MULTIVARIATE GENERALIZED FUNCTIONAL LINEAR MODELS WITH APPLICATIONS TO GENOMICS

By

Sneha Jadhav

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2017

ABSTRACT

MULTIVARIATE GENERALIZED FUNCTIONAL LINEAR MODELS WITH APPLICATIONS TO GENOMICS

By

Sneha Jadhav

This thesis is focused on developing functional data methodology with the aim of addressing problems that arise in genetic sequencing data. While significant progress has been made in identifying common genetic variants associated with diseases, these variants only explain a small proportion of heritability. Recent studies suggest that rare variants could account for this variability. With advancements in sequencing technology, large-scale sequencing studies are now being conducted to comprehensively investigate the contribution of rare variants to the genetic etiology of various diseases. Although these studies hold great potential for uncovering new disease-associated variants, the massive amount of data and complex structure of sequencing data poses great analytical challenges on association analysis. Advanced methods are needed to address these challenges and to facilitate the discovery process of new variants predisposing to various diseases. We use functional data analysis methods to capture the complexities of sequencing data.

In the first chapter we investigate the importance of considering the genetic structure of sequencing data. In association studies the effect of appropriately modeling genetic structure of sequencing data on association analysis have not been well studied. We compare three statistical approaches which use different strategies to model the genetic structure. They are a burden test, a burden test that considers pairwise correlation, and a functional analysis of variance (FANOVA) test that models the gene through fitting continuous curves on an individuals genotype profile. We find some evidence in favor of treating sequencing data as a function.

In the second chapter we present the definitions of some fundamental concepts in Functional

Data Analysis like the mean element, covariance operator and its eigen decomposition, and Karhunen-Loève expansion. Basis expansion and in particular Karhunen-Loève expansion play an important role in this thesis. We briefly discuss the estimators for the mean function, the covariance operator and their consistency. Results on the consistency of the eigenvalues and eigenfunctions of the sample covariance operator are also stated.

Several times genetic data is collected on families, where the response variable or the trait of the family members can be dependent on each other. Additionally, this trait of interest can be discrete or continuous. Thus there is a need for a functional model that can handle dependent data that may be continuous or discrete. The model proposed by Müller and Stadtmüller (2005) uses the generalized estimating equations approach that can handle both continuous and discrete data. However, they assume the response variable to be univariate and the sample to be independent. There are no existing functional methods that we know of that can be directly applied to the family data. In the third chapter we develop a framework for dependent generalized functional linear models where the response is multivariate, that can be used to test for a certain type of association between the genetic data and the trait of interest for family data.

In the fourth chapter we develop regression framework where the response variable has a normal distribution and there is measurement error in the regressor function. In this set-up, the true regressor function is not observable. Instead, we observe a surrogate variable and its replicates. The relation between the true function and the surrogate one is assumed to follow the additive classical measurement error model. We use the approach developed by Stefanski and Carroll (1987) to propose an estimating equation for the parameters and show asymptotic existence and consistency of the estimate obtained from this equation.

ACKNOWLEDGMENTS

I wish to express sincere gratitude to my advisor Prof. Hira L. Koul. I am thankful for his immeasurable support, guidance, encouragement and patience. I would also like to thank Prof. Qing Lu for giving me the opportunity to work with him on a variety of projects and broadening the scope of my work. He has been a very supportive mentor. I thank Prof. Sakhanenko for lending her ear to my ideas and offering her advice on several occasions and Prof Cui for serving on my dissertation committee.

A loving thanks to all my friends for making this a fun-filled journey. I would like to express my appreciation to the graduate secretary Sue Watson for looking out for us students. I am grateful to my husband Venkat Ramakrishnan for his unyielding patience, support and love. This thesis might not have been written without the support of my mum Hemlata Jadhav, my father Umakant Jadhav and my sister Shalmali Jadhav who have shaped me in to who I am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Modeling Sequencing Data

Genome-wide association studies have made substantial progress in identifying common variants associated with human diseases. Despite this, a large portion of heritability remains unexplained. Evolution theory and empirical human genetic studies suggest that rare mutations could play an important role in human diseases, which motivates comprehensive investigation of rare variants in sequencing studies. Advances in sequencing technology have enabled researchers to sequence exome regions or even the whole genome at affordable costs (Cirulli and Goldstein, 2010). The emerging sequencing data facilitates the study of massive amounts of single nucleotide variants, including both rare and commons variants for their potential role in complex human diseases. Although these studies hold great promise for identification of new disease-susceptibility variants, the extremely large number of single nucleotide variants (SNVs) brings significant challenge for association analysis. Each of these variants have a certain position on the chromosome that is also available to us. Given the linkage disequilibrium (Laird and Lange (2010a)) and that the genes that are closer together are more likely to be inherited together during meiosis, we should consider the positions of the variants in our analysis. This suggests that we can consider the genetic data as a function of the positions of the variants. Moreover, the sequencing data is usually high dimensional and we shall see in subsequent chapters that treating it as a function will help address the problems posed by its high dimensional nature. Treating this data a function does not require the data to be independent or assume a certain correlation structure. The functional approach can accommodate rare variants. To explore the association of rare variants with human diseases, many statistical approaches have been developed with different ways of modelling the genome so as to capture its properties to the fullest. Conventional single-locus analysis suffers from low power because of low frequency of SNVs and multiple testing issues. Grouping SNVs in a genetic region (e.g., gene) could aggregate the association signal and alleviate the multiple testing issues, and therefore has been widely used in statistical analysis of sequencing data (Lee et al. (2014)). Various statistical methods have been proposed to group SNVs with or without considering the underlying genetic structure. However, the impact of different strategies of modelling sequencing structure on the association results has rarely been investigated. If empirical evidence suggests no use of considering the sequencing structure in the association analysis, it gives us a basis for excluding this factor from statistical modelling. On the other hand, if it is important to consider the relationship among SNVs, then we need to investigate appropriate strategies for characterizing the underlying sequencing structure. As an initial step to investigate this issue, we choose three tests with different ways of modelling the correlation between SNVs: 1) A weighted burden test (BT) (Madsen and Browning, 2009); 2) A weighted burden test considering pairwise correlation (BTCOV) (Schaid et al., 2013); and 3) A functional analysis of variance test (FANOVA) (Vsevolozhskaya et al., 2014) that considers the relation among nearby loci and models the genotype profile of an individual as a continuous function.

We briefly present the three methods below.

1.1 Burden Test

We consider a burden test developed by Madsen and Browning (2009). The test summarizes the genetic score of multiple SNVs as $\gamma_j = \sum_{i=1}^{L} \frac{g_{ij}}{w_i}$, where g_{ij} is the number of low frequency allele of the SNV, i, for individual j. The weight is defined to emphasize the effect of rare variants i.e. $w_i = \sqrt{n_i q_i(1-q_i)}$, where q_i is the minor allele frequency (MAF) of the SNV, calculated from controls. Because the test simply adds the genotype of each SNV weighted by its MAF, it does not consider Linkage Disequilibrium (LD) or the correlation between SNVs.

1.2 Burden test that considers the pairwise Correlation

In addition to the above burden test, we also consider another type of burden test suggested by Schaid et.al Schaid et al. (2013), which considers pairwise covariance. We consider the following summary of genetic scores, $S_j = \sum_{i=1}^{L} \frac{g_{ij}}{w_i}$, w_l and g_{ij} are defined in the same manner in BT. However, unlike the conventional burden test, the test statistic of BTCOV is given by $T = \frac{((Y-\bar{Y})'S)^2}{(Y-\bar{Y})'V_s(Y-\bar{Y})}$, where $V_s = \sum_{k=1}^{L}\sum_{l=1}^{L} w_k w_l R_{kl}\sqrt{p_k(1-p_k)p_l(1-p_l)}$ and R_{kl} is the correlation between the SNPs.

1.3 Functional Analysis of Variance test

FANOVA fits a continuous function (curve) on the genotype data of an individual. ANOVA can then be used to test the curve difference in cases and controls. While various smoothing methods can be used to fit smooth functions on genotype data, we use the cubic B-splines to fit the smooth functions. After continuous functions g(t), are obtained, the functional analysis of variance can be used for association testing. FANOVA model is written as:

$$g_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t) \quad \epsilon_{ij}(t) \rightarrow G.P(0,\gamma) \quad i = 1,2,...,n_j \;\; j = 1,2,$$

where, t denotes the genomic position of the genetic variant, k denotes the case or control group and j denotes the individual, $G.P(0, \gamma)$ denotes Gaussian process with γ as the covariance function, ϵ_{kj} is the error term and μ_k is the mean function of group k.

To evaluate the association, we test the hypotheses: $H_0 : \mu_1(t) = \mu_2(t) \quad \forall t$ vs. $H_1 : \mu_1(t) \neq \mu_2(t)$ for some t. Similar to ANOVA, the test statistic for the hypothesis is,

$$F = \frac{\int \sum_{k=1}^{2} n_k(\hat{\mu}_k(t) - \hat{\mu}(t))^2 dt/(2-1)}{\int \sum_{k=1}^{2} \sum_{j=1}^{n_k} (g_{kj}(t) - \hat{\mu}_k(t))^2 dt/(n-2)},$$

where $\hat{\mu}_k(t) = n_k^{-1} \sum_{j=1}^{n_k} g_{kj}(t)$ and $\hat{\mu}(t) = n^{-1} \sum_{k=1}^{2} n_k \hat{\mu}_k(t)$. The numerator and the denominator in the F follow a mixture of chi-squared distributions. Satterwaite approximation is used to approximate the distribution of F as F-distribution. The details can be found in Vsevolozhskaya et al. (2014).

1.4 Simulation

We now report the finding of the simulation study that compares the performance of the three methods. We selected a one mb region from chromosome three of the unrelated real sequencing data provided by GAW19, which comprises of 8575 SNVs. For each replicate, we randomly selected a 30 kb segment from the one mb region. From this segment, we randomly selected a specified proportion of SNVs for generating phenotype. We used the logistic model to generate phenotype from these selected SNVs. The two types of effects were considered in the simulation. These effects refer to the coefficients of the SNVs in the logistic set up. We randomly generated the regression coefficients from $N(0, 1)$ for bidirectional effects and $N(2, 1)$ for unidirectional effects.

For each type of effect, we varied the proportion of disease-associated variants from 0.01 to 0.5. One thousand replicates were simulated for each scenario for power and type-I error calculation. For comparison, we adopted the same weight for both the burden tests. For FANOVA, we used the penalized cubic B-splines to determine the smoothness of functions.

These simulations only evaluated one genetic region. To investigate the performance of the three methods on regions with different genetic structures, we applied them to the unrelated simulated GAW data in order to evaluate the association of the 294 reported disease-associated genes with the simulated hypertension phenotype. For the association analysis, hypertension (HTN1) from the first simulation out of the 200 simulations was used. This data has a sample size of 142, out of which there were 24 cases.

1.5 Results

Type-I error rates of the three tests were well controlled at 0.05 level (0.046 for BT, 0.044 for BTCOV, and 0.047 for FANOVA). As we observe from Table 1, power of the three tests increases as we increase the proportion of disease-associated variants. Overall, FANOVA has better or comparable performance to BT and BTCOV, while BTCOV obtains similar power to BT. The same conclusion also holds when the effects are bidirectional (see Table 2). We also observe that the power of the three tests was slightly lower in the case of bidirectional effects than in the case of unidirectional effects. Table 3 summarizes the top ten genes with the smallest p-values from the association analysis. Consistent with the result from simulations, we find that in general FANOVA attains smaller p-values, while the p-values of BT and BTCOV were very similar. Here smaller p-values indicate better performance as these genes were from the 294 reported disease-associated genes.

Table 1.1: Power for unidirectional effect

%CV	.01	.05	.1	.15	.2	.25	.3	.5
BT	0.343	0.617	0.714	0.766	0.759	0.776	0.793	0.781
BTCOV	0.339	0.615	0.712	0.767	0.755	0.780	0.792	0.794
FANOVA	0.398	0.700	0.764	0.808	0.807	0.814	0.814	0.744

Table 1.2: Power for bidirectional effect

%CV	.01	.05	.1	.15	.2	.25	.3	.5
BT	0.208	0.508	0.585	0.621	0.665	0.678	0.703	0.683
BTCOV	0.200	0.509	0.579	0.618	0.663	0.668	0.698	0.680
FANOVA	0.217	0.597	0.683	0.732	0.765	0.799	0.809	0.815

Table 3 summarizes top 10 genes with the smallest p-values from the association analysis.

1.6 Discussion

Through this study, we observe that overall BT and BTCOV have a comparable performance. However, for one gene, THRA, BTCOV attained better performance than the other two tests. In the follow-up analysis, we observe a small LD block in this gene (see Figure 1). The plot of the fitted genotype curves reveal the association happens to lie in that LD block. Therefore, BTCOV, which models the LD pattern, outperforms the other two tests. Also, the effects in the LD block

Table 1.3: Summary of top 10 genes with the smallest p-values from the association analysis

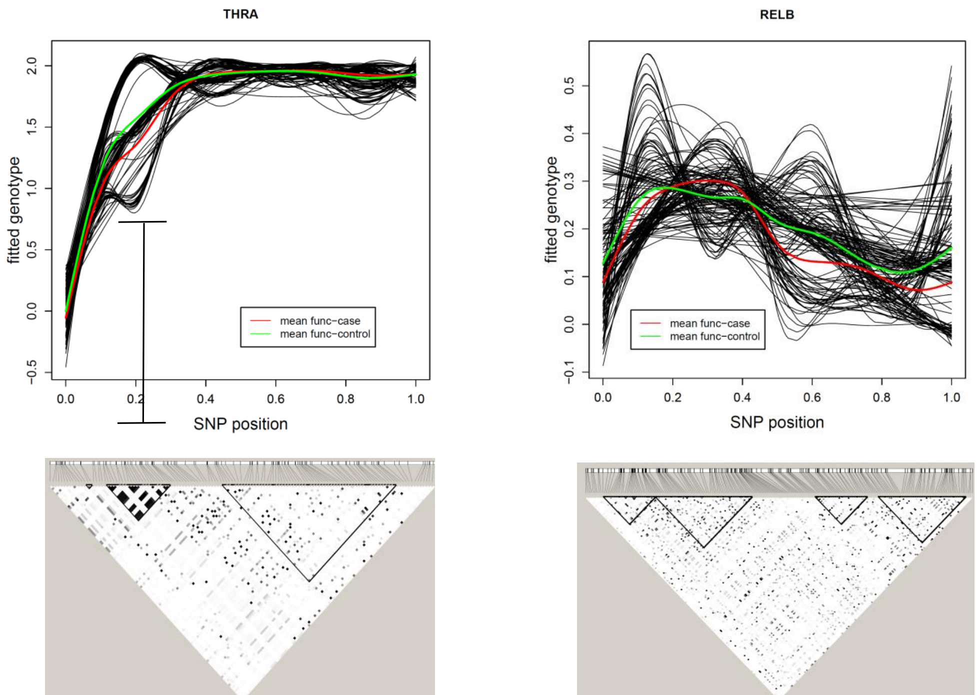
Gene	BT	FANOVA	BTCOV
SUMF1	8.75E-05	1.31E-05	8.43E-05
RELB	6.57E-02	4.35E-04	7.08E-02
HIF3A	2.12E-02	4.34E-03	2.19E-02
THRA	1.85E-02	3.62E-02	9.68E-03
TFDP1	1.50E-02	1.95E-02	1.13E-02
PROK2	1.23E-02	1.27E-01	1.32E-02
POLR2A	2.10E-02	1.31E-02	2.25E-02
CD1C	2.76E-02	5.27E-02	1.42E-02
CCL24	2.24E-02	1.92E-02	2.71E-02
MAP3K6	8.72E-02	3.27E-02	8.45E-02

were largely unidirectional, which is in favor of burden tests. The genetic structure of sequencing variants is usually more complex than pairwise LD. Functional methods provide tool to visualize this structure and explore the associated regions. Figure 1 reveals the interesting structure of RELB. We can see that the associated region lies in LD block but the correlation is not as strong as that for THRA. FANOVA performs much better for this gene.

1.7 Conclusion

Our observations indicate that the performance of tests depends on the underlying genetic structure and hence ignoring it in the association analysis may not be ideal. It is advisable to use function based approaches to explore and model the sequencing structure. As illustrated by Figure 1 (fitted genotype function vs variant position), the plot of fitted functional curves provides a reasonable way to explore the genetic structure. The disease-associated regions can also be visualized in the plot. If the underlying genetic structure tends to be complex, it is also advisable to use function based approaches, such as FANOVA, to adequately model the sequencing data.

Figure 1.1: LD plots and plots of the fitted smooth functions for THRA



Chapter 2

Preliminaries on Functional Data Analysis

Functional data can be viewed as realizations of a random variable that takes values in a Hilbert space. In this light, we briefly introduce the concept of mean element and covariance operator in Hilbert space and give some necessary definitions. We can also think of functional data as the sample paths of a stochastic process with smooth mean and covariance functions. Both these perspectives are discussed in a greater detail in Hsing and Eubank (2015). We briefly discuss the estimation of the mean function and the covariance operator in the space $L^2[0,1]$ of all square integrable functions on domain $[0,1]$. We also discuss the concept as well as the estimation of eigenvalues and eigenfunctions of the covariance operator. Further details of these concepts can be found in Horváth and Kokoszka (2012).

2.1 Definitions

Let f be a function on a measure space $(E, \mathcal{B}, \mu)$ that takes values in a separable Hilbert space $\mathcal{H}$.

Definition 2.1.1 *A function f is called simple if it can be represented as: $f(\omega) = \sum_{i=1}^{k} I_{E_i}(\omega) g_i$, for some finite k, $E_i \in \mathcal{B}$ and $g_i \in \mathcal{H}$.*

Definition 2.1.2 *Any simple function $f(\omega) = \sum_{i=1}^{k} I_{E_i}(\omega) g_i$ with $\mu(E_i) < \infty, \forall i$, is said to be integrable and it's Bochner integral is defined as: $\int f d\mu = \sum_{i=1}^{k} \mu(E_i) g_i$.*

Definition 2.1.3 *A measurable function f is said to be Bochner integrable if there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of integrable simple functions such that $\lim_{n\to\infty} \int ||f_n - f|| = 0$. The Bochner integral of f is defined as $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$.*

The existence of the sequence of functions f_n in the above definition is guaranteed by the following condition: $\int ||f|| d\mu < \infty$.

Definition 2.1.4 *Let $\{e_i\}$ be an orthonormal basis for Hilbert space $\mathcal{H}_1$ and X be a bounded linear function from $\mathcal{H}_1 \to \mathcal{H}_2$ where $\mathcal{H}_2$ is also a Hilbert space. If X satisfies $\sum_{i=1}^{\infty} ||Xe_i||_2^2 < \infty$, where $|| \quad ||_2$ is the norm on $\mathcal{H}_2$ then X is a Hilbert Schmidt operator. The collection of all such Hilbert Schmidt operators is denoted by $\mathcal{B}_{HS}(\mathcal{H}_1, \mathcal{H}_2)$. This is a Hilbert space. If $X_1, X_2 \in \mathcal{B}_{HS}(\mathcal{H}_1, \mathcal{H}_2)$ then $\langle X_1, X_2 \rangle_{HS} = \sum_{i=1}^{\infty} \langle X_1 e_i, X_2 e_i \rangle_2$.*

Definition 2.1.5 *Let $X_1 \in \mathcal{H}_1$ and $X_2 \in \mathcal{H}_2$. The tensor product $X_1 \otimes X_2$ is an operator defined from $\mathcal{H}_1 \to \mathcal{H}_2$ in the following way: $X_1 \otimes X_2(Y) = \langle X_1, Y \rangle X_2$, where $Y \in \mathcal{H}_1$.*

Let X be a random element of a separable Hilbert space $\mathcal{H}$ defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The norm on this space is denoted by $|| \quad ||$ and the inner product by $\langle \cdot, \cdot \rangle$.

Definition 2.1.6 *If $E||X|| < \infty$, the mean of X is defined as the Bochner integral $\mu = E(X) = \int X d\mathcal{P}$.*

Definition 2.1.7 *Assume that $E||X||^2 < \infty$. Then, the covariance operator for X is the element of $\mathcal{B}_{HS}(\mathcal{H})$ given by $\mathcal{K} = E[(X - \mu) \otimes (X - \mu)] = \int (X - \mu) \otimes (X - \mu) dP$.*

The Hilbert space of particular interest to us is the $L^2([0,1])$ space. We say that a function X belongs to the space $L^2 = L^2([0,1])$ if X is defined on $[0,1]$ and satisfies $\int_0^1 X^2(t)d(t) < \infty$. L^2 space is a separable Hilbert space with the following inner product: $\langle X, Y \rangle = \int_0^1 X(t)Y(t)dt$.

We consider a random curve $X(t), t \in [0,1]$ to be a random element of L^2. The mean function is $\mu(t) = E(X(t))$ and the covariance function is $K(s,t) = E(X(s)X(t))$. We can see that $\mathcal{K}(Y)(t) = E(\langle X, Y\rangle X(t)) = E \int X(s)Y(s)ds)X(t) = \int E(X(t)X(s))Y(s)ds = \int K(t,s)Y(s)ds$. We can show that the operator $\mathcal{K}$ is symmetric and positive definite.

Definition 2.1.8 *Suppose that there exists a λ such that $\mathcal{K}e = \lambda e$, then λ is the eigenvalue and e is the eigenfunction of $\mathcal{K}$.*

Note that we can show that the eigenfunctions of $\mathcal{K}$ are linearly independent. We can orthonormalize them using Gram Schmidt orthonormalization and use them as basis functions. We use this fact in our simulations extensively to generate basis functions.

2.2 Estimation

Now that we have presented some basic definitions, we turn to estimation of the defined constructs. Assume that we have a sample of curves $X_1, ...X_n$ from L^2, that are independent and have the same distribution as X. We assume that X is integrable.

The mean function estimate is given by the sample mean:

$$\hat{\mu}(t) = n^{-1} \sum_{i=1}^{n} X_i(t).$$

The covariance function estimate is given by its sample counterpart:

$$\hat{K}(t,s) = n^{-1} \sum_{i=1}^{n} (X_i(s) - \hat{\mu}(s))(X_i(t) - \hat{\mu}(t)).$$

Similarly the covariance operator estimate is given by

$$\hat{\mathcal{K}}(x) = n^{-1}\sum_{i=1}^{n}\langle X_i - \hat{\mu}, x\rangle (X_i - \hat{\mu}), \quad x \in L^2.$$

We state some results that establish the properties of these estimates.

Result 2.2.1 *Under the assumption that sample of curves is i.i.d, integrable and has the same distribution as X, $E(\hat{\mu}) = \mu$ and $E||\hat{\mu} - \mu||^2 = O(n^{-1})$.*

Thus, sample mean is an unbiased and consistent. Thus, from now on we can assume that the mean function $\mu = 0$. The estimate for covariance is biased just like in the multivariate case. This bias is asymptotically negligible.

Result 2.2.2 *If $E||X||^4 < \infty$, $EX = 0$ then $E||\hat{\mathcal{K}} - \mathcal{K}||^2_{HS} = n^{-1}E||X||^4$.*

We often need to estimate the eigenvalues and eigenfunctions of the the covariance operator. Thus, the estimates of these eigenvalues are given by $\hat{\mathcal{K}}\hat{v} = \hat{\lambda}\hat{v}$. Note that if v is an eigenfunction then av is also an eigenfunction. The eigenfunctions are usually normalized so that $||v|| = 1$. This does not determine the sign of v. If $\hat{k}_j = sign(\langle \hat{v}, v\rangle)$, then $\hat{k}_j$ cannot be determined from the data.

Result 2.2.3 *Suppose $E||X_i||^4 < \infty$, $EX = 0$ and $\lambda_1 >, ..., > 0$. Then, for each $j \geq 1$, $\lim\sup_{n\to\infty} \quad nE(||\hat{k}_j\hat{v}_j - v_j||^2) < \infty$, $\lim\sup_{n\to\infty} \quad nE(|\lambda - \hat{\lambda}_j|^2) < \infty$. If we assume that only the top $p+1$ eigenvalues are non zero then the above result holds for $1 \leq j \leq p$.*

We now state the Mercers theorem and Karhünen Loéve expansion. Let $(E, \mathcal{B}, \mu)$ be a measure space. Suppose that K is a measurable continuous function on $E \times E$ such that $\int\int K(s,t)d\mu(s)d\mu(t) < \infty$. Define, operator $\mathcal{K}$ by $\mathcal{K}f(\cdot) = \int K(s,\cdot)f(s)d\mu(s)$. $\mathcal{K}$ is the integral operator of K and K is the kernel function of $\mathcal{K}$.

Definition 2.2.1 *The kernel function K is symmetric if $K(s,t) = K(t,s)$.*

Definition 2.2.2 *The kernel function is non negative definite if $\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0$ holds for all $n \in \mathbb{N}, x_1, ...x_n \in E, c_1, ...c_n \in \mathbb{R}$. It's positive definite if the relation is strictly greater than zero.*

The following result is the Mercers Theorem. Notice that the covariance function has this representation.

Result 2.2.4 *Let K be a symmetric, non negative definite kernel function and $\mathcal{K}$ be its integral operator. If (λ_i, e_i) are the eigenvalue and eigenfunction pairs of $\mathcal{K}$, then K has the representation $K(s,t) = \sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t)$.*

The following result is the Karhünen Loéve expansion.

Result 2.2.5 *Let X be a zero-mean, square-integrable random function defined over a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and indexed over a closed and bounded interval $[a,b]$, with continuous covariance function $K(s,t)$. Let (λ_i, e_i) be the eigenvalue and eigenfunction pairs of the integral operator of the covariance function. Then, $X(t)$ admits the following decomposition: $X(t) = \sum_{i=1}^{\infty} Z_i e_i(t)$, where $Z_k = \int X(t) e_k(t) dt$. Furthermore, the random variables Z_k have zero mean, are uncorrelated and have variance λ_k.*

Chapter 3

Multivariate Generalized Functional Linear Models

3.1 Introduction

The focus of this chapter is to provide a large sample test for assessing if a functional covariate has a regression effect on real valued responses, when there is some dependence among responses. This kind of data typically arises in family based genetic studies when gene expression data or even DNA sequencing data is collected. The aim of these studies is often to test for regression relation between the gene region and the phenotype of interest. Sequencing data for a gene or a gene region consists of observations on a large number of single nucleotide variants. In the light of linkage disequilibrium (Laird and Lange, 2010b), we know that the variants that are closer to each other may have greater association than those that are farther from each other. This motivates us to treat the high dimensional sequencing data as a function of the single nucleotide variant positions and necessitates the use of a functional data based inference method for correlated data.

There is abundant literature on functional linear models (Cardot et al., 1999, 2003; Cardot and Sarda, 2005; Cardot and Johannes, 2010; Ramsay, 2006). Recent reviews of functional regression can be found in Morris (2015) and Wang et al. (2016). These linear models assume that the response is a univariate continuous variable and the regressor variable is a function. However, in genetic studies the phenotype or the response variable is often binary and few papers discuss generalized functional linear models (Müller and Stadtmüller, 2005; Li et al., 2010; Gertheiss et al., 2013) as needed for handling binary response variables. Existing methodology cannot be directly applied to genetic family data, where the response variable is a vector of dependent traits and the regressor is a vector of functions. We address this shortcoming using generalized estimating equations. The estimators thus obtained are shown to be consistent and asymptotically normal, under suitable conditions on the underlying entities. The latter result is used to propose an asymptotic test for regression relation between the functional covariates and the responses.

The chapter also includes a finite sample simulation study. Through this study, we first demonstrate the importance of addressing correlation structures in the data and then compare the performance of our proposed method with another method for family studies proposed by Wang et al. (2013). This method is based on a generalized estimating equations approach suitable for accommodating a large number of variables. We also present additional results demonstrating the effect of sample size, the dimension of parameter to be estimated and family (cluster) size on the power of the proposed test.

This chapter is organized as follows. Section 2 contains formal description of the original infinite dimensional model and a working truncated model based on the strategy proposed by Müller and Stadtmüller (2005). Section 3 contains the description of the estimators based on generalized estimating equations while section 4 describes the asymptotic normality results for these estimators. Section 5 describes findings of a simulation study while section 6 presents a real data application.

3.2 Model

Let m, n be positive integers. We observe n clusters $(X_i(t), t \in T = [0,1], Y_i), i = 1, ..., n,$ each of size m, where, for each i, $X_i = (X_{i1}, X_{i2}, \cdots, X_{im})^T$ is a an m-dimensional predicting process and $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{im})^T$ is a vector of m responses. We assume that these n clusters are independent and identically distributed and that all clusters have the same correlation structure. For jth subject in ith cluster the predicting process $X_{ij}(t), t \in T$, is assumed to be a square integrable random process on T. The corresponding response Y_{ij} can be continuous or discrete, real valued variable which is related to $X_{ij}(t), t \in T$ via the following generalized regression model, where ω is a positive measure on T. For a constant $\tilde{\beta}_0$, and a real valued function $\tilde{\beta}(t)$, $t \in T$, let

$$\tilde{\zeta}_{ij} = \tilde{\beta}_0 + \int \tilde{\beta}(t) X_{ij}(t) d\omega(t), \quad i = 1, ..., n, \quad j = 1, ..., m.$$

All integrals in this paper are taken over the interval T, unless specified otherwise. In the rest of this paper, $i = 1, ..., n$ and $j = 1, ..., m$. We model the regression of Y_{ij} on X_{ij} as

$$\begin{aligned} &Y_{ij} = g(\tilde{\zeta}_{ij}) + e_{ij}, \quad \mu_{ij} = E(Y_{ij} \mid X_{ij}(t), t \in T) = g(\tilde{\zeta}_{ij}), \qquad (3.2.1) \\ &\widetilde{\sigma}^2(\mu_{ij}) = \text{var}(Y_{ij} - \mu_{ij} \mid X_{ij}(T), t \in T) = \sigma^2(\tilde{\zeta}_{ij}), \quad e_i = (e_{i1}, e_{i2}, ..., e_{im})^T, \\ &R_0 = \text{cor}(e_i), \end{aligned}$$

for a known real valued link function g and a positive function σ, where e_{ij} have zero mean. R_0 is $m \times m$ true correlation matrix.

We now give another representation of the model (3.2.1). Let $\rho_j, j = 1, 2, ...,$ be an orthonormal basis of the functional space $L^2(T, \omega) = L^2(\omega)$. The predictor process and parameter function can be written as

$$X_{ij}(t) = \sum_{k=1}^{\infty} \varepsilon_{ij}^{(k)} \rho_k(t), \quad \tilde{\beta}(t) = \sum_{k=1}^{\infty} \tilde{\beta}_k \rho_k(t), \tag{3.2.2}$$

with random variables $\varepsilon_{ij}^{(k)}$=$\int X_{ij}(t)\rho_k(t)d\omega(t)$, and coefficients $\tilde{\beta}_k = \int \tilde{\beta}(t)\rho_k(t)d\omega(t)$. By assumption 3.4.1 below, we obtain $E(\varepsilon_{ij}^{(k)}) = 0$, for all i, j, and k. Note that the random variables $\varepsilon_{ij}^{(k)}$ and $\varepsilon_{ij}^{(l)}$ are uncorrelated for $k \neq l$. By the orthonormality of the basis functions,

$$\sum_{k=1}^{\infty} \widetilde{\sigma}_{ij}^{(k)2} = \int E(X_{ij}^2(t))d\omega(t) < \infty, \qquad \int \tilde{\beta}(t)X_{ij}(t)d\omega(t) = \sum_{k=1}^{\infty} \tilde{\beta}_k \varepsilon_{ij}^{(k)}.$$

Using the above representation we now address the infinite dimensionality issue or equivalently the issue of infinite number of predicting variables. Based on the truncation strategy proposed by Müller and Stadtmüller (2005) we replace model (3.2.1) with the following approximate sequence of finite dimensional models. Let $p = p_n$ be a sequence of positive integers tending to infinity and define our new approximate model as

$$\begin{aligned} &Y_{ij} = g\Big(\tilde{\beta}_0 + \sum_{k=1}^{p_n} \tilde{\beta}_k \varepsilon_{ij}^{(k)}\Big) + e_{ij}, \quad \tilde{\eta}_{ij}^{p_n} = \tilde{\beta}_0 + \sum_{k=1}^{p_n} \tilde{\beta}_k \varepsilon_{ij}^{(k)}, \\ &U_{ij}^{(p_n)} = \sum_{k=1}^{p_n} \varepsilon_{ij}^{(k)} \rho_k(t), \quad \sigma^2(\eta_{ij}^{(p_n)}) = \widetilde{\sigma}^2(\mu_{ij}^{(p_n)}) = \text{var}(e_{ij}|U_{ij}^{(p_n)}) \\ &e_i = (e_{i1}, e_{i2}, ..., e_{im})^T, \quad R_0 = \text{cor}(e_i). \end{aligned} \tag{3.2.3}$$

Let $\tilde{\beta} = (\tilde{\beta}_0, ..., \tilde{\beta}_{p_n})^T$. The superscript p_n indicates the number of parameters. We exhibit this superscript and subscript when necessary. We assume that the standardized error $e_{ij}\tilde{\sigma}(\mu_{ij})$ is independent of $\varepsilon_{ij}^{(k)}$ for all $i,\ j,\ k$.

In the sequel, $N_m(\mu, \Sigma)$ stands for m-dimensional normal distribution with the mean vector μ and covariance matrix Σ, $m \geq 1$, and all limits are taken as $n \to \infty$, unless specified otherwise.

For any vector or a finite dimensional matrix A, $\|A\|$ denotes the Frobenius norm of A.

3.3 Estimation

We use the generalized estimating equation set-up for estimating the parameter $\tilde{\beta}$. In most applications, we do not know the true correlation matrix R_0, so we use a working correlation matrix $R(\gamma)$ to specify the working covariance. The parameter γ can be estimated using the residuals. We use $\hat{R}$ to denote an estimated working correlation matrix. The estimate $\hat{R}$ that we use below was suggested by Balan et al. (2005). Denote the derivative of function g as g'. The estimator denoted by $\hat{\beta}$ is the solution to the following equation.

$$U^*(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i(\beta)^{-1/2} \hat{R}^{-1} A_i(\beta)^{-1/2} e_i(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) \hat{V}_i(\beta)^{-1} e_i(\beta) = 0, \tag{3.3.1}$$

where,

$$G_i(\beta) = \begin{bmatrix} g'_{i1}(\beta) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & g'_{im}(\beta) \end{bmatrix}, \ F_i^T = \begin{bmatrix} \varepsilon_{i1}^{(0)} & \varepsilon_{i2}^{(0)} & \dots & \varepsilon_{im}^{(0)} \\ \varepsilon_{i1}^{(2)} & \varepsilon_{i2}^{(2)} & \dots & \varepsilon_{im}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{i1}^{(p)} & \varepsilon_{i2}^{(p)} & \dots & \varepsilon_{im}^{(p)} \end{bmatrix}, \ e_i(\beta) = \begin{bmatrix} y_{i1} - g_{i1}(\beta) \\ \vdots \\ y_{im} - g_{im}(\beta) \end{bmatrix},$$

$$A_i(\beta) = diag\big(\sigma_{i1}^2(\beta), ..., \sigma_{im}^2(\beta)\big), \quad \hat{R} = \frac{1}{n}\sum_{i=1}^{n} A_i(\check{\beta})(Y_i - g_i(\check{\beta}))(Y_i - g_i(\check{\beta}))^T A_i(\check{\beta})$$

$$\hat{V}_i(\beta) = A_i(\beta)^{-1/2} \hat{R} A_i(\beta)^{-1/2}.$$

Let $\varepsilon_{ij} = 1, \varepsilon_{ij} = (\varepsilon_{ij}^{(0)}, ..., \varepsilon_{ij}^{(pn)})^T, g_{ij}(\beta) = g(\varepsilon_{ij}^T\beta), g_i(\beta) = (g_{ij}(\beta), ..., g_{ij}(\beta)), \sigma_{i1}(\beta) = \sigma(\varepsilon_{ij}^T\beta)$ and $\check{\beta}$ be a preliminary $\sqrt{n/p_n}$ consistent estimate of β. It can be obtained using the independence estimating equations. Further details of this estimator can be found in Wang et al. (2011).

3.4 Asymptotic Theory

3.4.1 Existence and Consistency

We first state the assumptions needed for existence of the solution of (3.3.1) and its consistency. Note that $\lambda_{max}(A), \lambda_{min}(A)$ denote the maximum and minimum eigenvalues of a matrix A respectively. Also $\rightarrow_D$ ($\rightarrow_P$) denote convergence in distribution (probability).

Assumption 3.4.1 *Assume that $E\{X(t)\} = 0$ for all t, $E\|X\|^4 < \infty$ and $\|\tilde{\beta}\|^2 < \infty$.*

Assumption 3.4.2 *Function g is monotone, invertible, and has two continuous bounded derivatives. The function σ^2 has a continuous bounded derivative and is bounded from below by $\delta > 0$.*

Assumption 3.4.3 *Assume that $p_n n^{-1/2} \rightarrow 0$.*

Assumption 3.4.4 *The true correlation matrix R_0 has positive eigenvalues. The estimated working correlation matrix $\hat{R}$ satisfies $\|\hat{R}^{-1} - \bar{R}^{-1}\| = O_p\{(p_n/n)^{1/2}\}$. $\bar{R}$ is a constant positive definite matrix with positive eigenvalues.*

Assumption 3.4.5 *There exist two positive constants b_1, b_2 such that $b_1 \leq \lambda_{min}\{E(F_1^T F_1)\} \leq \lambda_{max}\{E(F_i^T F_i)\} \leq b_2$.*

We assume that the mean function of the regressor functions is zero to ease some of the calculations. Note that since we can obtain a consistent estimator of the mean function this assumptions is easily satisfied. We assume that $E\|X\|^4 < \infty$ mainly to obtain consistent estimates of the eigenvalues (Horváth and Kokoszka, 2012). Assumption 4, 5 and 6 can also be found in Wang et al. (2011). We prove that $\hat{R}$ used in the score function satisfies the assumption 4.

Theorem 3.4.1 *Under the assumptions 3.4.1–3.4.5, the solution $\hat{\beta}$ to* (3.3.1) *exists and satisfies the following*

$$||\hat{\beta} - \tilde{\beta}|| = O_p(\sqrt{p_n/n})$$

The proof is along the same lines as Wang et al. (2011) and can be found in the appendix. We approximate $U^*(\beta)$ by $\bar{U}(\beta) = \sum_{i=1}^n F_i^T G_i(\beta) A_i(\beta)^{-1/2} \bar{R}^{-1} A_i(\beta)^{-1/2} e_i(\beta)$. Thus, we can consistently estimate the parameters even if the correlation structure is misspecified.

3.4.2 Asymptotic Normality

To show the asymptotic normality of the estimator we will approximate $U^*(\beta)$ by $U(\beta) = \sum\limits_{i=1}^{n} F_i^T G_i(\beta) A_i(\beta)^{-1/2} R_0^{-1} A_i(\beta)^{-1/2} e_i(\beta)$. We will now rewrite the $U(\beta)$ as

$$U(\beta) = D^T(\beta) V(\beta)^{-1/2} (Y - \mu(\beta)),$$

where,

$$V_i(\beta) = A_i(\beta) R_0^{-1} A_i(\beta), \quad i = 1, ..., n, \quad V(\beta) = \begin{bmatrix} V_1(\beta) & 0 & \dots & 0 \\ 0 & V_2(\beta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_n(\beta) \end{bmatrix}, \tag{3.4.1}$$

$$D(\beta)=\begin{bmatrix} V_1^{-1/2}(\beta) & 0 & \dots & 0 \\ 0 & V_2^{-1/2}(\beta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_n^{-1/2}(\beta) \end{bmatrix} \begin{bmatrix} G_1(\beta) & 0 & \dots & 0 \\ 0 & G_2(\beta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G_n(\beta) \end{bmatrix} \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix},$$

$$\begin{aligned} \Gamma(\beta) &= \frac{1}{n}E(D(\beta)^T D(\beta)) = (\gamma_{kl}(\beta))_{0\le k,l\le p}\,, \quad \Xi(\beta)=\Gamma(\beta)^{-1}=(\xi_{kl}(\beta))_{0\le k,l\le p}, \\ J(\beta) &= \frac{dU(\beta)}{d\beta}, \quad Y=(Y_{11},Y_{12},...,Y_{1m},...,Y_{nm})^T, \quad R_0^{-1}=\begin{bmatrix} l_{11} & \dots & l_{1m} \\ \vdots & \vdots & \ddots \\ l_{m1} & \dots & l_{mm} \end{bmatrix}. \end{aligned}$$

Next, we shall state the needed assumptions for establishing the asymptotic normality of $\hat{\beta}$.

Assumption 3.4.6 *Assume that* $p_n n^{-1/8} \to 0$.

Assumption 3.4.7 *The matrices* $\Gamma = \Gamma(\tilde{\beta}) = n^{-1}E\{D(\tilde{\beta})^T D(\tilde{\beta})\}$ *and* $n^{-1}D(\tilde{\beta})^T D(\tilde{\beta})$ *are non-singular for all* n.

Assumption 3.4.8 *The eigenvalues of* Γ *are bounded and* $\left\| \left\{ \frac{D(\tilde{\beta})^T D(\tilde{\beta})}{n} \right\}^{-1} \right\| = O_p(p_n^{1/2})$.

We are now ready to state the theorem regarding the normality of the estimate. Let $\tilde{\Gamma} = n^{-1}D(\tilde{\beta})^T D(\tilde{\beta})$ and $\hat{\Gamma} = n^{-1}D(\hat{\beta})^T D(\hat{\beta})$. Note that $\tilde{\Gamma}$ depends on the unknown parameter where as $\hat{\Gamma}$ is a plug in estimator of $\tilde{\Gamma}$.

Theorem 3.4.2 *Under the assumptions 1–9, the following statements hold.*

$$\frac{n(\hat{\beta}-\tilde{\beta})^T\Gamma(\hat{\beta}-\tilde{\beta})-(p_n+1)}{\sqrt{2(p_n+1)}} \to_D N(0,1). \tag{3.4.2}$$

$$\frac{n(\hat{\beta}-\tilde{\beta})^T\tilde{\Gamma}(\hat{\beta}-\tilde{\beta})-(p_n+1)}{\sqrt{2(p_n+1)}} \to_D N(0,1). \tag{3.4.3}$$

$$\frac{n(\hat{\beta}-\tilde{\beta})^T\hat{\Gamma}(\hat{\beta}-\tilde{\beta})-(p_n+1)}{\sqrt{2(p_n+1)}} \to_D N(0,1). \tag{3.4.4}$$

Note that this theorem can be used to construct $1-\alpha$ confidence bands. Refer to corollary 4.3 in Müller and Stadtmüller (2005). The proofs of both the theorems are given at the nd of this chapter.

We are now ready to describe a test for the problem of testing for no regression relation between a real valued response and a functional predicting variable. Referring to the model (3.2.1), testing for no association between the response and the predicting process is equivalent to testing for $H_0 : \tilde{\beta} = 0$. Since we use the sequence of approximate models proposed in (3.2.3), we test the following hypothesis instead: $H_0 : (\tilde{\beta}_1, ..., \tilde{\beta}_{p_n}) = (0, ..., 0)$, versus the alternative that H_0 is not true, for a given appropriate value of p_n. The proposed test rejects H_0 for the large absolute values of the statistic

$$D_n = \frac{n\hat{\beta}^T\hat{\Gamma}\hat{\beta}-(p_n+1)}{\sqrt{2(p_n+1)}}.$$

From Theorem 2 it follows that the test that rejects H_0 whenever $|D_n| > z_{\alpha/2}$ is of the asymptotic size α, where z_α is the $100(1-\alpha)$th percentile of standard normal distribution.

3.5 Simulation

In this section we report the finding of several simulation studies, which investigate different aspects of the proposed testing procedure. In the first study we demonstrate the importance of considering the correlation structure in the data. We then explore the applicability of our method to sequencing data. In this study, we use sequencing data in the simulation to demonstrate the applicability of our method to genetic studies. We also compare the empirical power of our test to that of gSKAT proposed by Wang et al. (2013). We then study several other aspects of the proposed test by varying the sample and cluster sizes and the dimensions p_n.

For the first simulation we generated pseudo-random regression functions using Fourier basis functions $\rho_j, j \geq 1$, and the following model: $X(t) = \sum_{k=1}^{p_n} \varepsilon^k \rho^k(t), \varepsilon^j$ are independent and identically distributed with $N(0,1)$ distribution. We then evaluated each function on a grid of finite points to obtain functional data that are realizations of an underlying function. We take this approach as in applications only finite discretization of functional data is available. We then applied a smoothing procedure to this discrete data to obtain smooth functions that are used in our proposed test. This smoothing procedure is described later in the section. We simulated the effect function $\widetilde{\beta}(t)$ in the following way: $\widetilde{\beta}(t) = \sum_{k=1}^{p_n} \tilde{\beta}_k \rho^k(t)$, $\tilde{\beta}_k = k^{-1}\delta, k \geq 1$. Here, δ determines the magnitude of the effect of the regression function on the response. The choice of the number of basis functions p_n selected to generate each of these functions is set to $10n^{1/7}$.

We studied the performance of our method when the response is continuous and binary. For cluster size $m = 3$ and the case of continuous response we used the following model to generate the response variables: $Y_i = \int_0^1 X_i(t)\widetilde{\beta}(t)dt + e_i$, $e_i \sim N_3(\bar{0}, R)$ where, $\bar{0}$ is a vector of zeroes of length 3 and R is a 3×3 correlation matrix with all off diagonal elements equal to γ, $|\gamma| < 1$.

For binary response, we generated the correlated responses using the function from the bindata package in R software with marginal probabilities given by $g\big(\int_0^1 X(t)\widetilde{\beta}(t)dt\big)$, where g is the logistic link function. The correlation matrix is the same as in the previous case.

We then studied the empirical power of the proposed test of no regression relation by considering the underlying correlation structure and treating the observations as being independent. We label these two approaches as F test and Ind test, respectively. In these power studies, we chose $n = 500$, replicated 1000 times and the significance level $\alpha = 0.05$. The value of δ is fixed at 0.08 for continuous traits and at 0.1 for binary ones. The number of parameters p_n was determined using five fold cross validation procedure (see the end of this section about how we chose the value of p_n). We can see from Table 3.1 that as the correlation increases between the individuals the empirical power of the F test increases whereas that of the Ind test remains more or less the same as expected. This demonstrates the need to factor in the correlation structure. The empirical level for this study can be found in Table 3.2. For the second study, we compared the performance in

Table 3.1: Empirical power as a function of γ

Corr(γ)	Continuous Trait		Binary Trait	
	F Test	Ind Test	F Test	Ind Test
0	0.421	0.418	0.137	0.127
0.05	0.456	0.45	0.130	0.124
0.30	0.567	0.457	0.165	0.124
0.50	.692	0.443	0.195	0.129
0.80	0.993	0.455	0.308	0.134

Table 3.2: Empirical level as a function of γ

Corr(γ)	Continuous Trait		Binary Trait	
	F Test	Ind Test	F Test	Ind Test
0	0.03	0.051	0.039	0.032
0.05	0.048	0.049	0.036	0.035
0.30	0.045	0.051	0.032	0.029
0.50	0.059	0.043	0.043	0.037
0.80	0.068	0.047	0.041	0.037

terms of power of our test and gSKAT for testing of no regression relation. Note that gSKAT is a sequence kernel association tests that is based on generalized estimating equations. It is not meant for functional variables though it can handle cases where the number of variables is larger than the sample size. To make the comparison fair and to investigate the treatment of sequencing data as a function we used real genetic data from the 1000 Genome Project (**?**) and then simulated response values using this data. In particular, we used a region of the genome (Chromosome 17) from this dataset for following simulation. Note that this a not a family data set and thus, all the individuals are assumed to be independent. The primary focus of this simulation is to investigate the treatment of sequencing data as a function.

In real applications, we do not observe the regressor function, but assume that the densely observed data is a realization of a function. So the first step in any functional data study is a smoothing step which involves constructing regressor functions from observed data.

Smoothing methods given in Ramsay (2006) typically use the following model to fit a single curve: The given curve X_{ij} is observed at l discrete points $(t_1, ..., t_l)$. Let $x_{ijk}, k = 1, ..., l$, denote these observed values. We then recover the function X_{ij} from these observed values by fitting the linear model $x_{ijk} = \sum_{q=1}^{r} c_q \phi_q(t_k)$, $k = 1, ..., l$, where ϕ_q's are basis functions. We choose a large value for r and use penalization techniques to ensure that the fitted function is not very rough. We penalize the integral of the squared second derivative, i.e., we choose $c_q, 1 \leq q \leq r$ to minimize $\sum_{k=1}^{l}(x_{ijk} - \sum_{q=1}^{r} c_q \phi_q(t_k))^2 - \lambda PEN_2(X_{ij})$. We take $PEN_2(X_{ij}) = \int D^2 X_{ij}(s)ds$, where D^2 denotes the second derivative. The underlying function X_{ij} is approximated by $X_{ij}(t) = \sum_{q=1}^{r} \hat{c}_q \phi_q(t)$. We used cubic B-spline basis functions and the smooth.spline function available in R software to carry out the smoothing procedure. We specified the knots to be at $(t_1, ..., t_l)$ leading to $r = l + 4 - 2$.

After obtaining the underlying function we simulated a discrete and continuous response using the sequencing data. Let X denote the sequencing data matrix of dimension $\sum_{i=1}^{n} m_i \times l$, where

n, m_i, and l represent the number of clusters, cluster size, and the number of variants, respectively. For the selected region from the 1000 Genome Project, $l = 800, n = 1092$ while the family size m_i is one for all the families. Then the responses are generated using the relations $Y_i = \sum_{k=1}^{l} \widetilde{\beta}_k X_{ik} + e_i, e_i \sim N_m(\bar{0}, R_i)$, $\widetilde{\beta}_k = k^{-1}\delta$, $k = 1, \cdots, l$, where now R_i is the $m_i \times m_i$ correlation matrix with all off diagonal terms equal to $\gamma = 0$. For the binary case we proceeded as before using the rmvbin function in R. The following empirical powers are based on 1000 replicates and level of significance of 0.05. We chose p_n by using the five fold cross validation procedure (see the end of this section). We can see in Table 3.3 shows that both the tests have comparable power but Type-I error for gSKAT for the binary case is inflated.

Table 3.3: Empirical power comparison with the gSKAT

Effect(δ)	Continuous Trait		Binary Trait	
	F Test	gSKAT	F Test	gSKAT
0.00	0.041	0.048	0.047	0.061
0.50	0.102	0.108	0.072	0.063
1.00	0.314	0.374	0.114	0.133
3.00	0.996	0.994	0.555	0.608
5.00	1.00	1.00	0.951	0.962

In the simulation study to investigate the effect of sample size and cluster size on the power of our test, the regression function was generated using the Fourier basis functions and the response variable was generated using the linear and logistic model as in the first simulation study. The choice of the number of basis functions p_n selected to generate the regression functions is again set to $10n^{1/7}$. Effect size δ used in this study is set equal to 0.1 and 0.05 for binary response and continuous response, respectively. The correlation for the sample size and the cluster study are set to 0.8. From Table 3.4 we can see that the power increases with the sample size for both types of the response variables.

Table 3.4: Empirical power as a function of sample size

Sample Size	Continuous Trait	Binary Trait
500	0.718	0.295
1000	0.970	0.65
2000	1	0.946

To demonstrate the effect of using large cluster size m, we used a sample size of 100 in the study. Results of this simulation can be seen in Table 3.5. We observe that the Type I error is inflated with the increase in the cluster size. For a cluster of size m the number of parameters in the correlation matrix is of order m^2. Large m will affect the consistency of correlation estimate ($\hat{R}$) used in the score equation (3.3.1) rendering our asymptotic results invalid .

Table 3.5: Empirical level as a function of cluster size

Cluster Size	Continuous Trait	Binary Trait
3	0.203	0.05
10	0.659	0.14
20	0.867	0.4

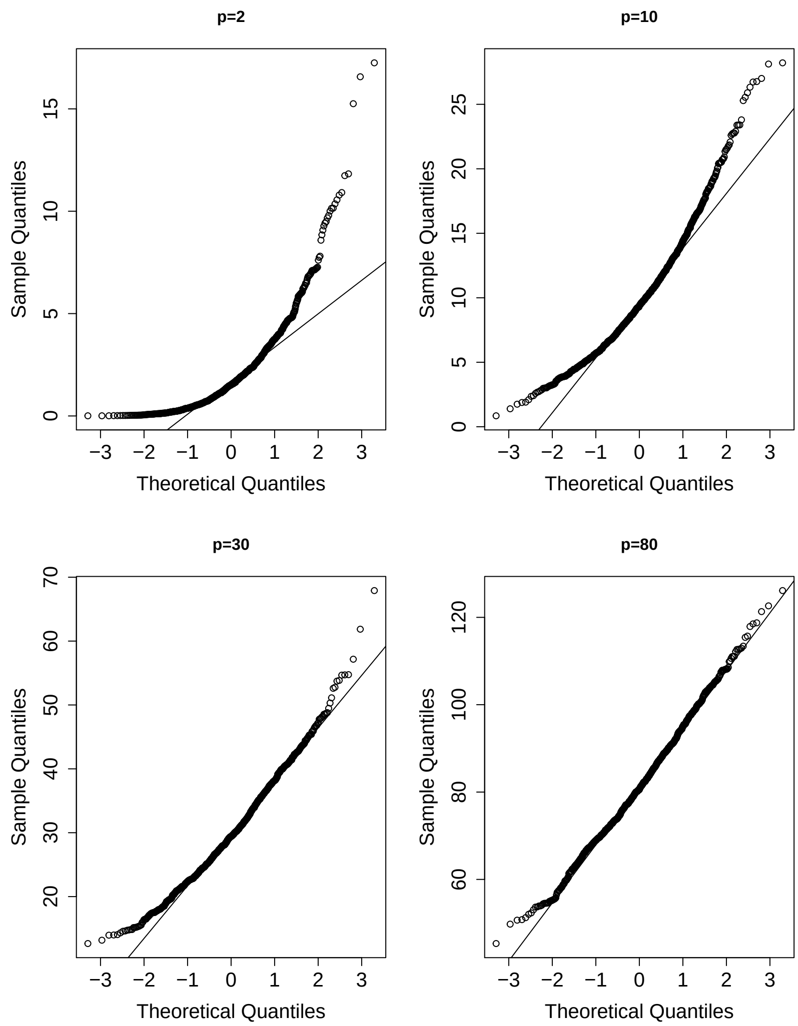
Table 3.6 below demonstrates the effect of increasing the number of parameters p_n in the model. The regressor function and the continuous response variable were simulated as before. The number of basis functions used to generate the regressor was set to $p_n = 30, 50, 90$. Sample size was 500, the effect size δ was 0.05, correlation parameter γ was taken to be 0.8. The number of parameters in the model p_n was chosen to be $30, 50, 90$ instead of using cross validation. We can see that as the number of parameters increases the Type-I error also increases.

Table 3.6: Effect of increasing p_n

p_n	Empirical level
30	0.068
50	0.065
90	0.107

The following figure displays Q-Q plots to check the normality of the test statistic. For small dimensions the test statistic follows a chi-square distribution with p degrees of freedom. As we increase p the normal approximation becomes more appropriate.

Figure 3.1: Normal Q-Q plots



An important problem that we need to address while applying our method is the determination of the number of parameters p_n. In all of the simulations reported here, we used the functional principal component analysis method for dimension reduction. We projected the infinite dimensional function on to a finite p_n dimensional subspace spanned by the first p_n eigenfunctions of the covariance operator of the regressor functions. This covariance operator is estimated from the regressor functions obtained by smoothing of the observed data. We first selected those values of $\tilde{p}$ for which the proportion of variance explained by the first $\tilde{p}$ principal components $\sum_{k=1}^{\tilde{p}} \lambda_k / \sum_{k=1}^{\infty} \lambda_k$ varied from 80% to 99% as the potential values for the number of parameters p_n. We then used 5 fold cross validation to select a final value for p_n.

3.6 Application to the sequencing data from the Minnesota Twin Study

Substance use disorders and related health conditions pose a remarkable burden on global health. Twin and family-based studies have suggested a substantial genetic contribution to substance dependence (e.g., nicotine and alcohol dependence). Studies have been successful in identifying genetic variants contributing to nicotine dependence. In this application, we applied the proposed method to assess the dependence between 15 neuronal nicotinic acetylcholine receptors (nAChRs) subunit genes and nicotine dependence using the sequencing data from Minnesota Twin Study (Vrieze et al., 2014). Genetic associations between nAChRs subunit genes and nicotine dependence has already been established. A comprehensive study on the these genes (Saccone et al., 2009) found associations for loci in the *CHRNA5*, *CHRNA3*,*CHRNA4*, *CHRNB4*,*CHRNB3*, *CHRNB1*, *CHRNA6*, *CHRND*, *CHRNG*, *CHRNB4* with nicotine dependence. Our aim is to confirm whether our test can replicate these associations. The sequencing data set we used for this analysis has 662 families and 1445 individuals.

Nicotine dependence is a continuous variable measured based on the protocols of the Substance Abuse Module of Composite International Diagnostic Interview (Hicks et al., 2011). It considers the frequency and quantity of nicotine use including cigarettes, cigars, pipes and chewing tobacco. Covariates of age and sex were also considered. We first fit a regression model with nicotine dependence as response and age and sex as predictors, and then used the residuals in the analysis.

We apply our proposed test to each of the 15 genes individually. As the response and regressors are centered, we omit the intercept from this analysis. The smooth functions obtained by the smoothing method discussed earlier in the simulation section serve as regressors in the test. Table 3 report the number of parameters i.e., p_n used for each gene. Note that as these values are small, normal approximation does not work as can be seem from Figure 3.1 . We instead use D_n^* as the statistic that follows χ_{p_n} and use this to report the un-adjusted p-values for the F-test. From this table, we find that the p-values for F-test are smaller than those of the gSKAT for many genes. Also, our method is able to detect several associations that are undetected by gSKAT, which might suggest that our method has better performance. Note that even after adjustment for multiple testing our method is able to find associations for genes, such as *CHRNB2*, *CHRNA6* and *CHRND*.

Table 3.7: p-values for Minnesota Twin Study

Gene	F Test	gSKAT	p_n	D_n^*
CHRNA1	0.929	0.621	2	0.147
CHRNA2	0.740	0.595	2	0.600
CHRNA3	0.057	0.611	2	5.713
CHRNA4	0.011	0.125	2	8.900
CHRNA5	0.148	0.415	2	3.810
CHRNA6	4E-05	0.216	6	29.765
CHRNA7	0.625	0.736	2	0.937
CHRNA9	0.400	0.524	2	1.831
CHRNB1	0.017	0.870	2	8.047
CHRNB2	2E-09	0.1351	7	53.838
CHRNB3	0.003	0.429	5	17.524
CHRNB4	0.382	0.561	2	1.922
CHRND	8E-04	0.175	2	14.230
CHRNE	0.207	0.675	2	3.142
CHRNG	0.044	0.270	2	6.219

3.7 Discussion

The proposed method can be easily extended to include additional finite dimensional covariates $Z = (Z_1, ..., Z_k)$ as well as finitely many functional regressors $X_1, ..., X_l$. The main question is to choose the dimension $p_{1n}, ..., p_{ln}$ for each of the functional regresors. We can use cross validation techniques to determine $p_{1n}, ..., p_{ln}$. Note that this is more computationally challenging than the case of a single functional regressor. The dimension of the truncated model (3.2.3) is $p_n = \sum_{j=1}^{l} p_{jn} + k + 1$.

3.8 Results

This section contains the consistency and existence results. Note that C denotes a constant in the details below.

Proposition 3.8.1 *The inverse of estimated correlation matrix $\hat{R}^{-1}$ converges to the inverse of the true correlation in probability i.e. $\|\hat{R}^{-1} - R_0^{-1}\| = O_p\{(p_n/n)^{1/2}\}$.*

Proof Let $R^* = n^{-1} \sum_{i=1}^{n} A_i^{-1/2}(\tilde{\beta})\{Y_i - g_i(\tilde{\beta})\}\{Y_i - g_i(\tilde{\beta})\}^T A_i^{-1/2}(\tilde{\beta})$. Central limit theorem gives us $\|R^* - R_0\| = O_p(n^{-1/2})$. We will show that $\|\hat{R} - R^*\| = O_p\{(p_n/n)^{1/2}\}$. Combining these two results will yield the proposition. Recall that $e_{ij}(\beta) = y_{ij} - g(\varepsilon_{ij}^T\beta)$, $\sigma_{ij}(\beta) = \sigma(\varepsilon_{ij}^T\beta)$. We have

$$\|\hat{R} - R^*\|^2 = \sum_{j_i, j_2 = 1}^{m} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{e_{ij_1}(\check{\beta}) e_{ij_2}(\check{\beta})}{\sigma_{ij_1}(\check{\beta})\sigma_{ij_2}(\check{\beta})} - \frac{e_{ij_1}(\tilde{\beta}) e_{ij_2}(\tilde{\beta})}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_2}(\tilde{\beta})} \right\}^2$$
$$\leq 2 \sum_{j_i, j_2 = 1}^{m} I_{j_1 j_2, 1}^2 + 2 \sum_{j_i, j_2 = 1}^{m} I_{j_1 j_2, 2}^2,$$

where

$$I_{j_1j_2,1} = \frac{1}{n}\sum_{i=1}^{n} \frac{e_{ij_1}(\check{\beta})e_{ij_2}(\check{\beta}) - e_{ij_1}(\tilde{\beta})e_{ij_2}(\tilde{\beta})}{\sigma_{ij_1}(\check{\beta})\sigma_{ij_2}(\check{\beta})},$$
$$I_{j_1j_2,2} = \frac{1}{n}\sum_{i=1}^{n} e_{ij_1}(\tilde{\beta})e_{ij_2}(\tilde{\beta})\left\{\frac{1}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_2}(\tilde{\beta})} - \frac{1}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_2}(\tilde{\beta})}\right\}.$$

We first show that $\sum_{j_i,j_2=1}^{m} I^2_{j_1j_2,1} = O_p(p_n/n)$. Using Assumption 3.4.2

$$\begin{aligned}
I^2_{j_1j_2,1} &\le C\left[\frac{1}{n}\sum_{i=1}^{n} e_{ij_1}(\check{\beta})\{e_{ij_2}(\check{\beta}) - e_{ij_2}(\tilde{\beta})\}\right]^2 + C\left(\frac{1}{n}\sum_{i=1}^{n} e_{ij_2}(\tilde{\beta})\{e_{ij_1}(\check{\beta}) - e_{ij_1}(\tilde{\beta})\}\right)^2 \\
&\le C\left[\frac{1}{n}\sum_{i=1}^{n} e_{ij_1}(\tilde{\beta})\{g_{ij_2}(\tilde{\beta}) - g_{ij_2}(\check{\beta})\}\right]^2 \\
&\quad + C\left[\frac{1}{n}\sum_{i=1}^{n}\{g_{ij_1}(\tilde{\beta}) - g_{ij_1}(\check{\beta})\}\{g_{ij_2}(\tilde{\beta}) - g_{ij_2}(\check{\beta})\}\right]^2 \\
&\quad + C\left[\frac{1}{n}\sum_{i=1}^{n} e_{ij_2}(\tilde{\beta})\{g_{ij_1}(\tilde{\beta}) - g_{ij_1}(\check{\beta})\}\right]^2.
\end{aligned}$$

Using Cauchy Schwarz inequality,

$$\begin{aligned}
\sum_{j_i,j_2=1}^{m} I^2_{j_1j_2,1} &\le C\sum_{j_i,j_2=1}^{m}\left\{\frac{1}{n}\sum_{i=1}^{n} e^2_{ij_1}(\tilde{\beta})\right\}\left[\sum_{i=1}^{n}\frac{1}{n}\{g_{ij_2}(\tilde{\beta}) - g_{ij_2}(\check{\beta})\}^2\right] \\
&\quad + C\sum_{j_i,j_2=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\{g_{ij_1}(\tilde{\beta}) - g_{ij_1}(\check{\beta})\}^2\right]\left[\sum_{i=1}^{n}\frac{1}{n}\{g_{ij_2}(\tilde{\beta}) - g_{ij_2}(\hat{\beta})\}^2\right] \\
&\quad + C\sum_{j_i,j_2=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\{g_{ij_1}(\tilde{\beta}) - g_{ij_1}(\check{\beta})\}^2\right]\left\{\sum_{i=1}^{n}\frac{1}{n}e^2_{ij_2}(\tilde{\beta})\right\} \\
&= I_{j_1j_2,11} + I_{j_1j_2,12} + I_{j_1j_2,13}.
\end{aligned}$$

Now,

$$I_{j_1j_2,11} = C\sum_{j_i,j_2=1}^{m}\left\{\frac{1}{n}\sum_{i=1}^{n}e_{ij_1}^2(\tilde{\beta})\right\}\left[\sum_{i=1}^{n}\frac{1}{n}\{g_{ij_2}(\tilde{\beta})-g_{ij_2}(\check{\beta})\}^2\right]$$
$$= C\left\{\sum_{j_i=1}^{m}\frac{1}{n}\sum_{i=1}^{n}e_{ij_1}^2(\tilde{\beta})\right\}\left[\sum_{j_2=1}^{m}\sum_{i=1}^{n}\frac{1}{n}\{g_{ij_2}(\tilde{\beta})-g_{ij_2}(\check{\beta})\}^2\right]$$

Taylors expansion gives us $\{g_{ij}(\check{\beta})-g_{ij}(\tilde{\beta})\}^2 \leq C(\check{\beta}-\tilde{\beta})^T\varepsilon_{ij}\varepsilon_{ij}^T(\check{\beta}-\tilde{\beta})$, yielding

$$\begin{aligned}
\sum_{j_2=1}^{m}\sum_{i=1}^{n}\frac{1}{n}\{g_{ij_2}(\tilde{\beta})-g_{ij_2}(\check{\beta})\}^2 &\leq C\frac{1}{n}\sum_{i=1}^{n}\sum_{j_2=1}^{m}(\check{\beta}-\tilde{\beta})^T\varepsilon_{ij_2}\varepsilon_{ij_2}^T(\check{\beta}-\tilde{\beta})\\
&\leq C(\check{\beta}-\tilde{\beta})^T\frac{1}{n}\sum_{i=1}^{n}F_i^TF_i(\check{\beta}-\tilde{\beta})\\
&\leq C(\check{\beta}-\tilde{\beta})^T\frac{1}{n}\left\{\sum_{i=1}^{n}F_i^TF_i-E(F_i^TF_i)+E(F_i^TF_i)\right\}(\check{\beta}-\tilde{\beta})\\
&\leq C(\check{\beta}-\tilde{\beta})^TE(F_i^TF_i)(\check{\beta}-\tilde{\beta})\\
&\leq \|\check{\beta}-\tilde{\beta}\|^2\lambda_{max}E(F_i^TF_i).\\
&= O_p(p_n/n)
\end{aligned}$$

We get the above using Assumption 3.4.4, 3.4.5. We have $n^{-1}\sum_{i=1}^{n}e_{ij_1}^2(\tilde{\beta}) = O_p(1)$. Using similar techniques for the remaining terms we can show $I_{j_1j_2,11}$, $I_{j_1j_2,12}$, $I_{j_1j_2,13}$ and thereby $\sum_{j_i,j_2=1}^{m}I_{j_1j_2,1}^2$ are all of order $O_p(p_n/n)$. The same holds for $\sum_{j_i,j_2=1}^{m}I_{j_1j_2,2}^2$ to complete the proof.

Define,

$$\bar{U}(\beta) = \sum_{i=1}^{n}F_i^TG_i(\beta)A_i^{-1/2}(\beta)\bar{R}^{-1}A_i^{-1/2}(\beta)e_i(\beta),$$
$$J(\beta) = \left[\frac{\partial U(\beta)}{\partial\beta}\right]_\beta, \quad J^*(\beta) = \left[\frac{\partial U^*(\beta)}{\partial\beta}\right]_\beta, \quad \bar{J}(\beta) = \left[\frac{\partial\bar{U}(\beta)}{\partial\beta}\right]_\beta.$$

Lemma 3.8.1 *The jacobian $J(\beta)$ can be decomposed as*

$$J(\beta) = -H(\beta) + R_1(\beta) - R_2(\beta) - R_3(\beta),$$

where

$$H(\beta) = -\sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) R_0^{-1} A_i^{-1/2}(\beta) G_i(\beta) F_i,$$
$$R_1(\beta) = \sum_{i=1}^{n} F_i^T \ddot{G}_i(\beta) A_i^{-1/2}(\beta) F_0(\beta) F_i,$$
$$R_2(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) R_0^{-1} \dot{A}_i^{1/2}(\beta) A_i^{-1}(\beta) \mathit{diag}\{e_i(\beta)\} F_i,$$
$$R_3(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1}(\beta) \dot{A}_i^{1/2}(\beta) F_0(\beta) F_i.$$

$\ddot{G}_i(\beta) = \mathit{diag}\{g''_{i1}(\beta), ..., g''_{im}(\beta)\}, \dot{A}_i = \mathit{diag}\{\sigma'_{i1}(\beta), ..., \sigma'_{im}(\beta)\}, F_0(\beta) = \mathit{diag}(R_0^{-1} A_i^{-1/2} e_i).$
Similar result holds for $\bar{J}(\beta)$ and $J^*(\beta)$ i.e.

$$\bar{J}(\beta) = -\bar{H}(\beta) + \bar{R}_1(\beta) - \bar{R}_2(\beta) - \bar{R}_3(\beta),$$

$$J^*(\beta) = -H^*(\beta) + R_1^*(\beta) - R_2^*(\beta) - R_3^*(\beta),$$

where
$\bar{H}(\beta) = -\sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) \bar{R}^{-1} A_i^{-1/2}(\beta) G_i(\beta) F_i$ *and*
$H^*(\beta) = -\sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) \hat{R}^{-1} A_i^{-1/2}(\beta) G_i(\beta) F_i$. *Remaining terms $\bar{R}_i(\beta), R_i^*(\beta), (i = 1, 2, 3)$ are defined in a similar fashion by using $\bar{R}$ and R^* respectively.*

Proof As the proof is simple but tedious we omit the details.

Lemma 3.8.2 *For all* $\Delta > 0$, $b_n \in R^{p_n}$,

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T(J^*(\beta) - \bar{J}(\beta))b_n| = O_p\{(np_n)^{1/2}\}$$

Proof Due to Lemma 3.8.1 it is sufficient to prove the following results:

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T(H^*(\beta) - \bar{H}(\beta))b_n| = O_p\{(np_n)^{1/2}\} \tag{3.8.1}$$

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T(R_i^*(\beta) - \bar{R}_i(\beta))b_n| = O_p\{(np_n)^{1/2}\}, \quad (i = 1, 2, 3). \tag{3.8.2}$$

We have

$$\begin{aligned}
|b_n^T\{H^*(\beta) - \bar{H}(\beta)\}b_n| &\leq \sum_{i=1}^{n} |b_n^T F_i^T G_i(\beta) A_i^{-1/2}(\beta)(\hat{R}^{-1} - \bar{R}^{-1})A_i^{-1/2}(\beta)G_i(\beta)F_i b_n| \\
&\leq C\|\hat{R}^{-1} - \bar{R}^{-1}\| \sum_{i=1}^{n} \|b_n^T F_i^T\|^2 \\
&= C\|\hat{R}^{-1} - \bar{R}^{-1}\| \sum_{i=1}^{n} b_n^T F_i^T F_i b_n \\
&\leq C O_p\{(p_n/n)^{1/2}\} n \left\{ \sum_{i=1}^{n} \frac{b_n^T F_i^T F_i b_n}{n} - b_n^T E(F_1^T F_1) b_n \right\} \\
&\quad + C O_p\{(p_n/n)^{1/2}\} n b_n^T E(F_1^T F_1) b_n \\
&\leq C O_p\{(p_n/n)^{1/2}\} n \lambda_{max}\{E(F_1^T F_1)\} \\
&\leq C O_p\{(p_n/n)^{1/2}\}.
\end{aligned}$$

Similarly we can prove (3.8.2).

Lemma 3.8.3 *We have*

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T\{\bar{J}(\beta)+\bar{H}(\beta)\}b_n| = O_p(n^{1/2}p_n).$$

Proof From Lemma 3.8.1 it is sufficient to show that

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T\bar{R}_i(\beta)b_n| = O_p(n^{1/2}p_n), \quad (i=1,2,3).$$

We first consider

$$\begin{aligned}
\bar{R}_2(\beta) &= \sum_{i=1}^n F_i^T\{G_i(\beta)-G_i(\tilde{\beta})\}A_i^{-1/2}(\beta)\bar{R}^{-1}\dot{A}_i^{1/2}(\beta)A_i^{-1}(\beta)\text{diag}\{e_i(\beta)\}F_i \\
&\quad + \sum_{i=1}^n F_i^T G_i(\tilde{\beta})\{A_i^{-1/2}(\beta)-A_i^{-1/2}(\tilde{\beta})\}\bar{R}^{-1}\dot{A}_i^{1/2}(\beta)A_i^{-1}(\beta)\text{diag}\{e_i(\beta)\}F_i \\
&\quad + \sum_{i=1}^n F_i^T G_i(\tilde{\beta})A_i^{-1/2}(\tilde{\beta})\bar{R}^{-1}\{\dot{A}_i^{1/2}(\beta)-\dot{A}_i^{1/2}(\tilde{\beta})\}A_i^{-1}(\beta)\text{diag}\{e_i(\beta)\}F_i \\
&\quad + \sum_{i=1}^n F_i^T G_i(\tilde{\beta})A_i^{-1/2}(\tilde{\beta})\bar{R}^{-1}\dot{A}_i^{1/2}(\tilde{\beta})\{A_i^{-1}(\beta)-A_i^{-1}(\tilde{\beta})\}\text{diag}\{e_i(\beta)\}F_i \\
&\quad + \sum_{i=1}^n F_i^T G_i(\tilde{\beta})A_i^{-1/2}(\tilde{\beta})\bar{R}^{-1}\dot{A}_i^{1/2}(\tilde{\beta})A_i^{-1}(\tilde{\beta})[\text{diag}\{e_i(\beta)\}-\text{diag}\{e_i(\tilde{\beta})\}]F_i \\
&\quad + \sum_{i=1}^n F_i^T G_i(\tilde{\beta})A_i^{-1/2}(\tilde{\beta})\bar{R}^{-1}\dot{A}_i^{1/2}(\tilde{\beta})A_i^{-1}(\tilde{\beta})\text{diag}\{e_i(\tilde{\beta})\}F_i \\
&= \sum_{i'=1}^6 \bar{R}_{2,i'}.
\end{aligned}$$

We investigate the first and last terms of the above equation and leave the rest to the reader. For the first term using Assumption 3.4.5, 3.4.2 we get

$$
\begin{aligned}
&\sup_{\|\beta-\tilde{\beta}\|\le\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} \sum_{i=1}^{n} |b_n^T \bar{R}_{2,1} b_n| \\
&\le \sup_{\|\beta-\tilde{\beta}\|\le\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} \sum_{i=1}^{n} |b_n^T F_i^T \{G_i(\beta) - G_i(\tilde{\beta})\} A_i^{-1/2}(\beta) \bar{R}^{-1} \dot{A}_i^{1/2}(\beta) A_i^{-1}(\beta) \\
&\quad \times \text{diag}\{e_i(\beta)\} F_i b_n| \\
&\le \sup_{\|\beta-\tilde{\beta}\|\le\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} C \sup_{i,j}\{g'_{ij}(\beta) - g'_{ij}(\hat{\beta})\} \sum_{i=1}^{n} \|b_n^T F_i\|^2 \\
&\le \sup_{\|\beta-\tilde{\beta}\|\le\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} C\|\beta - \hat{\beta}\| \sum_{i=1}^{n} b_n^T F_i^T F_i b_n \\
&\le O_p((np_n)^{1/2}).
\end{aligned}
$$

In a similar fashion we can show that

$$
\sup_{\|\beta-\tilde{\beta}\|\le\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} \sum_{i=1}^{n} |b_n^T \bar{R}_{2,i'} b_n| = O_p((np_n)^{1/2}), \quad (i' = 2, ..., 5).
$$

We now investigate the last term $\bar{R}_{i,6}$. Using Assumption 3.4.2 we have

$$
\|\bar{R}_{2,6}\|^2 \le C \sum_{k,l=0}^{p_n} \sum_{i_1,i_1'=1}^{n} \sum_{j_1,j_2,j_1',j_2'=1}^{m} \varepsilon_{i_1 j_2}^{(k)} \varepsilon_{i_1 j_1}^{(l)} \varepsilon_{i_1' j_2'}^{(k)} \varepsilon_{i_1' j_1'}^{(l)} e_{i_1 j_1}(\tilde{\beta}) e_{i_1' j_1'}(\tilde{\beta}).
$$

Taking expectation and using the independence of errors between clusters we get

$$
\|\bar{R}_{2,6}\|^2 = O_p(np_n^2). \tag{3.8.3}
$$

Thus we get,

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T(\bar{R}_2(\beta))b_n| = O_p(n^{1/2}p_n).$$

We can obtain the result for $\bar{R}_1$ and $\bar{R}_3$ in a similar manner to complete the proof.

Lemma 3.8.4 *We have*

$$\sup_{\|\beta-\tilde{\beta}\|\leq\Delta(p_n/n)^{1/2}} \sup_{\|b_n\|=1} |b_n^T(\bar{H}(\beta) - \bar{H}(\tilde{\beta})|b_n = O_p(n^{1/2}p_n).$$

The proof of this lemma is similar to the proofs of Lemma 3.8.2, 3.8.3 so we omit the details.

Lemma 3.8.5 *We have*

$$(\beta - \tilde{\beta})^T \bar{H}(\tilde{\beta})(\beta - \tilde{\beta}) \leq -Cn\|\beta - \tilde{\beta}\|^2.$$

Proof Using Assumptions 3.4.2, 3.4.4, 3.4.5 ,

$$\begin{aligned}(\beta - \tilde{\beta})^T \bar{H}(\tilde{\beta})(\beta - \tilde{\beta}) &= -\sum_{i=1}^{n}(\beta - \tilde{\beta})^T F_i^T G_i A_i^{-1/2} \bar{R}^{-1} A_i^{-1/2} G_i F_i(\beta - \tilde{\beta}) \\ &\leq -\lambda_{min}(\bar{R}) \min_i \lambda_{min}(A_i^{-1})\lambda_{min}(G_i^2) \sum_{i=1}^{n} F_i^T F_i \|\beta - \tilde{\beta}\|^2 \\ &\leq -Cn\|\beta - \tilde{\beta}\|^2.\end{aligned}$$

Lemma 3.8.6 *Let* $\bar{U}(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2} \bar{R}^{-1} A_i^{-1/2}(\beta)^{-1} e_i(\beta)$. *Then we can show that* $\|\bar{U}(\tilde{\beta})\| = O_p\{(np_n)^{1/2}\}$.

Proof Let $\bar{R}^{-1} = (\bar{l}_{ij})_{1 \le i,j \le m}$. Using Assumption 3.4.2,

$$\|\bar{U}(\tilde{\beta})\|^2 = \sum_{k=0}^{p_n} \left(\sum_{i=1}^{n} \sum_{j_1,j_2=1}^{m} \frac{g'_{ij_2}(\tilde{\beta}) \bar{l}_{j_2 j_1} \varepsilon_{ij_2}^{(k)} e_{ij_1}(\tilde{\beta})}{\sigma_{ij_1}(\tilde{\beta}) \sigma_{ij_2}(\tilde{\beta})} \right)^2$$

$$E(\|\bar{U}(\tilde{\beta})\|^2) \le CO(np_n).$$

Thus, the result is proved.

Lemma 3.8.7 *We have* $\|U^*(\tilde{\beta}) - \bar{U}(\tilde{\beta})\| = O_p(p_n)$.

Proof Let $Q = \{q_{j_1 j_2}\}_{1 \le j_1, j_2, m} = \hat{R}^{-1} - \bar{R}^{-1}$. We have,

$$U^*(\tilde{\beta}) - U(\tilde{\beta}) = \sum_{j_1 j_2 = 1}^{m} q_{j_1 j_2} \sum_{i=1}^{n} \frac{g'(\tilde{\eta}_{ij_2}) e(\tilde{\eta}_{ij_2})}{\sigma(\tilde{\eta}_{ij_2}) \sigma(\tilde{\eta}_{ij_1})} \varepsilon_{ij_1}.$$

Note that,

$$E \Big\| \sum_{i=1}^{n} \frac{g'(\tilde{\eta}_{ij_2}) e(\tilde{\eta}_{ij_2})}{\sigma(\tilde{\eta}_{ij_2}) \sigma(\tilde{\eta}_{ij_1})} \varepsilon_{ij_1} \Big\|^2 = O(np_n).$$

This implies, $\Big\| \sum_{i=1}^{n} \frac{g'(\tilde{\eta}_{ij_2}) e(\tilde{\eta}_{ij_2})}{\sigma(\tilde{\eta}_{ij_2}) \sigma(\tilde{\eta}_{ij_1})} \varepsilon_{ij_1} \Big\| = O_p\{(np_n)^{1/2}\}, (1 \le j_1, j_2 \le m)$. Assumption 3.4.4 yields the result.

We are now ready to prove consistency result.

Proof of Theorem 1

According to Theorem 6.3.4 from Ortega and Rheinboldt (2000), to prove the existence and consistency of the estimator $\hat{\beta}$, it is enough to verify the following condition: for all $\epsilon > 0$, there exists a constant $\Delta > 0$ such that, for all n sufficiently large,

$$\mathrm{P} \left(\sup_{\|\beta - \tilde{\beta}\| = \Delta (p_n/n)^{1/2}} (\beta - \tilde{\beta})^T U^*(\beta) < 0 \right) \ge 1 - \epsilon.$$

This same technique has been used previously by Wang et al. (2013). Using Taylors expansion we get,

$$(\beta-\tilde{\beta})^T U^*(\beta) = (\beta-\tilde{\beta})^T U^*(\tilde{\beta}) - (\beta-\tilde{\beta})^T\{-J^*(\beta^*)\}(\beta-\tilde{\beta}) \tag{3.8.4}$$
$$= A_1 + A_2,$$

where, β^* lies between β and $\tilde{\beta}$ i.e. $max\{\|\beta^*-\beta\|, \|\beta^*-\tilde{\beta}\|\} \leq \|\beta-\tilde{\beta}\|$. Also, $\|\beta-\tilde{\beta}\| = \Delta(p_n/n)^{1/2}$. By Lemma 3.8.6, 3.8.7,

$$A_1 = (\beta-\tilde{\beta})^T \bar{U}(\tilde{\beta}) + (\beta-\tilde{\beta})^T (U^*(\tilde{\beta}) - \bar{U}(\tilde{\beta}))$$
$$|A_1| \leq \Delta(p_n/n)^{1/2}(np_n)^{1/2} + \Delta(p_n/n)^{1/2}p_n$$
$$\leq \Delta O_p(p_n) + \Delta o_p(p_n). \tag{3.8.5}$$

Next,

$$A_2 = -(\beta-\tilde{\beta})^T J^*(\beta^*)(\beta-\tilde{\beta})$$
$$= -(\beta-\tilde{\beta})^T \bar{J}(\beta^*)(\beta-\tilde{\beta}) + (\beta-\tilde{\beta})^T\{-J^*(\beta^*) + \bar{J}(\beta^*)\}(\beta-\tilde{\beta})$$
$$= A_{21} + A_{22}.$$

From Lemma 3.8.2 we obtain $A_{22} = \Delta^2 o_p(p_n)$. Also

$$\begin{aligned}
A_{21} &= -(\beta - \tilde{\beta})^T \{\bar{J}(\beta^*) + \bar{H}(\beta^*)\}(\beta - \tilde{\beta}) \\
&\quad (\beta - \tilde{\beta})^T \{\bar{H}(\beta^*) - \bar{H}(\tilde{\beta})\}(\beta - \tilde{\beta}) \\
&\quad (\beta - \tilde{\beta})^T \bar{H}(\tilde{\beta})(\beta - \tilde{\beta}) \\
&\leq -C\Delta^2 p_n + o_p(p_n) + o_p(p_n).
\end{aligned} \tag{3.8.6}$$

From (3.8.4), (3.8.5), (3.8.6) we can see that $(\beta - \tilde{\beta})^T U^*(\beta) = \Delta O_p(p_n) - C\Delta^2 p_n$. Thus for a suitable Δ this term is negative in probability and we get the existence and consistency of $\hat{\beta}$.

We now turn to prove the asymptotic normality result.

Note that,

$$D(\beta)^T D(\beta) = -H(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) R_0^{-1} A_i^{-1/2}(\beta) G_i(\beta) F_i.$$

Similarly define,

$$D^*(\beta)^T D^*(\beta) = -H^*(\beta) = \sum_{i=1}^{n} F_i^T G_i(\beta) A_i^{-1/2}(\beta) \hat{R}^{-1} A_i^{-1/2}(\beta) G_i(\beta) F_i.$$

Proposition 3.8.2 *Let β^* be a fixed value of β between $\hat{\beta}$ and $\tilde{\beta}$,*

$$\|J^*(\beta) - J(\beta)\| = O_p(n^{1/2} p_n^{1.5}) = o_p(n).$$

Proof Using Lemma 3.8.1 we obtain,

$$\|J^*(\beta) - J(\beta)\| \leq \|D^*(\beta)^T D^*(\beta) - D(\beta)^T D(\beta)\| + \sum_{i=1}^{3} \|R_i^*(\beta) - R_i(\beta)\|,$$

$$= I_1 + I_2. \tag{3.8.7}$$

We now look at each of these terms, $I_1 = \Big\| \sum_{i=1}^{n} \sum_{j_1 j_2 = 1}^{m} - \frac{q_{j_1 j_2} g'_{ij_1}(\beta) g'_{ij_2}(\beta)}{\sigma_{ij_1}(\beta)\sigma_{ij_2}(\beta)} \varepsilon_{ij_1} \varepsilon_{ij_2}^T \Big\|$, where $q_{j_1 j_2}$ are the terms from $R_0^{-1} - \hat{R}^{-1}$. We use Lemma 3.8.1, Assumptions 3.4.2 and 3.4.6 to get $(I_1/n)^2 = O_p(p_n^3/n) = o_p(1)$. Similarly we can prove that $I_2 = O_p(n^{1/2} p_n^{1.5}) = o_p(n)$.

Proposition 3.8.3 *Let β^* be a fixed value of β between $\hat{\beta}$ and $\tilde{\beta}$,*

$$\|J(\beta^*) - J(\tilde{\beta})\| = O_p(n p_n^{1.5} / n^{1/4}) = o_p(n).$$

Proof

$$\|J(\beta^*) - J(\tilde{\beta})\| \leq \|D(\beta^*)^T D(\beta^*) - D(\beta)^T D(\tilde{\beta})\| + \sum_{i=1}^{n} \|R^*(\beta^*) - R(\tilde{\beta})\|$$

$$= I_1 + I_2$$

$$I_1 = \Big\| \sum_{i=1}^{n} \sum_{j_1 j_2 = 1}^{m} \left\{ \frac{g'_{ij_1}(\beta^*) g'_{ij_2}(\beta^*)}{\sigma_{ij_1}(\beta^*)\sigma_{ij_2}(\beta^*)} - \frac{g'_{ij_1}(\tilde{\beta}) g'_{ij_2}(\tilde{\beta})}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_2}(\tilde{\beta})} \right\} l_{j_1 j_2} \varepsilon_{ij_1} \varepsilon_{ij_2}^T \Big\|$$

Recall that $\tilde{\eta}_{ij} = \varepsilon_{ij}^T \tilde{\beta}$ and $\eta_{ij}^* = \varepsilon_{ij}^T \beta^*$. Using integral form of the remainder in Taylor's theorem,

Theorem 1 and Assumption 3.4.2,

$$g'_{ij}(\beta^*) - g'_{ij}(\tilde{\beta}) = \int g''\{\lambda\tilde{\eta}_{ij} - (1-\lambda)\eta^*_{ij}\}(\eta^*_{ij} - \tilde{\eta}_{ij})d\lambda$$
$$\leq c \times |\tilde{\eta}_{ij} - \eta^*_{ij}| \leq c \times \|\tilde{\beta}^T - \beta^{*T}\| \sup_{i,j} \|\varepsilon_{ij}\| \qquad (3.8.8)$$

Consider,

$$P(\sup_{i,j}\|\varepsilon_{ij}\| \geq n^{1/4}) \leq nmP(\|\varepsilon_1\| > n^{1/4})$$
$$\leq mE\|\varepsilon_1\|^4 = mE(\|\varepsilon_1\|^2)^2 = mE\left(\sum_{k=1}^{p_n} \varepsilon_{1_k}^2\right)^2$$
$$\leq mE\left(\sum_{k=1}^{\infty} \varepsilon_{1_k}^2\right)^2 = mE\|X_1\|^4 \leq \infty$$

Combining the above result with (3.8.8) we get $g'_{ij}(\beta^*) - g'_{ij}(\tilde{\beta}) = O_p(p_n^{1/2} n^{-1/4})$ This can be used to show that $I_1 = O_p(n^{3/4} p_n^{1.5}) = o_p(n)$. Similarly, we can prove the same result for I_2.

Proposition 3.8.4 *This proposition states that* $n^{1/2}(\hat{\beta} - \tilde{\beta}) \approx \left\{\frac{D(\tilde{\beta})^T D(\tilde{\beta})}{n}\right\}^{-1} \frac{U(\tilde{\beta})}{n^{1/2}}$

Proof Using Taylors expansion,

$$U^*(\hat{\beta}) = U^*(\tilde{\beta}) + J^*(\beta^*)(\hat{\beta} - \tilde{\beta})$$
$$U^*(\tilde{\beta}) = -J^*(\beta^*)(\hat{\beta} - \tilde{\beta})$$
$$\frac{U^*(\tilde{\beta}) - U(\tilde{\beta})}{n^{1/2}} + \frac{U(\tilde{\beta})}{n^{1/2}} = -\frac{J^*(\beta^*)}{n}(\hat{\beta} - \tilde{\beta})n^{1/2}$$

From Lemma 3.8.7 we obtain,

$$\frac{U(\tilde{\beta})}{n^{1/2}} = -\frac{J^*(\beta^*)}{n}(\hat{\beta} - \tilde{\beta})n^{1/2}$$

The right hand side can be written as

$$\begin{aligned} -\frac{J^*(\beta^*)}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} = \frac{-J^*(\beta^*) + J(\beta^*)}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} + \frac{-J(\beta^*) + J(\tilde{\beta})}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} \\ - \frac{J(\tilde{\beta}) + D(\tilde{\beta})^T D(\tilde{\beta})}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} + \frac{D(\tilde{\beta})^T D(\tilde{\beta})}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} \end{aligned}$$

We can show that $R_i(\tilde{\beta}) = O_p(n^{1/2}p_n), i = 1, 2, 3$ in the same manner as (3.8.3). Using Assumption 3.4.2, 3.4.6, Proposition 3.8.2, 3.8.3, Theorem 1 yields

$$-\frac{J^*(\beta^*)}{n}(\hat{\beta} - \tilde{\beta})n^{1/2} = \frac{D(\tilde{\beta})^T D(\tilde{\beta})}{n}(\hat{\beta} - \tilde{\beta})n^{1/2}$$

Thus, the Proposition is proved.

Let,

$$\begin{aligned} e'(\beta) = V(\beta)^{-1/2}e(\beta), \quad Z_n(\beta) = \left\{\frac{D(\beta)^T D(\beta)}{n}\right\}^{-1} \frac{D(\beta)^T e(\beta)'}{n^{1/2}}, \qquad (3.8.9) \\ \chi_n(\beta) = \frac{\Xi_n^{1/2}(\beta) D(\beta)^T e'(\beta)}{n^{1/2}}, \quad \Psi_n(\beta) = \Gamma(\beta)^{1/2} \left\{\frac{D(\beta)^T D(\beta)}{n}\right\}^{-1} \Gamma(\beta)^{1/2}. \end{aligned}$$

As the functions in Proposition 3.8.4 are evaluated at $\tilde{\beta}$, for ease of notation we denote $Z_n = Z_n(\tilde{\beta})$, $\chi_n = \chi_n(\tilde{\beta})$, $\Psi_n = \Psi_n(\tilde{\beta})$, $e' = e'(\tilde{\beta})$. From the statement of Theorem 2, 3.4.2, we are interested in $Z_n^T \Gamma Z_n$. We decompose it into the three terms as

$$\begin{aligned} Z_n^T \Gamma Z_n &= \chi_n^T \Psi_n^2 \chi_n \\ &= \chi_n^T \chi_n + 2\chi_n^T (\Psi_n - I_{p_n+1})\chi_n + \chi_n^T (\Psi_n - I_{p_n+1})(\Psi_n - I_{p_n+1})\chi_n \\ &= F_n + G_n + H_n, \qquad \text{say.} \end{aligned}$$

We shall show that $H_n/p_n^{1/2}$ and $G_n/p_n^{1/2}$ are asymptotically negligible. This goal is in part facilitated by the following proposition.

Proposition 3.8.5 *We have*

$$\|\Psi_n - I_{p_n+1}\|^2 = o_p(1/p_n), \tag{3.8.10}$$

$$(\chi_n^T \chi_n - (p_n + 1))/(2p_n)^{1/2} \xrightarrow{d} N(0,1). \tag{3.8.11}$$

We shall prove this lemma shortly. As a consequence of this lemma, we obtain

$$\begin{aligned} |\chi_n^T (\Psi_n - I_{p_n+1})\chi_n| &\le |\chi_n \chi_n^T| \, \|\Psi_n - I_{p_n+1}\|, \\ &= O_p(p_n) o_p(1/p_n^{1/2}) = o_p(p_n^{1/2}). \end{aligned}$$

This implies that $G_n/p_n^{1/2} \xrightarrow{p} 0$. We can similarly show that $H_n/p_n^{1/2} \xrightarrow{p} 0$. Hence

$$\frac{Z_n^T \Gamma Z_n}{(2p_n)^{1/2}} = \frac{F_n}{(2p_n)^{1/2}}. \tag{3.8.12}$$

Proof We shall first prove (3.8.10). Observe that

$$
\begin{aligned}
\|\Psi_n - I_{p_n+1}\| &\leq \|\Psi_n\| \|\Psi_n^{-1} - I_{p_n+1}\| \qquad (3.8.13)\\
\|\Psi_n\| &\leq \|I_{p_n+1}\| + \frac{\|\Psi_n^{-1} - I_{p_n+1}\|}{1 - \|\Psi_n^{-1} - I_{p_n+1}\|}.
\end{aligned}
$$

We shall show that

$$
\|\Psi_n^{-1} - I_{p_n+1}\| = o_p(p_n^{-1}) = o_p(1), \qquad (3.8.14)
$$

implying $\|\Psi_n\| \leq \|I_{p_n+1}\| = (p_n+1)^{1/2}$. This will prove (3.8.10), because

$$
\|\Psi_n - I_{p_n+1}\| = O_p(p_n^{1/2}) o_p\left(\frac{1}{p_n}\right) = o_p\left(\frac{1}{p_n^{1/2}}\right).
$$

We now turn to the proof of (3.8.14). Let, $\Xi = \Xi^{1/2}\Xi^{1/2}$, $\Xi^{1/2} = (\xi_{ks}^{(1/2)})_{0 \leq k,s \leq p_n}$. We can write

$$
\begin{aligned}
\Psi_n^{-1} &= \Xi^{1/2} \frac{1}{n} D(\tilde{\beta})^T D(\tilde{\beta}) \Xi^{1/2} \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j_1 j_2=1}^{m} \sum_{ts=1}^{p_n} \frac{l_{j_1 j_2} g'_{ij_2}(\tilde{\beta}) g'_{ij_1}(\tilde{\beta})}{\sigma_{ij_2}(\tilde{\beta}) \sigma_{ij_1}(\tilde{\beta})} \left[\xi_{ks}^{(1/2)} \varepsilon_{ij_1}^{(s)} \varepsilon_{ij_2}^{(t)} \xi_{tl}^{(1/2)}\right]_{0 \leq k,l \leq p_n}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E\|\Psi_n^{-1} - I_{p_n+1}\|^2 &= E \sum_{kl=0}^{p_n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j_1 j_2=1}^{m} \sum_{st=0}^{p_n} \frac{l_{j_1 j_2} g'_{ij_2}(\tilde{\beta}) g'_{ij_1}(\tilde{\beta})}{\sigma_{ij_2}(\tilde{\beta}) \sigma_{ij_1}(\tilde{\beta})} \xi_{ks}^{(1/2)} \varepsilon_{ij_1}^{(s)} \varepsilon_{ij_2}^{(t)} \xi_{tl}^{(1/2)} - \delta_{kl} \right\}^2 \\
&= E \sum_{kl=0}^{p_n} (a_{kl} - \delta_{kl})^2 \\
&= E \sum_{kl=0}^{p_n} (a_{kl}^2 - 2a_{kl}\delta_{kl} + \delta_{kl}^2), \quad \text{say.}
\end{aligned}
$$

We investigate each term separately. Write $\sum_{kl=0}^{p_n} a_{kl}^2 = B_1 + B_2$, where

$$B_1 = \frac{1}{n^2}\sum_{kl=0}^{p_n}\sum_{i=1}^{n}\sum_{j_1j_2j_1'j_2'=1}^{m}\sum_{s_1t_1s_2t_2=0}^{p_n}\frac{l_{j_1j_2}g'_{ij_2}(\tilde\beta)g'_{ij_1}(\tilde\beta)}{\sigma_{ij_2}(\tilde\beta)\sigma_{ij_1}(\tilde\beta)}\xi_{ks_1}^{(1/2)}\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_2}^{(t_1)}\xi_{t_1l}^{(1/2)}$$
$$\times\frac{l_{j_1'j_2'}g'_{ij_2'}(\tilde\beta)g'_{ij_1'}(\tilde\beta)}{\sigma_{ij_2'}(\tilde\beta)\sigma_{ij_1'}(\tilde\beta)}\xi_{ks_2}^{(1/2)}\varepsilon_{ij_1'}^{(s_2)}\varepsilon_{ij_2'}^{(t_2)}\xi_{t_2l}^{(1/2)},$$

$$B_2 = \frac{1}{n^2}\sum_{kl=0}^{p_n}\sum_{i_1\neq i_2=1}^{n}\sum_{j_1j_2j_1'j_2'=1}^{m}\sum_{s_1t_1s_2t_2=0}^{p_n}\frac{l_{j_1j_2}g'_{i_1j_2}(\tilde\beta)g'_{i_1j_1}(\tilde\beta)}{\sigma_{i_1j_2}(\tilde\beta)\sigma_{i_1j_1}(\tilde\beta)}\xi_{ks_1}^{(1/2)}\varepsilon_{i_1j_1}^{(s_1)}\varepsilon_{i_1j_2}^{(t_1)}\xi_{t_1l}^{(1/2)}$$
$$\times\frac{l_{j_1'j_2'}g'_{i_2j_2'}(\tilde\beta)g'_{i_2j_1'}(\tilde\beta)}{\sigma_{i_2j_2'}(\tilde\beta)\sigma_{i_2j_1'}(\tilde\beta)}\xi_{ks_2}^{(1/2)}\varepsilon_{i_2j_1'}^{(s_2)}\varepsilon_{i_2j_2'}^{(t_2)}\xi_{t_2l}^{(1/2)}.$$

Using the fact that $\Xi^{1/2}\Xi^{1/2} = \Xi$,

$$B_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j_1j_2j_1'j_2'=1}^{m}\sum_{s_1t_1s_2t_2=0}^{p_n}\frac{l_{j_1j_2}g'_{ij_2}(\tilde\beta)g'_{ij_1}(\tilde\beta)}{\sigma_{ij_2}(\tilde\beta)\sigma_{ij_1}(\tilde\beta)}\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_2}^{(t_1)}$$
$$\times\frac{l_{j_1'j_2'}g'_{ij_2'}(\tilde\beta)g'_{ij_1'}(\tilde\beta)}{\sigma_{ij_2'}(\tilde\beta)\sigma_{ij_1'}(\tilde\beta)}\varepsilon_{ij_1'}^{(s_2)}\varepsilon_{ij_2'}^{(t_2)}\xi_{t_1t_2}\xi_{s_1s_2}$$

$$E(B_1) = \frac{1}{n}o\left(\frac{n}{p_n^2}\right)\to 0$$

$$E(B_2) = \frac{1}{n^2}\sum_{i_1\neq i_2=1}^{n}\sum_{j_1j_2j_1'j_2'=1}^{m}\sum_{s_1t_1s_2t_2=0}^{p_n}E\left\{\frac{l_{j_1j_2}g'(\tilde\eta_{i_1j_2})g'_{i_1j_1}(\tilde\beta)}{\sigma_{i_1j_2}(\tilde\beta)\sigma_{i_1j_1}(\tilde\beta)}\varepsilon_{i_1j_1}^{(s_1)}\varepsilon_{i_1j_2}^{(t_1)}\right\}$$
$$\times E\left\{\frac{l_{j_1'j_2'}g'(\tilde\eta_{i_2j_2'})g'_{i_2j_1'}(\tilde\beta)}{\sigma_{i_2j_2'}(\tilde\beta)\sigma_{i_2j_1'}(\tilde\beta)}\varepsilon_{i_2j_1'}^{(s_2)}\varepsilon_{i_2j_2'}^{(t_2)}\right\}\xi_{t_1t_2}\xi_{s_1s_2}$$
$$= \frac{1}{n^2}\sum_{i_1\neq i_2=1}^{n}\sum_{s_1t_1s_2t_2=0}^{p_n}\gamma_{s_1t_1}\gamma_{s_2t_2}\xi_{t_1t_2}\xi_{s_1s_2}.$$

We observe that $\Gamma\Xi = I$ and that each of the matrices are symmetric giving, implying

$$\sum_{t_1=0}^{p_n} \gamma_{s_1 t_1} \xi_{t_1 s_2} = 1, \qquad s_1 = s_2 \tag{3.8.15}$$

$$= 0, \qquad s_1 \neq s_2. \tag{3.8.16}$$

Thus

$$\begin{aligned} E(B_2) &= \frac{1}{n^2} \sum_{i_1 \neq i_2 = 1}^{n} \sum_{s_1 t_2 s_2 = 0}^{p_n} \gamma_{s_2 t_2} \xi_{s_1 s_2} \sum_{t_1=0}^{p_n} \gamma_{s_1 t_1} \xi_{t_1 t_2} \qquad (3.8.17) \\ &= \frac{1}{n^2} \sum_{i_1 \neq i_2 = 1}^{n} \sum_{s_2=0}^{p_n} \sum_{s_1=0}^{p_n} \gamma_{s_2 s_1} \xi_{s_1 s_2} \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n^2} \sum_{i_1 \neq i_2 = 1}^{n} \sum_{s_2=0}^{p_n} 1 = \frac{p_n + 1}{n} 2 \binom{n}{2} \\ &= \frac{(n-1)(p_n+1)}{n}. \end{aligned}$$

In a similar fashion, we can show that,

$$-2 \sum_{kl=0}^{p_n} E(a_{kl} \delta_{kl}) = \frac{-2}{n} \sum_{i=1}^{n} \sum_{st=0}^{p_n} \gamma_{st} \sum_{k=0}^{p_n} \xi_{sk}^{(1/2)} \xi_{kt}^{(1/2)} = \frac{-2}{n} \sum_{i=1}^{n} \sum_{t=0}^{p_n} \sum_{s=0}^{p_n} \gamma_{ts} \xi_{st} = -2(p_n + 1)$$

and $\sum_{kl=0}^{p_n} \delta_{kl}^2 = p_n + 1$.

These results together imply that

$$\begin{aligned} E\|\Psi_n^{-1} - I_{p_n+1}\|^2 &= o\left(\frac{1}{p_n^2}\right) + \left\{\frac{(n-1)(p_n+1)}{n}\right\} - 2(p_n+1) + (p_n+1), \qquad (3.8.18) \\ &= o\left(\frac{1}{p_n^2}\right) + O\left\{\frac{p_n+1}{n}\right\}. \end{aligned}$$

Thus, we obtain $\|\Psi_n^{-1} - I_{p_n+1}\| = o_p\left(\frac{1}{p_n}\right)$ and conclude the proof of (3.8.10) in Proposition 3.8.5 .

Next we prove (3.8.11). Referring to (3.8.9),

$$\begin{aligned}
\chi_n &= \frac{\Xi^{1/2} D^T V^{-1/2} e}{n^{1/2}} = \sum_{i=1}^{n} \frac{\Xi^{1/2} F_i^T G_i A_i^{-1/2} R^{-1} A_i^{-1/2} e_i}{n^{1/2}} \\
&= \sum_{i=1}^{n} \frac{\Xi^{1/2} F_i^T G_i A_i^{-1/2} R^{-1} \tilde{e}_i}{n^{1/2}}, \qquad \tilde{e}_i = A_i^{-1/2}(\tilde{\beta}) e_i(\tilde{\beta})
\end{aligned}$$

$$= \frac{1}{n^{1/2}} \sum_{i=1}^{n} \sum_{j_1 j_2=1}^{m} \sum_{s=0}^{p_n} \begin{bmatrix} \dfrac{\xi_{0s}^{1/2} \varepsilon_{ij_1}^{(s)} g'_{ij_1}(\tilde{\beta}) l_{j_1 j_2} \tilde{e}_{ij_2}}{\sigma_{ij_1}(\tilde{\beta})} \\ \vdots \\ \dfrac{\xi_{p_n s}^{1/2} \varepsilon_{ij_1}^{(s)} g'_{ij_1}(\tilde{\beta}) l_{j_1 j_2} \tilde{e}_{ij_2}}{\sigma_{ij_1}(\tilde{\beta})} \end{bmatrix},$$

$$\begin{aligned}
\chi_n^T \chi_n &= \frac{1}{n} \sum_{t=0}^{p_n} \left\{ \sum_{i=1}^{n} \sum_{j_1 j_2=1}^{m} \sum_{s=0}^{p_n} \frac{\xi_{ts}^{1/2} \varepsilon_{ij_1}^{(s)} g'_{ij_1}(\tilde{\beta}) l_{j_1 j_2} \tilde{e}_{ij_2}}{\sigma_{ij_1}(\tilde{\beta})} \right\}^2 \\
&= A_n + B_n, \qquad (3.8.19) \\
A_n &= \frac{1}{n} \sum_{t=0}^{p_n} \sum_{i=1}^{n} \sum_{j_1 j_2 j_3 j_4=1}^{m} \sum_{s_1 s_2=0}^{p_n} \frac{\xi_{ts_1}^{(1/2)} \xi_{ts_2}^{(1/2)} \varepsilon_{ij_1}^{(s_1)} \varepsilon_{ij_3}^{(s_2)} g'_{ij_1}(\tilde{\beta}) g'_{ij_3}(\tilde{\beta}) l_{j_1 j_2} l_{j_3 j_4} \tilde{e}_{ij_2} \tilde{e}_{ij_4}}{\sigma_{ij_1}(\tilde{\beta}) \sigma_{ij_3}(\tilde{\beta})}, \\
B_n &= \sum_{t=0}^{p_n} \sum_{i_1 \neq i_2=1}^{n} \sum_{j_1 j_2 j_3 j_4=1}^{m} \sum_{s_1 s_2=0}^{p_n} \frac{\xi_{ts_1}^{(1/2)} \xi_{ts_2}^{(1/2)} \varepsilon_{i_1 j_1}^{(s_1)} \varepsilon_{i_2 j_3}^{(s_2)} g'_{i_1 j_1}(\tilde{\beta}) g'_{i_2 j_3}(\tilde{\beta}) l_{j_1 j_2} l_{j_3 j_4} \tilde{e}_{i_1 j_2} \tilde{e}_{i_2 j_4}}{n \sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_2 j_3}(\tilde{\beta})}.
\end{aligned}$$

We shall show that $\frac{A_n - (p_n+1)}{p_n^{1/2}}$ is asymptotically negligible in Lemma 3.8.8 and that B_n has desired distribution in Lemma 3.8.9.

Lemma 3.8.8 *We have* $\frac{A_n - (p_n+1)}{p_n^{1/2}} \xrightarrow{p} 0.$

Proof Using independence of $\tilde{e}$ with all ε,

$$E(A_n) = \sum_{t=0}^{p_n}\sum_{i=1}^{n}\sum_{j_1j_2j_3j_4=1}^{m}\sum_{s_1s_2=0}^{p_n} E\frac{\xi_{ts_1}^{(1/2)}\xi_{ts_2}^{(1/2)}\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_3}^{(s_2)}g'_{ij_1}(\tilde{\beta})g'_{ij_3}(\tilde{\beta})l_{j_1j_2}l_{j_3j_4}\text{cov}(\tilde{e}_{ij_2}\tilde{e}_{ij_4})}{n\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_3}(\tilde{\beta})}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{s_1s_2=0}^{p_n}\xi_{s_1s_2}\sum_{j_1j_2j_3=1}^{m} E\left\{\frac{\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_3}^{(s_2)}g'_{ij_1}(\tilde{\beta})g'_{ij_3}(\tilde{\beta})l_{j_1j_2}}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_3}(\tilde{\beta})}\sum_{j_4=1}^{m} l_{j_3j_4}\text{cov}(\tilde{e}_{ij_2}\tilde{e}_{ij_4}).\right\}$$

We note that $l_{j_3j_4}$ are elements from the inverse of the correlation matrix (R) and that covariance matrix of $\tilde{e_i} = A_i^{-1/2}e_i$ is R . Also for any invertible matrix M, $M^{-1}M = I$. Using (3.8.16), we obtain

$$\begin{aligned} E(A_n) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{s_1s_2=0}^{p_n}\xi_{s_1s_2}\sum_{j_1j_2=j_3=1}^{m} E\left\{\frac{\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_2}^{(s_2)}g'_{ij_1}(\tilde{\beta})g'_{ij_2}(\tilde{\beta})l_{j_1j_2}}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_2}(\tilde{\beta})}\right\} \quad (3.8.20) \\ &= \sum_{s_1s_2=0}^{p_n}\xi_{s_1s_2}\gamma_{s_1s_2} = \sum_{s_1=0}^{p_n} 1 = p_n + 1. \end{aligned}$$

We now evaluate

$$\begin{aligned} E(A_n^2) &= E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j_1j_2j_3j_4=1}^{m}\sum_{s_1s_2=0}^{p_n}\frac{\xi_{s_1s_2}\varepsilon_{ij_1}^{(s_1)}\varepsilon_{ij_3}^{(s_2)}g'_{ij_1}(\tilde{\beta})g'_{ij_2}(\tilde{\beta})l_{j_1j_2}l_{j_3j_4}\tilde{e}_{ij_2}\tilde{e}_{ij_4}}{\sigma_{ij_1}(\tilde{\beta})\sigma_{ij_3}(\tilde{\beta})}\right\}^2 \\ &= A_n^a + A_n^b, \quad \text{say.} \quad (3.8.21) \end{aligned}$$

Now consider,

$$A_n^a = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j_1 j_1' \ldots j_4 j_4'=1}^{m} \sum_{s_1 s_2 s_1' s_2'=0}^{p_n} \xi_{s_1 s_2} \xi_{s_1' s_2'}$$

$$\times E \left\{ \frac{\varepsilon_{ij_1}^{s_1} \varepsilon_{ij_1'}^{(s_1')} \varepsilon_{ij_3}^{(s_2)} \varepsilon_{ij_3'}^{(s_2')} g_{ij_1}'(\tilde{\beta}) g_{ij_1'}'(\tilde{\beta}) g_{ij_3}'(\tilde{\beta}) g_{ij_3'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{ij_2} \tilde{e}_{ij_2'} \tilde{e}_{ij_4} \tilde{e}_{ij_4'}}{\sigma_{ij_1}(\tilde{\beta}) \sigma_{ij_1'}(\tilde{\beta}) \sigma_{ij_3}(\tilde{\beta}) \sigma_{ij_3'}(\tilde{\beta})} \right\}$$

$$= O\left(\frac{1}{n}\right) = o\left(\frac{1}{p_n^2}\right). \tag{3.8.22}$$

Next, we have

$$A_n^b = \frac{1}{n^2} \sum_{i_1 \neq i_2 = 1}^{n} \sum_{j_1 j_1' \ldots j_4 j_4'=1}^{m} \sum_{s_1 s_2 s_1' s_2'=0}^{p_n} \xi_{s_1 s_2} \xi_{s_1' s_2'} l_{j_1 j_2} l_{j_1' j_2'} l_{j_3 j_4} l_{j_3' j_4'}$$

$$\times E \left\{ \frac{\varepsilon_{i_1 j_1}^{(s_1)} \varepsilon_{i_2 j_1'}^{(s_1')} \varepsilon_{i_1 j_3}^{(s_2)} \varepsilon_{i_2 j_3'}^{(s_2')} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) g_{i_1 j_3}'(\tilde{\beta}) g_{i_2 j_3'}'(\tilde{\beta}) \tilde{e}_{i_1 j_2} \tilde{e}_{i_2 j_2'} \tilde{e}_{i_1 j_4} \tilde{e}_{i_2 j_4'}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_2 j_1'}(\tilde{\beta}) \sigma_{i_1 j_3}(\tilde{\beta}) \sigma_{i_2 j_3'}(\tilde{\beta})} \right\}$$

$$= \frac{1}{n^2} \sum_{i_1 \neq i_2 = 1}^{n} \sum_{s_1 s_2} \xi_{s_1 s_2} \gamma_{s_1 s_2} \sum_{s_1' s_2'} \xi_{s_1' s_2'} \gamma_{s_1' s_2'}$$

$$= \frac{(n-1)(p_n+1)^2}{n}. \tag{3.8.23}$$

Upon combining this fact with (3.8.20), (3.8.21) and (3.8.22), we obtain

$$\mathrm{var}(A_n) \;=\; E(A_n^2) - E^2(A_n) = o\left(\frac{1}{p_n^2}\right) + \frac{(p_n+1)^2}{n}.$$

Thus, we have $\frac{A_n - (p_n+1)}{p_n^{1/2}} \xrightarrow{p} 0$.

From, (3.8.19) and Lemma 3.8.8 we obtain

$$\frac{\chi^T\chi - (p_n+1)}{(2p_n)^{1/2}} = \frac{B_n}{(2p_n)^{1/2}}.$$

So the next lemma will establish (3.8.11).

Lemma 3.8.9 *Asymptotic distribution of B_n is given as*

$$\frac{B_n}{(2p_n)^{1/2}} \to N(0,1).$$

Definition 3.8.1

$$\begin{aligned} W_{ni} &= \sum_{i_1=1}^{i-1} \sum_{j_1 j_2 j_3 j_4=1}^{m} \sum_{s_1 s_2=0}^{p_n} \frac{\xi_{s_1 s_2} \varepsilon_{i_1 j_1}^{(s_1)} \varepsilon_{i j_3}^{(s_2)} g'_{i_1 j_1}(\tilde{\beta}) g'_{i j_3}(\tilde{\beta}) l_{j_1 j_2} l_{j_3 j_4} \tilde{e}_{i_1 j_2} \tilde{e}_{i j_4}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i j_3}(\tilde{\beta})}, \\ \widetilde{W}_{ni} &= \frac{2}{n(2p_n)^{1/2}} W_{ni}. \end{aligned}$$

Note that,

$$B_n = \frac{2}{n}\sum_{i=1}^{n} W_{ni}.$$

The random variables W_{ni} form a triangular array of martingale differences w.r.t. the filtration $\mathcal{F}_{n,i} = \sigma(\varepsilon_{i_1 j}^{(k)}, e_{ij}, 1 \le i_1 \le i, 1 \le j \le m, 0 \le k \le p_n), (1 \le i \le n, n \in \mathrm{N})$. N denotes the set of all natural numbers. This implies that $\widetilde{W}_{ni}$ is also a martingale difference array.

Proof

To prove this lemma it is enough to prove the following:

$$\sum_{i=1}^{n} \widetilde{W}_{ni} \xrightarrow{d} N(0,1). \tag{3.8.24}$$

According to the CLT for the sums of martingale difference arrays, see Corollary 3.1 in Hall and Heyde (1980), it suffices to verify the following two conditions

$$\sum_{i=1}^{n} E(\widetilde{W}_{ni}^{2} \mid \mathcal{F}_{n,i-1}) \xrightarrow{p} 1, \tag{3.8.25}$$

$$\sum_{i=1}^{n} E\{\widetilde{W}_{ni}^{4} \mid \mathcal{F}_{n,i-1}\} \xrightarrow{p} 0, \qquad \text{for all } \epsilon > 0. \tag{3.8.26}$$

Consider,

$$\begin{aligned}
E(W_{ni}^{2} \mid \mathcal{F}_{n,i-1}) &= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1j_2j_3j_4=1}^{m} \sum_{j_1'j_2'j_3'j_4'=1}^{m} \sum_{s_1s_2s_1's_2'=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1s_2}\xi_{s_1's_2'}\varepsilon_{i_1j_1}^{(s_1)}\varepsilon_{i_2j_1'}^{(s_1')}g'_{i_1j_1}(\tilde{\beta})g'_{i_2j_1'}(\tilde{\beta})l_{j_1j_2}l_{j_1'j_2'}l_{j_3j_4}l_{j_3'j_4'}\tilde{e}_{i_1j_2}\tilde{e}_{i_2j_2'}}{\sigma_{i_1j_1}(\tilde{\beta})\sigma_{i_2j_1'}(\tilde{\beta})} \\
&\quad \times E\left\{\frac{\varepsilon_{ij_3}^{(s_2)}\varepsilon_{ij_3'}^{(s_2')}g'_{ij_3}(\tilde{\beta})g'_{ij_3'}(\tilde{\beta})\tilde{e}_{ij_4}\tilde{e}_{ij_4'}}{\sigma_{ij_3}(\tilde{\beta})\sigma_{ij_3'}(\tilde{\beta})} \mid \mathcal{F}_{n,i-1}\right\} \\
&= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1j_2j_3j_4=1}^{m} \sum_{j_1'j_2'j_3'=1}^{m} \sum_{s_1s_2s_1's_2'=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1s_2}\xi_{s_1's_2'}\varepsilon_{i_1j_1}^{(s_1)}\varepsilon_{i_2j_1'}^{(s_1')}g'_{i_1j_1}(\tilde{\beta})g'_{i_2j_1'}(\tilde{\beta})l_{j_1j_2}l_{j_1'j_2'}l_{j_3j_4}\tilde{e}_{i_1j_2}\tilde{e}_{i_2j_2'}}{\sigma_{i_1j_1}(\tilde{\beta})\sigma_{i_2j_1'}(\tilde{\beta})} \\
&\quad \times E\left\{\frac{\varepsilon_{ij_3}^{(s_2)}\varepsilon_{ij_3'}^{(s_2')}g'_{ij_3}(\tilde{\beta})g'_{ij_3'}(\tilde{\beta})}{\sigma_{ij_3}(\tilde{\beta})\sigma_{ij_3'}(\tilde{\beta})} \sum_{j_4'=1}^{m} l_{j_3'j_4'}\text{cov}(\tilde{e}_{ij_4}\tilde{e}_{ij_4'})\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1 j_2=1}^{m} \sum_{j_1' j_2'=1}^{m} \sum_{s_1 s_2 s_1' s_2'=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1 s_2}\xi_{s_1' s_2'}\varepsilon_{i_1 j_1}^{(s_1)}\varepsilon_{i_2 j_1'}^{(s_1')} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2}\tilde{e}_{i_2 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta})\sigma_{i_2 j_1'}(\tilde{\beta})} \\
&\quad \times \sum_{j_3 j_4=1}^{m} E\left\{ \frac{l_{j_3 j_4}\varepsilon_{ij_3}^{(s_2)}\varepsilon_{ij_4}^{(s_2')} g_{ij_3}'(\tilde{\beta}) g_{ij_4}'(\tilde{\beta})}{\sigma_{ij_3}(\tilde{\beta})\sigma_{ij_4}(\tilde{\beta})} \right\},
\end{aligned}
$$

$$
\begin{aligned}
E(W_{ni}^2 \mid \mathcal{F}_{n,i-1}) &= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1 j_2=1}^{m} \sum_{j_1' j_2'=1}^{m} \sum_{s_1 s_2 s_1' s_2'=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1 s_2}\xi_{s_1' s_2'}\varepsilon_{i_1 j_1}^{s_1}\varepsilon_{i_2 j_1'}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2}\tilde{e}_{i_2 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta})\sigma_{i_2 j_1'}(\tilde{\beta})} \times \gamma_{s_2 s_2'} \\
&= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{s_1' s_2' s_1 s_2=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1' s_2'}\varepsilon_{i_1 j_1}^{(s_1)}\varepsilon_{i_2 j_1'}^{(s_1')} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2}\tilde{e}_{i_2 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta})\sigma_{i_2 j_1'}(\tilde{\beta})} \\
&\quad \times \sum_{s_2=0}^{p_n} \xi_{s_1 s_2}\gamma_{s_2 s_2'} \\
&= \sum_{i_1,i_2=1}^{i-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{s_1' s_1=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1' s_1}\varepsilon_{i_1 j_1}^{(s_1)}\varepsilon_{i_2 j_1'}^{(s_1')} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2}\tilde{e}_{i_2 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta})\sigma_{i_2 j_1'}(\tilde{\beta})} = A_n^{(1)} + A_n^{(2)},
\end{aligned}
$$

where

$$A_n^{(1)} = \sum_{i_1=i_2=1}^{i-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{s_1' s_1=0}^{p_n} \frac{\xi_{s_1' s_1} \varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_1 j_1'}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_1 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2} \tilde{e}_{i_1 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_1 j_1'}(\tilde{\beta})},$$

$$A_n^{(2)} = \sum_{i_1 \neq i_2=1}^{i-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{s_1' s_1=0}^{p_n} \frac{\xi_{s_1' s_1} \varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_2 j_1'}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_2 j_1'}'(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{i_1 j_2} \tilde{e}_{i_2 j_2'}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_2 j_1'}(\tilde{\beta})}.$$

Note that

$$\begin{aligned}
E(A_n^{(1)}) &= \sum_{i_1=1}^{i-1} \sum_{j_1 j_2 j_1'=1}^{m} \sum_{s_1' s_1=0}^{p_n} \\
&\qquad \times E\left\{ \frac{\xi_{s_1' s_1} \varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_1 j_1'}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_1 j_1'}'(\tilde{\beta}) l_{j_1 j_2}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma(\tilde{\beta}_{i_1 j_1'})} \sum_{j_2'=1}^{m} l_{j_1' j_2'} \text{cov}(\tilde{e}_{i_1 j_2} \tilde{e}_{i_1 j_2}) \right\} \\
&= \sum_{i_1=1}^{i-1} \sum_{j_1 j_2=1}^{m} \sum_{s_1' s_1=0}^{p_n} E\left\{ \frac{\xi_{s_1' s_1} \varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_1 j_2}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_1 j_2}'(\tilde{\beta}) l_{j_1 j_2}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_1 j_2}(\tilde{\beta})} \right\} \\
&= \sum_{i_1=1}^{i-1} \sum_{s_1' s_1=0}^{p_n} \xi_{s_1' s_2'} \sum_{j_1 j_2=1}^{m} E\left\{ \frac{\varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_1 j_2}^{s_1'} g_{i_1 j_1}'(\tilde{\beta}) g_{i_1 j_2}'(\tilde{\beta}) l_{j_1 j_2}}{\sigma_{i_1 j_1}(\tilde{\beta}) \sigma_{i_1 j_2}(\tilde{\beta})} \right\} \\
&= \sum_{i_1=1}^{i-1} \sum_{s_1' s_1=0}^{p_n} \xi_{s_1' s_1} \gamma_{s_1 s_1'} = (i-1)(p_n+1), \\
E(A_n^{(2)}) &= 0.
\end{aligned}$$

The above facts together imply:

$$E\{E(W_{ni}^2 \mid \mathcal{F}_{n,i-1})\} = (i-1)(p_n+1).$$

Thus

$$E\{\sum_{i=1}^{n} E(\widetilde{W}_{ni}^2 \mid \mathcal{F}_{n,i-1})\} = \frac{2}{n^2 p_n} E\{\sum_{i=1}^{n} E(W_{ni}^2 \mid \mathcal{F}_{n,i-1})\} = \frac{2}{n^2 p_n} \sum_{i=1}^{n} (i-1)(p_n+1) \to 1. \tag{3.8.27}$$

Next, we shall show that

$$\text{var}\Big\{ \sum_{i=1}^{n} E(\widetilde{W}_{ni}^2 \mid \mathcal{F}_{n,i-1}) \Big\} \to 0. \tag{3.8.28}$$

To prove this claim, first consider

$$E\{\sum_{i=1}^{n} E(\widetilde{W}_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2 = \sum_{i=1}^{n} E\{E(\widetilde{W}_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2 + 2 \sum_{1 \le i_1 < i_2 \le n} E\{E(\widetilde{W}_{ni_1}^2 \mid \mathcal{F}_{n,i_1-1}) E(\widetilde{W}_{ni_2}^2 \mid \mathcal{F}_{n,i_2-1})\}. \tag{3.8.29}$$

The second term satisfies

$$\begin{aligned}
& E(W_{ni_1}^2 \mid \mathcal{F}_{n,i_1-1}) E(W_{ni_2}^2 \mid \mathcal{F}_{n,i_2-1}) \\
&= \sum_{a_1,a_2=1}^{i_1-1} \sum_{a_3,a_4=1}^{i_2-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{j_3 j_4 j_3' j_4'=1}^{m} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n} \\
&\quad \times \frac{\xi_{s_1' s_1} \varepsilon_{a_1 j_1}^{s_1} \varepsilon_{a_2 j_1'}^{s_1'} g'_{a_1 j_1}(\tilde{\beta}) g'_{a_2 j_1'}(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{a_1 j_2} \tilde{e}_{a_2 j_2'}}{\sigma_{a_1 j_1}(\tilde{\beta}) \sigma_{a_2 j_1'}(\tilde{\beta})} \\
&\quad \times \frac{\xi_{s_2 s_2'} \varepsilon_{a_3 j_3}^{s_2} \varepsilon_{a_4 j_3'}^{s_2'} g'(\tilde{\beta}_{a_3 j_3}) g'(\tilde{\beta}_{a_4 j_3'}) l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{a31 j_4} \tilde{e}_{a_4 j_4'}}{\sigma_{a_3 j_3}(\tilde{\beta}) \sigma_{a_4 j_3'}(\tilde{\beta})} \\
&= A_n^{(1)} + B_n^{(1)} + C_n^{(1)} + D_n^{(1)} + E_n^{(1)},
\end{aligned}$$

where,

$$A_n^{(1)} = \sum_{a=a_1=\ldots=a_4=1}^{i_1-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{j_3 j_4 j_3' j_4'=1}^{m} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n}$$

$$\times \frac{\xi_{s_1 s_1'} \varepsilon_{aj_1}^{s_2} \varepsilon_{aj_1'}^{s_2'} g'_{aj_1}(\tilde{\beta}) g'_{aj_1'}(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{aj_2} \tilde{e}_{aj_2'}}{\sigma_{aj_1}(\tilde{\beta}) \sigma_{aj_1'}(\tilde{\beta})}$$

$$\times \frac{\xi_{s_2 s_2'} \varepsilon_{aj_3}^{s_2} \varepsilon_{aj_3'}^{s_2'} g'_{aj_3}(\tilde{\beta}) g'_{aj_3'}(\tilde{\beta}) l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{aj_4} \tilde{e}_{aj_4'}}{\sigma_{aj_3}(\tilde{\beta}) \sigma_{aj_3'}(\tilde{\beta})},$$

$$B_n^{(1)} = \sum_{a_1=a_2=1}^{i_1-1} \sum_{a_3=a_4=1}^{i_2-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{j_3 j_4 j_3' j_4'=1}^{m} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n}$$

$$\times \frac{\xi_{s_1' s_1} \varepsilon_{a_1 j_1}^{s_1} \varepsilon_{a_1 j_1'}^{s_1'} g'_{a_1 j_1}(\tilde{\beta}) g'_{a_1 j_1'}(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{a_1 j_2} \tilde{e}_{a_1 j_2'}}{\sigma_{a_1 j_1}(\tilde{\beta}) \sigma_{a_1 j_1'}(\tilde{\beta})}$$

$$\times \frac{\xi_{s_2 s_2'} \varepsilon_{a_3 j_3}^{s_2} \varepsilon_{a_3 j_3'}^{s_2'} g'_{a_3 j_3}(\tilde{\beta}) g'_{a_3 j_3'}(\tilde{\beta}) l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{a31 j_4} \tilde{e}_{a_3 j_4'}}{\sigma_{a_3 j_3}(\tilde{\beta}) \sigma_{a_3 j_3'}(\tilde{\beta})},$$

$$C_n^{(1)} = \sum_{a_1=a_3\neq a_2=a_4=1}^{i_1-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{j_3 j_4 j_3' j_4'=1}^{m} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n}$$

$$\times \frac{\xi_{s_1' s_1} \varepsilon_{a_1 j_1}^{s_1} \varepsilon_{a_2 j_1'}^{s_1'} g'_{a_1 j_1}(\tilde{\beta}) g'_{a_2 j_1'}(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{a_1 j_2} \tilde{e}_{a_2 j_2'}}{\sigma_{a_1 j_1}(\tilde{\beta}) \sigma_{a_2 j_1'}(\tilde{\beta})}$$

$$\times \frac{\xi_{s_2 s_2'} \varepsilon_{a_1 j_3}^{s_2} \varepsilon_{a_2 j_3'}^{s_2'} g'_{a_1 j_3}(\tilde{\beta}) g'_{a_2 j_3'}(\tilde{\beta}) l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{a_1 j_4} \tilde{e}_{a_2 j_4'}}{\sigma_{a_1 j_3}(\tilde{\beta}) \sigma_{a_2 j_3'}(\tilde{\beta})}$$

$$
= \sum_{a_1=1}^{i_1-1} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n} \sum_{j_1 j_2 j_3 j_4=1}^{m} \xi_{s_1' s_1} \xi_{s_2 s_2'}
\frac{\varepsilon_{a_1 j_1}^{s_1} \varepsilon_{a_1 j_3}^{s_2} g'_{a_1 j_1}(\tilde{\beta}) l_{j_1 j_2} l_{j_3 j_4} \tilde{e}_{a_1 j_4} \tilde{e}_{a_1 j_2}}{\sigma_{a_1 j_1}(\tilde{\beta}) \sigma_{a_1 j_3}(\tilde{\beta})}
\times \sum_{a_2=1}^{i_1-1} \sum_{j_1' j_2' j_3' j_4'=1}^{m} \frac{\varepsilon_{a_2 j_1'}^{s_1'} \varepsilon_{a_2 j_3'}^{s_2'} g'_{a_2 j_1'}(\tilde{\beta}) l_{j_1' j_2'} l_{j_3' j_4'} \tilde{e}_{a_2 j_2'} \tilde{e}_{a_2 j_4'}}{\sigma_{a_2 j_1'}(\tilde{\beta}) \sigma_{a_2 j_3'}(\tilde{\beta})}.
$$

From (3.8.29) we can see that our interest is only in $E(A_n^{(1)} + B_n^{(1)} + C_n^{(1)} + D_n^{(1)})$. So it is sufficient talk about just the expectation of these terms i.e $A_n^{(1)}, B_n^{(1)}, C_n^{(1)}, D_n^{(1)}$ and $E_n^{(1)}$. $D_n^{(1)}$ has terms with $a_1 = a_4, a_2 = a_3$ and has same expectation as $C_n^{(1)}$. $E_n^{(1)}$ has all the terms like $a_1 \neq a_2 \neq a_3 \neq a_4$ and so on with all of them the expectation is 0 and so $E(E_n) = 0$. We now examine the expectations of the remaining terms i.e $A_n^{(1)}, B_n^{(1)}$ and $C_n^{(1)}$ i.e.

$$
\begin{aligned}
E(A_n^{(1)}) &= (i_1 - 1) o\left(\frac{n}{p_n^2}\right), \\
E(B_n^{(1)}) &= \sum_{a_1=1}^{i_1-1} \sum_{j_1 j_2 j_1' j_2'=1}^{m} \sum_{s_1' s_1=0}^{p_n} \\
&\quad \times E\left\{ \frac{\xi_{s_1' s_1} \varepsilon_{a_1 j_1}^{s_1} \varepsilon_{a_1 j_1'}^{s_1'} g'_{a_1 j_1}(\tilde{\beta}) g'_{a_1 j_1'}(\tilde{\beta}) l_{j_1 j_2} l_{j_1' j_2'} \tilde{e}_{a_1 j_2} \tilde{e}_{a_1 j_2'}}{\sigma_{a_1 j_1}(\tilde{\beta}) \sigma_{a_1 j_1'}(\tilde{\beta})} \right\} \\
&\quad \times \sum_{a_3=1}^{i_2-1} \sum_{j_3 j_4 j_3' j_4'=1}^{m} \sum_{s_2' s_2=0}^{p_n} \\
&\quad \times E\left\{ \frac{\xi_{s_2' s_2} \varepsilon_{a_3 j_3}^{s_2} \varepsilon_{a_3 j_3'}^{s_2'} g'_{a_3 j_3}(\tilde{\beta}) g'_{a_3 j_3'}(\tilde{\beta}) l_{j_3 j_4} l_{j_3' j_4'} \tilde{e}_{a_3 j_4} \tilde{e}_{a_3 j_4'}}{\sigma_{a_3 j_3}(\tilde{\beta}) \sigma_{a_3 j_3'}(\tilde{\beta})} \right\} \\
&= (i_1 - 1)(p_n + 1)(i_2 - 1)(p_n + 1),
\end{aligned}
$$

$$
\begin{aligned}
E(C_n^{(1)}) &= \sum_{a_1=1}^{i_1-1} \sum_{s_1' s_1 s_2 s_2'=0}^{p_n} \gamma_{s_1 s_2} \gamma_{s_1' s_2'} \sum_{a_2=1}^{i_1-1} \xi_{s_1 s_1'} \xi_{s_2 s_2'} \\
&= (i_1 - 1)^2 (p_n + 1).
\end{aligned}
$$

Thus, we get

$$
\begin{aligned}
E\{E(W_{ni_1}^2 \mid \mathcal{F}_{n,i_1-1}) E(W_{ni_2}^2 \mid \mathcal{F}_{n,i_2-1})\} &= E(A_n^{(1)} + B_n^{(1)} + C_n^{(1)} + D_n^{(1)} + E_n^{(1)}) \\
&= (i_1 - 1) o\left(\frac{n}{p_n^2}\right) + (i_1 - 1)(i_2 - 1)(p_n + 1)^2 \\
&\quad + 2(i_2 - 1)^2 (p_n + 1). \qquad (3.8.30)
\end{aligned}
$$

On the same lines we can show that

$$
E\{E(W_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2 = (i-1) o\left(\frac{n}{p_n^2}\right) + (i-1)^2 (p_n+1)^2 + 2(i-1)^2 (p_n+1). \qquad (3.8.31)
$$

Combining results (3.8.29), (3.8.30), (3.8.31) we obtain,

$$
\begin{aligned}
E\{\sum_{i=1}^{n} E(W_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2 &= \sum_{i=1}^{n} \left((i-1) o\left(\frac{n}{p_n^2}\right) + (i-1)^2 (p_n+1)^2 + 2(i-1)^2 (p_n+1) \right) \\
&\quad + 2 \sum_{i=1}^{n} \sum_{i_1=1}^{i-1} \left((i_1 - 1) o\left(\frac{n}{p_n^2}\right) + (i-1)(i_1-1)(p_n+1)^2 \right) \\
&\quad + 2 \sum_{i=1}^{n} \sum_{i_1=1}^{i-1} 2(i_1-1)^2 (p_n+1) \\
&= O(n^3) o\left(\frac{n}{p_n^2}\right) + n^4 (p_n+1)^2 \{1 + o(1)\} + n^4 (p_n+1)\{1 + o(1)\},
\end{aligned}
$$

$$E\{\sum_{i=1}^{n} E(\widetilde{W}_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2 = E\{\sum_{i=1}^{n} E(\frac{2}{n^2 p_n} W_{ni}^2 \mid \mathcal{F}_{n,i-1})\}^2$$
$$= 1 + o(1).$$

Thus, we have proved (3.8.28). This fact together with (3.8.27) completes the proof of (3.8.25).

Next, we shall sketch a proof of the fact

$$\sum_{i=1}^{n} E(\widetilde{W}_{ni}^4 \mid \mathcal{F}_{n,i-1}) \xrightarrow{P} 0. \tag{3.8.32}$$

This fact readily implies (3.8.26).

Note that

$$W_{ni}^2 = \sum_{i_1,i_2=1}^{i-1} \sum_{j_1 j_2 j_3 j_4=1}^{m} \sum_{j_5 j_6 j_7 j_8=1}^{m} \sum_{s_1 s_2 s_3 s_4=0}^{p_n} \times \frac{\xi_{s_1 s_2} \xi_{s_3 s_4} \varepsilon_{i_1 j_1}^{s_1} \varepsilon_{i_2 j_5}^{s_3} g'_{i_1 j_1}(\tilde{\beta}) g'_{i_2 j_5}(\tilde{\beta}) l_{j_1 j_2} l_{j_5 j_6} l_{j_3 j_4} l_{j_7 j_8} \tilde{e}_{i_1 j_2} \tilde{e}_{i_2 j_6}}{\sigma_{i_1 j_1} a(\tilde{\beta}) \sigma_{i_2 j_5}(\tilde{\beta})} \times \frac{\varepsilon_{ij_3}^{s_2} \varepsilon_{ij_7}^{s_4} g'_{ij_3}(\tilde{\beta}) g'_{ij_7}(\tilde{\beta}) \tilde{e}_{ij_4} \tilde{e}_{ij_8}}{\sigma_{ij_3}(\tilde{\beta}) \sigma_{ij_7}(\tilde{\beta})}.$$

Using techniques similar to the ones in the previous result, we can write terms in W_{ni}^4 in to 4 terms of type $i_1 = i_2 = i_3 = i_4$ that are of order $(i-1)o(n^2 p_n^2)$ and remaining terms that are of the order $i(i-1)p_n^4 o\left(n/p_n^2\right)$, to obtain

$$
\begin{aligned}
\sum_{i=1}^{n} E(W_{ni}^4) &= O(n^2)o(n^2p_n^2) + O(n^3)p_n^4 o\left(\frac{n}{p_n^2}\right), \\
\sum_{i=1}^{n} E(\widetilde{W}_{ni}^4) &= O\left(\frac{1}{n^4p_n^2}\right)\left\{O(n^2)o(n^2p_n^2) + O(n^3)p_n^4 o\left(\frac{n}{p_n^2}\right)\right\} \\
&= o(1). \qquad (3.8.33)
\end{aligned}
$$

Thus, from (3.8.33) we prove (3.8.26) and (3.8.24) and thereby Proposition 3.8.9. From Lemma 3.8.5 , we also prove 3.4.2 of Theorem 2. Next, we prove the remaining parts of Theorem 2.

We now prove 3.4.3 of Theorem2

Proof Recall, $D(\tilde{\beta}) = D$, $\Psi_n^{-1} = n\Gamma^{1/2}(D^TD)^{-1}\Gamma^{1/2}$ and $\tilde{\Gamma} = D^TD/n$.

From Lemma 3.8.4 and (3.8.9) we have,

$$
n(\hat{\beta} - \beta)^T\Gamma(\hat{\beta} - \beta) = Z_n^T\Gamma Z_n.
$$

Hence consider,

$$
\begin{aligned}
Z_n^T\Gamma Z_n - Z_n^T\tilde{\Gamma}Z_n &= Z_n^T\Gamma^{1/2}(I - \Psi_n^{-1})\Gamma^{1/2}Z_n \\
&= \chi_n^T\Psi_n(I - \Psi_n^{-1})\Psi_n\chi_n.
\end{aligned}
$$

Taking norm we get,

$$
\begin{aligned}
\left\|Z_n^T\Gamma Z_n - Z_n^T\tilde{\Gamma}Z_n\right\| &= \left\|\chi_n^T\Psi_n(I - \Psi_n^{-1})\Psi_n\chi_n\right\| \\
&\le |\chi_n\chi_n^T|\left\|\Psi_n(I - \Psi_n^{-1})\Psi_n\right\| = O_p(p_n)O_p\left(\frac{1}{p_n}\right).
\end{aligned}
$$

We use (3.8.11), (3.8.14) to get the above result. Thus, we obtain,

$$\left\| \frac{Z_n^T \Gamma Z_n - Z_n^T \tilde{\Gamma} Z_n}{(2p_n)^{1/2}} \right\| = o_p(1).$$

This concludes the proof of 3.4.3 from Theorem 2.

We next prove 3.4.4 Theorem 2

Proof Consider,

$$\left\| n(\hat{\beta} - \tilde{\beta})^T \frac{D^T D}{n} (\hat{\beta} - \tilde{\beta}) - n(\hat{\beta} - \tilde{\beta})^T \frac{D_{\hat{\beta}}^T D_{\hat{\beta}}}{n} (\hat{\beta} - \tilde{\beta}) \right\| \leq \left\| n(\hat{\beta} - \tilde{\beta}) \right\|^2 \left\| \frac{D^T D}{n} - \frac{D_{\hat{\beta}}^T D_{\hat{\beta}}}{n} \right\|.$$

We have $\left\| \frac{D^T D}{n} - \frac{D_{\hat{\beta}}^T D_{\hat{\beta}}}{n} \right\| = o_p(1)$ and from Theorem 1 we have $n\|\hat{\beta} - \tilde{\beta}\| = O_p(1)$. Thus, we get

$$\left\| n(\hat{\beta} - \tilde{\beta})^T \frac{D^T D}{n} (\hat{\beta} - \tilde{\beta}) - n(\hat{\beta} - \tilde{\beta})^T \frac{D_{\hat{\beta}}^T D_{\hat{\beta}}}{n} (\hat{\beta} - \tilde{\beta}) \right\| = o_p(1).$$

Thus, the result is proved.

Chapter 4

Measurement Error Model

In this chapter, we consider the problem of estimating the effect function in a functional regression set-up where the response is a scalar and the regressor is a function, in the presence of measurement error. This is as an extension of the functional model proposed in the previous chapter albeit with stronger assumptions. In several studies the covariates are contaminated with errors. Sometimes, this error can be easily determined and eliminated from the data. However, if it is random then we need to adjust our analysis to account for this measurement error as ignoring it can lead to bias in estimating parameters. We adapt the framework provided by Stefanski and Carroll (1987) to accommodate functional covariate. We then use the Karhunen-Loéve expansion to obtain estimating equations similar to those in Stefanski and Carroll (1987). The main difference between the two frameworks is that the number of parameters to be estimated in the functional case is diverging while in the non functional case is fixed. We prove the consistency of our estimates in the case that the response has normal distribution conditional on the functional regressor. We perform a limited simulation study to investigate the accuracy of our estimator.

There are two important features of measurement error models: the first one is the relationship between the original unobserved process $X(t)$ and the observed process $W(t)$ with error, and the type of data available to assess some characteristics of the measurement error. For the relationship between $X(t)$ and $W(t)$ we assume the classical additive error model i.e. $W(t) = X(t) + U(t)$, $E(U(t)|X(t), t \in T) = 0$, $T = [0, 1]$. We assume that replicates of $W(t)$ are available. Further details of this model will be given subsequently.

There is little literature available on measurement errors in functional model. Cardot et al. (2007) consider the linear functional model: $Y_i = \int \beta(t)X_i(t) + e_i, i = 1..., n$ and assume the following measurement error structure: $W_i(t_j) = X_i(t_j) + U_{ij}, i = 1, ..., n,\ j = 1, ..., p$ and $t_1 < t_2 <, ... < t_p$ are the discrete points at which the curves are observed. Noise components U_{ij} are assumed to be independent of each other and error e_i for all i, j. The noise components are not considered to be discrete realizations of a continuous time stochastic process and are interpreted as random measurement errors at finite discretization points. This particular measurement error framework is used because in practice we only observe curves at finite discrete points instead of the entire curve. Similar approach can be found in Cardot (2000). In her dissertation, Cai (2014) extends the simulation extrapolation (SIMEX) method developed by Carroll et al. (2006) to accommodate the measurement error at finite discrete points in linear as well as non-linear functional models. The measurement errors U_{ij} are allowed to have a sparse correlation structure like the autocorrelation structure and are assumed to be normally distributed. The main advantage of the method that we propose in this chapter is that it does not impose any restriction on the covariance function of the measurement error process $U(t), t \in T$.

4.1 Model

We observe the following independent sample $(W_i(t), t \in T = [0, 1], Y_i)$, $i = 1, ..., n$, where $W_i(t)$ is the surrogate of the true covariate $X_i(t)$ which is unobservable. We assume that $W_i(\cdot), 1 \leq i \leq n$ are random elements of $L^2 = L^2[0, 1]$ and that Y_i given the function $X_i(\cdot)$, has the following distribution with respect to a dominating measure m :

$$f_{Y_i}(y_i;\theta_1, x_i(\cdot)) = \exp\left(\frac{y_i(\beta_0 + \int_0^1 \tilde{\beta}(t)x_i(t)d(t)) - b(\beta_0 + \int_0^1 \tilde{\beta}(t)x_i(t)dt)}{a(\phi)} + c(y_i,\phi)\right). \tag{4.1.1}$$

All integrals hereafter are taken over $[0,1]$ and $X_i(\cdot), \forall i$ and $\tilde{\beta}(\cdot) \in L^2$, $\theta_1 = (\beta_0, \tilde{\beta}(\cdot), \phi)$.

Let $(\rho_k)_{k=1}^{\infty}$ be an orthonormal basis functions in L^2 space. Using basis expansion we obtain

$$E(Y_i;\theta_1, X_i(\cdot)) = \int \tilde{\beta}(t)X_i(t)dt = \sum_{k=1}^{\infty} X_{ik}\tilde{\beta}_k,$$

where $X_{ik} = \int X(t)\rho_k(t)dt$, $\tilde{\beta}_k = \int \tilde{\beta}(t)\rho_k(t)dt$. Model (4.1.1) has infinitely many parameters. We address this issue of infinite dimensions with a truncation strategy. Let $\tilde{\beta} = (\beta_1, ..., \beta_{p_n})^T$ and $\tilde{\theta} = (\beta_0, \tilde{\beta}, \phi)$. Instead of the model (4.1.1), we work with the following sequence of models with increasing dimension p_n:

$$f_{Y_i}(y_i;\tilde{\theta}, x_i) = \exp\left(\frac{y_i\left(\beta_0 + \sum\limits_{k=1}^{p_n} x_{ik}\tilde{\beta}_k\right) - b\left(\beta_0 + \sum\limits_{k=1}^{p_n} x_{ik}\tilde{\beta}_k\right)}{a(\phi)} + c(y_i,\phi)\right). \tag{4.1.2}$$

We assume the following relation between the true covariate $X_i(\cdot)$ and the observed surrogate $W_i(\cdot)$:

$$W_i(t) = X_i(t) + U_i(t),\ i = 1, ..., n,\ \forall t \in [0,1], \tag{4.1.3}$$

where U_i is a copy of Gaussian process U having mean function 0 and covariance function $K(\cdot,\cdot)$, $\forall i$. Take $(\rho_i)_{i=1}^{\infty}$ to be the basis formed from the eigenfunctions of the integral operator $\mathcal{K}$ associated with the covariance function $K(\cdot,\cdot)$. Let $W_{ik} = \int W_i(t)\rho_k(t)dt$ and $U_{ik} = \int U_i(t)\rho_k(t)dt$. This yields the following measurement error set-up for (4.1.2):

$$W_{ik} = X_{ik} + U_{ik},\ i = 1,..,n, k \geq 1.$$

Let, λ_k be the eigenvalue associated with the k^{th} eigenfunction of $\mathcal{K}$. From the Karhünen-Loève expansion we obtain that U_{ik} are independent and $U_{ik} \sim N(0, \lambda_k),\ i = 1,..,n, k \geq 1$. Thus, we have a additive measurement error set up where errors are independent with normal distribution.

This set-up is similar to that of Stefanski and Carroll (1987) where the authors proposed the sufficiency estimator for generalized linear models. We use their approach to propose estimators for the parameters. Denote $\mathbf{X}_i = (X_{i1}, ..., X_{ip_n})^T$. Let $\widetilde{\mathbf{W}}_i^{(l)}$ denote the l^{th} replicate of $\mathbf{W}_i = (W_{i1}, ..., W_{ip_n})^T$. The likelihood of the replicates $\widetilde{W}_i = (\widetilde{\mathbf{W}}_i^{(1)}, ..., \widetilde{\mathbf{W}}_i^{(m)})'$ is

$$f_{\widetilde{W}_i}(\widetilde{w}_i, \tilde{\theta}, \mathbf{x}_i) = \prod_{j=1}^{m} \frac{(2\pi)^{-p_n/2}}{|\Omega_1|} \exp\left(\frac{-1}{2} (\widetilde{\mathbf{w}}_i^{(j)} - \mathbf{x}_i)^T (\widetilde{\mathbf{w}}_i^{(j)} - \mathbf{x}_i) \right), \tag{4.1.4}$$

where, $\Omega_1 = \text{diag}(\lambda_1, ..., \lambda_{p_n})$ is the unknown covariance matrix of the measurement error vector and m denotes the number of replicates. We assume that $m = 1$ and $\Omega = \Omega_1/a(\phi)$ is known. The estimate of Ω_1 and $a(\phi)$ are given subsequently. Let $Y,\ \mathbf{W}$ denote random variables with the same distribution as $Y_i,\ \mathbf{W}_i,\ i = 1, ..., n$ respectively. For now we drop the subscript i and use it when necessary. We assume that $\mathbf{W}$ has no information on Y other than what is contained in $\mathbf{x}$ i.e

$$f_{Y,\mathbf{W}}(y, \mathbf{w}; \theta, \mathbf{x}) = f_Y(y; \theta, \mathbf{x}) f_{\mathbf{W}}(\mathbf{w}, \theta, \mathbf{x}). \tag{4.1.5}$$

Let $\mathbf{\Delta}(\tilde{\beta}) = (\Delta_1, ..., \Delta_{p_n})^T = \mathbf{\Delta} = \mathbf{W} + Y\Omega\tilde{\beta}$. The distribution of $Y|\mathbf{\Delta}$ is given by

$$f_{Y|\mathbf{\Delta}}(y|\mathbf{\Delta} = \delta; \theta) = \exp\left(y\eta - \frac{1}{2}y^2\tilde{\beta}^T\Omega\tilde{\beta}/a(\phi) + c(y, \phi) - log\left(S(\eta, \tilde{\beta}, \phi)\right)\right), \quad (4.1.6)$$

where $S(\eta, \tilde{\beta}, \phi) = \int \exp\left(y\eta - \frac{1}{2}y^2\tilde{\beta}^T\Omega\tilde{\beta}/a(\phi) + c(y, \phi)\right) dm(y)$, $\eta = (\beta_0 + \delta^T\tilde{\beta})/a(\phi)$. This distribution belongs to exponential family. This leads to the following estimating equations: $\Psi_s(y, \mathbf{w}, \theta) = (\partial/\partial\theta) log\, f_{Y/\mathbf{\Delta}}(y|\delta; \theta)$ evaluated at $\delta(\beta) = \mathbf{w} + y\Omega\beta$.

$$\Psi(y, \mathbf{w}, \theta) = \begin{bmatrix} (y - E(Y|\Delta(\beta) = \delta(\beta)))\,/a\,(\phi) \\ (y - E(Y|\mathbf{\Delta}(\beta) = \delta(\beta)))\,\delta(\beta)/a(\phi) - \left(y^2 - E(Y^2|\mathbf{\Delta}(\beta) = \delta(\beta))\right)\Omega\beta/a(\phi) \\ r(y, \mathbf{w}, \theta) - E(r(y, \mathbf{w}, \theta)|\mathbf{\Delta}(\beta) = \delta(\beta)) \end{bmatrix}. \quad (4.1.7)$$

where, $r(Y, \mathbf{w}, \theta) = \dfrac{\partial c(Y, \theta)}{\partial\phi} - Y\dfrac{\beta_0 + \delta(\beta)^T\beta}{a^2(\phi)}a'(\phi) + Y^2\dfrac{\beta^T\Omega\beta}{2a^2(\phi)}a'(\phi)$. Note that Ψ is unbiased for θ. Any $\hat{\theta}$ satisfying $\sum_{i=1}^{n}\Psi(\mathbf{w}_i, y_i, \hat{\theta}) = 0$ is called as a sufficiency estimator. This sufficiency estimator does not maximize the conditional likelihood.

We present the case when Y has normal distribution. In this case, (4.1.7) can be written as

$$\Psi(y, \mathbf{w}, \theta) = \begin{bmatrix} \frac{1}{\sigma^2}(y - \mu) \\ \frac{\Omega\beta}{1 + \beta^T\Omega\beta} - \frac{1}{\sigma^2}\left((y - \mu)^2\Omega\beta - (y - \mu)(\delta(\beta) - 2\mu\Omega\beta)\right) \\ \frac{-1}{2\sigma^2} + \frac{(y - \mu)^2(1 + \beta^T\Omega\beta)}{2\sigma^4} \end{bmatrix}. \quad (4.1.8)$$

Let $\mathbf{\Delta}^*(\beta) = (I + \Omega\beta\beta^T)^{-1}(\mathbf{\Delta}(\beta) - \beta_0\Omega\beta)$ and $\mu = (\beta_0 + \beta^T\delta(\beta))/(1 + \beta^T\Omega\beta)$. Consider the following equations:

$$U(\beta) = \sum_{i=1}^{n} (\mathbf{\Delta}_i^*(\beta) Y_i - \mathbf{\Delta}_i^*(\beta)\beta_0 - \mathbf{\Delta}_i^*(\beta)\beta^T \mathbf{\Delta}_i^*(\beta)) = 0, \tag{4.1.9}$$
$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta^T \mathbf{\Delta}_i^*(\beta)) = 0,$$
$$\sigma^2 = \frac{1 + \beta^T \Omega \beta}{n} \sum_{i=1}^{n} (Y_i - \mu_i)^2.$$

These equations resemble the usual normal equations. The solution to (4.1.9) is also a solution to (4.1.8). Note that the above equations are non linear in the parameters. Let $\mathbf{W}^* = \mathbf{W} - \overline{\mathbf{W}}, Y^* = Y - \overline{Y}$, where $\overline{Y}$ and $\overline{W}$ denotes the average of $Y_1, ..., Y_n$ and $\mathbf{W}_1, ..., \mathbf{W}_n$ respectively. The equations (4.1.9) can be re-written as

$$\beta_0 = \overline{Y} - \beta^T \overline{\mathbf{W}}, \tag{4.1.10}$$
$$U(\beta) = -\sum_{i=1}^{n} \mathbf{W}_i^* Y_i^* \beta^T \Omega \beta + \sum_{i=1}^{n} Y_i^{*2} \Omega \beta - \sum_{i=1}^{n} \mathbf{W}_i^* \mathbf{W}_i^{*T} \beta + \sum_{i=1}^{n} \mathbf{W}_i^* Y_i^* = 0,$$
$$\sigma^2 = \frac{1 + \beta^T \Omega \beta}{n} \sum_{i=1}^{n} (Y_i - \mu_i)^2.$$

The equation $U(\beta) = 0$ is a quadratic equation and thus has two roots. There is no way of determining which root is the correct. One way to handle this problem is to solve the equation iteratively starting from the naive estimator. Naive estimator can be obtained by treating W as the true covariate i.e. ignoring the measurement error. We state the assumptions needed to show that the equations (4.1.10) has a solution $\hat{\beta}$ and that this solution is consistent.

Assumption 1 $p = p_n \to \infty, \quad p_n n^{-1/8} \to 0.$

Assumption 2 $\sup_i \|\mathbf{X}_i\| = O(n^{1/4}).$

Assumption 3 *For all n we assume that $\|\Omega_1\| = O(\sqrt{p_n})$.*

Assumption 4 *There exist positive constants $b_1,\ b_2$ such that for all n,*

$$0 < b_1 \leq \lambda_{min}\left(\sum_{i=1}^{n}\frac{X_iX_i^T}{n}\right) \leq \lambda_{max}\left(\sum_{i=1}^{n}\frac{X_iX_i^T}{n}\right) \leq b_2 < \infty$$

Assumption 5 *There exists a constant c such that for all n,*

$$max\left\{\left|\lambda_{min}\left(\sum_{i_1\neq i_2}^{n}\frac{X_{i_1}X_{i_2}^T}{n^2}\right)\right|,\left|\lambda_{max}\left(\sum_{i_1\neq i_2}^{n}\frac{X_{i_1}X_{i_2}^T}{n^2}\right)\right|\right\} \leq c$$

Assumption 6 *For all n,*

$$\frac{\lambda_{max}(\Omega_1)}{\lambda_{min}(\Omega_1)} \leq \lambda_{min}\left(\sum_{i=1}^{n}\frac{X_iX_i^T}{n}\right)c^{-1},\ c \leq \lambda_{min}\left(\sum_{i=1}^{n}\frac{X_iX_i^T}{n}\right)$$

.

Lemma 4.1.1

$$\|U(\tilde{\beta})\| = \sqrt{np_n}.$$

Proof

We can show that the distribution of Y_i conditional on $\mathbf{\Delta}^*$ is normal with mean $\beta_0 + \tilde{\beta}^T\mathbf{\Delta}^*$.

$$\begin{aligned}
\|U(\tilde{\beta})\|^2 &= \sum_{k=1}^{p_n}\left(\sum_{i=1}^{n}(Y_i - E(Y_i|\mathbf{\Delta}_i^*))\Delta_{ik}^*\right)^2 \\
&= \sum_{k=1}^{p_n}\sum_{i=1}^{n}\left((Y_i - E(Y_i|\mathbf{\Delta}_i^*))\Delta_{ik}^*\right)^2 \\
&\quad + \sum_{k=1}^{p_n}\sum_{i_1\neq i_2=1}^{n}\left((Y_{i_1} - E(Y_{i1}|\Delta_{i1}^*))\Delta_{i_1k}^*\right)\left((Y_{i_2} - E(Y_{i_2}|\Delta_{i_2}^*))\Delta_{i_2k}^*\right)
\end{aligned}$$

Taking expectation and using the fact that that the samples are independent we get

$$E(||U(\tilde{\beta})||^2) = O(np_n).$$

As mentioned in chapter 3 the next theorem is sufficient to prove weak consistency of our estimator.

Theorem 4.1.1 *For all $\epsilon > 0$, there exists a constant $\zeta > 0$ such that for sufficiently large n,*

$$P\left(\sup_{||\beta-\tilde{\beta}||=\zeta\sqrt{p_n/n}} (\beta-\tilde{\beta})^T U(\beta) < 0\right) \geq 1-\epsilon.$$

Proof

Let $J(\beta) = \left(\frac{\partial U(\beta)}{\partial \beta}\right)$ and β^* be such that $\|\beta^* - \tilde{\beta}\| < \|\beta - \tilde{\beta}\|$. Using Taylors expansion we obtain,

$$\begin{aligned}(\beta-\tilde{\beta})^T U(\beta) &= (\beta-\tilde{\beta})^T U(\tilde{\beta}) + (\beta-\tilde{\beta})^T J(\beta^*)(\beta-\tilde{\beta}) \\ &= (\beta-\tilde{\beta})^T U(\tilde{\beta}) + (\beta-\tilde{\beta})^T (J(\beta^*) - J(\tilde{\beta}))(\beta-\tilde{\beta}) + (\beta-\tilde{\beta})^T J(\tilde{\beta})(\beta-\tilde{\beta}).\end{aligned}$$

Using representation (4.1.10) of $U(\beta)$ we get

$$J(\beta) = -\sum_{i=1}^{n} \Omega(\mathbf{W}^{*T}_i \beta) Y_i^* - \Omega\beta \mathbf{W}_i^{*T} Y_i^* + \Omega Y_i^{*2} - \mathbf{W}_i^* \mathbf{W}_i^{*T}.$$

Consider,

$$
\begin{aligned}
&(\beta-\tilde{\beta})^T(J(\beta^*)-J(\tilde{\beta}))(\beta-\tilde{\beta}) \\
&=(\beta-\tilde{\beta})^T\left(\sum_{i=1}^{n}-\Omega(\mathbf{W}_i^{*T}\beta^*)Y_i^*-\Omega\beta^*\mathbf{W}_i^{*T}Y_i^*+\Omega Y_i^{*2}-\mathbf{W}_i^*\mathbf{W}_i^{*T}\right)(\beta-\tilde{\beta}) \\
&\quad+(\beta-\tilde{\beta})^T\left(\sum_{i=1}^{n}\Omega(\mathbf{W}_i^{*T}\tilde{\beta})Y_i^*+\Omega\tilde{\beta}\mathbf{W}_i^{*T}Y_i^*-\Omega Y_i^{*2}+\mathbf{W}_i^*\mathbf{W}_i^{*T}\right)(\beta-\tilde{\beta}) \\
&=(\beta-\tilde{\beta})^T\sum_{i=1}^{n}\Omega Y_i^*(\mathbf{W}_i^{*T}\tilde{\beta}-\mathbf{W}_i^{*T}\beta^*)(\beta-\tilde{\beta})+(\beta-\tilde{\beta})^T\sum_{i=1}^{n}\Omega\mathbf{W}_i^{*T}Y_i^*(\tilde{\beta}-\beta^*)(\beta-\tilde{\beta}) \\
&=A_1+A_2
\end{aligned}
$$

Consider,

$$
\begin{aligned}
|A_1|&=|(\beta-\tilde{\beta})^T\sum_{i=1}^{n}\Omega Y_i^*(\mathbf{W}_i^{*T}\tilde{\beta}-\mathbf{W}_i^{*T}\beta^*)(\beta-\tilde{\beta})| \\
&\leq\sum_{i=1}^{n}|(\beta-\tilde{\beta})^T\Omega Y_i^*(\mathbf{W}_i^{*T}\tilde{\beta}-\mathbf{W}_i^{*T}\beta^*)(\beta-\tilde{\beta})| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^2\sum_{i=1}^{n}\|Y_i^*\mathbf{W}_i^{*T}(\tilde{\beta}-\beta^*)\| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^3\sum_{i=1}^{n}\|Y_i^*\mathbf{W}_i^{*T}\| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^3\sum_{i=1}^{n}\|Y_i^*(\mathbf{X}_i^{*T}+\mathbf{U}_i^{*T})\| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^3\sum_{i=1}^{n}\|Y_i^*\mathbf{X}_i^{*T}\|+Y_i^*\mathbf{U}_i^{*T}\| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^3\sum_{i=1}^{n}\|Y_i^*\mathbf{X}_i^{*T}\|+\|Y_i^*\mathbf{U}_i^{*T}\| \\
&\leq\|\Omega\|\|(\beta-\tilde{\beta})\|^3\sum_{i=1}^{n}\|Y_i^*\|\left(\sup_i\|X_i^*\|+\sup_i\|U_i^*\|\right)
\end{aligned}
$$

Taking norm and using Assumption $1, 2, 3$ we get,

$$\sup_{||\beta-\tilde{\beta}||=\zeta\sqrt{p_n/n}} ||(\beta-\tilde{\beta})^T(J(\beta^*)-J(\tilde{\beta}))(\beta-\tilde{\beta})|| = O_p\left(\frac{\tilde{\zeta}^3\sqrt{p_n}p_n^{1.5}nn^{1/4}}{n^{1.5}}\right) = o_p(1).$$

We can prove similarly for A_2. Now consider,

$$n^{-1}(\beta-\tilde{\beta})^T J(\tilde{\beta})(\beta-\tilde{\beta}) = n^{-1}(\beta-\tilde{\beta})^T\left(\sum_{i=1}^{n} -2\Omega(\mathbf{W}_i^{*T}\tilde{\beta})Y_i^* + \Omega Y_i^{*2} - \mathbf{W}_i^*\mathbf{W}_i^{*T}\right)(\beta-\tilde{\beta})$$

We have,

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n} -2\Omega(\mathbf{W}_i^{*T}\tilde{\beta})Y_i^* &\approx -2\Omega n^{-1}\sum_{i=1}^{n} E((\mathbf{W}_i^{*T}\tilde{\beta})Y_i^*) \\
&= -2\Omega\sum_{i=1}^{n}\frac{(X_i^T\tilde{\beta})^2}{n} + 2\Omega\sum_{i_1\neq i_2}^{n}\frac{(X_{i_1}^T\tilde{\beta})(X_{i_2}^T\tilde{\beta})}{n^2}, \\
n^{-1}\sum_{i=1}^{n}\Omega Y^{*2} &\approx n^{-1}\sum_{i=1}^{n}\Omega E Y^{*2} \\
&= \Omega + \sum_{i=1}^{n}\Omega\frac{(X_i^T\tilde{\beta})^2}{n} - \Omega\sum_{i_1\neq i_2}^{n}\frac{(X_{i_1}^T\tilde{\beta})(X_{i_2}^T\tilde{\beta})}{n^2}, \\
n^{-1}\sum_{i=1}^{n}\mathbf{W}_i^*\mathbf{W}_i^{*T} &\approx n^{-1}\sum_{i=1}^{n} E\mathbf{W}_i^*\mathbf{W}_i^{*T} \\
&= \sum_{i=1}^{n}\frac{X_iX_i^T}{n} - \Omega + \sum_{i_1\neq i_2}^{n}\frac{X_{i_1}X_{i_2}^T}{n^2}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&n^{-1}(\beta-\tilde{\beta})^T J(\tilde{\beta})(\beta-\tilde{\beta}) \\
&= (\beta-\tilde{\beta})^T \left(-\frac{\Omega_1}{\sigma^2}\tilde{\beta}^T \sum_{i=1}^n \frac{X_i X_i^T}{n}\tilde{\beta} + \frac{\Omega_1}{\sigma^2}\tilde{\beta}^T \sum_{i_1 \neq i_2}^n \frac{X_{i_1} X_{i_2}^T}{n^2}\tilde{\beta}\right)(\beta-\tilde{\beta}) \\
&\quad - (\beta-\tilde{\beta})^T \left(\sum_{i=1}^n \frac{X_i X_i^T}{n} + \sum_{i_1 \neq i_2}^n \frac{X_{i_1} X_{i_2}^T}{n^2}\right)(\beta-\tilde{\beta}).
\end{aligned}
$$

Let $\sum_{i=1}^n \frac{X_i X_i^T}{n} = Q$, $\sum_{i_1 \neq i_2}^n \frac{X_{i_1} X_{i_2}^T}{n^2} = R, max\{|\lambda_{max}(R)|, |\lambda_{min}(R)|\} = c$. Taking supremum

$$
\begin{aligned}
&\sup_{||\beta-\tilde{\beta}||=\zeta\sqrt{p_n/n}} (\beta-\tilde{\beta})^T J(\tilde{\beta})(\beta-\tilde{\beta}) \\
&\leq \zeta^2 p_n \left(\frac{-\lambda_{min}(\Omega_1)}{\sigma^2}\tilde{\beta}^T Q\tilde{\beta} + \frac{\lambda_{max}(\Omega_1)}{\sigma^2}\left|\tilde{\beta}^T R\tilde{\beta}\right| - \lambda_{min}(Q)\right) \\
&\quad + \zeta^2 p_n \lambda_{max}(R) \\
&\leq \zeta^2 p_n \left(\frac{-\lambda_{min}(\Omega_1)}{\sigma^2}\lambda_{min}(Q)\|\tilde{\beta}\|^2 + \frac{\lambda_{max}(\Omega_1)}{\sigma^2}c\|\tilde{\beta}\|^2 - \lambda_{min}(Q) + \lambda_{max}(R)\right) = \zeta^2 p_n A
\end{aligned}
$$

Assumptions $4, 5, 6$ ensure that $A \leq 0$. Thus, the theorem is proved.

4.2 Simulation

We report the details of a limited simulation study that investigates the accuracy of our proposed estimator in presence of measurement error. While our method does not put restrictions on the covariance function of the error process, given the challenges of generating a process with the specific covariance structure we considered the identity covariance function i.e. $\sigma_2 I_{\{s=t\}}$. This makes it easier to control the error in the model.

Following Xiaochen Cai (2014) we generated the functional covariate using the following:

$$X_i(t) = \sum_{k=1}^{p_n} \varepsilon_{ik}\rho_k(t),$$

where, $\varepsilon_{ik} \sim N(0,1)$ and are independent for all $i = 1, ..., n$ and $k = 1, ...p_n$. We set $p_n = 8$. We generated the basis function using the Canadian weather data set Ramsay (2006). This data set is available in the fda package in R software. We first smoothed the data using B-splines smoothing that is described in the simulation section in Chapter 3. We generated the response using $Y_i = \int X_i(t)\beta(t)dt + \epsilon_i$, where $\epsilon_i \sim N(0,1)$ and are independent for all i. We then generated the surrogate variable $W_i(t)$ using $W_i(t) = X_i(t) + U_i(t)$ where, $U_i(t)$ is a Gaussian process with 0 mean function and covariance function given as $K(s,t) = \sigma_2 I_{\{s=t\}}$. We vary the value of σ_2 to study the effect of measurement error on the estimation procedure. We use 100 replicates to estimate the covariance function. Let $\widetilde{\mathbf{W}}_{i1}, ..., \widetilde{\mathbf{W}}_{i100}$ denote the replicates of $\mathbf{W}_i$ and $\widetilde{\mathbf{W}}_{i.}$ denote their mean. The estimate of the covariance function $\mathcal{K}(\cdot,\cdot)$ of the measurement error is

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{100}(\widetilde{\mathbf{W}}_{ij} - \widetilde{\mathbf{W}}_{i.})(\widetilde{\mathbf{W}}_{ij} - \widetilde{\mathbf{W}}_{i.})^T}{n(100-1)}.$$

Let $\hat{\lambda}_k$ denote the k^{th} eigenvalue of the $\hat{\mathcal{K}}(\cdot,\cdot)$. Then, $\hat{\Omega}_1 = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_{p_n})$. The dimension p_n is chosen by using 5-fold cross validation. We solve the (4.1.10) iteratively with start value obtained from the naive estimator. The following table reports the result of the simulation. Error is calculated as $(\hat{\beta} - \tilde{\beta})^T(\hat{\beta} - \tilde{\beta})/\tilde{\beta}^T\tilde{\beta}$. In Table 4.1, the naive estimator is the one where we ignore the measurement error i.e treat $W(t)$ as the true covariate. The corrected estimator is the one obtained by solving (4.1.10). We can see that as the measurement error increases, the error

Table 4.1: Error in the estimator as a function of σ_2

Error(σ_2)	Naive	Corrected
0.1	0.009	0.001
0.3	0.056	0.006
0.4	0.08	0.008
0.5	0.115	0.02
0.8	0.202	0.08
1	0.256	0.855

in the estimation increases for both the estimators. We can also see that the corrected estimator improves the accuracy of the estimate. Recall that the estimating equation yields multiple solutions and we do not have a way of choosing the correct one. To avoid this problem, we solve the equation iteratively starting from the naive estimator. However, as the measurement error increases in the data, the start values are farther from the true value and hence, the iterative algorithm does not always converge to the true value. We observe that the corrected estimator performed better than the naive one only for a limited range of σ_2.

BIBLIOGRAPHY

BIBLIOGRAPHY

Balan, R. M., Schiopu-Kratina, I., et al. (2005). Asymptotic results with generalized estimating equations for longitudinal data. The Annals of Statistics, 33(2):522–541.

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. Journal of Nonparametric Statistics, 12(4):503–538.

Cardot, H., Crambes, C., Kneip, A., and Sarda, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. Computational statistics & data analysis, 51(10):4832–4848.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. Statistics & Probability Letters, 45(1):11–22.

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

Cardot, H. and Johannes, J. (2010). Thresholding projection estimators in functional linear models. Journal of Multivariate Analysis, 101(2):395–408.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. Journal of Multivariate Analysis, 92(1):24 – 41.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). Measurement error in nonlinear models: a modern perspective. CRC press.

Cirulli, E. T. and Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics, 11(6):415–425.

Gertheiss, J., Maity, A., and Staicu, A.-M. (2013). Variable selection in generalized functional linear models. Stat, 2(1):86–101.

Hall, P. and Heyde, C. (1980). Martingale limit theory and its application. Probability and mathematical statistics. Academic Press.

Hicks, B. M., Schalet, B. D., Malone, S. M., Iacono, W. G., and McGue, M. (2011). Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. Behavior genetics, 41(4):459–475.

Horváth, L. and Kokoszka, P. (2012). Inference for functional data with applications, volume 200. Springer Science & Business Media.

Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. John Wiley & Sons.

Laird, N. M. and Lange, C. (2010a). The fundamentals of modern statistical genetics. Springer Science & Business Media.

Laird, N. M. and Lange, C. (2010b). The Fundamentals of Modern Statistical Genetics. Springer Publishing Company, Incorporated, 1st edition.

Lee, S., Abecasis, G. R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. The American Journal of Human Genetics, 95(1):5–23.

Li, Y., Wang, N., and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. Journal of the American Statistical Association, 105(490):621–633.

Madsen, B. E. and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet, 5(2):e1000384.

Morris, J. S. (2015). Functional regression. Annual Review of Statistics and Its Application, 2:321–359.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. Ann. Statist., 33(2):774–805.

Ortega, J. M. and Rheinboldt, W. C. (2000). Iterative Solution of Nonlinear Equations in Several Variables. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Ramsay, J. O. (2006). Functional data analysis. Wiley Online Library.

Saccone, N. L., Saccone, S. F., Hinrichs, A. L., Stitzel, J. A., Duan, W., Pergadia, M. L., Agrawal, A., Breslau, N., Grucza, R. A., Hatsukami, D., et al. (2009). Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (chrn) genes. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150(4):453–466.

Schaid, D. J., McDonnell, S. K., Sinnwell, J. P., and Thibodeau, S. N. (2013). Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. Genetic epidemiology, 37(5):409–418.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. Biometrika, pages 703–716.

Vrieze, S. I., Malone, S. M., Vaidyanathan, U., Kwong, A., Kang, H. M., Zhan, X., Flickinger, M., Irons, D., Jun, G., Locke, A. E., et al. (2014). In search of rare variants: preliminary results from whole genome sequencing of 1,325 individuals with psychophysiological endophenotypes. Psychophysiology, 51(12):1309–1320.

Vsevolozhskaya, O. A., Zaykin, D. V., Greenwood, M. C., Wei, C., and Lu, Q. (2014). Functional analysis of variance for association studies. PloS one, 9(9):e105074.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. Annual Review of Statistics and Its Application, 3:257–295.

Wang, L. et al. (2011). Gee analysis of clustered binary data with diverging number of covariates. The Annals of Statistics, 39(1):389–417.

Wang, X., Lee, S., Zhu, X., Redline, S., and Lin, X. (2013). Gee-based snp set association test for continuous and discrete traits in family-based association studies. Genetic epidemiology, 37(8):778–786.