot) = \sum_{J=1-p}^{N_m} \hat{\boldsymbol{\beta}}_J B_J(\cdot)$, $\{\hat{\boldsymbol{\beta}}_J\}_{J=1-p}^{N_m}$ is defined in (3.2). Let $\boldsymbol{B}(\cdot) = (B_{1-p}(\cdot), \ldots, B_{N_m}(\cdot))^{\mathrm{T}}$, where $B_J(\cdot)$ are the B-spline basis functions, $J = 1-p, \ldots, N_m$, as defined in Section 3.3.1.

The following lemma provides an initial approximation of $m(x)$ by a spline function, resulting in an approximation bias. Notice that, although this bias is negligible for increasing N_m, the approximating spline function cannot be directly obtained from the sample curves, therefore it is an infeasible estimator.

Lemma 1. *Define $R_n(x) = \boldsymbol{B}^{\mathrm{T}}(x)\boldsymbol{\beta}^* - m(x)$. If Assumption (A1) holds, then there exists $\boldsymbol{\beta}^* \in \mathbb{R}^{N_m+p}$ such that*

$$\sup_{x\in[0,1]} |R_n(x)| = \sup_{x\in[0,1]} |\boldsymbol{B}^{\mathrm{T}}(x)\boldsymbol{\beta}^* - m(x)| = O(N_m^{-p}). \tag{3.5}$$

Proof. The proof follows from general properties of spline approximation, i.e., Corollary 6.21 in Schumaker (2007). □

The above result states that to prove Theorem 2, we can replace m by $m^*(\cdot) = \boldsymbol{B}^{\mathrm{T}}(\cdot)\boldsymbol{\beta}^*$ up to an order of N_m^{-p}. In particular, we just need to prove that

$$\|\widehat{m} - m^*\|_2^2 = O_P\left(n^{-1}N_m\right). \tag{3.6}$$

Define

$$\mathbb{S}_n = \frac{n}{N}\sum_{j=1}^{N} \boldsymbol{B}\left(\frac{j}{N}\right)\boldsymbol{B}^{\mathrm{T}}\left(\frac{j}{N}\right), \quad \tilde{\boldsymbol{B}}\left(\frac{j}{N}\right) = \mathbb{S}_n^{-1/2}\boldsymbol{B}\left(\frac{j}{N}\right), \text{ and } \boldsymbol{\theta} = \mathbb{S}_n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*), \tag{3.7}$$

where $\boldsymbol{\beta}^*$ is defined in Lemma 1.

Lemma 2 (Lemma A.3, Huang et al. (2004))**.** *Under Assumption (A1), there exist positive constants M_1 and M_2, such that all the eigenvalues of $(N_m/n)\mathbb{S}_n$ fall between M_1 and M_2, and $\mathbb{S}_n$ is invertible consequently.*

By Lemma 2, and the definition of $\tilde{\boldsymbol{B}}(\cdot)$ in (3.7), we have $Y_{ij} - \boldsymbol{B}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\beta} = e_{ij} + R_{nj} - \tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}$. This implies that

$$\begin{aligned} & \min_{\boldsymbol{\beta}\in\mathbb{R}^{N_m+p}} \sum_{i=1}^{n} \frac{1}{N} \sum_{i=1}^{N} \rho\left(Y_{ij} - \boldsymbol{B}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\beta}\right) \\ = & \min_{\boldsymbol{\theta}\in\mathbb{R}^{N_m+p}} \sum_{i=1}^{n} \frac{1}{N} \sum_{i=1}^{N} \left[\rho\left(e_{ij} + R_{nj} - \tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right) - \rho\left(e_{ij} + R_{nj}\right)\right], \quad (3.8) \end{aligned}$$

where $\rho(e_{ij} + R_{nj})$ is a constant term with respect to $\boldsymbol{\theta}$, so it does not change the minimum. Let $\Gamma_n(\theta)$ be the operator defined in (3.8), that is,

$$\Gamma_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \left[\rho\left(e_{ij} + R_{nj} - \tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right) - \rho\left(e_{ij} + R_{nj}\right)\right],$$

and define

$$\Delta_n(\boldsymbol{\theta}) = \Gamma_n(\boldsymbol{\theta}) - \mathbb{E}\left(\Gamma_n(\boldsymbol{\theta})\right) + \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \left[\psi(e_{ij})\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right]. \quad (3.9)$$

Before proving Theorem 2, we need to obtain asymptotic upper bounds on $\Gamma_n(\boldsymbol{\theta})$, $\mathbb{E}\left(\Gamma_n(\boldsymbol{\theta})\right)$ and $\Delta_n\left(\boldsymbol{\theta}\right)$ in the following lemma.

Lemma 3. *Under Assumptions (A1) - (A4) and (A6) and for a fixed constant $L > 1$,*

$$\sup_{\|\boldsymbol{\theta}\|_2<1} \left|\frac{1}{N_m}\Delta_n\left(N_m^{1/2} L\,\boldsymbol{\theta}\right)\right| = o_P(1). \quad (3.10)$$

Proof. We need to prove that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{\|\boldsymbol{\theta}\|_2 \leq 1} \frac{1}{N_m} \left| \Delta_n \left(N_m^{1/2} L\, \boldsymbol{\theta} \right) \right| \geq \varepsilon \right) \to 0. \tag{3.11}$$

This asymptotic bound will be proved using Bernstein's theorem. To do this, first define

$$\Pi = \left\{ \boldsymbol{\theta} \in \mathbb{R}^p; \|\boldsymbol{\theta}\|_2 \leq 1 \right\},$$

and find a decomposition

$$\Pi = \Pi_1 \cup \cdots \cup \Pi_{K_n},$$

where $\{\Pi_k\}_{k=1}^{K_n}$ are pairwise disjoint sets and, for any $1 \leq k \leq K_n$,

$$\operatorname{diam}(\Pi_k) = \max_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Pi_k} \{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2\} \leq q_o = \varepsilon C^{-1} N_m n^{-1}.$$

Notice that we can find such decomposition with

$$K_n \leq \left(2\frac{\sqrt{p}}{q_0} + 1 \right)^p. \tag{3.12}$$

For each $1 \leq k \leq K_n$, select $\boldsymbol{\theta}_k \in \Pi_k$. Then we have

$$\begin{aligned}
& \min_{1 \leq k \leq K_n} \frac{1}{N_m} \left| \Delta_n \left(N_m^{1/2} L\, \boldsymbol{\theta} \right) - \Delta_n \left(N_m^{1/2} L\, \boldsymbol{\theta}_k \right) \right| \\
\leq & \min_{1 \leq k \leq K_n} \frac{1}{N_m} \left| \Gamma_n (N_m^{1/2} L\, \boldsymbol{\theta}) - \Gamma_n (N_m^{1/2} L\, \boldsymbol{\theta}_k) \right| \\
+ & \min_{1 \leq k \leq K_n} \frac{1}{N_m} \left| \mathbb{E}\left(\Gamma_n (N_m^{1/2} L\, \boldsymbol{\theta}) \right) - \mathbb{E}\left(\Gamma_n (N_m^{1/2} L\, \boldsymbol{\theta}_k) \right) \right| \\
+ & \min_{1 \leq k \leq K_n} \frac{1}{N_m} \left| N_m^{1/2} L \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \left[\psi(e_{ij}) \cdot \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \boldsymbol{\theta} - \psi(e_{ij}) \cdot \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \boldsymbol{\theta}_k \right] \right| \\
= & \; I + II + III.
\end{aligned}$$

We proceed by obtaining an asymptotic upper bound for *I, II* and *III*. Using the definition of $\Gamma_n(\boldsymbol{\theta})$, we have

$$I = \min_{1\le k\le K_n} \frac{1}{N_m} \left| \sum_{i=1}^{n} \frac{1}{N} \sum_{i=1}^{N} \left[\rho\left(e_{ij} + R_{nj} - N_m^{1/2} L\, \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \boldsymbol{\theta} \right) - \rho\left(e_{ij} + R_{nj} \right) \right] \right.$$
$$\left. - \sum_{i=1}^{n} \frac{1}{N} \sum_{i=1}^{N} \left[\rho\left(e_{ij} + R_{nj} - N_m^{1/2} L\, \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \boldsymbol{\theta}_k \right) - \rho\left(e_{ij} + R_{nj} \right) \right] \right|,$$

and using the mean value theorem on $\rho(\cdot)$ and Assumption (A2), we have

$$I \le N_m^{-1} C n N_m^{1/2} L \max_{1\le j\le N} \left\| \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \right\|_2 \min_{1\le k\le K_n} \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_2.$$

Similarly, $\max\{II, III\} \le N_m^{-1} C n N_m^{1/2} L \max_{1\le j\le N} \left\| \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \right\|_2 \min_{1\le k\le K_n} \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_2$. Combining the previous results, we have

$$\begin{aligned} & \min_{1\le k\le K_n} \frac{1}{N_m} \left| \Delta_n\left(N_m^{1/2} L\, \boldsymbol{\theta} \right) - \Delta_n\left(N_m^{1/2} L\, \boldsymbol{\theta}_k \right) \right| \\ \le\ & 3\, N_m^{-1} C \min_{1\le k\le K_n} \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_2\, n N_m^{1/2} L \max_{1\le j\le N} \left\| \tilde{\boldsymbol{B}}^{\mathrm{T}} \left(\frac{j}{N} \right) \right\|_2 \\ \le\ & 3 N_m^{-1} C q_0 d_n = 3\varepsilon d_n, \end{aligned} \tag{3.13}$$

where

$$d_n = N_m \left[\max_{1\le j\le N} \left(\left\| \tilde{\boldsymbol{B}} \left(\frac{j}{N} \right) \right\|_2 L N_m^{1/2} + |R_{nj}| \right) \right].$$

By Lemma 2 one has $N_m \left[\max_{1\le j\le N} \left\| \tilde{\boldsymbol{B}} \left(\frac{j}{N} \right) \right\|_2 L N_m^{1/2} \right] = O\left(N_m^{3/2} n^{-1/2} \right) = o(1)$. According to Lemma 1, one has $N_m \max_{1\le j\le N} |R_{nj}| = O\left(N_m^{1-p} \right) = o(1)$. Combining these two upper bounds we have $d_n = o(1)$. In particular, by (3.13) and by choosing $d_n < 1/12$, we have

$$\min_{1\le k\le K_n} \frac{1}{N_m} \left| \Delta_n\left(N_m^{1/2} L\, \boldsymbol{\theta} \right) - \Delta_n\left(N_m^{1/2} L\, \boldsymbol{\theta}_k \right) \right| < \varepsilon/4. \tag{3.14}$$

For any $1 \leq i \leq n$, $1 \leq j \leq N$, and $\boldsymbol{\theta} \in \mathbb{R}^p$, define

$$\Omega_{ij}(\boldsymbol{\theta}) = \rho\left(e_{ij} + R_{nj} - N_m^{1/2} L\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right) - \rho\left(e_{ij} + R_{nj}\right) + \psi(e_{ij})\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}.$$

Using an argument similar to equation (3.13), we can prove that $\sup_{\|\boldsymbol{\theta}\|_2 \leq 1} |\Omega_{ij}(\boldsymbol{\theta})| = O(N_m^{-1})$, and consequently $\sup_{\|\boldsymbol{\theta}\|_2 \leq 1} |\Omega_{ij}(\boldsymbol{\theta}) - \mathbb{E}(\Omega_{ij}(\boldsymbol{\theta}))| = O(N_m^{-1})$. Using the previous equation, we have

$$\sup_{\|\boldsymbol{\theta}\|_2 \leq 1} \sum_{i=1}^{n} \mathrm{Var}\left(\frac{1}{N}\sum_{j=1}^{N} \Omega_{ij}(\boldsymbol{\theta})\right) = O\left(nN^{-1}N_m^{-1}\right). \tag{3.15}$$

Using the above three upper bounds and Bernstein's Inequality we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{\|\boldsymbol{\theta}\|_2 \leq 1} \frac{1}{N_m}\left|\Delta_n\left(N_m^{1/2} L\boldsymbol{\theta}\right)\right| \geq \varepsilon\right) &\leq \sum_{k=1}^{K_n} \mathbb{P}\left(\left|\Delta_n\left(N_m^{1/2} L\boldsymbol{\theta}_k\right)\right| \geq \frac{\varepsilon N_m}{2}\right) \\
&\leq \sum_{k=1}^{K_n} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N}\sum_{j=1}^{N}|\Omega_{ij}(\theta_k) - \mathbb{E}(\Omega_{ij}(\theta_k))|\right] \geq \frac{\varepsilon N_m}{2n}\right) \\
&\leq K_n \exp\left(-\frac{Cn\left(\varepsilon N_m/2n\right)^2}{N^{-1}N_m^{-1} + N_m^{-1}(\varepsilon N_m/2n)}\right) \\
&\leq K_n \exp\left(-\frac{C\varepsilon^2 N_m^2/n}{N^{-1}N_m^{-1} + \varepsilon n^{-1}}\right) \\
&\leq K_n \exp\left(-\frac{C\varepsilon^2 N N_m^3}{n + \varepsilon N N_m}\right) = o(1).
\end{aligned}$$

Therefore the Lemma 3 is proved. □

The second asymptotic bound we need is given by the following lemma.

Lemma 4. *Under the Assumptions (A1) - (A5) and for a fixed constant $L > 1$,*

$$\sup_{\|\boldsymbol{\theta}\|_2 < L}\left|N_m^{-1/2}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left[\psi(e_{ij})\cdot\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right]\right| = o_P(1). \tag{3.16}$$

Proof. Notice that by Assumption (A2) one has

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{N_m^{1/2}}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\left[\psi(e_{ij})\cdot\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right]\right) &\leq \frac{1}{N_m}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[\psi(e_{ij})^2\right]\left[\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right]^2\\ &\leq CN_m^{-1}\|\boldsymbol{\theta}\|_2.\end{aligned}$$

Using Tchebychev's Inequality, equation (3.16) is proved. □

The last asymptotic bound needed is given by the following lemma.

Lemma 5. *Under Assumptions (A1) - (A5) and for a fixed constant $L > 1$,*

$$\mathbb{P}\left(\inf_{\|\boldsymbol{\theta}\|_2=L}\left|\frac{1}{N_m}\mathbb{E}\left[\Gamma_n\left(N_m^{1/2}\boldsymbol{\theta}\right)\right]\right| > 0\right) \to 1.$$

Proof. By Lemma 2 we can assume that $\sup_{\|\boldsymbol{\theta}\|_2\leq L}\left(|R_{nj}| + N_m^{1/2}\left\|\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right\|_2^2\right) < C$. By Assumption (A4) we have

$$\begin{aligned}N_m^{-1}\mathbb{E}\left(\Gamma_n\left(N_m^{1/2}L\boldsymbol{\theta}\right)\right) &= N_m^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\int_{R_{nj}}^{R_{nj}-N_m^{1/2}\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}}\mathbb{E}\left(\psi(e_{ij}+u)\right)du\\ &= N_m^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\int_{R_{nj}}^{R_{nj}-N_m^{1/2}\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}}\delta\left(\frac{j}{N}\right)u+O(u^2)du\\ &= N_m^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\delta\left(\frac{j}{N}\right)\frac{1}{2}\left[\left(R_{nj}-N_m^{1/2}\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right)^2-R_{nj}^2\right]+o(1)\\ &= \sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\delta\left(\frac{j}{N}\right)\left[\frac{\left(\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\theta\right)^2}{2}-N_m^{-1/2}R_{nj}\boldsymbol{B}^{\mathrm{T}}\left(\frac{j}{N}\right)\theta\right]+o(1)\\ &\geq Cn\inf_{x\in\mathbb{R}}\delta(x)\|\boldsymbol{\theta}\|_2^2-\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}N_m^{-1/2}R_{nj}\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}+o(1)\\ &= CnL^2-CnL+o(1),\end{aligned}\tag{3.17}$$

which is positive for large enough L. This finishes the proof of the lemma. □

The following lemma is standard in the spline approximation theory and we omit the proof here.

Lemma 6 (Theorem 5.4.2, DeVore and Lorentz (1993))**.** *There is a constant $C_p > 0$, such that for any spline function $S(\cdot) = \sum_{J=1-p}^{N_m} \gamma_J B_J(\cdot)$ of order p, we have $C_p N_m^{-1} \|\boldsymbol{\gamma}\|_2^2 \leq \|S\|_2^2 \leq N_m^{-1} \|\boldsymbol{\gamma}\|_2^2$, where $\boldsymbol{\gamma} = (\gamma_{1-p}, \ldots, \gamma_{N_m})^{\mathrm{T}}$.*

Now, after presenting the required lemmas, and their proofs, we are ready to give the proof for Theorem 2.

Proof of Theorem 2. Combining Lemmas 3, 4 and 5, and using the convexity of $\rho(\cdot)$ we have

$$\mathbb{P}\left(\inf_{\|\boldsymbol{\theta}\|_2 \geq L} \frac{1}{N_m} \Gamma(N_m^{1/2}\,\boldsymbol{\theta}) > 0\right) = \mathbb{P}\left(\inf_{\|\boldsymbol{\theta}\|_2 = L} \frac{1}{N_m} \Gamma(N_m^{1/2}\,\boldsymbol{\theta}) > 0\right) \to 1.$$

This in turn implies

$$\mathbb{P}\left(\inf_{\|\boldsymbol{\theta}\|_2 \geq L N_m^{1/2}} \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \rho\left(e_{ij} + R_{nj} - N_m^{1/2} L_n \tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\boldsymbol{\theta}\right) > \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} \rho\left(e_{ij} + R_{nj}\right)\right) \to 1. \tag{3.18}$$

Define $\hat{\boldsymbol{\theta}} := \boldsymbol{S}_n^{-1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right) = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^{N_m + p}} \Gamma_n(\boldsymbol{\theta})$. By equation (3.18) one has $\|\hat{\boldsymbol{\theta}}\|_2 = O_P\left(N_m^{1/2}\right)$, and using Lemma 2, we obtain $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2^2 = O_P\left(n^{-1} N_m^2\right)$. The approximation property of B-Splines implies that $\|\hat{m}(\cdot) - \boldsymbol{B}(\cdot)^{\mathrm{T}}\boldsymbol{\beta}^*\|_2^2 = \|\boldsymbol{B}(\cdot)^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2^2 = O(N_m^{-1})\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2^2 = O_P\left(n^{-1} N_m\right)$, where the second-to-last equality comes from Lemma 6. Finally, using Lemma 1, and by Assumption (A2), we have

$$\|\hat{m} - m\|_2^2 = O_P\left(n^{-1} N_m + N_m^{-2p}\right), \tag{3.19}$$

which completes the proof of Theorem 2. □

Asymptotic Normality

Before presenting the second theorem, we need some additional notation. Let

$$\mathbb{W}_n = \frac{n}{N}\sum_{j=1}^{N} \delta\left(\frac{j}{N}\right) \boldsymbol{B}\left(\frac{j}{N}\right) \boldsymbol{B}^{\mathrm{T}}\left(\frac{j}{N}\right), \widetilde{\mathbb{W}}_n = \mathbb{S}_n^{-1/2}\mathbb{W}_n\mathbb{S}_n^{-1/2} \text{and}$$

$$\tilde{\boldsymbol{\theta}} = \widetilde{\mathbb{W}}_n^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\tilde{\boldsymbol{B}}\left(\frac{j}{N}\right)\psi(e_{ij}). \tag{3.20}$$

The second result answers the question of what is the asymptotic distribution of the B-Spline smoothed M-estimator. We prove that $\hat{m}(\cdot)$ converges to a normal distribution. Besides the convergence in distribution, this theorem provides an estimator for the variance of $\hat{m}(\cdot)$, which is of fundamental importance when constructing the robust simultaneous confidence band in the Section 3.3.3. Note also that the convergence provided in Theorem 3 is only true point-wise, that is, this result cannot be directly used to construct a simultaneous confidence band.

Theorem 3 (Asymptotic Normality)**.** *Under Assumptions (A1) - (A7) we have*

$$\frac{\widehat{m}(x) - m(x)}{\sqrt{D_n(x)}} \xrightarrow{d} N(0,1),\ 0 \leq x \leq 1, \tag{3.21}$$

where $D_n(x) = \boldsymbol{B}(x)^{\mathrm{T}}\mathbb{W}_n^{-1}\left(\sum_{i=1}^{n} N^{-2}\mathbb{B}^{\mathrm{T}}\mathbb{G}_i\mathbb{B}\right)\mathbb{W}_n^{-1}\boldsymbol{B}(x)$*, where* $\mathbb{G}_i$ *was defined in the Assumption A7 and* $\mathbb{B}$ *is the matrix with columns* $\boldsymbol{B}(j/N)$*,* $j = 1, \ldots, N$*, and* $\mathbb{W}_i$ *are defined in (3.20).*

Proof of Theorem 3

The first step in the proof of the Theorem 3 is to obtain an asymptotic upper bound on the difference between $\hat{\theta}$ and $\tilde{\theta}$, where $\tilde{\theta}$ is given in (3.20).

Lemma 7. *Under Assumptions (A1) - (A7),*

$$N_m^{-1/2}\|\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\|_2 = o_P(1). \tag{3.22}$$

Proof. By Assumption (A5), note that $\widetilde{\mathbb{W}}_n$ is invertible and, for all n, $\lambda_{\min}(\widetilde{\mathbb{W}}_n) > \tilde{\lambda}_0 > 0$ for some constant $\tilde{\lambda}_0$. We will use an argument similar to the proof of Theorem 2, and first show that, for any fixed $\varepsilon > 0$, $\mathbb{P}\left(\inf_{N_m^{-1/2}\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|_2 \geq \varepsilon} N_m^{-1}|\Gamma(\boldsymbol{\theta}) - \Gamma(\tilde{\boldsymbol{\theta}})| > 0\right) \to 1$. To prove this, using the convexity of $\rho(\cdot)$, we only need to show that

$$\mathbb{P}\left(\inf_{N_m^{-1/2}\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|_2 = \varepsilon} N_m^{-1}|\Gamma(\boldsymbol{\theta}) - \Gamma(\tilde{\boldsymbol{\theta}})| > 0;\ N_m^{-1/2}\|\tilde{\boldsymbol{\theta}}\|_2 < L\right) \to 1. \tag{3.23}$$

Using the equation (3.9) and the same argument to show the bound in equation (3.17), we have

$$\begin{aligned} N_m^{-1}\Gamma_n(\boldsymbol{\theta}) &= N_m^{-1}\left[\Delta_n(\theta) + \mathbb{E}(\Gamma_n(\theta)) - \sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\psi(e_{ij})\tilde{\boldsymbol{B}}^{\mathrm{T}}\left(\frac{j}{N}\right)\theta\right] \\ &= N_m^{-1}\left[\Delta_n(\boldsymbol{\theta}) + \frac{\boldsymbol{\theta}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\boldsymbol{\theta}}{2} - \tilde{\boldsymbol{\theta}}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\boldsymbol{\theta}\right] + o(1), \end{aligned} \tag{3.24}$$

using equation (3.17). Notice that $2\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\boldsymbol{\theta} = \boldsymbol{\theta}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\boldsymbol{\theta} + \tilde{\boldsymbol{\theta}}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\tilde{\boldsymbol{\theta}} - (\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}\widetilde{\mathbb{W}}_n(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})$. Substituting this into equation (3.24) we obtain

$$\frac{1}{N_m}\Gamma_n(\boldsymbol{\theta}) = \frac{1}{N_m}\left[\frac{(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}\widetilde{\mathbb{W}}_n(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})}{2} - \frac{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\tilde{\boldsymbol{\theta}}}{2} + \Delta_n(\boldsymbol{\theta})\right] + o(1). \tag{3.25}$$

In particular we have

$$\frac{1}{N_m}\Gamma_n(\tilde{\boldsymbol{\theta}}) = \frac{1}{N_m}\left[-\frac{\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\widetilde{\mathbb{W}}_n\tilde{\boldsymbol{\theta}}}{2} + \Delta_n(\tilde{\boldsymbol{\theta}})\right] + o(1). \tag{3.26}$$

Using Lemmas 2 and Assumption (A3) we have $\|\tilde{\boldsymbol{\theta}}\|_2 = O(N_m^{1/2})$, which implies that for a large enough constant $L > 0$, we can assume $N_m^{-1/2}\|\tilde{\boldsymbol{\theta}}\|_2 < L$. Note that if $N_m^{-1/2}\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|_2 = \varepsilon$ and $N_m^{-1/2}\|\tilde{\boldsymbol{\theta}}\|_2 < L$, then $N_m^{-1/2}\|\boldsymbol{\theta}\| \leq L + \varepsilon$. Subtracting, then, equation (3.26) from (3.25) we get

$$\begin{aligned}
& \inf_{N_m^{-1/2}\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|_2=\varepsilon, N_m^{-1/2}\|\tilde{\boldsymbol{\theta}}\|_2<L} \frac{1}{N_m}\left|\Gamma(\boldsymbol{\theta}) - \Gamma(\tilde{\boldsymbol{\theta}})\right| \\
= \;& N_m^{-1}\left[\frac{(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^{\mathrm{T}}\widetilde{\mathbb{W}}_n(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})}{2} + \Delta_n(\boldsymbol{\theta}) - \Delta_n(\tilde{\boldsymbol{\theta}})\right] + o(1) \\
\geq \;& \frac{\tilde{\lambda}_0\varepsilon^2}{2} - 2\sup_{N_m^{-1/2}\|\boldsymbol{\theta}\|_2 \leq L+\varepsilon} N_m^{-1}|\Delta_n(\boldsymbol{\theta})| + o(1) = \frac{\tilde{\lambda}\varepsilon^2}{2} + o(1),
\end{aligned}$$

where the last equality comes from Lemma 3. This proves equation (3.23) and implies that $N_m^{-1/2}\|\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\|_2 = o_P(1)$. This proves Lemma 7. □

After presenting the required lemmas, and their proofs, we are ready to give a proof for Theorem 3.

Proof of Theorem 3. By Lemmas 2 and 7, we have

$$\begin{aligned}
\left\|(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) - \boldsymbol{S}_n^{-1/2}\tilde{\boldsymbol{\theta}}\right\|_2 &= \left\|\boldsymbol{S}_n^{-1/2}\left[\boldsymbol{S}_n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) - \tilde{\boldsymbol{\theta}}\right]\right\|_2 \\
&= \left\|S_n^{-1/2}\right\|_2 \left\|\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right\|_2 = o_P\left(N_m n^{-1/2}\right),
\end{aligned}$$

which implies that, for any vector $\boldsymbol{\gamma} \in \mathbb{R}^{N_m+p}$, with $\|\boldsymbol{\gamma}\| \leq L$, for a fixed constant $L > 0$,

$$\begin{aligned}
\boldsymbol{\gamma}^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &= \boldsymbol{\gamma}^{\mathrm{T}}\mathbb{S}_n^{-1/2}\tilde{\boldsymbol{\theta}} + o_P\left(N_m n^{-1/2}\right) \\
&= \boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{B}\left(\frac{j}{N}\right)^{\mathrm{T}}\psi(e_{ij}) + o_P\left(N_m n^{-1/2}\right). \quad (3.27)
\end{aligned}$$

We can rewrite

$$\boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\sum_{i=1}^{n}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{B}\left(\frac{j}{N}\right)^{\mathrm{T}}\psi(e_{ij})=\sum_{i=1}^{n}\frac{1}{N}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\psi}(\boldsymbol{e}_i), \tag{3.28}$$

where $\boldsymbol{v}=\left(\boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\boldsymbol{B}\left(\frac{1}{N}\right),\cdots,\boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\boldsymbol{B}\left(\frac{N}{N}\right)\right)^{\mathrm{T}}$. Note also that $\mathrm{Var}\left(\sum_{i=1}^{n}N^{-1}\boldsymbol{v}^{\mathrm{T}}\psi(\boldsymbol{e_i})\right)=\sum_{i=1}^{n}N^{-2}\boldsymbol{v}^{\mathrm{T}}\mathbb{G}_i\boldsymbol{v}$. Using this calculation, we can rewrite equation (3.28) as $\sum_{i=1}^{n}a_i\xi_i$, where $a_i^2=N^{-2}\boldsymbol{v}^{\mathrm{T}}\mathbb{G}_i\boldsymbol{v}$, $1\leq i\leq n$, and $\{\xi_i\}_{i=1}^{n}$ are independent with mean zero and unit variance. Using the Lindeberg's central limit theorem, see Billingsley (2008), if $\max_{i=1,\ldots,n}a_i^2/\sum_{i=1}^{n}a_i^2=o(1)$, then $\sum_{i=1}^{n}a_i\xi_i/\sqrt{\sum_{i=1}^{n}a_i^2}$ converges in distribution to $N(0,1)$. We have

$$\begin{aligned}\max_{i=1,\ldots,n}a_i^2 &\leq \max_{i=1,\ldots,n}N^{-2}\|v\|_2^2\sum_{j=1}^{N}\mathbb{E}(\psi(e_{ij})^2)\\ &\leq CN^{-1}\sum_{j=1}^{N}\left(\boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\boldsymbol{B}\left(\frac{j}{N}\right)\right)^2=O(N_m^2n^{-2}),\end{aligned}$$

from Assumption (A4) and Lemma 2. We also have

$$\sum_{i=1}^{n}a_i^2=\sum_{i=1}^{n}N^{-2}\boldsymbol{v}^{\mathrm{T}}\mathbb{G}_i\boldsymbol{v}\geq N^{-1}\lambda_0\boldsymbol{\gamma}^{\mathrm{T}}\mathbb{W}_n^{-1}\mathbb{S}_n\mathbb{W}_n^{-1}\boldsymbol{\gamma}=O(N_mn^{-1}N^{-1}),$$

where the inequality comes from the Assumption (A7), and the last equation from Lemma 2. Collecting the previous bounds we have $\max_{i=1,\ldots,n}a_i^2/\sum_{i=1}^{n}a_i^2=O(N_mN/n)=o(1)$ due to Assumption (A1). This proves that the condition of Lindeberg's central limit theorem is satisfied. Setting $\boldsymbol{\gamma}=\boldsymbol{B}(x)$, $0\leq x\leq 1$, we obtain

$$\sum_{i=1}^{n}N^{-2}\boldsymbol{v}^{\mathrm{T}}\mathbb{G}_i\boldsymbol{v}=\boldsymbol{B}\left(x\right)^{\mathrm{T}}\mathbb{W}_n^{-1}\left(\sum_{i=1}^{n}N^{-2}\mathbb{B}^{\mathrm{T}}\mathbb{G}_i\mathbb{B}\right)\mathbb{W}_n^{-1}\boldsymbol{B}\left(x\right)=D_n(x),$$

and due to the Assumption (A1) and (3.27) we complete the proof of Theorem 3. □

3.3.3 Robust Simultaneous Confidence Band

In this section we propose a method to calculate a *robust simultaneous confidence band* (RSCB) for the mean function. Opposed to point-wise confidence intervals, for which the confidence level is maintained only for each individual point individually, a *simultaneous confidence band* (SCB) must maintain the confidence level uniformly for all the points, that is, the $1-\alpha$ SCB for the mean function, $(U(x), L(x))$ must satisfiy $\mathbb{P}\{L(x) \leq m(x) \leq U(x),\, \forall x \in [0,1]\} \geq 1-\alpha$.

Obtaining SCB for *homogeneous* functional datasets has been discussed in previous literature, such as Cao et al. (2012b). In that work, the SCB is obtained by first estimating the covariance function of the functional process, and then calculating the quantile of a Gaussian process with the same covariance structure. This procedure, though, is very sensitive to outliers, as discussed in the Section 3.4. The result of the Theorem 3 alone is not enough to provide a RSCB, but with the help of a modified pseudo-data transformation, we can translate the calculation of the RSCB to the simpler problem of obtaining a SCB for homogenous functional datasets.

Pseudo-data

The objective of this section is first to modify the dataset, possibly contaminated with outliers, to reduce the influence of any extreme observation, producing a more homogeneous dataset. We can then use this new dataset with a classical, non-robust method to calculate the SCB.

To accomplish the first objective, we can transform each curve in the original dataset into the new homogeneous curves using a modified version of the *pseudo-data*. This concept was first introduced in Cox (1983), where it was used to obtain asymptotic properties of M-based regression estimation for i.i.d. data by transforming the problem to the established properties of the non-robust least square regression of the

pseudo-data. We modify the definition of the pseudo-data to allow for heteroskedasticity of the random errors considered in our model.

Define the *pseudo-data* derived from the dataset Y_{ij} as

$$Z_{ij} = m\left(j/N\right) + \sqrt{2nD_n(j/N)}\psi\left(e_{ij}/\sqrt{2nD_n(j/N)}\right), i = 1, \ldots, n, j = 1, \ldots, N \tag{3.29}$$

where $\psi(\cdot)$, as defined before, is the derivative of loss function, and $D_n(\cdot)$ is the variance of $\hat{m}(\cdot)$ obtained from the Theorem 3.

To understand the multiplication by $\sqrt{2nD_n(x)}$ in (3.29), a simple example is helpful. Assume that the dataset $\{Y_{ij}\}_{i=1,\ldots,n,j=1,\ldots,N}$ is homogeneous, that is, assume that $e_{i\cdot}$ is i.i.d. and $\mathrm{Var}(e_{ij}) < \infty$, $i = 1, \ldots, n, j = 1, \ldots, N$. Also, consider the least square loss function, that is, $\rho_{LS}(x) = x^2, x \in \mathbb{R}$. Then, a direct calculation results in that $D_n(x)$ is the B-spline smoothing of $\{Y_{ij}\}_{i=1,\ldots,n,j=1,\ldots,N}$. That is, the pseudo-data defined in (3.29) is the original homogeneous dataset, $Z_{ij} \equiv Y_{ij}$, $i = 1, \ldots, n, j = 1, \ldots, N$.

In the general case, with outlier contaminated datasets, and robust loss functions, the pseudo-data Z_{ij} has a similar behavior as the original dataset, excluding the influence of the outlier curves. The extra multiplication by $\sqrt{2}$ in (3.29) was determined empirically and it is needed to make sure that, for each $j = 1, \ldots, N$, the variance of Z_{ij}, $i = 1, \ldots, n$ is similar to the variance of the original dataset, excluding the influence of the outlier curves. A simulation to support this is presented in the Section 3.4.

More recently, Oh et al. (2007) and Lim and Oh (2015) used the concept of pseudo-data to obtain simultaneous confidence bands for regression function with i.i.d. data. Here we will extend this idea to work with functional data and transform

the estimation of the RSCB for the mean function of contaminated data to the estimation of a SCB for homogeneous functional data.

More precisely, we will apply the estimation method for the SCB of functional datasets developed in Cao et al. (2012b) to the pseudo-dataset Z_{ij}, $1 \leq i \leq n$, $1 \leq j \leq N$. Since neither the mean function nor the random error are directly observable, we calculate the empirical pseudo-data as

$$\hat{Z}_{ij} = \hat{m}\left(j/N\right) + \sqrt{2nD_n(j/N)}\psi\left(\hat{e}_{ij}/\sqrt{2nD_n(j/N)}\right), i = 1, \ldots, n, j = 1, \ldots, N,$$

where $\hat{m}(\cdot)$ is the B-spline M-estimator of the mean function defined in the Section 3.3.1 and $\hat{e}_{ij}$ is the empirical estimator of the random error defined as $\hat{e}_{ij} = Y_{ij} - \hat{m}(j/N)$, $i = 1, \ldots, n$, $j = 1, \ldots, N$. The full method to estimate the RSCB is summarized in the Algorithm 1.

Algorithm 1: Robust Simultaneous Confidence Band for the mean function

---

Part 1: Pseudo-data preparation

1 Robust estimation of mean using B-Spline M-Estimator: $\hat{m}(\cdot)$;
2 Calculate asymptotic variance of $\hat{m}(\cdot)$: $D_n(x)$;
3 Evaluate the empirical random-error: $\hat{e}_{ij} = Y_{ij} - \hat{m}(x_{ij})$;
4 Calculate the empirical pseudo-data:
$\hat{Y}_{ij} = \hat{m}(x_{ij}) + \sqrt{2nD_n(x_{ij})}\psi\left(\hat{e}_{ij}/\sqrt{2nD_n(x_{ij})}\right)$;

Part 2: RSCB Construction

5 Estimate mean function using B-spline smooth least-square: $\hat{m}^{pd}$;
6 Estimate sample covariance function using B-spline smoothing: $\hat{G}^{pd}(\cdot, \cdot)$;
7 Calculate empirical quantile of Gaussian process with same covariance structure: $\hat{Q}^{pd}_{1-\alpha}$;
8 Construct Robust Simultaneous Confidence Band:

$$\hat{m}^{pd}(\cdot) \pm \hat{Q}^{pd}_{1-\alpha}\sqrt{\hat{G}^{pd}(\cdot, \cdot)/n}.$$

3.3.4 The RSCB for the difference of two mean functions

The framework proposed here to obtain a RSCB for the mean function can be extended to obtain a RSCB for the difference of the mean functions of two populations. Denote for $d = 1, 2$ the samples coming from each population, satisfying the model defined in (3.1)

$$Y_{dij} = m_d\left(\frac{j}{N}\right) + e_{dij}, 1 \leq i \leq n_d,\ 1 \leq j \leq N.$$

Define the ratio of two-sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1\to\infty} \hat{r} = r > 0$. For each group we can obtain the B-spline smooth M-estimator for the mean function as described in the Section 3.3.1.

Following the procedure in the Section 3.3.3, we can obtain empirical pseudo-samples $\hat{Z}_{dij}$ for each population, $d = 1, 2$. We can then use the empirical pseudo-samples to obtain the RSCB for the difference of the mean functions using a method for non-contaminated datasets. First, we obtain estimators for the covariance function of each group, $\hat{G}_d^{pd}(\cdot,\cdot)$, $d = 1, 2$, then proceed to obtain the empirical quantile, $\hat{Q}_{12,1-\alpha}$, of a Gaussian process having covariance structure defined by

$$\frac{\hat{G}_1^{pd}(x,x') + \hat{r}\hat{G}_2^{pd}(x,x')}{\left\{\hat{G}_1^{pd}(x,x) + \hat{r}\hat{G}_2^{pd}(x,x)\right\}^{1/2}\left\{\hat{G}_1^{pd}(x',x') + \hat{r}\hat{G}_2^{pd}(x',x')\right\}^{1/2}},$$

where $x, x' \in [0, 1]$. The RSCB for $m_1(x) - m_2(x)$ is then given as

$$(\hat{m}_1(x) - \hat{m}_2(x)) \pm n_1^{1/2}\left[\hat{G}_1^{pd}(x,x) + \hat{r}\hat{G}_2^{pd}(x,x)\right]^{1/2}\hat{Q}_{12,1-\alpha}.$$

The confidence band for the difference of the mean functions can be used to perform a hypothesis test of the form $H_0 : m_1(x) \equiv m_2(x), \forall x \in [0, 1]$ vs. H_A :

$m_1(x) \neq m_2(x), \exists\, x \in [0,1]$. The test can be performed by calculating the appropriate $(1-\alpha)\times 100\%$ RSCB and checking if the horizontal line $y = 0$ is fully contained in the RSCB. Although the *p-value* for the test cannot be calculated directly, it can be estimated by finding the largest α for which H_0 is rejected.

3.4 Simulation

In this section we perform a numerical study to analyze the performance of methods proposed in this paper. We investigate the consistency of the B-Spline M-estimator for the mean function and the empirical coverage and band area of the RSCB. We use the SCB (non-robust) proposed in Cao et al. (2012b) as a baseline for comparison. Since outlier curves often have different types of outlying behaviors in functional data, we consider several types of outliers in the assessment of the performance of the RSCB.

3.4.1 Simulation setting

Based on the model proposed in Cao et al. (2012b), we decompose the random component of the model (3.1) as

$$e_{ij} = \sum_{k=1}^{2} \xi_{ik}\phi_k\left(\frac{j}{N}\right) + \sigma\epsilon_{ij},\ 1 \leq j \leq N,\ 1 \leq i \leq n.$$

This decomposition corresponds to two different sources of randomness. The first component, $\sum_{k=1}^{2} \xi_{ik}\phi_k\,(j/N)$, corresponds to the intrinsic error of the functional data, while the second component, $\sigma\epsilon_{ij}$, corresponds to an additive random measurement error with a noise level σ.

For this simulation we generate ξ_{ik} for $k = 1, 2$ and ϵ_{ij} for $1 \leq j \leq N$, $1 \leq i \leq n$ from $N(0,1)$. The number of subjects is n, the number of observations

per curve is taken as $N = \lfloor 1.22n^{(2p-2)/2p+1}\log(n))\rfloor$ and the number of knots is $N_m = \lfloor n^{1/2p}\log(n)\rfloor$, where p, the order of the B-splines, is chosen as 4. The mean function, eigenvector functions and the noise level are taken as $m(x) = 10 + \sin\{2\pi(x-1/2)\}$, $\phi_1(x) = -2\cos\{(\pi(x-1/2)\}$, $\phi_2(x) = \sin\{\pi(x-1/2)\}$ and $\sigma = 0.5$, respectively.

Under this functional model we introduce outlier curves, Y_{ij}^o, to the generated functional sample in two different ways. In the first method we contaminate a subset, I_O, of the original functional sample. The contamination proportion varies from 0.00 to 0.20, at 0.05 increment. These outliers are described in items (1) to (6) below. The second type of contamination considers changing the distribution of the additive measurement error to a heavy-tailed distribution, described in (7) below.

1. *Peak Outliers.* To simulate an outlier with a punctual influence, each curve was contaminated at a single measurement point, j^*/N, by adding a random value s_i taken from a uniform distribution on $[-s_u, -s_l] \cup [s_l, s_u]$, that is,

$$Y_{ij^*}^o = Y_{ij^*} + s_i, i \in I_O,\ j^* = \lfloor 0.05N \rfloor.$$

   This produces a peak outlier curve with a *peak* at the point j^*/N. The parameters s_l and s_u controls the strength of outliers.

2. *Bump Outliers.* This type is an extension of the peak outliers and the contamination occurs in an interval, $[b_0, b_1]$, rather than at a single point, that is,

$$Y_{ij^*}^o = Y_{ij^*} + s_i, i \in I_O, j^*/N \in [b_0, b_1].$$

   In the simulation, the interval is chosen as $[b_0, b_1] = [0.5, 0.53]$.

3. *Step Outliers.* A further extension of the bump outliers is created by contaminating the curve in the interval $[c_i, 1]$, where c_i is randomly chosen from $[0.5, 1]$ for each outlying curve, that is,

$$Y^o_{ij^*} = Y_{ij^*} + s_i, i \in I_O, j^*/N \geq c_i.$$

4. *Mixture Normal–Laplace.* Similar to the previous type of outliers, but using a mixture of a normal distribution $N(0, \sigma^2)$ and a Laplace distribution with mean 0 and scale σ.

5. *Mixture Normal–Slash.* Similar to item (4), but using a mixture of a normal distribution $N(0, \sigma^2)$ and a Slash distribution with location 0 and scale σ.

6. *Mixture Normal–Cauchy.* Similar to item (4), but using a mixture of a normal distribution $N(0, \sigma^2)$ and a Cauchy distribution with location 0 and scale σ.

7. *Laplace Errors.* For this type of outlier, we replace the distribution of the measurement errors for all random curves ϵ_{ij} by a Laplace distribution with mean 0 and scale σ.

The *Huber loss function*, described in the Section 3.3, was used as loss function for the M-estimation. This choice was guided by the good performance of the Huber loss, added to its the well studied robust theoretical properties. These two qualities make the Huber loss function well suited for many of the robust estimation, see Huber et al. (1964). The choice of the parameter k was based on an empirical selection for maximizing the empirical coverage rate of the RSCB, while keep its total area small. For outlier types *1* to *6*, we used $k = 2.50$. For outlier type *7*, a value of $k = 3.5$ produced best results. A thorough analysis of data-based choices for k, such as a

cross-validation and generalized cross-validation, could also have been used, but it is beyond the scope of this work.

3.4.2 An illustrative example

We first illustrate the performance of the proposed RSCB using a toy example. We generate a functional sample of $n = 50$ curves from the model defined in the Section 3.4.1 and contaminate the data by using all types of outliers defined previously, using a contamination proportion of 20% for outlier types *1* to *6*, choosing $s_l = 10$ and $s_u = 20$ for outlier types *1* to *3*. We first construct the 95% confidence band using the proposed RSCB (black) and non-robust (red) SCB for the mean function for each outlier type. We also construct the non-robust SCB and RSCB for the mean function when the sample does not have outlier curves to assess the consistency of the proposed RSCB. Figure 3.1 depicts the effects of outlier curves on the non-robust SCB and RSCB methods.

The first graph (top left in Figure 3.1) for no outlier case shows that the proposed RSCB behaves the same as the non-robust SCB when there are no outlier curves in the data. For peak, bump and step outliers (center and right, first row, left second row in Figure 3.1) the width of the non-robust SCB is strongly deformed around the outlier location, resulting in a non-informative SCB. The RSCB is less influenced by the outliers, resulting in a SCB with similar characteristics to the SCB for clean dataset.

For the mixture outliers (center and right in second row, and left in third row in Figure 3.1) the influence of the outliers is most notable, with a strong deformation of the non-robust mean estimation and SCB for Normal-Slash and Normal-Cauchy mixtures. The RSCB is not much influenced by the presence of outliers, albeit a slight increase in the band width when compared to the clean dataset.

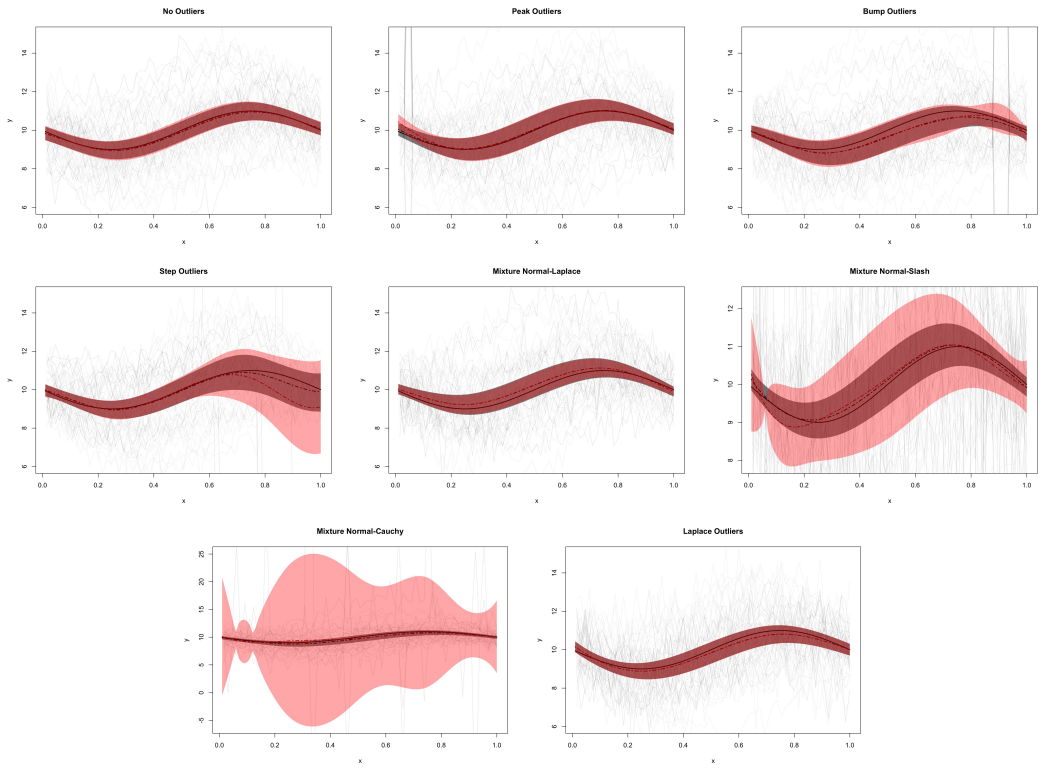

Figure 3.1: Comparison between RSCB (Black) and non-robust SCB (Red) for a simulated dataset.

For the pure Laplace outlier, (right third row in Figure 3.1) the resulting band is similar for the toy example, but as we will see in the Section 3.4.5, the RSCB performs better than the non robust SCB.

This illustrative example provides evidence that when there are outlier curves in a functional dataset, estimate of the mean function and non-robust SCB are both strongly affected while the proposed RSCB based on the R mean estimator performs well for different types of outlier curves.

3.4.3 Asymptotic consistency of the mean function estimator

To evaluate the performance of the B-spline M-Estimator for the estimation of the mean function, we generated functional sample from the model in the Section

3.4.1 for sample sizes $n = 50, 100$ and 200, with $s_l = 5$ and $s_u = 7$ for outlier types peak, bump and step. Each simulation is repeated 500 times.

The average L^2 distance between the real mean function and the B-Spline smoothed M-estimator was calculated. As a baseline comparison, the least square method used in Cao et al. (2012b) was also calculated. The results are presented in Tables 3.1-3.4.

The B-Spline Smoothed M-estimator shows a similar or faster convergence rate, as measured by the average L^2 distance between the estimator and the real mean function, than the non-robust method for all outliers, with a smaller or similar standard deviation. For localized outliers, the results are similar, but robust method shows better results for large contamination, as can be seen in Table 3.1, for *step* outliers, with $n = 200$ and 20% contamination proportion, the average L^2 distance between the robust estimator and the real mean function is 0.120, while the non-robust is 0.133. The standard deviation of the L^2 distance is also smaller for the robust (0.058), when compared to the non-robust (0.063). The improvement of the robust method is made clearer for mixtures of heavy tailed distributions, such as the mixture Normal-Slash and mixture Normal-Cauchy, when the convergence of the non-robust estimator is most influenced by the outlying curves. From Table 3.2, the average L^2 distance between the robust estimator and the real mean function for *mixture Normal-Cauchy* outliers, with $n = 200$ and 20% contamination is 0.100, with a standard deviation of 0.060, while for the non-robust the average L^2 distance is 0.915, with a standard deviation of 2.322, an increase of approximately 9 times for the average distance, and 38 times for the standard deviation. Notice also that the results for the robust case are very similar to the results for the *clean dataset*, Table 3.4, for which the average distance is 0.094 and the standard deviation 0.054. This

Table 3.1: $\|\hat{m} - m\|_2$ - Mean (SD) L^2 distance between the real mean function and the estimated mean function for contamination type outliers. Comparison between Robust (R) and Non-Robust (NR) methods.

Outlier Type	n	Method	Contamination Prop. *0.05*	*0.10*	*0.15*	*0.20*
Peak	50	R	0.200 (0.114)	0.193 (0.109)	0.193 (0.105)	0.196 (0.110)
		NR	0.198 (0.114)	0.189 (0.103)	0.192 (0.104)	0.199 (0.107)
	100	R	0.141 (0.081)	0.126 (0.076)	0.133 (0.078)	0.129 (0.072)
		NR	0.140 (0.083)	0.139 (0.079)	0.138 (0.079)	0.140 (0.076)
	200	R	0.094 (0.054)	0.099 (0.057)	0.103 (0.060)	0.098 (0.053)
		NR	0.097 (0.053)	0.097 (0.059)	0.096 (0.056)	0.092 (0.053)
Bump	50	R	0.200 (0.114)	0.195 (0.113)	0.200 (0.107)	0.200 (0.112)
		NR	0.200 (0.112)	0.197 (0.112)	0.202 (0.106)	0.202 (0.111)
	100	R	0.138 (0.081)	0.145 (0.082)	0.137 (0.076)	0.143 (0.080)
		NR	0.139 (0.081)	0.146 (0.082)	0.138 (0.076)	0.145 (0.079)
	200	R	0.096 (0.053)	0.098 (0.055)	0.099 (0.056)	0.102 (0.056)
		NR	0.096 (0.052)	0.099 (0.055)	0.100 (0.055)	0.103 (0.056)
Step	50	R	0.207 (0.113)	0.217 (0.110)	0.224 (0.114)	0.249 (0.119)
		NR	0.208 (0.109)	0.232 (0.114)	0.253 (0.127)	0.272 (0.134)
	100	R	0.146 (0.077)	0.148 (0.078)	0.166 (0.081)	0.176 (0.088)
		NR	0.153 (0.075)	0.165 (0.078)	0.174 (0.085)	0.187 (0.088)
	200	R	0.104 (0.052)	0.110 (0.055)	0.113 (0.053)	0.120 (0.058)
		NR	0.111 (0.056)	0.114 (0.052)	0.121 (0.055)	0.133 (0.063)

further indicates that the B-spline smoothed M-estimator is successfully diminishing the influence of the outliers in the estimation of the mean function.

3.4.4 Variance of Pseudo-data

The definition of the pseudo-data given in equation (3.29), that is,

$$\hat{Z}_{ij} = \hat{m}\left(j/N\right) + \sqrt{2nD_n(j/N)}\psi\left(\hat{e}_{ij}/\sqrt{2nD_n(j/N)}\right), i = 1, \ldots, n, j = 1, \ldots, N,$$

aims to produce a new (pseudo) dataset that has the same variance structure as the non-contaminated dataset.

Table 3.2: $\|\hat{m}-m\|_2$ - Mean (SD) L^2 distance between the real mean function and the estimated mean function for mixture model outliers. Comparison between Robust (R) and Non-Robust (NR) methods.

Outlier Type	n	Method	Contamination Prop. *0.05*	*0.10*	*0.15*	*0.20*
Mixture Normal-Laplace	50	R	0.200 (0.120)	0.196 (0.108)	0.192 (0.106)	0.191 (0.113)
		NR	0.202 (0.115)	0.194 (0.108)	0.198 (0.119)	0.193 (0.111)
	100	R	0.132 (0.077)	0.146 (0.084)	0.144 (0.082)	0.141 (0.083)
		NR	0.137 (0.079)	0.139 (0.080)	0.134 (0.078)	0.136 (0.081)
	200	R	0.093 (0.052)	0.095 (0.056)	0.099 (0.056)	0.092 (0.053)
		NR	0.096 (0.056)	0.099 (0.061)	0.096 (0.052)	0.094 (0.054)
Mixture Normal-Slash	50	R	0.192 (0.110)	0.193 (0.105)	0.193 (0.107)	0.200 (0.109)
		NR	0.362 (0.684)	1.128 (7.103)	1.796 (13.37)	5.940 (109.9)
	100	R	0.140 (0.081)	0.141 (0.078)	0.137 (0.078)	0.144 (0.082)
		NR	0.458 (2.896)	5.060 (60.09)	6.983 (98.27)	1.510 (5.676)
	200	R	0.095 (0.054)	0.101 (0.059)	0.095 (0.057)	0.100 (0.051)
		NR	0.545 (3.406)	0.616 (1.676)	0.948 (2.655)	1.602 (10.99)
Mixture Normal-Cauchy	50	R	0.199 (0.111)	0.200 (0.116)	0.194 (0.112)	0.205 (0.111)
		NR	0.534 (3.067)	0.529 (1.213)	1.820 (13.74)	1.147 (5.073)
	100	R	0.132 (0.080)	0.140 (0.077)	0.143 (0.079)	0.143 (0.079)
		NR	0.416 (1.615)	1.561 (15.47)	0.786 (2.189)	3.621 (56.44)
	200	R	0.095 (0.057)	0.095 (0.056)	0.097 (0.054)	0.100 (0.060)
		NR	0.750 (8.707)	0.633 (3.040)	0.823 (2.279)	0.915 (2.322)

Table 3.3: $\|\hat{m}-m\|_2$ - Mean (SD) L^2 distance between the real mean function and the estimated mean function for heavy tailed distribution. Comparison between Robust (R) and Non-Robust (NR) methods.

Outlier Type	Method	n *50*	*100*	*200*
Laplace Errors	R	0.193 (0.106)	0.133 (0.077)	0.097 (0.056)
	NR	0.201 (0.112)	0.131 (0.075)	0.097 (0.056)

Table 3.4: Comparison of Robust (R) and non-robust (NR) average L^2 distance between the real mean function and the estimated mean function for datasets with no outliers.

	Method	n *50*	*100*	*200*
Clean Dataset	R	0.195 (0.109)	0.135 (0.076)	0.094 (0.054)
	NR	0.198 (0.114)	0.128 (0.077)	0.098 (0.057)

In order to evaluate the efficacy of this method, we compare the sample variance of the pseudo-data with the real variance of the uncontaminated model defined in the Section 3.4.1. To further emphasize the influence of outliers in the calculation of the variance, we also compared the results with the sample variance of the outlier contaminated dataset, using the least squares method as the estimator for the mean function. We generate a functional dataset from the model in the Section 3.4.1 for sample size $n = 200$, with $s_l = 5$ and $s_u = 7$ for outlier types *1* to *3*. Each simulation is repeated 500 times. The results are presented in the Figure 3.2. We only present the results for Peak outliers, as all other have similar results.

The results show that the variance of the pseudo-data is very close to the real variance of the uncontaminated model, while the non-robust estimation of the variance of the outlier contaminated dataset is strongly affected by the outlier curves.

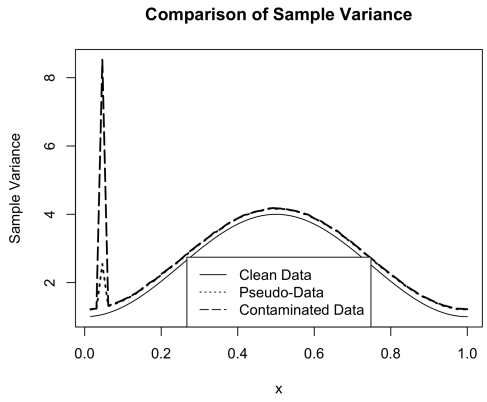

Figure 3.2: Comparison of the real variance (solid), sample variance of pseudo data (dashed) and sample variance of contaminated data (dash-dot). Peak outlier with 5% contamination.

Table 3.5: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for contamination type outliers. In parenthesis, the average area SCB.

Outlier Type	n	Method	Contamination Prop. *0.05*	*0.10*	*0.15*	*0.20*
Peak	50	R	0.902 (1.053)	0.924 (1.045)	0.890 (1.053)	0.880 (1.049)
		NR	0.898 (1.055)	0.912 (1.047)	0.878 (1.056)	0.866 (1.054)
	100	R	0.926 (0.743)	0.934 (0.748)	0.914 (0.746)	0.934 (0.748)
		NR	0.928 (0.743)	0.934 (0.749)	0.914 (0.747)	0.936 (0.750)
	200	R	0.952 (0.530)	0.938 (0.529)	0.904 (0.528)	0.958 (0.531)
		NR	0.954 (0.530)	0.938 (0.530)	0.904 (0.528)	0.958 (0.531)
Bump	50	R	0.910 (1.047)	0.876 (1.036)	0.884 (1.054)	0.884 (1.048)
		NR	0.906 (1.050)	0.870 (1.040)	0.872 (1.059)	0.872 (1.056)
	100	R	0.898 (0.745)	0.886 (0.743)	0.908 (0.746)	0.874 (0.744)
		NR	0.894 (0.747)	0.886 (0.747)	0.916 (0.751)	0.876 (0.751)
	200	R	0.938 (0.529)	0.914 (0.529)	0.902 (0.530)	0.876 (0.529)
		NR	0.930 (0.530)	0.898 (0.531)	0.886 (0.533)	0.838 (0.533)
Step	50	R	0.888 (1.080)	0.878 (1.165)	0.892 (1.232)	0.878 (1.324)
		NR	0.906 (1.101)	0.896 (1.223)	0.874 (1.278)	0.868 (1.357)
	100	R	0.872 (0.772)	0.924 (0.832)	0.894 (0.899)	0.890 (0.943)
		NR	0.914 (0.794)	0.932 (0.873)	0.918 (0.928)	0.910 (0.965)
	200	R	0.902 (0.545)	0.910 (0.592)	0.914 (0.637)	0.908 (0.669)
		NR	0.890 (0.560)	0.926 (0.625)	0.936 (0.659)	0.928 (0.686)

3.4.5 Simulation of the SCB for the mean function and the difference of two mean functions

Case I: SCB for the mean function, $m(x)$

To evaluate the performance of the proposed RSCB method for the mean function, we calculate the empirical coverage rate. We generate functional samples from the model in the Section 3.4.1 for sample sizes $n = 50, 100$ and 200, with $s_l = 5$ and $s_u = 7$ for outlier types *1* to *3*. Each simulation is repeated 500 times.

The empirical coverage rates for contamination proportions varying from 0.05 to 0.20 are presented in Tables 3.4.5 to 3.8 The results for clean datasets are similar for robust and non-robust methods, with the empirical coverage for both methods approaching 95% and the area of RSCB smaller than the area of non-robust SCB.

Table 3.6: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for mixture model types of outliers. In parenthesis, the average area of SCB.

Outlier Type	n	Method	Contamination Prop. *0.05*	*0.10*	*0.15*	*0.20*
Mixture Normal-Laplace	50	R	0.890 (1.055)	0.896 (1.050)	0.920 (1.050)	0.920 (1.054)
		NR	0.906 (1.048)	0.894 (1.040)	0.910 (1.046)	0.908 (1.048)
	100	R	0.944 (0.748)	0.906 (0.746)	0.926 (0.748)	0.916 (0.747)
		NR	0.930 (0.748)	0.934 (0.748)	0.918 (0.748)	0.906 (0.744)
	200	R	0.956 (0.530)	0.942 (0.531)	0.928 (0.530)	0.948 (0.531)
		NR	0.940 (0.532)	0.924 (0.531)	0.944 (0.531)	0.946 (0.530)
Mixture Normal-Slash	50	R	0.908 (1.049)	0.888 (1.042)	0.890 (1.046)	0.822 (1.042)
		NR	0.676 (2.397)	0.470 (4.326)	0.394 (5.299)	0.338 (6.205)
	100	R	0.920 (0.745)	0.904 (0.744)	0.912 (0.740)	0.870 (0.739)
		NR	0.604 (2.805)	0.436 (10.192)	0.384 (5.332)	0.292 (7.718)
	200	R	0.944 (0.529)	0.910 (0.526)	0.918 (0.525)	0.888 (0.523)
		NR	0.490 (2.613)	0.386 (3.929)	0.300 (5.919)	0.326 (7.606)
Mixture Normal-Cauchy	50	R	0.912 (1.047)	0.856 (1.046)	0.884 (1.043)	0.862 (1.042)
		NR	0.680 (2.137)	0.556 (3.063)	0.442 (4.958)	0.348 (5.208)
	100	R	0.920 (0.744)	0.906 (0.746)	0.910 (0.740)	0.910 (0.737)
		NR	0.630 (4.167)	0.440 (5.504)	0.366 (4.401)	0.348 (18.69)
	200	R	0.912 (0.528)	0.944 (0.529)	0.926 (0.526)	0.924 (0.524)
		NR	0.494 (2.446)	0.414 (3.552)	0.304 (4.417)	0.318 (5.617)

Table 3.7: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for heavy tailed distribution. In parenthesis, the average area of SCB.

Outlier Type	Method	n *50*	*100*	*200*
Laplace Errors	R	0.904 (1.052)	0.924 (0.749)	0.936 (0.532)
	NR	0.896 (1.052)	0.916 (0.748)	0.930 (0.532)

Table 3.8: Comparison of Robust (R) and non-robust (NR) empirical coverage rates of 95% SCB for datasets with no outliers. In parenthesis, the average area of SCB.

	Method	n *50*	*100*	*200*
Clean Dataset	R	0.922 (1.046)	0.932 (0.746)	0.946 (0.529)
	NR	0.906 (1.050)	0.922 (0.747)	0.922 (0.530)

For outlier contaminated datasets, the advantage of the RSCB becomes clear, with a breakdown point of around 20%, while the non-robust SCB has a breakdown point at 5% to 10%. The precision in the RSCB is also greater, with the area of the RSCB smaller than the non-robust SCB. Peak outliers produce similar empirical convergence rate for large sample size, but for $n = 50$, the advantage of the RSCB is apparent. For the less localized outliers, Bump and Step, the robust method has better empirical coverage and maintains the area of the RSCB reasonably constant among contamination levels. The non-robust method shows a quicker decay of the empirical coverage, for bump and at the same time has wider SCB. The mixture models more clearly shows the better results of the robust method, with the heavy-tailed mixture models, Slash and Cauchy, showing the significant advantage of the RSCB over the non-robust SCB. The outliers with a heavy-tailed distributions exhibits the most extreme difference between the non-robust and robust SCBs, while the precision of the RSCB is kept at the same level as all other outlier and the non-contaminated dataset. This provides strong evidence that the proposed RSCB is less sensitive to the presence of outliers in the dataset, maintaining both a good confidence level and precision.

Case II: SCB for the difference of two mean functions, $m_1(x) - m_2(x)$

We also conducted a simulation to evaluate the performance of the RSCB method for the difference between two mean functions, by testing the hypotheses described in the Section 3.3.4,

$$H_0 : m_1(x) = m_2(x),\ \forall x \in [0,1] \text{ vs. } H_A : m_1(x) \neq m_2(x), \exists\, x \in [0,1]. \tag{3.30}$$

We employ the same model in the Section 3.4.1 for the one sample case. In this simulation setup, $n_1 = 100$, and $n_2 = 130$ correspond to the sample sizes for the first

and the second population, respectively, $N = 100$ are the number of measurement points for both samples, and outlier curves are introduced to the first population. The number of knots used for each population is the same as in the Section 3.4.1, that is, $N_{m,i} = \lfloor n_i^{1/2p} \log(n) \rfloor$, $i = 1, 2$. For the first population, we selected the parameter for the Huber loss as $k_1 = 2.50$, and for the second, $k_2 = 1.345$.

The results of the simulation are presented in Table 3.9 for the all types of outliers. For peak outliers the empirical type I error for the robust method is kept close to the nominal value, $\alpha = 5\%$, decreasing with the rise in the contamination proportion. For the non-robust test, the empirical type I error is much smaller than the nominal value, which indicates that the non-robust SCB is over-estimated. For less localized outliers, Step and Bump, the robust method produces an empirical type I error closer to the nominal value then the non-robust method for all contamination proportions. For the mixture models the results for the heavy tailed distributions, Slash and Cauchy, highlight the most the advantage of the robust method. The non-robust method has empirical type I error equals to 1.0, while the robust method keeps the empirical type I error similar to the non-contaminated dataset results. For the Laplace error type, the non-robust test repeat the previous results, with a empirical type I error much smaller than the nominal value.

3.5 Applications

We illustrate our approach on two datasets: Octane dataset for the one sample case and Ground level Ozone concentration dataset for the two sample case.

3.5.1 Octane dataset

This dataset is the same used in the Section 2.5, it consists of 39 near infrared (NIR) spectra of gasoline sample, obtained from Esbensen et al. (1996). It is known

Table 3.9: Empirical Type I Error. Nominal Type I Error set at $\alpha = 0.05$. Hypothesis test for the difference of mean functions. Sample size fixed at $n_1 = 100$, $n_2 = 130$, with $N = 100$ measurement points for both groups.

Outlier Type	Method	*Contamination Prop.* *0.05*	*0.10*	*0.15*	*0.20*
Peak	R	0.044	0.038	0.032	0.036
	NR	0.036	0.030	0.024	0.032
Bump	R	0.050	0.044	0.050	0.052
	NR	0.040	0.034	0.042	0.046
Step	R	0.064	0.050	0.066	0.098
	NR	0.044	0.060	0.082	0.096
Mixture Laplace	R	0.024	0.040	0.040	0.044
	NR	0.024	0.028	0.042	0.030
Mixture Slash	R	0.038	0.056	0.046	0.042
	NR	1.000	1.000	1.000	1.000
Mixture Cauchy	R	0.030	0.042	0.058	0.062
	NR	1.000	1.000	1.000	1.000

Outlier Type	Method	Empirical Error
Laplace Errors	R	0.046
	NR	0.032

	Method	Empirical Error
Clean Dataset	R	0.048
	NR	0.052

that 6 of the samples contain added ethanol, which corresponds to an upward translation on the upper wavelength, 1390 onward, interval of the spectrum. This is considered as the step outliers described in the Section 3.4.

The robust estimation of the mean and the 95% RSCB are calculated for this dataset, using $k = 1.65$ as the empirically chosen parameter for the Huber loss, as well as the mean estimator and confidence band using the method in Cao et al. (2012b). The results are presented in Figure 3.3, showing the full spectrum measure (left panel) and magnified on the second half of the spectrum to display the differences more apparently between the non-robust and robust SCBs (right panel).

We observe that the robust mean estimator remains close to the non-outlying curves, while the non-robust estimate of the mean function is heavily influenced by the outliers, resulting in an upward shift. The non-robust SCB is also heavily influenced by the outliers, translating in a very wide band on the second half of the spectrum. However the proposed RSCB shows a smaller variation of the width across the spectrum.

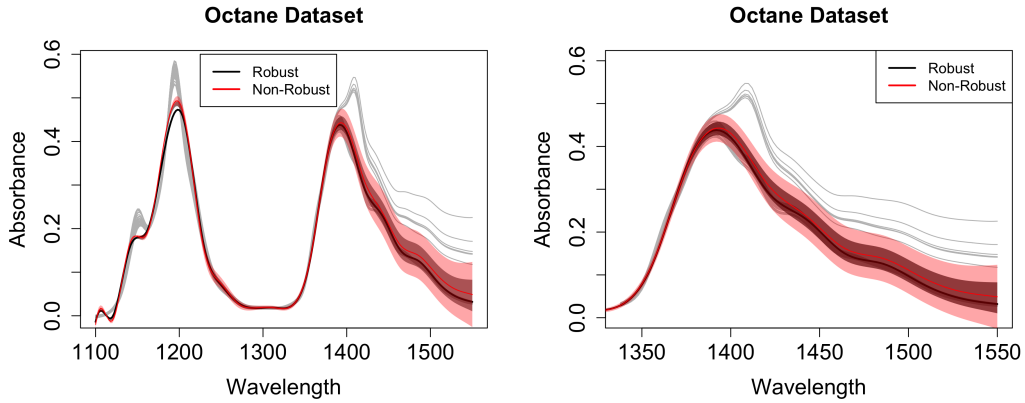

Figure 3.3: 95% SCB comparison for the octane dataset. Non-Robust (Red) vs. Robust (Black) methods. Left: full spectrum. Right: magnified on the second half of the spectrum.

3.5.2 Ground level Ozone concentration dataset

This dataset is the same used in the Section 2.5, and it consists of hourly average measurements of ground level ozone (O_3) concentrations from a monitoring station in Richmond, BC, Canada, from the years of 2004 to 2012. The presence of Ozone at ground level is highly undesirable, and considered a serious air pollutant. Since the concentration of ground level Ozone typically peaks at summer months, only the month of August is analyzed, resulting in 31 samples, with 24 measurement points for each sample.

The plot of the ground level O_3 concentration for years 2005 and 2007 is presented in Figure 3.4, top panel, with the year of 2005 in gray/black, and the year of 2007 in red. The outliers detected by Boente and Salibian-Barrera (2015) are highlighted.

We set up our hypotheses for testing if there is a difference between the ozone mean functions of the years 2005 and 2007 in Richmond, Canada. The outliers in the dataset are similar to the bump outliers described in the Section 3.4, but they are asymmetrical, localized only in the upper portion of the dataset. We chose the Huber loss parameter empirically as $k_1 = 2.0$ for the 2005 sample and $k_2 = 1.45$ for the 2007 sample. The 95% SCB of the difference between the mean functions of the ground level O_3 concentration in years of 2005 and 2007 is presented in the bottom left panel of Figure 3.4. We also calculate the 95% SCB for the difference between the mean functions with the outliers kept for the RSCB, and excluding the outliers for the non-robust SCB. This is presented in the bottom right panel of Figure 3.4. This plot provides a comparison of the SCB between the robust and non-robust methods. The RSCB has a much smaller width around the location of the outliers, but due to the asymmetrical disposition of the outlying curves, the estimated difference of mean functions is shifted slightly upwards.

Notice that the robust method does not reject the null hypothesis at a significance level $\alpha = 0.05$, while the non-robust test rejects the null hypothesis (Figure 3.4 bottom left). The result for the non-robust test is contradictory with the result of the hypothesis test using the dataset with outlier curves removed, which does not reject the null hypothesis at $\alpha = 0.05$ (Figure 3.4 bottom right).

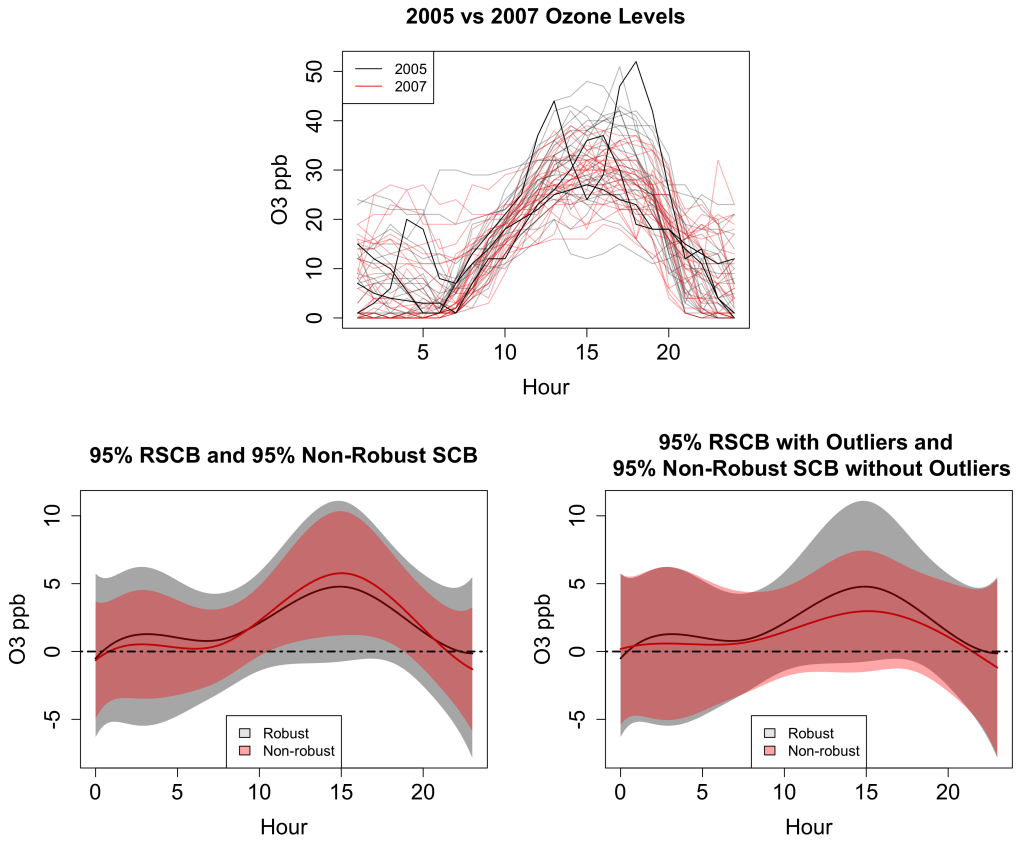

Figure 3.4: Top: O_3 Levels in years of 2005 (Gray and black) and 2007 (Red) in Richmond. Black lines are the outliers which are determined in Boente and Salibian-Barrera (2015). Bottom Left: 95% non-robust SCB (Red) and RSCB (Gray) for the difference between the mean functions of the two years. Bottom Right: 95% non-robust SCB (Red) and RSCB (Gray) for the difference between the mean functions of the two years, keeping outliers for RSCB, excluding outliers for non-robust SCB.

3.6 Conclusion

In this chapter we presented a new method of robustly estimating the mean function of the functional data, using B-Spline Smoothing and M-Based estimator. We proved that the proposed estimator is consistent and asymptotically normal. We also presented a new method to obtain a RSCB for the mean function, using a modified Pseudo-data. Additionally, we presented an extension of the RSCB for the difference of two mean functions, which can be used as a test statistic for the hypothesis test of the difference of the mean function of two populations.

A thorough simulation study was performed, showing that the proposed method performs well under the presence of the seven different types of outliers, which are typical types of contamination in the robust functional data analysis.

Chapter 4

Conclusion and Discussion

In this dissertation we presented two methods to obtain a robust estimator and robust simultaneous confidence band for the mean function of functional data. In Chapter 2, we proposed an estimation estimation procedure for the mean function based on the Least Absolute Deviation (LAD) and the B-Spline smoothing techniques. We proposed a robust estimator the mean function of outlier contaminated functional datasets by extending the *LAD* estimator to functional data analysis, and using *B-Spline smoothing*. We also proposed a robust estimator for the covariance function of contaminated functional datasets, using a three-step procedure. First, we used the *spherical principal components* to robustly estimate the eigenfunctions of the covariance function. Second, we robustly estimated the eigenvalues of the covariance function by using a robust estimator of the variance, the Huber's M-estimator, of the projection of the centralized data onto each robustly estimated eigenfunction. Finally, we reconstruct the covariance function using the estimated eigenfunctions and eigenvalues. Using the two previous estimators, we also proposed a method for constructing a *robust simultaneous confidence band* for the mean function of non-homogenous functional datasets. We first used a Monte-Carlo simulation to estimate the quantile of the absolute maximum of a Gaussian process with the same covariance function as the dataset, using the proposed robust estimation of the covariance function. Second, a correction factor was implemented, using empirical simulations, to better improve the efficiency of the RSCB. Finally, the RSCB was calculated using the robust mean estimator, the estimated empirical quantile, the robust estimator of

the covariance function and the empirically estimated correction factor. The robust methods was also extended for calculating a robust estimator for the difference of the mean function of two functional populations, and for calculating a RSCB for the difference of the mean function of two functional populations. Based on the RSCB, we proposed a robust test statistics for the hypothesis that the mean function of two populations are different.

A simulation study was performed, in the Section 2.4, and the performance of the LAD-based RSCB was compared with a non-robust SCB based on the least squares. The results of the simulation showed that the RSCB outperforms the non-robust SCB by a large margin for different types of simulated outliers. We also performed a simulation study to compare the robust and non-robust empirical type I error of the hypothesis test for the difference of the mean function of two functional populations. Similar to the results for the one-sample SCB, the robust test statistic outperforms the non-robust test statistic. The robust test kept the empirical type I error close to the nominal value for large contamination proportion of outliers, between 20% and 45%, depending on the type of outliers, while the non-robust test resulted in a empirical type I close to 1.0 for small contamination proportions, between 5% and 10%, depending on the type of outliers. We also compared the robust and non-robust SCB for two real datasets, the octane dataset for comparing one-sample robust and non-robust SCB, and the ground level O_3 concentration dataset for comparing the two-sample robust and non-robust SCB, and the robust and non-robust test statistics for hypothesis test for the difference of the mean functions. The results were inline with the simulation results, with the robust methods performing better then non-robust methods.

We also proved that the robust LAD-based B-spline smoothed estimator is asymptotically consistent, in the Section 2.3.3.

In Chapter 3, we proposed an extension of the LAD-based robust methods from Chapter 2. We first proposed the B-spline Smoothed M-estimator for the mean function of functional data. We researched the asymptotic theoretical properties of the proposed estimator, obtaining its asymptotic consistency and asymptotic normality, with full proofs provided in the Section 3.3.2. We also obtained a asymptotically correct estimation for the variance of the B-spline smoothed M-estimator of the mean function, in Theorem 3. Using the estimated variance of the mean function estimator, in the Section 3.3.3 we proposed a modification of the *pseudo-data* idea, obtaining a pseudo-dataset that diminishes that influence of the outliers on the covariance function. We also proposed a robust simultaneous confidence interval for the mean function of outlier contaminated datasets, by using the pseudo-data and previous non-robust methods for constructing the SCB for the mean function of homogenous functional datasets. We also extended the robust methods for constructing the RSCB for the difference of the mean functions of two outlier contaminated functional populations. We also extended the two-sample RSCB to obtaining a robust test statistic for the hypothesis of the difference of the mean functions of two non-homogenous functional populations.

A simulation study was performed, presented in the Section 3.4, using several types of outliers, with local contamination models, mixture models and contamination with pure heavy-tailed distributions. We evaluated the convergence of the B-spline smoothed M-estimator, and compared the results with the non-robust least square estimator. We calculated the average and the standard deviation of the L^2 distance between the estimator produced by each method and the real mean function. The results of the simulation showed that the robust estimator is better, or similar, to the non-robust estimator for all different types of outliers considered. We also evaluated the proposed pseudo-data definition, comparing the sample variance

of the pseudo-data, with the sample variance of the contaminated dataset, and with the variance of the clean dataset. The results showed that the pseudo-data limits the influence of the outliers in the estimation of the variance. We also performed a simulation comparing the RSCB with the non-robust SCB. The results showed the advantage of using the RSCB over the non-robust SCB. The robust method performed better than the non-robust method for the outliers considered, with the mixture models for outliers showing a great advantage for the RSCB over the non-robust SCB. We also performed a simulation comparing the empirical type I error for the hypothesis test of the difference for the mean functions of two heterogeneous functional populations. The results were inline with the one-sample SCB. For the robust test statistic, the empirical type I error was kept close to the nominal value, while the empirical type I error for the non-robust test statistic was deteriorated, specially for the mixture model type of outliers. We also used the same real datasets from the Chapter 2, namely, the octane dataset and the ground level O_3 concentration dataset, to evaluate the performance of the RSCB and robust test statistic, respectively. The results, when compared with the non-robust equivalent methods, show the advantage of the proposed robust methods.

When comparing the LAD-based method proposed in the Chapter 2 and M-based method proposed in the Chapter 3, the results of a numerical simulation performed, but not included in this dissertation, indicates that the M-based method produces more precise results, in exchange for a slightly smaller confidence. The results showed that the RSCB produced by the LAD-based method are, in average, twice as large as the RSCB produced by the M-based method. This trade-off between the accuracy and confidence level might be used to decide between the use of the LAD-based method and the M-based method, depending the functional dataset being considered, and the results desired.

References

Adrover, J., Salibian-Barrera, M., and Zamar, R. (2004). Globally robust inference for the location and simple linear regression models. *Journal of statistical planning and inference*, 119(2):353–375.

Bali, J. L., Boente, G., Tyler, D. E., Wang, J.-L., et al. (2011). Robust functional principal components: A projection-pursuit approach. *Annals of Statistics*, 39(6):2852–2882.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Boente, G., Barrera, M. S., and Tyler, D. E. (2014). A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis*, 131:254–264.

Boente, G. and Salibian-Barrera, M. (2015). S-estimators for functional principal component analysis. *Journal of the American Statistical Association*, 110(511):1100–1111.

Cao, G., Wang, J., Wang, L., and Todem, D. (2012a). Spline confidence bands for functional derivatives. *Journal of Statistical Planning and Inference*, 142(6):1557–1570.

Cao, G., Yang, L., and Todem, D. (2012b). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2):359–377.

Cox, D. D. (1983). Asymptotics for m-type smoothing splines. *Annals of Statistics*, pages 530–551.

Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., and Walczak, B. (2007). Robust statistics in data analysis—a review: basic concepts. *Chemometrics and Intelligent Laboratory Systems*, 85(2):203–219.

De Boor, C. (1978). A practical guide to splines. *Mathematics of Computation*.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation*, volume 303. Springer Science & Business Media.

Embling, C. B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K. C., and Scott, B. E. (2012). Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. *Journal of Applied Ecology*, 49(2):481–492.

Esbensen, K., Schönkopf, S., Midtgaard, T., and Guyot, D. (1996). *Multivariate Analysis in Practice: A Training Package*. Camo As.

Febrero, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19(4):331–345.

Ferraty, F., Rabhi, A., and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic "el niño" phenomenon. *Sankhyā: The Indian Journal of Statistics*, 67(2):378–398.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368.

Fraiman, R., Yohai, V. J., and Zamar, R. H. (2001). Optimal robust M-estimates of location. *Annals of statistics*, 29(1):194–223.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions*, volume 114. John Wiley & Sons.

Haque, M. E. and Khan, J. A. (2012). Globally robust confidence intervals for location. *Dhaka University Journal of Science*, 60(1):109–113.

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

Huber, P. and Ronchetti, E. (2011). *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley.

Huber, P. J. et al. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *Annals of Statistics*, 35(1):92–108.

Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99(4):813–832.

Lee, S., Shin, H., and Billor, N. (2013). M-type smoothing spline estimators for principal functions. *Computational Statistics & Data Analysis*, 66:89–100.

Lim, Y. and Oh, H.-S. (2015). Simultaneous confidence interval for quantile regression. *Computational Statistics*, 30(2):345–358.

Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., Cohen, K., Boente, G., Fraiman, R., Brumback, B., Croux, C., et al. (1999). Robust principal component analysis for functional data. *Test*, 8(1):1–73.

Maronna, R. A. and Yohai, V. J. (2013). Robust functional linear regression based on splines. *Computational Statistics & Data Analysis*, 65:46–55.

Oh, H.-S., Nychka, D. W., and Lee, T. C. (2007). The role of pseudo data for robust smoothing with application to wavelet regression. *Biometrika*.

Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge University Press.

Shin, H. and Lee, S. (2016). An RKHS approach to robust functional linear regression. *Statistica Sinica*, 26:255–272.

Silverman, B. and Ramsay, J. (2005). *Functional Data Analysis*. Springer-Verlag New York, 2 edition.

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

Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of california boys and girls from birth to eighteen years. *Publications in child development. University of California, Berkeley*, 1(2):183.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Wang, H. J., Zhu, Z., and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *Annals of Statistics*, 37(6b):3841–3866.

Wei, Y. and He, X. (2006). Conditional growth charts. *Annals of Statistics*, 34(5):2069–2097.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. Academic Press.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.