Received: 7 January 2021 | Revised: 24 September 2021 | Accepted: 7 January 2022

DOI: 10.1111/sjos.12583

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Minimax powerful functional analysis of covariance tests with application to longitudinal genome-wide association studies

Weicheng Zhu[1] | **Sheng Xu**[2] | **Catherine C. Liu**[3] | **Yehua Li**[4]

[1]Amazon Inc., Seattle, Washington, USA

[2]Global Statistics and Data Science, BeiGene Co., Ltd., Beijing, China

[3]Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

[4]Department of Statistics, University of California at Riverside, Riverside, California, USA

Correspondence
Yehua Li, Department of Statistics, University of California at Riverside, Riverside, CA 92521, USA.
Email: yehuali@ucr.edu

Abstract

We model the Alzheimer's disease-related phenotype response variables observed on irregular time points in longitudinal Genome-Wide Association Studies as sparse functional data and propose nonparametric test procedures to detect functional genotype effects while controlling the confounding effects of environmental covariates. Our new functional analysis of covariance tests are based on a seemingly unrelated kernel smoother, which takes into account the within-subject temporal correlations, and thus enjoy improved power over existing functional tests. We show that the proposed test combined with a uniformly consistent nonparametric covariance function estimator enjoys the Wilks phenomenon and is minimax most powerful. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database, where an application of the proposed test lead to the discovery of new genes that may be related to Alzheimer's disease.

KEYWORDS

functional data, GWAS, hypothesis testing, kernel smoothing, longitudinal data, minimax power

1 | INTRODUCTION

Genome-wide association studies, GWAS, have been successfully used to associate diseases or traits with genetic variants defined by Single Nucleotide Polymorphisms (SNPs) (Visscher et al., 2017). A commonly used approach is to perform SNP-level hypothesis tests with multiple comparison adjustments (Fadista et al., 2016). The vast majority of the GWAS literature focuses on analyzing phenotypes measured at a single time, however, in many aging studies phenotypes are repeatedly measured over years where the measurement times are irregular and subject specific. One example of such studies comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI), where both longitudinal Alzheimer phenotypes and SNP-level genotypes are available for all subjects. The longitudinal phenotype responses can be naturally modeled as functional data (Ramsay & Silverman, 2005) and it is of scientific interest to test if the mean phenotype trajectories differ across different genotypes.

Let $Y_{k,i}(t)$ be the phenotype of the ith subject in the kth genotype group, observed at time $t \in \mathcal{T}$, $i = 1, \dots, n_k$, $k = 1, \dots, q$, where $\mathcal{T}$ is a closed time interval and q is the number of genotypes. Denote $n = \sum_{k=1}^{q} n_k$ as the total sample size. Suppose $\boldsymbol{X}_{k,i}(t)$ is a p-dimensional subject-specific covariate vector which represents confounding environmental effects and can be time dependent. Denote $\mu_{k,i}(t) = E\left\{Y_{k,i}(t) | \boldsymbol{X}_{k,i}(t)\right\}$ and assume

$$g\{\mu_{k,i}(t)\} = \boldsymbol{X}_{k,i}^{\mathrm{T}}(t)\boldsymbol{\beta} + \theta_k(t), \tag{1}$$

where $g(\cdot)$ is a known monotonic and differentiable link function, $\boldsymbol{\beta}$ is a p-vector of unknown coefficients and $\theta_k(\cdot)$ is an unknown smooth function representing the mean trend of the phenotype in the kth genotype group. In the ADNI data, one of the most important Alzheimer-related phenotypes is the hippocampal volume, the decay of which is known to be related to memory loss (Schuff et al., 2009), and the genotypes are AA, AB, or BB defined by the two alleles of a SNP.

Model (1) is closely related to functional analysis of variance models since the treatment effect for genotype k is represented by a nonparametric function $\theta_k(t)$. Some recent literature on functional analysis of variance models under various designs includes (Brumback & Rice, 1998; Zhou et al., 2010; Xu et al., 2018). Unlike most papers on semiparametric regression problems, which focus on inference on the parametric component (Wang et al., 2005), the parameter $\boldsymbol{\beta}$ in (1) is mainly used to control the confounding effects of the covariates, and our primary interest is to make inference on the functional genotype effects θ_k's. Specifically, we are interested in testing the following nonparametric hypotheses

$$H_0 : \theta_1 = \cdots = \theta_q \quad \text{versus} \quad H_1 : \text{not all } \ \theta_k' s \ \text{ are the same.} \tag{2}$$

In the ADNI example, a SNP is disease related if $\theta_k(t)$'s are different across the genotypes defined by the SNP.

Most existing work on functional analysis of variance tests consider dense functional data with Gaussian-type responses, where observations on each curve are made on a dense grid. A good summary of these methods are provided by Zhang (2013) (Reimherr and Nicolae (2014) and Huang et al. (2017) also applied similar test procedures in genetic studies. There are a few common restrictions of these methods. First, their test statistics were based on the integrated square error rather than the likelihood; second, within-subject temporal correlations were not taken into account in their test statistics, which leads to loss of statistical power; third, the available asymptotic theories were developed for dense functional data, which lead to χ^2 mixture limiting distributions for the test statistics, and are not applicable to sparse longitudinal data.

Many longitudinal data, observed on irregular time points and with substantial measurement errors, can be treated as sparse functional data (Yao et al., 2005; Hall et al., 2006; Zhang and Wang, 2016). Existence of within-subject correlations is a fundamental issue in functional or longitudinal data analysis, since methods taking into account correlation are generally more efficient than those do not. There has been a lot of recent work on modeling covariance of longitudinal data (Fan & Wu, 2008), and improving estimation efficiency of nonparametric regression using correlation information (Wang et al., 2005). However, these results have not been used to improve the power of nonparametric tests. In longitudinal GWAS, SNP-level tests are performed for hundreds of thousands of SNPs, and the multiple comparison adjustments often lead to conservative test results. It is therefore even more critical to improve the power of functional tests in order to achieve the genome-wide significance level (Liu & Lin, 2017).

More recently, Tang et al. (2016) studied model (1) under a longitudinal clinical trial setting and proposed a generalized quasi-likelihood ratio (GQLR) test for hypotheses (2). Their test is an extension of the generalized likelihood ratio test, originally proposed for various nonparametric models on independent data, including the varying coefficient models (Fan et al., 2001; Li and Liang, 2008) and the additive models (Fan & Jiang, 2005). See González-Manteiga and Crujeiras (2013) for a comprehensive review of these test procedures. The test proposed by Tang et al. (2016) is based on the classic kernel estimators, also referred to as "working independent" (WI) estimators by Lin and Carroll (2001), and suffers from low statistical power as we will demonstrate in this paper. When applied to longitudinal GWAS data, the WI test of Tang et al. (2016) fails to detect many important genes that have already been documented in the literature.

We propose to estimate model (1) by a profile estimating equation method based on the seemingly unrelated kernel (Wang et al., 2005) and to build a nonparametric test for (2) that takes into account the within-subject correlation. It is known that the seemingly unrelated kernel leads to more efficient nonparametric estimators, but so far it has been neither applied to the functional analysis of variance models nor used to build a nonparametric test. We show the proposed test is minimax most powerful when the covariance structure is correctly specified and is more powerful than existing tests. We provide practical strategies to estimate the within-subject covariance nonparametrically and a bootstrap procedure to consistently estimate the null distribution of the test statistic. We also show that the proposed test enjoys a property called the Wilks phenomenon (Fan et al., 2001), that the null distribution of the test statistic does not depend on the unknown model parameters. This important property makes it practical to perform functional tests to longitudinal GWAS data, saving us from repeating the bootstrap procedure on hundreds of thousands of SNPs.

There has also been some recent work applying functional data analysis on the ADNI brain image data. Wang and Zhu (2017) studied a regression model using the 2-dim magnetic resonance image of the brain as a functional predictor to predict the Alzheimer's disease (AD) status; Li et al. (2017) modeled the DTI fractional anisotropy on corpus callosum, a fiber tract in human brain, as a functional response and regressed it against disease status and other subject-specific covariates. Both papers considered dense functional data on regular grid points, extracted from a brain image during one clinic visit, and neither considered genetic information. In contrast, the data we analyze are longitudinal phenotypes, measured on sparse, irregular time points, which provide more information on the decline of cognitive functions and thus increase the chance of identifying Alzheimer-related genes. Xu et al. (2014) also proposed SNP-level hypothesis testing methods for longitudinal GWAS data, but their methods were based on linear mixed models with strong parametric assumptions. Song et al. (2014) and Chu et al. (2020) also considered similar models as (1) using spline methods. However, they focused on feature selection rather than hypothesis testing.

The rest of the paper is organized as follows. In Section 2, we describe the estimation procedures under both the null and alternative hypotheses. In Section 3, we propose the seemly unrelated kernel-based functional analysis of covariance test and study its asymptotic properties, including its asymptotic null distribution, Wilks phenomenon, local power, and its minimax optimality. We discuss some implementation issues in Section 4, including covariance estimation, bootstrap procedure and its consistency. We then illustrate the proposed methodology by simulation studies in Section 5, and analyze the ADNI data in Section 6. Finally, some concluding remarks and discussions are provided in Section 7. Technical proofs and additional data analysis results are collected in Appendix S1.

2 | ESTIMATION PROCEDURE

Although Model (1) is defined in continuum, observations on $Y_{k,i}(t)$ and $\boldsymbol{X}_{k,i}(t)$ are, in practice, made on discrete and subject-specific time points. Let $\boldsymbol{T}_{k,i} = (T_{k,i1}, \cdots, T_{k,im_{k,i}})^{\mathrm{T}}$ be the random observation time points for subject i with genotype k, where $m_{k,i}$ is the number of repeated measurements. Denote $\boldsymbol{Y}_{k,i} = (Y_{k,i1}, \ldots, Y_{k,im_{k,i}})^{\mathrm{T}}$, $\boldsymbol{\mu}_{k,i} = (\mu_{k,i1}, \ldots, \mu_{k,im_{k,i}})^{\mathrm{T}}$, $\boldsymbol{X}_{k,i} = (\boldsymbol{X}_{k,i1}, \ldots, \boldsymbol{X}_{k,im_{k,i}})^{\mathrm{T}}$, where $Y_{k,ij} = Y_{k,i}(T_{k,ij})$, $\mu_{k,ij} = \mu_{k,i}(T_{k,ij})$ and $\boldsymbol{X}_{k,ij} = \boldsymbol{X}_{k,i}(T_{k,ij})$. Define $\epsilon_{k,ij} = Y_{k,ij} - \mu_{k,ij}$, and consider $\boldsymbol{\epsilon}_{k,i} = (\epsilon_{k,i1}, \ldots, \epsilon_{k,im_{k,i}})^{\mathrm{T}}$ as discrete observations on a longitudinal process $\epsilon_{k,i}(t)$.

We assume the conditional covariance of $Y_{k,i}(t)$ is a bivariate function

$$\mathcal{R}(t_1, t_2) = \operatorname{cov}\left\{\epsilon_{k,i}(t_1), \epsilon_{k,i}(t_2)\right\}, \quad \text{for any } \; t_1, t_2 \in \mathcal{T}. \tag{3}$$

Note that the assumption of the covariance structures being the same across treatment groups is common in analysis of variance. Let $\boldsymbol{\Sigma}_{k,i} = \operatorname{cov}(\boldsymbol{Y}_{k,i} | \boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i}) = \{\mathcal{R}(T_{k,ij}, T_{k,ij'})\}_{j,j'=1}^{m_{k,i}}$ be the subject-specific covariance matrices. Since the true covariance function $\mathcal{R}$ is unknown, the covariance model $\mathcal{V}(t_1, t_2)$ adopted in data analysis is commonly referred to as a "working" covariance, which is subject to misspecification. Historically, a working covariance model is usually assumed to be a member of a parametric family, such as the Matérn family. Let $\boldsymbol{V}_{k,i} = \{\mathcal{V}(T_{k,ij}, T_{k,ij'})\}_{j,j'=1}^{m_{k,i}}$ be the "working" covariance matrix for subject (k, i), which is the interpolation of the continuous covariance function $\mathcal{V}$ on the subject-specific time points. The simplest working covariance is working independence, that is, $\boldsymbol{\Sigma}_{k,i} = \boldsymbol{I}_{m_{k,i}}$. It is known that misspecified working covariance can still lead to consistent but inefficient estimators (Wang et al., 2005). More discussions on covariance modeling and estimation for irregular longitudinal data are provided in Section 4.1. We refer to the models under the null and alternative hypotheses in (2) as the reduced and full models, respectively.

2.1 | Estimation under both the null and alternative hypotheses

To estimate the full model, we extend the profile kernel estimating equation approach of Wang et al. (2005) to the multiple treatment group setting. Let $K(\cdot)$ be a kernel function, h be the bandwidth, and denote $K_h(t) = h^{-1}K(t/h)$. Following Wang et al. (2005), let $\boldsymbol{G}_{k,ij}(t)$ be an $m_{k,i} \times 2$ matrix with the jth row being $\{1, (T_{k,ij} - t)/h\}$ and rest of the entries 0. The full model estimators are obtained by iterating between the following two steps.

Step 1 (Seemingly unrelated kernel estimator): By Taylor's expansion, for any $T_{k,ij}$ in a neighborhood h of t, $\theta_k(T_{ij}) \approx \alpha_0 + \alpha_1(T_{ij} - t)/h$, where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^{\mathrm{T}} = \{\theta_k(t), h\theta_k^{(1)}(t)\}^{\mathrm{T}}$ are the coefficients of the local polynomial. Let $\widetilde{\theta}_{F,k}(\cdot)$ be the current estimator of $\theta_k(\cdot)$. For a given $\boldsymbol{\beta}$, update $\widehat{\theta}_{F,k}(t; \boldsymbol{\beta})$ by $\widehat{\alpha}_0(t; \boldsymbol{\beta})$, where $\widehat{\boldsymbol{\alpha}} = \{\widehat{\alpha}_0(t, \boldsymbol{\beta}), \widehat{\alpha}_1(t, \boldsymbol{\beta})\}^{\mathrm{T}}$ is the solution of

$$\mathbf{0} = \sum_{i=1}^{n_k}\sum_{j=1}^{m_{k,i}} K_h(T_{k,ij} - t)\mu_{k,ij}^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha})\boldsymbol{G}_{k,ij}^{\mathrm{T}}(t)\boldsymbol{V}_{k,i}^{-1}\left[\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}^*\left\{t, \boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \widetilde{\theta}_{F,k}(\boldsymbol{T}_{k,i}; \boldsymbol{\beta})\right\}\right]. \tag{4}$$

Here, $\boldsymbol{\mu}^*\left\{t, \boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \widetilde{\theta}_{F,k}(\boldsymbol{T}_{k,i}; \boldsymbol{\beta})\right\}$ is a vector of dimension $m_{k,i}$ with the lth element being $\mu[\boldsymbol{X}_{k,il}^{\mathrm{T}}\boldsymbol{\beta} + I(l = j)\{\alpha_0 + \alpha_1(T_{k,ij} - t)/h\} + I(l \neq j)\widetilde{\theta}_{F,k}(T_{k,il}, \boldsymbol{\beta})]$; $\mu_{k,ij}^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is the first derivative of the function $\mu(\cdot) = g^{-1}(\cdot)$ evaluated at $\boldsymbol{X}_{k,ij}^{\mathrm{T}}\boldsymbol{\beta} + \alpha_0 + \alpha_1\{(T_{k,ij} - t)/h\}$.

Step 2 (Profile estimating equation): Then $\widehat{\boldsymbol{\beta}}_F$ is updated by solving the estimating equation pooling all treatment groups together

$$\mathbf{0} = \sum_{k=1}^{q}\sum_{i=1}^{n_k} \frac{\partial\boldsymbol{\mu}\{\boldsymbol{X}_{k,i}\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}_{F,k}(\boldsymbol{T}_{k,i}; \boldsymbol{\beta})\}^{\mathrm{T}}}{\partial\boldsymbol{\beta}}\boldsymbol{V}_{k,i}^{-1}\left[\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}\left\{\boldsymbol{X}_{k,i}\boldsymbol{\beta} + \widehat{\boldsymbol{\theta}}_{F,k}(\boldsymbol{T}_{k,i}; \boldsymbol{\beta})\right\}\right]. \tag{5}$$

One may use the working independent estimators of Lin and Carroll (2001) as the initial values. At convergence, denote the final estimators as $\widehat{\boldsymbol{\beta}}_F$ and $\widehat{\theta}_{F,k}(\cdot) = \widehat{\theta}_{F,k}(\cdot\,; \widehat{\boldsymbol{\beta}}_F)$.

Solving local estimating Equation (4) requires iteratively reweighted least square steps for each t, but when $g(\cdot)$ is the identity link a closed form solution is given by

$$\widehat{\theta}_{F,k}(t; \boldsymbol{\beta}) = \boldsymbol{H}_k^{\mathrm{T}}(t)(\mathbb{Y}_k - \mathbb{X}_k\boldsymbol{\beta}), \tag{6}$$

where $\mathbb{Y}_k$ and $\mathbb{X}_k$ are the response vector and the covariate design matrix pooling all subjects within group k together and $\boldsymbol{H}_k^{\mathrm{T}}(t)$ is a linear smoother described in proposition 1 in Lin et al. (2004). Denote $v_{k,i}^{j\ell}$ as the (j,ℓ)th element in $\boldsymbol{V}_{k,i}^{-1}$, $N_k = \sum_{i=1}^{n_k} m_{k,i}$, $\boldsymbol{T}_k = \{T_{k,ij},\ j = 1, \dots, m_{k,i};\ i = 1, \dots, n_k\}$ and $\mathbb{V}_k = \mathrm{diag}(\boldsymbol{V}_{k,1}, \dots, \boldsymbol{V}_{k,n_k})$. Let $\mathbb{V}_k^d = \mathrm{diag}(\mathbb{V}_k^{-1})$ be the diagonal matrix containing all diagonal elements in $\mathbb{V}_k^{-1}$. Define $\boldsymbol{K}_{wh}(t) = \{\sum_{i=1}^{n_k}\sum_{j=1}^{m_{k,i}} K_h(T_{k,ij} - t)v_{k,i}^{jj}\}^{-1}K_h(\boldsymbol{T}_k - t)$ as an N_k-vector and $\mathbb{K}_{wh} = \{\boldsymbol{K}_{wh}(T_{k,i1}), \dots, \boldsymbol{K}_{wh}(T_{k,n_k,m_{k,n_k}})\}$ as an $N_k \times N_k$ matrix evaluating $\boldsymbol{K}_{wh}(t)$ on $\boldsymbol{T}_k$, then $\boldsymbol{H}_k(t) = \boldsymbol{K}_{wh}^{\mathrm{T}}(t)\{\mathbb{I} + (\mathbb{V}_k^{-1} - \mathbb{V}_k^d)\mathbb{K}_{wh}\}^{-1}\mathbb{V}_k^{-1}$. By (6), it is easy to see $(\partial\widehat{\theta}_{F,k}/\partial\boldsymbol{\beta})(t; \boldsymbol{\beta}) = -\boldsymbol{H}_k^{\mathrm{T}}(t)\mathbb{X}_k$, and the solution for (5) is

$$\widehat{\boldsymbol{\beta}}_F = \left(\sum_{k=1}^{q}\widetilde{\mathbb{X}}_k^{\mathrm{T}}\mathbb{V}_k^{-1}\widetilde{\mathbb{X}}_k\right)^{-1}\left(\sum_{k=1}^{q}\widetilde{\mathbb{X}}_k^{\mathrm{T}}\mathbb{V}_k^{-1}\widetilde{\mathbb{Y}}_k\right), \tag{7}$$

where $\widetilde{\mathbb{X}}_k = (\mathbb{I} - \mathbb{H}_k)\mathbb{X}_k$, $\widetilde{\mathbb{Y}}_k = (\mathbb{I} - \mathbb{H}_k)\mathbb{Y}_k$, and $\mathbb{H}_k$ is the matrix evaluating $\boldsymbol{H}_k(\cdot)$ at $\boldsymbol{T}_k$.

The reduced model estimators, $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\theta}_R(t)$, can be estimated by the same procedure assigning all subject into the same group.

2.2 | Asymptotic properties of the estimators

We first investigate the asymptotic properties of the profile-kernel estimators of $\boldsymbol{\beta}$ and $\theta_k(t)$ under both the full and reduced models. Denote the true parameters as $\boldsymbol{\beta}_0$ and $\theta_{k0}(t)$, $k = 1, \dots, q$. For ease of exposition, we assume $m_{k,i} = m < \infty$ for all k and i. For situations where the numbers 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 $\mathcal{T}$, with a density $f(t) > 0$ for all $t \in \mathcal{T}$, and there exist constants $0 < \rho_1, \dots, \rho_q < 1$ such that $\sum_{k=1}^q \rho_k = 1$ and $n_k/n - \rho_k = O_p(n^{-1/2})$. In an observational study such as the ADNI, suppose the subjects are random samples from a target population, ρ_k is the population portion of genotype k, then $(n_1, \dots, n_q)$ follows a multinomial distribution, $\mathrm{E}(n_k/n - \rho_k)^2 = \rho_k(1-\rho_k)/n$ and hence the assumption $n_k/n - \rho_k = O_p(n^{-1/2})$ is satisfied.

Under our framework, the true and working covariance matrices $\boldsymbol{\Sigma}_{k,i}$ and $\boldsymbol{V}_{k,i}$ are independent realizations of random matrices $\boldsymbol{\Sigma}_k$ and $\boldsymbol{V}_k$, respectively, because they are the same covariance functions $\mathcal{R}$ and $\mathcal{V}$ interpolated on independent and identically distributed time vectors $\boldsymbol{T}_{k,i}$. Similarly, $\boldsymbol{\Delta}_{k,i} = \mathrm{diag}\{\mu_{k,ij}^{(1)}\}_{j=1}^m$ are independent copies of the random matrix $\boldsymbol{\Delta}_k$. Denote $\sigma_{k,j\ell}$, $v_k^{j\ell}$ and $\Delta_{k,j\ell}$ as the (j,ℓ)th element in $\boldsymbol{\Sigma}_k$, $\boldsymbol{V}_k^{-1}$ and $\boldsymbol{\Delta}_k$, respectively. When the response variables are non-Gaussian, $\boldsymbol{\Sigma}_k$, $\boldsymbol{V}_k$ and $\boldsymbol{\Delta}_k$ depend on the mean structure and hence might be different between treatment groups. Under the null hypothesis in (2), however, all groups are identical and $\boldsymbol{\Sigma}_k \equiv \boldsymbol{\Sigma}$, $\boldsymbol{V}_k \equiv \boldsymbol{V}$ and $\boldsymbol{\Delta}_k \equiv \boldsymbol{\Delta}$ for all k. In addition, we make the following assumptions.

- (C1) Assume that $\theta_{k0}(\cdot), k = 1, \cdots, q$, are twice continuously differentiable on $\mathcal{T}$. Define $B_{1k}(t) = \sum_{j=1}^m E[\Delta_{k,jj}^2 v_k^{jj} | T_{k,1j} = t] f(t)$ and $B_1(t) = \sum_{k=1}^q \rho_k B_{1k}(t)$, and assume these functions are Lipschitz continuous.
- (C2) The kernel function $K(\cdot)$ is a symmetric continuous probability density function on $[-1, 1]$ with $\int K(t)t^2 dt = 1$ and $\nu_K = \int K^2(t)dt < \infty$.
- (C3) Assume $h \to 0$ as $n \to \infty$, such that $nh^8 \to 0$ and $nh/\log(1/h) \to \infty$.

The reduced model under the null hypothesis in (2) is a generalized partially linear model, the properties of which are studied in Wang et al. (2005) and summarized in the following proposition.

Proposition 1. *Under $H_0 : \theta_{10}(t) = \cdots = \theta_{q0}(t) \equiv \theta_0(t)$ and assumptions above,*

$$\hat{\theta}_R(t) - \theta_0(t) = \frac{h^2}{2} b_*(t) - \boldsymbol{\varphi}_R(t)(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) + \mathcal{U}_R(t) + \mathcal{M}_R(t) + o_p[h^2 + \{\log(n)/nh\}^{1/2} + n^{-1/2}], \tag{8}$$

where $\mathcal{U}_R(t) = \{nB_1(t)\}^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^m K_h(T_{k,ij} - t) \mu_{k,ij}^{(1)} \left\{ \sum_{l=1}^m v_{k,i}^{jl} \epsilon_{k,il} \right\}$, $\mathcal{M}_R(t) = \{nB_1(t)\}^{-1} \times \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^m \mu_{k,ij}^{(1)} \{ Q_{1,*}(t, T_{k,ij}) \sum_{l=1}^m v_{k,i}^{jl} \epsilon_{k,il} + v_{k,i}^{jj} Q_{2,*}(t, T_{k,ij}) \epsilon_{k,ij} \}$, *and* $\boldsymbol{\varphi}_R(t)$, $b_*(t)$, $Q_{1,*}$ *and* $Q_{2,*}$ *are defined in Appendix A.1. In addition,*

$$\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 = \boldsymbol{D}_R^{-1} \mathcal{E}_R + o_p(n^{-1/2}), \tag{9}$$

where $\boldsymbol{D}_R = E(\widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Delta} \boldsymbol{V}^{-1} \boldsymbol{\Delta} \widetilde{\boldsymbol{X}})$, $\mathcal{E}_R = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} \boldsymbol{\Delta}_{k,i} \boldsymbol{V}_{k,i}^{-1} \boldsymbol{\epsilon}_{k,i}$ *and* $\widetilde{\boldsymbol{X}}_{k,i} = \boldsymbol{X}_{k,i} - \boldsymbol{\varphi}_R(\boldsymbol{T}_{k,i})$.

Adopting similar derivations in the multiple-group setting, we get the asymptotic expansions of the full model estimators.

Proposition 2. *Under the full model and assumptions above,*

$$\hat{\theta}_{F,k}(t) - \theta_{k0}(t) = \frac{h^2}{2} b_{k*}(t) - \boldsymbol{\varphi}_{F,k}(t)(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + \mathcal{U}_{F,k}(t) + \mathcal{M}_{F,k}(t) + o_p[h^2 + \{\log(n)/nh\}^{1/2} + n^{-1/2}], \tag{10}$$

where $\mathcal{U}_{F,k}(t) = \{n_k B_{1k}(t)\}^{-1} \sum_{i=1}^{n_k} \sum_{j=1}^{m} K_h(T_{k,ij} - t) \mu_{k,ij}^{(1)} \left(\sum_{l=1}^{m} v_{k,i}^{jl} \epsilon_{k,il} \right)$, $\mathcal{M}_{F,k}(t) = \{n_k B_{1k}(t)\}^{-1} \times \sum_{i=1}^{n_k} \sum_{j=1}^{m} \mu_{k,ij}^{(1)} \{ Q_{k1,*}(t, T_{k,ij}) \sum_{l=1}^{m} v_{k,i}^{jl} \epsilon_{k,il} + v_{k,i}^{jj} Q_{k2,*}(t, T_{k,ij}) \epsilon_{k,ij} \}$,*and* $\boldsymbol{\varphi}_{F,k}(t)$, $b_{k*}(t)$, $Q_{k1,*}$ *and* $Q_{k2,*}$ *are defined in Appendix A.1. In addition,*

$$\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0 = \boldsymbol{D}_F^{-1} \mathcal{E}_F + o_p(n^{-1/2}), \tag{11}$$

where $\boldsymbol{D}_F = \sum_{k=1}^{q} \rho_k E(\widetilde{\boldsymbol{X}}_k^{\mathrm{T}} \boldsymbol{\Delta}_k \boldsymbol{V}_k^{-1} \boldsymbol{\Delta}_k \widetilde{\boldsymbol{X}}_k)$, $\mathcal{E}_F = n^{-1} \sum_{k=1}^{q} \sum_{i=1}^{n_k} \widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}} \boldsymbol{\Delta}_{k,i} \boldsymbol{V}_{k,i}^{-1} \epsilon_{k,i}$ *and* $\widetilde{\boldsymbol{X}}_{k,i} = \boldsymbol{X}_{k,i} - \boldsymbol{\varphi}_{F,k}(\boldsymbol{T}_{k,i})$.

When H_0 in (2) holds, the full and reduced models are identical, $\hat{\boldsymbol{\beta}}_F$ and $\hat{\boldsymbol{\beta}}_R$ have the same first order asymptotic expansion and $\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_R = o_p(n^{-1/2})$.

3 | FUNCTIONAL ANALYSIS OF COVARIANCE TEST

3.1 | Test procedure and the asymptotic null distribution

We now address the hypothesis testing problem in (2). Our test procedure is based on quasi-likelihoods (McCullagh & Nelder, 1989), which only require correctly specifying the mean structure. A quasi-likelihood function $\mathcal{Q}$ satisfies

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

where $\boldsymbol{Y}$ is the response vector within a subject, $\boldsymbol{\mu} = g^{-1}\{\boldsymbol{X\beta} + \theta(\boldsymbol{T})\}$ is the conditional mean vector in model (1) and $\boldsymbol{V}$ is the working covariance assumed to be the same as the one used in estimation. Our proposed test is based on a GQLR test statistic

$$\lambda_n = \sum_{k=1}^{q} \sum_{i=1}^{n_k} \left(\mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\hat{\boldsymbol{\beta}}_F + \hat{\theta}_{F,k}(\boldsymbol{T}_{k,i})\}, \boldsymbol{Y}_{k,i}] - \mathcal{Q}[g^{-1}\{\boldsymbol{X}_{k,i}\hat{\boldsymbol{\beta}}_R + \hat{\theta}_R(\boldsymbol{T}_{k,i})\}, \boldsymbol{Y}_{k,i}] \right). \tag{12}$$

The following theorem provides the asymptotic distribution of λ_n under H_0, the proof of which is provided in Appendix A.2.

Theorem 1. *Under the assumptions outlined above, further assume*

$$B_2(t) = \sum_{j=1}^{m} E\{\Delta_{jj}^2 (\boldsymbol{V}^{-1} \boldsymbol{\Sigma} \boldsymbol{V}^{-1})_{jj} | T_j = t\} f(t), \tag{13}$$

is Lipschitz continuous in t, then under the null hypothesis H_0 in (2),

$$\sigma_n^{-1}\{\lambda_n - \mu_n - d_n\} \to \text{Normal }(0,1) \quad \text{in distribution, as } n \to \infty,$$

where $d_n = o_p(h^{-1/2})$, $\mu_n = (q-1)h^{-1}\{K(0) - \nu_k/2\}\int_{\mathcal{T}} B_2(t)/B_1(t)dt + O_p(1)$, $\sigma_n^2 = 2(q-1)h^{-1} \times \varpi_K \int_{\mathcal{T}} B_2^2(t)/B_1^2(t)dt + O_p(1)$, $\varpi_K = \int \{K(u) - \frac{1}{2}K * K(u)\}^2 du$, *and* $K * K(t) = \int_{-\infty}^{\infty} K(s)K(t-s)ds$ *is the convolution of the kernel function.*

Remark 1. Fan et al. (2001) showed that the generalized likelihood ratio test enjoys a property called the Wilks phenomenon, that is, the asymptotic distribution of the test statistic under the null hypothesis does not depend on the value of the unknown parameters. Indeed, when the likelihood function is used and correctly specified, this property holds for a wide range of problems. As shown in Theorem 1, the asymptotic distribution of λ_n in our problem, however, depends on parameters in the true and working covariance structures. Since the working covariance is often misspecified under longitudinal/ functional data settings, the Wilks phenomenon does not hold in general for the GQLR test.

Remark 2. Under the special case where the working covariance is equal to the true covariance, the asymptotic distribution of λ_n does not depend on nuisance parameters, as shown in the following corollary, and hence the Wilks phenomenon holds. The issue of consistently estimating the covariance function is deferred to Section 4.

Corollary 1. *Under the setting of Theorem 1, if* $\boldsymbol{V} = \boldsymbol{\Sigma}$,

$$\sigma_{n*}^{-1}\{\lambda_n - \mu_{n*} - d_{n*}\} \to \text{Normal }(0,1) \quad \text{in distribution, as } n \to \infty,$$

where $d_{n*} = o_p(h^{-1/2})$, $\mu_{n*} = (q-1)h^{-1}|\mathcal{T}|\{K(0) - \nu_k/2\}$, $\sigma_{n*}^2 = 2(q-1)h^{-1}|\mathcal{T}|\varpi_K$.

Remark 3. The result in Corollary 1 implies a rescaled version of λ_n can be approximated by a χ^2 distribution with a degree of freedom diverging to infinity in the rate of $O(h^{-1})$. Specifically, $r_K\lambda_n$ follows an asymptotic χ^2 distribution with $r_K\mu_{n*}$ degrees of freedom, where $r_K = \frac{K(0)-\nu_K/2}{\int\{K(t)-0.5K*K(t)\}^2dt}$.

3.2 | Minimax power of the functional analysis of variance test

To study the power of the proposed test, consider a local alternative hypothesis

$$H_{1n}: \theta_k(t) = \theta_0(t) + S_{kn}(t), \quad k = 1, \dots, q, \quad \text{with } \sum_{k=1}^{q} \rho_k S_{kn}(t) = 0, \tag{14}$$

where $S_{kn}(t)$ are twice continuously differentiable functions with $\sup_{t\in\mathcal{T}}|S_{kn}(t)| \to 0$ as $n \to \infty$. The asymptotic distribution of the GQLR test statistic λ_n under the local alternative H_{1n} is given in Theorem 2.

Theorem 2. *Under assumptions in Section 3.1 and the local alternative (14), denote* $\mu_{1n} = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k} \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i})\}$, *and assume*

$$h \times \mu_{1n} \to C_S < \infty. \tag{15}$$

for some constant $C_S > 0$*, then*

$$\sigma_{1n}^{-1}\{\lambda_n - \mu_n - \mu_{1n}\} \to \text{Normal }(0,1) \quad \text{in distribution, as } n \to \infty,$$

where $\sigma_{1n}^2 = \sigma_n^2 + \sum_{k=1}^{q}\sum_{i=1}^{n_k} \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Sigma}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i})\}$*, and* μ_n *and* σ_n^2 *are defined in Theorem 1.*

An approximate level-α test is to reject the null hypothesis if $\lambda_n - \mu_n > z_\alpha \sigma_n$, where z_α is the upper $100 \times \alpha$ percentile of Normal$(0,1)$. Define the class of functions

$$\mathcal{S}_n(\varrho) = [\boldsymbol{S}_n = (S_{1n}, \ldots, S_{qn})^{\mathrm{T}} : \sum_{k=1}^{q} \rho_k \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\Delta_k \boldsymbol{\Sigma}_k^{-1}\Delta_k S_{kn}(\boldsymbol{T}_k)\} \geq \varrho^2],$$

where ϱ measures the size of the local signal. As shown in Fan et al. (2001), these nonparametric tests have nontrivial power for local signals of size $\varrho_n^* = n^{-4/9}$ when the bandwidth is $h_n^* = c^* n^{-2/9}$ for a constant c^*. The rate ϱ_n^* is referred to as the minimax rate (Ingster, 1993). Following similar arguments, it is easy to show that the class of quasi-likelihood ratio tests proposed in this paper enjoy the same minimax power rate.

We now want to show that, within the class of proposed tests, the power of the test is minimax optimal when the working covariance is correctly specified. To simplify our arguments, we focus on the case where $g(\cdot)$ is an identity link and hence $\Delta_k = I$. Suppose $\mathcal{V}$ is a bivariate working covariance function and $\boldsymbol{V}_{k,i}$ is the working covariance matrix evaluating $\mathcal{V}$ on the observation times $\boldsymbol{T}_{k,i}$. For a local signal $\boldsymbol{S}_n = (S_{1n}, \ldots, S_{qn})^{\mathrm{T}} \in \mathcal{S}_n(\varrho_n^*)$ and bandwidth chosen at h_n^*, we have $\sigma_{1n}^2 = \sigma_n^2 \times \{1 + o_p(1)\}$ and the Type II error of the test based on working covariance $\mathcal{V}$ is

$$P\{\lambda_n - \mu_n < z_\alpha \sigma_n\} = \mathcal{B}_\alpha(\boldsymbol{S}_n, \mathcal{V}) + o_p(1), \tag{16}$$

where $\mathcal{B}_\alpha(\boldsymbol{S}_n, \mathcal{V}) = \Phi(z_\alpha - \sigma_n^{-1}\mu_{1n})$ with $\Phi(\cdot)$ being the cumulative distribution function of Normal$(0,1)$. The minimax optimality of our test under correctly specified working covariance is shown in the following theorem, the proof of which is in Appendix A.4.

Theorem 3. *Under assumptions of Theorem 2 and identity link, with bandwidth chosen at* $h^* = c^* n^{-2/9}$ *for a constant* c^**,*

$$\min_{\mathcal{V}} \max_{\boldsymbol{S}_n \in \mathcal{S}_n(\varrho_n^*)} \mathcal{B}_\alpha(\boldsymbol{S}_n, \mathcal{V}) = \max_{\boldsymbol{S}_n \in \mathcal{S}_n(\varrho_n^*)} \mathcal{B}_\alpha(\boldsymbol{S}_n, \mathcal{R}),$$

where $\mathcal{R}$ *is the true covariance as described in (3).*

Theorem 3 implies that, among all working covariance models, the maximum asymptotic Type II error is minimized when the true covariance function is used. Since the working independent test advocated by Tang et al. (2016) is a special case of our test with $\boldsymbol{V} = \boldsymbol{\Sigma}^d$, where $\boldsymbol{\Sigma}^d$ is a diagonal matrix with correctly specified variance on the diagonal, our test is more powerful than that of Tang et al. (2016) in the minimax sense.

Remark 4. Theorem 3 also implies that, in order to enjoy the minimax optimal power, some degree of undersmoothing is needed. It is well-known that cross-validation estimates the optimal bandwidth for estimation, which is of order $n^{-1/5}$ (Xia & Li, 2002). To make the bandwidth follow the optimal order $n^{-2/9}$ for hypothesis testing, we propose to multiply the cross-validated bandwidth by a factor $n^{-1/45}$. As shown in the empirical studies of Fan and Jiang (2005), the hypothesis test results are quite robust against the choice of h as long as it is in the right order.

4 | IMPLEMENTATION ISSUES

4.1 | Covariance estimation

Theorem 3 stresses the importance of correctly specifying the within-subject covariance structure in order to achieve the optimal power. Wang et al. (2005) limited their discussions to parametric covariance structures. Since then, there have been a lot of new developments on semiparametric and nonparametric covariance estimation methods, that can more flexibly model covariance functions of longitudinal data with irregular time. We now briefly describe two mainstream methods, the performance of which will be further evaluated in our simulation studies.

Fan and Wu (2008) proposed a semiparametric model for $\mathcal{R}$ via the decomposition $\mathcal{R}(t_1, t_2) = \sigma(t_1)\sigma(t_2)\rho(t_1, t_2; \boldsymbol{\gamma})$, where $\sigma^2(t)$ is a nonparametric variance function and $\rho(\cdot, \cdot)$ is a correlation function from a known parametric family with parameter $\boldsymbol{\gamma}$. An example of parametric correlation function is the ARMA$(1, 1)$ correlation

$$\rho(s, t; \boldsymbol{\gamma}) = \gamma_1 \exp(-|s - t|/\gamma_2) I(s \neq t) + I(s = t), \tag{17}$$

which is also a member of the Matérn family with a nugget effect. They proposed to estimate variance function by a kernel estimator $\hat{\sigma}^2(t)$ smoothing the squared residuals of a pilot fit, and then estimate the correlation parameter $\boldsymbol{\gamma}$ using a quasi-maximum likelihood estimator (QMLE)

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmax}_{\gamma} -\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\log|\boldsymbol{V}_{k,i}(\gamma)| + (\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i})^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}(\gamma)(\boldsymbol{Y}_{k,i} - \boldsymbol{\mu}_{k,i})\right\}, \tag{18}$$

where $\boldsymbol{\mu}_{k,i}$ are substituted by consistent pilot estimators (e.g. the working independent estimators), $\boldsymbol{V}_{k,i}(\gamma)$ is a within-subject covariance matrix with the (j, j')th entry $\hat{\sigma}(T_{k,ij})\hat{\sigma}(T_{k,ij'})\rho(T_{k,ij}, T_{k,ij'}; \boldsymbol{\gamma})$. When the parametric model on ρ is correctly specified, Fan and Wu (2008) showed that $\hat{\boldsymbol{\gamma}}$ in (18) is root-n consistent, and hence estimator of the covariance function $\mathcal{R}$ is uniformly consistent. However, as shown in Li (2011), a misspecified model on ρ can lead to loss of statistical efficiency.

An alternative method is to model the covariance function nonparametrically, which is the mainstream method in functional data analysis (Li & Hsing, 2010; Yao et al., 2005; Zhang & Wang, 2016). We assume $\mathcal{R}(t_1, t_2) = \mathcal{R}_0(t_1, t_2) + \sigma^2_{\text{nug}}(t_1)I(t_1 = t_2)$, where $\mathcal{R}_0(t_1, t_2)$ is a smooth, positive semi-definite, bivariate function and $\sigma^2_{\text{nug}}(t)$ is the nugget effect representing the variance function of measurement errors. Both the smooth covariance $\mathcal{R}_0(t_1, t_2)$ and the marginal variance $\sigma^2(t) = \mathcal{R}_0(t, t) + \sigma^2_{\text{nug}}(t)$ can be estimated using kernel smoothing on the residuals from a pilot fit of the full model. We refer the readers to Li (2011) for detailed algorithms on nonparametric covariance estimation. We then interpolate the estimated covariance function on the subject-specific observation times to get the within-subject covariance matrices $\hat{\boldsymbol{\Sigma}}_{k,i} = \{\hat{\mathcal{R}}_0(T_{k,ij}, T_{k,ij'})I(j \neq j') + \hat{\sigma}^2(T_{k,ij})I(j = j')\}_{j,j'=1}^{m_{k,i}}$. Following the arguments in Li (2011), the estimated covariance $\hat{\mathcal{R}}(t_1, t_2)$ is uniformly consistent to the true covariance on $\mathcal{T}^2$, and, by substituting $\boldsymbol{V}$ with $\hat{\boldsymbol{\Sigma}}$, the proposed GQLR test has the optimal asymptotic power as if the true covariance is known.

4.2 | Evaluating the null distribution with bootstrap

As demonstrated by Mammen (1993), the null distribution of a nonparametric test statistic converges to its limit very slowly; under a moderate sample size, resampling methods are recommended to evaluate the distribution of λ_n. We extend the wild bootstrap procedure of Mammen (1993) to our longitudinal/ functional data setting:

Step 1. Obtain a pilot fit of the full model assuming working independence, and estimate the variance and covariance functions from the residuals, using a method described in Section 4.1.

Step 2. Estimate both the full and reduced models using the seemingly unrelated kernel profile estimators described in Section 2, substituting $\boldsymbol{V}_{k,i}$ with estimated covariance $\widehat{\boldsymbol{\Sigma}}_{k,i}$, and evaluate the test statistic λ_n.

Step 3. For the bth bootstrap sample, regenerate the response from the reduced model $\boldsymbol{Y}^b_{k,i} = g^{-1}\{\boldsymbol{X}^{\mathrm{T}}_{k,i}\hat{\boldsymbol{\beta}}_R + \hat{\theta}_R(\boldsymbol{T}_{k,i})\} + \boldsymbol{\epsilon}^b_{k,i}$, where $\boldsymbol{\epsilon}^b_{k,i} = \omega_{k,i}\epsilon_{k,i}$, $\epsilon_{k,i}$'s are the full model residuals obtained from Step 2 the and $\omega_{k,i}$'s are independent Rademacher variables with $P(\omega_{k,i} = 1) = P(\omega_{k,i} = -1) = 0.5$.

Step 4. Calculate the test statistic λ^b_n from the bootstrap samples $\{\boldsymbol{Y}^b_{k,i}, \boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i}\}$ using the same procedure as for the original data, and repeat the bootstrap a large number B times.

Step 5. The estimated p-value is the percentage of λ^b_n that are greater than λ_n.

Note that in Step 3 we preserve the within-subject covariance structure by multiplying the residuals within a subject with the same perturbation factor. The following theorem establishes the consistency of our bootstrap procedure by showing that, conditioning on observed data $\mathcal{X} = \left\{(\boldsymbol{X}_{k,i}, \boldsymbol{T}_{k,i})\right\}$, the bootstrap test statistic λ^*_n follows the same asymptotic distribution as λ_n in Theorem 1. The proof of Theorem 4 is relegated to Appendix A.5.

Theorem 4. *Under the same assumptions for Theorem 1,*

$$P\left[\sigma_n^{-1}\left\{\lambda^*_n - \mu_n - d_n\right\} < x|\mathcal{X}\right] \to \Phi(x) \quad \text{in probability for all } x,$$

where σ_n, μ_n, and d_n are the same as defined in Theorem 1.

Remark 5. The bootstrap procedure proposed above is applicable to test a single hypothesis. We encounter two technical difficulties when applying this procedure to multiple hypotheses testing in GWAS data. First, it is computationally infeasible to run bootstrap for hundreds of thousands of SNPs. Second, it requires a gigantic bootstrap sample to reach genome-wide significance levels 10^{-7} (Fadista et al., 2016; Huang et al., 2017). To overcome these difficulties, we evoke the Wilk's phenomenon described in Corollary 1, which implies that the null distribution of the proposed functional analysis of variance test is the same for all SNPs. We can perform the proposed wild bootstrap procedure on some randomly selected SNPs, fit a χ^2 distribution to the bootstrap sample using maximum likelihood estimation, and use the fitted χ^2 distribution to determine the p-values for all SNPs.

5 | SIMULATION STUDIES

5.1 | Null distribution and Wilks phenomenon

We use simulations to demonstrate the proposed methods and validate the theoretical findings, especially the Wilks phenomenon under the null hypothesis. We generate data from $q = 4$ treatment groups with $n_k = 50$ subjects in each group and $m = 5$ repeated measurements per subject.

The responses are generated as $Y_{k,ij} = X_{1,k,ij}\beta_1 + X_{2,k,i}\beta_2 + \theta_k(T_{k,ij}) + \epsilon_{k,ij}$, where $T_{k,ij} \sim \text{Unif}(0,1)$, $X_{1,k,ij} = T_{k,ij} + U_{k,ij}$ is a time varying covariate with $U_{k,ij} \sim \text{Unif}(-1,1)$, and $X_{2,k,i}$ is a binary, time-invariant covariate that equals 0 or 1 with probability 0.5.

We examine the behavior of the proposed test under the null hypothesis $\theta_k(t) \equiv \theta_0(t)$ for all k, and consider the following three scenarios with different settings of $\boldsymbol{\beta}$ and θ_0:

$$\begin{aligned}
&\text{Scenario I:} \quad \beta_1 = 1,\ \beta_2 = 1,\ \theta_0(t) = \sin(2\pi t);\\
&\text{Scenario II:} \quad \beta_1 = 1,\ \beta_2 = -1,\ \theta_0(t) = \sin(2\pi t);\\
&\text{Scenario III:} \quad \beta_1 = -1,\ \beta_2 = 1,\ \theta_0(t) = \cos(2\pi t).
\end{aligned}$$

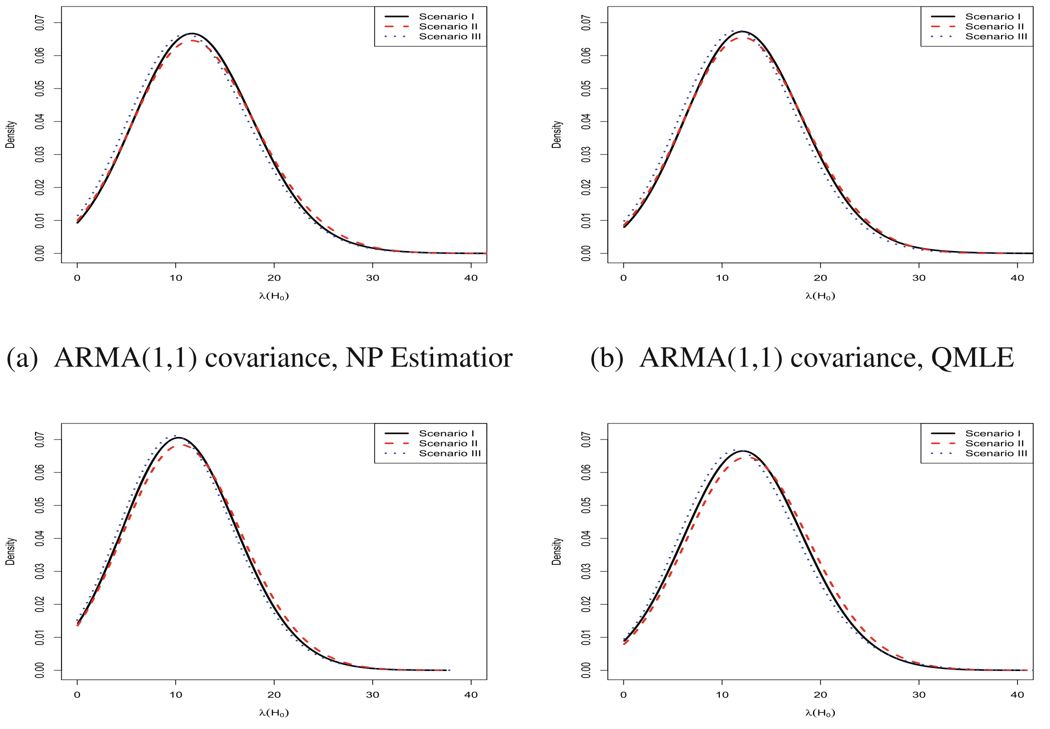
We generate the errors $\epsilon_{k,ij}$ as discrete observations on a zero-mean Gaussian process $\epsilon_{k,i}(t)$ and consider two covariance settings: (i) ARMA(1,1) covariance with $\sigma^2(t) = 0.5$ and correlation (17) with $\gamma = 0.75$ and $\nu = 1$; (ii) a nonparametric covariance induced by the mixed model $\epsilon_{k,ij} = \xi_{0,k,ij} + \sum_{l=1}^{3} \xi_{l,k,i}\phi_l(T_{k,ij})$, where $\xi_{0,k,ij}, \xi_{l,k,i} \sim N(0, 0.3)$ are independent random effects and $\phi_1(t) = t^2 + 0.5$, $\phi_2(t) = \sin(3\pi t)$, $\phi_3(t) = \cos(3\pi t)$. Note that the covariance under setting (ii) can be written as $\mathcal{R}(t_1, t_2) = \sum_{l=1}^{3} \omega_l \phi_l(t_1)\phi_l(t_2) + \sigma^2_{\text{nugg}} I(t_1 = t_2)$, which is nonstationary and cannot be represented by any "off-the-shelf" parametric covariance model, such as those in the Matérn family.

For each combination of mean and covariance settings, we generate 200 datasets and apply the proposed estimation and test procedures. For covariance estimation, we apply both methods described in Section 4.1 for comparison: the semiparametric QMLE (Fan & Wu, 2008) assuming that the correlation is from the ARMA (1,1) family and the nonparametric covariance estimation method. The bandwidths are selected by the procedure described in Remark 4 using small scale simulations and are then held fixed for massive simulations. For the test statistic, we use a Gaussian quasi-likelihood $\mathcal{Q}(\boldsymbol{\mu}, \boldsymbol{Y}) = -(\boldsymbol{Y} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu})/2$, where $\boldsymbol{V}$ is replaced with estimated covariance.

Figure 1 shows the estimated density for λ using kernel smoothing, under various mean and covariance settings and using different covariance estimators. The top two panels show results under covariance setting (i) where the true covariance is a member of the ARMA (1,1) family, and the bottom panels correspond to covariance setting (ii). Panels in the left column are results using nonparametric covariance estimators and the panels on the right hand side are based on semiparametric covariance estimators using the QMLE method of Fan and Wu (2008).

In all four panels, the null distributions under the three scenarios are almost identical, which corroborates our results in Theorem 1 that the null distribution does not depend on the true values of $\boldsymbol{\beta}$ and $\theta_0(t)$. Under setting (i), both covariance estimators consistently estimate the true covariance function, and all densities in Panels (a) and (b) are almost identical, which corroborates our results in Corollary 1 that, when the true covariance is used, the test enjoys the Wilks property and the null distribution does not depend on the values of the nuisance parameters. In Figure S1, we overlay the six curves in Panels (a) and (b) together, which clearly shows that they are very close. We also perform a k-sample Anderson–Darling test which finds no significant difference between the six distributions.

Under setting (ii), the semiparametric covariance estimator uses a misspecified covariance structure. As a result, the densities in Panel (d) are different from those in Panel (c), which is confirmed by the k-sample Anderson–Darling test. This difference shows more clearly in Figure S2 where we overlay the two groups of curves in a single plot. This result also agrees with Theorem 1 that the distribution of λ depends on the working covariance, should it be misspecified.



(a) ARMA(1,1) covariance, NP Estimatior

(b) ARMA(1,1) covariance, QMLE

(c) NP covariance, NP estimator

(d) NP covariance, QMLE

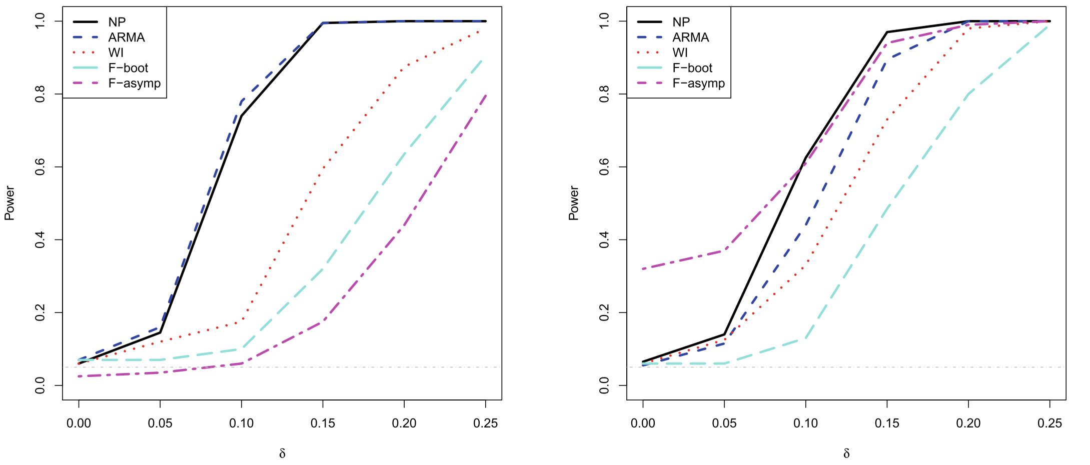
FIGURE 1 Estimated densities for λ under different settings and scenarios. Panels (a) and (b) are results under covariance setting (i) where the true covariance is ARMA(1,1); Panels (c) and (d) are under setting (ii) where the errors were generated from a mixed model with nonparametric factors. Panels (a) and (c) are based on nonparametric covariance estimator; Panels (b) and (d) are based on quasi-maximum likelihood (QMLE) assuming ARMA(1,1) covariance

5.2 | Power of the functional analysis of variance tests

To study the power of the proposed test, we adopt a similar setting as Scenario I in Section 5.1 and generate data from local alternative models with $\theta_1(t) = \theta_0(t) - 2\delta S(t)$, $\theta_2(t) = \theta_0(t) - \delta S(t)$, $\theta_3(t) = \theta_0(t) + \delta S(t)$ and $\theta_4(t) = \theta_0(t) + 2\delta S(t)$, where $S(t) = \sin(6\pi t)$. The null hypothesis is true when $\delta = 0$ and the model deviates further away from H_0 as δ increases.

We set $\delta = \{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$. For each value of δ and each of the two covariance settings described in Section 5.1, we simulate 200 datasets and apply the proposed tests based on seemingly unrelated kernel smoothing under three different covariance structures: working independence (equivalent to the test of Tang et al., 2016), semiparametric ARMA(1,1) covariance and nonparametric covariance function estimated by kernel smoothing. For each test, the nominal size is set at $\alpha = 0.05$ and the critical value is estimated by the wild bootstrap procedure in Section 4.2 based on 1000 bootstrap samples. For comparison, we also adopt the functional F-test of Zhang (2013) into our setting

$$F = \frac{\int_{\mathcal{T}} \sum_{k=1}^{q} n_k \{\widehat{\theta}_{F,k}(t) - \widehat{\theta}_R(t)\}^2 dt/(q-1)}{\int_{\mathcal{T}} \widehat{\mathcal{R}}(t,t)dt},$$



(a) Covariance setting (i) (b) Covariance setting (ii)

FIGURE 2 Empirical power of competing test procedures. NP, ARMA, and WI are the proposed generalized quasi-likelihood ratio test based on different working covariance: nonparametric covariance estimator, ARMA(1,1) and working independence. F-boot and F-asymp are F-tests based on bootstrap and asymptotic theory. The horizontal dotted line is set at 0.05. (a) Covariance setting (i) (b) Covariance setting (ii)

where $\widehat{\mathcal{R}}(t,t)$ is the kernel estimator of the variance function described in Section 4.1. We implement two versions of the F-test: **F-asymp** adopts its critical value from the asymptotic F distribution described in chapter 5 of Zhang (2013), whereas the critical value for **F-boot** is estimated by the bootstrap procedure described in Section 4.2. The empirical powers as functions of δ are the five tests as under consideration are shown in Figure 2, where the two panels correspond to the two true covariance settings.

As we can see, the F-test based on asymptotic F distribution cannot hold its nominal size: the real size of this test is lower than 0.05 under setting (i) and much higher than 0.05 under setting (ii). These results show that the asymptotic distribution of Zhang (2013) developed for dense functional data does not apply to sparse longitudinal data. All tests based on bootstrap hold their nominal size and are therefore legitimate. Among the four bootstrap-based tests, **F-boot** is not based on the likelihood principle and is least powerful, followed by the working independent version of the proposed test, which ignores the within-subject correlations.

The two versions of the proposed test considering correlation are most powerful under both settings. Under setting (i), both the nonparametric and the semiparametric covariance estimator consistently estimate the true covariance. The two proposed tests based on different covariance estimators have almost identical power curves and both are much higher than the working independent test. Specifically, at $\delta = 0.1$, the powers of the two proposed tests are over four times that of the working independent test, indicating a huge power gain by taking into account of correlations. Under setting (ii), the proposed test combined with nonparametric covariance estimator has the highest power; the semiparametric covariance estimator based on a misspecified correlation model leads to reduced power, even though the power of this test is still significantly higher than the other tests.

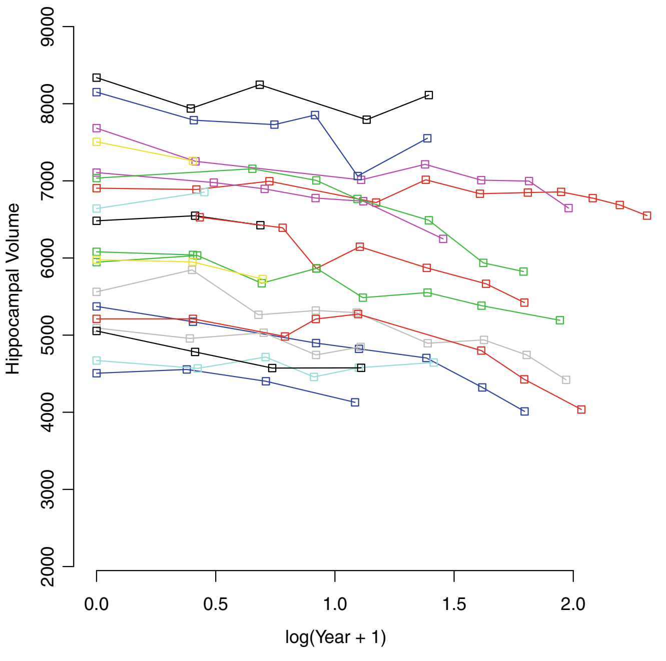


FIGURE 3 Twenty randomly selected hippocampal volume trajectories from the Alzheimer's Disease Neuroimaging Initiative cohort with log-transformed time

6 | APPLICATION TO THE ADNI DATA

The AD is an irreversible, progressive brain disorder that affects about 35.6 million people around the world (Weiner et al., 2013). The ADNI, first launched in 2004 and then renewed in 2009, is an NIH-funded longitudinal observational study, the goal of which is to develop biomarkers to detect and track AD. The original ADNI cohort included a total of 800 subjects, many of whom have repeated measurements on AD-related biomarkers over 10 years of followups. More information on ADNI data collection protocol and open data access are available at http://adni.loni.usc.edu.

Among the biomarkers considered in ADNI, there has been some documented evidence that loss of hippocampal volume in human brain may be associated with memory loss and AD (Schuff et al., 2009). In the ADNI cohort, 629 subjects have repeatedly measured hippocampal volume using neuroimaging methods during the 10-year follow-up. The measurement times are irregular and random, and the number of repeated measures per subject ranges between 2 and 11 with a median of 4. The distribution of observation time is highly skewed and observations become increasingly sparse after year 6, we therefore take a log-transformation to time and let $t = \log(1{+}\text{actual visit time})$, which brings the time domain to $\mathcal{T} = [0, 2.4]$. In Figure 3, we show 20 randomly selected hippocampal volume trajectories in log-transformed time.

Genotype (AA, AB, or BB) of 311,417 SNPs were measured for the ADNI subjects. Our goal is to identify the SNPs related to hippocampal volume loss by testing hypothesis (2) for each SNP. Demographical variables including age, gender, years of education, race, and marital status are considered as covariates in Model (1). Summary statistics of these covariates are provided in Table S1 in Appendix S1.

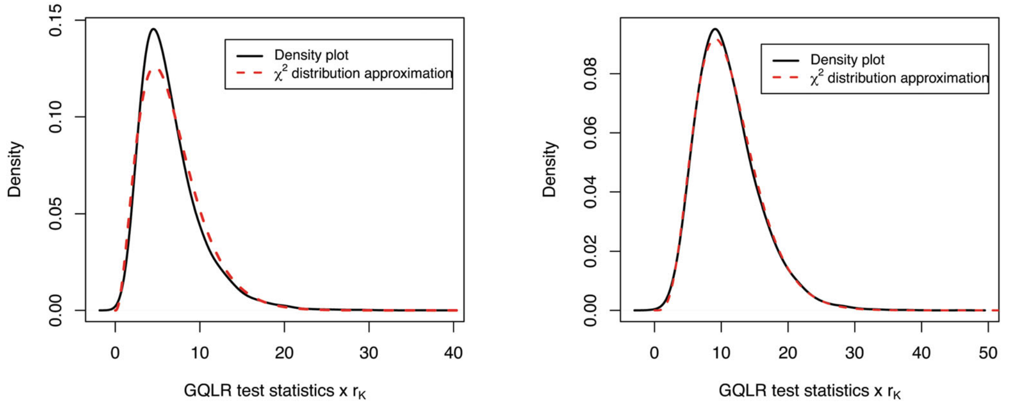
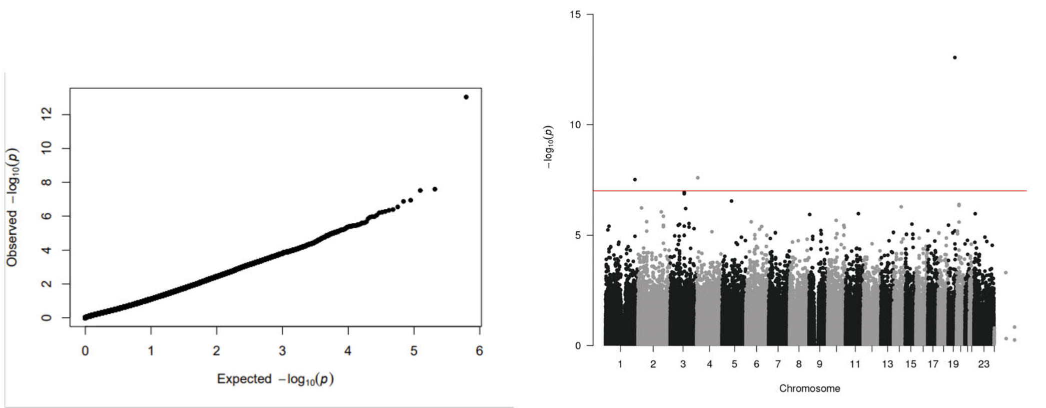


FIGURE 4 The empirical distributions (black sold line) and their χ^2 approximations (red dashed line) by the working independent method (the left panel) and nonparametric method (the right panel)



(a) (b)

FIGURE 5 Single nucleotide polymorphisms screening for the Alzheimer's disease neuroimaging initiative hippocampal volume data: (a) QQ plot of the p values, (b) the Manhatton plot

We first apply the working independent functional analysis of variance test of Tang et al. (2016) to screen for the important SNPs. The bandwidth is selected using cross validations on 20 randomly selected SNPs, the average of these selected bandwidths is adjusted by the procedure in Remark 4 and then fixed for all SNPs. Following the procedure described in Section 4.2, we perform wild bootstrap on 20 randomly selected SNPs, with 1000 bootstrap samples for each SNP, and fit a χ^2 distribution to the combined bootstrap sample using maximum likelihood estimation. The left panel of Figure 4 shows the empirical distribution of the combined bootstrap sample for the working independent test statistic and its χ^2 approximation. We then use the fitted χ^2 distribution to evaluate the p-values for all SNPs. At the 10^{-7} significance level, the working independent test detects three SNPs associated with hippocampal volume. Following Huang et al. (2017), we also provide the QQ-plot and the Manhattan plot for the p-values in Figure 5.

Next, we apply the proposed seemingly unrelated functional analysis of variance test to the top 2000 SNPs screened by the working independent test. We adopt the same bandwidth for the mean estimation as the working independent procedure, estimate the covariance function separately for each SNP using the nonparametric procedure described in Section 4.1, where the bandwidth for covariance estimation is chosen by cross-validation in 20 randomly selected SNPs. To estimate the null distribution, we run wild bootstrap on 20 randomly selected SNPs; the empirical distributions of $r_K \lambda_n^*$ from the combined bootstrap sample and its χ^2 approximation are shown in the right panel of Figure 4. The closeness of the two distributions corroborates with the results in Corollary 1. At significance level 10^{-7}, the proposed test detects 177 SNPs that are associated with hippocampal volume. These SNPs deserve further investigation using independent studies. We summarize the top 50 SNPs detected by the proposed test in Table S2. The SNPs are ranked by their significance level. We provide the names of the SNPs, the chromosomes they are on, and the gene names for SNPs located in known genes.

The most significant SNP is rs2075650 located in gene APOE and some other top genes include MCF2L, OPCML, TLE1, FAM111A, and ALDH1L1, all of which have been identified by multiple independent studies to be related to hippocampal volume and AD. References of these genes are listed in Appendix S1. On the other hand, the proposed method also finds some new genes, such as LOCI107986777 and KAZN, which we could not find in existing literature and merit further investigation. Figure S3 in Appendix S1 shows the estimated functional genotype effects for the top three SNPs, rs2075650, rs2722385, and rs3817959, located in genes APOE, LOCI107986777, and KAZN, respectively. In each panel of Figure S3, the solid curve is the overall mean function, while the dashed, dotted and dash-dot curves are the estimated mean functions for different genotypes.

7 | DISCUSSION

In longitudinal GWAS, the main effects of genotypes can be modeled as nonparametric functions of time. The conservative nature of multiple comparison in GWAS makes it crucial to improve the power of the SNP level tests. Commonly used kernel estimators do not take into account the within-subject correlation, which leads to reduced power in statistical tests. Our strategy is to build our functional analysis of variance test based on the class of seemly unrelated kernel smoother of Wang et al. (2005), and we show the power of our test is minimax optimal when the true covariance structure is used. We propose a wild bootstrap procedure to consistently estimate the null distribution of the test statistic. To perform large-scale multiple hypotheses testing in longitudinal GWAS, we propose a χ^2 approximation to the wild bootstrap samples, which can be justified by the Wilks property of the proposed test procedure. In our simulation studies, the proposed test combined with consistent covariance estimators have significant higher power than the working independent test of Tang et al. (2016) and other competing test procedures. For the ADNI data, the proposed test detects not only some well-known AD-related genes, which the working independent test misses, but also some new genes that are worth further investigation. It is also worth noting, even though our work is motivated by longitudinal GWAS data, the statistical issues we address, including semiparametric modeling, covariance estimation, and improving power of nonparametric statistical tests, are generally applicable to all longitudinal data.

Covariance function modeling and estimation play a critical role in our methodology and there are various way to extend our work. Equation (3) is an equal covariance assumption commonly used in the analysis of variance literature, where the consensus is that ANOVA test is generally robust against mild violation of this assumption. In the functional data literature, there has been some recent work on testing equal covariance functions across different treatment groups (Guo et al., 2019); however, their discussion was limited to densely observed function data. It is not yet clear how their methods can be extended to longitudinal data or sparse functional data. When the assumption in (3) is seriously violated, the asymptotic χ^2 distribution described in Corollary 1 and Remark 3 may no longer hold, a new test taking into account the heterogeneity is necessary, and further investigation is needed.

The covariance estimation methods described in Section 4.1 are commonly used in functional data and longitudinal data analysis, when the covariance and the mean functions are not directly related. However, in some non-Gaussian generalized linear model setting (e.g., Poisson regression), the variance/covariance structure of the longitudinal process naturally depends on the mean function. For Poisson type of longitudinal data, Lin (2007) proposed a Poisson Mixed Model, where the within-subject temporal correlation is accommodated by introducing a few latent Gaussian random effects. If the mean function is modeled by a semiparametric regression model as (1), the resulting covariance function is similar to the semiparametric covariance model described in Section 4.1 in the sense that the covariance depends on the semiparametric model in the mean and a few additional variance parameters of the random effects. For such a model, Lin (2007) proposed to estimate the covariance parameters by maximizing a Gaussian quasi-likelihood similar to the QMLE loss (18). Other types of non-Gaussian longitudinal data may also be modeled similarly through the generalized linear mixed model framework. These possible extensions will be further pursued in our future work.

ACKNOWLEDGMENTS

Data used in preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this paper. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Xu's research is supported by the UGC funded scholarship of Hong Kong SAR. Liu's research is supported in part by the General Research Fund 15301519, RGC, UGC, Hong Kong SAR. Li's research is supported in part by the US National Institutes of Health, grant 5R21AG058198. The authors would like to thank the two anonymous referees for their constructive comments and helpful suggestions, which lead to great improvements in the quality of this paper.

ORCID

Yehua Li https://orcid.org/0000-0002-5945-378X

REFERENCES

Brumback, B., & Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, *93*, 961–994.

Chu, W., Li, R., Liu, J., & Reimherr, M. (2020). Feature selection for generalized varying coefficient mixed-effect models with application to obesity gwas. *Annals of Applied Statistics*, *14*, 276–298.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, *75*(2), 261–277.

Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in) famous gwas p-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, *24*, 1202.

Fan, J., & Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of the American Statistical Association*, *100*, 890–907.

Fan, J., & Wu, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data. *Journal of the American Statistical Association*, *103*, 1520–1533.

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics*, *29*, 153–193.

González-Manteiga, W., & Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, *22*, 361–411.

Guo, J., Zhou, B., & Zhang, J.-T. (2019). New tests for equality of several covariance functions for functional data. *Journal of the American Statistical Association*, *114*(527), 1251–1263.

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

Huang, C., Thompson, P., Wang, Y., Yu, Y., Zhang, J., Kong, D., Colen, R. R., Knickmeyer, R. C., Zhu, H., & The Alzheimer's Disease Neuroimaging Initiative. (2017). FGWAS: Functional genome wide association analysis. *NeuroImage*, *159*, 107–121.

Ingster, Y. I. (1993). Asymptotic minimax hypothesis testing for nonparametric alternatives. *Mathematical Methods in Statistics*, *2*, 85–114.

Li, J., Huang, C., & Zhu, H. (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, *112*, 1169–1181.

Li, R., & Liang, H. (2008). Variable selection in semiparametric regression modeling. *Annals of Statistics*, *36*, 261–286.

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, *98*, 355–370.

Li, Y., & Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, *38*, 3321–3351.

Lin, X. (2007). Estimation using penalized quasilikelihood and quasi-pseudo-likelihood in Poisson mixed models. *Lifetime Data Analysis*, *13*, 533–544.

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

Lin, X., Wang, N., Welsh, A. H., & Carroll, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika*, *91*, 177–193.

Liu, Z., & Lin, X. (2017). Multiple phenotype association tests using summary statistics in genome-wide association studies. *Biometrics*, *74*, 165–175.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, *21*, 255–285.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer.

Reimherr, M., & Nicolae, D. (2014). A functional data analysis approach for genetic association studies. *Annals of Applied Statistics*, *8*, 406–429.

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L., Trojanowski, J., Thompson, P., Jack, C., Jr., Weiner, M., & Initiative, A. D. N. (2009). MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*, *132*, 1067–1077.

Song, R., Yi, F., & Zou, H. (2014). On varying-coefficient independence screening for high-dimensional varying-coefficient models. *Statistica Sinica*, *24*, 1735–1752.

Tang, J., Li, Y., & Guan, Y. (2016). Generalized quasi-likelihood ratio tests for semiparametric analysis of covariance models in longitudinal data. *Journal of the American Statistical Association*, *111*, 736–747.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of gwas discovery: Biology, function, and translation. *The American Journal of Human Genetics*, *101*, 5–22.

Wang, N., Carroll, R. J., & Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, *100*, 147–157.

Wang, X., & Zhu, H. (2017). Generalized scalar-on-image regression models via total variation. *Journal of the American Statistical Association*, *112*, 1156–1168.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., & Initiative, A. D. N. (2013). The Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *The Journal of the Alzheimer's Association*, *9*, e111–e194.

Xia, Y., & Li, W. K. (2002). Asymptotic behavior of bandwidth selected by cross-validation method under dependence. *Journal Multivariate Analysis*, *83*, 265–287.

Xu, Y., Li, Y., & Nettleton, D. (2018). Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes. *Journal of the American Statistical Association*, *113*, 593–606.
Xu, Z., Shen, X., & Pan, W. (2014). Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS One*, *9*(8), e102312.
Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590.
Zhang, J. T. (2013). *Analysis of variance for functional data*. CRC Press.
Zhang, X., & Wang, J. L. (2016). From sparse to dense functional data and beyond. *Annals of Statistics*, *44*(5), 2281–2321.
Zhou, L., Huang, J., Martinez, J. G., Maity, A., Baladandayuthapani, V., & Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association*, *105*, 390–400.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhu, W., Xu, S., Liu, C. C., & Li, Y. (2023). Minimax powerful functional analysis of covariance tests with application to longitudinal genome-wide association studies. *Scandinavian Journal of Statistics*, *50*(1), 266–295. https://doi.org/10.1111/sjos.12583

APPENDIX

A.1 Notation

The following are additional notations used in Propositions 1 and 2. Define $\boldsymbol{\varphi}_{F,k}(t)$ as the solution of the integral equation

$$\sum_{j=1}^{m}\sum_{l=1}^{m} E[\Delta_{k,jj} v_k^{jl} \Delta_{k,ll}\{\boldsymbol{X}_{k,l} - \boldsymbol{\varphi}(T_{k,1l})\} | T_{k,1j} = t] f(t) = 0, \tag{A1}$$

for $k = 1, \ldots, q$, and $\boldsymbol{\varphi}_R(t)$ the solution of

$$\sum_{k=1}^{q} \rho_k \sum_{j=1}^{m}\sum_{l=1}^{m} E[\Delta_{k,jj} v_k^{jl} \Delta_{k,ll}\{\boldsymbol{X}_{k,l} - \boldsymbol{\varphi}(T_{k,1l})\} | T_{k,1j} = t] f(t) = 0. \tag{A2}$$

Define $Q_k(t,s) = \sum_{j=1}^{m} \sum_{l \neq j} E\left\{ \Delta_{k,jj} v_k^{jl} \Delta_{k,ll} B_{1k}^{-1}(T_{k,1l}) | T_{k,1j} = t, T_{k,1l} = s \right\} f(t) f(s)$, for $k = 1, \ldots, q$, and $Q(t,s) = \sum_{k=1}^{q} \rho_k Q_k(t,s)$. In addition, define operators

$$\ddot{A}_k(G; t, s) = -\sum_{j=1}^{m} \sum_{l \neq j} E\left\{ \Delta_{k,jj} v_k^{jl} \Delta_{k,ll} B_{1k}^{-1}(T_{k,1l}) G(T_{k,1l}, s) | T_{k,1j} = t \right\} f(t),$$

for any bivariate function G, and $\ddot{A}(G; t, s) = \sum_{k=1}^{q} \rho_k \ddot{A}_k(G; t, s)$.

Let $b_{k*}(t)$, $Q_{k1,*}(t)$ and $Q_{k2,*}(t)$ be the solutions of integration equations $b_{k*}(t) = \theta_{k0}^{(2)}(t) - \{B_{1k}(t)\}^{-1}\sum_{j=1}^{m}\sum_{l\neq j} E\left\{\Delta_{k,jj}v_k^{jl}\Delta_{k,ll}b_{k*}(T_{k,1l})|T_{k,1j}=t\right\}f(t)$, $Q_{k1,*}(t,s) = -Q_k(t,s) + \ddot{A}_k(Q_{k1,*};t,s)$, and $Q_{k2,*}(t,s) = \ddot{A}_k(Q_{k2,*};t,s)$, respectively. Likewise, b_*, $Q_{1,*}$ and $Q_{2,*}$ are the solutions of

$$b_*(t) = \theta_0^{(2)}(t) - \{B_1(t)\}^{-1}\sum_{k=1}^{q}\rho_k\sum_{j=1}^{m}\sum_{l\neq j} E\left\{\Delta_{k,jj}v_k^{jl}\Delta_{k,ll}b_*(T_{k,1l})|T_{k,1j}=t\right\}f(t),$$

$Q_{1,*}(t,s) = -Q(t,s) + \ddot{A}(Q_{1,*};t,s)$, and $Q_{2,*}(t,s) = \ddot{A}(Q_{2,*};t,s)$, respectively.

A.2 Proof of Theorem 1

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

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

where $\boldsymbol{\Delta}(\boldsymbol{x}) = \mathrm{diag}\{\frac{dg^{-1}}{dx}(x_j)\}_{j=1}^{m}$, $\mathcal{D}_j = \partial(\boldsymbol{V}^{j\bullet}\boldsymbol{\Delta})/\partial\boldsymbol{x}$, and $\boldsymbol{V}^{j\bullet}$ is the jth row of $\boldsymbol{V}^{-1}$. Denote $\boldsymbol{\eta}_{0k,i} = \boldsymbol{X}_{k,i}\boldsymbol{\beta}_0 + \theta_0(\boldsymbol{T}_{k,i})$, $\boldsymbol{\mu}_{0k,i} = g^{-1}(\boldsymbol{\eta}_{0k,i})$ 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

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

For any vector $\boldsymbol{a}$ and a symmetric matrix $\boldsymbol{A}$, define $\|\boldsymbol{a}\|_{\boldsymbol{A}}^2 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{a}$.

By the Taylor expansion above, the test statistic can be decomposed into

$$\lambda_n(H_0) = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + o_p(1), \tag{A3}$$

where

$$\begin{aligned}
J_1 &= \sum_{k=1}^{q}\sum_{i=1}^{n_k}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\{\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)\}, \\
J_2 &= \sum_{k=1}^{q}\sum_{i=1}^{n_k}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}(\widehat{\boldsymbol{\beta}}_F - \widehat{\boldsymbol{\beta}}_R), \\
J_3 &= \sum_{k=1}^{q}\sum_{i=1}^{n_k}\left[(\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{\widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_0(\boldsymbol{T}_{k,i})\right\}\right. \\
&\quad \left. -(\widehat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i}) - \theta_0(\boldsymbol{T}_{k,i})\right\}\right],
\end{aligned}$$

$$J_4 = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}} - \|\hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}}\right\},$$

$$J_5 = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0\|_{\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}} - \|\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0\|_{\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}}\right\},$$

$$J_6 = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)+\hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}\right.$$
$$\left. - \|\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)+\hat{\theta}_{R}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}\right\}.$$

By Lemma 1, $J_2+J_3+J_5+J_6 = o_p(h^{-1/2})$, the asymptotic distribution of $\lambda_n(H_0)$ follows directly from the asymptotic distribution of J_1+J_4 from Lemma 2.

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

Proof. (i) Under and null hypothesis and by the asymptotic expansions in Propositions 1 and 2, $\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0 = O_p(n^{-1/2})$, $\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0 = O_p(n^{-1/2})$ and $\hat{\boldsymbol{\beta}}_F-\hat{\boldsymbol{\beta}}_R = o_p(n^{-1/2})$. Therefore, $J_2 = \left\{\sum_{k=1}^q\sum_{i=1}^{n_k}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}\right\}(\hat{\boldsymbol{\beta}}_F-\hat{\boldsymbol{\beta}}_R) = O_p(n^{1/2})\times o_p(n^{-1/2}) = o_p(1)$, and

$$J_5 = \sum_{k=1}^{q}\sum_{i=1}^{n_k}(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\boldsymbol{V}_{k,i}^{-1}\Delta_{k,i}\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}}_F)+(\hat{\boldsymbol{\beta}}_R-\hat{\boldsymbol{\beta}}_F)\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\Delta_{k,i}\boldsymbol{V}_{k,i}^{-1}\Delta_{k,i}\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)$$
$$= o_p(1).$$

(iii) By similar arguments as on p. 156 of Wang, Carroll and Lin (2005),

$$\sum_{k=1}^{q}\sum_{i=1}^{n_k}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{\hat{\theta}_R(\boldsymbol{T}_{k,i})-\theta_0(\boldsymbol{T}_{k,i})\right\} = o_p(n^{1/2}),$$

hence the first term of J_3 is of order $o_p(1)$. By similar arguments, the second term in J_3 is of the same order.

(iv) We decompose J_6 into three parts,

$$J_{61} = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}} - \|\widetilde{\boldsymbol{X}}_{k,i}(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}\right\},$$

$$J_{62} = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}} - \|\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}\right\},$$

$$J_{63} = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\left[(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\left(\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}\right)\left\{\hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\right\}\right.$$
$$\left. -(\hat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\left(\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}\right)\left\{\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_0(\boldsymbol{T}_{k,i})\right\}\right].$$

It can easily show that $J_{61} = O_p(n^{-1/2})$. By Propositions 1 and 2, under the null hypothesis both $\hat{\theta}_R(t)-\theta_0(t)$ and $\hat{\theta}_{F,k}(t)-\theta_0(t)$ are of order $O_p(h^2+n^{-1/2}h^{-1/2})$. Since $\mathrm{corr}(\epsilon_{k,ij},\epsilon_{k',i'j'})\neq 0$ only if

$k = k'$ and $i = i'$, by lengthy moment calculations $\mathrm{E}J_{62}^2 = O\{n \times (h^2 + n^{-1/2}h^{-1/2})^4\}$. Therefore, we conclude $J_{62} = O_p(n^{1/2}h^4 + n^{-1/2}h^{-1})$. By similar arguments, we find $J_{63} = O_p(h^2 + n^{-1/2}h^{-1/2})$. Combining the three parts, we have $J_6 = O_p(n^{1/2}h^4 + n^{-1/2}h^{-1}) = o_p(1)$ by condition (C.3). ∎

Lemma 2. *Under H_0 in (2) and assumptions in Theorem 1,*

$$\sigma_n^{-1}(J_1 + J_4 - \mu_n) \to N(0,1) \quad \text{in distribution,}$$

where μ_n and σ_n^2 are defined in Theorem 1.

Proof. Under the null hypothesis, $B_{1k}(t) = B_1(t)$ for $k = 1, \dots, q$. Rewrite J_1 as

$$J_1 = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\sum_{l=1}^{m} \epsilon_{k,ij} v_{k,i}^{jl} \mu_{k,il}^{(1)} \{\hat{\theta}_{F,k}(T_{k,il};\boldsymbol{\beta}_0) - \hat{\theta}_R(T_{k,il})\}.$$

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

$$R_1 = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\sum_{l=1}^{m} \epsilon_{k,ij} v_{k,i}^{jl} \mu_{k,il}^{(1)} B_1^{-1}(T_{k,il}) \left(\frac{1}{n_k} - \frac{1}{n}\right) \sum_{j'} \left[K_h(T_{k,ij'} - T_{k,il})\mu_{k,ij'}^{(1)} \left\{\sum_{l'} v_{k,i}^{j'l'} \epsilon_{k,il'}\right\}\right],$$

$$R_2 = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{i'\neq i}^{n_k}\sum_{j=1}^{m}\sum_{l=1}^{m} \epsilon_{k,ij} v_{k,i}^{jl} \mu_{k,il}^{(1)} B_1^{-1}(T_{k,il}) \left(\frac{1}{n_k} - \frac{1}{n}\right) \sum_{j'} \left[K_h(T_{k,i'j'} - T_{k,il})\mu_{k,i'j'}^{(1)} \left\{\sum_{l'} v_{k,i'}^{j'l'} \epsilon_{k,i'l'}\right\}\right],$$

$$R_3 = -\sum_{k=1}^{q}\sum_{k'\neq k}^{q}\sum_{i=1}^{n_k}\sum_{i'=1}^{n'_k}\sum_{j=1}^{m}\sum_{l=1}^{m} \epsilon_{k,ij} v_{k,i}^{jl} \mu_{k,il}^{(1)} B_1^{-1}(T_{k,il}) \frac{1}{n} \sum_{j'} \left[K_h(T_{k',i'j'} - T_{k,il})\mu_{k',i'j'}^{(1)} \left\{\sum_{l'} v_{k',i'}^{j'l'} \epsilon_{k',i'l'}\right\}\right].$$

By straightforward moment calculations,

$$R_1 = \frac{q-1}{h} K(0) E\left\{B_2(T) B_1^{-1}(T) f^{-1}(T)\right\} + O_p(1).$$

It is easy to see that R_2 and R_3 have mean zero and therefore only contribute to the variance of the test statistic. Similarly, we have

$$\begin{aligned}
J_4 &= \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k} \left[\left\{\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_0(\boldsymbol{T}_{k,i})\right\}^{\mathrm{T}} \boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}\boldsymbol{\Delta}_{k,i} \left\{\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)\right\}\right. \\
&\quad \left. + \left\{\hat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)\right\}^{\mathrm{T}} \boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}\boldsymbol{\Delta}_{k,i} \left\{\hat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_0(\boldsymbol{T}_{k,i})\right\}\right] \\
&= \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{l_1,l_2=1}^{m} \mu_{k,il_1}^{(1)} v_{k,i}^{l_1 l_2} \mu_{k,il_2}^{(1)} \\
&\quad \times \left[\frac{h^2}{2} b_*(T_{k,il_1}) + \{B_1(T_{k,il_1})\}^{-1} \frac{1}{n} \sum_{k_1=1}^{q}\sum_{i_1=1}^{n_{k_1}}\sum_{j_1=1}^{m} K_h(T_{k_1,i_1j_1} - T_{k,il_1}) \mu_{k_1,i_1j_1}^{(1)} \left\{\sum_{l=1}^{m} v_{k_1,i_1}^{j_1 l} \epsilon_{k_1,i_1 l}\right\}\right] \\
&\quad \times \left[\{B_1(T_{k,il_2})\}^{-1} \frac{1}{n} \sum_{k_2=1}^{q}\sum_{i_2=1}^{n_{k_2}}\sum_{j_2=1}^{m} K_h(T_{k_2,i_2j_2} - T_{k,il_2}) \mu_{k_2,i_2j_2}^{(1)} \left\{\sum_{l=1}^{m} v_{k_2,i_2}^{j_2 l} \epsilon_{k_2,i_2 l}\right\}\right.
\end{aligned}$$

$$-\{B_{1k}(T_{k,il_2})\}^{-1}\frac{1}{n_k}\sum_{i_2=1}^{n_k}\sum_{j_2=1}^{m}K_h(T_{k,i_2j_2}-T_{k,il_2})\mu^{(1)}_{k,i_2j_2}\left\{\sum_{l=1}^{m}v^{j_2l}_{k,i_2}\epsilon_{k,i_2l}\right\}\Bigg]$$

$$+\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{l_1,l_2=1}^{m}\mu^{(1)}_{k,il_1}v^{l_1l_2}_{k,i}\mu^{(1)}_{k,il_2}$$

$$\times\left[\frac{h^2}{2}b_{k*}(T_{k,il_2})+\{B_{1k}(T_{k,il_1})\}^{-1}\frac{1}{n_k}\sum_{i_1=1}^{n_k}\sum_{j_1=1}^{m}K_h(T_{k,i_1j_1}-T_{k,il_1})\mu^{(1)}_{k,i_1j_1}\left\{\sum_{l=1}^{m}v^{j_1l}_{k,i_1}\epsilon_{k,i_1l}\right\}\right]$$

$$\times\Bigg[\{B_1(T_{k,il_2})\}^{-1}\frac{1}{n}\sum_{k_2=1}^{q}\sum_{i_2=1}^{n_{k_2}}\sum_{j_2=1}^{m}K_h(T_{k_2,i_2j_2}-T_{k,il_2})\mu^{(1)}_{k_2,i_2j_2}\left\{\sum_{l=1}^{m}v^{j_2l}_{k_2,i_2}\epsilon_{k_2,i_2l}\right\}$$

$$-\{B_{1k}(T_{k,il_2})\}^{-1}\frac{1}{n_k}\sum_{i_2=1}^{n_k}\sum_{j_2=1}^{m}K_h(T_{k,i_2j_2}-T_{k,il_2})\mu^{(1)}_{k,i_2j_2}\left\{\sum_{l=1}^{m}v^{j_2l}_{k,i_2}\epsilon_{k,i_2l}\right\}\Bigg].$$

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

$$R_4=\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{\mu^{(1)}_{k,il_1}v^{l_1l_2}_{k,i}\mu^{(1)}_{k,il_2}}{B_1(T_{k,il_1})B_1(T_{k,il_2})}\Bigg[\sum_{k_1}^{q}\left\{I(k_1=k)\left(\frac{1}{n^2}-\frac{1}{n_k^2}\right)+I(k_1\neq k)\frac{1}{n^2}\right\}$$

$$\times\sum_{i_1=1}^{n_{k_1}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\Bigg\{K_h(T_{k_1,i_1j_1}-T_{k,il_1})K_h(T_{k_1,i_1j_2}-T_{k,il_2})\mu^{(1)}_{k_1,i_1j_1}\mu^{(1)}_{k_1,i_1j_2}$$

$$\times\left(\sum_{l_3=1}^{m}v^{j_1l_3}_{k_1,i_1}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}v^{j_2l_4}_{k_1,i_1}\epsilon_{k_1,i_1l_4}\right)\Bigg\}\Bigg]$$

$$=\frac{1}{2n_{k_1}}\sum_{k_1=1}^{q}\sum_{i_1=1}^{n_{k_1}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\mu^{(1)}_{k_1,i_1j_1}\mu^{(1)}_{k_1,i_1j_2}\left(\sum_{l_3=1}^{m}v^{j_1l_3}_{k_1,i_1}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}v^{j_2l_4}_{k_1,i_1}\epsilon_{k_1,i_1l_4}\right)$$

$$\times\sum_{k=1}^{q}\Bigg[\left\{I(k_1=k)(\rho_k^2-1)+I(k_1\neq k)\rho_k\rho_{k_1}\right\}$$

$$\times E\left\{\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{\mu^{(1)}_{k,il_1}v^{l_1l_2}_{k,i}\mu^{(1)}_{k,il_2}}{B_1(T_{k,il_1})B_1(T_{k,il_2})}K_h(T_{k_1,i_1j_1}-T_{k,il_1})K_h(T_{k_1,i_1j_2}-T_{k,il_2})\right\}\Bigg]+O_p(1)$$

$$=\frac{1}{2n_{k_1}}\sum_{k_1=1}^{q}\sum_{i_1=1}^{n_{k_1}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\mu^{(1)}_{k_1,i_1j_1}\mu^{(1)}_{k_1,i_1j_2}\left(\sum_{l_3=1}^{m}v^{j_1l_3}_{k_1,i_1}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}v^{j_2l_4}_{k_1,i_1}\epsilon_{k_1,i_1l_4}\right)$$

$$\times\sum_{k=1}^{q}\Bigg[\Bigg\{I(k_1=k)(\rho_k^2-1)+I(k_1\neq k)\rho_k\rho_{k_1}$$

$$\times\frac{1}{h}\frac{1}{B_1(T_{k_1,i_1j_1})}\left\{K*K\left(\frac{T_{k_1,i_1j_2}-T_{k_1,i_1j_1}}{h}\right)I(j_1\neq j_2)+\nu_kI(j_1=j_2)\right\}\Bigg]+O_p(1)$$

$$=\frac{1}{2h}\sum_{k_1=1}^{q}\left(\rho_k^2-1+\sum_{k\neq k_1}\rho_k\rho_{k_1}\right)E\left[\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\mu^{(1)}_{k_1,i_1j_1}\mu^{(1)}_{k_1,i_1j_2}\left(\sum_{l_3=1}^{m}v^{j_1l_3}_{k_1,i_1}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}v^{j_2l_4}_{k_1,i_1}\epsilon_{k_1,i_1l_4}\right)\right.$$

$$\times\frac{1}{B_1(T_{k_1,i_1j_1})}\left\{K*K\left(\frac{T_{k_1,i_1j_2}-T_{k_1,i_1j_1}}{h}\right)I(j_1\neq j_2)+\nu_kI(j_1=j_2)\right\}\Bigg]+O_p(1)$$

$$=\frac{1-q}{2h}\nu_kE\left\{B_2(T)B_1(T)^{-1}f^{-1}(T)\right\}+O_p(1),$$

$$R_5=\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{\mu_{k,il_1}^{(1)}\nu_{k,i}^{l_1l_2}\mu_{k,il_2}^{(1)}}{B_1(T_{k,il_1})B_1(T_{k,il_2})}\left[\sum_{k_1}^{q}\left\{I(k_1=k)(\frac{1}{n^2}-\frac{1}{n_k^2})+I(k_1\neq k)\frac{1}{n^2}\right\}\right.$$

$$\times\sum_{i_1=1}^{n_{k_1}}\sum_{i_2\neq i_1}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\left\{K_h(T_{k_1,i_1j_1}-T_{k,il_1})K_h(T_{k_1,i_2j_2}-T_{k,il_2})\mu_{k_1,i_1j_1}^{(1)}\mu_{k_1,i_2j_2}^{(1)}\right.$$

$$\left.\left.\times\left(\sum_{l_3=1}^{m}\nu_{k_1,i_1}^{j_1l_3}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}\nu_{k_1,i_2}^{j_2l_4}\epsilon_{k_1,i_2l_4}\right)\right\}\right]$$

$$=\frac{1}{2}\sum_{k_1=1}^{q}\frac{\rho_{k_1}-1}{n_{k_1}}\sum_{i_1=1}^{n_{k_1}}\sum_{i_2\neq i_1}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\frac{\mu_{k_1,i_1j_1}^{(1)}\mu_{k_1,i_2j_2}^{(1)}}{B_1(T_{k_1,i_1j_1})}K_h*K_h(T_{k_1,i_2j_2}-T_{k_1,i_1j_1})$$

$$\times\left(\sum_{l_3=1}^{m}\nu_{k_1,i_1}^{j_1l_3}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}\nu_{k_1,i_2}^{j_2l_4}\epsilon_{k_1,i_2l_4}\right)+O_p(1),\quad\text{and}$$

$$R_6=\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m}\frac{\mu_{k,il_1}^{(1)}\nu_{k,i}^{l_1l_2}\mu_{k,il_2}^{(1)}}{B_1(T_{k,il_1})B_1(T_{k,il_2})}\left[\sum_{k_1}^{q}\sum_{k_2\neq k_1}\frac{1}{n^2}\sum_{i_1=1}^{n_{k_1}}\sum_{i_2=1}^{n_{k_2}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\left\{K_h(T_{k_1,i_1j_1}-T_{k,il_1})\right.\right.$$

$$\left.\left.\times K_h(T_{k_2,i_2j_2}-T_{k,il_2})\mu_{k_1,i_1j_1}^{(1)}\mu_{k_2,i_2j_2}^{(1)}\left(\sum_{l_3=1}^{m}\nu_{k_1,i_1}^{j_1l_3}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}\nu_{k_2,i_2}^{j_2l_4}\epsilon_{k_2,i_2l_4}\right)\right\}\right]$$

$$=\frac{1}{2n}\sum_{k_1=1}^{q}\sum_{k_2\neq k_1}\sum_{i_1=1}^{n_{k_1}}\sum_{i_2=1}^{n_{k_2}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\frac{\mu_{k_1,i_1j_1}^{(1)}\mu_{k_2,i_2j_2}^{(1)}}{B_1(T_{k_1,i_1j_1})}K_h*K_h(T_{k_2,i_2j_2}-T_{k_1,i_1j_1})$$

$$\times\left(\sum_{l_3=1}^{m}\nu_{k_1,i_1}^{j_1l_3}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}\nu_{k_2,i_2}^{j_2l_4}\epsilon_{k_2,i_2l_4}\right)+O_p(1).$$

It is easy to see that R_4 is the leading term in the mean of J_4. R_5 and R_6 have mean zero and only contribute to the variance of J_4.

We first combine the mean components in J_1 and J_4 in

$$\mu_n=R_1+R_4=\frac{q-1}{h}\left\{K(0)-\frac{\nu_k}{2}\right\}E\left\{\frac{B_2(T)}{B_1(T)f(T)}\right\}+O_p(1).$$

Next, we collect the remaining terms into $R_2+R_5+R_3+R_6=W_n+O_p(1)$, where $W_n=W_{n1}+W_{n2}$,

$$W_{n1}=\sum_{k_1=1}^{q}\frac{1-\rho_{k_1}}{n_{k_1}}\sum_{i_1=1}^{n_{k_1}}\sum_{i_2\neq i_1}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\frac{\mu_{k_1,i_1j_1}^{(1)}\mu_{k_1,i_2j_2}^{(1)}}{B_1(T_{k_1,i_1j_1})}\left(\sum_{l_3=1}^{m}\nu_{k_1,i_1}^{j_1l_3}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}\nu_{k_1,i_2}^{j_2l_4}\epsilon_{k_1,i_2l_4}\right)$$

$$\times\left\{K_h(T_{k_1,i_2j_2}-T_{k_1,i_1j_1})-\frac{1}{2}K_h*K_h(T_{k_1,i_2j_2}-T_{k_1,i_1j_1})\right\},$$

and

$$W_{n2} = -\frac{1}{n}\sum_{k_1=1}^{q}\sum_{k_2\neq k_1}\sum_{i_1=1}^{n_{k_1}}\sum_{i_2=1}^{n_{k_2}}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\frac{\mu^{(1)}_{k_1,i_1j_1}\mu^{(1)}_{k_2,i_2j_2}}{B_1(T_{k_1,i_1j_1})}\left(\sum_{l_3=1}^{m}v^{j_1l_3}_{k_1,i_1}\epsilon_{k_1,i_1l_3}\right)\left(\sum_{l_4=1}^{m}v^{j_2l_4}_{k_2,i_2}\epsilon_{k_2,i_2l_4}\right)$$
$$\times\left\{K_h(T_{k_2,i_2j_2}-T_{k_1,i_1j_1})-\frac{1}{2}K_h*K_h(T_{k_2,i_2j_2}-T_{k_1,i_1j_1})\right\}.$$

It is easy to see W_{n1} and W_{n2} are uncorrelated to each other, and hence $\mathrm{var}(W_n)=\mathrm{E}W_{n1}^2+\mathrm{E}W_{n2}^2$. Straightforward calculations show that

$$\mathrm{E}W_{n1}^2 = \frac{2}{h}\sum_{k_1}^{q}(1-\rho_{k1})^2\mathrm{E}\left(\frac{B_2^2(T)}{B_1^2(T)f(T)}\times\int\left\{K(u)-\frac{1}{2}K*K(u)\right\}^2du\right)+O(1),$$
$$\mathrm{E}W_{n2}^2 = \frac{2}{h}\sum_{k_1}^{q}\sum_{k_2\neq K_1}^{q}\rho_{k1}\rho_{k2}\mathrm{E}\left(\frac{B_2^2(T)}{B_1^2(T)f(T)}\times\int\left\{K(u)-\frac{1}{2}K*K(u)\right\}^2du\right)+O(1),$$

and hence

$$\mathrm{var}(W_n)=\frac{2(q-1)}{h}\mathrm{E}\left(\frac{B_2^2(T)}{B_1^2(T)f(T)}\times\int\left\{K(u)-\frac{1}{2}K*K(u)\right\}^2du\right)+O(1).$$

Since $J_1+J_4=\mu_n+W_n+O_p(1)$, the asymptotic distribution in the lemma directly follows from proposition 3.2 in de Jong (1987). ∎

A.3 Proof of Theorem 2

Lemma 3. *Suppose Assumptions (C1)–(C3) and the local alternative described in (14) and (15) hold, $\hat{\boldsymbol{\beta}}_R$ is still root-n consistent to $\boldsymbol{\beta}_0$, and $\hat{\boldsymbol{\beta}}_F-\hat{\boldsymbol{\beta}}_R=o_p(n^{-1/2})$. The nonparametric estimator $\hat{\theta}_R(t)$ has the same asymptotic expansion as in (8).*

Proof. Under the local alternative hypothesis described in (14) and (15), $\theta_k(t)=\theta_0(t)+S_{kn}(t)$ with $S_{kn}(T)=O_p(n^{-1/2}h^{-1/2})$, we have $\epsilon_{k,i}(t)=Y_{k,i}(t)-\mu_{k,i}(t)=Y_{k,i}(t)-\mu\{\boldsymbol{X}^{\mathrm{T}}_{k,i}(t)\boldsymbol{\beta}+\theta_k(t)\}\approx Y_{k,i}(t)-\mu\{\boldsymbol{X}^{\mathrm{T}}_{k,i}(t)\boldsymbol{\beta}+\theta_0(t)\}-\mu^{(1)}_{k,i}(t)S_{kn}(t)$.

For a fixed $\boldsymbol{\beta}$, we derive the asymptotic expansion of profile kernel estimator $\hat{\theta}_R(t;\boldsymbol{\beta})$ using similar derivations as in Wang et al., (2005) and get

$$\begin{aligned}\hat{\theta}_R(t;\boldsymbol{\beta})-\theta_0(t) &= \frac{h^2}{2}b_*(t)+\frac{1}{nB_1(t)}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}K_h(T_{k,ij}-t)\mu^{(1)}_{k,ij}\left[\sum_{l=1}^{m}v^{jl}_{k,i}\{\epsilon_{k,il}+\mu^{(1)}_{k,il}S_{kn}(T_{k,il})\}\right]\\
&\quad+\frac{1}{nB_1(t)}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}\mu^{(1)}_{k,ij}\left[Q_{1,*}(t,T_{k,ij})\sum_{l=1}^{m}v^{jl}_{k,i}\{\epsilon_{k,il}+\mu^{(1)}_{k,il}S_{kn}(T_{k,il})\}\right.\\
&\quad\left.+v^{jj}_{k,i}Q_{2,*}(t,T_{k,ij})\{\epsilon_{k,ij}+\mu^{(1)}_{k,ij}S_{kn}(T_{k,ij})\}\right]\\
&\quad-\boldsymbol{\varphi}_R(t)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p[h^2+\{\log(n)/nh\}^{1/2}+n^{-1/2}+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|]\\
&= \frac{h^2}{2}b_*(t)+\frac{1}{nB_1(t)}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\sum_{j=1}^{m}K_h(T_{k,ij}-t)\mu^{(1)}_{k,ij}\sum_{l=1}^{m}v^{jl}_{k,i}\epsilon_{k,il}+\mathcal{U}_R(t)+\mathcal{M}_R(t)\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{B_1(t)}\sum_{k=1}^{q}\rho_k\left[\sum_{j=1}^{m}\sum_{l=1}^{m}\mathrm{E}\{Q_{1,*}(t,T_{k,j})\Delta_{k,jj}v_k^{jl}\Delta_{k,ll}S_{kn}(T_{k,l})\}\right.\\
&\left.+\sum_{j=1}^{m}\mathrm{E}\{\Delta_{k,jj}^2v_k^{jj}Q_{2,*}(t,T_{k,j})S_{kn}(T_{k,j})\}\right]\\
&-\boldsymbol{\varphi}_R(t)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o_p[h^2+\{\log(n)/nh\}^{1/2}+n^{-1/2}+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|]. \qquad \text{(A4)}
\end{aligned}
$$

Since $S_{kn}(T)=O_p\{(nh)^{-1/2}\}$ and $\boldsymbol{\Delta}_k$ and $\boldsymbol{V}_k$ are continuous functions of θ_k, $\Delta_{k,jj}-\Delta_{jj}=O_p\{(nh)^{-1/2}\}$ and $v_k^{jl}-v^{jl}=O_p\{(nh)^{-1/2}\}$. By the assumption $\sum_k\rho_kS_{kn}(t)=0$ for all t, the additional terms in (A4) are negligible. Therefore, if $\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0=O_p(n^{-1/2})$, the expansion of $\widehat{\theta}_R(t)$ follows directly from (A4) and the leading terms are identical to those in (8).

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

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

where $\mathcal{E}_n^{\dagger}=n^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\{\boldsymbol{\epsilon}_{k,i}+\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i})\}$. It is easy to see that the additional term is

$$
\begin{aligned}
n^{-1}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i})&=\sum_{k=1}^{q}\rho_k\mathrm{E}\{\widetilde{\boldsymbol{X}}_k^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{V}_k^{-1}\boldsymbol{\Delta}_kS_{kn}(\boldsymbol{T}_k)\}+O_p\{n^{-1/2}\times(nh)^{-1/2}\}\\
&=O_p\{n^{-1/2}\times(nh)^{-1/2}\}.
\end{aligned}
$$

Therefore, $\widehat{\boldsymbol{\beta}}_R$ still have the same leading asymptotic expansion as (9) and is still root-n consistent to $\boldsymbol{\beta}_0$.

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

Proof of Theorem 2. The local alternative described in (14) and (15) are close to the null hypothesis, with the size of the local signal $S_{kn}(T)=O_p\{(nh)^{-1/2}\}$, $k=1,\dots,q$. By Lemma 3, $\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0=O_p(n^{-1/2})$, $\widehat{\boldsymbol{\beta}}_F-\widehat{\boldsymbol{\beta}}_R=o_p(n^{-1/2})$ and $\widehat{\theta}_R(t)$ still has the same asymptotic expansion as in (8).

The test statistic has a similar expansion as in (A3),

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

where

$$
\begin{aligned}
J_1^{\dagger}&=\sum_{k=1}^{q}\sum_{i=1}^{n_k}\epsilon_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\{\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)\},\\
J_2^{\dagger}&=\sum_{k=1}^{q}\sum_{i=1}^{n_k}\epsilon_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}(\widehat{\boldsymbol{\beta}}_F-\widehat{\boldsymbol{\beta}}_R),\\
J_3^{\dagger}&=\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left[(\widehat{\boldsymbol{\beta}}_R-\boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{\widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0)-\theta_k(\boldsymbol{T}_{k,i})\right\}\right.
\end{aligned}
$$

$$- (\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0)^{\mathrm{T}}\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\right\}\Big],$$

$$J_4^\dagger = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}} - \|\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}}\right\},$$

$$J_5^\dagger = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0\|_{\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}} - \|\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0\|_{\widetilde{\boldsymbol{X}}_{k,i}^{\mathrm{T}}\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\widetilde{\boldsymbol{X}}_{k,i}}\right\},$$

$$J_6^\dagger = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\Big\{\|\widetilde{\boldsymbol{X}}_{k,i}(\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + \widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i},\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}$$
$$- \|\widetilde{\boldsymbol{X}}_{k,i}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) + \widehat{\theta}_R(\boldsymbol{T}_{k,i},\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\|^2_{\sum_{j=1}^m \epsilon_{k,ij}\mathcal{D}_{k,ij}}\Big\}.$$

By similar calculations in Lemma 1, $J_2^\dagger + J_3^\dagger + J_5^\dagger + J_6^\dagger = o_p(h^{-1/2})$, hence $J_1^\dagger$ and $J_4^\dagger$ are the dominating terms in $\lambda_n(H_{1n})$.

By straightforward calculations,

$$J_1^\dagger = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\left\{S_{kn}(\boldsymbol{T}_{k,i}) + \mathcal{U}_{F,k}(\boldsymbol{T}_{k,i}) - \mathcal{U}_R(\boldsymbol{T}_{k,i})\right\} + o_p(h^{-1/2})$$
$$= J_1 + R_1^\dagger + o_p(h^{-1/2}),$$

and

$$J_4^\dagger = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}\left\{\|\widehat{\theta}_R(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_0(\boldsymbol{T}_{k,i}) - S_{kn}(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}} - \|\widehat{\theta}_{F,k}(\boldsymbol{T}_{k,i};\boldsymbol{\beta}_0) - \theta_k(\boldsymbol{T}_{k,i})\|_{\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}}\right\}$$
$$= J_4 + R_2^\dagger + R_3^\dagger + o_p(h^{-1/2}),$$

where

$$R_1^\dagger = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i}),$$
$$R_2^\dagger = \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{n_k}S_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i}),$$
$$R_3^\dagger = -\sum_{k=1}^{q}\sum_{i=1}^{n_k}S_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}\mathcal{U}_R(\boldsymbol{T}_{k,i}).$$

It can be shown that $R_3^\dagger = o_p(h^{-1/2})$ and $R_2^\dagger = \mu_{1n} + o_p(h^{-1/2})$, therefore

$$\lambda_n(H_{1n}) = \mu_n + \mu_{1n} + W_n + R_1^\dagger + o_p(h^{-1/2}),$$

where μ_n and W_n are the same as defined in Theorem 1 and Lemma 2. It is easy to see that $\mathrm{E}(R_1^\dagger) = 0$ and

$$\mathrm{var}(R_1^\dagger) = \sum_{k=1}^{q}\sum_{i=1}^{n_k}\mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_{k,i})\boldsymbol{\Delta}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Sigma}_{k,i}\boldsymbol{V}_{k,i}^{-1}\boldsymbol{\Delta}_{k,i}S_{kn}(\boldsymbol{T}_{k,i})\}.$$

Since $R_1^\dagger$ is a linear combination of $\epsilon_{k,ij}$ and W_n only consists of quadratic terms, $R_1^\dagger$ and W_n are uncorrelated and hence asymptotically independent. As a result, $\mathrm{var}(W_n + R_1^\dagger) = \mathrm{var}(W_n) + \mathrm{var}(R_1^\dagger) = \sigma_{1n}^2$, and the asymptotic normal distribution in Theorem 2 follows from de Jong (1987). ∎

A.4 Proof of Theorem 3

We only need to show, for all working covariance $\mathcal{V}$,

$$\min_{\boldsymbol{S}\in\overline{S}_n(\varrho_n^*)} \frac{\sum_{k=1}^q \rho_k \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\boldsymbol{V}_k^{-1}S_{kn}(\boldsymbol{T}_k)\}}{\{\int_{\mathcal{T}} B_2^2(t)/B_1^2(t)dt\}^{1/2}} \le \min_{\boldsymbol{S}\in\overline{S}_n(\varrho_n^*)} \frac{\sum_{k=1}^q \rho_k \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\boldsymbol{\Sigma}_k^{-1}S_{kn}(\boldsymbol{T}_k)\}}{|\mathcal{T}|^{1/2}}, \tag{A5}$$

where $\overline{S}_n(\varrho) = \{\boldsymbol{S}: \sum_{k=1}^q S_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\boldsymbol{\Sigma}_k^{-1}S_{kn}(\boldsymbol{T}_k) = \varrho^2\}$ is the boundary of $\mathcal{S}_n(\varrho)$.

With a change of variable $\mathcal{S}(\boldsymbol{T}) = \boldsymbol{\Sigma}^{-1/2}S(\boldsymbol{T})/\varrho_n^*$,

$$\begin{aligned}(\varrho_n^*)^{-2}\min_{\boldsymbol{S}\in\overline{S}_n(\varrho_n^*)}\sum_{k=1}^q \rho_k \mathrm{E}\{S_{kn}^{\mathrm{T}}(\boldsymbol{T}_k)\boldsymbol{V}_k^{-1}S_{kn}(\boldsymbol{T}_k)\} &= \min_{\mathrm{E}\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\mathcal{S}(\boldsymbol{T})=1} \mathrm{E}\{\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}^{1/2}\mathcal{S}(\boldsymbol{T})\}\\ &\le |\mathcal{T}|^{-1}\int_{\mathcal{T}} \mathscr{A}(t)dt,\end{aligned}$$

where

$$\begin{aligned}\mathscr{A}(t) &= \min_{\mathcal{S}} \frac{\mathrm{E}\{\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}^{1/2}\mathcal{S}(\boldsymbol{T})|T_1=t\}}{\mathrm{E}\{\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\mathcal{S}(\boldsymbol{T})|T_1=t\}}\\ &= \min_{\mathcal{S}} \frac{\mathrm{E}\{\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\boldsymbol{V}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{V}^{-1/2}\mathcal{S}(\boldsymbol{T})|T_1=t\}}{\mathrm{E}\{\mathcal{S}^{\mathrm{T}}(\boldsymbol{T})\mathcal{S}(\boldsymbol{T})|T_1=t\}}.\end{aligned} \tag{A6}$$

Now, to show (A5), we only need

$$[\int_{\mathcal{T}} \mathscr{A}(t)dt\Big]^2 \le |\mathcal{T}|\int_{\mathcal{T}}\left[\frac{\mathrm{E}\{(\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1})_{11}|T_1=t\}}{\mathrm{E}\{(\boldsymbol{V}^{-1})_{11}|T_1=t\}}\right]^2 dt. \tag{A7}$$

Inequality (A7) can be easily shown by the Cauchy–Schwartz inequality if we can show

$$\mathscr{A}(t) \le \frac{\mathrm{E}\{(\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1})_{11}|T_1=t\}}{\mathrm{E}\{(\boldsymbol{V}^{-1})_{11}|T_1=t\}} \quad \text{for all } t\in\mathcal{T}. \tag{A8}$$

Realizing the right-hand side of (A8) is objective function in (A6) evaluated at $\mathcal{S} = \boldsymbol{V}^{-1/2}\boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is a m-dim vector with 1 on the first entry and 0 everywhere else, inequality (A8) holds by the definition of $\mathscr{A}(t)$ in (A6).

A.5 Proof of Theorem 4

By Proposition 2, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$ and $\sup_{t\in\mathcal{T}}|\widehat{\theta}_{F,k}(t) - \theta_k(t)| = O_p[h^2 + \{\log(n)/nh\}^{1/2}]$, therefore the residuals of the full model satisfy $\widehat{\epsilon}_{k,ij} - \epsilon_{k,ij} = O_p[h^2 + \{\log(n)/nh\}^{1/2}]$ uniformly for all k, i and j. Denote $\boldsymbol{\epsilon}_{k,i}^* = \omega_{k,i}\boldsymbol{\epsilon}_{k,i}$, where $P(\omega_{k,i}=1) = P(\omega_{k,i}=-1) = 0.5$. Then we have $\mathrm{E}(\epsilon_{k,ij}^*|\mathcal{X}) = 0$ and $\mathrm{cov}(\boldsymbol{\epsilon}_{k,i}^*, \boldsymbol{\epsilon}_{k',i'}^*|\mathcal{X}) = \boldsymbol{\epsilon}_{k,i}\boldsymbol{\epsilon}_{k,i}^{\mathrm{T}}$ if $(k,i) = (k',i')$ and 0 otherwise. Thus, the bootstrap sample

$\left\{ (Y^*_{k,ij}, \boldsymbol{X}_{k,ij}, T_{k,ij}) \right\}$ satisfy model (1) and the null hypothesis H_0 with the true parameters $\boldsymbol{\beta}^*_0 = \widehat{\boldsymbol{\beta}}_R$ and $\theta^*_0(\cdot) = \widehat{\theta}_R(\cdot)$.

Using the same arguments in Theorem 1, we can show that $\lambda^*_n(H_0) = R^*_1 + R^*_4 + W^*_n + o_p(h^{-1/2})$ where R^*_1, R^*_4 and W^*_n are the same as R_1, R_4, and W_n in Lemma 2 except that $\boldsymbol{\epsilon}_{k,i}$ are replaced by $\boldsymbol{\epsilon}^*_{k,i}$. By similar calculations as in Lemma 2, we have $R^*_1 + R^*_4 = \mu_n + O_p(1)$ and $\mathrm{var}(W^*_n|\mathcal{X}) = \sigma^2_n \times \{1 + o_p(1)\}$. By proposition 3.2 in de Jong (1987), $[\lambda^*_n(H_0)|\mathcal{X}]$ has the same asymptotic normal distribution as $\lambda_n(H_0)$ in Theorem 1 for every event defined on $\mathcal{X}$.