PREDICTION OF REMAINING LIFETIME DISTRIBUTION FROM FUNCTIONAL TRAJECTORIES BASED ON CENSORED OBSERVATIONS

By
IZZET SOZUCOK

Presented to Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
August 2019

Supervising Committee:
Shan Sun-Mitchell Supervising Professor
Andrzej Korzeniowski
Suvra Pal
Jonghyun Yun

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Shan Sun Mitchell for the continuous support of my Ph.D study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Andrzej Korzeniowski Dr. Suvra Pal and Dr. Jonghyun Yun for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I would especially like to thank my family. My wife, Emine has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point. My children, Mehmet Akif, Rana have continually provided the requisite breaks from philosophy and the motivation to finish my degree with expediency. My parents, Sadik and Raziye, deserve special thanks for their continued support and encouragement.

ABSTRACT

Prediction of Remaining Lifetime Distribution from Functional Trajectories Under Censoring Data

Izzet Sozucok Ph.D.

The University of Texas at Arlington, 2019

Supervising Professor: Shan Sun Mitchell

The goal in functional data studies on failure time or on death time of the objects is to find a relationship between age-at-death (failure time) and current values of a functional predictors. In this study, a novel technique is applied to predict the failure time of devices (such as bearings in a mechanical system) and to try to predict the "age-at-death" distributions under censoring data. We concern ourselves with circumstances where all co-variate trajectories are observed until a current time t. The predictors observed up to current time can be shown by time-varying principal component scores which is continuously updated as time progresses. We establish the estimation of modified survival function for longitudinal trajectories by inspiring Kaplan-Meire method in order to predict mean residual life distribution. Projecting behavior of co-variate trajectories on single index we reduce their dimension to get predictions for each individual object. Furthermore, the uniform convergence rate is proved for mean and co-variance function for censored functional data based on some specified conditions. The proposed method is validated as the leave-one- out method and the approach is illustrated using the simulation study as well

Contents

Abstract

The goal in functional data studies on failure time or on death time of the objects is to find a relationship between age-at-death (failure time) and current values of a functional predictors.In this study, a novel technique is applied to predict the failure time of devices (such as bearings in a mechanical system) and to try to predict the "age-at-death" distributions under censoring data. We concern ourselves with circumstances where all co-variate trajectories are observed until a current time t. The predictors observed up to current time can be shown by time-varying principal component scores which is continuously updated as time progresses. We establish the estimation of modified survival function for longitudinal trajectories by inspiring Kaplan-Meire method in order to predict mean residual life distribution.Projecting the behavior of co-variate trajectories on single index we reduce their dimension to get predictions for each individual object. Furthermore, the uniform convergence rate is proved for mean and co-variance function for censored functional data based on some specified condiotions. The proposed method is validated as the leave-one-out method and the approach is illustrated using the simulation study as well.

1. INTRODUCTION

In functional data analysis, measurements obtained usually depend on time-co-variate and a time-to-event for each subject. The relationship between these co-variate and remaining life is of interest in bio-demography in engineering systems and competing risk studies. The data obtained in those fields can be uncensored (complete), censored and sparse/fragmented (incomplete) signals. While the studies in biology and in bio-demography characterize the relationship between longitudinal co-variate and time-to-event, the studies in engineering systems deal with gradual and irreversible accumulation of damages that occurs during a system's life cycle. A classical framework to identify the association between functional predictors and time-to-event is the proportional hazards regression model [1].Using this model with current information of functional predictors, the hazard rate is estimated. However, our model in this paper depends on the entire event history as obtained by the co-variate trajectory, and not just on current information levels. On the other hand in engineering systems when the co-variate trajectories are assumed as degradation signal under observed condition-based signal, the evolution of some manifestation can be monitored using sensor technology. Some example of degradation signals include vibration signals in order to monitor excessive wear in rotating machinery, acoustic emission to monitor crack propagation, temperature changes and oil debris for monitoring engine lubrication etc. The main goal of these signals analysis is to estimate remaining lifetime distribution to keep safe of the systems.

There is a significant number of research on mean residual function and remaining life distribution; for example biologically reasonable study was demonstrated in [2],where a parametric model summarizing egg-laying trajectories of female med-flies (Mediterranean fruit fly, Ceralititis capitata) was shown to define remaining egg-laying potential.Therefore, a connection between the entire egg-laying trajectory up to current time and remaining life time was obtained by using parametric model. In some degradation studies, the models used to characterize the evolution of sensor-based degradation signals are the parametric models.Some common approaches are to model with random coefficients [3, 4, 5]and[6, 7]. The most of these researches rely on a sample of uncensored data (completed data).Here we mean that a completed data is continuously observed data which captures the trajectories from initial time to failed time as seem in Figure1(a).In contrast to the researches with uncensored data, in practice we have many different kinds of data to be analyzed such as sparse/fragmented data,censored data. For example, random censorship, in brief, means that X is not always completely observable but is restricted to the form $X = min(Y_1, Y_2, ..., Y_k)$, where $Y_1, Y_2, ..., Y_k$ are independent non-negative random variables as depicted in Figure1(d) . The model arises from many practical situations such as medical follow-up studies competing risk [8, 9],[10, 11, 12]. Another example for different data in applications consisting of relatively static structures such as bridge's degradation usually takes places very slowly (several tens years). Since the system is relatively static it suffices to observe the degradation process at intermittent discrete time points. The result is a sparsely observed degradation signal.On the contrary, the application such as turbine, generator and degradation machine cannot be reasonable assessed by sparse measurements.In naval maritime applications, powered generating units of aircraft are removed tested for a short period of time, and put pack into operation, That result is collection of fragmented degradation signals as depicted in Figure1(b)[13].

In this study, we develop a functional model that applies to incomplete data as well as complete degradation signal by using methods and tools from functional data analysis. Our main goal here is to predict the remaining lifetime distributions of non-parametric predictors. and especially mean remaining lifetime by extracting information from the available co-variate trajectories. We assume that the trajectories follow as non-parametric setting. Other model approaches assume that trajectories follow a Brownian motion process [14, 15] or a Gaussian Process with known co-variance structure[12, 7] The co-variance function is decomposed ussing the Karhunen-Loeve decomposition[16] and estimated by

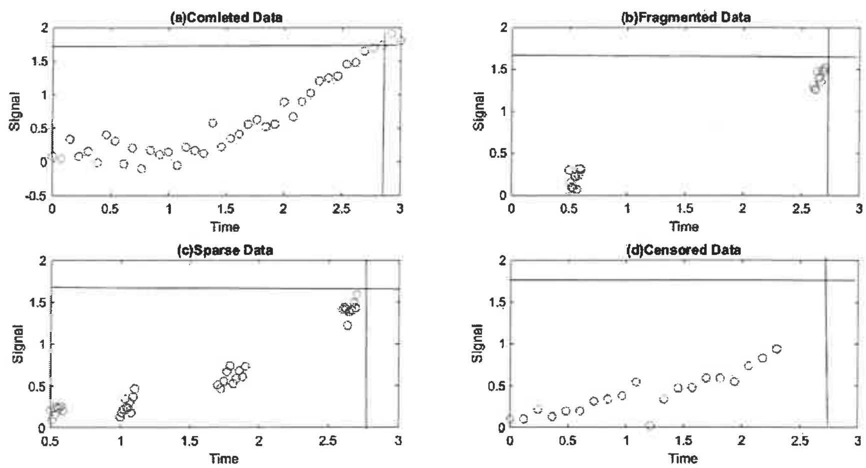

Figure 1: The Examples of Different Data

using the functional Principle Component Analysis(FPCA) method indroduced by Yao, Muller and Wang (2005).W.J Hall and Jon A. Wellner estimated mean Residual life on whole half line and as well as variance of the limiting process[17]

In non-parametric model, one condition for accurate estimation of the mean and covariance functions is that age-at-death process is densely observed throughout its support.However,in many applications where the trajectories are incompletely sampled,not all trajectories are observed up to the point of failure. Consequently, the trajectories are commonly under-sampled close to the upper bound of its support. To overcome this issue, Zhou,Serban and Geraeel[13] introduced a nonuniform sampling procedure for collecting incomplete data such as sparse data and fragmented data. For censoring data, we introduce a proposition which helps us to deal with none dense domain of incomplete trajectories by converting the unobserved data from non-parametric model.

We aim two main subjects in this study; first,the method we introduce is applicable from incomplete data to complete data. This will allow the estimation of quantiles and prediction intervals for remaining lifetimes, which are highly desirable for survival analysis. We evaluate the performance of our methodology using a growth data set taken

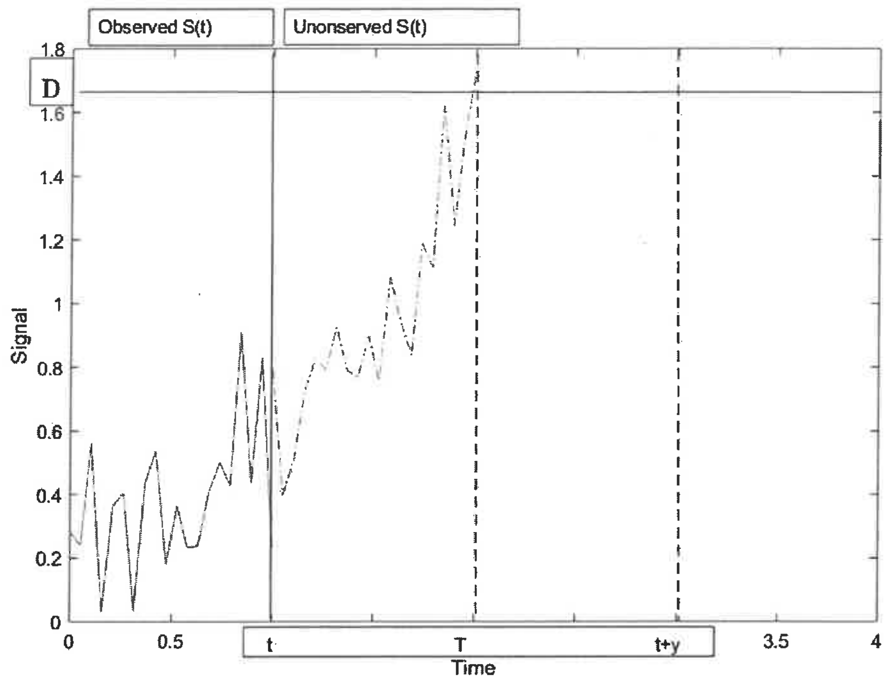

Figure 2: The Examples of An Independent Subject

from the engineering system.After assessment of the accuracy of the estimation of the remaining lifetime for incomplete and complete data,it seems that there is no a significant difference between both types of data. Second, the predictor trajectories for each subject are only observed until failure time and follow either uniform or nonuniform sampling procedures.At each given time only a random number of subjects is still alive for who the co-variate trajectories can be observed as depicted in Figure2.

The article is organized as follows: In Chapter 2, we introduce assumptions, the model and how we estimate mean residual life function by giving a new methods result in survival outcomes.The estimation of component of the model is presented in Section 3. We give the uniform convergence rate for mean and co-variance function in Chapter 4 In Chapter 5, we discuss the performance of our method by applying the the degradation signal under censored data taken in engineering system and simulated degradation signals, followed by discussion and concluding remarks in Section5.

2. Proposed Model.

2.1. Introduction of Co-variate Trajectories

We denote lifetime or age-at-time of subject by T and observed trajectories $X_i(s_{ij})$ for $j = 1, ..., m_i$(the number of observation time points for individual i) and $i = 1, ..., n$ (number of subjects), where$\{s_{ij}\}_{j=1,...,m_i}$ are observation time points in a bounded time domain $[0, M]$ for trajectory i. However, the available data for estimation of residual life distribution is not $(T_i)_{i=1,...,n}$ but the list of pairs $(\hat{T}_i, c_i)$, where $c_i \geq 0$ is a fixed deterministic censoring time of i and $\hat{T}_i = min(T_i, c_i)$. Then $(X_i(.), \hat{T}_i)'s$ are available data for each subject, where $X_i(.)$ is co-variate trajectory with domain $[0.T]$. We will show the trajectory for subjects that is still alive at time t by $\hat{X}(t, s), 0 \leq s \leq t$.

$$\hat{X}(t,s) = \begin{cases} X(s) & ,\hat{T} \geq t, 0 \leq s \leq t, \\ unobserved & ,\hat{T} < t \end{cases}$$

We decompose the co-variate trajectories as

$$\hat{X}_i(t,s) = \mu(s,t) + S_i(s,t) + \sigma\epsilon_i(s,t) \tag{2.1}$$

Where for $t \in [0, M]$ with some large define M, we define $\mu(s,t) = E\{X(s)|T > t\}$ is assumed to be fixed, but unknown, and $S_i(s,t)$ represents the random deviation from the underlying trend. We also assume $S_i(s,t)$ and $\epsilon_i(s,t)$ are independent.In this study, we assume that $\epsilon_i(s,t)$ is standard normal distribution.σ is deviance of $\epsilon_i(s,t)$

Let's define the eigenfunctions or principal component functions of the conditional covariance as solutions of the eigenequations are given as follows

$$\int_0^t cov[\{X(s), X(u)\}|T > t]\rho_{it}(u)du = \lambda_{jt}(s)\rho_j t(s) \tag{2.2}$$

where $\lambda_{1t} \geq \lambda_{2t} \geq ... \geq 0$ are eigenvalues and $\rho_{1t}(.), \rho_{2t}(.), ..., \rho_{mt}(.)$ are orthonormal eigenfunctions associated with these eigenvalues. Then ,one has the representation for $0 \leq s_1, s_2 \leq t$

$$cov(s_1, s_2|T > t) = \sum_{j=1}^{\infty} \lambda_{jt}\rho_{jt}(s_1)\rho_{jt}(s_2) \quad (2.3)$$

where

$$cov(s_1, s_2|T > t) = cov[\{X(s_1), X(s_2)\}|T > t] = cov(\hat{X}(s_1, t) - \mu(s_1, t), \hat{X}(s_2, t) - \mu(s_2, t))$$

Therefore, the observed trajectories $\hat{X}(s, t)$ for individuals with $T > t$ can be decomposed by the Karhunen–Loève extension [16].

$$\hat{X}(s, t) = \mu(s, t) + \sum_{j=1}^{\infty} \xi_{jt}\rho_{jt(s)}, 0 \leq s \leq t, t \geq 0 \quad (2.4)$$

where $\xi_{it}$are random scores , they are uncorrelated random effect with $E(\xi_{it}) = 0$ and $E(\xi_{it}^2) = \lambda_i$

The decomposition in equation (2.4) is infinite sum. Normally only a small number of eigenvalues are commonly significantly nonzero. For the eigenvalues which approximately zero the corresponding scores will also be approximately zero.Consequently, we will use a truncated version of this decomposition.Thus,

$$\hat{X}(s, t) = \mu(s, t) + \sum_{j=1}^{K} \xi_{it}\rho_{jt}(s), 0 \leq s \leq t, t \geq 0 \quad (2.5)$$

where K is the number of significantly nonzero eigenvalues. We select K to minimize the modified Akaike Criterion defined by Yao, Muller, Wang[18]. In the statistical terminology, this method is called Functional principle Component Analysis(FPCA).The main source for FPCA is seen by Ramsay-Silverman[19]. On the other hand, Yao, Muller and Wang [18] derived theoretical results for model parameter consistency, asymptotic (n large) distribution result under the assumption that the scores follow a normal distribution.

The number N_t of trajectories $\hat{X}(s,t)$ observable up to time t is random. Assuming $\overline{F}(t) = P(T > t)$, we have $N_t \sim Binomial(n, \overline{F}(t))$ where n is total number of subjects.Doneting by risk set at time t, $R(t) = \{i : T_i > t\}$ then $\hat{X}_i(s,t) \sim \hat{X}(s,t)$ for all $i \in R(t)$. In survival analysis, the remaining lifetime function at t is

$$e(t) = E(T - t|0 \leq t \leq T) \tag{2.6}$$

and the corresponding distribution function of remaining lifetime at y, where $y > 0$, is

$$F_t(y) = P(T - t < y|0 \leq t \leq T) \tag{2.7}$$

so that $e(t) = \int_t^\infty y dF_t(y)$. It is well known (Cox,1972) that the corresponding survival $\overline{F}(.)$ and hazard $\lambda(.)$ functions are

$$\overline{F}(t) = \frac{e(0)}{e(t)} exp(-\int_0^t \frac{1}{e(u)} du)$$

and

$$\lambda(t) = \{\frac{d}{dt} e(t) + 1\}/e(t)$$

2.2. Modeling Mean Remaining Lifetime for Censoring Data

Our aim is to relate the remaining life time T-t for given arbitrary $t \in [0, M]$ to the observed trajectory $\hat{X}$ in $[0, t]$ which may complete or incomplete that is to estimate

$$e_{\hat{X}}(t) = E(T - t|\hat{X}(s,t), 0 \leq s \leq t, 0 \leq t \leq T) \tag{2.8}$$

and corresponding distribution function of remaining lifetime

$$F_{\hat{X}(t),t}(y) = P(T - t < y|\hat{X}(s,t), 0 \leq s \leq t, 0 \leq t \leq T). \tag{2.9}$$

Let's define modified Survival function as $\overline{F}_{\hat{X}(t)}(t) = P(T > t|\hat{X}(s,t))$

we get the mean remaining lifetime function as the following;

$$\begin{aligned} e_{\hat{X}}(t) &= \int_t^{\infty} (y-t) dF_{\hat{X}(t)}(y) \\ &= \int_t^{\infty} \int_t^{y} du dF_{\hat{X}(t)}(y) \\ &= \int_t^{\infty} \int_t^{y} du d\frac{P(T < y|\hat{X}(t))}{P(T > t|\hat{X}(t))} \end{aligned} \tag{2.10}$$

When we apply Fubini's Theorem to the integrals above, we obtain $e_{\hat{X}(t)}$ as

$$\begin{aligned} e_{\hat{X}}(t) &= \frac{\int_t^{\infty} \int_u^{\infty} dF_{\hat{X}(t)}(y) du}{\overline{F}_{\hat{x}(t)}(t)} \\ &= \frac{\int_t^{\infty} (1 - F_{\hat{X}(t)}(u)) du}{\overline{F}_{\hat{X}(t)}(t)} \\ &= \frac{\int_t^{\infty} (\overline{F}_{\hat{X}(t)}(u)) du}{\overline{F}_{\hat{X}(t)}(t)} \end{aligned} \tag{2.11}$$

To predict mean residual lifetime function, we need to estimate at least one of probabilities; $P(T > t|\hat{X}(t))$ or $P(T < t|\hat{X}(t))$. However , for censoring data we don't have $(T_i)_{i=1,\dots,n}$ but the list of pairs $(\hat{T}_i, c_i)$. To overcome this issue we give the proposition below.

Proposition 2.1

If censoring time c_k of event k exceeds t ($c_k \geq$t) then

$\hat{T}_k = t$ hold true if only if $T_k = t$,

$\hat{T}_k \geq t$ hold true if only if $T_k \geq t$,

Let k be such that $c_k > t$, then $P(T > t|\hat{X}(t)) = P(\hat{T} > t|\hat{X}(t)) = \overline{F}_{\hat{X}(t)}(t)$ The Proposition 2.1 helps the residual lifetime function transit from the unknown random T variables to be able to estimate probability.Therefore, we need to estimate at least one of the probabilities;$P(\hat{T} > t|\hat{X}(t))$ or $P(\hat{T} < t|\hat{X}(t))$ which is similar distribution (2.9). Thus,random scores ξ_{it} are key random variable which were assumed as normal distribution by Zhou,Serban,Gebraeel [13] with mean $E(\xi_{it}) = 0$ and variance $E(\xi_{it}^2) = \lambda_i$ and they estimated the distribution in (2.9) under given assumption A1,A2. as

$$P(T - t \leq y|\hat{X}(s,t), T \geq t) = \frac{\Phi_Z(g(y|t)) - \Phi_Z(g(0|t))}{1 - \Phi_Z(g(0|t))}, \tag{2.12}$$

where Φ_Z represents the standard normal cumulative distribution function and $g(y|t) = \frac{\mu(t+y)-D}{\sqrt{V_t(t+y)}}$, D is threshold for degradation signal and

$$\mu_t(t+y) = \mu(t+y) + (Cd)'p(t+y)$$

$$V_t(t+y) = \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}[C_{k_1,k_2}\rho_{k1}(t+y)\rho_{k_2}(t+y)]$$

In the above equations,$p(t+y) = (\rho_1(t+y), ..., \rho_K(t+y))'$ and C_{k_1,k_2} refers to the (k1,k2) element of matrix C.

Under the assumptions, A1,A2, We can obtain the distribution we need to estimate in equation (2.11) as $P(\hat{T} \geq y|\hat{X}(t)) = 1 - \Phi(g(y|t)$.

On the other hand, If we choose random scores as $\xi_{jt} = \int_0^t \hat{X}(t,s) - \mu(s,t))\rho_j t(s)ds$, then we can estimate mean remaining lifetime in(2.8) as follows

$$r_{\hat{X}(t)}(t) = E(\hat{T} - t|\hat{X}(s,t), 0 \leq s \leq t, 0 \leq t \leq \hat{T})$$

and corresponding distribution

$$F_{\hat{X}(t),t}(y) = P(\hat{T} - t < y|\hat{X}(s,t), 0 \leq s \leq t, 0 \leq t \leq \hat{T})$$

as the follows:

We assume that there exists a family of smooth link functions h_t with $h_t(s) = H(s,t)$: $[0,M]x[0,M] \to R$ for a function H that is continuous in s and t associated evaluation function $\beta(s,t)$ satisfying $\beta \in L^2(C_t)$, $C_t = \{(s,t).0 \leq s \leq t, 0 \leq t \leq M\}$ such that

$$r_{\hat{X}}(t) = h_t(\int_0^t \hat{X}(s,t)\beta(s,t)ds)$$

This assumption puts mean remaining lifetime function into the framework of an extension of function regression [19, 20, 6].

For given t and orthonormal basis $\psi_{jt}(.), j = 1, 2, .., onL^2[0,t]$, the evaluation function $\beta(.,t)$ can represented by $\beta(s,t) = \sum_{j=1}^{\infty} \psi_{jt}\phi_{jt}(s), 0 \leq s \leq t, 0 \leq t \leq M$ with varying coefficient β_{jt}. A special choice for the basis are the eigenfunctions ρ_{jt} of $cov(X(s_1), X(s_2))$. Then we obtain remaining lifetime

$$r_{\hat{X}}(t) = h_t(r_0(t) + \int_0^t (\hat{X}(s,t) - \mu(s,t))\beta(s,t)ds)$$

where $r_0(t)$ is nonrandom function ,so we can introduce another link function $g_t(z(t)) = r_0(t) + z(t)$. We get

$$r_{\hat{X}}(t) = g_t(\int_0^t (\hat{X}(s,t) - \mu(s,t))\beta(s,t)ds).$$

I f $\hat{X}(.,t)$ and $\beta(.,t)$ are expressed in terms of the same orthonormal basis $\rho_{1t}, \rho_{2t}, ...$,and assuming the link function as identity function, we obtain remaining lifetime as

$$r_{\hat{X}(t)}(t) = \sum_{1}^{K} \xi_{jt}\beta_{jt} \tag{2.13}$$

where K is a finite number of component such that the trajectories $\hat{X}(.)$ can spanned by first K eigenfunctions and

$$\beta_{jt} = \int_0^t \beta(s,t)\rho_j(s)ds.$$

Having summarized the co-variate trajectories $\hat{X}(.,t)$ by linear predictor function $r_{\hat{X}}(.)$, we assume that the linear predictor function determines the conditional distribution

$$\hat{F}_{\hat{X}}(y) = P(\hat{T} - t \leq y|\hat{X}(s,t)) = P(\hat{T} - t \leq y|r_{\hat{X}}(t)) \tag{2.14}$$

Hence estimating conditional remaining life time distributions then is equivalent to estimating function $e_X(t)$.

3. Estimating the Model Components

3.1. Preliminaries

We use local polynomial kernel regression for shooting purposes to estimate $E(y|X = x)$. Given data$\{(x_i, y_i) \in R^2, i = 1, ..., n\}$ and Let $B = \{b_0, b_1, ..., b_p\}$ minimize

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}K(\frac{x-x_i}{h})\{y_i - b_0-, ..., -b_p(x-x_i)^p\}^2$$

where K is non-negative kernel function, h is a convenient bandwidth. Assuming the invertibility of $P_x^{'}W_xP_x$, standard weighted least squares theory leads to the solution

$$\hat{B} = (P_x^{'}W_xP_x)^{-1}P_x^{'}W_xY \tag{3.1}$$

where $Y = \{y_1, ...y_2\}$, is a vector of response, $P_x = \begin{bmatrix} 1 & x_1 - x & \cdots & (x_1-x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & \cdots & (x_n - x)^p \end{bmatrix}$ is an $nx(p+1)$ design matrix and

$W_x = dig\{K(\frac{x_1-x}{h}), K(\frac{x_2-x}{h}), ..., K(\frac{x_n-x}{h})\}$ is nxn diagonal matrix of weights[21] When we specify $\hat{B}$ as $p = 1$, we obtain that the non-parametric regression estimate is

$$\begin{aligned} m\{x_i; (x_i, y_i)_{i=1,..,n}; h\} &= \hat{E}(y|X = x) = \hat{b_0} \\ &= n^{-1}\sum_{i}^{n}\frac{\{\hat{s}_2(x;h) - \hat{s}_1(x;h)(x_i - x)\}K_h(x_1 - x)y_i}{\hat{s}_2(x;h)\hat{s}_0(x;h) - \hat{s}_1(x;h)^2} \end{aligned} \tag{3.2}$$

where $\hat{s}_r(x;h) = n^{-1}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}(x_i - x)^r K_h(x_i - x)$.

Our first step is to estimate the function $\mu(.,.)$ in (2.1) at current time t for all subject who are at risk at t. The estimate of $\mu(s,t)$ is

$$\hat{\mu}(s,t) = m[s; (t_{ij}, \hat{X}_i(t_{ij}, t)_{i\in R(t), j\in(j:0\leq t_{ij}\leq t}; h,] \tag{3.3}$$

where R(t) is the risk set at time t. the t_{ij}' s are the pooled time points of all observation. Second step is that the co-variance surface is estimated using the demeaned data $\hat{X}_i(s,t) - \hat{\mu}(s,t)$. The raw co-variances is defined by $R_i^t(s_{ij}, s_{il}) = Cov[(X_i(s_{ij},t) - \hat{\mu}(s_{ij},t), (X_i(s_{il},t) - \hat{\mu}(s_{il},t))]$ which can be expressed as following;

$$R^t(s,s') = \sum_{j=1}^{\infty} = \lambda_{jt}\rho_{jt}(s)\rho_{jt}(s') \tag{3.4}$$

where eigen-function and eigenvalues are solutions of estimated eigen-equations,

$$\int_0^t R^t(r,s)\rho_{jt}(r)dr = \lambda_{jt}\rho_{jt}(s) \tag{3.5}$$

with the constraints $\int_0^t \rho_{jt}(s)^2 ds = 1$ and$\int_0^t \rho_{jt}(s)\rho_{lt}(s)ds = 0$, for $j < 0$. we obtain these solutions by discretizing (3.4). Details can be found in Yao et al[22].After then we estimate the functional principal component scores are then determined by

$$\hat{\xi}_{ijt} = \int_0^t (\hat{X}_i(s,t) - \hat{\mu}(s,t))\hat{\rho}_{jt}(s)ds \tag{3.6}$$

Here the random scores can be determined by using numerical integration.Consistency results for $\hat{\rho}_{jt}, \hat{\mu}, \hat{\xi}_{ijt}$ can be found by P.Hall andM.Hossein-Hassab(2003).

3.2. Estimating Mean Remaining Lifetime

We purpose to estimate remaining lifetime for incomplete data in (2.11),so we use the least squares estimates of $\beta_t = (\beta_{0t}, \beta_{1t}, ..., \beta_{Kt})$ in model (2.13),

$$\hat{\beta}_t = argmin_{\beta_t} \sum_{i \in R(t)} \{T_i - t - (\beta_{0t} + \sum_{j=1}^{K} \hat{\xi}_{ijt}\beta_{jt})\}^2$$

the fitted model for expected value $r_{\hat{X}}(t)$ is obtained as;

$$\hat{r}_{\hat{X}(t)}(t) = \hat{\beta}_0(t) + \sum_{1}^{K} \hat{\xi}_{jt}\hat{\beta}_{jt} \tag{3.7}$$

where β_{0t} the mean remaining lifetime function.Finally, we estimate the conditional distribution in (2.14) after $r_{\hat{X}}(t)$ is estimated via (3.6) ,using kernel polynomial regression for smooth estimate. Consider i.i.d pairs $\{(X_1, Y_2), ..., (X_n, Y_n)\}$ such that $(X, Y) \in R^2$. The conditional distribution $F(y|x) = P(Y \leq y|X = x)$ from this sample equals to $E(I(Y_i \leq y)|X = x)$,where is I indicator function.Therefore,we can estimate it as regression problem by using the kernel polynomial regression for p=0 in (3.1),then we obtain the Nadara-Watson kernel,

$$E(Y|X) = \sum_{i=1}^{n} \frac{Y_i K_h((x_i - x)}{\sum_i^n K_h(x_i - x)}$$

When we apply the conditional distribution we get

$$\hat{F}(y|X = x) = \sum_{i:Y_i \leq y} \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \tag{3.8}$$

where $K_h(.) = \frac{1}{h}K(\frac{\cdot}{h})$ for a bandwidth h. Combining estimate (3.7) with (2.14),

$$\begin{aligned} \hat{F}(y|\hat{X}(s,t)) &= \hat{P}(\hat{T} - t \leq y|\hat{r}(t)) \\ &= \frac{\sum_{i \in \{i:\hat{T}_i - t \leq y\} \cap R(t)} K_h(r - \hat{r}_i(t))}{\sum_{j \in R(t)} K_h(r - \hat{r}_j(t))} \end{aligned} \tag{3.9}$$

Where $\hat{r}_i(t)$ (3.7) is the linear predictor for the ith individual and R(t) the risk set at time t. Here we always use the Epanechnikov kernels for estimating the conditional distribution. To estimate conditional density function for more details ,see Yu and Jones,(1998),[17]. Finally, we estimate mean remaining lifetime by using its complement for estimator of modified survival function $\hat{\overline{F}}_{\hat{X}(t)}(y) = 1 - \hat{P}(\hat{T} - t \leq y|\hat{r}(t))$ in order to obtain mean residual life function in(2.11) as

$$\hat{e}_{\hat{X}}(t) = \frac{\int_t^{\infty}(\hat{\overline{F}}_{\hat{X}(t)}(u))du}{\hat{\overline{F}}_{\hat{X}(t)}(t)}, \tag{3.10}$$

where we use the numerical integration to calculate the integral.

4. Uniform Convergence Rates of The Mean and Covariance Function for Censored Functional Data

4.1. Introduction

In this section, we consider the convergence rate of mean $\mu(s,t) = E\{X(s)|T > t\}$ and co-variance function $R^t(s_1, s_2) = cov[\{X(s_1), X(s_2)\}|T > t]$ as we introduced in section 2.1. Strong uniform convergence rates are developed for estimator which are local-linear smooths.We obtain the result in unified/non-unified framework where the number of observation may depend on the sample size.We show that the convergence rate depends on both of the number observations and sample size on each trajectory.Also, in sparse data this rate is equivalent to optimal non-parametric regression rate.Many recent scientist focused on the non-parametric estimation in order to model mean and co-variance estimations.Some of such work includes Ramsay and Silverman (2005)[19],Lin and Lee (2006), Hall, Muller and Wang (2006) and Yehua Li and Tailen Hsing(2010)[23].On other hand, The studies on kernel smoothing [Yao,Muller and Wang(2005a)[24], Hall,Muller and Wang(2006)].

In the Section 4.2 we review the model and data structure as well as all of the estimation procedures. We introduce the asymptotic theory of the procedures in Section 4.3, where we also discuss the results and their connections to prominent results in the literature.In section 4.4 we prove two given theorems in Section 4.3.

4.2. Model and Methodology

Let $X_i(s), s \in [0, T]$ be a stochastic process defined on a fixed interval $[a, b]$. As we denoted mean co-variance function of process by

$$\mu(s,t) = E\{X(s)|T > t\}, \qquad \text{and} \qquad R^t(s, s^{'}) = cov(X(s), X(s'))$$

which are assumed to exist.The model (2.1) can be rewritten as

$$\hat{X}_{ij} = X_i(S_{ij}) + \epsilon_{ij}, \qquad i = 1, ...n, j = 1, ..., m_i,$$

where the S_{ij}'s are random observations points with density function$f_S(.)$ and the ϵ_{ij} are identically distributed random errors with mean zero and finite variance σ.Assume $m_i > 2$ and let $N_i = m_i(m_i - 1)$.

Our approach is based on the local-linear-smoother ; see for example, Fan and Gijbels(1995). As we mentioned in section 3.1, let K(.) i be a symmetric kernel density function on[0,1]. The estimator for mean $\hat{\mu}(s,t)$ was obtained in Section 3.1 can been seen easily as

$$\hat{\mu}(t) = \frac{G_0 S_2 - G_1 S_1}{S_0 S_2 - S_1^2} \tag{4.1}$$

where

$$S_r = \frac{1}{n}\sum_{i=i}^{n}\frac{1}{m_i}\sum_{j=i}^{m_i} K_{h_\mu}(S_{ij} - s)\{(S_{ij} - s)/h_\mu\}^r \tag{4.2}$$

$$G_r = \frac{1}{n}\sum_{i=i}^{n}\frac{1}{m_i}\sum_{j=i}^{m_i} K_{h_\mu}(S_{ij} - s)\{(S_{ij} - s)/h_\mu\}^r \hat{X}_{ij} \tag{4.3}$$

To show estimator for co-variance function $\mathbb{R}^t(s, s^{'})$, we first estimate $C(s, s') = E\{X(s), X(s')\}$ explicitly as $\hat{C}(s, s') = \hat{a}_0$ by minimizing

$$\begin{aligned}(\hat{a}_0, \hat{a}_1, \hat{a}_2) = argmin\frac{1}{n}\sum_{i=1}^{n}\Bigg[\frac{1}{N_i}\sum_{k\neq j}\{\hat{X}_{ij}\hat{X}_{ik} - a_0 - a_1(S_{ij} - s) - a_2(S_{ik} - s)\}^2 \\ K_{h_{Gt}}(S_{ij} - s)K_{h_{Gt}}(S_{ik} - s)\Bigg]\end{aligned} \tag{4.4}$$

with $\sum_{k\neq j}$ denoting sum over all $k, j = 1, ..., m_i$ such that $k \neq j$. It follows that

$$\hat{C}(s, s') = (A_1 G_{00} - A_2 G_{10} - A_3 G_{01})B^{-1}, \tag{4.5}$$

where

$$A_1 = S_{20}S_{02} - S_{11}^2, A_2 = S_{10}S_{02} - S_{01}S_{11} \tag{4.6}$$

$$A_3 = S_{01}S_{20} - S_{10}S_{11}, B = A_1S_{00} - A_2S_{10} - A_3S_{01} \tag{4.7}$$

Also

$$Spq = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\left(\frac{S_{ij}-s}{h_R}\right)^p\left(\frac{S_{ij}-s'}{h_R}\right)^q K_{h_{Rt}}(S_{ij}-s)K_{h_{Rt}}(S_{ik}-s)$$

$$Gpq = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\hat{X}_{ij}\hat{X}_{ik}\left(\frac{S_{ij}-s}{h_R}\right)^p\left(\frac{S_{ij}-s'}{h_R}\right)^q K_{h_{Rt}}(S_{ij}-s)K_{h_{Rt}}(S_{ik}-s)$$

Then we estimate $R^t(s,s')$ by

$$\hat{R}^t(s,s') = \hat{C}(s,s') - \hat{\mu}(s,t)\hat{\mu}(s,t)$$

We concentrate the mean and co-variance estimation for dense and sparse functional.The sparse case roughly refers to the situation where each m_i is essentially bounded by some finite number M. The local-linear smoothers in these estimation procedures was studied by Yao,Muller and Wang (2005a) and Hall, Muller and Wang(2006).We attach the weigths m_i andN_i to each curve i in order to optimize the estimations.As will be seen, our approach is suitable for dense functional data and sparse data.The one benefit of our method is that we don't have to discern data type dense, sparse or mixed and decide which methodology should be used accordingly.Almost-sure(a.s) uniform rates of convergence for $\mu(s,t)$ and $R^t(s,s')$ over the entire range of s,s' will be proved. The sample size m_i for each trajectory will be completely flexible.These rates match the best known rates.

4.3. Asymptotic Theory

In asymptotic approach, we assume that m_i may depend on n as well, namely, $m_i = m_{in}$. However, for simplicity we continue to use the notation m_i Define

$$\gamma_{nk} = \left(n^{-1}\sum_{i=1}^{n} m_i^{-k}\right)^{-1} \qquad k = 1, 2, ...$$

which is the kth order harmonic mean of m_i and for any bandwidth h,

$$\delta_{n1}(h) = [\{1 + \frac{1}{h\gamma_{n1})}\}\frac{logn}{n}]^{1/2}$$

and

$$\delta_{n2}(h) = [\{1 + \frac{1}{h\gamma_{n1}} + \frac{1}{h\gamma_{n2}}\}\frac{logn}{n}]^{1/2}$$

We prove the uniform convergence rate of mean and convergence functions under the following conditions which h_μ *and* h_{R^t} are bandwidths.

(a) Let $m > 0$ and $M > 0$ be constants $m < f_S(s) < M$ for all $s \in [a, b]$ .Further, f_S is differential with derivative with a bounded derivative.

(b) The kernel function K(.) is a symmetric probability density function on $[-1, 1]$ and is of bounded variation on $[-1, 1], v_2 = \int_1^1 s^2 K(s)ds < \infty$

(c) $\mu(.)$ is twice differentiable and the second derivative is bounded on $[a, b]$

(d) All second-order partial derivatives of $R^t(s, s')$ exist and are bounded on $[a, b]^2$.

(e) $E(|\epsilon_{ij}|^{\lambda_\mu}) < \infty$ and $E(sup_{s\in[a,b]}|X(s)|^{\lambda_\mu}) < \infty$ for some $\lambda_\mu \in (2, \infty); h_\mu \to 0$ and $(h_\mu^2 + h_\mu/\gamma_{n1})^{-1}(logn/n)^{1-2/\lambda_R} \to 0 \quad as \quad n \to \infty.$

(f) $E(|\epsilon_{ij}|^{2\lambda_{R^t}}) < \infty$ and $E(sup_{s\in[a,b]}|X(s)|^{2\lambda_R}) < \infty$ for some $\lambda_R \in (2, \infty); h_R \to 0$ and $(h_R^4 + h_R^3/\gamma_{n1} + h_R^2/\gamma_{n2})^{-1}(logn/n)^{1-2/\lambda_R} \to 0 \quad as \quad n \to \infty.$

The conditions (e) and (f) hold generally, where they hold for normal process with continuous sample paths for all $\lambda > 0$. Hall, Muller and Wang(2006) adopted those conditions as well.

THEOREM 4.1. (Convergence rate of Mean Estimation)
Assume that (a), (b),(c) and (e) hold. Then

$$sup_{t\in[a,b]}|\hat{\mu}(s,t) - \mu(s,t)| = O(h_\mu^2 + \delta_{n1}(h_\mu)) \qquad a.s. \tag{4.8}$$

While $\delta_{n1}(h_\mu)$ is bound for $sup_{t\in[a,b]}|\hat{\mu}(s,t) - E(\hat{\mu}(s,t))|$,$O(h_\mu^2)$ is bound for the bias whose derivation is easy to figure out and is essentially the same as in classical non-parametric regression.The second bound of the last is derived more difficult and represents our main contribution in this result. We obtain uniform bound for $sup_{t\in[a,b]}|\hat{\mu}(s,t) - E(\hat{\mu}(s,t))|$ over $[a,b]$ and obtained a uniform bound over a finite grid on $[a,b]$ where the grid grows increasingly dense with n and then that the difference between two uniform bound is asymptotic negligible.The main difficulty of our approaches is that it is necessary to deal with curve dependence .Note that the dependence between $X(s,t)$ and $X(s',t)$ typically becomes stronger as $|s-s'|$ becomes smaller.For dense functional data, Hall,Muller and Wang(2006) and Zhang Chen(2007) address the dense functional data by setting as following ;
If $min_{1<i<n}m_i > M_n$ for some sequence M_n where $M_n^{-1} < h-\mu < (logn/n)^{1/4}$ is bounded away from 0 then

$$sup_{t\in[a,b]}|\hat{\mu}(s,t) - \mu(s,t)| = O(\{logn/n\}^{1/2})$$

where both papers take the approach of first fitting a smooth curve to $\hat{X}_{ij}(s,t), 1 < j < m_i$ for each i and then estimating $\mu(s,t)$ and$R(s,s')$ by sample mean and co-variance functions respectively of fitted curves. However, their methods have two drawbacks are:

- Differentiate of the sample curves is required. Therefore,this approach of first will not suitable for Brownian motion, which has continuous but non-differentiable sample paths.

- The sample curves that are included in the analysis need to be all densely observed; those that do not meet denseness criterion are dropped even though they may contain useful information.

Our approach does not require sample-path differentialbility and all of the data are used in the analysis

THEOREM 4.2. (Convergence rate of Co-variance Estimation)

Assume that (a), (b),(c),(e) and (f) hold. Then

$$sup_{t\in[a,b]}|\hat{R}(s,s') - R(s,s')| = O(h_\mu^2 + \delta_{n1}(h_\mu) + h_R^2 + \delta_{n2}(h_R)) \qquad a.s. \qquad (4.9)$$

- The rate in(4.9) is the classical non-parametric rate for estimating a surface(bi-variate function) which will be referred to as a two-dimensional rate. Note that $\hat{\sigma}$ has one dimensional rate in sparse setting, while both $\hat{R}(s,s')$ and $\hat{\sigma}^2$ have root-n rates in the dense setting. Most of the discussion in Sections 4.3 obviously apply here will not be repeated.

- Yao, Muller and Wang (2005a) smoothed the products of residuals instead of $X_{ij}X_{jk}$ in the local linear smoothing algorithm in (4.4) .There is some evidence that a slightly better rate can be achieved in that procedure.

4.4. Proofs

4.4.1. The Proof of mean Convergence rate of Mean Estimation

For simplicity,throughout this subsection, we abbreviate h_μ as h.Also

- let $s_1 \wedge s_2 = min(s_1, s_2)$
- $s_1 \vee s_2 = max(s_1, s_2)$
- $K_{(l)} = s^l K(s)$
- $K_{h,(l)}(v) = (1/h)K_{(l)}(v/h)$

Before starting the proof of theorem we will give two lemmas.

LEMMA 1: Assume that

$$E(sup|X(s,t)^{\lambda}) < \infty \qquad and \qquad E|\epsilon^{\lambda}| < \infty \qquad for \quad some \quad \lambda \in (2, \infty) \tag{4.10}$$

Let $\zeta_{ij} = X(S_{ij})$ or ϵ_{ij} $\quad for \quad 1 \leq i \leq n \quad 1 \leq j \leq m_i$. Let c_n any positive sequence tending to 0 and $\beta_n = c_n^2 + c_n/\gamma_{n1}$. Assume that $\beta_n^{-1}(logn/n)^{1-2/\lambda} = o(1)$.

Let

$$R_n(s_1, s_2) = \frac{1}{n}\sum_{i=1}^{n}\left\{\zeta_{ij} I(S_{ij} \in [s_1 \wedge s_2, s_1 \vee s_2]\right\} \tag{4.11}$$

$$R(s_1, s_2) = E(R_n(s_1, s_2))$$

and

$$V_n(s, c) = sup_{|u|\leq c}|R_n(s, s+u) - R(s, s+u)| \qquad c > 0$$

Then

$$sup_{s\in[a,b]} V_n(s, c_n) = O(n^{-1/2}\beta_n logn^{1/2}) \qquad a.s \tag{4.12}$$

Proof of Lemma1:

Assume ζ_{ij} is non-negative, define equally space grid $\theta = \{v_k\}$ with $v_k = a + kc_n$ for $k = 0, 1, 2, .., [(a-b)/c_n]$ and for any $s \in [a, b]$ and $|u| \leq c_n$.

Let v_k be a grid point that is with c_n of both s and $s+u$,which exists

$$|R_n(s, s+u) - R(s, s+u)| \leq |R_n(v_k, s+u) - R(v_k, s+u)| + |R_n(v_k, s) - R(v_k, s)|$$

$$|R_n(s,s+u) - R(s,s+u)| \leq 2sup_{s\in\theta(v_k)}V_n(s,c_n)$$

thus

$$sup_{s\in[a,b]}V_n(s,c_n) \leq 2sup_{s\in\theta}V_n(s,c_n) \tag{4.13}$$

From now we focus on the rigth-hand side of the inequality above . Let

$$a_n = n^{-1/2}\{\beta_n logn\}^{1/2} \qquad and \qquad Q_n = \beta_n/a_n, \tag{4.14}$$

If we define new functions $R_n^*(s_1,s_2)$, $R^*(s_1,s_2)$ *and* $V_n^*(s,c_n)$ in the same way as $R_n(s_1,s_2)$, $R(s_1,s_2)$ *and* $V_n(s,c_n)$, respectively, except with $\zeta_{ij}I(\zeta_{ij}\leq Q_n)$ replacing ζ_{ij}. Then

$$sup_{s\in\theta}V_n(s,c_n) \leq sup_{s\in\theta}V_n^*(s,c_n) + A_{n1} + A_{n2} \tag{4.15}$$

where

$$A_{n1} = sup_{s\in\theta}sup_{|u|\leq c_n}(R_n(s,s+u) - R_n^*(s,s+u))$$

$$A_{n2} = sup_{s\in\theta}sup_{|u|\leq c_n}(R(s,s+u) - R^*(s,s+u))$$

We firstly focus on A_{n1} *and* A_{n2}. when we plug the equality of a_n and Q_n into $a_n^{-1}Q_n^{1-n}$ we get

$$a_n^{-1}Q_n^{1-n} = \{\beta_n^{-1}(logn/n)^{1-2/\lambda}\}^{\lambda/2} = o(1) \tag{4.16}$$

For all s and u, by Markov's inequality,

$$\begin{aligned} a_n^{-1}(R_n(s,s+u) - R_n^*(s,s+u)) &= a_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\zeta_{ij}I(\zeta_{ij}>Q_n)\right\} \\ &\leq a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\zeta_{ij}^{\lambda}I(\zeta_{ij}>Q_n)\right\} \\ &\leq a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\zeta_{ij}^{\lambda}\right\} \end{aligned} \tag{4.17}$$

Consider the case $\zeta_{ij} = X(S_{ij})$, the other case being simpler.It follows that

$$\frac{1}{m_i}\sum_{j=1}^{m_i}\zeta_{ij}^{\lambda} \le W_i \qquad where \quad W_i = sup_{t\in[a,b]}|X_i(s,t)^{\lambda}|.$$

Thus

$$a_n^{-1}(R_n(s,s+u) - R_n^*(s,s+u)) \le a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}W_i \tag{4.18}$$

By the SLLN, $n^{-1}\sum_{i=1}^{n}W_i \to^{a.s} E(sup_{s\in[a,b]}|X(s)^{\lambda}) < \infty$. By(4.16) and (4.17) $a_n^{-1}A_{n1} \to^{a.s}$ 0. By(4.16) and (4.17) again $a_n^{-1}A_{n2} \to^{a.s} 0$.. Therefore, we have proofed that

$$\lim_{n\to\infty}(A_{n1} + A_{n2}) = o(a_n) \qquad a.s \tag{4.19}$$

To bound $V^*(s,c_n)$ for a fixed $s \in \theta$, we need to get new partition. Define $w_n = [Q_n c_n/a_n + 1]$ and $u_r = rc_n/w_n$ for $r = -w_n, -w_n+1, ..., w_n$. Note that $R_n^*(s,s+u)$ is monotone in $|u|$ since $\zeta_{ij} \ge 0$. Suppose that $0 \le u_r \le u \le u_{r+1}$ Then

$$\begin{aligned}
&R_n^*(s,s+u_r) - R^*(s,s+u_r) + R^*(s,s+u_r) + R^*(s,s+u_{r+1})\\
&\le R_n^*(s,s+u) - R^*(s,s+u)\\
&\le R_n^*(s,s+u_{r+1}) - R^*(s,s+u_{r+1}) + R^*(s,s+u_{r+1}) - R^*(s,s+u_r)
\end{aligned} \tag{4.20}$$

By defining $\xi_{nr} = R_n^*(s,s+u_r) - R^*(s,s+u_r)$

$$|R_n^*(s,s+u) - R^*(s,s+u)| \le max(\xi_{nr}, \xi_{n,r+1}) - R^*(s+u_r, s+u_{r+1})$$

The same holds if $u_r \le u_{r+1} \le 0$. Thus we get

$$V_n^*(s,c_n) \le \max_{-w_n\le r\le w_n}(\xi_{nr}) + \max_{-w_n\le r\le w_n} R^*(s+u_r, s+u_{r+1})$$

for all r,

$$R_n^*(s+u_r, s+u_{r+1}) \leq Q_n P(s+u_r \leq S \leq s+u_{r+1}) \leq M_S Q_n (u_{r+1}-u_r) \leq M_S a_n$$

since $f(s) < M_S < \infty$ and $u_{r+1} - u_r = c_n/w_n$

Therefore for any B.

$$P(V^*(s,c_n) \geq Ba_n) \leq P(max\xi_{nr} \geq (B-M_S)a_n) \tag{4.21}$$

Now let $\mathbb{Z}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} \zeta_{ij} I(\zeta_{ij} \leq Q_n) I(S_{ij} \in (s, s+u_r))$ so that $\xi_{nr} = |\frac{1}{n}\sum_{i=1}^{n}(Z_i - E(Z_i))|$.We have $|Z_i - E(Z_i)| \leq Q_n$ and

$$\sum_{i=1}^{n} var(Z_i) \leq \sum_{i=1}^{n} E(Z_i^2) \leq M \sum_{i=1}^{n}(c_n^2 + c_n/m_i) \leq Mn\beta_n$$

for some finite M. Bernstein's inequality is that ,

Let $X_1, ..., X_n$ be independent zero-mean random variables. Suppose that $|X_i| \leq M$ almost surely, for all i. Then, for all positive t,

$$P(\sum X_i > t) \leq exp\Big(- \frac{1/2t^2}{\sum X_i^2 + 1/3Mt}\Big)$$

if we replace t by $(B - M_S)a_n$ and $\sum X_i$ by ξ_{nr} since $\sum_{i=1}^{n} EX_i^2 = Var(\xi_{nr}) = \frac{var(Z_i)}{n^2}$ we get

$$\begin{aligned} P(\xi_{nr} \geq (B-M_S)a_n) &\leq exp\Big\{ - \frac{(B-M_S)^2 a_n^2 n^2}{2\sum_{i=1}^{n} var(Z_i) + (2/3)(B-M_S)Q_n n a_n}\Big\} \\ &\leq exp\Big\{ - \frac{(B-M_S)^2 a_n^2 n^2}{2Mn\beta_n + (2/3)(B-M_S)n\beta_n}\Big\} \\ &\leq n^{-B^*} \end{aligned} \tag{4.22}$$

where $B^* = \frac{(B-M_T)^2}{2M+(2/3)(B-M_T)}$ By (4.21) and (4.22) and Boole's inequality

$$P(supV_n^*(s,c_n) \geq Ba_n) \leq ([\frac{b-a}{c_n}]+1)(2[\frac{Q_n c_n}{a_n}+1]+1)n^{-B^*} \leq C\frac{Q_n}{a_n}n^{-B^*}$$

for some finite C. Consider $\frac{Q_n}{a_n} = \frac{\beta_n}{a_n^2} = \frac{n}{logn}$ so $P(V_n^*(s,c_n) \leq Ba_n)$ is sum able in n if we select B large enough such that $B^* > 2$. By the Borel-Cantelli Lemma

$$supV_n^*(s,c_n) = O(a_n) \qquad a.s \tag{4.23}$$

Thus we get(4.12) by considering the expressions in(4.13),(4.15),(4.19) and (4.23).

LEMMA 2:

Let ζ_{ij} be as in lemma 1 and assume that the conditions of lemma1 holds.Let $h = h_n$ be a bandwith and let $\beta_n = h^2 + h/\gamma_{n1}$. Assume that $h \to 0$ and $\beta_n^{-1}(logn/n)^{1-2/\lambda} = o(1)$. For any nonnegative integer p, let

$$D_{p,n}(s) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m_i}\sum_{j=1}^{m_i} K_{h,(p)}(S_{ij}-s)\zeta_{ij}\right]$$

then we have

$$\sup_{s\in[a,b]} \sqrt{nh^2/(\beta_n logn)}|D_{p,n}(s) - ED_{p,n}(s)| = O(1) \qquad a.s \tag{4.24}$$

Proof Of Lemma 2: Since both K and t^p are bounded variations.Thus we can write $K_{(p)} = K_{(p,1)} - K_{(p,2)}$ where $K_{(p,1)}$ *and* $K_{(p,2)}$ are both increasing functions; without loss of generality, assume that $K_{(p,1)}(-1) = K_{(p,2)}(-1) = 0$. Below, we apply Lemma1 by letting $c_n = 2h$. It is clear that the assumptions of Lemma 1 hold here.Write.

$$\begin{aligned} D_{p,n}(s) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m_i}\sum_{j=1}^{m_i} K_{h,(p)}(S_{ij}-s)\zeta_{ij}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m_i}\sum_{j=1}^{m_i} \zeta_{ij} I(-h \le S_{ij}-s \le h)\int_{-h}^{S_{ij}-s} dK_{h,(p)}(v)\right] \\ &= \int_{-h}^{h}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\zeta_{ij} I(v \le S_{ij}-s \le h\right\} dK_{h,(p)}(v) \\ &= \int_{-h}^{h} R_n(s+v, s+h)\, dK_{h,(p)}(v) \end{aligned} \tag{4.25}$$

where R_n is as defined in (4.11).Therefore , we have

$$\begin{aligned} &\sup_{t\in[a,b]} |D_{p,n}(s) - E(D_{p,n}(s))| \\ &= \sup_{s\in[a,b]} \left|\int_{-h}^{h} R_n(s+v,s+h)dK_{h,(p)}(v) - E\int_{-h}^{h} R_n(s+v,s+h)dK_{h,(p)}(v)\right| \\ &= \sup_{s\in[a,b]} \left|\int_{-h}^{h} (R_n(s+v,s+h) - ER_n(s+v,s+h))dK_{h,(p)}(v)\right| \\ &= \{k_{(p,1)}(1) + K_{(p,2)}(1)\}h^{-1} \sup_{s\in[a,b]} V_n(s,2h) \\ &< \{k_{(p,1)}(1) + K_{(p,2)}(1)\}h^{-1} O\Big((n^{-1/2}(\beta_n logn)^{1/2}\Big) \end{aligned} \tag{4.26}$$

Thus we get

$$\sup_{t\in[a,b]} |D_{p,n}(s) - E(D_{p,n}(s))| = O(\delta_{n1}(h))$$

or

$$\sup_{t\in[a,b]} \sqrt{nh^2/(\beta_n logn)}\,|D_{p,n}(s) - E(D_{p,n}(s))| = O(1)$$

where $\delta_{n1} = \left[(1+(h\gamma_{n1})^{-1})logn/n\right]^{1/2}$ and $\sqrt{nh^2/(\beta_n logn)} = \delta_{n1}(h)^{-1}$.

Proof of Theorem 4.1 From the equality in (4.2) and in (4.3), we have

$$S_r = \frac{1}{n}\sum_{i=i}^{n}\frac{1}{m_i}\sum_{j=i}^{m_i} K_{h_\mu}(S_{ij}-s)\{(S_{ij}-s)/h_\mu\}^r \tag{4.27}$$

$$G_r = \frac{1}{n}\sum_{i=i}^{n}\frac{1}{m_i}\sum_{j=i}^{m_i} K_{h_\mu}(S_{ij}-s)\{(S_{ij}-s)/h_\mu\}^r \hat{X}_{ij}$$

and

$$\hat{\mu}(t) = \frac{G_0S_2 - G_1S_1}{S_0S_2 - S_1^2}$$

Thus, if we define a new function

$$G_r^* = G_r - \mu(s,t)S_r - h\mu^{(1)}(s,t)S_{r+1}$$

By straightforward calculations, we have

$$\hat{\mu}(t) - \mu(s,t) = \frac{G_0^*S_2 - G_1^*S_1}{S_0S_2 - S_1^2}$$

where S_0, S_1, S_3 are defined in the equation (4.27).

$$\begin{aligned} G_r^* &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i} K_n(S_{ij}-s)\{(S_{ij}-s)/h\}^r\{X_{ij}(s,t) - \mu(s,t) - \mu^{(1)}(s,t)(S_{ij}-s)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i} K_n(S_{ij}-s)\{(S_{ij}-s)/h\}^r\{\epsilon_{ij} + \mu(S_{ij}) - \mu(s,t) - \mu^{(1)}(s,t)(S_{ij}-s)\} \end{aligned} \tag{4.28}$$

By Taylor's expansion and Lemma2, uniformly in it.

$$G_r^* = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i} K_n(S_{ij}-s)\{(S_{ij}-s)/h\}^r\epsilon_{ij} + O(h^2) \tag{4.29}$$

and it follows from Lemma 2 that

$$G_i^* = O(h^2 + \delta_{n1}(h)) \qquad a.s \tag{4.30}$$

where $\delta_{n1} = \left[(1 + (h\gamma_{n1})^{-1})logn/n\right]^{1/2}$ Now , at any interior point $s \in [a+h, b-h]$ since f has a bounded derivative

$$E(S_0) = \int_{-1}^{1} K(v)f(s+hv)dv = \int_{-1}^{1} K(v)[f(s) + f'(s)h]dv = f(s) + O(h)$$

$$E(S_1) = \int_{-1}^{1} K(v)Vf(s+hv)dv = \int_{-1}^{1} K(v)v[f(s) + f'(s)h]dv = O(h)$$

Similarly,We can get

$$E(S_2) = f(s)v_2 + O(h)$$

where $v_2 = \int v^2 K(v)dv$

By Lemma2 uniformly for $s \in [a+h, b-h]$ direclty we can get uniformly converges rate for S_0, S_1, S_2 as

$$S_0 = f(s) + O(h + \delta_{n1}(h)$$

$$S_1 = O(h + \delta_{n1}(h)$$

and

$$S_2 = f(s)v_2 + O(h + \delta_{n1}(h))$$

Thus, we get

$$\sup_{s \in [a,b]} |\hat{\mu}(s) - \mu(s)| = O(h_\mu^2 + \delta_{n1}(h_\mu))$$

4.4.2. The Proof of Theorem 4.2(Convergence Rate for Co-variance Functions

Before starting the proof we give two lemma;Lemma4.3 and Lemma 4.4.

LEMMA 3

Assume that $E(sup|X(s,t)|^{2\lambda}) < \infty$ and $E|\epsilon|^{2\lambda} < \infty$ for some $\lambda \in (2, \infty)$.

Let Z_{ijk} *be* $X(\hat{S}_{ij})\hat{X}(S_{ik})$. Let $c_n \to 0$ on a $c_n > 0$ and $\beta_n = (c_n^4 + c_n^3/\gamma_{n1} + c_n^2/\gamma_{n2}) = O(1)$.Let $N_i = m_i(m_i - 1)$ Define a new function as

$$R_n(s_1, s_1', s_2, s_2') = \frac{1}{n}\sum_{i=1}^{n}\left\{Z_{ij}I(S_{ij} \in [s_1 \wedge s_2, s_1 \vee s_2], S_{ik} \in [s_1' \wedge s_2', s_1' \vee s_2'])\right\} \quad (4.31)$$

$$R(s_1, s_1', s_2, s_2') = E\{R_n(s_1, s_1', s_2, s_2')\}$$

and

$$V_n(s, s', \delta) = \sup_{|u_1|,|u_2|<\delta} |R_n(s, s', s + u_1, s' + u_2) - R(s, s', s + u_1, s' + u_2)$$

Then

$$\sup_{s,s' \in [a,b]} V_n(s, s', c_n) = O(n^{-1/2}(\beta_n logn)^{1/2})| \qquad a.s$$

Proof of Lemma 3

Let $a_n = n^{-1/2}(\beta_n logn)^{1/2}$ and $Q_n = \beta_n/a_n$.Let P be a two-dimension grid on $[a, b]^2$ with mesh c_n that is $P = \{(v_{k_1}, v_{k_2})\}$ where v_k is defined as in the proof of Lemma4.1

$$\sup_{(s,s') \in [a,b]} V(s, s', c_n) \leq 4 \sup_{(s,s') \in P} (s, s', c_n)$$

Define new functions $R_n^*(s_1, s_1', s_2, s_2')$, $R^*(s_1, s_1', s_2, s_2')$ *and* $V_n^*(s, s', c_n)$ in the same way as $R_n(s_1, s_1', s_2, s_2')$ $R(s_1, s_1', s_2, s_2')$ *and* $V_n(s, s', c_n)$, respectively, except with

$Z_{ijk}I(Z_{ijk} \leq Q_n)$ replacing Z_{ijk}. Then

$$\sup_{(s,s')\in P} V_n(s,s',c_n) \leq \sup_{(s,s')\in P} V_n^*(s,c_n) + A_{n1} + A_{n2} \tag{4.32}$$

where

$$A_{n1} = \sup_{(s,s')\in P} \sup_{|u_1|||u_2|\leq c_n} (R_n(s,s',s+u_1,s'+u_2) - R_n^*(s,s',s+u_1,s'+u_2)$$

$$A_{n2} = \sup_{(s,s')\in P} \sup_{|u_1|||u_2|\leq c_n} (R(s,s',s+u_1,s'+u_2) - R^*(s,s',s+u_1,s'+u_2)$$

We firstly focus on A_{n1} *and* A_{n2}. when we plug the equality of a_n and Q_n into $a_n^{-1}Q_n^{1-n}$ we get

$$a_n^{-1}Q_n^{1-n} = \{\beta_n^{-1}(logn/n)^{1-2/\lambda}\}^{\lambda/2} = o(1) \tag{4.33}$$

For all s, s' and u, by Markov's inequality,

$$\begin{aligned} a_n^{-1}(R_n(s,s',s+u_1,s'+u_2 - R_n^*(s,s',s+u_1,s'+u_2) \\ = a_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ijk}I(Z_{ijk} > Q_n)\right\} \\ \leq a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ijk}^{\lambda}I(Z_{ijk} > Q_n)\right\} \\ \leq a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ijk}^{\lambda}\right\} \end{aligned} \tag{4.34}$$

Consider the case $Z_{ijk} = X(S_{ij})X(S_{ik})$, the other case being simpler.It follows that

$$\frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ijk}^{\lambda} \leq W_i \qquad where \quad W_i = sup_{(s,s')\in[a,b]}|(X_i(s)X_i(s'))^{\lambda}|.$$

Thus

$$a_n^{-1}(R_n(s,s',s+u_1,s'+u_2 - R_n^*(s,s',s+u_1,s'+u_2) \leq a_n^{-1}Q_n^{1-\lambda}\frac{1}{n}\sum_{i=1}^{n} W_i \tag{4.35}$$

By the SLNN, $n^{-1}\sum_{i=1}^{n} W_i \to^{a.s} E(sup_{s\in[a,b]}|X(s)^{\lambda}) < \infty$. By(4.33) and (4.35) $a_n^{-1}A_{n1} \to 0 \quad a.s.$ By(4.33) and (4.35) again $a_n^{-1}A_{n2} \to 0 \quad a.s.$ Therefore, we have proofed that

$$\lim_{n\to\infty}(A_{n1}+A_{n2}) = o(a_n) \qquad a.s \tag{4.36}$$

To bound $V^*(s,s',c_n)$ for a fixed $(s,s') \in P$, we need to get new partition. Define $w_n = [Q_n c_n/a_n + 1]$ and $u_r = rc_n/w_n$ for $r = -w_n, -w_n+1, ..., w_n$. Note that $R_n^*(s,s',s+u_1,s'+u_2)$ is monotone in $|u_1||u_2|$ since $Z_{ijk} \geq 0$. Suppose that $0 \leq u_{r_1} \leq u_1 \leq u_{r_1+1}$ and $0 \leq u_{r_2} \leq u_2 \leq u_{r_2+1}$ Then

$$\begin{aligned}
&R_n^*(s,s',s+u_{r_1},s'+u_{r_2}) - R^*(s,s',s+u_{r_1},s'+u_{r_2})\\
&+R^*(s,s',s+u_{r_1+1},s'+u_{r_2+1}) - R^*(s,s',s+u_{r_1},s'+u_{r_2})\\
&\leq R_n^*(s,s',s+u_1,s'+u_2) - R^*(s,s',s+u_1,s'+u_2)\\
&\leq R_n^*(s,s',s+u_{r_1+1},s'+u_{r_2+1} - R^*(s,s',s+u_{r_1+1},s'+u_{r_2+1})\\
&+R^*(s,s',s+u_{r_1+1},s'+u_{r_2+1}) - R^*(s,s',s+u_{r_1},s'+u_{r_2})
\end{aligned} \tag{4.37}$$

By defining

$$\xi_{n,r_1,r_2} = |R_n^*(s,s',s+u_{r_1},s'+u_{r_2}) - R^*(s,s',s+u_{r_1},s'+u_{r_2})|$$

$$\begin{aligned}
|R_n^*(s,s',s+u_1,s'+u_2) - R^*(s,s',s+u_1,s'+u_2)| &\leq max(\xi_{n,r_1,r_2},\xi_{n,r_1+1,r_2+1})\\
&+R_n^*(s+u_{r_1},s'+u_{r_2},s+u_{r_1+1},s'+u_{r_2+1})
\end{aligned}$$

The same holds if $u_{r_1} \le u_1 \le u_{r_1+1} \le 0$ and $u_{r_2} \le u_2 \le u_{r_2+1} \le 0$. Thus we get

$$V_n^*(s, s', c_n) \le \max_{-w_n \le r_1, r_2 \le w_n} (\xi_{n,r_1,r_2}) + \max_{-w_n \le r_1, r_2 \le w_n} R_n^*(s + u_{r_1}, s' + u_{r_2}, s + u_{r_1+1}, s' + u_{r_2+1})$$

For all r_1, r_2,

$$R_n^*(s+u_{r_1}, s'+u_{r_2}, s+u_{r_1+1}, s'+u_{r_2+1}) \le Q_n P\bigg(s+u_{r_1} \le S \le s+u_{r_1+1} ands'+u_{r_2} \le S \le s'+u_{r_2+1}\bigg)$$

$$\le M_S Q_n (u_{r_1+1} - u_{r_1})(u_{r_2+1} - u_{r_2}) \le M_S a_n$$

Therefore for any B.

$$P(V^*(s, s', c_n) \ge B a_n) \le P(\max_{-w_n \le r_1, r_2 \le w_n} \xi_{n,r_1,r_2} \ge (B - M_T) a_n) \tag{4.38}$$

Now let $\mathbb{Z}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbb{Z}_{ijk} I(Z_{ijk} \le Q_n) I(S_{ij} \in (s, s + u_{r_1}), S_{ik} \in (s', s' + u_{r_2}))$ so that $\xi_{n,r_1,r_2} = |\frac{1}{n} \sum_{i=1}^{n} (Z_i - E(Z_i))|$.We have $|Z_i - E(Z_i)| \le Q_n$ and

$$\sum_{i=1}^{n} var(Z_i) \le \sum_{i=1}^{n} E(Z_i^2) \le M \sum_{i=1}^{n} (c_n^2 + c_n/N_i) \le M n \beta_n$$

For some finite M, using Bernstein's inequality, We get;

$$\begin{aligned} P(\xi_{n,r_1,r_2} \ge (B - M_S) a_n) &\le exp\bigg\{ - \frac{(B - M_S)^2 a_n^2 n^2}{2 \sum_{i=1}^{n} var(Z_i) + (2/3)(B - M_S) Q_n n a_n} \bigg\} \\ &\le exp\bigg\{ - \frac{(B - M_S)^2 a_n^2 n^2}{2 M n \beta_n + (2/3)(B - M_S) n \beta_n} \bigg\} \\ &\le n^{-B^*} \end{aligned} \tag{4.39}$$

By (4.39) and Boole's inequality

$$P(supV_n^*(s,s',c_n) \geq Ba_n) \leq ([\frac{b-a}{c_n}]+1)(2[\frac{Q_nc_n}{a_n}n^{-B^*}+1])n^{-B^*} \leq C\frac{Q_n}{a_n}n^{-B^*}$$

for some finite C. Consider $\frac{Q_n}{a_n} = \frac{\beta_n}{a_n^2} = \frac{n}{logn}$so $P(V_n^*(s,s',c_n) \leq Ba_n)$ is sum able in n if we select B large enough such that $B^* > 2$. By the Borel-Cantelli Lemma

$$\sup_{(s,s')\in P} V_n^*(s,s',c_n) = O(a_n) \qquad a.s \tag{4.40}$$

Thus we obtain the result of Lemma3 by considering the expressions in(4.32),(4.36)and (4.40).

LEMMA 4:

Assume that $E(sup|X(s,t)|^{2\lambda}) < \infty$ and $E|\epsilon|^{2\lambda} < \infty$ for some $\lambda \in (2,\infty)$.

Let Z_{ijk} *be* $X(S_{ij})X(S_{ik})$, $X(S_{ij})\epsilon_{ik}$, *or* $\epsilon_{ik}\epsilon_{ik}$. Let $c_n \to 0$ on a $c_n > 0$ and $\beta_n = (c_n^4 + c_n^3/\gamma_{n1} + c_n^2/\gamma_{n2}) = O(1)$. Let $N_i = m_i(m_i-1)$. For any $p,q > 0$ let

$$D_{p,q}(s,s') = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{j\neq k} Z_{ijk}K_{h(p)}(S_{ij}-s)K_{h(q)}(S_{ik}-s')\right]$$

Then

$$\sup_{(s,s')\in[a,b]} \sqrt{nh^4/(\beta_n logn)}|D_{p,q,n}(s,s') - E(D_{p,q,n}(s,s'))| = O(1)$$

Proof of Lemma 4.4:

$$D_{p,q}(s,s') = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{j\neq k} Z_{ijk}I(S_{ij} \leq s+h)I(S_{ik} \leq s'+h)K_{h(p)}(S_{ij}-s)K_{h(q)}(S_{ik}-s')\right]$$

$$= \int\int_{(u,v)\in[-h,h]^2} \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{j\neq k} Z_{ijk} I(S_{ij}\in[s+u,s+h] I(S_{ik}\in[s'+v,s'+h])\right.$$

$$\left. K_{h(p)}(S_{ij}-s)K_{h(q)}(S_{ik}-s')\right] dK_{h,(p)}(u)dK_{h,(q)}(v)$$

$$= \int\int_{(u,v)\in[-h,h]^2} \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{j\neq k} R_n(s+u,s'+v,s+h,s'+h)\right.$$

$$\left. K_{h(p)}(S_{ij}-s)K_{h(q)}(S_{ik}-s')\right] dK_{h,(p)}(u)dK_{h,(q)}(v)$$

where R_n is as(4.31).Now

$$\sup_{(s,s')\in[a,b]} |D_{p,q,n}(s,s') - E(D_{p,q,n}(s,s'))|$$

$$\leq \sup_{(s,s')\in[a,b]} V_n(s,s',2h)\int\int_{(u,v)\in[-h,h]^2} |K_{h(p)}(u)||K_{h(q)}(v)|$$

$$= O((\beta_n logn/nh^4)^{1/2}) \qquad a.s$$

Proof of Theorem 2:

Define a new function

$G^*_{pq} = G_{pq} - C(s,s')S_{pq} - h_G C^{(1,0)}(s,t)S_{p+1,q} - h_R C^{(0,1)}(s,t)S_{p,q+1}$ Considering the variables

$$G_{20} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\hat{X}_{ij}\hat{X}_{ik}\left(\frac{S_{ij}-s}{h_R}\right)^2 K_{h_{R^t}}(S_{ij}-s)K_{h_{R^t}}(S_{ik}-s')$$

$$G_{02} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\hat{X}_{ij}\hat{X}_{ik}\left(\frac{S_{ij}-s'}{h_R}\right)^2 K_{h_{R^t}}(S_{ij}-s)K_{h_{R^t}}(S_{ik}-s')$$

$$G_{11} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{k\neq j}\hat{X}_{ij}\hat{X}_{ik}\left(\frac{S_{ij}-s}{h_R}\right)\left(\frac{S_{ij}-s'}{h_R}\right) K_{h_{R^t}}(S_{ij}-s)K_{h_{R^t}}(S_{ik}-s')$$

Therefore

$$(\hat{C}-C)(s,s') = (A_1 G^*_{00} - A_2 G^*_{10} - A_3 G^*_{01})B^{-1} \qquad (4.41)$$

By standard calculation, we have the following notes uniformly on $[a+h_R, b-h_R]^2$;

$$E(S_{00}) = \int_{-1}^{1}\int_{-1}^{1} K_{h_R}(S_{ij}-s)K_{h_R}(S_{ik}-s')f(s)f(s')dsds'$$

$$= \int_{-1}^{1}\int_{-1}^{1} K_{h_R}(u)K_{h_R}(v)f(uh+s)f(vh+s')dsds'$$

$$= f(s)f(s') + O(h_R)$$

$$E(S_{01}) = \int_{-1}^{1}\int_{-1}^{1} K_{h_R}(S_{ij}-s)K_{h_R}(S_{ik}-s')\left(\frac{S_{ij}-s}{h_R}\right)f(s)f(s')dsds'$$

$$= \int_{-1}^{1}\int_{-1}^{1} K_{h_R}(u)K_{h_R}(v)uf(uh+s)f(vh+s')dsds'$$

$$= O(h_R)$$

Similarly, $E(S_{10}) = O(h_R)$, $E(S_{11}) = O(h_R)$

$E(S_{02}) = f(s)f(s')v_2 + O(h_R)$ and $E(S_{20}) = f(s)f(s')v_2 + O(h_R)$

By Lemma 4 we get;

$$S_{10} = O(h_R + \delta_{n2}(h_R))$$

$$S_{11} = O(h_R + \delta_{n2}(h_R))$$

$$S_{00} = f(s)f(s') + O(h_R + \delta_{n2}(h_R)))$$

$$S_{20} = f(s)f(s')v_2 + O(h_R + \delta_{n2}(h_R)))$$

$$S_{02} = f(s)f(s')v_2 + O(h_R + \delta_{n2}(h_R)))$$

$$B = f^3(s)f^3(s')v_2^2 + O(h_R + \delta_{n2}(h_R)))$$

When we plug these values into A_1, A_2, A_3 and B we can reach almost sure uniform rates:

$$\begin{aligned} A_1 &= f^2(s)f^2(s')v_2^2 + O(h_R + \delta_{n2}(h_R))) \\ A_2 &= O(h_R + \delta_{n2}(h_R))) \\ A_1 &= O(h_R + \delta_{n2}(h_R))) \\ B &= f^3(s)f^3(s')v_2^2 + O(h_R + \delta_{n2}(h_R))) \end{aligned} \tag{4.42}$$

To analyze the behaviour of the components of (4.41) it suffices now to analyze G^*_{pq}. Write

$$G^*_{00} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N-i}\sum_{k\neq j}\{X_{ij}X_{ik} - C(s,s') - C^{(1,0)}(s,s')(S_{ij}-s) - C^{(0,1)}(s,s')(S_{ij}-s'\}) \right.$$

$$\left. K_{h_R}(S_{ij}-s)K_{h_R}(S_{ij}-s')\right]$$

Let $\xi_{ijk} = X_{ij}X_{ik} - C(S_{ij}, S_{ik})$, By Taylor's expansions

$$X_{ij}X_{ik} - C(s,s') - C^{(1,0)}(s,s')(S_{ij}-s) - C^{(0,1)}(s,s')(S_{ij}-s')$$

$$= X_{ij}X_{ik} - C(s,s') - C(S_{ij},S_{ik}) + C(S_{ij},S_{ik}) - C^{(1,0)}(s,s')(S_{ij}-s) - C^{(0,1)}(s,s')(S_{ij}-s'$$

$$= \xi^*_{ijk} + O(h_R^2) \qquad a.s$$

Thus we get

$$G^*_{00} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N-i}\sum_{k\neq j}\xi_{ijk})K_{h_R}(S_{ij}-s)K_{h_R}(S_{ij}-s')\right] + O(h_R^2)$$

Applying Lemma4 we obtain uniformly rate in s, s'

$$G^*_{00} = +O(\delta_{n2}(h_R) + h_R^2) \qquad a.s$$

By(4.42)

$$A_1B^{-1} = [f(s)f(s')]^{-1} + O(h_R + \delta_{n2}(h_R))) \qquad a.s$$

We reach uniform rate of

$$G^*_{00}A_1B^{-1} = O(h_R + \delta_{n2}(h_R))) \qquad a.s$$

Similarly we can reach the uniform rate of $G^*_0 A_2 B^{-1} = O(h_R + \delta_{n2}(h_R)))$, $G^*_{01} A_3 B^{-1} = O(h_R + \delta_{n2}(h_R)))$. Therefore, we have obtained the uniform rate claimed in theorem 4.2 for $(s, s') \in [a + h_R, b - h_R]$.

5. Performance of Our Model

5.1. Simulation Study

The non-parametric degradation modeling framework introduce in this paper applies to both complete as well incomplete degradation signal.To ensures accurate estimation of the mean function and the co-variance surface,it is significant to have a sampling plan. Yao,Muller and Wang(2005) provide theoretical results on the estimation of the co-variance surface using FPCA under large n but small m_i for i=1,...,n. and Zhou, Serban and Gebraeel(2011) introduce a new sampling scheme for sparse and fragment data by considering uniform sampling and nonuniform sampling methods.For censoring data we assume that the sample size of sample data is enough large size to reach dense time-line. For fitted model(3.10), the one-leave out prediction for the ith subject is

$$\hat{e}_{\hat{X}}^{-i}(t) = \frac{\int_t^{\infty}(\hat{\bar{F}}_{\hat{X}(t)}^{-i}(u))du}{\hat{\bar{F}}_{\hat{X}(t)}^{-i}(t)} \tag{5.1}$$

Where is $\hat{\bar{F}}_{\hat{X}(t)}^{-i}(y) = 1 - \hat{P}(\hat{T} - t \leq y|\hat{r}^{-i}(t))$. The estimation of $\hat{r}^{-i}(t)$ is obtained by coefficientsξ_{ijt} for eigenfunctions $\rho_j^{-i}(t), j = 1, .., M$ which is estimated eigenfunctions after removing the ith subject's trajectory.The one-leave-out predictions lead to the rood squared prediction error at t,

$$RSPE(t) = \{\frac{1}{N_t}\sum_{i \in R(t)}(\hat{e}^{-i}(t) - (\hat{T}_i - t))^2\}^{\frac{1}{2}} \tag{5.2}$$

Where N_t is the number of subjects in the risk set R(t). Root squared prediction error functions for various completed predictors is displayed below for 3 different data types ; complete, sparse and fragmented data.All of them are special cases of the general model (2.1).More specifically:

- For complete data, we choose $\hat{X}_i(s,t) = X_i(s,t) + \sigma\epsilon(s,t)$ such that $\sigma = 2, \epsilon(s,t) \sim N(0,1)$ and $X_1(s,t) = exp(s^2), X_2(s,t) = 3s^2, X_3(s,t) = 2s^3, X_4(s,t) = 1/2exp(s)$, $X_5(s,t) = exp(2s), X_6(s,t) = exp(s), X_7(s,t) = 2s^2, X_8(s,t) = 2T^3, X_9(s,t) = t^3, X_{10}(s,t) = 3s^3$ we run simulation 4 times for each trajectory on $[0,2]$ which is made space grid $c_0, ..., c_{100}$ with $c_0 = 0, and c_{100} = 2$ .Then we obtain 40 different co-variate trajectories, which estimate mean function $\mu(s,t)$ as we introduce in (3,3).

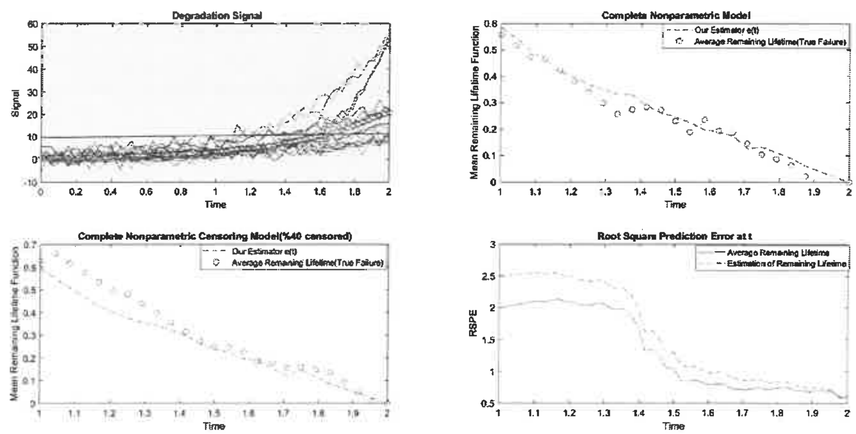

Figure 3: Complete Degradation Signal: The first graph shows the growth of degradation signal. Top right plot compares the averages of remaining lifetime for true failure time and the estimator in (3.10). In left bottom plot compares this mean remaining lifetime for censoring version.The last graph shows average root squared prediction errors(RSPE)(4.2) for censoring and non-censoring completed data

- For sparse data, we generate the observations from complete signal by dividing the interval $[0,2]$ to 8 part randomly and we choose just 4 parts for each trajectory. The stopping time for each signal is genarated from Uniform distribution [Uniform(0.8,2)]

Across all the types of data, the failure threshold is set to D=10 for uncensored data and is set to D=7 for censored data.We use local linear weighted least squares for smoothing the mean function and the variance-co-variance surface as describe above(3.3)

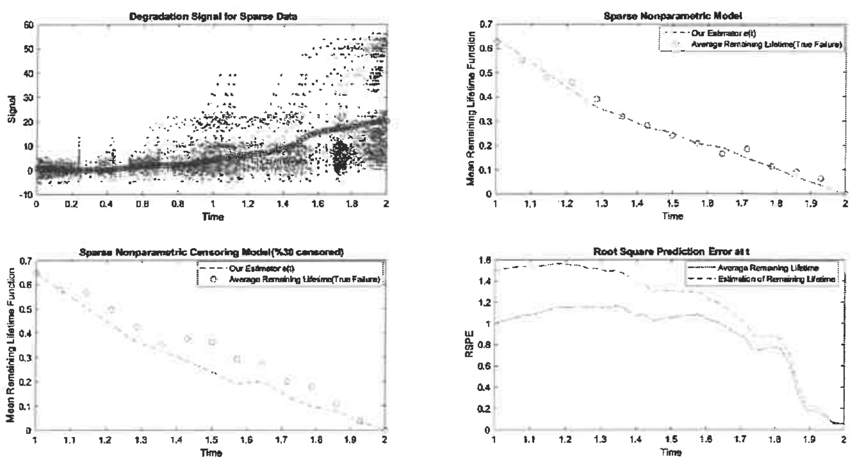

Figure 4: Sparse Degradation Signal: The first graph shows the growth of degradation signal. Top right plot compares the averages of remaining lifetime for true failure time and the estimator in (3.10). In left bottom plot compares this mean remaining lifetime for censoring version.The last graph shows average root squared prediction errors(RSPE)(4.2) for censoring and non-censoring sparse data

- For fragmented data, we generate the observations from complete signal by dividing the interval $[0, 2]$ to 8 part randomly and we choose just 2 parts for each trajectory. The stopping time for each signal is genarated from Uniform distribution [Uniform(0.7,2)]

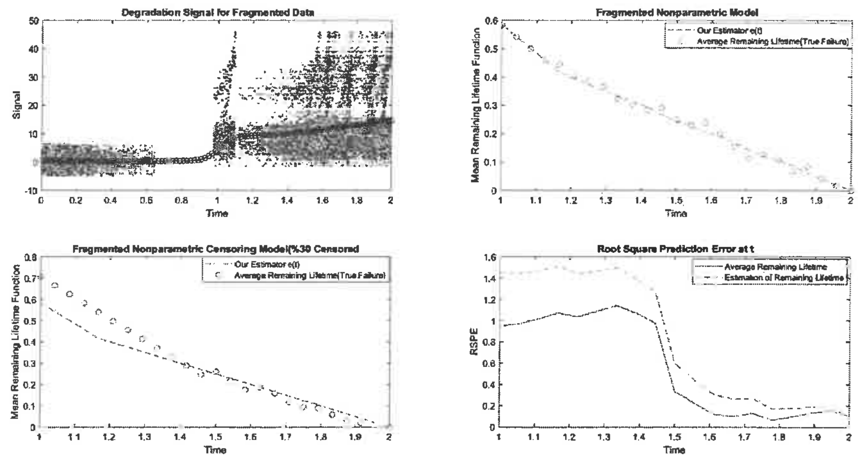

Figure 5: Fragmented Degradation Signal: The first graph shows the growth of degradation signal. Top right plot compares the averages of remaining lifetime for true failure time and the estimator in (3.10). In left bottom plot compares this mean remaining lifetime for censoring version.The last graph shows average root squared prediction errors(RSPE)(4.2) for censoring and non-censoring fregmented data

and(3.5). The bandwidth for smoothing the co-variance function from one-curve-leave-out cross validation was h=1.3.The bandwidths for smoothing the mean function were visually chosen as h=.6. The number of significantly nonzero eigenvalues was chosen as K=2. by minimizing the modified Akaike Criterion defined by Yao, Muller, Wang[18].Once the evolution of mean and eigenfunctions has been determined, estimated functional principal component scores $\hat{\xi}_{ijt}$ for each trajectory $0 \leq t \leq T$ are obtained via (3.6).These serve as predictors in various regression models that can be consider for predicting remaining lifetime.

5.2. Result and Analysis

In Figure 3, we present the new estimator of mean residual distribution and the mean remaining lifetime of the actual co-variate trajectories for complete non-censoring and censoring case.Also, we compare the root squared prediction errors (RSPE) for complete non-censoring and censoring data using non-parametric model. The first observation is that there is a small difference in the predictor error between censored non-parametric model and uncensored non-parametric model. Difference is larger for high degradation percentiles. This framework is consistent the result of Zhou, Serban, and Gebraeel[13] even if the data they used wasn't contaminated by right censoring data and they assumed the random scores as normal distribution in parametric case. As expected, the root Squared Prediction errors (RSPE) for Censoring complete data is higher than the root squared prediction errors (RSPE) for non-censoring complete data when the prediction include many failure time since each calculation of used numerical integral and estimator includes errors.The second observation is that the nonuniform or uniform sampling methods can be applicable to our proposed method. Overall, when we look at the comparison of the our estimator and average remaining lifetime of true trajectories, and when we look at the root squared prediction errors for complete censored/uncensored data, the performance of our method is good as much as the result of Zhou,Serban and Gebraeel(2011). In Figure 6 the evolution of mean degradation signal function for complete data and of the first two eigenfunction are displayed (for t=1.2, 1.5 and 2 respectively). These components describe the time-evolution of degradation signal trajectories.We find that mean and first eigenfunction quite smooth but second eigenfunction fluctuates more than first eigenfunction.While increased the time t,the smoothness of the second eigenfunction increases as same as the first eigenfunction does .

In Figure 4, we present the observations, mean remaining lifetime function and root squared prediction for sparse data using non-uniform sampling method for each independent trajectory. The mean residual lifetime function of uncensored sparse data approach the true value of average remaining lifetime better than for censoring data as expected since censoring data includes missing some failure time. Therefore, the right graph at bottom in Figure 4 shows RSPE for censored sparse data is higher than the RSPE of uncensored sparse data.

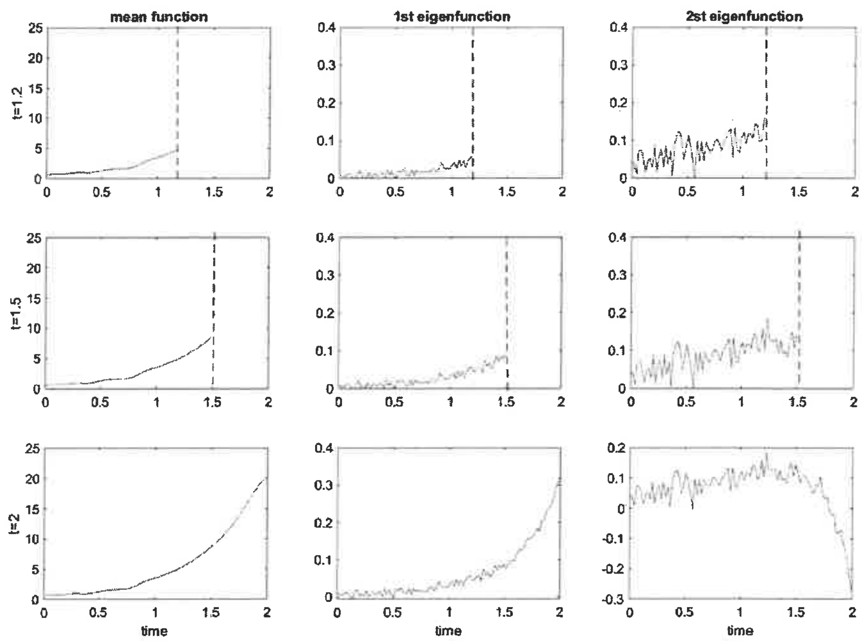

Figure 6: Evolution of mean functions(left column), and of first(middle column) and second(right column) eigenfunctions for current times t=1.2(first row),t=1.5(second row),t=2(third row) for degradation signal data

The observations , mean residual function and root squared prediction errors are dis-

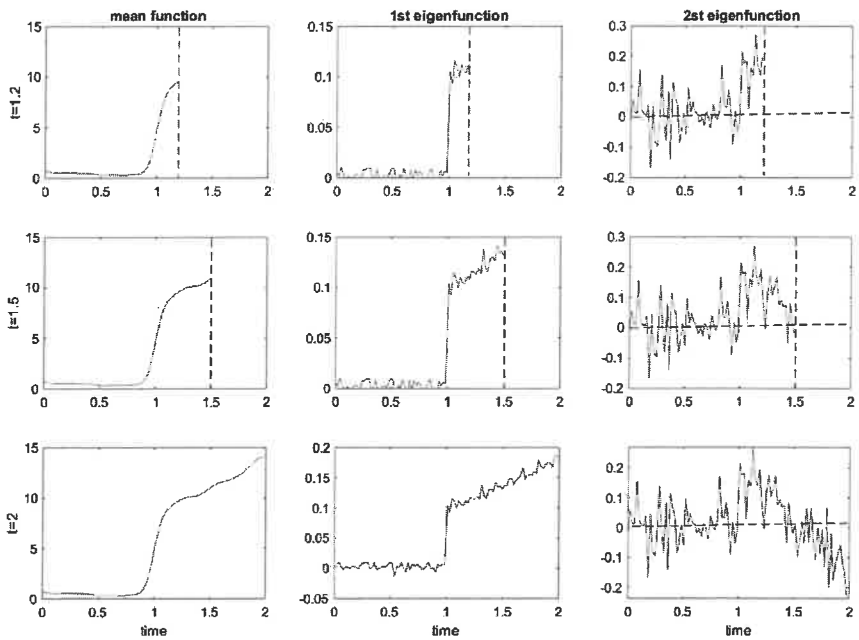

Figure 7: Evolution of mean functions(left column)for censoring fragmented data, and of first(middle column) and second(right column) eigenfunctions for current times t=1.2(first row),t=1.5(second row),t=2(third row) for degradation signal data

played in Figure 5 for fragmented data.As seemed, mean remaining lifetime function for non-censoring data approaches very well to true mean remaining lifetime function. However,although initial mean residual function for censoring data has gap with true average remaining lifetime, while time increases, that gap decreases. The root squared prediction error verifies the fact that non-uniform sampling models for fragmented data is can be applied to our methods. Furthermore, the Figure 7 proves that reduction of the dimension to just two eigenfunction simply the model sufficiently.

6. Conclusion and Discussion

The time-evolution of mean and eigenfunction is a concept that provides a stepping stones to extend to reach of functional data analysis to the analysis of trajectories that are truncated by death or other events.The interpretation of Figure 6-7 depicting this time-evolution and implying an reduction of the durability of each component of system, which proposed analysis tools can lead to interesting insight that would be hard to come for censored data in traditional methods. In studies on aging, predicted remaining lifetime is a useful measure for security of the system.The proposed methods yield estimates for such measures base on observed co-variate trajectories.

Under right censored Sparse/Fragmented sampling, the selection of the observation times of the degradation signals impacts the accuracy of the degradation censored modeling.For example, if the degradation signals are uniformly but fragmented censored sampled, the degradation process will not be adequately observed at the later extreme time point M, since few component will survive up to this time point. To apply the proposed method like this data, the time points at which the degradation signals have been observed cover the domain [0,M].Nonuniform sampling guarantees to be dense-sty of domain enough.

The proposed methods allow for straightforward inclusion of more than one predictor trajectory per subject and also additional multivariate co-varieties that may be available for each subject in some applications.

References

[1] D. R Cox. Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society*, series B 34:187–200, 1972.

[2] Carey J. R. Wu D. Liedo P. Vaupel J. W. Muller, H. G. Reproductive potential predicts longevity of female Mediterranean fruit flies. *Proceedings of the Royal Society*, B 268:445–450., 2001.

[3] C. J. Lu and W. Q. Meeker. Using degradation measures to estimate a time-to-failure distribution.. *Technometrics*, 35:161–174, 1993.

[4] Lawley M. Li R. Gebraeel, N. and J. Ryan. Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37:543–557, 2005.

[5] N. Gebraeel. Sensory updated residual life distribution for components with exponential degradation patterns.. *IEEE Transactions on Automation Science and Engineering.*, 3:382–393, 2006.

[6] H.-G. Muller and Y. Zhang. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories.. *Biometrics*, 61:1064–1075, 2005.

[7] C. Park and W. J. Padgett. Stochastic degradation models with several accelerating variables.. *IEEE Transactions on Reliability*, 55:379–390, 2006.

[8] E.L. Kaplan and P. Meier. Non-parametric estimation from incomplete observation. *JASA*, 282:457–841, 1967.

[9] V. Blum, J.R. Maximal deviation theory of density and failure rate function estimate

based on censored data.. *Multivariate Analysis(North-Holland,New York)*, 5:213–222., 1980.

[10] J. Mielniczuk. Kernel estimators of a density function in case of censored data., 1985.

[11] D.T. Mcnichols and W.J. Padgett. A modified kernel estimator for randomly right censored data. *South Africa statist.*, j.18:13–27, 1984.

[12] W. J. Padgett and M. A. Tomlinson. Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Anal*, 10:191–206, 2004.

[13] N. Serban N. Gebraeel R., Zhou. Degradation modeling applied to residual lifetime prediction using fucntional data analysis. *The annals of Applied statistics*, 5(2B):1586–1610, 2011.

[14] K. A. Doksum and A. Hoyland. Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution. *Technometrics.*, 34:174–82, 1992.

[15] L. I. Pettit and K. D. S. Young. Bayesian analysis for inverse Gaussian lifetime data with measures of degradation.. *J. Statist. Comput.*, Simulation 63:217–234, 2004.

[16] K. Karhunen. *Uber lineare Methoden in der Wahrscheinlichkeitsrechnung.* Suomalainen Tiedeakatemia,Finland, 1947.

[17] Muller H. G. Hall, P. Order-preserving nonparametric regression, with applications to conditional distribution and quantile function estimation. *Journal of the American Statistical Association*, 98:598–608, 2003.

[18] Muller H.-G. Yao, F. and J.-L. Wang. Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.*, 100:577–590, 2005.

[19] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer,New York, MR1964455, 2003.

[20] Ferraty F.-Mas A. Cardot, H. and P. Sarda. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, 30:241–255, 2003.

[21] K. M. Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93:228–237., 1998.

[22] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer,New York, MR1964455, 2003.

[23] Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional data. *Institute of Meathematical Statistics*, 38:3321–3351, 2010.

[24] H.-G Yao, F.Muller and J. Wang. Functional linear Regression analysis for longitudinal data. *Ann.Stat*, 33:2873–2903, 2005b.