Generalized Quasi-Likelihood Ratio Test for Semiparametric Analysis of Covariance Models in Longitudinal Data

by

Jin Tang

(Under the direction of Yehua Li)

Abstract

Semiparametric regression models have been wildly applied into the longitudinal data. In this dissertation, we model generalized longitudinal data from multiple treatment groups by a class of semiparametric analysis of covariance models, which take into account the parametric effects of time dependent covariates and the nonparametric time effects. In these models, the treatment effects are represented by nonparametric functions of time and we propose a generalized quasi-likelihood ratio (GQLR) test procedure to test if these functions are the same. We first consider an estimation approach for our semiparametric regression model based on profile estimation equations combined with local linear smoothers. Next, we describe the proposed GQLR test procedure and study the asymptotic null distribution of test statistic. We find that the much celebrated Wilks phenomenon which is well established for independent data still holds for longitudinal data if variance is estimated consistently, even though the working correlation structure is misspecifed. However, this property does not hold in general, especially when the wrong working variance function is assumed. As for the power of the proposed GQLR test, our empirical study shows that incorporating correlation into the GQLR test does not necessarily improve the power of the test. A more extensive simulation study is conducted in which the Wilks Phenomenon is investigated under both Gaussian and Non-Gaussian longitudinal data and a wider variety of scenarios. The proposed methods are also illustrated with two real applications from AIDS clinical trial and opioid agonist treatment.

INDEX WORDS: Generalized quasi-likelihood ratio test, Semiparametric regression, Longitudinal data, Kernel smoothing, Bootstrap, Hypothesis testing, Analysis of variance, Functional data.

# Generalized Quasi-Likelihood Ratio Test for Semiparametric Analysis of Covariance Models in Longitudinal Data

by

Jin Tang

B.S., Southwest University, 2003

M.S., East China Normal University, 2006

M.S., University of Georgia, 2009

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2014

# Generalized Quasi-Likelihood Ratio Test for Semiparametric Analysis of Covariance Models in Longitudinal Data

by

Jin Tang

Approved:

Major Professor: Yehua Li

Committee: William McCormick
Jaxk Reeves
Lily Wang
Xiangrong Yin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2014

# Dedication

To my parents, Qi and Jason

# Acknowledgments

First and foremost, I would like to express my sincere appreciation to my advisor Dr. Yehua Li, whose expertise, constant guidance, endless encouragement and precious advice make it possible for me to pursue this interesting topic and complete my dissertation. I am sincerely grateful for his long-term support and help, especially for his patience in helping me improve my writing skills. It has been a great pleasure to be his student. His kindness, genuine caring and concern will always be remembered.

I would also like to thank my committee members Dr. William McCormick, Dr. Jaxk Reeves, Dr. Lily Wang and Dr. Xiangrong Yin for their time and effort in serving on my doctoral committee despite their heavy loads of responsibilities. Special thanks go to Dr. Jaxk Reeves for his kind help and understanding and for providing me an invaluable opportunity to work in the Statistical Consulting Center.

I am grateful to all the professors, staff members and friends I have met in the Department of Statistics, who made our department such a great place to work in and made my years at UGA enjoyable and memorable.

Last, but certainly not least, I want to express my deepest gratitude to my parents. Without their unconditional love and support, I would not have come this far. I wish I could show them how much I love and appreciate them. I am also deeply indebted to my husband, Qi Li, whose love and encouragement allowed me to overcome all the difficulties and finish this journey.

# Table of Contents

## List of Figures

## List of Tables

# Chapter 1

## Introduction and Literature Review

### 1.1 Introduction

In many longitudinal studies, the response variable is repeatedly measured during the follow-up and the treatment effects are represented as functions over time. The responses within the same subject are usually strongly correlated, while the variation between subjects, even within the same treatment group, is large. One example in our research is from an AIDS clinical trial study. This data involved 1309 AIDS patients with CD4 counts of less than or equal to 50 cells/$mm^3$. They were randomized to into four treatment groups and their CD4 counts were measured at baseline and at 8-week intervals during the 40 weeks of follow-up. The measurement times are unbalanced and irregular. We can describe the treatment effects as smooth functions over time. Therefore, it is desirable to model the time effect nonparametrically, while modeling all other covariate effects parametrically.

In our study, we consider a marginal semiparametric model that consists of $q$ treatment groups, and the $k$th group comprises $n_k$ independent clusters (subjects) with the $i$th cluster consisting $m_{k,i}$ repeated measures over a time interval $\mathcal{T}$. All subjects are independent, but the observations within a subject are correlated. Let $Y_{k,ij}$ and $(\boldsymbol{X}_{k,ij}, T_{k,ij})$ be the response variable and covariates for the $j$th visit of the $i$th subject in the $k$th group, where $\boldsymbol{X}_{k,ij}$ is a $p$-dim covariate vector whose effect is modeled parametrically and $T_{k,ij}$ is a scalar covariate (e.g. the visiting time) whose effect is modeled nonparametrically. Denote $\boldsymbol{Y}_{k,i} = (Y_{k,i1}, \ldots, Y_{k,im_{k,i}})^{\mathrm{T}}$, $\boldsymbol{X}_{k,i} = (\boldsymbol{X}_{k,i1}, \ldots, \boldsymbol{X}_{k,im_{k,i}})^{\mathrm{T}}$ and $\boldsymbol{T}_{k,i} = (T_{k,i1}, \ldots, T_{k,im_{k,i}})^{\mathrm{T}}$, and let the conditional mean and variance of the response be $\mathrm{E}(Y_{k,ij}|\boldsymbol{X}_{k,ij}, T_{k,ij}) = \mu_{k,ij}$ and $\mathrm{var}(Y_{k,ij}|\boldsymbol{X}_{k,ij}, T_{k,ij}) = \sigma^2(\mu_{k,ij})$, where $\sigma^2(\cdot)$ is a conditional variance function. Given that the subject belongs to the $k$th treatment group, the relationship between $Y$ and $(\boldsymbol{X}, T)$ is modeled by

$$g(\boldsymbol{\mu}_{k,i}) = \boldsymbol{X}_{k,i}\boldsymbol{\beta} + \theta_k(\boldsymbol{T}_{k,i}), \quad k = 1, \ldots, q, \quad i = 1, \ldots, n_k, \tag{1.1}$$

where $g(\cdot)$ is a known link function, $\boldsymbol{\beta}$ is a $p$-dim regression coefficient vector, and $\theta_k(\cdot)$'s are nonparametric functions.

The covariate vector $\boldsymbol{X}$ often contains the baseline information about the subjects (e.g. age and gender) which need to be controlled in order to have a fair comparison of the treatment effects, and therefore its effect is of less interest for our particular problem. Moreover, it is well known that, in a semiparametric model like (1.1), the parametric component $\boldsymbol{\beta}$ can be estimated with a root-$n$ convergence rate and the estimator is asymptotically normal. One can easily construct Wald type of tests for various hypotheses regarding $\boldsymbol{\beta}$. It is of primary interest to us to test the treatment effects, which are represented by $\theta_k(t)$'s. Specifically the hypotheses of interest are

$$H_0 : \theta_1 = \cdots = \theta_q \quad \text{vs.} \quad H_1 : \text{ not all } \theta_k\text{'s are the same.} \tag{1.2}$$

Since the treatment effects are represented by nonparametric functions, model (1.1) can be considered as a class of functional ANOVA model. Some related literature includes, Brumback and Rice (1998), Morris and Carroll (2006) and more recently Zhou et al. (2011). A common approach in these papers is to express each trajectory as a linear combination of some basis functions and adapt the coefficients into a linear mixed model. Such hierarchical models usually require a relatively large number of repeated measures in each cluster and many random effects to capture the features of each longitudinal trajectory. In contrast, our method does not require a large number of repeated measures and is applicable to the so-called sparse functional data (Yao, Muller and Wang, 2005). More importantly, these conditional methods provide pointwise confidence intervals for $\theta_k$'s, which can not replace a rigorous test as we are about to propose.

## 1.2 Literature review

### 1.2.1 Estimation of marginal semiparametric regression models

There is a vast volume of work on semiparametric models for longitudinal data. Suppose that $Y_{ij}$ and $(\boldsymbol{X}_{ij}, T_{ij})$ are the response variable and covariates of the $j$th observation of $i$th subject. Consider a marginal semiparametric generalized partially linear model

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta(T_{ij}),$$

where $\mu_{ij} = \mathrm{E}(Y_{ij}|\boldsymbol{X}_{ij}, T_{ij})$, $g(\cdot)$ is a link function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and $\theta(t)$ is a smooth function. This model is a special case of our model (1.1) with only one treatment group.

Zeger and Diggle (1994) considered this semiparametric model in the Gaussian case. They estimated $\theta(t)$ by local polynomial kernel smoothing ignoring the within-cluster correlation, and estimated $\boldsymbol{\beta}$ by a weighted least square accounting for the within-cluster correlations. Severini and Staniswalis (1994) proposed an approach based on profile-kernel Generalized

Estimating Equations (GEEs). Lin and Carroll (2001) considered another class of profile-kernel estimating equations allowing different working correlation matrices in estimating $\boldsymbol{\beta}$ and $\theta(t)$ and using local linear kernel estimating equations for $\theta(t)$ instead of local average kernel equations. They showed that the estimator $\widehat{\boldsymbol{\beta}}$ is $\sqrt{n}$-consistent when the correlation structure is ignored completely or the nonparametric component $\theta(t)$ is undersmoothed. In fact, such an estimator of $\boldsymbol{\beta}$ is not semiparametric efficient if one accounts for correlations. More efficient kernel estimators were proposed by Wang et al. (2003, 2005). They estimated $\theta(t)$ by solving a kernel estimating equation accounting for within-subject correlations, and estimated $\boldsymbol{\beta}$ by solving a profile-type equation, which requires an iterative procedure.

Here we briefly describe the method proposed by Lin and Carroll (2001) since our estimation procedure is based on it. Denote $K_h(t) = h^{-1}K(t/h)$, where $K(\cdot)$ is a symmetric probability density function and $h$ is the bandwidth. Denote $\mathbf{U}_i(t) = (\mathbf{U}_{i1}(t), \ldots, \mathbf{U}_{im_i}(t))^T$ with $\mathbf{U}_{ij}(t) = \{1, (T_{ij} - t)/h\}^T$, $\mu_{ij}(\boldsymbol{X}_{ij}, t) = g^{-1}\{\boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \mathbf{U}_{ij}^{\mathrm{T}}(t)\boldsymbol{\alpha}\}$ with $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^{\mathrm{T}}$, and $\Delta_i(\boldsymbol{X}_i, t) = \mathrm{diag}\{\mu_{ij}^{(1)}\}_{j=1}^{m_i}$, where $\mu_{ij}^{(1)}$ is the first derivative of $\mu(\cdot) = g^{-1}(\cdot)$ evaluated at $\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{U}_{ij}^{\mathrm{T}}(t)\boldsymbol{\alpha}$. To get estimators of the finite-dimensional parameter $\boldsymbol{\beta}$ and the infinite-dimensional parameter $\theta(t)$, two estimating equations need to be solved.

Given $\boldsymbol{\beta}$, let $\widehat{\boldsymbol{\alpha}}(t) = (\widehat{\alpha}_0, \widehat{\alpha}_1)^{\mathrm{T}}$ be the solution of

$$\sum_{i=1}^{n} \mathbf{U}_i(t)^{\mathrm{T}} \Delta_i(\boldsymbol{X}_i, t) \mathbf{K}_h^{1/2}(\boldsymbol{T}_i - t) \mathbf{V}_{1i}^{-1} \mathbf{K}_h^{1/2}(\boldsymbol{T}_i - t)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{X}_i, t)\} = 0, \qquad (1.3)$$

where $\mathbf{K}_h(\boldsymbol{T}_i - t) = \mathrm{diag}\{K_h(T_{ij} - t)\}_{j=1}^{m_i}$, $\boldsymbol{\mu}_i = \{\mu_{ij}\}_{j=1}^{m_i}$, and $\mathbf{V}_{1i}$ is a working covariance. The kernel GEE estimator of $\theta(t)$ is $\widehat{\theta}(t; \boldsymbol{\beta}) = \widehat{\alpha}_0$.

Next, we can estimate $\boldsymbol{\beta}$ by solving the following profile estimating equation

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}(\boldsymbol{T}_i; \boldsymbol{\beta})\}^{\mathrm{T}}}{\partial \boldsymbol{\beta}} \mathbf{V}_{2i}^{-1}[\boldsymbol{Y}_i - \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}(\boldsymbol{T}_i; \boldsymbol{\beta})\}] = 0, \qquad (1.4)$$

where $\mathbf{V}_{2i}$ is another working covariance matrix, which can be different from $\mathbf{V}_{1i}$.

Lin and Carroll (2001) derived the asymptotic properties of the estimators described in equations (1.3) and (1.4). In our study, we are interested in extending these results to model (1.1) with multiple treatment groups and testing if the nonparametric functions from different treatment groups are the same.

### 1.2.2 Generalized likelihood ratio tests on nonparametric models

The generalized likelihood ratio (GLR) test was proposed by Fan et al. (2001) and Fan and Jiang (2005), who showed that it is a general methodology for testing various nonparametric hypotheses in many useful models, including nonparametric regression models, varying coefficient models and additive models. We briefly review the local linear approach proposed by Fan et al. (2001) to construct a GLR test statistic in varying-coefficient models.

Suppose that $\{Y_i, X_i, T_i\}_{i=1}^n$ are a random sample from the model

$$Y_i = A(T_i)X_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\{\epsilon_i\}$ are i.i.d. random variables from $N(0, \sigma^2)$. For simplicity, suppose that $X_i$ is a 1-dim covariate, and $A(\cdot)$ is an unspecified smooth function.

First, we construct the estimators of the nonparametric component $A(t)$ and the variance $\sigma^2$. Denote $\boldsymbol{\alpha}(t) = \{\alpha_0(t), \alpha_1(t)\}^{\mathrm{T}}$, and $\mathbf{U}_i = \{X_i, X_i(T_i - t)/h\}^{\mathrm{T}}$. For each given t, the local maximum likelihood estimator $\widehat{\boldsymbol{\alpha}}(t)$ can be obtained by maximizing the local log-likelihood

$$\ell(\boldsymbol{\alpha}) = -n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}_i)^2 K_h(T_i - t)$$

with respect to $\boldsymbol{\alpha}$. The local linear kernel estimator of $A(t)$ is $\widehat{A}(t) = \widehat{\alpha}_0(t)$. The estimated variance $\widehat{\sigma}^2$ maximizes

$$\ell = -n\log(\sqrt{2\pi}\sigma) - \frac{\sum_i \{Y_i - \widehat{A}(T_i)X_i\}^2}{2\sigma^2}$$

with respect to $\sigma^2$.

Consider the hypothesis testing problem:

$$H_0 : A = A_0, \quad H_1 : A \neq A_0,$$

where $A_0$ is a known function. Define sums of squares $RSS_0 = \sum_i \{Y_i - A_0(T_i)X_i\}^2$ and $RSS_1 = \sum_i \{Y_i - \widehat{A}(T_i)X_i\}^2$, the maximum log-likelihoods under the null and alternative hypotheses are

$$\ell_n(H_0) = -(n/2)\log(2\pi/n) - (n/2)\log(RSS_0) - n/2,$$

and

$$\ell_n(H_1) = -(n/2)\log(2\pi/n) - (n/2)\log(RSS_1) - n/2.$$

Taking the difference between $\ell_n(H_0)$ and $\ell_n(H_1)$ leads to the generalized likelihood ratio test statistic

$$\lambda_n(A_0) = \ell_n(H_1) - \ell_n(H_0) = \frac{n}{2}\log\frac{RSS_0}{RSS_1} \approx \frac{n}{2}\frac{RSS_0 - RSS_1}{RSS_1}.$$

Fan et al. (2001) showed that the asymptotic distribution of $\lambda_n(A_0)$ is a $\chi^2$ distribution independent of the unknown parameter $A(t)$. This allows us to easily obtain critical values for the GLR tests using either the asymptotic chi-square distribution or a bootstrap method. In general, the above approach can be extended to various composite hypothesis tests, such as testing linearity of the coefficient functions (Fan et al., 2001) and testing the significance of variables in additive models (Fan et al., 2005).

### 1.2.3 Semiparametric generalized likelihood ratio test for variable selection

Variable selection is fundamental in statistical modeling. A variety of variable selection procedures for parametric models have been studied, including bridge regression (Frank and Friedman, 1993), LASSO (Tibshirani, 1996) and nonconcave penalized likelihood (Fan and Li, 2001). For semiparametric regression models, a key challenge of variable selection is that it includes selection of significant variables in both parametric and nonparametric components. Li and Liang (2008) considered a generalized varying-coefficient partially linear model. They adopted a penalized likelihood approach to select the parametric components and used generalized likelihood ratio tests to select the nonparametric components, thus extended the generarlized likelihood ratio test from fully nonparametric models to semiparametric models for variable selection.

Let $Y$ be a response variable and $\{\boldsymbol{X}, \mathbf{Z}, T\}$ be the covariates. The generalized varying-coefficient partially linear model based on $n$ i.i.d. samples is given by

$$g\{\mu(T_i, \boldsymbol{X}_i, \mathbf{Z}_i)\} = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(T_i), \quad i = 1, \cdots, n,$$

where $\mu = \mathrm{E}(Y|\boldsymbol{X}, \mathbf{Z}, T)$, $g(\cdot)$ is a link function, $\boldsymbol{\beta}$ is a $q \times 1$ unknown coefficient vector and $\boldsymbol{\theta}(T_i)$ is a $p \times 1$ vector of unspecified smooth functions.

To estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}(t)$, Li and Liang(2008) proposed the following procedure:

- *Step I.* Denote $\boldsymbol{a} = (a_1, \cdots, a_p)^{\mathrm{T}}$ and $\boldsymbol{b} = (b_1, \cdots, b_p)^{\mathrm{T}}$, then the initial local estimators $(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{a}}^{\mathrm{T}}, \widehat{\boldsymbol{b}}^{\mathrm{T}})^{\mathrm{T}}$ maximize

$$\sum_i \mathcal{Q}[g^{-1}\{\boldsymbol{a}^{\mathrm{T}}\mathbf{Z}_i + \boldsymbol{b}^{\mathrm{T}}\mathbf{Z}_i(T_i - t) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}\}, Y_i]K_h(T_i - t), \tag{1.5}$$

where $\mathcal{Q}(\cdot)$ is a quasi-likelihood function (McCullagh, 1983). The local linear estimator is $\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{a}}$.

- *Step II.* The initial estimator of $\boldsymbol{\beta}$ was estimated locally, hence not efficient. In this step, we update $\widehat{\boldsymbol{\beta}}$ by maximizing the global penalized likelihood

$$\sum_{i=1}^{n} \mathcal{Q}[g^{-1}\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(T_i) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}\}, Y_i] - n\sum_{j=1}^{q} P_{\lambda_j}(|\beta_j|) \tag{1.6}$$

  with respect to $\boldsymbol{\beta}$, where $\widehat{\boldsymbol{\theta}}(T_i)$ is the initial estimator obtained from step I, and $P_{\lambda_j}(\cdot)$ is a prespecified penalty function with a regularization parameter $\lambda_j$ chosen by generalized cross validation (GCV).

- *Step III.* Replacing $\boldsymbol{\beta}$ in (1.5) by estimator $\widehat{\boldsymbol{\beta}}$, we can update $\widehat{\boldsymbol{\theta}}(t)$ by maximizing the local likelihood function (1.5) again.

Now, we consider the following hypotheses:

$$H_0 : \theta_1(t) = \cdots = \theta_p(t) = 0, \quad \text{vs.} \quad H_1 : \text{ not all } \theta_k\text{'s are } 0.$$

Define $\widehat{\boldsymbol{\beta}}_F$ and $\widehat{\boldsymbol{\theta}}_F(t)$ as the estimators under the alternative hypothesis, and $\widehat{\boldsymbol{\beta}}_R$ as the estimator under the null hypothesis. Denote $K * K$ as the convolution of the kernel function $K$, so that $K * K(t) = \int_{-\infty}^{\infty} K(s)K(t-s)ds$. The generalized quasi-likelihood ratio test statistic can be expressed as

$$T_{GLR} = r_k\{\ell(H_1) - \ell(H_0)\},$$

where

$$r_k = \left\{K(0) - \frac{1}{2}\int K^2(t)dt\right\}\left\{\int \{K(t) - \frac{1}{2}K * K(t)\}dt\right\}^{-1},$$

$$\ell(H_0) = \sum_i \mathcal{Q}\{g^{-1}(\boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_R), Y_i\},$$

$$\ell(H_1) = \sum_i \mathcal{Q}\{g^{-1}(\widehat{\boldsymbol{\theta}}_F^{\mathrm{T}}(T_i)\mathbf{Z}_i + \boldsymbol{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_F), Y_i\}.$$

The penalized quasi likelihood (1.6) selects the non-zero components in $\boldsymbol{\beta}$ by forcing the insignificant components to be 0. The nonparametric test has the same theoretical properties as in Fan et al. (2001). The nonparametric tests in Fan et al. (2001, 2005) and Li and Liang (2008) were proposed for independent data. Extensions of these results to longitudinal data have not been studied and we want to investigate the effect of correlation on the null distribution and power of the generalized quasi-likelihood ratio tests.

## 1.3 Outline of the dissertation

As discussed above, the development of GLR tests has provided us general and powerful tools to address various nonparametric testing problems. The contributions in this dissertation is to develop a generalized quasi-likelihood ratio (GQLR) test to detect the treatment effects for longitudinal data with multiple treatment groups.

In Chapter 2, we propose profile-kernel equation methods to estimate the parameters in model (1.1) under both the null and alternative hypotheses in (1.2), and study the asymptotic properties of the proposed estimators. Under some regularity conditions, the proposed estimators of parametric parameters are $\sqrt{n}$-consistent and asymptotically normal. The asymptotic expansions of the nonparametric components are derived as well.

In Chapter 3, we propose a new GQLR test and study its null distribution. Our theoretical results indicate that the Wilks phenomenon proved by Fan et al. (2001) for independent data continues to hold in our hypothesis testing if the variance function is correctly specified and consistently estimated, while the within-cluster correlation can be completely misspecified. However, the Wilks phenomenon does not hold in general. In addition, we derive the local power of the GQLR test and show that the proposed test achieves the minimax power rate. We also describe the estimating algorithm for our semiparametric model and bootstrap procedures to estimate the p-value of the test.

In Chapter 4, we conduct three simulations to demonstrate the null distribution and the power of the proposed GQLR tests. The three scenarios in our simulation studies are Gaussian data with homogenous variance, Gaussian data with heterogeneous variance and binary data. Our numerical results provide strong evidence that corroborates our asymptotic theory.

In Chapter 5, we present two applications in AIDS clinical trial and opioid agonist treatment. In both applications, the proposed GQLR test detect significant differences among treatment groups.

# Chapter 2

## Estimation of the Semiparametric Analysis of Covariance Model

In this chapter, we will describe the estimation procedure for model (1.1) using a profile-kernel method similar to the one proposed by Lin and Carroll (2001). We use a working independence (WI) estimator because it is still one of the most widely used estimators in longitudinal and functional data analysis, see Fan and Li (2004), Yao et al. (2005), Hall et al. (2006, 2008). It is therefore of practical value to study the nonparametric test procedures based on the working independence estimators.

### 2.1 Estimation under both null and alternative models

We refer to the model under the null hypothesis as the reduced model and that under the alternative hypothesis as the full model. We denote $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\theta}_R(t)$ as the estimators under the reduced model and $\widehat{\boldsymbol{\beta}}_F$ and $\widehat{\theta}_{F,k}(t)$, $k = 1, \ldots, q$, as the estimators under the full model. Our estimation procedures under both models are based on profile-kernel estimating equations.

We first consider estimation under the reduced model, where the treatment effects are all the same. Based on the Taylor's expansion, for any $T_{k,ij}$ in a neighborhood $h$ of t, $\theta(T_{k,ij})$ can be approximated locally by a linear polynomial

$$\theta(T_{k,ij}) \approx \theta(t) + \theta'(t)(T_{k,ij} - t) = \alpha_0 + \alpha_1(T_{k,ij} - t)/h.$$

For a given $\boldsymbol{\beta}$, $\widehat{\theta}_R(t)$ can be obtained by solving the following local linear kernel estimating equation regarding $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^{\mathrm{T}}$,

$$\sum_{k=1}^{q}\sum_{i=1}^{n_k} \mathbf{U}_{k,i}(t)^T \Delta_{k,i}(\boldsymbol{X}_{k,i}, t)\mathcal{W}_{k,i}^{-1}\mathbf{K}_h(\boldsymbol{T}_{k,i} - t)\{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t)\} = 0, \qquad (2.1)$$

where $\mathbf{K}_h(\boldsymbol{T}_{k,i} - t), \mathbf{U}_{k,i}(t), \Delta_{k,i}(\boldsymbol{X}_{k,i}, t), \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t)$ are the same as those in (1.3) and (1.4) and $\mathcal{W}_{k,i}$ is a diagonal weight matrix set to be $\mathcal{W}_{k,i} \equiv \mathcal{W}(\boldsymbol{T}_{k,i}) = \mathrm{diag}\{\omega(\mu_{k,ij})\}_{j=1}^{m_{k,i}}$ and $\omega(\cdot)$ is a working variance function. The kernel estimator is given by $\widehat{\theta}_R(t;\boldsymbol{\beta}) = \widehat{\alpha}_0$, where $(\widehat{\alpha}_0, \widehat{\alpha}_1)$ is the solution of (2.1). Then $\widehat{\boldsymbol{\beta}}_R$ is obtained by solving

$$\sum_{k=1}^{q}\sum_{i=1}^{n_k}\{\boldsymbol{X}_{k,i}^{\mathrm{T}} + \frac{\partial\widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\}\Delta_{k,i}(\boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i})\mathcal{W}_{k,i}^{-1}\{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i})\} = 0, \qquad (2.2)$$

where $\boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i}) = g^{-1}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta})\}$. In real life, the weight matrix need to be estimated and the estimation of $\mathcal{W}$ will be discussed in Section 3.3.2.

Next, we consider estimation under the full model, where we need to estimate $\theta_k(\cdot)$ using the $k$th treatment group only. Given $\boldsymbol{\beta}$, $\widehat{\theta}_{F,k}(t, \boldsymbol{\beta}) = \widehat{\alpha}_0$, where $\widehat{\boldsymbol{\alpha}} = (\alpha_0, \alpha_1)^{\mathrm{T}}$ is the solution of

$$\sum_{i=1}^{n_k} \mathbf{U}_{k,i}(t)^T \Delta_{k,i}(\boldsymbol{X}_{k,i}, t)\mathcal{W}_{k,i}^{-1}\mathbf{K}_h(\boldsymbol{T}_{k,i} - t)\{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t)\} = 0. \qquad (2.3)$$

To obtain $\widehat{\boldsymbol{\beta}}_F$, we will again solve an estimating equation by pooling all treatment groups together

$$\sum_{k=1}^{q}\sum_{i=1}^{n_k}\{\boldsymbol{X}_{k,i}^{\mathrm{T}} + \frac{\partial\widehat{\boldsymbol{\theta}}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i})\} = 0. \qquad (2.4)$$

The nonparametric components are then updated as $\widehat{\theta}_{F,k}(t) = \widehat{\theta}_{F,k}(t, \widehat{\boldsymbol{\beta}}_F)$, for $k = 1, \ldots, q$.

## 2.2 Assumptions and asymptotic properties

In this section, we investigate the asymptotic properties of the profile-kernel estimators of $\boldsymbol{\beta}$ and $\theta(t)$ under both the full and reduced models. Denote the true parameters as $\boldsymbol{\beta}_0$ and $\theta_{k0}(t)$, $k = 1, \ldots, q$. Under the reduced model, $\theta_{10} = \cdots = \theta_{q0} \equiv \theta_0$. For ease of exposition, we assume the number of observations per subject is a constant in our theoretical derivations, i.e. $m_{k,i} = m < \infty$ for all $k$ and $i$. For the situations where the number of repeated measurements are unequal, a common practice is to model $m_{k,i}$ as independent realizations of a positive random variable $m$, and essentially the same results can be derived. We assume that the observation times $T_{k,ij}$ are independent random variables on a compact interval $\mathcal{T} = [a, b]$, with a density $f(t)$ and allow the covariate $\boldsymbol{X}$ to be time dependent.

What makes our problem fundamentally differently from those in the GLR test literature is the existence of within-cluster correlation. Define the errors as $\varepsilon_{k,ij} = Y_{k,ij} - g^{-1}\{\boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \theta_{k0}(T_{k,ij})\}$. It is helpful to consider $\boldsymbol{X}_{k,ij}$ and $\varepsilon_{k,ij}$ as discrete observations on continuous longitudinal processes $\boldsymbol{X}(t)$ and $\varepsilon(t)$, $t \in \mathcal{T}$. Define the conditional variance and correlation functions of the error process as

$$\sigma^2(\mu) = \mathrm{var}\left[\varepsilon(t) \middle| g^{-1}\{\boldsymbol{X}^{\mathrm{T}}(t)\boldsymbol{\beta} + \theta(t)\} = \mu\right], \quad \mathcal{R}(s, t; \boldsymbol{\tau}) = \mathrm{corr}\{\varepsilon(s), \varepsilon(t) \mid \boldsymbol{\tau}\}, \tag{2.5}$$

where $\boldsymbol{\tau}$ is a vector of unknown correlation parameters. Note that many authors model the variance function $\sigma^2(\cdot)$ as a nonparametric function while model the correlation function $\mathcal{R}(s, t; \boldsymbol{\tau})$ as a member of a parametric family, such as the AR or ARMA correlations (see Fan et al., 2007 and Fan and Wu, 2008). The within-cluster covariance matrix is

$$\Sigma_{k,i} = \boldsymbol{S}_{k,i}\mathcal{R}_{k,i}(\boldsymbol{\tau})\boldsymbol{S}_{k,i}, \tag{2.6}$$

where $\boldsymbol{S}_{k,i} = \text{diag}\{\sigma(\boldsymbol{\mu}_{k,i})\}$ and $\mathcal{R}_{k,i}(\boldsymbol{\tau}) = \{\mathcal{R}(T_{k,ij}, T_{k,ij'}; \boldsymbol{\tau})\}_{j,j'=1}^{m_{k,i}}$. We allow the covariance to depend on the mean and the parameters $\boldsymbol{\beta}$ and $\theta(\cdot)$, since this is usually the case for generalized longitudinal data, e.g. binary data.

We first collect all the key assumptions for the asymptotic theory.

(C.1) Let the total number of clusters be $n = \sum_{k=1}^{q} n_k$, and we assume that there exist constants $0 < \rho_1, \ldots, \rho_q < 1$ such that $\sum_{k=1}^{q} \rho_k = 1$ and $n_k/n - \rho_k = o(n^{-1/2})$.

(C.2) Let $T$ be a generic copy of the random observation time with a continuous density $f(t)$ so that $f(t) > 0$ for all $t \in \mathcal{T}$.

(C.3) We assume $\boldsymbol{X}$ and $\varepsilon$ are stochastic processes. Let $\boldsymbol{X}$ be a generic copy of $\boldsymbol{X}(T)$ and $\Sigma_k = (\sigma^2_{k,jj'})_{j,j'=1}^{m}$ be a generic copy of the true within-subject covariance matrix, which might depend on the mean parameter $\theta_{k0}$.

(C.4) Assume that the true mean functions $\theta_{k0}(\cdot)$, $k = 1, \ldots, q$, are smooth and twice continuously differentiable. Using the shorthand notation $\mu_k(\boldsymbol{X}, T) = g^{-1}\{\boldsymbol{X}^{\mathrm{T}}(T)\boldsymbol{\beta}_0 + \theta_{k0}(T)\}$, $\omega_k(\boldsymbol{X}, T) = \omega\{\mu_k(\boldsymbol{X}, T)\}$ and $\mu_k^{(1)}(\boldsymbol{X}, T) = \mu^{(1)}\{\boldsymbol{X}^{\mathrm{T}}(T)\boldsymbol{\beta}_0 + \theta_{k0}(T)\}$, we define

$$B_{1k}(t) = E[\{\mu_k^{(1)}(\boldsymbol{X}, T)\}^2 \omega_k^{-1}(\boldsymbol{X}, T) | T = t] f(t), \qquad B_1(t) = \sum_{k=1}^{q} \rho_k B_{1k}(t),$$
$$\boldsymbol{\mu}_{X,k}(t) = E[\{\mu_k^{(1)}(\boldsymbol{X}, T)\}^2 \omega_k^{-1}(\boldsymbol{X}, T) \boldsymbol{X}(T) | T = t] f(t)/B_{1k}(t),$$
$$\boldsymbol{\mu}_X(t) = B_1^{-1}(t) \sum_{k=1}^{q} \rho_k \mathrm{E}[\{\mu_k^{(1)}(\boldsymbol{X}, T)\}^2 \omega_k^{-1}(\boldsymbol{X}, T) X(T) \mid T = t] f(t).$$

Assume all functions defined above are Lipschitz continuous.

(C.5) The kernel density function $K(\cdot)$ is a symmetric continuous function with mean 0 and unit variance, i.e. $\int tK(t)dt = 0$ and $\int t^2 K(t)dt = 1$.

(C.6) Assume that $h \to 0$ as $n \to \infty$, $nh^2 \to \infty$, $nh^8 \to 0$.

<u>Proposition</u> **1.** *Suppose the null hypothesis in (1.2) is true and conditions (C.1)-(C.6) hold. Let* $\mathcal{W}^{-1} = diag(\omega^{jj})_{j=1}^m$ *and* $\Delta = diag(\Delta_{\ell\ell})_{\ell=1}^m$ *be generic copies of* $\mathcal{W}_{k,i}^{-1}$ *and* $\Delta_{k,i}$*, and let* $\widetilde{\boldsymbol{X}}$ *be a* $m \times p$ *matrix with the* $j$*th row being* $\widetilde{\boldsymbol{X}}_j = \boldsymbol{X}(T_j) - \boldsymbol{\mu}_X(T_j)$*. Then the parametric estimator has the following asymptotic expansion*

$$\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 = \boldsymbol{D}^{-1}\mathcal{E}_n + o_p(n^{-1/2}), \tag{2.7}$$

*where* $\boldsymbol{D} = \mathrm{E}(\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Delta\widetilde{\boldsymbol{X}})$ *and* $\mathcal{E}_n = n^{-1}\sum_{k=1}^q\sum_{i=1}^{n_k}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\boldsymbol{\varepsilon}_{k,i}$.

*The nonparametric estimator has the following asymptotic expansion*

$$\widehat{\theta}_R(t) - \theta_0(t) \quad = \quad \frac{1}{2}\theta_0^{(2)}(t)h^2 + \mathcal{U}_R(t) - \boldsymbol{\mu}_X^{\mathrm{T}}(t)(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) + o_p(n^{-1/2}), \tag{2.8}$$

*where* $\mathcal{U}_R(t) = \{nmB_1(t)\}^{-1}\sum_{k=1}^q\sum_{i=1}^{n_k}\sum_{j=1}^m \mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij}-t)\varepsilon_{k,ij}$*, and* $\mu_{k,ij}^{(1)}$ *and* $\omega_{k,i}^{jj}$ *are shorthands for* $\mu^{(1)}\{\boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \theta_{k0}(T_{k,ij})\}$ *and* $\omega^{-1}(\mu_{k,ij})$.

The asymptotic results in Proposition 1 are standard for generalized partially linear models (Lin and Carroll, 2001, Fan and Li, 2004, and Wang et al., 2005). Note that under the null hypothesis, $B_1(t)$ and $\boldsymbol{\mu}_X(t)$ can be simplified as $B_1(t) = \mathrm{E}[\{\mu^{(1)}(\boldsymbol{X},T)\}^2\omega^{-1}(\boldsymbol{X},T)|T = t]f(t)$ and $\boldsymbol{\mu}_X(t) = B_1^{-1}(t)\mathrm{E}[\{\mu^{(1)}(\boldsymbol{X},T)\}^2\omega^{-1}(\boldsymbol{X},T)X(T)|T = t]f(t)$. The proof is given in Appendix 2.3. With similar arguments, one can easily show the following results regarding the estimators under the full model.

<u>Proposition</u> **2.** *Under the full model (1.1) and suppose conditions (C.1)-(C.6) hold,*

$$\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0 = \boldsymbol{D}_*^{-1}\mathcal{E}_{n*} + o_p(n^{-1/2}), \tag{2.9}$$

*where* $\boldsymbol{D}_* = \sum_{k=1}^q \rho_k \mathrm{E}\{\widetilde{\boldsymbol{X}}_k^{\mathrm{T}} \Delta_k \mathcal{W}_k^{-1} \Delta_k \widetilde{\boldsymbol{X}}_k\}$, $\mathcal{E}_{n*} = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} \Delta_{k,i} \mathcal{W}_{k,i}^{-1} \boldsymbol{\varepsilon}_{k,i}$, $\widetilde{\boldsymbol{X}}_k = \{(\boldsymbol{X} - \boldsymbol{\mu}_{X,k})(T_\ell)\}_{\ell=1}^m$ *is a* $m \times p$ *matrix containing the centered covariate vectors in a generic cluster in group* $k$*, and* $\Delta_k$ *and* $\mathcal{W}_k^{-1}$ *are similarly defined as in Proposition 1 by replacing* $\theta_0$ *with* $\theta_{k0}$.

*The nonparametric estimator* $\widehat{\theta}_{F,k}(t)$ *has the following asymptotic expansion,*

$$\widehat{\theta}_{F,k}(t) - \theta_{k0}(t) = \frac{\theta_{k0}^{(2)}(t) h^2}{2} + \mathcal{U}_{F,k}(t) - \boldsymbol{\mu}_{X,k}^{\mathrm{T}}(t)(\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + o_p(n_k^{-1/2}), \qquad (2.10)$$

*for* $k = 1, \ldots, q$*, where* $\mathcal{U}_{F,k}(t) = \{n_k m B_{1k}(t)\}^{-1} \sum_{i=1}^{n_k} \sum_{j=1}^m \mu_{k,ij}^{(1)} \omega_{k,i}^{jj} K_h(T_{k,ij} - t) \varepsilon_{k,ij}$.

The asymptotic results in Proposition 2 are derived in a broader setting. When the null hypothesis in (1.2) holds, one can easily see that the first order expansions of $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\boldsymbol{\beta}}_F$ are identical, and hence $\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}_F = o_p(n^{-1/2})$.

## 2.3 Appendix: Technical Proofs

**Proof of Proporsition 1**

Define $\boldsymbol{\beta}_0$ and $\theta_0(t)$ are the true values of the parameter $\boldsymbol{\beta}$ and $\theta_k(t)$ under the null hypothesis and $\boldsymbol{\alpha} = \{\theta_0(t), h\theta_0^{(1)}(t)\}^{\mathrm{T}}$. Let $\boldsymbol{X}$ and $\boldsymbol{T}$ be generic copies of $\boldsymbol{X}_{k,i}$ and $\boldsymbol{T}_{k,i}$, similarly, let $\Delta = \mathrm{diag}(\Delta_{jj})_{j=1}^m$ and $\mathcal{W} = \mathrm{diag}(\omega^{jj})_{j=1}^m$ be generic copies of $\Delta_{k,i}$ and $\mathcal{W}_{k,i}^{-1}$. Following Lin and Carroll (2001), given a fixed $\boldsymbol{\beta}$, a linear Taylor expansion of (2.1) gives

$$\widehat{\boldsymbol{\alpha}}(t, \boldsymbol{\beta}) - \boldsymbol{\alpha}(t) = B_n^{-1} A_n + o_p(1),$$

where

$$B_n = n^{-1}\{\sum_{k=1}^q \sum_{i=1}^{n_k} \mathbf{U}_{k,i}(t)^{\mathrm{T}} \Delta_{k,i}(\boldsymbol{X}_{k,i}, t) \mathcal{W}_{k,i}^{-1} \mathbf{K}_h(\boldsymbol{T}_{k,i} - t) \Delta_{k,i}(\boldsymbol{X}_{k,i}, t) \mathbf{U}_{k,i}(t)\}$$

$$A_n = n^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\mathbf{U}_{k,i}(t)^{\mathrm{T}}\Delta_{k,i}(\boldsymbol{X}_{k,i},t)\mathcal{W}_{k,i}^{-1}\mathbf{K}_h(\boldsymbol{T}_{k,i}-t)[\boldsymbol{Y}_{k,i}-g^{-1}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta}+\mathbf{U}_{k,i}^{\mathrm{T}}\boldsymbol{\alpha}(t)\}]$$

Let $\varepsilon_{k,ij} = Y_{k,ij} - g^{-1}\{\boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \theta_0(T_{k,ij})\}$ under the null hypothesis, some calculations show that:

$$\begin{aligned}
\widehat{\theta}_R(t,\boldsymbol{\beta}) - \theta_0(t) &= \{nmB_1(t)\}^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij}-t)\Big[\varepsilon_{k,ij}+\mu_{k,ij}^{(1)}\boldsymbol{X}_{k,ij}^{\mathrm{T}}(\boldsymbol{\beta}_0-\boldsymbol{\beta}) \\
&\quad + \mu_{k,ij}^{(1)}\{\theta_0(T_{k,ij})-\theta_0(t)\}\Big] + o_p(n^{-1/2}) \\
&= \{nmB_1(t)\}^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij}-t)\varepsilon_{k,ij} - \boldsymbol{\mu}_X^{\mathrm{T}}(t)(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \\
&\quad + \frac{\theta_0^{(2)}(t)h^2}{2} + o_p(n^{-1/2})
\end{aligned}$$

where $B_1(t) = \mathrm{E}[\{\mu^{(1)}(\boldsymbol{X},T)\}^2\omega^{-1}(\boldsymbol{X},T)|T=t]f(t)$ and $\boldsymbol{\mu}_X(t) = B_1^{-1}(t)\mathrm{E}[\{\mu^{(1)}(\boldsymbol{X},T)\}^2\omega^{-1}(\boldsymbol{X},T)X(T)|T=t]f(t)$.

To study the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_R$, define that $\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} = \boldsymbol{X}_{k,i}^{\mathrm{T}} + \partial\widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \boldsymbol{X}_{k,i}^{\mathrm{T}} - \boldsymbol{\mu}_X^{\mathrm{T}}(\boldsymbol{T}_{k,i})$, a linear Taylor expansion of the profile estimating equation (2.2) gives

$$\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 = \boldsymbol{D}_n^{-1}\boldsymbol{C}_n + o_p(1),$$

where

$$\boldsymbol{D}_n = \frac{1}{n}\sum_k\sum_i\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\Delta_{k,i}\widetilde{\boldsymbol{X}}_{k,i},$$

and

$$\boldsymbol{C}_n = \frac{1}{n}\sum_k\sum_i\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\Big[\boldsymbol{Y}_{k,i} - g^{-1}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta}_0 + \widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i})\}\Big].$$

We can rewrite $\boldsymbol{D}_n$ as $\boldsymbol{D}_n = \boldsymbol{D} + o_p(1)$, where $\boldsymbol{D} = \mathrm{E}(\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\Delta_{k,i}\widetilde{\boldsymbol{X}}_{k,i})$. Furthermore, $\boldsymbol{C}_n$ can be expressed as

$$\boldsymbol{C}_n = \frac{1}{n}\sum_k\sum_i\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\Big[\boldsymbol{\varepsilon}_{k,i} - \Delta_{k,i}\{\widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}) - \boldsymbol{\theta}_0(\boldsymbol{T}_{k,i})\}\Big] + o_p(1)$$

Note that

$$
\begin{aligned}
&\frac{1}{n}\sum_k\sum_i \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\Delta_{k,i}\{\widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta})-\boldsymbol{\theta}_0(\boldsymbol{T}_{k,i})\}\\
&\quad=\frac{1}{n}\sum_{k,i,j}\widetilde{\boldsymbol{X}}_{k,ij}\Delta_{k,ij}\mathcal{W}_{k,ij}^{-1}\Delta_{k,ij}\Bigg[\frac{1}{nmB_1(T_{k,ij})}\sum_{k',i',j'}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}K_h(T_{k',i'j'}-T_{k,ij})\varepsilon_{k',i'j'}\\
&\qquad\qquad -\boldsymbol{\mu}_X^{\mathrm{T}}(T_{k,ij})(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\frac{\theta_0^{(2)}(T_{k,ij})h^2}{2}\Bigg]+o_p(n^{-1/2})\\
&\quad=\frac{1}{n}\sum_{k,i,j}\frac{1}{B_1(T_{k,ij})}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\varepsilon_{k,ij}\frac{1}{m}\sum_{l}^{m}\mathrm{E}(\widetilde{X}_l\Delta_{ll}^2\omega^{ll}|T_l=T_{k,ij})f(T_{k,ij})\\
&\qquad -\mathrm{E}\{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Delta\boldsymbol{\mu}_X(\boldsymbol{T})\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\frac{h^2}{2}\mathrm{E}\{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Delta\boldsymbol{\theta}_0^{(2)}(\boldsymbol{T})\}+o_p(n^{-1/2})
\end{aligned}
$$

Under the assumption that the marginal density of $(\boldsymbol{X}_{k,il},T_{k,il})$ is the same for all $k,i$ and $l$, we have

$$
\begin{aligned}
\mathrm{E}(\widetilde{X}_l\Delta_{ll}^2\omega^{ll}|T_l=t) &= \mathrm{E}[\{X_l-\boldsymbol{\mu}_X(T_l)\}\Delta_{ll}^2\omega^{ll}|T_l=t]\\
&= \mathrm{E}\left[\{X_l-B_1^{-1}(T_l)\mathrm{E}(\Delta_{jj}^2\omega^{jj}X_j|T_j=T_l)\}\Delta_{ll}^2\omega^{ll}|T_l=t\right] \quad (2.11)
\end{aligned}
$$

It is obvious to see that $\mathrm{E}(\widetilde{X}_l\Delta_{ll}^2\omega^{ll}|T_l=t)$=0. Similar calculation shows that $\mathrm{E}\{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Delta\boldsymbol{\mu}_X(\boldsymbol{T})\}=0$ and $\mathrm{E}\{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Delta\boldsymbol{\theta}_0^{(2)}(\boldsymbol{T})\}=0$.

Finally, we have the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$

$$
\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0=\boldsymbol{D}^{-1}\mathcal{E}_n+o_p(n^{-1/2}).
$$

where $\mathcal{E}_n=\frac{1}{n}\sum_k\sum_i\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\boldsymbol{\varepsilon}_{k,i}$. Equivalently,

$$
\sqrt{n}(\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)\to N(0,\mathbf{V}_{\boldsymbol{\beta}})
$$

where $\mathbf{V}_{\boldsymbol{\beta}}=\boldsymbol{D}^{-1}\mathrm{E}(\widetilde{\boldsymbol{X}}^{\mathrm{T}}\Delta\mathcal{W}^{-1}\Sigma\mathcal{W}^{-1}\Delta\widetilde{\boldsymbol{X}})\boldsymbol{D}^{-1}$.

Proposition 2 can be shown using similar arguments.

# Chapter 3

# Generalized Quasi-Likelihood Ratio Test

## 3.1 Quasi-likelihood function in longitudinal data

The quasi-likelihood function is an important extension of the likelihood function used for estimation in generalized linear model. As we know, to define a likelihood, we need to specify the form of distribution of the data. However, to define a quasi-likelihood function, we need only to specify the model structures in the mean and the variance. It requires a much weaker assumption for estimation in various models. Hence, the quasi-likelihood function is widely used for the situations where there is insufficient information to construct a likelihood function.

Wedderburn (1974) and McCullagh (1983, 1989) proposed the quasi-likelihood function in the case where the response variables are independent. Let $Y_i$, $i = 1, \cdots, n$, be independent variables with mean $\mu_i$ and variance $\text{var}(Y_i) \propto \mathbf{V}(\mu_i)$, where $\mathbf{V}(\cdot)$ is a specific variance function. The quasi-likelihood $\mathcal{Q}$, considered as a function of $\mu_i$, is given by the system of partial differential equations

$$\frac{\partial \mathcal{Q}(\mu_i; Y_i)}{\partial \mu_i} = \mathbf{V}^{-1}(\mu_i)(Y_i - \mu_i). \tag{3.1}$$

Liang and Zeger (1986) extended the quasi-likelihood function from independent data to longitudinal data by taking the within-subject correlation into account. Their generalized estimating equation method for parametric models is an important approach in longitudinal data analysis. Consider the longitudinal data $\{\boldsymbol{Y}_i, \boldsymbol{X}_i\}_{i=1}^n$ with $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{in_i})^{\mathrm{T}}$ and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}^{\mathrm{T}}, \cdots, \boldsymbol{X}_{in_i}^{\mathrm{T}})^{\mathrm{T}}$. The marginal mean of $\boldsymbol{Y}_i$ is $\boldsymbol{\mu}_i$ which depends on the covariates $\boldsymbol{X}_i$ through a known link function $g(\cdot)$, i.e., $g(\boldsymbol{\mu}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$. The variance function in (3.1) is defined as $\mathbf{V}_i^{-1} = \mathcal{S}_i\mathcal{R}(\boldsymbol{\tau})\mathcal{S}_i$, where $\mathcal{S}_i = \mathrm{diag}\{\sqrt{\mathrm{var}(Y_{i1})}, \cdots, \sqrt{\mathrm{var}(Y_{in_i})}\}$ and $\mathcal{R}(\boldsymbol{\tau})$ is a correlation structure. By taking the derivative of quasi-likelihood function $\mathcal{Q}$ with respect to $\boldsymbol{\beta}$, we can get the GEEs

$$\sum_{i=1}^{n} \mathcal{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\varepsilon}_i = 0,$$

where $\mathcal{D}_i = \partial\mu_i/\partial\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$. Besides GEEs, semiparametric regression modeling is also useful for longitudinal data analysis. See, for example, Lin and Carroll (2001, 2006), He, Zhu, and Fuang (2002), Fan and Li (2004), Wang et al. (2005) and He Fung, and Zhu (2005).

Since the variance function is an essential element of the quasi-likelihood function, modeling and estimating the variance and correlation structures become important issues in the quasi-likelihood approach. Nonparametric approaches have gained popularity in estimating the covariance structure. Wu and Pourahmadi (2003) adopted Fan and zhang's (2000) two-step estimation procedure of functional linear models and proposed nonparametric estimatiors of large covariance matrices. Hall et al. (2006) considered a bivariate smoothing to estimated the covariance function. Huang, Liu, Pourahmadi, and Liu (2006) introduced a penalized likelihood method to estimate a covariance matrix. Fan, Huang and Li (2007) proposed a kernel estimator of the nonparametric variance function and introduced a quasi-likelihood method and a minimum generalized variance method to estimate correlation parameters. We will describe Fan, Huang and Li's (2007) estimating procedure in details in section 3.3.2.

## 3.2 Test procedure and null distribution

We now direct our focus back to testing the hypotheses in (1.2). In this section, we introduce our test procedure which is based on a quasi-likelihood function, and study the asymptotic distribution of the test statistic under the null hypothesis.

The quasi-likelihood function $\mathcal{Q}$ satisfies

$$\frac{\partial \mathcal{Q}(\boldsymbol{\mu}, \boldsymbol{Y})}{\partial \boldsymbol{\mu}} = V(\boldsymbol{\mu})^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}), \tag{3.2}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{Y}$ are $m$-vectors for the conditional mean and response within a cluster, $\boldsymbol{\mu} = g^{-1}\{\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\theta}(\boldsymbol{T})\}$ and $\mathbf{V}(\boldsymbol{\mu})$ is a working covariance matrix not necessarily the same as the true covariance $\Sigma(\boldsymbol{\mu})$. Define the quasi-likelihood function for the data as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{q} \sum_{i=1}^{n_k} \mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta} + \boldsymbol{\theta}_k(\boldsymbol{T}_{k,i})\}, \boldsymbol{Y}_{k,i}], \tag{3.3}$$

and the generalized quasi-likelihood test statistics is by taking the difference of the quasi-likelihoods under the full and reduced models

$$\lambda_n(H_0) = \sum_{k=1}^{q} \sum_{i=1}^{n_k} \mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\widehat{\boldsymbol{\beta}}_F + \widehat{\boldsymbol{\theta}}_{F,k}(\boldsymbol{T}_{k,i})\}, \boldsymbol{Y}_{k,i}] - \mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\widehat{\boldsymbol{\beta}}_R + \widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i})\}, \boldsymbol{Y}_{k,i}]. \tag{3.4}$$

Let $\Sigma$ and $\mathbf{V}$ be generic copies of $\Sigma_{k,i}$ and $\mathbf{V}_{k,i}$, and denote $\sigma_{j\ell}$ and $\nu^{j\ell}$ as the $(j, \ell)$th entry of $\Sigma$ and $\mathbf{V}^{-1}$ respectively. Denote $\nu_K = \int K^2(t)dt$, and

$$B_2(t) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}(\Delta_{jj}^2 \nu^{jj} | T_j = t) f(t), \quad B_3(t) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\{\sigma_{jj}\Delta_{jj}^2(\omega^{jj})^2 | T_j = t\} f(t),$$

$$B_4(t) = \frac{1}{m}\sum_{j_1=1}^{m} \mathrm{E}\left\{ \sum_{j_2=1}^{m}\sum_{j_3=1}^{m} \nu^{j_1 j_2}\sigma_{j_2 j_3}\nu^{j_3 j_1}\Delta_{j_1 j_1}^2 \middle| T_{j_1} = t \right\} f(t),$$

$$B_5(t) = \frac{1}{m}\sum_{j_1=1}^{m} \mathrm{E}\bigg\{\frac{1}{m}\sum_{j=1}^{m}\sigma_{jj_1}\nu^{jj_1}\Delta^2_{j_1j_1}\omega^{j_1j_1}\bigg|T_{j_1}=t\bigg\}f(t). \tag{3.5}$$

<u>Theorem</u> **1.** *Suppose conditions (C.1)- (C.6) hold, $B_2(t)$- $B_5(t)$ defined in (3.5) exist and are Lipschitz continuous in $t$. Under the null hypothesis in (1.2)*

$$\sigma_n^{-1}(\lambda_n(H_0) - \mu_n - d_{1n}) \xrightarrow{d} N(0,1),$$

*where $d_{1n} = o_p(h^{-1/2})$,*

$$\begin{aligned} \mu_n &= \frac{q-1}{h}\mathrm{E}\bigg\{K(0)\frac{B_5(T)}{B_1(T)f(T)} - \frac{\nu_K}{2}\frac{B_2(T)B_3(T)}{B_1^2(T)f(T)}\bigg\} + O_p(1), \\ \sigma_n^2 &= \frac{q-1}{h}\bigg[\mathrm{E}\bigg\{\frac{B_3(T)B_4(T)+B_5^2(T)}{B_1^2(T)f(T)}\int K^2(t)dt + \frac{B_2^2(T)B_3^2(T)}{2B_1^4(T)f(T)}\int(K*K)^2(t)dt \\ &\qquad -2\frac{B_2(T)B_3(T)B_5(T)}{B_1^3(T)f(T)}\int K(t)K*K(t)dt\bigg] + O_p(1). \end{aligned}$$

**Remark:** For nonparametric hypothesis testing in independent data, Fan et al. (2001) established a nice property called the Wilks phenomenon for the generalized likelihood ratio test, i.e. the asymptotic distribution of the test statistic under the null hypothesis does not depend on the unknown true parameters. Indeed, when the likelihood function is used and correctly specified, this property holds for a wide range of problems. However, for generalized longitudinal data, working covariance matrices, generalized estimating equations and quasi-likelihoods are commonly used, and in many situations these models does not need to be correctly specified. When the variance (covariance) of $\varepsilon$ depends on the mean, and if $\mathbf{V}$, $\mathcal{W}$ and $\Sigma$ are different, $B_1(t)$ - $B_5(t)$ and thus the asymptotic distribution of $\lambda_n(H_0)$ in Theorem 1 depend on the true parameters $\boldsymbol{\beta}_0$ and $\theta_0(t)$. In this case, the Wilks phenomena does not hold in general for the test in (3.4).

The following Corollaries provide special cases of Theorem 1, where the asymptotic distribution of $\lambda_n(H_0)$ does not depend on the true parameters $\boldsymbol{\beta}_0$ and $\theta_0(t)$. Recall that the true covariance matrix $\Sigma$ has the structure as in (2.6) and denote $\Sigma_d = \boldsymbol{S}^2$ as the diagonal variance matrix. We now investigate the situation that the variance function is correctly specified for both estimation and test, but the working correlation for test can be misspecified. In other words, we assume $\mathcal{W} = \Sigma_d$ and $\mathbf{V} = \boldsymbol{S}\mathcal{C}(\boldsymbol{\tau})\boldsymbol{S}$, where $\mathcal{C}(\boldsymbol{\tau})$ is working correlation matrix depending on an unknown parameter vector $\boldsymbol{\tau}$ and not necessarily equals to the true correlation matrix $\mathcal{R}(\boldsymbol{\tau})$. Under this circumstance, one can show $B_j(t) = B_{j\dagger}(t)B_1(t)$ for $j = 2, \ldots, 5$, where

$$\begin{aligned}
B_{2\dagger}(t) &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t], \quad B_{3\dagger}(t) = 1, \\
B_{4\dagger}(t) &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{R}(\boldsymbol{\tau})\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t\right], \\
B_{5\dagger}(t) &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[\{\mathcal{R}(\boldsymbol{\tau})\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}\middle|T_j = t\right].
\end{aligned} \tag{3.6}$$

<u>Corollary</u> **1.** *Let $\mathcal{W} = \Sigma_d$, $\mathbf{V} = \mathcal{S}\mathcal{C}(\boldsymbol{\tau})\mathcal{S}$ and $\Sigma = \mathcal{S}\mathcal{R}(\boldsymbol{\tau})\mathcal{S}$, then under conditions (C.1) - (C.6) and the null hypothesis in (1.2)*

$$\sigma_{n\dagger}^{-1}\{\lambda_n(H_0) - \mu_{n\dagger} - d_{n\dagger}\} \xrightarrow{d} \mathit{Normal}(0,1),$$

*where $d_{n\dagger} = o_p(h^{-1/2})$,*

$$\begin{aligned}
\mu_{n\dagger} &= \frac{q-1}{h}\mathrm{E}\left[\{K(0)B_{5\dagger}(T) - \frac{\nu_K}{2}B_{2\dagger}(T)\}/f(T)\right] + O_p(1), \\
\sigma_{n\dagger}^2 &= \frac{q-1}{h}\mathrm{E}\left\{\frac{B_{4\dagger}(T) + B_{5\dagger}^2(T)}{f(T)}\int K^2(t)dt + \frac{B_{2\dagger}^2(T)}{2f(T)}\int (K*K)^2(t)dt \right. \\
&\qquad \left. -2\frac{B_{2\dagger}(T)B_{5\dagger}(T)}{f(T)}\int K(t)K*K(t)dt\right\} + O_p(1).
\end{aligned}$$

Since $B_{j\dagger}(t)$, $j = 2, \ldots, 5$ do not depend on $\boldsymbol{\beta}_0$ and $\theta_0$, it follows that the asymptotic distribution of $\lambda_n(H_0)$ does not depend on the true value of these parameters when the variance is correctly specified but the correlation is misspecified. However, since the asymptotic distribution of $\lambda_n(H_0)$ depends on the true correlation function $\mathcal{R}(\boldsymbol{\tau})$ which is generally unknown, it is difficult to simulate the asymptotic distribution in Corollary 1.

We now consider another important special case where working independence is assumed for both estimation and test.

<u>Corollary</u> **2.** *Under the setting of Theorem 1, if $\mathbf{V} = \mathcal{W} = \Sigma_d$, the asymptotic distribution $\lambda_n(H_0)$ can be simplified to*

$$\sigma_{n*}^{-1}\{\lambda_n(H_0) - \mu_{n*} - d_{1n*}\} \xrightarrow{d} Normal(0,1),$$

*where $d_{1n*} = o_p(h^{-1/2})$, $\mu_{n*} = (q-1)|\mathcal{T}|h^{-1}\{K(0) - \nu_K/2\}$, $\sigma_{n*}^2 = 2(q-1)|\mathcal{T}|h^{-1}\int\{K(t) - \frac{1}{2}K * K(t)\}^2 dt$, and $|\mathcal{T}|$ is the length of the time domain. This result implies $r_K\lambda_n(H_0) \sim_a \chi^2_{r_K\mu_n^*}$ where*

$$r_K = \frac{K(0) - \nu_K/2}{\int\{K(t) - \frac{1}{2}K * K(t)\}^2 dt}.$$

Corollary 2 implies that, if working independence covariance is used in both estimation and hypothesis testing and if the variance function is correctly specified, the asymptotic distribution of $\lambda_n(H_0)$ does not depend on $\boldsymbol{\beta}_0$, $\theta_0(t)$ and the true correlation structure $\mathcal{R}(\boldsymbol{\tau})$. This Wilks result makes it easy to assess the distribution of $\lambda_n(H_0)$ using bootstrap. In practice, a bootstrap method usually provides a better estimate of the critical value than using the asymptotic null distribution because the asymptotic distribution only describes behavior of the leading term in the test statistic.

### 3.2.1 Power of the generalized quasi-likelihood ratio test

To study the local power of the generalized quasi-likelihood ratio test, we consider a contiguous alternative hypothesis

$$H_{1n} : \theta_k(t) = \theta_0(t) + G_{kn}(t), \quad k = 1, \dots, q, \quad \text{with} \sum_{k=1}^{q} \rho_k G_{kn}(t) = 0, \tag{3.7}$$

where $G_{kn}(t)$ are twice continuously differentiable smooth functions with $\sup_{t\in\mathcal{T}} G_{kn}(t) \to 0$ as $n \to \infty$.

Consider the test statistic in (3.4), and call it $\lambda(H_{1n})$ instead. The following theorem gives the asymptotic distribution of the test statistic under the local alternative (3.7).

<u>Theorem</u> **2.** *Suppose that assumptions (C.1) – (C.6) and the local alternative (3.7) hold, $nh^5 \to 0$, and the functions $G_{kn}(t)$'s are twice continuously differentiable. Denote $\mu_{1n} = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k} E G_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}G_{kn}(\boldsymbol{T}_{k,i})$ and we assume there exists a constant $C_G$ such that*

$$h\mu_{1n} \to C_G < \infty. \tag{3.8}$$

*Then the test statistic has the following limiting distribution*

$$\sigma_{n\ddagger}^{-1}(\lambda_n(H_{1n}) - \mu_n - \mu_{1n}) \xrightarrow{d} N(0,1),$$

*where $\sigma_{n\ddagger}^2 = \sigma_n^2 + d_{2n}$, $d_{2n} = \sum_{k=1}^{q}\sum_{i=1}^{n_k} E G_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Sigma_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}G_{kn}(\boldsymbol{T}_{k,i})$ and $\mu_n$ and $\sigma_n^2$ are as defined in Theorem 1.*

An approximate level-$\alpha$ test based on the setting of Corollary 2 is $\phi_h = I(\lambda_n(H_{1n}) - \mu_{n*} > z_\alpha\sigma_{n*})$, where $z_\alpha$ is the upper $\alpha 100\%$ percentile of $N(0,1)$, and we reject the null hypothesis if $\phi_h = 1$. Let $\Phi(\cdot)$ be the cumulative distribution function of $N(0,1)$, the type Ⅱ error of the test is

$$\beta(\alpha, \boldsymbol{G}_n) = P(\lambda_n(H_{1n}) - \mu_n < z_\alpha \sigma_n) \approx \Phi(\sigma_{n\ddagger}^{-1}\sigma_n z_\alpha - \sigma_{n\ddagger}^{-1}\mu_{1n}), \tag{3.9}$$

where $\boldsymbol{G}_n = (G_{1n}, \ldots, G_{qn})^{\mathrm{T}}$. Define the class of functions

$$\mathcal{G}_n(\varrho) = \{\boldsymbol{G}_n = (G_{1n}, \ldots, G_{qn})^{\mathrm{T}}; \textstyle\sum_{k=1}^{q} \rho_k \mathrm{E} G_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\Delta_k \mathbf{V}_k^{-1}\Delta_k G_{kn}(\boldsymbol{T}_k) \geq \varrho^2\},$$

and the maximum probability of type Ⅱ errors as

$$\beta(\alpha, \varrho) = \sup_{\boldsymbol{G}_n \in \mathcal{G}_n(\varrho)} \beta(\alpha, \boldsymbol{G}_n).$$

Following Ingster (1993) and Fan et al. (2001), define the minimax rate of test as $\varrho_n$ such that

(a) for any $\varrho > \varrho_n$, $\alpha > 0$, and $\beta > 0$, there exists a constant $c$ such that $\beta(\alpha, c\varrho) \leq \beta + o(1)$;

(b) for any sequence $\varrho_n^* = o(\varrho_n)$, there exist $\alpha > 0$ and $\beta > 0$ such that for any $c > 0$ $P(\phi = 1|H_0) = \alpha + o(1)$ and $\liminf_{n\to\infty} \beta(\alpha, c\varrho_n^*) > \beta$.

The following theorem provides the minimax rate for the test procedure.

<u>Theorem</u> **3.** *Under conditions (C.1) - (C.6), the minimax rate of the GQLR test is* $\varrho_n(h) = n^{-4/9}$ *with* $h = c^* n^{2/9}$.

The proof of Theorem 3 is provided in the Appendix 3.4. Theorem 3 shows that the GQLR test based on working independence covariance matrices achieves the minimax optimal rate of Ingster (1993).

## 3.3 Implementation of the Method

### 3.3.1 Estimation algorithm

We now describe the procedure of computing the estimates of model (1.1) under the null hypothesis. We consider an iterative algorithm to solve the profile-kernel estimating equations (2.1) and (2.2).

1. To obtain an initial values of $\boldsymbol{\beta}$ and $\theta(t)$, assume a parametric form for $\theta(t)$ such as $\theta(t) = \alpha_0 + \alpha_1 t$, and fit a generalized linear model $g(\mu_{k,ij}) = \boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta} + \alpha_0 + \alpha_1 T_{k,ij}$.

2. Fix the value of $\boldsymbol{\beta}$ and let $\widehat{\boldsymbol{\alpha}}_{pre}$ be the value of $\boldsymbol{\alpha}$ from the previous iteration, we update $\widehat{\boldsymbol{\alpha}}$ by

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{curr} &= \widehat{\boldsymbol{\alpha}}_{pre} + \left\{ \sum_{k,i} \mathbf{U}_{k,i}(t)^T \Delta_{k,i}(\boldsymbol{X}_{k,i}, t) \mathcal{W}_{k,i}^{-1} \mathbf{K}_h(\boldsymbol{T}_{k,i} - t) \Delta_{k,i}(\boldsymbol{X}_{k,i}, t) \mathbf{U}_{k,i}(t) \right\}^{-1} \\ & \left[ \sum_{k,i} \mathbf{U}_{k,i}(t)^{\mathrm{T}} \Delta_{k,i}(\boldsymbol{X}_{k,i}, t) \mathcal{W}_{k,i}^{-1} \mathbf{K}_h(\boldsymbol{T}_{k,i} - t) \{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t)\} \right],\end{aligned}$$

   where $\boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t) = g^{-1}(\boldsymbol{X}_{k,i}\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}_{pre}(t))$. When $g(\cdot)$ is a logistic link, $\boldsymbol{\mu}_{k,i}(\boldsymbol{X}_{k,i}, t) = \{1 + \exp(-\boldsymbol{X}_{k,i}\boldsymbol{\beta} - \widehat{\boldsymbol{\theta}}_{pre}(t))\}^{-1}$, and the first derivative of $\boldsymbol{\mu}(\cdot)$ is $\boldsymbol{\mu}(\cdot) \times \{1 - \boldsymbol{\mu}(\cdot)\}$.

3. Let $\boldsymbol{\beta}_{pre}$ be the value of $\boldsymbol{\beta}$ from the previous iteration and $\widehat{\theta}_{curr}(t)$ be the updated nonparametric estimator, we updated $\widehat{\boldsymbol{\beta}}$ by

$$\widehat{\boldsymbol{\beta}}_{curr} = \widehat{\boldsymbol{\beta}}_{pre} + \left\{ \sum_{k,i} \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} \Delta_{k,i} \mathcal{W}_{k,i}^{-1} \Delta_{k,i} \widetilde{\boldsymbol{X}}_{k,i} \right\}^{-1} \left[ \sum_{k,i} \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} \Delta_{k,i} \mathcal{W}_{k,i}^{-1} \{\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i}\} \right],$$

   where $\boldsymbol{\mu}_{k,i} = g^{-1}\{\boldsymbol{X}_{k,i}\widehat{\boldsymbol{\beta}}_{pre} + \widehat{\boldsymbol{\theta}}_{curr}(\boldsymbol{T}_{k,i})\}$ and $\widetilde{\boldsymbol{X}}_{k,i} = \boldsymbol{X}_{k,i} - \boldsymbol{\mu}_X(\boldsymbol{T}_{k,i})$. A consistent estimator of $\boldsymbol{\mu}_X(t)$ is

$$\left[ \sum_{k,i} \{\mu_{k,ij}^{(1)}\}^2 \omega_{k,i}^{jj} K_h(T_{k,ij} - t) \boldsymbol{X}_{k,ij} \right] \left[ \sum_{k,i} \{\mu_{k,ij}^{(1)}\}^2 \omega_{k,i}^{jj} K_h(T_{k,ij} - t) \right]^{-1}.$$

4. Iterate between steps 3 and 4 until convergence.

The estimation procedure under the full model is similar to that described above except using only the $k$th treatment group to estimate $\theta_k(t)$, $k = 1, \cdots, q$.

For the special case where the link function is identity, we can build a noniterative algorithm to obtain the estimators directly. Let's first consider estimation under the null hypothesis. For any value $t$, define $\mathbf{U}(t) = (\mathbf{U}_{1,1}(t)^{\mathrm{T}}, \cdots, \mathbf{U}_{q,n_q}(t)^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{Y}$ and $\boldsymbol{X}$ be the vector and matrix by stacking all $\boldsymbol{Y}_{k,i}$'s and $\boldsymbol{X}_{k,i}$'s together, where $\mathbf{U}_{k,i}(t)$, $\boldsymbol{Y}_{k,i}$ and $\boldsymbol{X}_{k,i}$ are the same as those in (2.1). The estimator of $\boldsymbol{\beta}$ can be obtained by:

$$\widehat{\boldsymbol{\beta}}_R = \left(\widetilde{\boldsymbol{X}}^{\mathrm{T}}\mathcal{W}^{-1}\widetilde{\boldsymbol{X}}\right)^{-1}\left(\widetilde{\boldsymbol{X}}^{\mathrm{T}}\mathcal{W}^{-1}\widetilde{\boldsymbol{Y}}\right), \tag{3.10}$$

where $\widetilde{\boldsymbol{X}} = \{I - \mathcal{S}(\boldsymbol{T})\}\boldsymbol{X}$, $\widetilde{\boldsymbol{Y}} = \{I - \mathcal{S}(\boldsymbol{T})\}\boldsymbol{Y}$, $\mathcal{S}(t) = e^{\mathrm{T}}\{\mathbf{U}(t)^{\mathrm{T}}\mathcal{W}^{-1}\mathbf{K}_h(\boldsymbol{T} - t)\mathbf{U}(t)\}^{-1}$ $\mathbf{U}(t)^{\mathrm{T}}\mathcal{W}^{-1}\mathbf{K}_h(\boldsymbol{T} - t)\}$, $e = (1, 0)^{\mathrm{T}}$, $\mathbf{K}_h(\boldsymbol{T} - t) = \mathrm{diag}\{\mathbf{K}_h(\boldsymbol{T}_{1,1} - t), \cdots, \mathbf{K}_h(\boldsymbol{T}_{q,n_q} - t)\}$ and $\mathcal{W}^{-1} = \mathrm{diag}(\mathcal{W}_{1,1}^{-1}, \cdots, \mathcal{W}_{q,n_q}^{-1})$.

The estimator of $\theta(t)$ is given by

$$\widehat{\theta}_R(t) = \mathcal{S}(t)(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

For the full model, the estimators $\widehat{\boldsymbol{\beta}}_F$ and $\widehat{\theta}_{F,k}(t)$ are

$$\widehat{\boldsymbol{\beta}}_F = \left(\sum_k \widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\mathcal{W}_k^{-1}\widetilde{\boldsymbol{X}}_k\right)^{-1}\left(\sum_k \widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\mathcal{W}_k^{-1}\widetilde{\boldsymbol{Y}}_k\right),$$

and

$$\widehat{\theta}_{F,k}(t) = \mathcal{S}_k(t)(\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}),$$

where $\widetilde{\boldsymbol{X}}_k = \{I - \mathcal{S}_k(\boldsymbol{T}_k)\}\boldsymbol{X}_k$, $\widetilde{\boldsymbol{Y}}_k = \{I - \mathcal{S}_k(\boldsymbol{T}_k)\}\boldsymbol{Y}_k$, and $\mathcal{S}_k(t) = e^{\mathrm{T}}\{\mathbf{U}_k(t)^{\mathrm{T}}\mathcal{W}_k^{-1}\mathbf{K}_h(\boldsymbol{T}_k - t)\mathbf{U}_k(t)\}^{-1}\mathbf{U}_k(t)^{\mathrm{T}}\mathcal{W}_k^{-1}\mathbf{K}_h(\boldsymbol{T}_k - t)\}$.

### 3.3.2 Estimation of the covariance function

As pointed out in the remarks on Theorem 1 and Corollary 2, in order for the GQLT procedure to enjoy the much celebrated Wilks property, one need to correctly specify the variance and correlation function. Even if working independence is assumed in both estimation and test, we still need to correctly specify the variance function. In real life, both the variance and the correlation functions are unknown and need to be estimated from the data.

When the error process $\varepsilon(t)$ is heteroscedastic, one popular approach in recent literature on longitudinal data analysis and functional data analysis is to model the variance as a nonparametric function of observation time or the mean value (Yao et al., 2005; Fan et al., 2007; Li 2011). In what follows, we assume the variance is a smooth function of the mean value, denoted as $\sigma^2(\mu)$, and estimate this function using a local linear smoother

$$\widehat{\sigma}^2(\mu) = [1,0]\Big\{\sum_{k,i} \boldsymbol{D}_{k,i}(\mu)^{\mathrm{T}}\mathbf{K}_{k,ih}(\mu)\boldsymbol{D}_{k,i}(\mu)\Big\}^{-1}\Big\{\sum_{k,i} \boldsymbol{D}_{k,i}(\mu)^{\mathrm{T}}\mathbf{K}_{k,ih}(\mu)\widetilde{\boldsymbol{\varepsilon}}^2_{k,i}\Big\}, \tag{3.11}$$

where $\boldsymbol{D}_{k,i}(\mu) = \{\boldsymbol{D}_{k,i1}(\mu), \cdots, \boldsymbol{D}_{k,im_{k,i}}(\mu)\}^{\mathrm{T}}$ with $\boldsymbol{D}_{k,ij}(\mu) = (1, (\widehat{\mu}_{k,ij} - \mu)/h)^{\mathrm{T}}$, $\mathbf{K}_{k,ih}(\mu) = \mathrm{diag}\{K_h(\widehat{\mu}_{k,ij} - \mu)\}_{j=1}^{m_{k,i}}$, and $\widetilde{\boldsymbol{\varepsilon}}_{k,i}$ are the residuals from a pilot estimation of the full model by setting $\mathcal{W}_{k,i}$ to be identity matrices.

Fan et al. (2007) proposed to model the correlation function as a member of a known parametric family, i.e. $\mathrm{corr}\{\varepsilon(s), \varepsilon(t)\} = \rho(s,t,\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is an unknown parameter vector independent of $\boldsymbol{\beta}$ and $\theta_0$. Examples of such correlation families include the AR and ARMA correlations. They also proposed to estimate the correlation parameter vector $\boldsymbol{\tau}$ by maximizing the following quasi-likelihood

$$-\frac{1}{2}\sum_k\sum_i \Big\{\log|\mathcal{C}_{k,i}(\boldsymbol{\tau})| + \widetilde{\boldsymbol{\varepsilon}}^{\mathrm{T}}_{k,i}\widehat{\boldsymbol{S}}^{-1}_{k,i}\mathcal{C}^{-1}_{k,i}(\boldsymbol{\tau})\widehat{\boldsymbol{S}}^{-1}_{k,i}\widetilde{\boldsymbol{\varepsilon}}_{k,i}\Big\}, \tag{3.12}$$

where $\widehat{\boldsymbol{S}}_{k,i} = \text{diag}\{\widehat{\sigma}(\mu_{k,ij})\}_{j=1}^{m_{k,i}}$ and $\mathcal{C}_{k,i}(\boldsymbol{\tau})$ is the correlation matrix for the $(k,i)$th cluster under the correlation function $\rho$.

Li (2011) showed that the local linear variance estimator is uniformly consistent, one can use the estimated variance function to further improve the estimation efficiency, and the refined estimator is as efficient as when the true variance function is known. Using similar arguments, we can show that plugging the estimated variance (and correlation) in the test statistic only incurs an asymptotically negligible error. The asymptotic distribution of $\lambda_n(H_0)$ is the same as if $\sigma^2(\mu)$ is known.

### 3.3.3 Bootstrap procedure for models with an identity link function

As suggested by Fan et al. (2001) and Fan and Jiang (2005), it is preferable to evaluate the null distribution of the test statistic by bootstrap, since the asymptotic distribution only capture the randomness in the leading term of $\lambda_n(H_0)$. For Gaussian longitudinal data, Li (2011) proposed a stratified conditional bootstrap procedure by taking the residuals of the full model and resampling the clusters within each treatment group. Here we describe a stratified conditional bootstrap procedure for models with an identity link function, which is similar to the bootstrap method proposed by Li (2011).

1. Estimate $\boldsymbol{\beta}$ and $\theta_k(t)$ under both the reduced and full models and calculate the GQLR test statistic $\lambda_n(H_0)$.

2. Compute $Y^*_{k,ij} = Y_{k,ij} - \widehat{\theta}_{F,k}(T_{k,ij}) + \widehat{\theta}_R(T_{k,ij})$, draw with replacement $n_k$ subjects from the $k$th group to form a bootstrap sample.

3. Calculate the generalized likelihood ratio test statistic $\lambda_n^*(H_0)$ for the bootstrap sample, using the same bandwidth as for the real data.

4. Repeat step 2 and 3 a large number of times to obtain the bootstrap replicates $\lambda_n^*(H_0)$, and the estimated p-value is the percentage of $\lambda_n^*(H_0)$ that are greater than $\lambda_n(H_0)$.

### 3.3.4 Bootstrap procedure for models with non-identity link functions

As we can see, the bootstrap procedure described in section 3.3.3 is only useful for Gaussian longitudinal data, since the residuals of binary or count data are not binary or count any more. In this section, we proposed a stratified parametric bootstrap procedure for models with non-identity link functions, e.g. logistic models.

1. Fix the bandwidth, obtain estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\theta}_k(t)$ under both the reduced and full semiparametric models from the original data, and estimate the working covariance structure $\mathcal{C}(\boldsymbol{\tau})$ from the full model.

2. Compute the GQLR test statistic $\lambda_n(H_0)$.

3. Draw with replacement $n_k$ subjects from the $k$th group to form a bootstrap sample $\{\boldsymbol{X}_{k,i}^*, \boldsymbol{T}_{k,i}^*\}$, and calculate $\boldsymbol{\mu}_{k,i}^* = g^{-1}\{\boldsymbol{X}_{k,i}^*\widehat{\boldsymbol{\beta}}_R + \widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i}^*)\}$, where $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\boldsymbol{\theta}}_R(\boldsymbol{T}_{k,i})$ are the estimates under the reduced model in step 1. Use the correlation $\mathcal{C}_{k,i}^*(\widehat{\boldsymbol{\tau}})$ and the conditional mean $\boldsymbol{\mu}_{k,i}^*$ to generate the correlated binary data $\boldsymbol{Y}_{k,i}^*$.

4. Calculate the generalized likelihood ratio test statistic $\lambda_n^*(H_0)$ from the bootstrap sample $\{Y_{k,i}^*, \boldsymbol{X}_{k,i}^*, \boldsymbol{T}_{k,i}^*\}$.

5. Repeat steps 3 and 4 a large number of times to obtain the bootstrap replicates $\lambda_n^*(H_0)$, the estimated p-value is the percentage of $\lambda_n^*(H_0)$ that are greater than $\lambda_n(H_0)$.

Generating a non-Gaussian random vector is often challenging. There are various methods based on marginal models, which require only specification of the marginal means and within-subject correlation structures. For example, Emrich and Piedmonte (1991) proposed a method based on the multivariate probit model in which they generate correlated standard normal variables and then dichotomize each coordinate. Park, Park and Shin (1996) proposed a simple algorithm for generating non-negatively correlated binary variables based on the sums of correlated Poisson variables. However, most existing methods are subject to general restrictions on $\{\mu, \mathcal{R}(\boldsymbol{\tau})\}$(e.g. the correlation should be positive definite; the correlation $\rho_{jk}$ must satisfy some pairwise bounds and is not free over [-1, 1].) These restrictions could be violated during the process of generating the response vectors for different clusters. It is also a time consuming process to generate binary data with large cluster size. Therefore, conducting the GQLR test by using the bootstrap method may leave users a lot of computational issues. One alternative approach is to assume working independence for both estimation and test as described in Corollary 2. By setting $\mathcal{C}_{k,i}^* = I$ for all $k$ and $i$, we can generate $Y_{k,ij}$ as if they are independent, and Corollary 2 insures that the asymptotic distribution of $\lambda_n(H_0)$ based on independent bootstrap samples is the equivalent to the one based on the real data.

## 3.4 Appendix: Technical Proofs

### 3.4.1 Proof of null distribution

For any $m$-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the first two partial derivatives of $\mathcal{Q}\{g^{-1}(\boldsymbol{x}),\boldsymbol{y}\}$ regarding $\boldsymbol{x}$ are

$$\frac{\partial\mathcal{Q}}{\partial\boldsymbol{x}}\{g^{-1}(\boldsymbol{x}),\boldsymbol{y}\} = \Delta(\boldsymbol{x})\mathbf{V}^{-1}\{g^{-1}(\boldsymbol{x})\}\{\boldsymbol{y}-g^{-1}(\boldsymbol{x})\},$$
$$\frac{\partial^2\mathcal{Q}}{\partial\boldsymbol{x}\partial\boldsymbol{x}^{\mathrm{T}}}\{g^{-1}(\boldsymbol{x}),\boldsymbol{y}\} = -\Delta(\boldsymbol{x})\mathbf{V}^{-1}\{g^{-1}(\boldsymbol{x})\}\Delta(\boldsymbol{x}) + \sum_{j=1}^{m}\{y_j - g^{-1}(x_j)\}\mathcal{D}_j(\boldsymbol{x}),$$

where $\Delta(\boldsymbol{x}) = \mathrm{diag}\{\frac{dg^{-1}}{dx}(x_j)\}_{j=1}^{m}$, $\mathcal{D}_j = \partial(\mathbf{V}^{j\bullet}\Delta)/\partial\boldsymbol{x}$, and $\mathbf{V}^{j\bullet}$ is the $j$-th row of $\mathbf{V}^{-1}$. Denote $\boldsymbol{\eta}_{0k,i} = \boldsymbol{X}_{k,i}\boldsymbol{\beta}_0 + \boldsymbol{\theta}_0(\boldsymbol{T}_{k,i})$, $\boldsymbol{\mu}_{0k,i} = g(\boldsymbol{\eta}_{0k,i})$ and $\boldsymbol{\epsilon}_{k,i} = \boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{0k,i}$. By taking a Taylor's expansion at $\boldsymbol{\eta}_{0k,i}$, we have

$$\begin{aligned}\mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\widehat{\boldsymbol{\beta}} + \widehat{\theta}(\boldsymbol{T}_{k,i})\},\boldsymbol{Y}_{k,i}] &= \mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta}_0 + \theta_0(\boldsymbol{T}_{k,i})\},\boldsymbol{Y}_{k,i}]\\
&+\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}\{\boldsymbol{X}_{k,i}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \widehat{\theta}(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\}\\
&+\frac{1}{2}\{\boldsymbol{X}_{k,i}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \widehat{\theta}(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\}^{\mathrm{T}}\{\sum_{j=1}^{m}\epsilon_{k,ij}\mathcal{D}_{k,ij} - \Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}\}\\
&\times\{\boldsymbol{X}_{k,i}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \widehat{\theta}(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\} + O\{(n^{-1/2} + h^2 + n^{-1/2}h^{-1/2})^3\}.\end{aligned}$$

For any vector $\boldsymbol{a}$ and a symmetric matrix $\boldsymbol{A}$, define $\|\boldsymbol{a}\|_{\boldsymbol{A}}^2 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{a}$. Denote $\boldsymbol{X}_k = (\boldsymbol{X}_{k,1}^{\mathrm{T}},\ldots,\boldsymbol{X}_{k,n_k}^{\mathrm{T}})^{\mathrm{T}}$, $\Delta_k = \mathrm{diag}(\Delta_{k,1},\ldots,\Delta_{k,n_k})$, and $\boldsymbol{\epsilon}_k = (\boldsymbol{\epsilon}_{k,1}^{\mathrm{T}},\ldots,\boldsymbol{\epsilon}_{k,n_k}^{\mathrm{T}})^{\mathrm{T}}$. By straight forward calculations,

$$\lambda_n(H_0) = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + o_p(1), \tag{3.13}$$

where

$$J_1 = \sum_k \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\Delta_k\{\widehat{\theta}_{F,k}(\boldsymbol{T}_k) - \widehat{\theta}_R(\boldsymbol{T}_k)\},\quad J_2 = \sum_k \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\Delta_k\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_F - \widehat{\boldsymbol{\beta}}_R),$$
$$J_3 = \sum_k(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)^{\mathrm{T}}\boldsymbol{X}_k^{\mathrm{T}}\Delta_k\mathbf{V}_K^{-1}\Delta_k\{\widehat{\theta}_R(\boldsymbol{T}_k) - \theta_0(\boldsymbol{T}_k)\}$$

$$-(\widehat{\boldsymbol\beta}_F-\boldsymbol\beta_0)^{\rm T}\boldsymbol X_k^{\rm T}\Delta_k\mathbf V_K^{-1}\Delta_k\{\widehat\theta_{F,k}(\boldsymbol T_k)-\theta_0(\boldsymbol T_k)\},$$

$$J_4=\frac{1}{2}\sum_k\|\widehat\theta_R(\boldsymbol T_k)-\theta_0(\boldsymbol T_k)\|^2_{\Delta_k\mathbf V_K^{-1}\Delta_k}-\|\widehat\theta_{F,k}(\boldsymbol T_k)-\theta_0(\boldsymbol T_k)\|^2_{\Delta_k\mathbf V_K^{-1}\Delta_k},$$

$$J_5=\frac{1}{2}\sum_k\|\widehat{\boldsymbol\beta}_c/R-\boldsymbol\beta_0\|^2_{\boldsymbol X_k^{\rm T}\Delta_k\mathbf V_k^{-1}\Delta_k\boldsymbol X_k}-\|\widehat{\boldsymbol\beta}_F-\boldsymbol\beta_0\|^2_{\boldsymbol X_k^{\rm T}\Delta_k\mathbf V_k^{-1}\Delta_k\boldsymbol X_k},$$

$$J_6=\frac{1}{2}\sum_k\sum_{i=1}^{n_k}\|\boldsymbol X_{k,i}(\widehat{\boldsymbol\beta}_F-\boldsymbol\beta_0)+\widehat\theta_{F,k}(\boldsymbol T_{k,i})-\theta_0(\boldsymbol T_{k,i})\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}$$
$$-\|\boldsymbol X_{k,i}(\widehat{\boldsymbol\beta}_R-\boldsymbol\beta_0)+\widehat\theta_R(\boldsymbol T_{k,i})-\theta_0(\boldsymbol T_{k,i})\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}.$$

By Lemma 4, $d_{1n}=J_2+J_3+J_5+J_6=o_p(h^{-1/2})$. Theorem follows the asymptotic distribution of $J_1+J_4$ given in Lemma 5.

Lemma **4.** *Under the null hypothesis in (1.2) and all assumptions in Theorem 1, $J_2=o_p(1)$, $J_3=o_p(h^{-1/2})$, $J_5=o_p(1)$ and $J_6=O_p(n^{1/2}h^4+n^{-1}h^{-2})$.*

**Proof:** (i) By the asymptotic expansions in Propositions 1 and 2, $(\widehat{\boldsymbol\beta}_F-\widehat{\boldsymbol\beta}_R)=o_p(n^{-1/2})$. Therefore $J_2=(\sum_k\boldsymbol\epsilon_k^{\rm T}\mathbf V_k^{-1}\Delta_k\boldsymbol X_k)(\widehat{\boldsymbol\beta}_F-\widehat{\boldsymbol\beta}_R)=O_p(n^{1/2})\times o_p(n^{-1/2})=o_p(1)$. Similarly,

$$J_5=\sum_k(\widehat{\boldsymbol\beta}_R-\boldsymbol\beta_0)\boldsymbol X_k^{\rm T}\Delta_k\mathbf V_k^{-1}\Delta_k\boldsymbol X_k(\widehat{\boldsymbol\beta}_R-\widehat{\boldsymbol\beta}_F)+(\widehat{\boldsymbol\beta}_R-\widehat{\boldsymbol\beta}_F)\boldsymbol X_k^{\rm T}\Delta_k\mathbf V_k^{-1}\Delta_k\boldsymbol X_k(\widehat{\boldsymbol\beta}_F-\boldsymbol\beta_0)=o_p(1).$$

(ii) Next, we derive the order for $J_3$. Using similar arguments as in page 1054 of Lin and Carroll (2001), the first term in $J_3$ is

$$(\widehat{\boldsymbol\beta}_R-\boldsymbol\beta_0)^{\rm T}\sum_{k=1}^q\sum_{i=1}^{n_k}\boldsymbol X_{k,i}^{\rm T}\Delta_{k,i}\mathbf V_{k,i}^{-1}\Delta_{k,i}\{\widehat\theta_R(\boldsymbol T_{k,i})-\theta_0(\boldsymbol T_{k,i})\}=O_p(1+n^{1/2}h^2)=o_p(h^{-1/2}).$$

Similarly, the second term and hence $J_3$ itself are of order $o_p(h^{-1/2})$.

(iii) We decompose $J_6$ into three parts,

$$J_{61}=\frac{1}{2}\sum_k\sum_{i=1}^{n_k}\|\boldsymbol X_{k,i}(\widehat{\boldsymbol\beta}_F-\boldsymbol\beta_0)\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}-\|\boldsymbol X_{k,i}(\widehat{\boldsymbol\beta}_R-\boldsymbol\beta_0)\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}$$

$$J_{62}=\frac{1}{2}\sum_k\sum_{i=1}^{n_k}\|\widehat\theta_{F,k}(\boldsymbol T_{k,i})-\theta_0(\boldsymbol T_{k,i})\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}-\|\widehat\theta_R(\boldsymbol T_{k,i})-\theta_0(\boldsymbol T_{k,i})\|^2_{\sum_{j=1}^m\epsilon_{k,ij}\mathcal D_{k,ij}}$$

$$J_{63} = \sum_k \sum_{i=1}^{n_k} (\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0)^{\mathrm{T}} \boldsymbol{X}_{k,i}^{\mathrm{T}} (\sum_{j=1}^{m} \epsilon_{k,ij} \mathcal{D}_{k,ij}) \{\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\}$$
$$- (\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)^{\mathrm{T}} \boldsymbol{X}_{k,i}^{\mathrm{T}} (\sum_{j=1}^{m} \epsilon_{k,ij} \mathcal{D}_{k,ij}) \{\widehat{\theta}_R(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\}$$

It can easily show that $J_{61} = O_p(n^{-1/2})$. Now, we need to proof that $J_{62} = O_p(n^{1/2}h^4 + h + n^{-1}h^{-2})$. Recall that by Proposition 1, $\widehat{\theta}_R(t) - \theta_0(t) = O_p(h^2 + n^{-1/2}h^{-1/2})$. Under the null hypothesis, Proposition 2 provides the same convergence rate for $\widehat{\theta}_{F,k}(t) - \theta_0(t)$. Since the error $\boldsymbol{\epsilon}_{k,i}$ is independent in our model, the correlation $\mathrm{corr}(\varepsilon_{k,ij}, \varepsilon_{k',i'j'}) \neq 0$ only if $k = k'$ and $i = i'$, by tedious calculation, the first part of $J_{62}$ can be written as

$$\begin{aligned}
&\sum_k \sum_{i=1} \|\widehat{\theta}_R(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}} \\
&= \sum_{k,i} \| \frac{\theta_0^{(2)}(\boldsymbol{T}_{k,i})h^2}{2} + \mathcal{U}_R(\boldsymbol{T}_{k,i}) - \boldsymbol{\mu}_{0X}^{\mathrm{T}}(\boldsymbol{T}_{k,i})(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}} + o_p(1) \\
&= O_p(n^{1/2}h^4 + n^{-1}h^{-2})
\end{aligned}$$

The second term and $J_{62}$ itself are of order $O_p(n^{1/2}h^4 + n^{-1}h^{-2})$. By using the similar argument for $J_3$, we find $J_{63} = O_p(h^2 + n^{-1/2}h^{-1})$. Combining all three parts, we have $J_6 = O_p(n^{1/2}h^4 + n^{-1}h^{-2}) = o_p(1)$ by Condition (C.6) . $\diamondsuit$

We now consider the asymptotic null distribution of the generalized quasi-likelihood test statistics. The following notation will be used in the proof of the lemmas and theorems. Denote $\sum_{k=1}^q \sum_{i=1}^{n_k}$ as $\sum_{k,i}$, and let $\nu^{jl}$ and $\sigma_{jl}$ be the $(j,l)$th elements of $\mathbf{V}^{-1}$ and $\Sigma$.

<u>Lemma</u> **5.** *Suppose all assumptions in Theorem 1 hold, then*

$$\sigma_n^{-1}(J_1 + J_4 - \mu_n) \xrightarrow{d} Normal(0,1),$$

*where $\mu_n$ and $\sigma_n^2$ are defined as in Theorem 1.*

**Proof:** It is easy to see that

$$J_1 \;=\; \sum_k \sum_i \sum_j \sum_\ell \varepsilon_{k,ij} \nu_{k,i}^{j\ell} \mu_{k,i\ell}^{(1)} \{\widehat{\theta}_{F,k}(T_{k,i\ell}) - \widehat{\theta}_R(T_{k,i\ell})\}.$$

Using the asymptotic expansions of $\widehat{\theta}_R(t)$ and $\widehat{\theta}_{F,k}(t)$ in Propositions 1 and 2, we have $J_1 = (R_1 + R_2 + R_3) \times \{1 + o_p(1)\}$ where

$$R_1 = \sum_{k=1}^{q} \sum_{i=1}^{n_k} \sum_{j,j',\ell=1}^{m} (\frac{1}{n_k} - \frac{1}{n}) \varepsilon_{k,ij} \varepsilon_{k,ij'} \nu_{k,i}^{jl} \mu_{k,il}^{(1)} \frac{\mu_{k,ij'}^{(1)} \omega_{k,i}^{j'j'}}{mB_1(T_{k,il})} K_h(T_{k,ij'} - T_{k,il}),$$

$$R_2 = \sum_{k=1}^{q} \sum_{i=1}^{n_k} \sum_{i' \neq i} \sum_{j,j',\ell=1}^{m} (\frac{1}{n_k} - \frac{1}{n}) \varepsilon_{k,ij} \varepsilon_{k,i'j'} \nu_{k,i}^{jl} \mu_{k,il}^{(1)} \frac{\mu_{k,i'j'}^{(1)} \omega_{k,i'}^{j'j'}}{mB_1(T_{k,il})} K_h(T_{k,i'j'} - T_{k,il}),$$

$$R_3 = -\sum_{k} \sum_{k' \neq k} \sum_{i=1}^{n_k} \sum_{i'=1}^{n_{k'}} \sum_{j,j',\ell=1}^{m} \frac{1}{nm} \varepsilon_{k,ij} \varepsilon_{k',i'j'} \nu_{k,i}^{jl} \mu_{k,il}^{(1)} \frac{\mu_{k',i'j'}^{(1)} \omega_{k',i'}^{j'j'}}{B_1(T_{k,il})} K_h(T_{k',i'j'} - T_{k,il}).$$

By straightforward calculation,

$$\begin{aligned} R_1 &= \sum_k (1-\rho_k) \mathrm{E}\Big\{ \sum_{j,j',\ell=1}^{m} \sigma_{jj'} \nu_{k,i}^{jl} \mu_{k,il}^{(1)} \frac{\mu_{k,ij'}^{(1)} \omega_{k,i}^{j'j'}}{mB_1(T_{k,il})} K_h(T_{k,ij'} - T_{k,il}) \Big\} \times \{1 + O_p(n^{-1/2})\} \\ &= \frac{q-1}{mh} K(0) \sum_{j,l} E\{\sigma_{jl} \nu^{jl} \Delta_{ll}^2 \omega^{ll} B_1^{-1}(T_l)\} + O_p(1). \end{aligned}$$

It can also easily to see that the terms $R_2$ and $R_3$ have mean zero and only contribute to the variance.

By similar calculations,

$$\begin{aligned} J_4 &= \frac{1}{2} \sum_k \sum_i \Big\{ (\widehat{\theta}_R - \theta_0)^{\mathrm{T}} (\boldsymbol{T}_{k,i}) \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} (\widehat{\theta}_R - \theta_0)(\boldsymbol{T}_{k,i}) \\ &\qquad - (\widehat{\theta}_{F,k} - \widehat{\theta}_0)^{\mathrm{T}} (\boldsymbol{T}_{k,i}) \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} (\widehat{\theta}_{F,k} - \theta_0)(\boldsymbol{T}_{k,i}) \Big\} \\ &= \frac{1}{2} \sum_k \sum_i \sum_{j,l}^{m} \mu_{k,ij}^{(1)} \nu_{k,i}^{jl} \mu_{k,il}^{(1)} \Big\{ \frac{\theta_0^{(2)}(T_{k,ij}) h^2}{2} - \boldsymbol{\mu}_X^{\mathrm{T}}(T_{k,ij})(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \\ &\qquad + \frac{1}{nmB_1(T_{k,ij})} \sum_{k'}^{q} \sum_{i'}^{n_{k'}} \sum_{j'}^{m} \mu_{k',i'j'}^{(1)} \omega_{k',i'}^{j'j'} K_h(T_{k',i'j'} - T_{k,ij}) \varepsilon_{k',i'j'} \Big\} \\ &\qquad \times \Big\{ \frac{\theta_0^{(2)}(T_{k,il}) h^2}{2} - \boldsymbol{\mu}_X^{\mathrm{T}}(T_{k,ij})(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \end{aligned}$$

$$+\frac{1}{nmB_1(T_{k,il})}\sum_{k_1}^{q}\sum_{i_1}^{n_{k_1}}\sum_{j_1}^{m}\mu_{k_1,i_1j_1}^{(1)}\omega_{k_1,i_1}^{j_1j_1}K_h(T_{k_1,i_1j_1}-T_{k,il})\varepsilon_{k_1,i_1j_1}\Big\}$$

$$-\frac{1}{2}\sum_k\sum_i\sum_{j,l}^{m}\mu_{k,ij}^{(1)}\nu_{k,i}^{jl}\mu_{k,il}^{(1)}\Big\{\frac{\theta_0^{(2)}(T_{k,ij})h^2}{2}-\boldsymbol{\mu}_X^{\mathrm{T}}(T_{k,ij})(\widehat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)$$

$$+\frac{1}{n_kmB_1(T_{k,ij})}\sum_{i'}^{n_k}\sum_{j'}^{m}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}K_h(T_{k,i'j'}-T_{k,ij})\varepsilon_{k,i'j'}\Big\}$$

$$\times\Big\{\frac{\theta_0^{(2)}(T_{k,il})h^2}{2}-\boldsymbol{\mu}_X^{\mathrm{T}}(T_{k,ij})(\widehat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)$$

$$+\frac{1}{n_kmB_1(T_{k,il})}\sum_{i_1}^{n_k}\sum_{j_1}^{m}\mu_{k,i_1j_1}^{(1)}\omega_{k,i_1}^{j_1j_1}K_h(T_{k,i_1j_1}-T_{k,il})\varepsilon_{k,i_1j_1}\Big\}+o_p(h^{-1/2}).$$

A more detailed calculation shows that $J_4=R_4+R_5+R_6+o_p(h^{-1/2})$ with

$$R_4 = \frac{1}{2}\sum_k\sum_i\sum_{j,j'}^{m}\varepsilon_{k,ij}\varepsilon_{k,ij'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,ij'}^{(1)}\omega_{k,i}^{j'j'}\Big[\sum_{k_1}\Big\{\frac{1}{n^2}I(k_1\neq k)+(\frac{1}{n^2}-\frac{1}{n_k^2})I(k_1=k)\Big\}$$
$$\sum_{i_1}\sum_{j_1,l}^{m}\frac{\mu_{k_1,i_1j_1}^{(1)}\nu_{k_1,i_1}^{j_1l}\mu_{k_1,i_1l}^{(1)}}{m^2B_1(T_{k_1,i_1j_1})B_1(T_{k_1,i_1l})}K_h(T_{k,ij}-T_{k_1,i_1j_1})K_h(T_{k,ij'}-T_{k_1,i_1l})\Big]$$

$$R_5 = \frac{1}{2}\sum_k\sum_i\sum_{i'\neq i}\sum_{j,j'}^{m}\varepsilon_{k,ij}\varepsilon_{k,i'j'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}\sum_{k_1}\Big\{I(k_1=k)(\frac{1}{n^2}-\frac{1}{n_k^2})$$
$$+I(k_1\neq k)\frac{1}{n^2}\Big\}\sum_{i_1}\sum_{j_1,l}^{m}\frac{1}{m^2}\frac{\mu_{k_1,i_1j_1}^{(1)}\nu_{k_1,i_1}^{j_1l}\mu_{k_1,i_1l}^{(1)}}{B_1(T_{k_1,i_1j_1})B_1(T_{k_1,i_1l})}$$
$$\times K_h(T_{k,ij}-T_{k_1,i_1j_1})K_h(T_{k,i'j'}-T_{k_1,i_1l})$$

$$R_6 = \frac{1}{2n^2}\sum_k\sum_{k'\neq k}\sum_{i,i'}\sum_{j,j'}^{m}\varepsilon_{k,ij}\varepsilon_{k',i'j'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}\sum_{k_1}\sum_{i_1}\sum_{j_1,l}^{m}\frac{1}{m^2}\frac{\mu_{k_1,i_1j_1}^{(1)}\nu_{k_1,i_1}^{j_1l}\mu_{k_1,i_1l}^{(1)}}{B_1(T_{k_1,i_1j_1})B_1(T_{k_1,i_1l})}$$
$$\times\ K_h(T_{k,ij}-T_{k_1,i_1j_1})K_h(T_{k',i'j'}-T_{k_1,i_1l})$$

Next, we need to simplify $R_4$ - $R_6$ further. First consider $R_4$, we use the following decomposition $R_4=R_{41}+R_{42}$ with

$$R_{41} = \sum_{k,i}\sum_{j,j'}^{m}\frac{n-n_k}{2n^2}\varepsilon_{k,ij}\varepsilon_{k,ij'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,ij'}^{(1)}\omega_{k,i}^{j'j'}\Big\{\frac{1}{mh}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})}K*K(\frac{T_{k,ij}-T_{k,ij'}}{h})$$
$$+\frac{1}{m^2}\mathrm{E}(\sum_{j_1}^{m}\sum_{l\neq j_1}\Delta_{j_1j_1}\nu^{j_1l}\Delta_{ll}|T_{j_1}=T_{k,ij},T_l=T_{k,ij'})$$

$$\times \frac{f(T_{k,ij})f(T_{k,ij'})}{B_1(T_{k,ij})B_1(T_{k,ij'})}\Big\} \times \{1+o_p(1)\}$$

$$= \frac{1}{2h}\sum_k(\rho_k-\rho_k^2)\frac{1}{m}\mathrm{E}\Big\{\sum_j^m \sigma_{jj}\Delta_{jj}^2(\omega^{jj})^2\frac{B_2(T_j)}{B_1^2(T_j)}\Big\}\int K^2(t)dt + O_p(1)$$

$$= \frac{1}{2h}\sum_k(\rho_k-\rho_k^2)\nu_K\mathrm{E}\{B_3(T)B_2(T)B_1^{-2}(T)f^{-1}(T)\} + O_p(1) \tag{3.14}$$

and

$$\begin{aligned} R_{42} &= \sum_{k,i,j,j'}\frac{1}{2m^2}(\frac{1}{n_k^2}-\frac{1}{n^2})\varepsilon_{k,ij}\varepsilon_{k,ij'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,ij'}^{(1)}\omega_{k,i}^{j'j'}\Big\{\sum_{i_1\neq i}\sum_{j_1,l}\frac{\mu_{k,i_1j_1}^{(1)}\nu_{k,i_1}^{j_1l}\mu_{k,i_1l}^{(1)}}{B_1(T_{k,i_1j_1})B_1(T_{k,i_1l})} \\ &\quad\times K_h(T_{k,ij}-T_{k,i_1j_1})K_h(T_{k,ij'}-T_{k,i_1l}) + \sum_{j_1,l}\frac{\mu_{k,ij_1}^{(1)}\nu_{k,i}^{j_1l}\mu_{k,il}^{(1)}}{B_1(T_{k,ij_1})B_1(T_{k,il})} \\ &\quad\times K_h(T_{k,ij}-T_{k,ij_1})K_h(T_{k,ij'}-T_{k,il})\Big\} \\ &= \frac{1}{2h}\sum_k(1-\rho_k^2)\nu_K\mathrm{E}\{B_3(T)B_2(T)B_1^{-2}(T)f^{-1}(T)\} + O_p(1+\frac{1}{nh^2}) \end{aligned} \tag{3.15}$$

Combining (3.14) and (3.15), we have

$$R_4 = R_{41}+R_{42} = \frac{1-q}{2h}\nu_K\mathrm{E}\{B_3(T)B_2(T)B_1^{-2}(T)f^{-1}(T)\} + O_p(1).$$

Similarly, the second term $R_5$ can be decomposed into $R_{51}$ and $R_{52}$ with

$$\begin{aligned} R_{51} &= \sum_{k,i}\sum_{i'\neq i}\sum_{j,j'}(\frac{1}{n^2}-\frac{1}{n_k^2})\varepsilon_{k,ij}\varepsilon_{k,i'j'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}\Big\{\sum_{i_1\neq i,i'}\sum_{j_1,l}\frac{1}{2m^2}\frac{\mu_{k,i_1j_1}^{(1)}\nu_{k,i_1}^{j_1l}\mu_{k,i_1l}^{(1)}}{B_1(T_{k,i_1j_1})B_1(T_{k,i_1l})} \\ &\quad\times K_h(T_{k,ij}-T_{k,i_1j_1})K_h(T_{k,i'j'}-T_{k,i_1l}) + \sum_{j_1,l}\frac{1}{m^2}\frac{\mu_{k,ij_1}^{(1)}\nu_{k,i}^{j_1l}\mu_{k,il}^{(1)}}{B_1(T_{k,ij_1})B_1(T_{k,il})} \\ &\quad\times K_h(T_{k,ij}-T_{k,ij_1})K_h(T_{k,i'j'}-T_{k,il})\Big\} \\ &= \frac{1}{2mh}\sum_{k,i}n_k(\frac{1}{n^2}-\frac{1}{n_k^2})\sum_{i'\neq i}\sum_{j,j'}\varepsilon_{k,ij}\varepsilon_{k,i'j'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})} \\ &\quad\times K*K(\frac{T_{k,ij}-T_{k,i'j'}}{h}) + O_p(1+\frac{1}{nh^{3/2}}) \end{aligned} \tag{3.16}$$

$$R_{52} = \sum_{k,i}\frac{n-n_k}{2n^2}\sum_{i'\neq i}\sum_{j,j'}^m \varepsilon_{k,ij}\varepsilon_{k,i'j'}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}\Big\{\frac{B_2(T_{k,ij})}{mhB_1(T_{k,ij})}K*K(\frac{T_{k,ij}-T_{k,i'j'}}{h})$$

$$+\frac{f(T_{k,ij})f(T_{k,i'j'})}{m^2 B_1(T_{k,ij})B_1(T_{k,i'j'})}\mathrm{E}(\sum_{j_1}^{m}\sum_{l\neq j_1}\Delta_{j_1j_1}\nu^{j_1l}\Delta_{ll}|T_{j_1}=T_{k,ij},T_l=T_{k,i'j'})\Big\}$$
$$\times\{1+o_p(1)\} \quad (3.17)$$

Combining (3.16) and (3.17) together, we get

$$R_5=\frac{1}{2mh}\sum_k(\rho_k-1)\frac{1}{n_k}\sum_i\sum_{i'\neq i}\sum_{j,j'}^{m}\varepsilon_{k,ij}\varepsilon_{k,i'j'}\mu^{(1)}_{k,ij}\omega^{jj}_{k,i}\mu^{(1)}_{k,i'j'}\omega^{j'j'}_{k,i'}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})}$$
$$\times K*K(\frac{T_{k,ij}-T_{k,i'j'}}{h})+O_p(1)$$

Finally, for the third term $R_6$,

$$\begin{aligned}
R_6 &= \frac{1}{2n^2}\sum_k\sum_{k'\neq k}\sum_{i,i'}\sum_{j,j'}^{m}\varepsilon_{k,ij}\varepsilon_{k',i'j'}\mu^{(1)}_{k,ij}\omega^{jj}_{k,i}\mu^{(1)}_{k',i'j'}\omega^{j'j'}_{k',i'}\sum_{k_1}\Big\{I(k_1\neq k\ \&\ k_1\neq k')\\
&\quad +I(k_1=k)+I(k_1=k')\Big\}\sum_{i_1}\sum_{j_1,l}^{m}\frac{1}{m^2}\frac{\mu^{(1)}_{k_1,i_1j_1}\nu^{j_1l}_{k_1,i_1}\mu^{(1)}_{k_1,i_1l}}{B_1(T_{k_1,i_1j_1})B_1(T_{k_1,i_1l})}\\
&\quad \times K_h(T_{k,ij}-T_{k_1,i_1j_1})K_h(T_{k',i'j'}-T_{k_1,i_1l})\\
&= \frac{1}{2nmh}\sum_{k=1}^{q}\sum_{k'\neq k}\sum_{i=1}^{n_k}\sum_{i'=1}^{n_{k'}}\sum_{j,j'}\varepsilon_{k,ij}\varepsilon_{k',i'j'}\mu^{(1)}_{k,ij}\omega^{jj}_{k,i}\mu^{(1)}_{k',i'j'}\omega^{j'j'}_{k',i'}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})}\\
&\quad \times K*K(\frac{T_{k,ij}-T_{k',i'j'}}{h})+O_p(1)
\end{aligned}$$

where $B_1(t)$ is as defined in Condition (C.4) and $B_2(t)$ and $B_3(t)$ are defined in (3.5).

Similar as the decomposition of $J_1$, we find that $R_4$ is the leading term in the mean of $J_4$, and the terms $R_5$ and $R_6$ have mean zero and only contribute to the variance of $J_4$. We now collect the mean components in $J_1+J_4$ as

$$\begin{aligned}
\mu_n=R_1+R_4 &= \frac{q-1}{mh}\mathrm{E}\Big\{K(0)\sum_j\sum_l\sigma_{jl}\nu^{jl}\Delta^2_{ll}\omega^{ll}B_1^{-1}(T_l)\\
&\quad -\frac{\nu_K}{2}\sum_{j=1}^{m}\Delta^2_{jj}(\omega^{jj})^2\sigma_{jj}B_2(T_j)B_1^{-2}(T_j)\Big\}+O_p(1)\\
&= \frac{q-1}{h}\mathrm{E}\Big\{K(0)\frac{B_5(T)}{B_1(T)f(T)}-\frac{\nu_K}{2}\frac{B_2(T)B_3(T)}{B_1^2(T)f(T)}\Big\}+O_p(1). \quad (3.18)
\end{aligned}$$

Next, we collect the leading terms that contribute to the variance as $R_2 + R_3 + R_5 + R_6 = W_n + O_p(1)$ where

$$\begin{aligned} W_n &= \sum_{k=1}^{q}(1-\rho_k)\frac{1}{n_k m}\sum_{i}\sum_{i'\neq i}\sum_{j}\sum_{j'}\varepsilon_{k,ij}\varepsilon_{k,i'j'}\Big\{\sum_{l}\frac{\nu_{k,i}^{jl}\mu_{k,il}^{(1)}}{B_1(T_{k,il})}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}K_h(T_{k,i'j'}-T_{k,il}) \\ &\qquad -\frac{1}{2h}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k,i'j'}^{(1)}\omega_{k,i'}^{j'j'}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})}K*K(\frac{T_{k,ij}-T_{k,i'j'}}{h})\Big\} \\ &\quad -\sum_{k}\sum_{k'\neq k}\frac{1}{nm}\sum_{i}\sum_{i'}\sum_{j}\sum_{j'}\varepsilon_{k,ij}\varepsilon_{k',i'j'}\Big\{\sum_{l}\frac{\nu_{k,i}^{jl}\mu_{k,il}^{(1)}}{B_1(T_{k,il})}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}K_h(T_{k',i'j'}-T_{k,il}) \\ &\qquad -\frac{1}{2}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}\frac{B_2(T_{k,ij})}{B_1^2(T_{k,ij})}K*K(\frac{T_{k,ij}-T_{k',i'j'}}{h})\Big\}. \end{aligned}$$

We can see that

$$\begin{aligned} \mathrm{var}(W_n) &= \Big\{\sum_{k_1}(1-\rho_{k_1})^2V_w + \sum_{k_1}\sum_{k_2\neq k_1}\rho_{k_1}\rho_{k_2}V_w\Big\}\times\{1+o(1)\} \\ &= (q-1)V_w\times\{1+o(1)\}, \end{aligned}$$

where

$$\begin{aligned} V_w &= \mathrm{E}\Big[\frac{1}{m^2}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\sum_{j_3=1}^{m}\sum_{j_4=1}^{m}\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\varepsilon_{i_1j_3}\varepsilon_{i_2j_4}\Big\{\sum_{\ell_1=1}^{m}\nu_{i_1}^{j_1l_1}\mu_{i_1l_1}^{(1)}\frac{\mu_{i_2j_2}^{(1)}\omega_{i_2}^{j_2j_2}}{B_1(T_{i_1l_1})}K_h(T_{i_2j_2}-T_{i_1l_1}) \\ &\qquad -\frac{1}{2}\mu_{i_1j_1}^{(1)}\omega_{i_1}^{j_1j_1}\mu_{i_2j_2}^{(1)}\omega_{i_2}^{j_2j_2}\frac{B_2(T_{i_2j_2})}{B_1^2(T_{i_2j_2})}K_h*K_h(T_{i_1j_1}-T_{i_2j_2})\Big\} \\ &\quad \times\Big\{\sum_{\ell_2=1}^{m}\nu_{i_1}^{j_3l_2}\mu_{i_1l_2}^{(1)}\frac{\mu_{i_2j_4}^{(1)}\omega_{i_2}^{j_4j_4}}{B_1(T_{i_1l_2})}K_h(T_{i_2j_4}-T_{i_1l_2}) \\ &\qquad -\frac{1}{2}\mu_{i_1j_3}^{(1)}\omega_{i_1}^{j_3j_3}\mu_{i_2j_4}^{(1)}\omega_{i_2}^{j_4j_4}\frac{B_2(T_{i_2j_4})}{B_1^2(T_{i_2j_4})}K_h*K_h(T_{i_1j_3}-T_{i_2j_4})\Big\}\Big] \\ &\quad +\mathrm{E}\Big[\frac{1}{m^2}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\sum_{j_3=1}^{m}\sum_{j_4=1}^{m}\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\varepsilon_{i_2j_3}\varepsilon_{i_1j_4}\Big\{\sum_{\ell_1=1}^{m}\nu_{i_1}^{j_1l_1}\mu_{i_1l_1}^{(1)}\frac{\mu_{i_2j_2}^{(1)}\omega_{i_2}^{j_2j_2}}{B_1(T_{i_1l_1})}K_h(T_{i_2j_2}-T_{i_1l_1}) \\ &\qquad -\frac{1}{2}\mu_{i_1j_1}^{(1)}\omega_{i_1}^{j_1j_1}\mu_{i_2j_2}^{(1)}\omega_{i_2}^{j_2j_2}\frac{B_2(T_{i_2j_2})}{B_1^2(T_{i_2j_2})}K_h*K_h(T_{i_1j_1}-T_{i_2j_2})\Big\} \\ &\quad \times\Big\{\sum_{\ell_2=1}^{m}\nu_{i_2}^{j_3l_2}\mu_{i_2l_2}^{(1)}\frac{\mu_{i_1j_4}^{(1)}\omega_{i_1}^{j_4j_4}}{B_1(T_{i_2l_2})}K_h(T_{i_1j_4}-T_{i_2l_2}) \\ &\qquad -\frac{1}{2}\mu_{i_2j_3}^{(1)}\omega_{i_2}^{j_3j_3}\mu_{i_1j_4}^{(1)}\omega_{i_1}^{j_4j_4}\frac{B_2(T_{i_1j_4})}{B_1^2(T_{i_1j_4})}K_h*K_h(T_{i_2j_3}-T_{i_1j_4})\Big\}\Big] \end{aligned}$$

$$= \mathrm{E}\left\{\frac{B_3(T)B_4(T)+B_5^2(T)}{B_1^2(T)f(T)}\int K^2(t)dt + \frac{B_2^2(T)B_3^2(T)}{2B_1^4(T)f(T)}\int (K*K)^2(t)dt \right.$$
$$\left. -2\frac{B_2(T)B_3(T)B_5(T)}{B_1^3(T)f(T)}\int K(t)K*K(t)dt\right\} + O(1).$$

Hence, $\mathrm{var}(W_n) = \sigma_n^2 + O(1)$. Since $J_1 + J_4 = \mu_n + W_n + O_p(1)$, the asymptotic distribution in the lemma follows directly from Proposition 3.2 in de Jong (1987). ◊

### 3.4.2 Proof of Corolary 1

Under the null hypothesis, recall that $B_1(t) = \frac{1}{m}\sum_{j=1}^m \mathrm{E}(\Delta_{jj}^2\omega^{jj}|T_j = t)f(t)$ and $B_2(t) = \frac{1}{m}\sum_{j=1}^m \mathrm{E}(\Delta_{jj}^2\nu^{jj}|T_j = t)f(t)$, we can easily show that $\nu^{jj} = \{\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}\omega^{jj}$, where $\mathcal{C}(\boldsymbol{\tau})$ is the working correlation matrix depending on an unknown parameter vector $\boldsymbol{\tau}$. Since both $\Delta_{jj}$ and $\omega^{jj}$ are functions of only $\mathrm{var}(\mu_{jj})$, which means that $\mathrm{var}(\mu_{jj}|\boldsymbol{X}_j, T_j, T_k) = \mathrm{var}(\mu_{jj}|\boldsymbol{X}_j, T_j)$ for all $jandk$, we can get

$$\begin{aligned} B_2(t) &= \frac{1}{m}\sum_{j=1}^m \mathrm{E}\left[\{\mathcal{S}^{-1}\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{S}^{-1}\}_{jj}\Delta_{jj}^2|T_j = t\right] \\ &= \frac{1}{m}\sum_{j=1}^m \mathrm{E}\left[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}\Delta_{jj}^2\omega^{jj}|T_j = t\right] \\ &= \frac{1}{m}\sum_{j=1}^m \mathrm{E}\left[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t\right]\mathrm{E}(\Delta_{jj}^2\omega^{jj}|T_j = t) \\ &= B_{2\dagger}(t)B_1(t) \end{aligned}$$

Next, we need to simplify $B_3(t)$, $B_4(t)$ and $B_5(t)$. According to the Corollary 1, $\mathcal{W} = \Sigma_d = \mathcal{S}^2$, a straightforward calculation shows that

$$\begin{aligned} B_3(t) &= \frac{1}{m}\sum_{j=1}^m \mathrm{E}\{\sigma_{jj}\Delta_{jj}^2(\omega^{jj})^2|T_j = t\}f(t) \\ &= \frac{1}{m}\sum_{j=1}^m \mathrm{E}\{\sigma_{jj}\omega^{jj}|T_j = t\}\mathrm{E}\{\Delta_{jj}^2\omega^{jj}|T_j = t\}f(t) = B_1(t) \end{aligned}$$

$$
\begin{aligned}
B_4(t) &= \frac{1}{m}\sum_{j_1=1}^{m} \mathrm{E}\left\{ \sum_{j_2=1}^{m}\sum_{j_3=1}^{m} \nu^{j_1j_2}\sigma_{j_2j_3}\nu^{j_3j_1}\Delta^2_{j_1j_1} \middle| T_{j_1}=t \right\} f(t) \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[ \{\mathcal{S}^{-1}\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{R}\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{S}^{-1}\}_{jj}\Delta^2_{jj}|T_j=t \right] f(t) \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[ \{\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{R}\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}\Delta^2_{jj}\omega^{jj}|T_j=t \right] f(t) \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[ \{\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{R}\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j=t \right] B_1(t) \\
&= B_{4\dagger}(t)B_1(t) \\
B_5(t) &= \frac{1}{m}\sum_{j_1=1}^{m} \mathrm{E}\left\{ \sum_{j=1}^{m} \sigma_{jj_1}\nu^{jj_1}\Delta^2_{j_1j_1}\omega^{j_1j_1} \middle| T_{j_1}=t \right\} f(t). \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[ \{\mathcal{S}\mathcal{R}\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{S}^{-1}\}_{jj}\Delta^2_{jj}\omega^{jj} \middle| T_j=t \right] f(t) \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{E}\left[ \{\mathcal{R}\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj} \middle| T_j=t \right] B_1(t) \\
&= B_{5\dagger}(t)B_1(t)
\end{aligned}
$$

where $B_{2\dagger}(t)$ - $B_{5\dagger}(t)$ are defined in (3.6).

By pluging $B_1(t)$ - $B_5(t)$ into the asymptotic distribution of $\lambda_n(H_0)$ in Theorem 1, $\lambda_n(H_0)$ follows an asymptotic normality given in Corollary 1. $\diamondsuit$

### 3.4.3 Proof of Theorem 2

Lemma **6.** *Suppose assumptions (C.1) – (C.6) and the local alternative described in (3.7) and (3.8) hold, $\widehat{\boldsymbol{\beta}}_R$ is still root-n consistent to $\boldsymbol{\beta}_0$, and $\widehat{\boldsymbol{\beta}}_F - \widehat{\boldsymbol{\beta}}_R = o_p(n^{-1/2})$. The nonparametric estimator $\widehat{\theta}_R(t)$ has the same asymptotic expansion as in (2.8).*

**Proof:** For a fixed $\boldsymbol{\beta}$, we derive the asymptotic expansion of profile kernel estimator $\widehat{\theta}_R(t;\boldsymbol{\beta})$ using standard derivations (Lin and Carroll, 2001) and get

$$\begin{aligned}\widehat{\theta}_R(t;\boldsymbol{\beta})-\theta_0(t) &= \frac{1}{nmB_1(t)}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij}-t)\Big[\varepsilon_{k,ij}+\mu_{k,ij}^{(1)}\boldsymbol{X}_{k,ij}^{\mathrm{T}}(\boldsymbol{\beta}_0-\boldsymbol{\beta})\\ &\quad+\mu_{k,ij}^{(1)}\{\theta_{k0}(T_{k,ij})-\theta_0(t)\}\Big]+o_p(n^{-1/2})\\ &= \frac{1}{mnB_1(t)}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij}-t)\{\varepsilon_{k,ij}+\mu_{k,ij}^{(1)}G_{kn}(T_{k,ij})\}\\ &\quad+\frac{h^2}{2}\theta_0^{(2)}(t)-\boldsymbol{\mu}_X^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p(n^{-1/2}+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|). \qquad (3.19)\end{aligned}$$

Since $\sum_k \rho_k G_{kn}(t)=0$ for all $t$ and $G_{kn}(T)=O_p(n^{-1/2}h^{-1/2})$, $B_{1k}(t)=B_1(t)+O(n^{-1/2}h^{-1/2})$ and

$$\begin{aligned}&\frac{1}{mn}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}(\mu_{k,ij}^{(1)})^2\omega_{k,i}^{jj}K_h(T_{k,ij}-t)G_{kn}(T_{k,ij})\\ =&\sum_{k=1}^{q}\rho_k B_{1k}(t)G_{kn}(t)+O_p\{(nh)^{-1/2}\times(h^2+n^{-1/2}h^{-1/2})\}=o_p(n^{-1/2}).\end{aligned}$$

Finally, we get

$$\widehat{\theta}_R(t;\boldsymbol{\beta})-\theta_0(t) = \frac{h^2}{2}\theta_0^{(2)}(t)+\mathcal{U}_R(t)-\boldsymbol{\mu}_X^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p(n^{-1/2}+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|). \qquad (3.20)$$

Therefore, if $\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0=O_p(n^{-1/2})$, the expansion of $\widehat{\theta}_R(t)$ follows directly from (3.20) and the leading terms are identical to those in (2.8).

We next derive the asymptotic expansion of $\widehat{\boldsymbol{\beta}}_R$. By standard profile estimator arguments,

$$\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0=\boldsymbol{D}_{\dagger}^{-1}\mathcal{E}_{n\dagger}+o_p(n^{1/2}),$$

where $\boldsymbol{D}_{\dagger}=\sum_k\rho_k\mathrm{E}\{\widetilde{\boldsymbol{X}}_{k\dagger}^{\mathrm{T}}\Delta_k\mathcal{W}_k^{-1}\Delta_k\widetilde{\boldsymbol{X}}_{k\dagger}\}$, $\mathcal{E}_{n\dagger}=n^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\widetilde{\boldsymbol{X}}_{k,i\dagger}^{\mathrm{T}}\Delta_{k,i}\mathcal{W}_{k,i}^{-1}\boldsymbol{\epsilon}_{k,i}$, $\widetilde{\boldsymbol{X}}_{k\dagger}=\{(\boldsymbol{X}-\boldsymbol{\mu}_X)(T_\ell)\}_{\ell=1}^{m}$. Therefore, $\widehat{\boldsymbol{\beta}}_R$ is still root-$n$ consistent to $\boldsymbol{\beta}_0$.

By the assumption that $G_{kn}(T) = O_p\{(nh)^{-1/2}\}$, $k = 1, \ldots, q$, we can see that $\mathcal{D}_* - \mathcal{D}_\dagger = o(1)$, and $\mathcal{E}_{n*} - \mathcal{E}_{n\dagger} = o(n^{-1/2})$, and hence $\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}_F = o_p(n^{-1/2})$.

$\diamondsuit$

**Proof of Theorem 2:** The test statistic has a similar decomposition as (3.13)

$$\lambda_{1n}(H_0) = J_1^\dagger + J_2^\dagger + J_3^\dagger + J_4^\dagger + J_5^\dagger + J_6^\dagger + o_p(1),$$

where

$$\begin{aligned}
J_1^\dagger &= \sum_k \boldsymbol{\epsilon}_k^{\mathrm{T}} \mathbf{V}_k^{-1} \Delta_k \{\widehat{\theta}_{F,k}(\boldsymbol{T}_k) - \widehat{\theta}_R(\boldsymbol{T}_k)\}, \quad J_2^\dagger = \sum_k \boldsymbol{\epsilon}_k^{\mathrm{T}} \mathbf{V}_k^{-1} \Delta_k \boldsymbol{X}_k (\widehat{\boldsymbol{\beta}}_F - \widehat{\boldsymbol{\beta}}_R), \\
J_3^\dagger &= \sum_k (\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{X}_k^{\mathrm{T}} \Delta_k \mathbf{V}_K^{-1} \Delta_k \{\widehat{\theta}_R(\boldsymbol{T}_k) - \theta_{k0}(\boldsymbol{T}_k)\} \\
&\qquad - (\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{X}_k^{\mathrm{T}} \Delta_k \mathbf{V}_K^{-1} \Delta_k \{\widehat{\theta}_{F,k}(\boldsymbol{T}_k) - \theta_{k0}(\boldsymbol{T}_k)\}, \\
J_4^\dagger &= \frac{1}{2} \sum_k \|\widehat{\theta}_R(\boldsymbol{T}_k) - \theta_{k0}(\boldsymbol{T}_k)\|^2_{\Delta_k \mathbf{V}_K^{-1} \Delta_k} - \|\widehat{\theta}_{F,k}(\boldsymbol{T}_k) - \theta_{k0}(\boldsymbol{T}_k)\|^2_{\Delta_k \mathbf{V}_K^{-1} \Delta_k}, \\
J_5^\dagger &= \frac{1}{2} \sum_k \|\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\|^2_{\boldsymbol{X}_k^{\mathrm{T}} \Delta_k \mathbf{V}_k^{-1} \Delta_k \boldsymbol{X}_k} - \|\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0\|^2_{\boldsymbol{X}_k^{\mathrm{T}} \Delta_k \mathbf{V}_k^{-1} \Delta_k \boldsymbol{X}_k}, \\
J_6^\dagger &= \frac{1}{2} \sum_k \sum_{i=1}^{n_k} \|\boldsymbol{X}_{k,i}(\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + \widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i}) - \theta_{k0}(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij} \mathcal{D}_{k,ij}} \\
&\qquad - \|\boldsymbol{X}_{k,i}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) + \widehat{\theta}_R(\boldsymbol{T}_{k,i}) - \theta_{k0}(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij} \mathcal{D}_{k,ij}}.
\end{aligned}$$

By similar derivations as in Lemma 4, we can show that $J_2^\dagger + J_3^\dagger + J_5^\dagger + J_6^\dagger = o_p(h^{-1/2})$ and hence the dominant terms in $\lambda_{1n}(H_0)$ are $J_1^\dagger$ and $J_4^\dagger$.

Under the local alternative, $G_{kn}(T) = O_p\{(nh)^{-1/2}\}$, one can show $\boldsymbol{\mu}_{X,k}(t) - \boldsymbol{\mu}_k(t) = O\{(nh)^{-1/2}\}$. By Proposition 2 and Lemma 6,

$$\begin{aligned}
\widehat{\theta}_{F,k}(t) - \widehat{\theta}_R(t) &= G_{kn}(t) + \frac{1}{2} G_{kn}^{(2)}(t) h^2 + \mathcal{U}_{F,k}(t) - \mathcal{U}_R(t) + \boldsymbol{\mu}_X^{\mathrm{T}}(t)(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \\
&\qquad - \boldsymbol{\mu}_{X,k}^{\mathrm{T}}(t)(\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + o_p(n^{-1/2}) \\
&= G_{kn}(t) + \mathcal{U}_{F,k}(t) - \mathcal{U}_R(t) + o_p(n^{-1/2}).
\end{aligned}$$

By straightforward calculations, $J_1^\dagger = J_1 + R_1^\dagger + o_p(h^{-1/2})$ and $J_4^\dagger = J_4 + R_2^\dagger + R_3^\dagger + R_4^\dagger + o_p(h^{-1/2})$, where

$$\begin{aligned}
R_1^\dagger &= \sum_{k=1}^{q}\sum_{i=1}^{n_k} \boldsymbol{\varepsilon}_{k,i}^{\mathrm{T}} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} G_{kn}(\boldsymbol{T}_{k,i}),\\
R_2^\dagger &= \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k} G_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}G_{kn}(\boldsymbol{T}_{k,i}),\\
R_3^\dagger &= -\sum_{k=1}^{q}\sum_{i=1}^{n_k} G_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}\mathcal{U}_R(\boldsymbol{T}_{k,i}),\\
R_4^\dagger &= \sum_{k=1}^{q}\sum_{i=1}^{n_k} G_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\Delta_{k,i}\mathbf{V}_{k,i}^{-1}\Delta_{k,i}\boldsymbol{\mu}_X(\boldsymbol{T}_{k,i})(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0).
\end{aligned}$$

More detailed calculation shows

$$\begin{aligned}
R_3^\dagger &= -\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\sum_{\ell=1}^{m} G_{kn}(T_{k,ij})\nu_{k,i}^{j\ell}\mu_{k,ij}^{(1)}\mu_{k,i\ell}^{(1)}\\
&\qquad \times\left\{\frac{1}{nmB_1(T_{k,i\ell})}\sum_{k'=1}^{q}\sum_{i'=1}^{n_{k'}}\sum_{j'=1}^{m}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}K_h(T_{k',i'j'} - T_{k,i\ell})\varepsilon_{k',i'j'}\right\}\\
&= -\sum_{k'=1}^{q}\sum_{i'=1}^{n_{k'}}\sum_{j'=1}^{m}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}\varepsilon_{k',i'j'}\\
&\qquad \times\frac{1}{nm}\left\{\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\sum_{\ell=1}^{m} G_{kn}(T_{k,ij})\nu_{k,i}^{j\ell}\mu_{k,ij}^{(1)}\mu_{k,i\ell}^{(1)}B_1^{-1}(T_{k,i\ell})K_h(T_{k',i'j'} - T_{k,i\ell})\right\}\\
&= -\sum_{k'=1}^{q}\sum_{i'=1}^{n_{k'}}\sum_{j'=1}^{m}\mu_{k',i'j'}^{(1)}\omega_{k',i'}^{j'j'}\varepsilon_{k',i'j'}[B_G(T_{k',i'j'}) + O_p\{n^{-1/2}h^{3/2} + (nh)^{-1}\}],
\end{aligned}$$

where $B_G(t) = \sum_{k=1}^{q}\rho_k \mathrm{E}\{\sum_{j=1}^{m} G_{kn}(T_j)\nu_k^{j1}\mu_{k,1j}^{(1)}\mu_{k,11}^{(1)}|T_1 = t\}/B_1(t)$. Since $\sum_k \rho_k G_{kn}(t) = 0$ and $G_{kn}(t) = O_p\{(nh)^{-1/2}\}$, we have $\nu_k^{j1} = \nu_1^{j1} + O_p\{(nh)^{-1/2}\}$ and $\mu_{k,1j}^{(1)} = \mu_{1,1j}^{(1)} + O_p\{(nh)^{-1/2}\}$ for $k = 2, \ldots, q$. Consequently, $B_G(t) = O\{(nh)^{-1}\}$ and $R_3^\dagger = O_p(n^{-1/2}h^{-1} + h^{3/2}) = o_p(h^{-1/2})$. Similarly,

$$\begin{aligned}
R_4^\dagger &= n(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)^{\mathrm{T}}\left[\sum_{k=1}^{q}\rho_k \mathrm{E}\left\{\sum_{j=1}^{m}\sum_{\ell=1}^{m} G_{kn}(T_j)\nu_{k,1}^{j\ell}\mu_{k,1j}^{(1)}\mu_{k,1\ell}^{(1)}\right\} + O_p(n^{-1}h^{-1/2})\right]\\
&= O_p(n^{-1/2}h^{-1}) = o_p(h^{-1/2}).
\end{aligned}$$

Therefore $\lambda_n(H_{1n}) = \mu_n + W_n + R_1^\dagger + R_2^\dagger + o_p(h^{-1/2})$, where $\mu_n$ and $W_n$ are as defined in Theorem 1. By the assumption in (3.8), $R_2^\dagger = \mu_{1n} + O_p(n^{-1/2}h^{-1}) = \mu_{1n} + o_p(h^{-1/2})$. Since $R_1^\dagger$ is a linear combination of $\varepsilon_{k,ij}$ and $W_n$ only consists of quadratic terms, it is easy to see that $R_1^\dagger$ and $W_n$ are uncorrelated and hence asymptotically independent. Therefore, $\mathrm{E}\{\lambda_n(H_{1n})\} = \mu_n + \mu_{1n} + o_p(h^{-1/2})$ and $\mathrm{var}\{\lambda_n(H_{1n})\} = \mathrm{var}(W_n) + \mathrm{var}(R_1^\dagger) + o_p(h^{-1}) = \sigma_{n\ddagger}^2$. The asymptotic normality of $\lambda_n(H_{1n})$ follows from those of $W_n$ and $R_1^\dagger$. $\diamondsuit$

### 3.4.4 Proof of Theorem 3

Following the proof of Theorem 2, the probability of type II error under the local alternative is

$$\beta(\alpha, \boldsymbol{G}_n) = \Phi\{(\sigma_n^2 + d_{2n})^{-1/2}(z_\alpha \sigma_n - \mu_{1n})\},$$

where $\mu_{1n}$ and $d_{2n}$ are defined as in Theorem 2. With a slight abuse of notation, define the squared norm of the functional vector $\boldsymbol{G}_n$ as $\varrho^2(\boldsymbol{G}_n) = \sum_{k=1}^q \rho_k \mathrm{E}\{G_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\Delta_k \mathbf{V}_{d,k}^{-1}\Delta_k G_{kn}(\boldsymbol{T}_k)\}$. With $h = c^* n^{-2/9}$, we have $\sigma_n^2 = C_1 n^{2/9} + O(1)$, $d_{2n} = C_2 n \varrho^2(\boldsymbol{G}_n) \times \{1 + O_p(n^{-1/2})\}$ for some constants $0 < C_1, C_2 < \infty$, and $\mu_{1n} = n\varrho^2(\boldsymbol{G}_n) \times \{1 + O_p(n^{-1/2})\}$.

For any $\varrho(\boldsymbol{G}_n) = cn^{-4/9}$, we have $\beta(\alpha, \boldsymbol{G}_n) = \Phi\{(C_1 n^{2/9} + C_2 c^2 n^{1/9})^{-1/2}(z_\alpha C_1^{1/2} n^{1/9} - c^2 n^{1/9})\} + o(1) = \Phi\{z_\alpha - c^2 C_1^{-1/2}\} + o(1)$. For any $\beta > 0$, we can choose $c$ to be large enough so that $\beta(\alpha, \boldsymbol{G}_n) < \beta$. Therefore, $\beta(\alpha, cn^{-4/9}) < \beta + o(1)$.

For any $\varrho_{n*} = o(n^{-4/9})$ and any $\boldsymbol{G}_n$ satisfying $\varrho(\boldsymbol{G}_n) = c\varrho_{n*}$ for some $c > 0$, we have $\beta(\alpha, \boldsymbol{G}_n) = \Phi\{(C_1 n^{2/9} + C_2 c^2 n \rho_{n*}^2)^{-1/2}(z_\alpha C_1^{1/2} n^{1/9} - c^2 n \varrho_{n*}^2)\} + o(1) = 1 - \alpha + o(1)$. Therefore there exists $\beta < 1 - \alpha$ so that $\beta(\alpha, \boldsymbol{G}_n) > \beta$ and hence $\liminf_n \beta(\alpha, c\varrho_{n*}) > \beta$. We have now verified that $\varrho_n(h) = n^{-4/9}$ satisfies both conditions for the minimax rate. $\diamondsuit$

# Chapter 4

# Simulation Studies

To investigate the performance of our proposed GQLR test, we consider three simulation settings: Gaussian longitudinal data with homogeneous variance, Gaussian longitudinal data with heterogenous variance and binary longitudinal data. Throughout the simulation studies, we first fit the full model to obtain initial estimates of $\boldsymbol{\beta}$ and $\theta_k(t)$, and then estimate the variance function $\sigma^2(\boldsymbol{\mu})$ and the working correlation structure $\mathcal{C}(\boldsymbol{\tau})$ by applying the methods described in Section 3.3.2. The weight is the inverse of the estimated variance function for both the reduced and full models.

## 4.1 Simulation 1: Gaussian data with homogenous variance

In this simulation, consider the following model

$$Y_{k,ij} = X_{k,ij}\beta + \theta_k(T_{k,ij}) + \varepsilon_{k,ij}, \quad k = 1, 2, \quad i = 1, \cdots, 100, \quad j = 1, \cdots, 4, \tag{4.1}$$

where $T_{k,ij}$ are generated as i.i.d. random variables from a uniform distribution on $[0, 1]$, $\varepsilon_{k,ij}$ are i.i.d. N(0, 1) random variables and the time dependent covariate $X_{k,ij} = T_{k,ij} + U(-1, 1)$. This setting implies that the marginal density of $(X_j, T_j)$ is the same for any $j$ and $\mathrm{E}(X_j|T_j, T_l) = \mathrm{E}(X_j|T_j)$ for $j \neq l$. Let the true correlation structure within a cluster to be ARMA(1,1) , i.e. $\mathrm{corr}\{\varepsilon(s), \varepsilon(t)\} = 1$ for $s = t$ and $\gamma \exp(-|s - t|/\nu)$ for $s \neq t$. We set $\gamma = 0.6$ and $\nu = 1$.

To examine the Wilks phenomenon of the GQLR test, we assume the null hypothesis is true that $\theta_1(t) = \theta_2(t) = \theta_0(t)$, and generate 400 datasets from each of the following three different scenarios,

$$\text{Scenario I}: \quad \beta = 1,\ \theta_0(t) = 2\sin(2t)^2;$$

$$\text{Scenario II}: \quad \beta = -0.5,\ \theta_0(t) = 2\sin(2t)^2;$$

$$\text{Scenario III}: \quad \beta = 1,\ \theta_0(t) = \sin(2\pi t).$$

For the local linear estimator, we adopt the Epanechnikov kernel

$$K(t/h) = \frac{3}{4}\{1 - (t/h)^2\}\mathbf{1}_{\{|t/h| \leqslant 1\}}, \tag{4.2}$$

where $h$ is the bandwidth set to be 0.12 and is fixed for both reduced and full models across all simulations to eliminate the variation in $\lambda_n(H_0)$ caused by bandwidth selection. We calculate the the initial estimates of $\beta$ and $\theta_k(t)$ in the full model setting $\mathcal{W}_{k,i}$'s to be identity matrices. To construct the GQLR test statistic, we use a Gaussian quasi-likelihood

$$\mathcal{Q}(\boldsymbol{\mu}, \boldsymbol{Y}) = (\boldsymbol{Y} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}(\boldsymbol{\mu})^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}), \tag{4.3}$$

where $\mathbf{V}(\boldsymbol{\mu}) = \mathcal{S}(\boldsymbol{\mu})\mathcal{C}(\boldsymbol{\tau})\mathcal{S}(\boldsymbol{\mu})$, $\mathcal{S}(\boldsymbol{\mu}) = \mathrm{diag}\{\widehat{\sigma}(\mu_j)\}_{j=1}^{m}$ and $\mathcal{C}(\boldsymbol{\tau})$ is a working correlation matrix.

*Asymptotic distribution of $r_k\lambda_n(H_0)$ under working independence:* Figure 4.1 shows the estimated kernel density of $r_k\lambda_n(H_0)$ (solid line) from scenario I, when the working correlation structure $\mathcal{C}(\boldsymbol{\tau})$ is set to be identity. The dashed curve is the density of a $\chi^2$ distribution with the degree of freedom set to be the empirical mean of $r_k\lambda_n(H_0)$. Both curves are estimated by using the 'density' function in **R**. The fact that the distribution of the test statistic is closely approximated by a $\chi^2$ density corroborates our theory in Corollary 2.

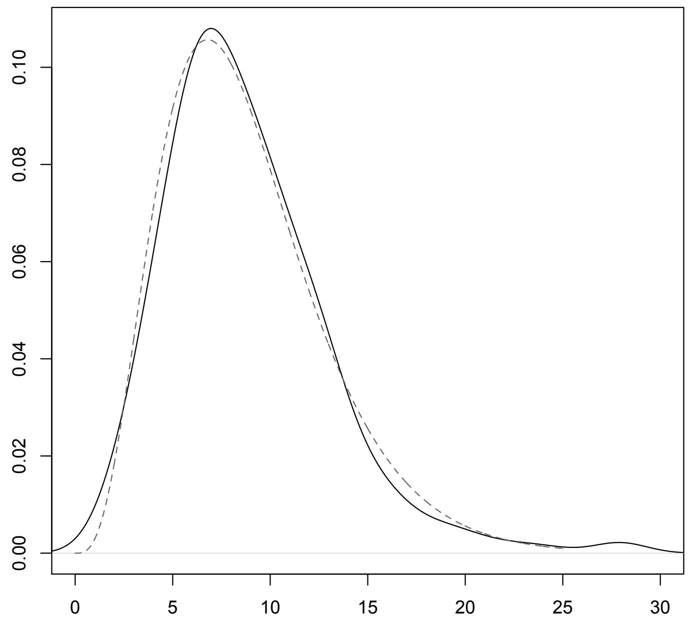



Figure 4.1: *Simulation 1: the estimated kernel density function of $r_K \lambda_n(H_0)$ (solid line) under Scenario I and the $\chi^2$ distribution with the degree freedom matching the sample mean of $r_K \lambda_n(H_0)$ (dashed line). The working correlation matrix $\mathcal{C}(\boldsymbol{\tau})$ is set to be identity.*

*Wilks Phenomenon under misspecified correlation structure:* To confirm that the asymptotic null distribution of $\lambda_n(H_0)$ is independent of the parameters $\beta$ and $\theta(t)$, we consider both a mis-specified AR(1) working correlation structure (i.e. $\rho(s, t; \tau) = exp(-|s - t|/\nu)$ if $s \neq t$ and 1 if $s = t$ ) and working independence. We examine the empirical distributions of $\lambda_n(H_0)$ under the three different scenarios when using each correlation structure respectively. Compared to scenario I, scenario Ⅱ has the same smoothing function $\theta(t)$ but a different coefficient $\beta$; scenario Ⅲ has the same coefficient $\beta$ but a different smoothing function $\theta(t)$. The three null distributions of $\lambda_n(H_0)$ under scenarios I - Ⅲ are depicted in Figure 4.2 under a misspecified AR(1) correlation and Figure 4.3 under WI. The solid, dashed and dotted curves correspond to the senarios I, II and III. Both Figures demonstrate that the null densities are almost the same either $\beta$ or $\theta$ varies, which confirms our theoretical results that, when the variance function is correctly specified, the asymptotic distribution of $\lambda_n(H_0)$ does not depends on the value of $\beta$ and $\theta_0(t)$. By comparing the curves between the Figures, we find the distribution of $\lambda_n(H_0)$ depends on the working correlation structure $\mathcal{C}(\boldsymbol{\tau})$.

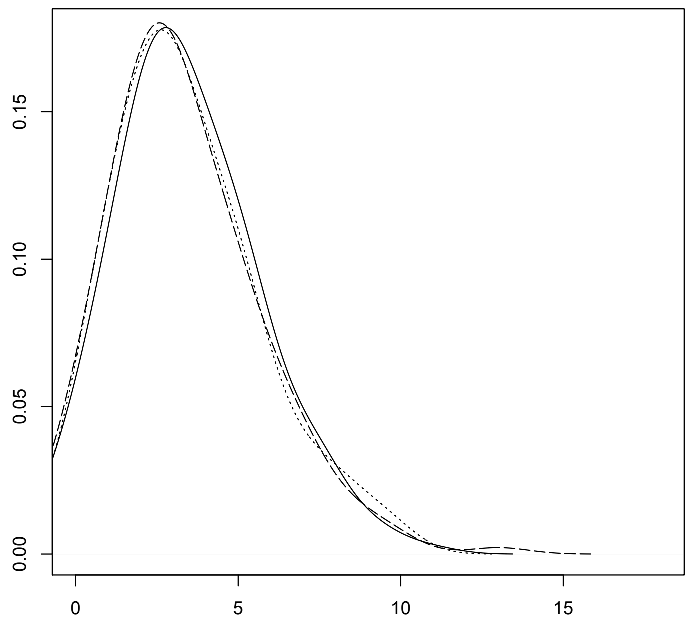



Figure 4.2: *Simulation 1: the empirical distribution of $\lambda_n(H_0)$ with a misspecified AR(1) correlation structure when the variance function is consistently estimated. (solid line: scenario I; dashed line: scenario II; dotted line: scenario III).*

*Power of the GQLR tests:* For the power assessment, we focuse on scenario I, fixing $\theta_1(t)$ at $\theta_0(t)$, while changing $\theta_2(t)$ to $\theta_{2,\phi}(t)$ where

$$\theta_{2,\phi}(t) = 2\sin(2t)^2 + \phi\sin(t), \quad 0 < \phi < 1.$$

We set $\phi$=0, 0.2, 0.4, 0.6, and 0.8, respectively. As $\phi$ increases, the model deviates further away from $H_0$. The discrepancy among the smooth functions is displayed in Figure 4.4.

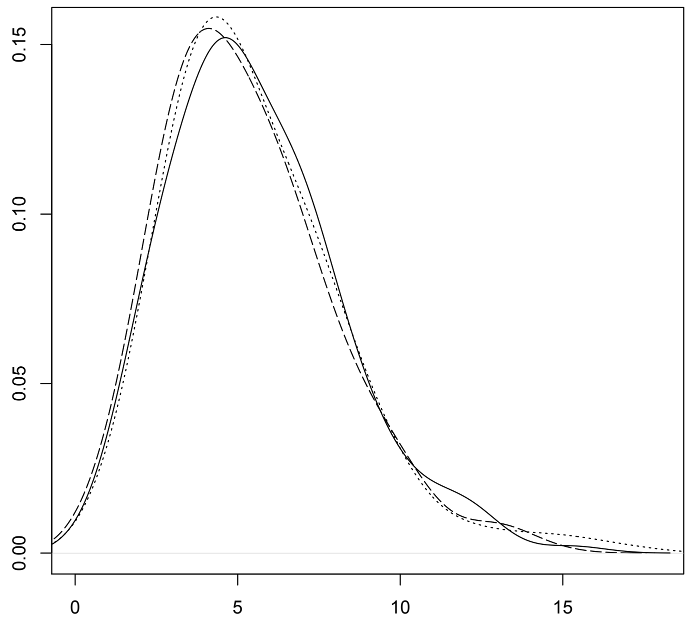



Figure 4.3: *Simulation 1: the empirical distribution of* $\lambda_n(H_0)$ *with WI working covariance when the variance function is consistently estimated. (solid line: scenario I; dashed line: scenario II; dotted line: scenario III).*

For each value of $\phi$, we generate 500 datasets from model (4.1). The true within-cluster correlation is ARMA(1,1) as described before, but we perform the GQLR test based on working independence. We set the significance level at $\alpha = 0.05$, use the distribution of $\lambda_n$ under $\phi = 0$ to decide the critical values, and calculate the rejection frequencies for the other $\phi$ values. The results are depicted in Figure 4.5. The rejection rate of $H_0$ gets higher as $\phi$ increases.

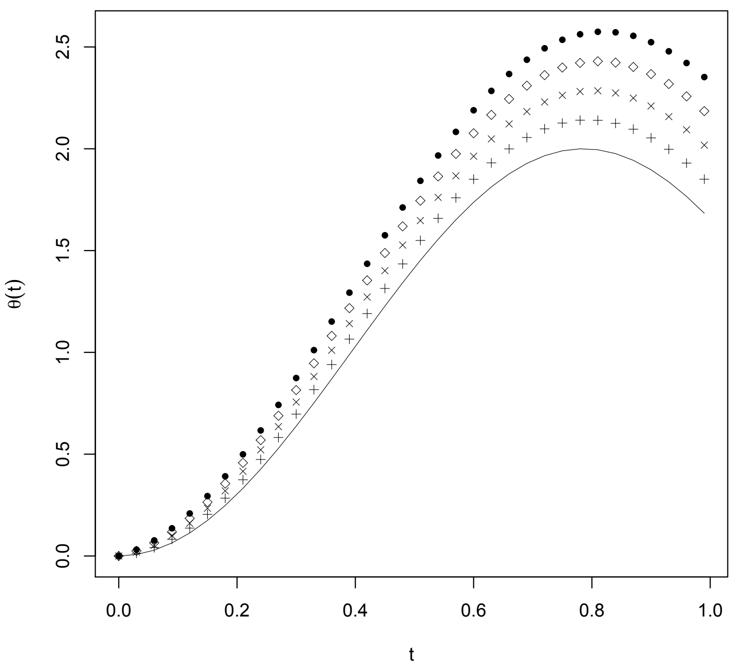



Figure 4.4: *Simulation 1: plot of* $\theta_{2,\phi}(t)$*:* — $\phi = 0$, $+++: \phi = 0.2$, $\times\times\times: \phi = 0.4$, $\diamond\diamond\diamond: \phi = 0.6$, $\cdots: \phi = 0.8$.

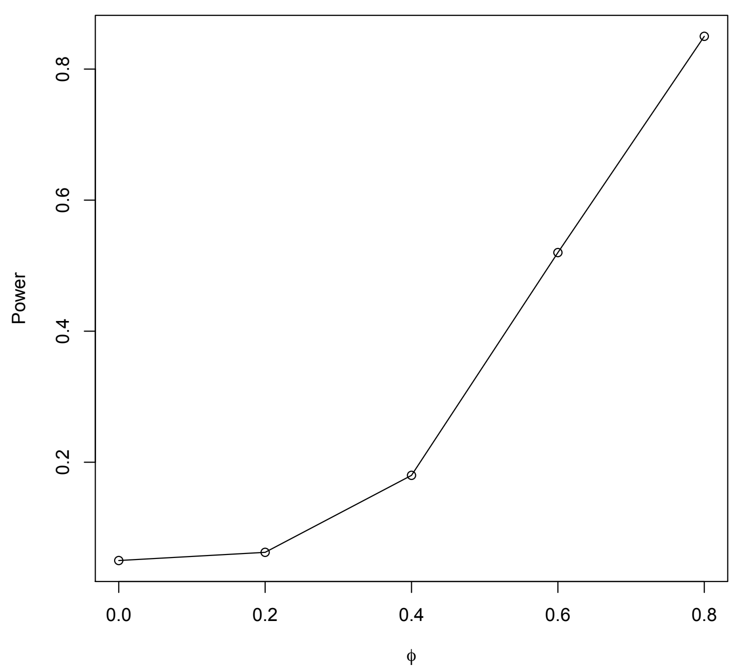



Figure 4.5: *Simulation 1: power curve of the GQLR test where the significance level is* $\alpha = 0.05$.

### 4.2 Simulation 2: Gaussian data with heterogeneous variance

As Corollary 1 shows, the Wilks phenomenon holds as long as the variance is estimated correctly, even though the correlation structure is misspecified. We here considere two particular cases where the variance structure is specified correctly and incorrectly. Consider simulation setups similar to the three scenarios in Section 4.1, except that the variance of $Y_{k,ij}$ depends on the mean such that

$$\mathrm{Var}(Y_{k,ij}|X_{k,ij}, T_{k,ij}) = 0.3 * \mu_{k,ij}^2 + 0.3,$$

where $\mu_{k,ij} = X_{k,ij}\beta + \theta_k(T_{k,ij})$.

*Wilks phenomenon when the variance function is consistently estimated:* We estimate the conditional variance function by a local linear estimator described in Section 3.3.2, and plug the estimated variance into the GQLR test statistic. To demonstrate the Wilks phenomenon, we compare the null distributions of $\lambda_n(H_0)$ under scenarios I, Ⅱ and Ⅲ based on 300 simulations. An exchangeable (i.e. $\rho(s,t;\tau) = \tau$ for some $-m^{-1} < \tau < 1$ if $s \neq t$) and working independence correlation structures are used.

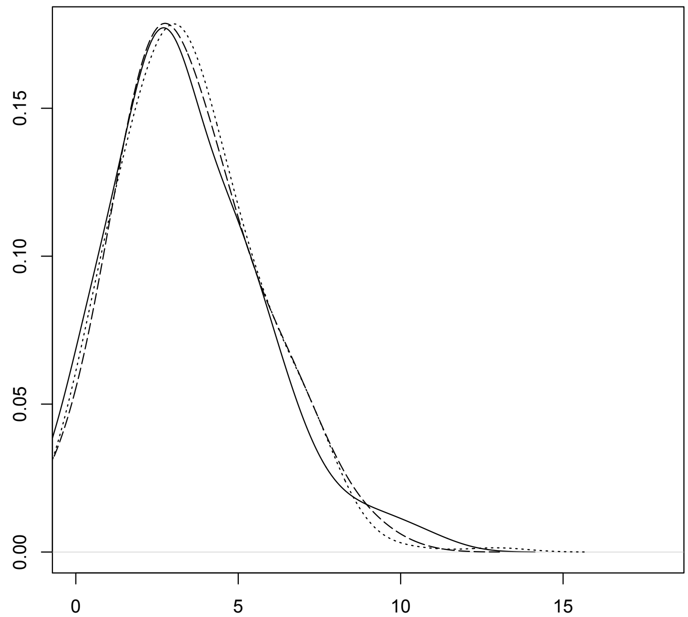
The empirical distributions of $\lambda_n(H_0)$ under different settings are shown in Figure 4.6. Panel (a) shows the distributions of $\lambda_n(H_0)$ when an exchangeable correlation structure is used; Panel (b) shows the same distributions when working independence is adopted. In each panel, the three estimated density functions correspond to the three scenarios. These three density functions in the same panel are almost the same when $\beta$ and $\theta_0(t)$ change. By comparing the density curves between the two panels, the distribution of $\lambda_n(H_0)$ depends on the working correlation structure.

| Working Correlation | Exchangeable | | Working Independence | | | |
|---|---|---|---|---|---|---|
| Working Variance | Nonpara. Est. | | Nonpara. Est. | | constant | |
| | Mean | SE | Mean | SE | Mean | SE |
| Scenario I | 3.38 | 2.34 | 5.03 | 2.47 | 5.27 | 3.03 |
| Scenario II | 3.47 | 2.15 | 5.17 | 2.43 | 5.24 | 2.66 |
| Scenario III | 3.45 | 2.17 | 5.10 | 2.29 | 5.30 | 2.42 |

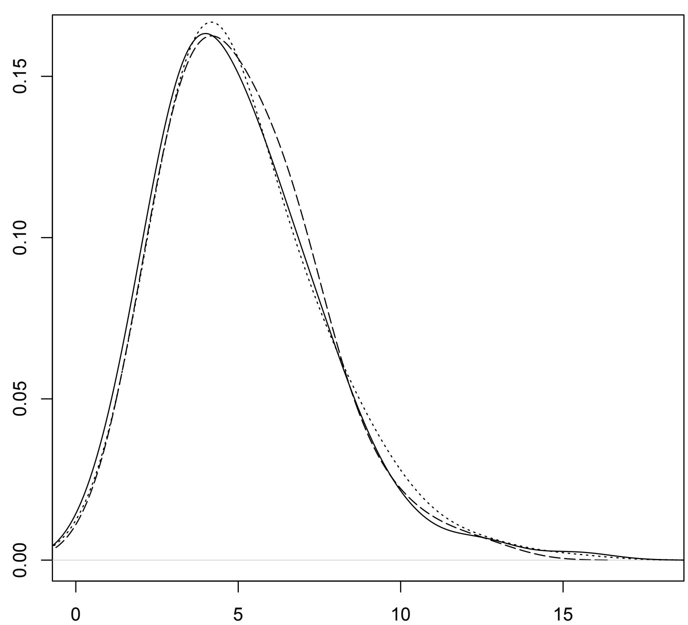
Table 4.1: Simulation 2: Mean and Standard Errors of $\lambda_n(H_0)$ for Gaussian data with heterogeneous variance. The working variance used in the test is either a nonparametric estimator using local polynomial or misspecified as a constant. The working correlation used in the test is either a mis-specified exchangeable correlation structure or working independence.

Table 4.1 presents the mean and standard error (SE) of $\lambda_n(H_0)$ under different working correlations. To verify that the distributions in each panel are indeed the same, we conduct the two-sample t-tests and the F-tests to examine the equality of means and variances of $\lambda_n(H_0)$ among different scenarios under the same working correlation structure. For example, when an exchangeable correlation is applied, the p-values are 0.624 and 0.144 for scenario I versus II, 0.704 and 0.193 for scenario I versus III, and 0.910 and 0.873 for scenario II versus III. These results confirm that the null distribution of $\lambda_n(H_0)$ is independent of parameters $\beta$ and $\theta(t)$. Similar test results can be obtained when working independence is used.

*Wilks phenomenon under variance misspecification:* To better understand how the null distributions change when the variance is misspecified, we consider the same simulations as above but misspecify the variance as a constant. That is, the assumed working variance is $\sigma^2(\mu) = \sigma^2$ and the estimator is the mean squared error of the residuals. Figure 4.7 displays the empirical distributions of the GQLR test statistics under the three scenarios using constant variance estimators and working independence correlations. We can clearly see the differences between these distributions, which indicates the failure of the Wilks phenomenon.




(a) Exchangeable working correlation




(b) Working independence

Figure 4.6: *Simulation 2: The empirical distributions of $\lambda_n(H_0)$ when the variance function is consistently estimated using a local linear estimator. Panel (a): an exchangeable correlation structure is assumed for the test, where the correlation parameter is estimated by the quasi maximum likelihood method described in Section . Panel (b): working independence is assumed for both estimation and test. (solid line: scenario I; dashed line: scenario II; dotted line: scenario III).*

The last two columns in Table 4.1 summarize the mean and SE of $\lambda_n(H_0)$ under a misspecified constant variance when the working correlation $\mathcal{C}(\boldsymbol{\tau})$ is identity. Again, we conduct the two-sample t-tests and the F-tests to check the significant differences between the three scenarios. The p-values for the t-tests are 0.897 for scenario I versus Ⅱ, 0.893 for I versus Ⅲ and 0.773 for Ⅱ versus Ⅲ. For the F-tests, the p-values are 0.02 for scenario I versus Ⅱ, 0.0001 for I versus Ⅲ and 0.1 for Ⅱ versus Ⅲ. These results confirm our theory that when the variance function is misspecified, the Wilks phenomenon does not hold in general.

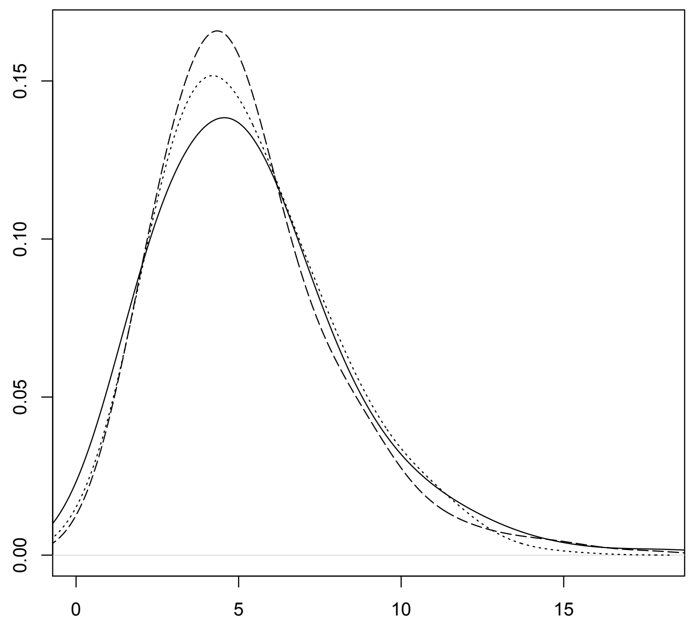


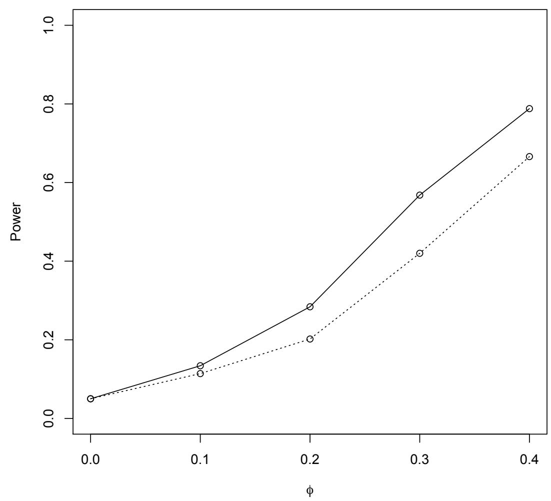

Figure 4.7: *Simulation 2: The empirical distribution of $\lambda_n(H_0)$ when the variance function is misspecified as a constant and the correlation is working independence. (solid line: scenario I; dashed line: scenario Ⅱ; dotted line: scenario Ⅲ)*

*Power of the GQLR tests:* To access the power of the GQLR test, we focus on the simulation setting described in scenario I and consider local alternatives with $\theta_1(t) = \theta_0(t) - \phi G(t)$ and $\theta_2(t) = \theta_0(t) + \phi G(t)$. We consider two cases: 1): $G(t) = \sin(t)$ with $\phi = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4$; and 2): $G(t) = \sin(3\pi t)$ with $\phi = 0,\ 0.05,\ 0.1,\ 0.15,\ 0.2$.

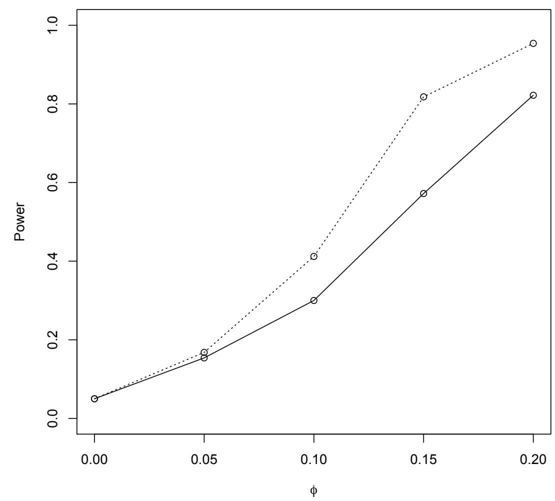
For each value of $\phi$, we generate 500 datasets from the full model and use local variance estimator as the working variance for both estimation and test. The true correlation is ARMA (1,1) as described before. We conduct the GQLR test under three different working correlation: working independence and true correlation, and set the significant level at $\alpha = 0.5$.

The power curves of the GQLR tests under the two settings of $G(t)$ and three correlation structures are depicted in the two panels of Figure 4.8. As we can see, the power of the GQLR tests gets higher as $\phi$ increases in all simulations. Interestingly, the simulation results suggest that using the true covariance does not always increase the power. When $G(t) = \sin(t)$, the proposed GQLR test based on working independence are more powerful than that using the true correlation, which is illustrated by panel (a) of Figure 4.8. On the other hand, from panel (b) of Figure 4.8, we show that using the true correlation results in a more powerful test than using working independence when $G(t) = \sin(3\pi t)$. These results demonstrate the power of our proposed GQLR test does depend on a complicated interaction between $G(t)$ and the correlation structure.

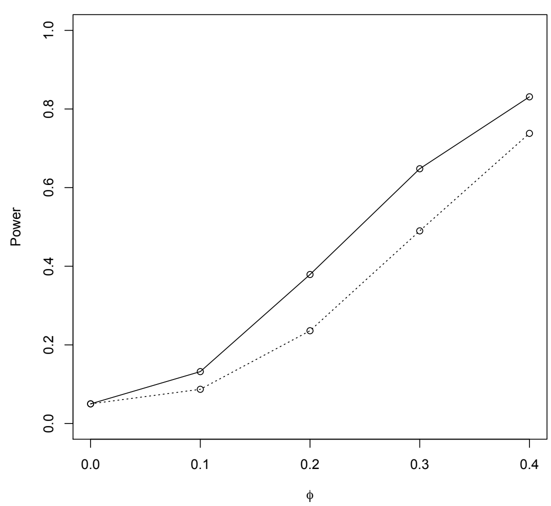
To validate this finding using our theoretical results, we also calculate the theoretical power of the test given in (3.9). The value of $\sigma_n$, $d_{2n}$, $\mu_n$ and $\mu_{1n}$ are estimated by replacing expectations with sample means. These estimated theoretical power curves under different choices of $G(t)$ and different working correlation structures are presented in panels (c) and (d) of Figure 4.8. The theoretical power curves are similar to the empirical ones and confirm that using the true correlation structure in the test does not necessarily increase the power of the GQLR test.



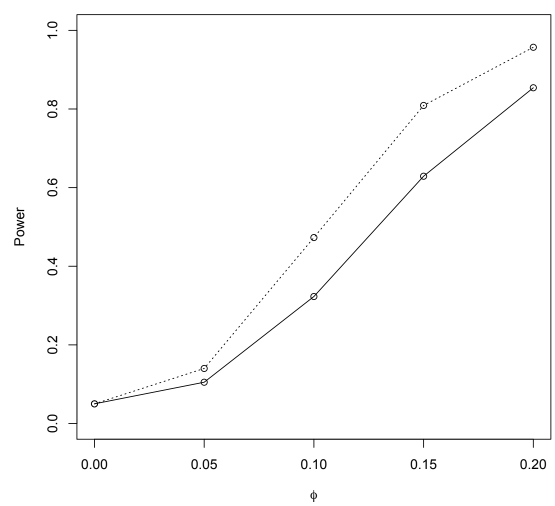

(a) Empirical power, $G(t) = \sin(t)$.




(b)Empirical power, $G(t) = \sin(3\pi t)$.




(c) Theoretical power, $G(t) = \sin(t)$.




(d) Theoretical power, $G(t) = \sin(3\pi t)$.

Figure 4.8: *Simulation 1: power of the GQLR test under the local alternative* $\theta_1(t) = \theta_0(t) - \phi G(t)$ *and* $\theta_2(t) = \theta_0(t) + \phi G(t)$*. The theoretical powers are calculated using equation (3.9). In each panel, the solid curve is the power under working independence and the dotted curve is the power when the true correlation is used.*

## 4.3 Simulation 3: Binary longitudinal data

In this simulation, we will illustrate the performance of the GQLR test for non-Gaussian data. We consider a binary longitudinal data with $k = 2$ treatment groups, each group has $n_k = 150$ subjects with $m = 4$ observations per subject. The response variable $Y_{k,ij}$ follows a marginal distribution of Binomial $(1, p_{k,ij})$ where

$$\text{logit}(p_{k,ij}) = \boldsymbol{X}_{k,ij}^{\mathrm{T}}\beta + \theta_k(T_{k,ij}). \tag{4.4}$$

We generate $T_{k,ij}$ as a random variable from a uniform distribution on [0, 1] and time dependent covariate $X_{k,ij}$ as the sums of $T_{k,ij}$ and a normal [0, 0.3] random variable, and assume an exchangeable within-subject correlation structure such that $\text{corr}(Y_{k,ij}, Y_{k,ij'}) = \rho_{j,j'} = 0.3$ for $j \neq j'$. To generate binary responses with the desired mean and correlation structure, we use a truncated Bahadur representation (Bahadur, 1961) ignoring expansions of order three and higher. Specifically, $\boldsymbol{Y}_{k,i}$ is generated from the following joint distribution

$$f(y_1, \ldots, y_m) = \left\{\prod_{j=1}^{m} p_j^{y_j}(1-p_j)^{1-y_j}\right\}\left\{1 + \sum_{1\leq j<j'\leq m} \rho_{jj'}\widetilde{y}_j\widetilde{y}_{j'}\right\},$$

where $p_j$ is the probability that $Y_j$ is equal to 1 and $\widetilde{y}_j = (y_j - p_j)/\sqrt{p_j(1-p_j)}$ is a standardized version of $y_j$.

To verify the Wilks results, we study the empirical distribution of the test statistic under the null hypothesis. We generate 300 datasets from each of the following three scenarios

$$\begin{aligned}
&\text{Scenario IV}: \beta = 0.5,\ \theta_0(t) = \frac{1}{2}\sin(\frac{3}{4}\pi t),\\
&\text{Scenario V}: \ \beta = -1,\ \theta_0(t) = \frac{1}{2}\sin(\frac{3}{4}\pi t),\\
&\text{Scenario VI}: \beta = 0.5,\ \theta_0(t) = \sin(\pi t) - 0.5.
\end{aligned}$$

For each simulated dataset, we estimate the variance by $\hat{p}_{k,ij}(1-\hat{p}_{k,ij})$, where $\hat{p}_{k,ij}$ is estimated as $\exp\{\boldsymbol{X}_{k,ij}^{\mathrm{T}}\widehat{\beta}_F+\widehat{\theta}_k(T_{k,ij})\}/[1+\exp\{\boldsymbol{X}_{k,ij}^{\mathrm{T}}\widehat{\beta}_F+\widehat{\theta}_k(T_{k,ij})\}]$. Epanechnikov kernel with a bandwidth $h=0.1$ is used. For binary responses, it is natural to use a binary quasi-likelihood for the test

$$\mathcal{Q}_{\text{binary}}(\boldsymbol{\mu},\boldsymbol{Y})=\sum_{j=1}^{m}Y_j\log\{\mu_j/(1-\mu_j)\}+\log(1-\mu_j). \tag{4.5}$$

In such a quasi-likelihood, it is difficult to incorporate within-cluster correlation. More importantly, some empirical evidence from our previous simulation shows, when working independence estimator is used, incorporating correlation into test does not always increase the power. Therefore, we focus on a working independence GQLR test using the quasi-likelihood in (4.5).

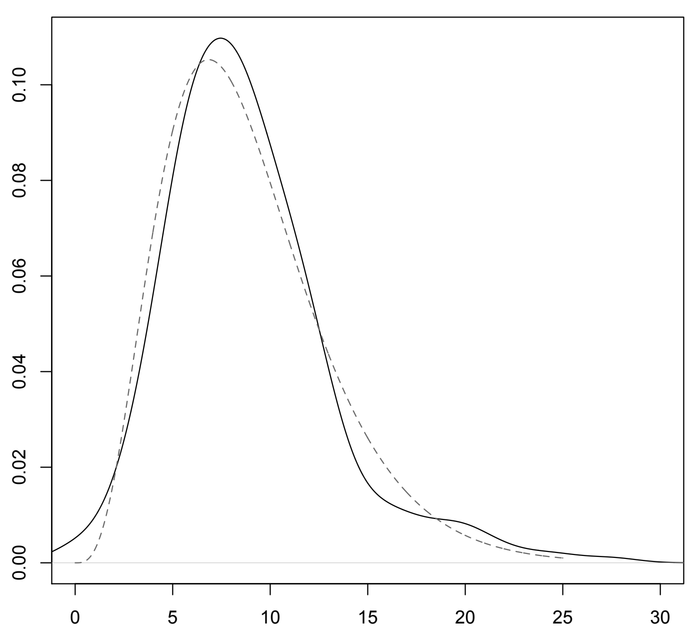



Figure 4.9: *Simulation 3: the empirical distribution of $r_K\lambda_n(H_0)$ (solid line) from scenario V under working independence and a density of the $\chi^2$ distribution with the degree freedom equaling the sample mean of $r_K\lambda_n(H_0)$ (dashed line).*

*Asymptotic distribution of $r_k\lambda_n(H_0)$ under working independence:* Figure 4.9 demonstrates the estimated density function of $r_K\lambda_n(H_0)$ from scenario IV under working independence assumption. As expected, it looks like the density of a $\chi^2$ distribution (dashed line) with the degrees of freedom equal to the sample mean of $r_k\lambda_n(H_0)$.

*GQLR tests on correlated and independent bootstrap samples under WI assumption:* As our theory shows, when the working independence used for both estimation and hypothesis test, the distribution of $\lambda_n(H_0)$ does not depend on the correlation structure, hence is identical to the case when the data are independent. Based on this result, we can simplify our bootstrap method in section 3.3.4 by simulating independent responses. To support this claim, we also compare the empirical distribution of $\lambda_n(H_0)$ to a case where the responses are truely independent. We generate 300 datasets from scenario VI with independent $Y_{k,ij}$.

The empirical distribution of $\lambda_n(H_0)$ under Scenarios IV - VI with correlated responses and that under Scenario VI with independent responses are shown in Figure 4.10. As we can see, the four distributions are almost identical. The means and standard errors of $\lambda_n(H_0)$ under different scenarios are displayed in Table 4.3. We performed the two-sample t-tests and F-tests to detect the differences for the correlated responses among the different scenarios. The p-values of t-tests and F-test are 0.558 and 0.37 for scenario IV vs V, 0.353 and 0.37 for scenario IV vs VI and 0.12 and 0.994 for scenario V vs VI. We also compared the distributions of $\lambda_n(H_0)$ under the correlated and independent responses for scenario VI. The p-values of t-tests and F-test are 0.585 and 0.731. We conclude that these four distributions are almost the same, which also corroborates our theory and the proposed bootstrap procedure.

Table 4.2: Simulation 3: means and standard error (SE) of $\lambda_n(H_0)$ for binary longitudinal data under scenario IV - VI and those of $\lambda_n(H_0)$ under scenario VI and independent response.

| | Correlated data | | | Indepadent data |
|---|---|---|---|---|
| | Scenario IV | Scenario V | Scenario VI | Senario VI |
| Mean | 5.87 | 5.99 | 5.68 | 5.57 |
| SE | 2.57 | 2.442 | 2.441 | 2.49 |

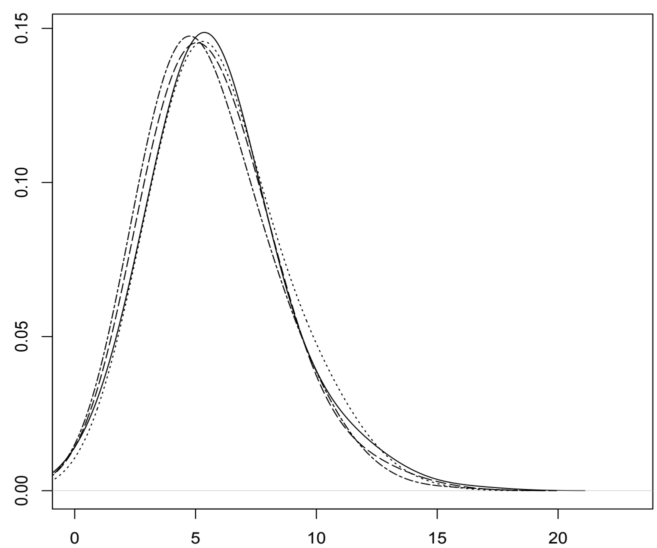



Figure 4.10: *Simulation 3: empirical distributions of $\lambda_n(H_0)$ for correlated datasets (scenario IV: solid line; scenario V: dotted line; scenario VI: dashed line) and independent datasets (scenario VI: dotdash line) using binary likelihood function under WI assumption.*

*Wilks Phenomenon under misspecified correlation structure:* We now examine the null distributions of $\lambda_n(H_0)$ using a misspecified AR(1) correlation under three different scenarios. Note that, the quasi-likelihood equation in (4.5) can not be used when taking the within-subject correlation into account. Since the quasi-likelihood used in the test statistic does not have to be a real likelihood, we use a Gaussian quasi-likelihood in (4.3) on binary longitudinal data. Simulation results are presented in Table 4.3, which are similar to those in the Gaussian cases. With the working correlation being AR(1), there are only small differences in both means and variances of $\lambda_n(H_0)$ among the three scenarios. We also conduct the t-tests and F-tests to determine if these differences are significant. It turns out that no significant differences are detected based on the p-values. Figure 4.11 also displays the empirical distributions of $\lambda_n(H_0)$ under the three scenarios.

Table 4.3: Simulation 3: means and standard error of $\lambda_n(H_0)$ for binary longitudinal data under scenario IV - VI with the working correlation being AR(1)

| Working correlation | AR(1) | | |
|---|---|---|---|
| | Scenario IV | Scenario V | Scenario VI |
| Mean | 4.92 | 4.65 | 4.73 |
| SE | 2.54 | 2.41 | 2.33 |

SE stand for standerd error.

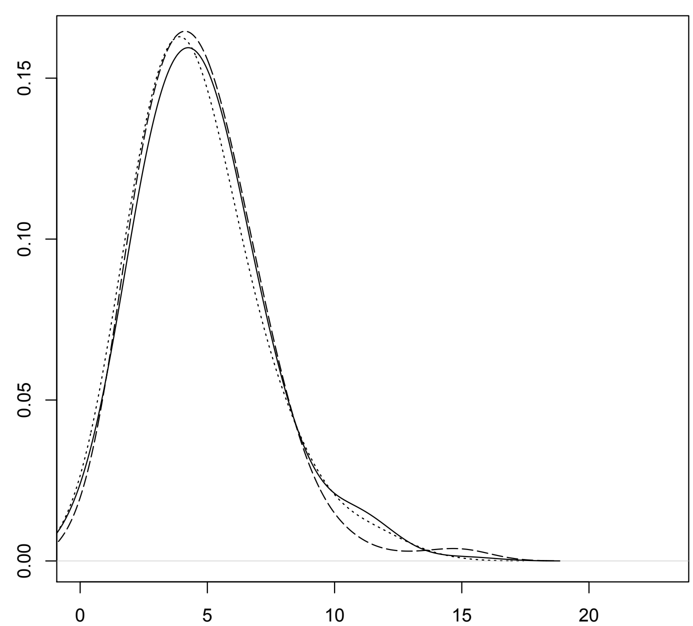



Figure 4.11: *Simulation 3: the estimated kernel density functions of $\lambda_n(H_0)$ using the Gaussian quasi-likelihood and a misspecified AR(1) correlation (solid line: scenario IV; dashed line: scenario V; dotted line: scenario VI).*

*Power of the GQLR tests:* As in the previous simulations, we focused on scenario IV, fixing $\theta_1(t)$ at $\theta_0(t)$ and changing $\theta_2(t)$ to $\theta_{2,\phi}(t)$ where

$$\theta_{2,\phi}(t) = \frac{1}{2} sin(\frac{3}{4}\pi t) + \phi exp(\frac{t}{2}), \quad \phi = 0.1, 0.2, 0.3, 0.4, 0.5.$$

According to our discussion above, the GQLR test under woking independence is easy to conduct in practice, and avoids computational issues in generating longitudinal binary data. Therefore, for each value of $\phi$, we generated 500 datasets from the model and ignore the within-subject correlation in both estimation and hypothesis test. The power is 0.068 for $\phi = 0.1$, 0.128 for $\phi = 0.2$, 0.385 for $\phi = 0.3$, 0.557 for $\phi = 0.4$ and 0.593 for $\phi = 0.5$, as shown in Figure 4.13. This is not surprising, the results confirms that the GQLR test is more powerful as $\phi$ increases.

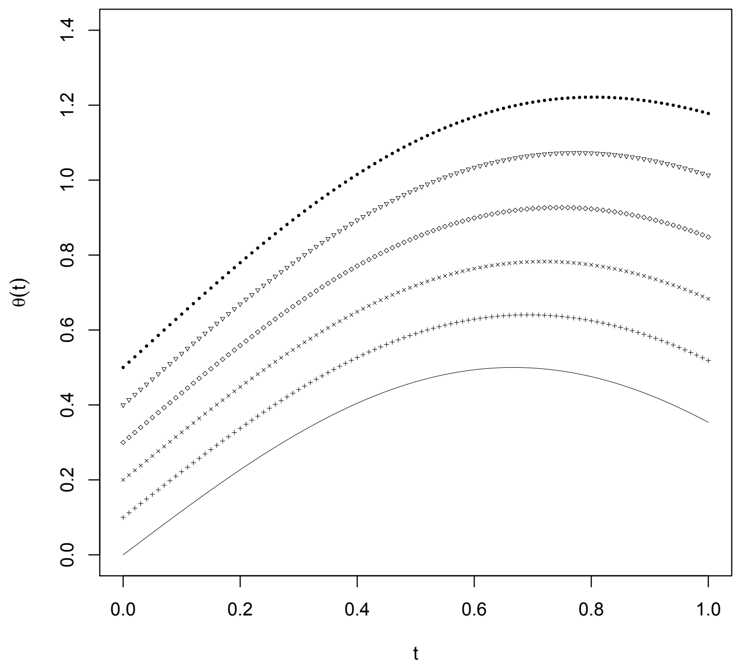



Figure 4.12: *Simulation 3: plot of* $\theta_{2,\phi}(t)$*:* — $\phi = 0$, $+\!+\!+ : \phi = 0.1$, $\times\times\times : \phi = 0.2$, $\diamond\diamond\diamond : \phi = 0.3$, $\triangleright\triangleright\triangleright : \phi = 0.4$, $\cdots : \phi = 0.5$.

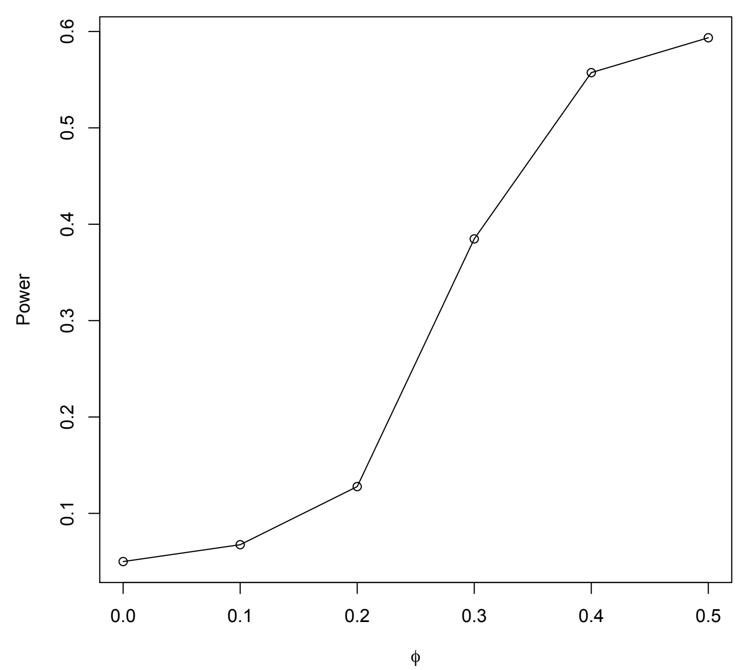



Figure 4.13: *Simulation 3: power of the GQLR tests for binary datasets, significance level is* $\alpha = 0.05$.

# Chapter 5

# Real Data Analysis

## 5.1 Application to CD4 count data

In this section, we perform the quasi-likelihood ratio test on repeated CD4 (cluster of differentiation 4) count data from AIDS Clinical Trial Group 193A Study Team. This data set is from a randomized, double-blind study of AIDS patients with advanced immune suppression. The patients in this study had CD4 counts less than or equal to 50 cells/$mm^3$, and were randomized to one of four daily regimens containing 600mg of zidovudine: treatment 1 is zidovudine alternating monthly with 400mg didanosine; treatment 2 is zidovudine plus 2.25mg of zalcitabine; treatment 3 is zidovudine plus 400mg of didanosine; treatment 4 is zidovudine plus 400mg of didanosine and 400mg of nevirapine.

Measurements of CD4 counts were scheduled to be collected at baseline and at 8-week intervals during 40 weeks of follow-up. The number of measurements of CD4 counts varied from 1 to 9, with a median of 4. There are totally 1309 patients enrolled in this study, including 162 females and 1147 males. After eliminating 122 patients who dropped out immediately after the baseline measurement, 1044 males and 143 females are used for our analysis. The response variable is the log transformed CD4 counts, log(CD4 counts + 1), and the covariates are age(years), gender (1=M, 0=F) and measurement time (weeks). Figure 5.1 shows the scatter plot of log CD4 count versus observation times.

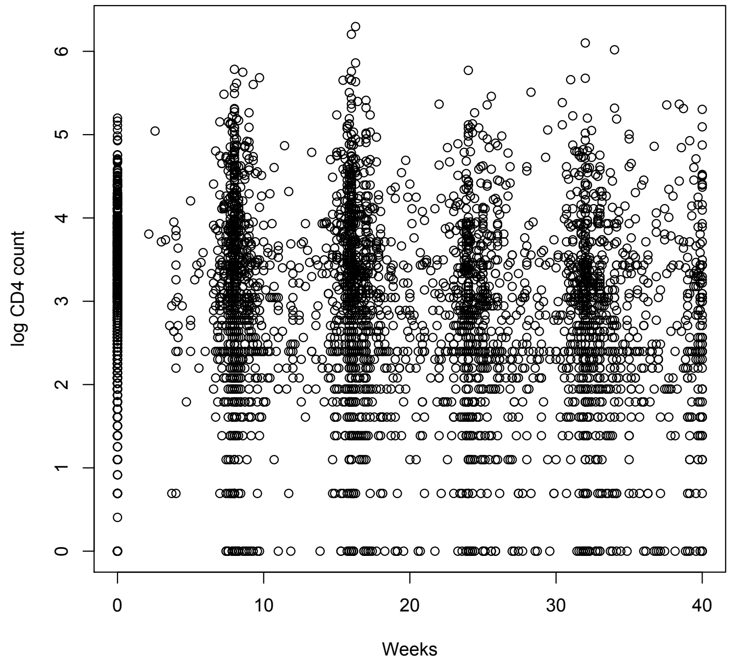



Figure 5.1: *Log CD4 count over time*

To apply the semiparametric model (1.1), denote $Y_{k,ij}$ as the log CD4 count for the $j$th visit of the $i$th patient in the $k$th treatment group, and $\boldsymbol{X}_{k,ij}$ as the 2-dimentional covariate vector consisting of age and gender. The time variable $T_{k,ij}$ is a scaler, and $\theta_k(t)$ is the treatment effect for the $k$th group. Assume that the within-subject covariance structure is the same for all subjects, the variance is a smooth function of time $T$, and the working correlation is ARMA(1,1). We chose the bandwidth $h$ by the generalized cross-validation criterion. Figure 5.2 shows the estimated smooth functions $\theta_k(t)$ for the four treatment groups. It is obvious to see the differences between treatment groups. All treatment groups have almost the same mean CD4 count at the baseline. As time goes on in the follow up,

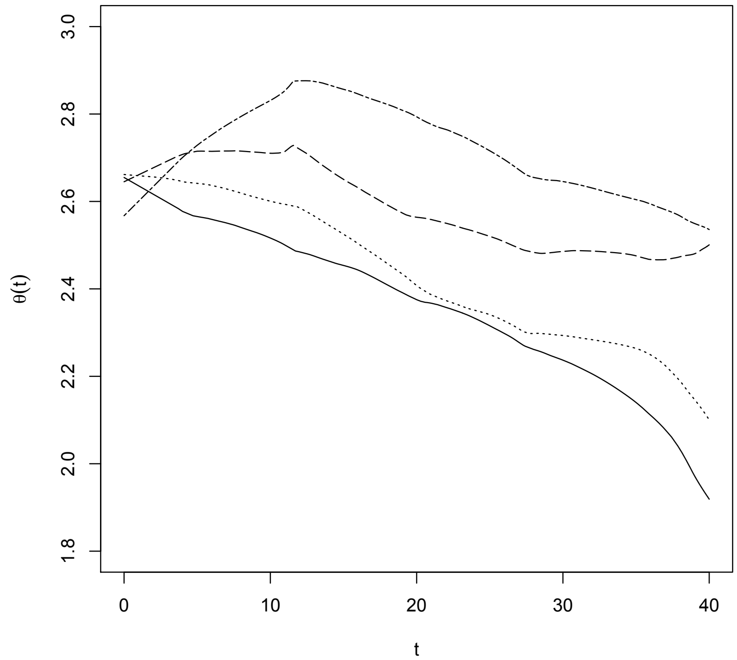



Figure 5.2: *The estimated time effects $\theta_k(t)$: treatment 1 (solid line); treatment 2 (dotted line); treatment 3(dashed line); treatment 4 (dot dash line)*

the CD4 count drops monotonically for treatment 1, while it increases during first 12 weeks and drops thereafter for treatment 4.

Consider the nonparametric hypothesis test:

$$H_0 : \theta_1(t) = \cdots = \theta_4(t) \quad vs. \quad H_1 : \textit{not all } \theta_k' s \textit{ are the same}, \tag{5.1}$$

The log-likelihood function for this data is

$$\ell(\beta, \theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_k \log|\mathbf{V}_k| - \frac{1}{2}\sum_k \{\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta} - \theta_k(\boldsymbol{T}_k)\}^T \mathbf{V}_k^{-1}\{\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta} - \theta_k(\boldsymbol{T}_k)\},$$

where $\boldsymbol{Y}_k$, $\boldsymbol{X}_k$ and $\mathbf{V}_k$ are the response vector and covariate matrix and working covariance matrix in the $k$th treatment group. Denote the maximum log-likelihood under $H_0$ and $H_1$

as $\widehat{\ell}_R$ and $\widehat{\ell}_F$ respectively. Then, the generalized likelihood ratio statistic is

$$\begin{aligned}\lambda_n(H_0) &= \widehat{\ell}_F - \widehat{\ell}_R \\ &= \frac{1}{2}\sum_{k=1}^{q}\{\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_R - \widehat{\theta}_R(\boldsymbol{T}_k)\}^T \mathbf{V}_k^{-1}\{\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_R - \widehat{\theta}_R(\boldsymbol{T}_k)\} \\ &\quad -\frac{1}{2}\sum_{k=1}^{q}\{\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_F - \widehat{\theta}_{F,k}(\boldsymbol{T}_k)\}^T \mathbf{V}_k^{-1}\{\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_F - \widehat{\theta}_{F,k}(\boldsymbol{T}_k)\}. \quad (5.2)\end{aligned}$$

We calculate the $p$ value of GQLR test based on the bootstrap procedure described in section 3.3, and the $p$ value based on 500 bootstrap samples is $< 0.002$. Therefore, we conclude that there are significant differences among four treatment groups.

## 5.2 Application to Opioid Agonist Treatment data

We now illustrate the application of our proposed GQLR test to a data set on opioid agonist treatment. The data involved 140 patients who received primary care-based buprenorphine, which is a commonly prescribed medication for treating opioid dependence, at the Primary Care Center of Yale-New Haven Hospital. Each patient first went through a two-week induction and stabilization period and then was prescribed with daily medication of buprenorphine for 24 weeks. Prior research has shown that adding counseling to buprenorphine can help increase opioid abstinence rate (Amato et al., 2011). The main objective of the study was to investigate the impact of adding cognitive behavioral therapy, which is a counseling intervention with demonstrated efficacy for a variety of psychiatric conditions and substance use disorders (Crits-Christoph et al., 1999; Beck, 2005; Butler et al., 2006; McHugh et al., 2010), to the efficacy of primary case-based buprenorphine to treat opioid dependence. The patients were randomly assigned to receive one of two treatments: physician management (PM) or physician management and cognitive behavioral therapy (PMCBT). Physician management was provided in the form of 15- to 20-minute sessions by internal medicine physicians who had experience with providing buprenorphine but had no training in cognitive behavioral therapy. These sessions were given weekly for the first two weeks, every two weeks for the next four weeks, and then monthly afterward. Patients in the PMCBT group were offered the additional opportunity to participate in up to twelve fifty-minute weekly cognitive behavioral therapy sessions during the first twelve weeks of treatment. All counseling sessions were given by well-trained masters and doctoral-level clinicians. The main components of counseling focused on developing behavioral skills such as promoting behavioral activation and identifying and coping with opioid craving.

Illicit opioid use was measured weekly by both self-reported frequency of opioid use and urine toxicology testing. The latter was conducted with the use of a semiquantitative homogeneous enzyme immunoassay for opioids and other substances such as cocaine and oxycodone. The accuracy of self-reported opioid use can be questionable. As a result, we considered only the urine data. The time points when the urine testings were done were unbalanced and irregular, because the patients did not provide urine samples on a strict weekly basis. Some of these subjects also had follow-up measurements going up to 195 days. The number of observations per patient is between 1 and 27, with a median of 24. The covariates we use include age, gender (1=female / 0=male) and the highest level of education completed (1= High School or Higher and 0= otherwise); the time variable is day with the range from day 0 to day 195. The response variable is urine toxicology testing result (1=positive / 0= negative).

Table 5.1: Summary statistics of covariates in Opioid Agonist Treatment data

| Covariates | Mean | SE | Median | Min | Max |
|---|---|---|---|---|---|
| age | 33.90 | 9.54 | 33.44 | 18.11 | 62.57 |
| Gender | Frequency | Percent | Education | Frequency | Percent |
| 1 | 37 | 0.264 | 1 | 118 | 0.843 |
| 0 | 103 | 0.736 | 0 | 22 | 0.157 |

A total of 140 patients are involved in our analysis, 69 patients are in PM group and 71 in PMCBT group. We analyze the dataset via the following logistic model

$$\text{logit}\{\Pr(Y_{k,ij}=1)\} = \boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta_k(T_{k,ij}), \tag{5.3}$$

where $\theta_k(T_{k,ij})$ is a smooth function of the time variable for the $k$th group.

We first employ a K-fold cross-validation method to select the bandwidth. We randomly partition the original data into $\iota = 6$ groups. For the $\iota$th group of data denoted as $\mathcal{G}_\iota$, $\iota = 1, \cdots, 6$, we fit the model (5.3) to the remaining 5 groups and compute the fitted value of $\boldsymbol{Y}_i$ $(i \in \mathcal{G}_\iota)$ defined as $\widehat{\mu}_{-\iota}(\boldsymbol{X}_i, \boldsymbol{T}_i)$. This leads to the cross-validation criterion

$$CV(h) = \sum_{\iota}\sum_{i\in\mathcal{G}_\iota}\{Y_i - \widehat{\mu}_{-\iota}(\boldsymbol{X}_i, \boldsymbol{T}_i)\}^2, \tag{5.4}$$

where $\widehat{\mu}_{-\iota}(\boldsymbol{X}_i, \boldsymbol{T}_i) = \text{logit}^{-1}\{\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{-\iota} + \widehat{\boldsymbol{\theta}}_{-\iota}(\boldsymbol{T}_i)\}$, and $\widehat{\boldsymbol{\beta}}_{-\iota}$ and $\widehat{\boldsymbol{\theta}}_{-\iota}(\boldsymbol{T}_i)$ are the estimates obtained without including data from $\mathcal{G}_\iota$.

The covariance of $\widehat{\boldsymbol{\beta}}$ is estimated using the sandwich formula

$$\begin{aligned}\text{cov}(\widehat{\boldsymbol{\beta}}) \;=\; & \left\{\sum_{k}^{q}\widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\Delta_k\mathcal{W}_k^{-1}\Delta_k\widetilde{\boldsymbol{X}}_k\right\}^{-1}\left\{\sum_{k=1}^{q}\widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\Delta_k\mathcal{W}_k^{-1}(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)^{\mathrm{T}}\right.\\ & \left.\times\mathcal{W}_k^{-1}\Delta_k\widetilde{\boldsymbol{X}}_k\right\}\left\{\sum_{k}^{q}\widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\Delta_k\mathcal{W}_k^{-1}\Delta_k\widetilde{\boldsymbol{X}}_k\right\}^{-1},\end{aligned}$$

where $\boldsymbol{\mu}_k = \text{logit}^{-1}\{\boldsymbol{X}_k\widehat{\boldsymbol{\beta}}_F + \widehat{\boldsymbol{\theta}}_{F,k}(\boldsymbol{T}_k)\}$ and $\widetilde{\boldsymbol{X}}_k = \boldsymbol{X}_k - \boldsymbol{\mu}_{X,k}(\boldsymbol{T}_k)$. To calculate the $\widetilde{\boldsymbol{X}}_k$, we need to estimate $\boldsymbol{\mu}_{X,k}(t)$. A consistent estimator of $\boldsymbol{\mu}_{X,k}(t)$ is

$$\left[\sum_i \{\mu_{k,ij}^{(1)}\}^2 \omega_{k,i}^{jj} K_h(T_{k,ij} - t)\boldsymbol{X}_{k,ij}\right]\left[\sum_i \{\mu_{k,ij}^{(1)}\}^2 \omega_{k,i}^{jj} K_h(T_{k,ij} - t)\right]^{-1}.$$

Table 5.2: Regression coefficient estimates in analysis of Opioid Agonist Treatment data.

| Parameter | Estimate | Standard error | P-value |
|---|---|---|---|
| Age | -0.022 | 0.014 | 0.116 |
| Gender | 0.345 | 0.270 | 0.201 |
| Education | 0.791 | 0.350 | 0.024 |

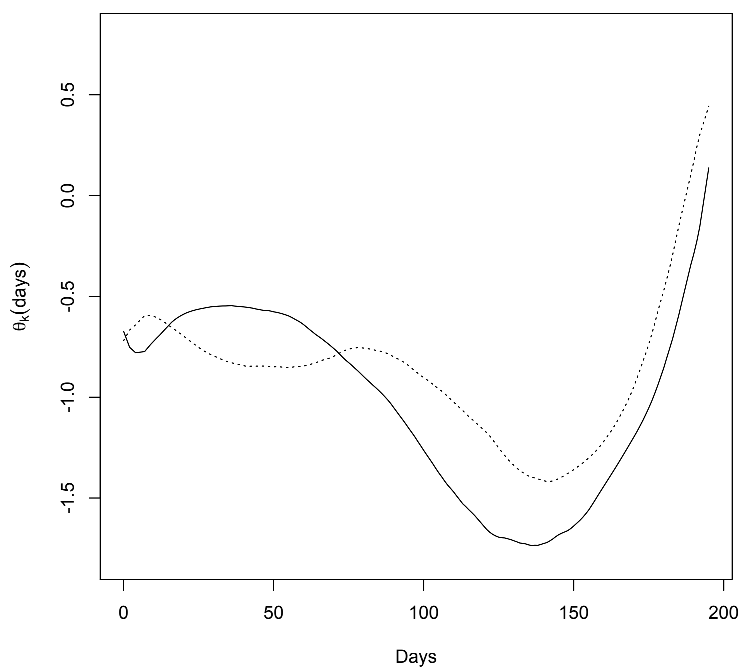



Figure 5.3: *The estimated time effect $\theta_k(t)$: PMCBT treatment (solid line); PM treatment: (dotted line)*

Table 5.2 presents the estimates and standard errors of the regression coefficients $\boldsymbol{\beta}$ under the full model. The variable education has the smallest p-value of less than 0.05, which indicates education makes a significant contribution to our model. Furthermore, The estimated mean curves for the two treatment groups are presented in Figure 5.3. At the beginning of the treatment, the estimated curves of the two groups are almost the same. After about 70 days, patients in the PMCBT group have a lower probability of opioid use. Near the end of treatment, both groups have increased opioid use rate, but patients with additional cognitive behavioral therapy seemed to have lower overall rate of use. In addition, the spaghetti plots of the estimated probability of opioid positive urines versus times for PM treatment and PMCBT treatment are illustrated in Figure 5.4.

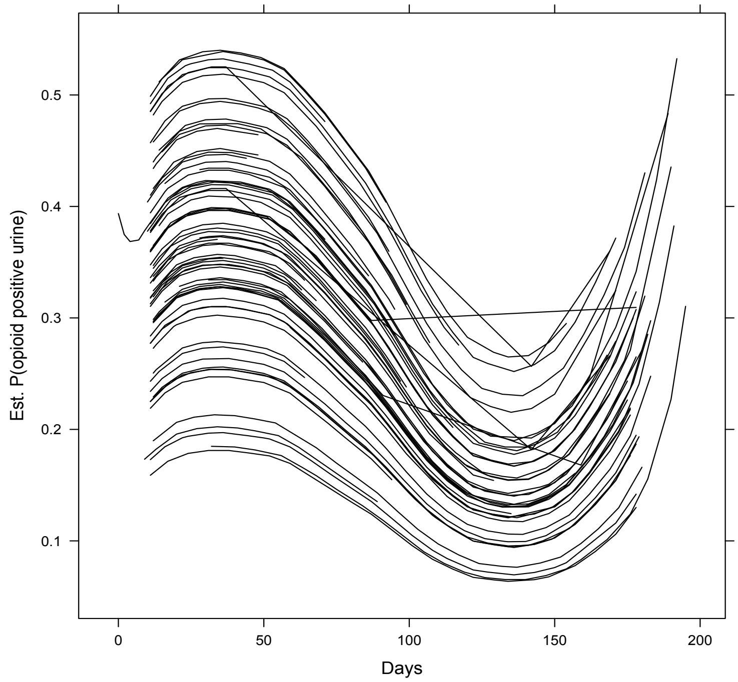
Our primary interest is to evaluate the impact of adding cognitive behavioral therapy to physician management. In other words, our interest is to test the nonparametric hypothesis

$$H_0 : \theta_1(t) = \theta_2(t) \quad vs. \quad H_1 : \theta_1(t) \neq \theta_2(t). \tag{5.5}$$

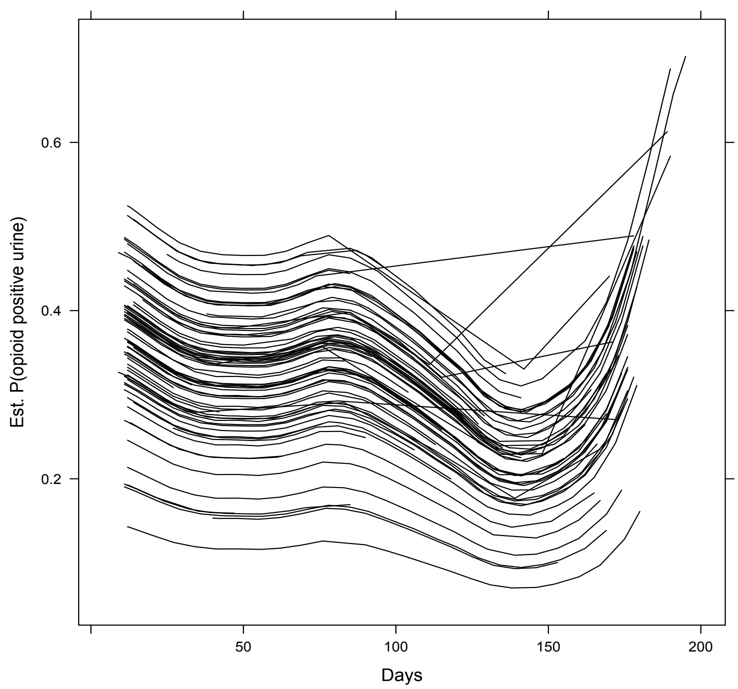
We apply the proposed GQLR test based on working independence correlation and quasi-likelihood function

$$\mathcal{Q} = \sum_{k,i} \sum_{j} log[1 + exp\{\boldsymbol{X}_{k,ij}\boldsymbol{\beta} + \theta_k(T_{k,ij})\}] + Y_{k,ij}\{\boldsymbol{X}_{k,ij}\boldsymbol{\beta} + \theta_k(T_{k,ij})\}. \tag{5.6}$$

The p-value of the test is 0.023 based on 1000 bootstrap replicates. We therefore conclude that there is a significant difference between the two treatment groups. In particular, Figure 5.3 suggests that cognitive behavioral therapy improved the opioid abstinence over time. Fiellin et al. (2013) analyzed the same data, using self-reported frequency of opioid use and the maximum number of consecutive weeks of abstinence from illicit opioids in the two 12-week periods as the primary outcome measures, but did not find any evidence supporting




(a) PMCBT treatment group




(b) PM treatment group

Figure 5.4: *The estimated probability of opioid positive urines versus observation time (in days) for each patient*

the benefit of adding cognitive behavioral therapy. Their outcome measures are aggregated and capture only certain features of the data. In contrast, our analysis is based on the entire longitudinal trajectory and hence more powerful in detecting the difference between the two treatments.

# Chapter 6

## Summary

We investigate a class of semiparametric analysis of covariance models for generalized longitudinal data, where the treatment effects are represented as nonparametric functions over time and covariates are incorporated to account for the variability caused by confounders. We propose to test the treatment effects using a generalized quasi-likelihood ratio test. Our theoretical study reveals that when the variance structure is correctly specified, the asymptotic distribution of GQLR test statistic does not depends on the parameters, but does depend on the true and working correlation structures. When the variance is mis-specified, the Wilks phenomenon might completely fail in that the distribution of the test statistic depends on all nuisance parameters. In particular, when working independence is assumed in both estimation and test and when the variance structure is correctly specified, the much celebrated Wilks phenomenon known to hold for independent data also holds for longitudinal data. Replacing the true variance with a consistent estimator, such as the nonparametric estimator based on local polynomial, only causes an asymptotically negligible error to the test. We have also shown that GQLR test assuming working independence yields the minimax optimal power rate.

It is a common practice to evaluate the null distribution by bootstrap. For Gaussian data, one can take residuals of the full model and resample the entire clusters within each treatment group. Such a procedure can preserve the correlation structure of the real data. However, such a procedure can not be used for non-Gaussian data, such as binary data, since the residuals are not the same type of data any more. A parametric bootstrap is also difficult to implement since generating non-Gaussian longitudinal data with the same correlation structure as the real data is challenging. For non-Gaussian data, the GQLR test assuming working independence is particularly appealing, since the asymptotic distribution of the test statistic does not depend on the correlation structure and one can simulate the null distribution using independent samples.

Our procedure is based on the working independence estimators of Lin and Carroll (2001), and it is easy to implement. More complicated but also more efficient estimators were proposed in Wang et al. (2005) and Lin and Carroll (2006). Tests based on those estimators might improve the power of our test by a fraction, but can not improve the rate of the power. However, how to incorporate correlation into the test statistics and how to implement bootstrap for non-Gaussian correlated data remains unclear and calls for future research.

# Chapter 7

# References

[1] Amato L, Minozzi S, Davoli M, Vecchi S. (2011). Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence, *Cochrane Database Syst Rev.*, **10**: CD004147

[2] Bahadur, R. R. (1961). A representation of the joint distribution of responses to $n$ dichotomous items, *Studies in Item Analysis and Prediction*, Ed. H. Solomon, Stanford University Press, Stanford, California,158-168.

[3] Beck, A. T. (2005). The current state of cognitive therapy: a 40-year retrospective, *Arch Gen Psychiatry*, **62**, 953-959.

[4] Brumback, B., and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion), *Journal of the American Statistical Association*, **93**, 961-994.

[5] Butler, A. C., Chapman, J. E., Forman, E. M., Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses, *Clin Psychol Rev.*, **26**, 17-31.

[6] Crits-Christoph, P., Siqueland, L., Blaine, J., et al. (1999). Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study, *Arch Gen Psychiatry*, **56**, 493-502

[7] De Jong, P. (1987). A central limit theorem for generalized quadratic forms, *Probability Theory and Its Related Fields*, **75**, 261-277.

[8] Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *American Statistician*, **45**, 302304.

[9] Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function, *Journal of the American Statistical Association*, **102**, 632-641.

[10] Fan, J. and Jiang, J.,(2005). Nonparametric inferences for additive models, *Journal of the American Statistical Association*, **100**, 890-907.

[11] Fan, J. and Jiang, J. (2001). Variable selection via non concave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348-1360.

[12] Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests (with discussion), *Test*, **16**, 409-478.

[13] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data, *Journal of the American Statistical Association*, **99**, 710-723.

[14] Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data, *Journal of the American Statistical Association*, **103**, 1520-1533.

[15] Fan, J., Zhang, C. and Zhang, J., (2001). Generalized likelihood ratio statistics and wilks phenomenon, *Annals of Statistics*, **29**, 153-193.

[16] Fan, J., and Zhang, J. (2000). Two-step estimation of functional linear models with applications to longitudinal data, *Journal of the Royal Statistical Society, Series B*, **62**, 303-322.

[17] Fiellin, D. A., Barry, D. T., Sullivan, L. E., Cutter, C. J., Moore, B. A., O'Connor, P. G., MPH, Schottenfeld, R. S. (2013). A Randomized Trial of Cognitive Behavioral Therapy in Primary Care-based Buprenorphine, *The American Journal of Medicine*, **126**, 74.e11-74.e17.

[18] Hall, P., Müller, H. -G. and Wang, J. -L. (2006). Properties of principal component methods for functional and longitudinal data analysis, *Annals of Statistics*, **34**, 1493-1517.

[19] Hall, P., Müller, H. G. and Yao, F. (2008). Modeling sparse generalized longitudinal observations with latent Gaussian processes, *Journal of the Royal Statistical Society, Series B*, **70**, 703-23.

[20] He, X., Fung, W. K., and Zhu, Z. Y. (2005). Robust estimation in generalized partial linear models for clustered data, *Journal of the American Statistical Association*, **472**, 1176-1184.

[21] He, X., Zhu, Z. Y., and Fung, W. K. (2002). Estimating in a semiparametric model for longitudinal data with unspecified dependence structure, *Biometrika*, **89**, 579-590.

[22] Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006), Covariance selection and estimation via penalized normal likelihood, *Biometrika*, **93**, 85-98.

[23] Hunsberger, S. (1994). Semiparametric regression in likelihood-based models, *Journal of the American Statistical Association*, **89**, 1354-1365.

[24] Ingster, Y. I. (1993), Asymptotic Minimax Hypothesis Testing for Nonparametric Alternatives IIII, *Mathematical Methods in Statistics*, 2, 85 - 114; 3, 171 - 189; 4, 249 - 268.

[25] Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13-22.

[26] Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations, *Journal of the American Statistical Association*, **96**, 1045-1056.

[27] Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society, Series B*, **68**, 69-88.

[28] Li, R., and Liang, H. (2008). Variable selection in semiparametric regression modeling, *Annals of Statistics*, **36**, 261-286.

[29] Li, Y. (2011). Efficient semiparametric regression in longitudinal data with nonparametric covariance estimation, *Biometrika*, **98**, 2, 355-370.

[30] McCullagh, P. (1983). Quasi-likelihood functions, *The annals of Statistics*, **11**, 59-67.

[31] McCullagh, P., and Nelder, J. A. (1989). Generalized linear models (2nd ed.), London: Chapman and Hall.

[32] McHugh, P. K., Hearon, B. A., Otto, M. W. (2010). Cognitive behavioral therapy for substance use disorders, *Psychiatr Clin North Am.*, **33**, 511-525.

[33] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models, *Journal of the Royal Statistical Society, Series B*, **68**, 179-199.

[34] Park, C.G., Park, T. and Shin, D.W. (1996). A simple method for generating correlated binary variates, *American Statistician*, **50**, 306310.

[35] Severini, T. A. and Staniswalis, J.G. (1994). Quasi-likelihood estimation in semiparametric models, *Journal of the American Statistical Association*, **89**, 501-511.

[36] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288.

[37] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation, *Biometrika*, **90**, 43-52.

[38] Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data, *Journal of the American Statistical Association*, **100**, **469**, 147–57.

[39] Wu, W. B., and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data, *Biometrika*, **90**, 831-844.

[40] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear meddles, and the Gauss-Newton method, *Biometrika*, **61**, 439-447.

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

[42] Zeger, S. L., and Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689-699.

[43] Zhou, L., Huang, J., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data, *Journal of the American Statistical Association*, **105**, 390-400.