

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

Jin Tang

Statistical Analyst, Pactera, 152nd Avenue Northeast, Redmond, WA, 98052,  
tjcandy@gmail.com

Yehua Li

Associate Professor, Department of Statistics & Statistical Laboratory, Iowa State  
University, Ames, IA 50011, yehuali@iastate.edu

Yongtao Guan

Professor, Department of Management Science, University of Miami, Coral Gables, FL  
33124, yguan@bus.miami.edu

## Abstract

We model generalized longitudinal data from multiple treatment groups by a class of semiparametric analysis of covariance models, which take into account the parametric effects of time dependent covariates and the nonparametric time effects. In these models, the treatment effects are represented by nonparametric functions of time and we propose a generalized quasi-likelihood ratio test procedure to test if these functions are identical. Our estimation procedure is based on profile estimating equations combined with local linear smoothers. We find that the much celebrated Wilks phenomenon which is well established for independent data still holds for longitudinal data if a working independence correlation structure is assumed in the test statistic. However, this property does not hold in general, especially when the working variance function is mis-specified. Our empirical study also shows that incorporating correlation into the test statistic does not necessarily improve the power of the test. The proposed methods are illustrated with simulation studies and a real application from opioid dependence treatments.

**Key Words:** Analysis of variance; Functional data; Longitudinal data; Hypothesis testing; Kernel smoothing; Semiparametric.

**Short title:** Nonparametric Test in Longitudinal Data

# 1 Introduction

In a longitudinal study of comparing efficacy of different treatments, suppose that there are  $q$  treatment groups. Let  $Y_{k,i}(t)$  be the primary variable of interest observed at time  $t$  for the  $i$ th subject receiving the  $k$ th treatment, and let  $\mu_{k,i}(t)$  be the expected value of  $Y_{k,i}(t)$ . We assume that

$$g\{\mu_{k,i}(t)\} = \mathbf{X}_{k,i}^T(t)\boldsymbol{\beta} + \theta_k(t), \quad k = 1, \dots, q, \quad i = 1, \dots, n_k, \quad (1)$$

where  $g(\cdot)$  is a known link function,  $\mathbf{X}_{k,i}(\cdot)$  is a  $p$ -dim covariate vector that may be time dependent,  $\boldsymbol{\beta}$  is a  $p$ -dim regression coefficient,  $\theta_k(\cdot)$  is a nonparametric function, and  $n_k$  is the number of subjects in the  $k$ th treatment group. Denote  $n = \sum_{k=1}^q n_k$  as the total number of subjects in the study. Our primary interest is to compare the treatment effects, which are represented by  $\theta_k$ 's. Specifically, we want to test the following hypotheses:

$$H_0 : \theta_1 = \dots = \theta_q \quad \text{vs.} \quad H_1 : \text{not all } \theta_k \text{'s are the same.} \quad (2)$$

In the context of our application to be detailed in Section 6, we consider two opioid agonist treatments: physician management and physician management plus cognitive behavioral therapy. The response variable  $Y_{k,i}(t)$  for the  $i$ th subject in the  $k$ th treatment group is a binary indicator of opioid use at day  $t$  in a followup period and the associated covariate vector  $\mathbf{X}_{k,i}$  contains baseline information (e.g., age, gender and education) for the same subject. We are interested in studying how the two treatments affect one's opioid use behavior over time differently, after having accounted for effects due to the baseline covariates  $\mathbf{X}_{k,i}$ . To answer this question, we will test hypotheses in the form of (2). We do not require any parametric form for  $\theta_k$ 's in order to maintain modeling flexibility.

The generalized likelihood ratio (GLR) test proposed by Fan et al. (2001) and Fan and Jiang (2005) is a very general approach to test various nonparametric hypotheses under varying coefficient and additive models. Some recent developments and reviews on GLR tests

include Fan and Jiang (2007), Li and Liang (2008) and González-Manteiga and Crujeiras (2013). All these papers studied independent Gaussian data. In our setting, however, the responses within the same subject are longitudinal outcomes and can therefore be strongly correlated. It remains unclear how the GLR test can be extended to dependent data. In this paper, we fill in this gap in literature by developing a generalized quasi-likelihood ratio (GQLR) test for (2), based on longitudinal data obtained from multiple treatment groups.

There is also a vast volume of work on semiparametric modeling of longitudinal data. Our work is most related to the literature on marginal semiparametric regression methods, including Lin and Carroll (2001, 2006) and Wang et al. (2005). However, most of these papers do not model data from multiple treatment groups and focus on estimation rather than testing hypotheses on the nonparametric components. Since the treatment effects are represented by nonparametric functions, model (1) is also related to functional ANOVA models, see, for example, Brumback and Rice (1998), Morris and Carroll (2006) and more recently Zhou et al. (2011). A common approach in these papers is to first express each trajectory as a linear combination of some basis functions and then treat the associated coefficients as random effects in a linear mixed model. Such a hierarchical modeling approach usually requires a relatively large number of repeated measures in each subject as well as many random effects in the model in order to fully capture the features in each longitudinal trajectory. In contrast, our method does not require many repeated measures and is applicable to the so-called sparse functional data (Yao, Müller and Wang, 2005). More importantly, existing methods only produce pointwise confidence intervals for  $\theta_k$ 's, which cannot replace a rigorous test as we are about to propose.

The rest of the paper is organized in the following way. In Section 2, we describe estimation procedures under both the null and alternative hypotheses, and study their asymptotic properties. In Section 3, we propose the new GQLR test and study its null distribution and local power. In Section 4, we discuss implementation issues such as variance and correlation

estimation, approximating the null distribution using bootstrap and bandwidth selection. We illustrate the methodology through simulations in Section 5 and an application on opioid dependence treatment in Section 6. Some concluding remarks are provided in Section 7. Technical proofs and additional simulation results are contained in the supplementary material.

## 2 Estimation and preliminary results

Although model (1) is defined in continuum, in real life we only observe the response variable at discrete time points within a study period  $\mathcal{T}$ . Let  $\mathbf{T}_{k,i} = (T_{k,i1}, \dots, T_{k,im_{k,i}})^T$  be the observation time on the  $i$ th subject in the  $k$ th treatment group, where  $m_{k,i}$  is a positive integer. These time points are usually irregular and subject specific. Denote  $\mathbf{Y}_{k,i} = (Y_{k,i1}, \dots, Y_{k,im_{k,i}})^T$ ,  $\boldsymbol{\mu}_{k,i} = (\mu_{k,i1}, \dots, \mu_{k,im_{k,i}})^T$ ,  $\mathbf{X}_{k,i} = (\mathbf{X}_{k,i1}, \dots, \mathbf{X}_{k,im_{k,i}})^T$ , where  $Y_{k,ij} = Y_{k,i}(T_{k,ij})$ ,  $\mu_{k,ij} = \mu_{k,i}(T_{k,ij})$  and  $\mathbf{X}_{k,ij} = \mathbf{X}_{k,i}(T_{k,ij})$ ,  $j = 1, \dots, m_{k,i}$ . Suppose that the conditional variance of the response is  $\text{var}(Y_{k,ij} | \mathbf{X}_{k,ij}, T_{k,ij}) = \sigma^2(\mu_{k,ij})$ , where  $\sigma^2(\cdot)$  is a conditional variance function. All subjects are assumed to be independent, but the observations within a subject are correlated.

### 2.1 Estimation under both null and alternative hypotheses

We refer to the model under the null hypothesis in (2) as the reduced model and that under the alternative hypothesis as the full model. Denote  $\widehat{\boldsymbol{\beta}}_R$  and  $\widehat{\theta}_R(t)$  as the estimators under the reduced model and  $\widehat{\boldsymbol{\beta}}_F$  and  $\widehat{\theta}_{F,k}(t)$ ,  $k = 1, \dots, q$ , as the estimators under the full model. Our estimation procedures under both models are based on profile-kernel estimating equations.

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

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

Let  $K(\cdot)$  be a symmetric probability density function and denote  $K_h(t) = h^{-1}K(t/h)$  where  $h$  is the bandwidth. Put  $\mathbf{U}_{k,i}(t) = \{\mathbf{U}_{k,i1}(t), \dots, \mathbf{U}_{k,im_i}(t)\}^T$  with  $\mathbf{U}_{k,ij}(t) = \{1, (T_{k,ij} - t)/h\}^T$ . For a given  $\boldsymbol{\beta}$ ,  $\theta(t)$  is estimated by solving the following local linear estimating equation regarding  $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ ,

$$\sum_{k=1}^q \sum_{i=1}^{n_k} \mathbf{U}_{k,i}(t)^T \Delta_{k,i}(t) \mathcal{W}_{k,i}^{-1} \mathbf{K}_h(\mathbf{T}_{k,i} - t) \{\mathbf{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(t)\} = 0, \quad (3)$$

where  $\mathbf{K}_h(\mathbf{T}_{k,i} - t) = \text{diag}\{K_h(T_{k,ij} - t)\}_{j=1}^{m_{k,i}}$ ,  $\boldsymbol{\mu}_{k,i}(t) = (\mu_{k,i1}, \dots, \mu_{k,im_{k,i}})^T(t)$ ,  $\mu_{k,ij}(t) = g^{-1}\{\mathbf{X}_{k,ij}^T \boldsymbol{\beta} + \mathbf{U}_{k,ij}^T(t) \boldsymbol{\alpha}\}$ ,  $\Delta_{k,i}(t) = \text{diag}\{\mu_{k,ij}^{(1)}(t)\}_{j=1}^{m_{k,i}}$ ,  $\mu_{k,ij}^{(1)}(t)$  is the first derivative of  $\mu(\cdot) = g^{-1}(\cdot)$  evaluated at  $\mathbf{X}_{k,ij}^T \boldsymbol{\beta} + \mathbf{U}_{k,ij}^T(t) \boldsymbol{\alpha}$ , and  $\mathcal{W}_{k,i}$  is a weight matrix to be specified below. The local linear estimator is given by  $\hat{\theta}_R(t; \boldsymbol{\beta}) = \hat{\alpha}_0$ , where  $(\hat{\alpha}_0, \hat{\alpha}_1)$  is the solution of (3). Then  $\hat{\boldsymbol{\beta}}_R$  is obtained by solving

$$\sum_{k=1}^q \sum_{i=1}^{n_k} \left\{ \mathbf{X}_{k,i}^T + \frac{\partial \hat{\theta}_R(\mathbf{T}_{k,i}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right\} \Delta_{k,i}(\mathbf{T}_{k,i}) \mathcal{W}_{k,i}^{-1} \{\mathbf{Y}_{k,i} - g^{-1}\{\mathbf{X}_{k,i} \boldsymbol{\beta} + \hat{\theta}_R(\mathbf{T}_{k,i}; \boldsymbol{\beta})\}\} = 0. \quad (4)$$

At convergence of the algorithm described above, the nonparametric estimator needs a final update as  $\hat{\theta}_R(t) = \hat{\theta}_R(t, \hat{\boldsymbol{\beta}}_R)$ .

As shown by Lin and Carroll (2001), the most efficient estimators within the class defined by (3) are obtained by setting  $\mathcal{W}_{k,i} = \text{diag}\{\omega(\mu_{k,ij})\}_{j=1}^{m_{k,i}}$  where  $\omega(\cdot)$  is a working variance function. Similar kernel estimators are widely used in longitudinal and functional data analysis, see Fan and Li (2004), Yao et al. (2005), Hall et al. (2006, 2008) and Li and Hsing (2010). The working variance function  $\omega(\cdot)$  can be replaced by a nonparametric variance estimator described in Section 4.1.

We now consider estimation under the full model, where we need to estimate  $\theta_k(\cdot)$  using the  $k$ th treatment group only. For a given  $\boldsymbol{\beta}$ , the profile local linear estimator for  $\theta_k(t)$  is given by  $\hat{\theta}_{F,k}(t, \boldsymbol{\beta}) = \hat{\alpha}_0$ , where  $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_0, \hat{\alpha}_1)^T$  is the solution of

$$\sum_{i=1}^{n_k} \mathbf{U}_{k,i}(t)^T \Delta_{k,i}(t) \mathcal{W}_{k,i}^{-1} \mathbf{K}_h(\mathbf{T}_{k,i} - t) \{\mathbf{Y}_{k,i} - \boldsymbol{\mu}_{k,i}(t)\} = 0. \quad (5)$$

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

$$\sum_{k=1}^q \sum_{i=1}^{n_k} \left\{ \mathbf{X}_{k,i}^T + \frac{\partial \widehat{\theta}_{F,k}(\mathbf{T}_{k,i}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right\} \Delta_{k,i}(\mathbf{T}_{k,i}) \mathcal{W}_{k,i}^{-1} [\mathbf{Y}_{k,i} - g^{-1}\{\mathbf{X}_{k,i} \boldsymbol{\beta} + \widehat{\theta}_{F,k}(\mathbf{T}_{k,i}; \boldsymbol{\beta})\}] = 0. \quad (6)$$

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

## 2.2 Asymptotic properties of the estimators

Before proposing a test procedure for the hypotheses in (2), we first investigate the asymptotic properties of the profile-kernel estimators of  $\boldsymbol{\beta}$  and  $\theta(t)$  under both the full and reduced models. Denote the true parameters as  $\boldsymbol{\beta}_0$  and  $\theta_{k0}(t)$ ,  $k = 1, \dots, q$ . Under the reduced model,  $\theta_{10} = \dots = \theta_{q0} \equiv \theta_0$ . For ease of exposition, we assume that the number of observations per subject is a constant in our theoretical derivations, i.e.  $m_{k,i} = m < \infty$  for all  $k$  and  $i$ . For 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  $t \in \mathcal{T}$ . We assume that there exist constants  $0 < \rho_1, \dots, \rho_q < 1$  such that  $\sum_{k=1}^q \rho_k = 1$  and  $n_k/n - \rho_k = o(n^{-1/2})$ .

Our problem is fundamentally different from those in the GLR test literature because of the presence of within-subject correlation. Define the errors as  $\varepsilon_{k,ij} = Y_{k,ij} - g^{-1}\{\mathbf{X}_{k,ij}^T \boldsymbol{\beta}_0 + \theta_{k0}(T_{k,ij})\}$  and the error vector within the  $(k, i)$ th subject as  $\boldsymbol{\varepsilon}_{k,i} = (\varepsilon_{k,i1}, \dots, \varepsilon_{k,im})^T$ . We consider  $\mathbf{X}_{k,ij}$  and  $\varepsilon_{k,ij}$  as discrete realizations of the continuous longitudinal processes  $\mathbf{X}(t)$  and  $\varepsilon(t)$ ,  $t \in \mathcal{T}$ . Define the conditional variance and correlation functions of  $\varepsilon(t)$  as

$$\sigma^2(\mu) = \text{var} \left[ \varepsilon(t) \middle| g^{-1}\{\mathbf{X}^T(t) \boldsymbol{\beta} + \theta(t)\} = \mu \right], \quad \mathcal{R}(s, t; \boldsymbol{\tau}) = \text{corr}\{\varepsilon(s), \varepsilon(t) \mid \boldsymbol{\tau}\}, \quad (7)$$

where  $\boldsymbol{\tau}$  is an unknown correlation parameter. Many authors modeled the variance function  $\sigma^2(\cdot)$  as a nonparametric function but the correlation function  $\mathcal{R}(s, t; \boldsymbol{\tau})$  as a member of a

parametric family, such as the autoregressive (AR) or autoregressive and moving-average (ARMA) correlations (see Fan et al., 2007 and Fan and Wu, 2008). The within-subject covariance matrix is

$$\boldsymbol{\Sigma}_{k,i} = \mathbf{S}_{k,i} \mathcal{R}_{k,i}(\boldsymbol{\tau}) \mathbf{S}_{k,i}, \quad (8)$$

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

In addition to the asymptotic framework described above, we make the following assumptions for the asymptotic theory.

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

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

Assume all functions defined above are Lipschitz continuous.

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

(C.3) Assume that  $h \rightarrow 0$  as  $n \rightarrow \infty$ ,  $nh^2 \rightarrow \infty$ ,  $nh^8 \rightarrow 0$ .

Let  $\tilde{\mathbf{X}}_{k,i}$  be an  $m \times p$  matrix with the  $j$ th row being  $\mathbf{X}_{k,ij} - \boldsymbol{\mu}_{X,k}(T_{k,ij})$ , where  $\boldsymbol{\mu}_{X,k}(\cdot)$  is defined in condition (C.1). For ease of exposition, we often suppress the index  $i$ , and write  $\boldsymbol{\Sigma}_k$ ,  $\Delta_k$ ,  $\mathcal{W}_k$  and  $\tilde{\mathbf{X}}_k$  as a generic version of these quantities for a typical subject receiving

the  $k$ th treatment. When the null hypothesis in (2) is true,  $\Sigma_k \equiv \Sigma$ ,  $\Delta_k \equiv \Delta$ ,  $\mathcal{W}_k \equiv \mathcal{W}$  and  $\tilde{\mathbf{X}}_k \equiv \tilde{\mathbf{X}}$  for  $k = 1, \dots, q$ . The following proposition provides asymptotic expansions of the estimators under the reduced model.

**PROPOSITION 1** *Suppose the null hypothesis in (2) is true and the asymptotic framework described above and conditions (C.1)-(C.3) hold. Then the parametric estimator has the following asymptotic expansion*

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

where  $\mathbf{D}_R = \text{E}(\tilde{\mathbf{X}}^T \Delta \mathcal{W}^{-1} \Delta \tilde{\mathbf{X}})$  and  $\boldsymbol{\mathcal{E}}_R = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \tilde{\mathbf{X}}_{k,i}^T \Delta_{k,i} \mathcal{W}_{k,i}^{-1} \boldsymbol{\varepsilon}_{k,i}$ .

The nonparametric estimator has the following asymptotic expansion

$$\hat{\theta}_R(t) - \theta_0(t) = \frac{1}{2} \theta_0^{(2)}(t) h^2 + \mathcal{U}_R(t) - \boldsymbol{\mu}_X^T(t) (\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) + o_p(n^{-1/2}), \quad (10)$$

where  $\mathcal{U}_R(t) = \{nmB_1(t)\}^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^m \mu_{k,ij}^{(1)} \omega_{k,i}^{jj} K_h(T_{k,ij} - t) \varepsilon_{k,ij}$ ,  $\mu_{k,ij}^{(1)}$  and  $\omega_{k,i}^{jj}$  are shorthands for  $\mu^{(1)}\{\mathbf{X}_{k,ij}^T \boldsymbol{\beta}_0 + \theta_{k0}(T_{k,ij})\}$  and  $\omega(\mu_{k,ij})^{-1}$  with  $\mu_{k,ij}$  being evaluated at the true parameter values  $\boldsymbol{\beta}_0$  and  $\theta_{k0}$ , and  $B_1(t)$  is defined in condition (C.1).

When the null hypothesis is true, Model (1) reduces to a generalized partially linear model, the asymptotic results in Proposition 1 are standard (Lin and Carroll, 2001, Fan and Li, 2004, and Wang et al., 2005) and hence the proof is omitted. With similar arguments, one can easily show the following results regarding the estimators under the full model.

**PROPOSITION 2** *Under the full model (1) and conditions (C.1)-(C.3),*

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

where  $\mathbf{D}_F = \sum_{k=1}^q \rho_k \text{E}(\tilde{\mathbf{X}}_k^T \Delta_k \mathcal{W}_k^{-1} \Delta_k \tilde{\mathbf{X}}_k)$  and  $\boldsymbol{\mathcal{E}}_F = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \tilde{\mathbf{X}}_{k,i}^T \Delta_{k,i} \mathcal{W}_{k,i}^{-1} \boldsymbol{\varepsilon}_{k,i}$ .

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

$$\hat{\theta}_{F,k}(t) - \theta_{k0}(t) = \frac{1}{2} \theta_{k0}^{(2)}(t) h^2 + \mathcal{U}_{F,k}(t) - \boldsymbol{\mu}_{X,k}^T(t) (\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) + o_p(n_k^{-1/2}), \quad (12)$$



for  $k = 1, \dots, q$  where  $\mathcal{U}_{F,k}(t) = \{n_k m B_{1k}(t)\}^{-1} \sum_{i=1}^{n_k} \sum_{j=1}^m \mu_{k,ij}^{(1)} \omega_{k,i}^{jj} K_h(T_{k,ij} - t) \varepsilon_{k,ij}$  and  $B_{1k}(t)$  is defined in condition (C.1).

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

### 3 Generalized quasi-likelihood ratio test

#### 3.1 Test procedure and null distribution

We now direct our focus back to testing the hypotheses in (2). In this section, we introduce our test procedure based on a quasi-likelihood function, and study the asymptotic distribution of the test statistic under the null hypothesis.

A quasi-likelihood function (McCullagh and Nelder, 1989)  $\mathcal{Q}$  satisfies

$$\frac{\partial \mathcal{Q}(\boldsymbol{\mu}, \mathbf{Y})}{\partial \boldsymbol{\mu}} = \mathbf{V}(\boldsymbol{\mu})^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \quad (13)$$

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

$$\ell(\theta, \boldsymbol{\beta}) = \sum_{k=1}^q \sum_{i=1}^{n_k} \mathcal{Q}[g^{-1}\{\mathbf{X}_{k,i}\boldsymbol{\beta} + \theta_k(\mathbf{T}_{k,i})\}, \mathbf{Y}_{k,i}]. \quad (14)$$

The GQLR test statistic is defined as the difference of the quasi-likelihoods under the full and reduced models

$$\lambda_n(H_0) = \sum_{k=1}^q \sum_{i=1}^{n_k} \left( \mathcal{Q}[g^{-1}\{\mathbf{X}_{k,i}\widehat{\boldsymbol{\beta}}_F + \widehat{\theta}_{F,k}(\mathbf{T}_{k,i})\}, \mathbf{Y}_{k,i}] - \mathcal{Q}[g^{-1}\{\mathbf{X}_{k,i}\widehat{\boldsymbol{\beta}}_R + \widehat{\theta}_R(\mathbf{T}_{k,i})\}, \mathbf{Y}_{k,i}] \right). \quad (15)$$

Under the null hypothesis, let  $\boldsymbol{\Sigma}$  and  $\mathbf{V}$  be generic copies of  $\boldsymbol{\Sigma}_{k,i}$  and  $\mathbf{V}_{k,i}$  in a typical subject. These matrices are random because the within-subject correlation might depend on the subject-specific time vector  $\mathbf{T}$  and the conditional variance might depend on the

conditional mean vector  $\boldsymbol{\mu} = g^{-1}\{\mathbf{X}\boldsymbol{\beta}_0 + \theta_0(\mathbf{T})\}$ . Denote  $\sigma_{j\ell}$ ,  $\nu^{j\ell}$  and  $\Delta_{j\ell}$  as the  $(j, \ell)$ th entry of  $\boldsymbol{\Sigma}$ ,  $\mathbf{V}^{-1}$  and  $\Delta$  respectively and they are considered random as well. Define

$$\begin{aligned} B_2(t) &= \frac{1}{m} \sum_{j=1}^m \mathbb{E}(\Delta_{jj}^2 \nu^{jj} | T_j = t) f(t), & B_3(t) &= \frac{1}{m} \sum_{j=1}^m \mathbb{E}\{\sigma_{jj} \Delta_{jj}^2 (\omega^{jj})^2 | T_j = t\} f(t), \\ B_4(t) &= \frac{1}{m} \sum_{j_1=1}^m \mathbb{E}\left(\sum_{j_2=1}^m \sum_{j_3=1}^m \nu^{j_1 j_2} \sigma_{j_2 j_3} \nu^{j_3 j_1} \Delta_{j_1 j_1}^2 \Big| T_{j_1} = t\right) f(t), \\ B_5(t) &= \frac{1}{m} \sum_{j'=1}^m \mathbb{E}\left(\sum_{j=1}^m \sigma_{jj'} \nu^{jj'} \Delta_{j'j'}^2 \omega^{j'j'} \Big| T_{j'} = t\right) f(t). \end{aligned} \quad (16)$$

Denote  $\nu_K = \int K^2(t) dt$  and  $K * K$  as the convolution of the kernel function  $K$ , so that  $K * K(t) = \int_{-\infty}^{\infty} K(s) K(t-s) ds$ . The following theorem provides the asymptotic distribution of the GQLR test statistic under the null hypothesis, where the proof is provided in Section W.1 of the supplementary material .

**THEOREM 1** *Under the asymptotic framework described in Section 2.2 and conditions (C.1)-(C.3), we assume that  $B_2(t) - B_5(t)$  defined in (16) exist and are Lipschitz continuous in  $t$ . In addition, assume that  $nh^5 \rightarrow 0$ , then under the null hypothesis in (2)*

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

where  $d_n = o_p(h^{-1/2})$ ,

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

**Remarks:** 1. The asymptotic normal distribution in Theorem 1 is derived using the central limit theorem for generalized quadratic forms developed by de Jong (1987), which is the same theoretical tool that is used to develop the asymptotic distribution of the GLR test for independent Gaussian data. Compared with the GLR test, the main theoretical challenges

in our problem are: 1) the data are correlated; 2) the covariance matrices  $\boldsymbol{\Sigma}_{k,i}$  and the local expansion of the link function  $\Delta_{k,i}$  are random and subject specific; 3) we allow the working covariances  $\mathcal{W}_{k,i}$  and  $\mathbf{V}_{k,i}$  to be mis-specified. To be more specific on the second challenge,  $\boldsymbol{\Sigma}_{k,i}$  and  $\Delta_{k,i}$  depend on the mean vector  $\boldsymbol{\mu}_{k,i}$  and the within-subject correlation may depend on the subject specific time vector  $\mathbf{T}_{k,i}$ . As a result, our proof is more involved when calculating the asymptotic mean and variance of the test statistic.

2. For nonparametric hypothesis testing in independent data, Fan et al. (2001) established a property called the Wilks phenomenon for the GLR test, i.e. the asymptotic distribution of the test statistic under the null hypothesis does not depend on the unknown true parameters. Indeed, when the likelihood function is used and correctly specified, this property holds for a wide range of problems. However, for generalized longitudinal data, working covariance matrices, generalized estimating equations and quasi-likelihoods are commonly used, and in many situations these models do not need to be correctly specified. When the variance (covariance) of  $\varepsilon$  depends on the mean, and if  $\mathbf{V}$ ,  $\mathcal{W}$  and  $\boldsymbol{\Sigma}$  are different,  $B_1(t) - B_5(t)$  and thus the asymptotic distribution of  $\lambda_n(H_0)$  in Theorem 1 depend on the true parameters  $\boldsymbol{\beta}_0$  and  $\theta_0(t)$ . In this case, the Wilks phenomenon does not hold in general for the test statistic defined in (15).

Next we provide two corollaries of Theorem 1, where the asymptotic distribution of  $\lambda_n(H_0)$  does not depend on  $\boldsymbol{\beta}_0$  and  $\theta_0(t)$ . Recall that the true covariance matrix  $\boldsymbol{\Sigma}$  has the structure as in (8) and denote  $\boldsymbol{\Sigma}_d = \mathbf{S}^2$  as the diagonal variance matrix. We now investigate the situation when the variance function is correctly specified for both the estimation and test procedures but the working correlation for the test is misspecified. In other words, we assume  $\mathcal{W} = \boldsymbol{\Sigma}_d$  and  $\mathbf{V} = \mathbf{S}\mathcal{C}(\boldsymbol{\tau})\mathbf{S}$ , where  $\mathcal{C}(\boldsymbol{\tau})$  is a working correlation matrix that may be different from the true correlation matrix  $\mathcal{R}(\boldsymbol{\tau})$ .

When  $\boldsymbol{\Sigma} = \mathbf{S}\mathcal{R}(\boldsymbol{\tau})\mathbf{S}$ ,  $\mathcal{W} = \boldsymbol{\Sigma}_d$  and  $\mathbf{V} = \mathbf{S}\mathcal{C}(\boldsymbol{\tau})\mathbf{S}$ , one can show  $B_j(t) = B_{j^\dagger}(t)B_1(t)$  for

$j = 2, \dots, 5$ , where

$$\begin{aligned}
B_{2\uparrow}(t) &= \frac{1}{m} \sum_{j=1}^m \mathbb{E}[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t], & B_{3\uparrow}(t) &= 1, \\
B_{4\uparrow}(t) &= \frac{1}{m} \sum_{j=1}^m \mathbb{E}[\{\mathcal{C}^{-1}(\boldsymbol{\tau})\mathcal{R}(\boldsymbol{\tau})\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t], \\
B_{5\uparrow}(t) &= \frac{1}{m} \sum_{j=1}^m \mathbb{E}[\{\mathcal{R}(\boldsymbol{\tau})\mathcal{C}^{-1}(\boldsymbol{\tau})\}_{jj}|T_j = t].
\end{aligned} \tag{17}$$

The asymptotic null distribution of the test statistic under this special case is given below.

**COROLLARY 1** *Under the setting of Theorem 1, suppose  $\mathcal{W} = \boldsymbol{\Sigma}_d$ ,  $\mathbf{V} = \mathbf{S}\mathcal{C}(\boldsymbol{\tau})\mathbf{S}$  and  $\boldsymbol{\Sigma} = \mathbf{S}\mathcal{R}(\boldsymbol{\tau})\mathbf{S}$ , then the asymptotic distribution of  $\lambda_n(H_0)$  under the null hypothesis in (2) is*

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

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

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

The functions  $B_{j\uparrow}$ ,  $j = 2, \dots, 5$ , are defined in (17).

From the expressions of  $B_{j\uparrow}(t)$ ,  $j = 2, \dots, 5$ , it is easy to see that they do not depend on  $\boldsymbol{\beta}_0$  and  $\theta_0$ . It follows that the asymptotic distribution of  $\lambda_n(H_0)$  does not depend on the true values of these parameters when the variance is correctly specified but the correlation is misspecified. However, the asymptotic distribution of  $\lambda_n(H_0)$  depends on the true correlation function  $\mathcal{R}(\boldsymbol{\tau})$  which is generally unknown. Thus, it is difficult to simulate the asymptotic distribution in Corollary 1.

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

**COROLLARY 2** Under the setting of Theorem 1, if  $\mathbf{V} = \mathcal{W} = \boldsymbol{\Sigma}_d$ , the asymptotic distribution  $\lambda_n(H_0)$  can be simplified to

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

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

Corollary 2 implies that, if a working independence covariance is used in both estimation and hypothesis testing and if the true variance function is used, the asymptotic distribution of  $\lambda_n(H_0)$  does not depend on  $\boldsymbol{\beta}_0$ ,  $\theta_0(t)$  or the true correlation structure  $\mathcal{R}(\boldsymbol{\tau})$ . This Wilks result makes it easy to assess the distribution of  $\lambda_n(H_0)$  using bootstrap, see our discussions in Section 4.2.

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

To study the local power of the GQLR test, we consider a contiguous alternative hypothesis

$$H_{1n} : \theta_k(t) = \theta_0(t) + G_{kn}(t), \quad k = 1, \dots, q, \quad \text{with} \quad \sum_{k=1}^q \rho_k G_{kn}(t) = 0, \quad (18)$$

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

Consider the test statistic in (15), and call it  $\lambda_n(H_{1n})$  instead. The following theorem gives the asymptotic distribution of the test statistic under the local alternative (18) where the proof is given the supplementary material.

**THEOREM 2** Suppose that assumptions (C.1) – (C.3) and the local alternative (18) hold,  $nh^5 \rightarrow 0$ , and the functions  $G_{kn}(t)$ 's are twice continuously differentiable. Denote  $\mu_{1n} = \frac{1}{2} \sum_{k=1}^q \sum_{i=1}^{n_k} \text{E}\{G_{kn}^T(\mathbf{T}_{k,i}) \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} G_{kn}(\mathbf{T}_{k,i})\}$  and we assume there exists a constant  $C_G$  such that

$$h \times \mu_{1n} \rightarrow C_G < \infty. \quad (19)$$

Then the test statistic has the following limiting distribution

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

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

An approximate level- $\alpha$  test based on the GQLR test statistic is  $\varphi_n = I\{\lambda_n(H_{1n}) - \mu_n > z_\alpha\sigma_n\}$ , where  $z_\alpha$  is the upper  $100 \times \alpha$  percentile of  $N(0, 1)$ , and we reject the null hypothesis if  $\varphi_n = 1$ . Let  $\Phi(\cdot)$  be the cumulative distribution function of  $N(0, 1)$ . Then the type II error of the test is

$$\beta(\alpha, \mathbf{G}_n) = P(\lambda_n(H_{1n}) - \mu_n < z_\alpha\sigma_n) \approx \Phi(\sigma_{1n}^{-1}\sigma_n z_\alpha - \sigma_{1n}^{-1}\mu_{1n}), \quad (20)$$

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

$$\mathcal{G}_n(\varrho) = [\mathbf{G}_n = (G_{1n}, \dots, G_{qn})^T : \sum_{k=1}^q \rho_k \mathbb{E}\{G_{kn}^T(\mathbf{T}_k)\Delta_k\mathbf{V}_k^{-1}\Delta_k G_{kn}(\mathbf{T}_k)\} \geq \varrho^2],$$

and the maximum probability of type II errors as

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

Following Ingster (1993) and Fan et al. (2001), define the minimax rate of a test  $\varphi$  with a type II error  $\beta(\alpha, \varrho)$  as  $\varrho_n$  such that

- (a) for any  $\varrho > \varrho_n$ ,  $\alpha > 0$ , and  $\beta > 0$ , there exists a constant  $c$  such that  $\beta(\alpha, c\varrho) \leq \beta + o(1)$ ;
- (b) for any sequence  $\varrho_n^* = o(\varrho_n)$ , there exist  $\alpha > 0$  and  $\beta > 0$  such that for any  $c > 0$ 

$$P(\varphi = 1|H_0) = \alpha + o(1) \text{ and } \liminf_{n \rightarrow \infty} \beta(\alpha, c\varrho_n^*) > \beta.$$

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

**THEOREM 3** *Under conditions (C.1) - (C.3), the minimax rate of the GQLR test is  $\varrho_n(h) = n^{-4/9}$  with  $h = c^*n^{-2/9}$  for a constant  $c^*$ .*

The proof of Theorem 3 is provided in the supplementary material. Theorem 3 shows that the GQLR test achieves the minimax optimal rate of Ingster (1993).

## 4 Implementation issues

### 4.1 Estimation of the variance and correlation functions

The proposed estimation and test procedures involve a working variance function and also a working correlation function if working independence is not assumed. In practice, both the working variance and working correlation functions need to be estimated from the data.

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

$$\begin{aligned} \widehat{\sigma}^2(\mu) &= \mathbf{e}^T \left\{ \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^{m_{k,i}} \mathbb{U}_{k,ij}(\mu) \mathbb{U}_{k,ij}^T(\mu) K_h(\tilde{\mu}_{k,ij} - \mu) \right\}^{-1} \\ &\quad \times \left\{ \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^{m_{k,i}} \mathbb{U}_{k,ij}(\mu) K_h(\tilde{\mu}_{k,ij} - \mu) \tilde{\varepsilon}_{k,ij}^2 \right\}, \end{aligned}$$

where  $\mathbf{e} = (1, 0)^T$ ,  $\mathbb{U}_{k,ij}(\mu) = \{1, (\tilde{\mu}_{k,ij} - \mu)/h\}^T$ ,  $\tilde{\mu}_{k,ij}$  are pilot estimators of  $\mu_{k,ij}$  from the full model by setting  $\mathcal{W}_{k,i}$  to be identity matrices and  $\tilde{\varepsilon}_{k,ij}$  are the residuals from the pilot estimates. Li (2011) showed that the local linear variance estimator is uniformly consistent to the true variance function and using the local linear variance estimator as the working variance in the estimators defined in Section 2 is asymptotically as efficient as using the true variance function. Using similar arguments, we can show that, if the nonparametric variance estimator  $\widehat{\sigma}^2(\mu)$  is used as the working variance in the GQLR test statistic, the asymptotic distribution of  $\lambda_n(H_0)$  is the same as if  $\sigma^2(\mu)$  was known.

Assuming that the working covariance matrices have the structure  $\mathbf{V}_{k,i} = \mathbf{S}_{k,i} \mathcal{C}_{k,i}(\boldsymbol{\tau}) \mathbf{S}_{k,i}$ , where  $\mathcal{C}_{k,i}(\boldsymbol{\tau})$  is the working correlation matrix for the  $(k, i)$ th subject with a correlation parameter  $\boldsymbol{\tau}$ , then  $\boldsymbol{\tau}$  can be estimated by the quasi-maximum likelihood (QMLE) method of

Fan et al. (2007). The estimator  $\hat{\boldsymbol{\tau}}$  is the maximizer of

$$-\frac{1}{2} \sum_{k=1}^q \sum_{i=1}^{n_k} \left\{ \log |\mathcal{C}_{k,i}(\boldsymbol{\tau})| + \tilde{\boldsymbol{\varepsilon}}_{k,i}^{\top} \hat{\boldsymbol{S}}_{k,i}^{-1} \mathcal{C}_{k,i}^{-1}(\boldsymbol{\tau}) \hat{\boldsymbol{S}}_{k,i}^{-1} \tilde{\boldsymbol{\varepsilon}}_{k,i} \right\},$$

where  $\hat{\boldsymbol{S}}_{k,i} = \text{diag}\{\hat{\sigma}(\mu_{k,ij})\}_{j=1}^{m_{k,i}}$ . When the working correlation model is correctly specified, Fan and Wu (2008) showed that  $\hat{\boldsymbol{\tau}}$  is root- $n$  consistent to the true correlation parameter.

## 4.2 Evaluating the null distribution with bootstrap

As indicated in Theorem 1, the null distribution of the GQLR test statistic is asymptotically normal. This seems to suggest that we can estimate the various terms in the functions  $B_1(t)$  -  $B_5(t)$  and use the normal distribution with the estimated asymptotic mean and variance to decide the rejection region of the test. However, the numerical studies of Härdle and Mammen (1993) show that such a “plug-in” method (i.e. plug-in estimates into the mean and variance of  $\lambda_n(H_0)$ ) does not work well even for independent data. See their Figures 1 - 4 for a comparison between the true density of the test statistic, normal density with estimated mean and variance and the bootstrap density. The weak performance of the plug-in method is due to 1) the distribution of a nonparametric test statistic converges in a rather slow rate and normal approximation may not work well under small sample size; 2) estimators of the asymptotic mean and variance of  $\lambda_n(H_0)$  also converge in a slow nonparametric rate. In our setting, estimating  $B_1(t)$  -  $B_5(t)$  consistently also requires estimating the covariance structure consistently, but in real data analysis the covariance structure is often misspecified. For these reasons, most authors advocate bootstrap procedures for nonparametric tests, e.g. Härdle and Mammen (1993) and Fan and Jiang (2005).

To estimate the null distribution using bootstrap, we need to generate data that satisfy the null hypothesis and closely resemble the true data. Because the GQLR test is in general lack of the Wilks property, a consistent bootstrap procedure requires the bootstrap samples to have the same correlation structure as the real data. Since the real correlation structure



is unknown and often misspecified, such a bootstrap procedure can lead to an inconsistent estimator of the null distribution and hence an invalid test.

One alternative approach is to assume working independence for both estimation and test as described in Corollary 2, in which case the distribution of  $\lambda_n(H_0)$  enjoys the Wilks property. Since, under this situation, the asymptotic distribution of  $\lambda_n(H_0)$  is the same as for independent data, we can use a simple wild bootstrap procedure proposed by Mammen (1993) for independent data. The detailed procedure is given below.

- (i) Estimate both the full and reduced models from the original data and estimate the variance function by the nonparametric estimator  $\hat{\sigma}^2(\mu)$  described in Section 4.1 using residuals from the full model. Evaluate the GQLR test statistic  $\lambda_n(H_0)$  under working independence.
- (ii) For each subject, regenerate the response vector  $Y_{k,ij}^* = g^{-1}(\mu_{k,ij}^*) + \varepsilon_{k,ij}^*$ , where  $\mu_{k,ij}^* = \mathbf{X}_{k,ij}^T \hat{\boldsymbol{\beta}}_R + \hat{\theta}_R(T_{k,ij})$ ,  $\varepsilon_{k,ij}^* = \xi_{k,ij} \hat{\varepsilon}_{k,ij}$ ,  $\hat{\varepsilon}_{k,ij}$ 's are the residuals from the full model and  $\xi_{k,ij}$ 's are independent random perturbation factors with mean 0 and variance 1.
- (iii) Calculate the GQLR test statistic  $\lambda_n^*(H_0)$  from the bootstrap sample  $\{\mathbf{Y}_{k,i}^*, \mathbf{X}_{k,i}, \mathbf{T}_{k,i}\}$  using the exact same procedure for the original data.
- (iv) Repeat Steps (ii) and (iii) a large number of times to obtain the bootstrap replicates  $\lambda_n^*(H_0)$ . The estimated p-value is the percentage of  $\lambda_n^*(H_0)$  that are greater than  $\lambda_n(H_0)$ .

Davidson and Flachaire (2008) advocate to generate the perturbation factor  $\xi_{k,ij}$  in Step (ii) from a simple Rademacher distribution, i.e.  $P(\xi = 1) = P(\xi = -1) = 1/2$ , which is also what we use in our numerical studies. Put  $\mathcal{X} = \{(\mathbf{X}_{k,i}, \mathbf{T}_{k,i}), i = 1, \dots, m_k, k = 1, \dots, q\}$ . The following theorem shows the consistency of the bootstrap procedure above. It states that, conditioning on  $\mathcal{X}$ ,  $\lambda_n^*(H_0)$  from the working independent bootstrap procedure above

has the same asymptotic normal distribution as  $\lambda_n(H_0)$  in Corollary 2. The proof of the theorem is provided in the supplementary material.

**THEOREM 4** *Under the same assumptions as Corollary 2,*

$$P[\sigma_{n^*}^{-1}\{\lambda_n^*(H_0) - \mu_{n^*} - d_{n^*}\} < x \mid \mathcal{X}] \rightarrow \Phi(x) \quad \text{in probability for any } x,$$

where  $\sigma_{n^*}$ ,  $\mu_{n^*}$  and  $d_{n^*}$  are as defined in Corollary 2.

### 4.3 Bandwidth selection

For bandwidth selection, we adopt a leave-one-subject-out cross-validation (Rice and Silverman, 1991) that is tailored to our data structure. Let  $\hat{h}_{cv}$  be the minimizer of

$$CV(h) = \sum_{k=1}^q \sum_{i=1}^{n_k} \mathcal{Q}[g^{-1}\{\mathbf{X}_{k,i} \hat{\boldsymbol{\beta}}_h^{-(k,i)} + \hat{\theta}_{k,h}^{-(k,i)}(\mathbf{T}_{k,i})\}, \mathbf{Y}_{k,i}],$$

where  $\mathcal{Q}(\cdot)$  is the quasi-likelihood function and  $\hat{\boldsymbol{\beta}}_h^{-(k,i)}$  and  $\hat{\theta}_{k,h}^{-(k,i)}(\cdot)$  are the full model estimators with bandwidth  $h$  by removing the  $(k, i)$ th subject from the data set. It is well-known that cross-validation estimates the optimal bandwidth for estimation, which is of order  $n^{-1/5}$  (Xia and Li, 2002). To make the bandwidth follow the optimal order for hypothesis test suggested in Theorem 3, we propose to use  $\hat{h} = \hat{h}_{cv} \times n^{-1/45}$ . 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.

## 5 Simulation study

To illustrate the performance of the GQLR procedure and the Wilks phenomenon, we consider two simulation settings: Gaussian longitudinal data with heterogenous variance and binary longitudinal data. To save space, we only present the simulation results on Gaussian data. Results of the binary case can found in the online supplementary material.

We generate data from the following model:

$$Y_{k,ij} = X_{k,ij}\beta + \theta_k(T_{k,ij}) + \varepsilon_{k,ij}, \quad k = 1, \dots, 4, \quad i = 1, \dots, 50, \quad j = 1, \dots, 4,$$

where  $T_{k,ij}$  are independent random variables following a uniform distribution on  $[0, 1]$  and  $X_{k,ij} = T_{k,ij} + U_{k,ij}$ ,  $U_{k,ij} \sim \text{i.i.d. Uniform}(-1, 1)$ . Under this setting, the marginal density of  $(X_{k,ij}, T_{k,ij})$  is the same for all  $(k, i, j)$  and  $E(X_{k,ij} | T_{k,ij}, T_{k,il}) = E(X_{k,ij} | T_{k,ij})$  for  $j \neq l$ . We assume that the variance of  $\varepsilon_{k,ij}$  depends on the conditional mean and is given as follows

$$\sigma^2(\mu) = \text{var}(Y|X, T) = 0.5\mu^2 + 0.3, \quad \text{where } \mu = E(Y|X, T).$$

We also assume an ARMA(1,1) correlation structure for the errors within the same subject, i.e.  $\text{corr}\{\varepsilon(s), \varepsilon(t)\} = 1$  for  $s = t$  and  $\gamma \exp(-|s - t|/\nu)$  for  $s \neq t$ , and set the correlation parameters at  $\gamma = 0.6$  and  $\nu = 1$ .

To examine the Wilks phenomenon of the GQLR test, we first assume that the null hypothesis is true, i.e.  $\theta_1(t) = \dots = \theta_4(t) = \theta_0(t)$ , and consider the following three scenarios:

$$\begin{aligned} \text{Scenario I: } & \beta = 1, \theta_0(t) = \cos(2\pi t); \\ \text{Scenario II: } & \beta = -1, \theta_0(t) = \cos(2\pi t); \\ \text{Scenario III: } & \beta = 1, \theta_0(t) = \sin(2\pi t). \end{aligned} \tag{21}$$

We simulate 200 data sets for each scenario and perform estimation and test under different working variance and correlation structures.

## 5.1 Wilks phenomenon when the variance function is consistently estimated.

For each simulated data set, we first fit the full model in a pilot analysis setting  $\mathcal{W}_{k,i}$ 's to be identity matrices, and then estimate the variance function by applying the local linear estimator described in Section 4.1 to the residuals. The inverse of the estimated variance function is used as the weight for fitting both the reduced and full models. To construct the GQLR test statistic, we use a Gaussian quasi-likelihood

$$\mathcal{Q}(\boldsymbol{\mu}, \mathbf{Y}) = -\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}(\boldsymbol{\mu})^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \tag{22}$$

where  $\mathbf{V}(\boldsymbol{\mu}) = \mathcal{S}(\boldsymbol{\mu})\mathcal{C}(\boldsymbol{\tau})\mathcal{S}(\boldsymbol{\mu})$ ,  $\mathcal{S}(\boldsymbol{\mu}) = \text{diag}\{\widehat{\sigma}(\mu_j)\}_{j=1}^m$  and  $\mathcal{C}(\boldsymbol{\tau})$  is a working correlation matrix. For the working correlation  $\mathcal{C}(\boldsymbol{\tau})$ , we consider both a mis-specified exchangeable working correlation structure (i.e.  $\text{corr}\{\varepsilon(s), \varepsilon(t)\} = \tau$  for some  $-m^{-1} < \tau < 1$  if  $s \neq t$ ) and working independence. When the exchangeable correlation structure is assumed, the correlation parameter is estimated by the QMLE of Fan et al. (2007). The bandwidth for the local linear estimators is chosen using the method described in Section 4.3 on a pilot data set and is then fixed for both the reduced and full models in all simulations.

The empirical distributions of  $\lambda_n(H_0)$  under different settings are shown in Figure 1, where the curves are estimated densities using the “density” function in  $\mathbf{R}$  based on 200 replicates of the test statistic. The left panel of Figure 1 shows the distributions of  $\lambda_n(H_0)$  when a mis-specified exchangeable correlation structure is used, and the right panel shows the same distributions when working independence is adopted. In each panel, the three estimated density functions correspond to the three scenarios in (21). The density functions in the same panel being almost identical confirms our theoretical results that, when the variance function is correctly specified, the asymptotic distribution of  $\lambda_n(H_0)$  does not depend on the values of  $\boldsymbol{\beta}$  and  $\theta_0(t)$ . By comparing the curves between the two panels, we find that the distribution of  $\lambda_n(H_0)$  depends on the working correlation structure  $\mathcal{C}(\boldsymbol{\tau})$ . These results corroborate our theoretical findings in Corollary 1.

To confirm these findings numerically, we perform Kolmogorov-Smirnov (KS) tests to test if  $\lambda_n(H_0)$  under different scenarios have the same distribution. When a mis-specified exchangeable correlation is assumed, the Kolmogorov-Smirnov  $p$ -values are 0.63 for Scenario I versus II, 0.18 for Scenario I versus III and 0.47 for Scenario II versus III. Similar results are obtained when working independence is used. We also test if  $\lambda_n(H_0)$  under different working correlation structures have the same distribution and the  $p$ -values are all in the magnitude of  $10^{-16}$ .

## 5.2 Wilks phenomenon fails to hold under variance mis-specification.

To better understand the behavior of the GQLR test when the variance is mis-specified, we consider the same simulation as above but mis-specify the variance as a constant. In other words, we assume  $\sigma^2(\mu) = \sigma^2$  and estimate  $\sigma^2$  using the mean squared error of the residuals in the full model. In Figure 2, we present the estimated densities of  $\lambda_n(H_0)$  under the mis-specified variance structure and working independence correlation. As we can see, there is a visible difference between the distributions of the test statistics under the three scenarios. This difference indicates the failure of the Wilks phenomenon.

Again, to confirm this point numerically, we perform KS tests to test if  $\lambda_n(H_0)$  under different scenarios have the same distribution. The KS  $p$ -values are  $1.5 \times 10^{-3}$  for Scenario I versus II,  $1.3 \times 10^{-4}$  for Scenario I versus III and 0.39 for Scenario II versus III.

## 5.3 Power of the GQLR test

We now study the power of the GQLR test. In particular, we are interested in the effect of working correlation on the power of the test. We focus on the simulation setting described in Scenario I and consider local alternatives with  $\theta_1(t) = \theta_0(t) - 2\phi G(t)$ ,  $\theta_2(t) = \theta_0(t) - \phi G(t)$ ,  $\theta_3(t) = \theta_0(t) + \phi G(t)$  and  $\theta_4(t) = \theta_0(t) + 2\phi G(t)$ . We consider two cases with  $G(t) = \sin(t)$  or  $G(t) = \sin(3\pi t)$  and set  $\phi$  at different values. Obviously, the null hypothesis is true when  $\phi = 0$  and as  $\phi$  increases the model deviates further away from  $H_0$ . The true within-subject correlation is ARMA(1,1) as described before.

For each value of  $\phi$ , we generate 200 datasets from the model. We use the local linear variance estimator as the working variance for both estimation and test, and we perform the GQLR test under two different working correlation structures: working independence and the true ARMA(1,1) correlation structure. The issue of estimating the critical value of the test using bootstrap will be addressed in Section 5.4. For the current study, we assume the critical values are consistently estimated and focus on the effect of correlation structure

on the power of the test. Specifically, we set the significance level at  $\alpha = 0.05$  and use another independent set of simulations to determine the 95% percentile of  $\lambda_n(H_0)$  under either working independence or the true correlation structure.

The power curves under the two settings of  $G(t)$  and the two correlation structures are depicted in the top two panels of Figure 3. We can also see from the graphs that the power of the GQLR test gets higher as  $\phi$  increases in all situations that we consider. Interestingly, the simulation results suggest that using the true correlation in the test does not always increase the power. From panel (a) of Figure 3 with  $G(t) = \sin(t)$ , the test based on working independence is more powerful than that using the true correlation, while the opposite is true when  $G(t) = \sin(3\pi t)$  as suggested by panel (b). It seems that the power of the test rather depends on a complicated interaction between  $G(t)$  and the true correlation, as suggested in Theorem 2. To validate this finding using our theoretical results, we also calculate the theoretical power of the test given in (20). The quantities  $\sigma_n^2$ ,  $\sigma_{1n}^2$ ,  $\mu_n$  and  $\mu_{1n}$  are estimated by replacing expectations with sample means. These theoretical power curves also confirm that the working independence test is more powerful when  $G(t) = \sin(t)$  and the opposite is true when  $G(t) = \sin(3\pi t)$ .

## 5.4 Estimating null distribution with wild bootstrap

Our study on the power of the test in the previous subsection suggests that, even if we consistently estimate the null distribution of the test statistic, using the true correlation structure may not increase the power of the GLQR test. In this subsection, we will focus on estimating the null distribution of  $\lambda_n(H_0)$  under working independence using the wild bootstrap procedure described in Section 4.2.

We use the data generated for Scenario I and apply the wild bootstrap procedure with a sample of 500 to each of the 200 data sets. We compare the bootstrap distributions with the true distribution of the test statistic in the left panel of Figure 4. The solid curve in the graph

is the true distribution of  $\lambda_n(H_0)$ , the dashed curve is the mean of the bootstrap densities from the 200 replicates, and the two dotted curves are the pointwise 1% and 99% percentiles of the bootstrap densities. For comparison, we also plot the asymptotic normal distribution in Corollary 2 as the dash-dot curve in the same plot. As we can see, for a moderate sample size of 50 subjects per group, there is still a large discrepancy between true distribution of  $\lambda_n(H_0)$  and the asymptotic normal distribution, which indicates a slow convergence to the limiting distribution for this nonparametric test. This result is in agreement with the simulation studies in Härdle and Mammen (1993). The bootstrap distribution does a much better job on approximating the truth, even though it still tends to underestimate the variation of  $\lambda_n(H_0)$ .

To use bootstrap in tests, it is more relevant to check the performance of the bootstrap quantiles. In the right panel of Figure 4, the dark dots indicate the true quantiles  $Q_{1-\alpha}$  of  $\lambda_n(H_0)$ , and the vertical bars are the percentile bands of the bootstrap estimator  $\widehat{Q}_{1-\alpha}$ . For each  $1 - \alpha$ , the lower end of the vertical bar is the 2.5% percentile of the bootstrap quantile  $\widehat{Q}_{1-\alpha}$ , and the upper end of the bar is the 97.5% percentile of  $\widehat{Q}_{1-\alpha}$ . As we can see, the true quantiles are covered by these bands.

## 6 Analysis of the opioid dependence treatment data

We now illustrate the application of our proposed GQLR test to a data set on opioid agonist treatment. The data were collected from 140 patients who received primary care-based buprenorphine, a commonly prescribed medication for treating opioid dependence, at the Primary Care Center of Yale-New Haven Hospital. Each patient first went through a two-week induction and stabilization period and then was prescribed with daily medication of buprenorphine for 24 weeks. Prior research has shown that adding counseling to buprenorphine can help increase opioid abstinence rate (Amato et al., 2011). The main objective of the study was to investigate the impact of adding cognitive behavioral therapy, which is a

counseling intervention for treating a variety of psychiatric conditions and substance use disorders (Crits-Christoph et al., 1999; Beck, 2005; Butler et al., 2006; McHugh et al., 2010), on the efficacy of primary case-based buprenorphine to treat opioid dependence. The patients were randomly assigned to receive one of two treatments: physician management (PM) or physician management and cognitive behavioral therapy (PMCBT). Physician management was provided in the form of 15- to 20-minute sessions by internal medicine physicians who had experience with providing buprenorphine but had no training in cognitive behavioral therapy. These sessions were given weekly for the first two weeks, every two weeks for the next four weeks, and then monthly afterward. Patients in the PMCBT group were offered the additional opportunity to participate in up to twelve fifty-minute weekly cognitive behavioral therapy sessions during the first twelve weeks of treatment. All counseling sessions were given by well-trained masters and doctoral-level clinicians. The main components of counseling focused on developing behavioral skills such as promoting behavioral activation and identifying and coping with opioid craving.

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



Among the 140 patients, 69 of them were in the ‘PM’ group and 71 in the ‘PMCBT’ group. We now analyze the data using the proposed model (1), where the link function  $g(\cdot)$  is set to be a logistic link. The bandwidth in the estimation is selected using the method described in Section 4.3. The estimated regression coefficients for the covariates are reported in Table 1, where the standard errors of these estimates are calculated using a sandwich formula similar to that on page 1056 of Lin and Carroll (2001). Among the three covariates, education level is significantly related to the response variable. The estimated mean curves for the two groups are presented in Figure 5. At the beginning of the treatment, the mean curves of the two groups were almost the same, while after about 70 days, patients in the ‘PMCBT’ group had a lower probability of opioid use. Near the end of the treatment, both groups had increased opioid use rate, but patients with additional cognitive behavioral therapy seemed to have a lower overall rate of use.

Our goal is to test whether the mean curves of the two groups are significantly different after taking into account the covariate effects. There has been some literature on testing the difference between two mean curves for independent Gaussian-type data, such as the methods proposed by Hall and Hart (1990). However, these classic methods are not applicable to this semiparametric setting with a nonlinear link function. Instead, we apply our proposed GQLR test to the data using a working independence binary quasi-likelihood

$$\mathcal{Q}_{\text{binary}}(\boldsymbol{\mu}, \mathbf{Y}) = \sum_{j=1}^m Y_j \log\{\mu_j/(1 - \mu_j)\} + \log(1 - \mu_j). \quad (23)$$

The p-value of the test is 0.044 based on 1000 bootstrap replicates. We therefore conclude that there is a significant difference between the two treatment groups. In particular, Figure 5 suggests that cognitive behavioral therapy improved the opioid abstinence over time. Fiellin et al. (2013) analyzed the same data, using self-reported frequency of opioid use and the maximum number of consecutive weeks of abstinence from illicit opioids in the two 12-week periods as the primary outcome measures, but did not find any evidence supporting the benefit of adding cognitive behavioral therapy. Their outcome measures were aggregated

and captured only certain features of the data. In contrast, our analysis is based on the entire longitudinal trajectory and hence is more powerful in detecting the difference between the two treatments.

## 7 Summary

We investigate a class of semiparametric analysis of covariance models for generalized longitudinal data, where the treatment effects are represented as nonparametric functions over time and covariates are incorporated to account for the variability caused by confounders. We propose to test the treatment effects using a generalized quasi-likelihood ratio test. Our theoretical study reveals that when the variance structure is correctly specified, the asymptotic distribution of the GQLR test statistic does not depend on the mean parameters but still depends on the true and working correlation structures. When the variance is mis-specified, the Wilks property might fail completely in the sense that the distribution of the test statistic depends on all nuisance parameters. In these situations, the bootstrap distribution of the test statistic might be inconsistent, which in turn results in invalid inferences.

However, if working independence is assumed in both estimation and test and if the variance structure is correctly specified, the Wilks phenomenon known to hold for independent data also holds for longitudinal data. Replacing the true variance with a consistent estimator, such as a nonparametric local linear estimator, only causes an asymptotically negligible error to the test. In this case, we propose a working independence wild bootstrap procedure that can consistently estimate the null distribution. We have also shown that the GQLR test yields the minimax optimal power rate.

Our test procedure is based on estimators similar to those in Lin and Carroll (2001). More sophisticated kernel estimators were proposed in Wang et al. (2005) and Lin and Carroll (2006). It is unclear how much power one can gain by building tests based on these more sophisticated estimators since our proposed test already attains the minimax optimal

power rate. Tests based on other smoothers, such as penalized splines (Crainiceanu, et al., 2006), or other principles, such as empirical likelihood (Chen and Van Keilegom, 2009), can potentially be extended to our problem. These possibility are open questions that call for future investigations.

## Acknowledgment

This paper is based on the first author's Ph.D. dissertation under Li's supervision. Li's research was supported by the National Science Foundation awards DMS-1105634 and DMS-1317118. Guan's research was supported by the National Institute of Health grant 7R01DA029081 and the National Science Foundation grant DMS-0845368. We thank the Associate Editor and an anonymous referee for their helpful and constructive comments, which lead to significant improvement in this paper.

## References

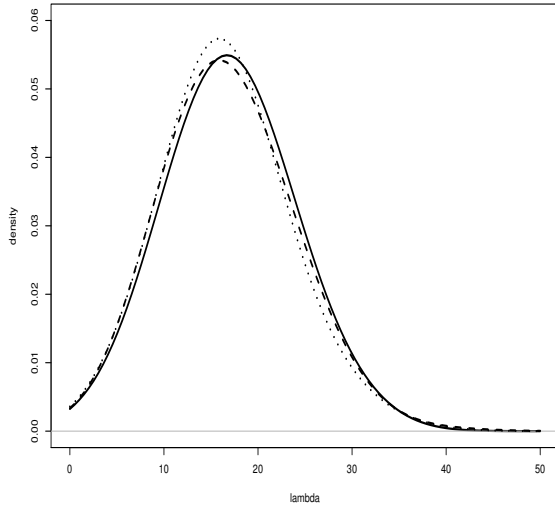
- Amato L., Minozzi S., Davoli M., and Vecchi S. (2011). Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence, *Cochrane Database Syst Rev.*, **10**: CD004147.
- Bahadur, R. R. (1961). A representation of the joint distribution of responses to  $n$  dichotomous items, *Studies in Item Analysis and Prediction*, Ed. H. Solomon, Stanford University Press, Stanford, California, 1961, pp.158-168.
- Beck, A. T. (2005). The current state of cognitive therapy: a 40-year retrospective, *Arch Gen Psychiatry*, **62**, 953-959.
- Brumback, B., and Rice, J. A. (1998), Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves (with discussion), *Journal of the American Statistical Association*, **93**, 961-994.
- Butler, A. C., Chapman, J. E., Forman, E. M., and Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses, *Clin Psychol Rev.*, **26**, 17-31.
- Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415-447.
- Crainiceanu, C. M., Ruppert, D, Claeskens, G. and Wand M.P. (2005). Exact likelihood ratio tests for penalized splines. *Biometrika*, **92**, 91-103.
- Crits-Christoph, P., Siqueland, L., Blaine, J., et al. (1999). Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study, *Arch Gen Psychiatry*, **56**, 493-502

- de Jong, P. (1987), A Central Limit Theorem for Generalized Quadratic Forms, *Probability Theory and Its Related Fields*, 75, 261 -277.
- Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last, *Journal of Econometrics*, 146, 162169.
- Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function, *Journal of the American Statistical Association*, 102, 632-641.
- Fan, J. and Jiang, J.,(2005). Nonparametric inferences for additive models, *Journal of the American Statistical Association*, **100**, 890-907.
- Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests (with discussion), *Test*, 16, 409-478.
- Fan, J., and Li, R., (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99, 710-723.
- Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data. *Journal of the American Statistical Association*, 103, 1520 - 1533.
- Fan, J., Zhang, C. and Zhang, J., (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics*, 29, 153 - 193.
- Fiellin, D. A., Barry, D. T., Sullivan, L. E., Cutter, C. J., Moore, B. A., O'Connor, P. G., MPH, Schottenfeld, R. S. (2013). A Randomized Trial of Cognitive Behavioral Therapy in Primary Care-based Buprenorphine, *The American Journal of Medicine*, **126**, 74, 11-17.
- González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models, *Test*, 22, 361-411.
- Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression, *Journal of the American Statistical Association*, 85, 1039-1049.
- Hall, P., Müller, H. -G. and Wang, J. -L. (2006). Properties of principal component methods for functional and longitudinal data analysis, *Annals of Statistics*, 34, 1493-1517.
- Hall, P., Müller, H. G. & Yao, F. (2008). Modeling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society, Series B*, **70**, 703-23.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, 21, 1926-1947.
- Ingster, Y. I. (1993), Asymptotic Minimax Hypothesis Testing for Nonparametric Alternatives III, *Mathematical Methods in Statistics*, 2, 85 - 114; 3, 171 - 189; 4, 249 - 268.

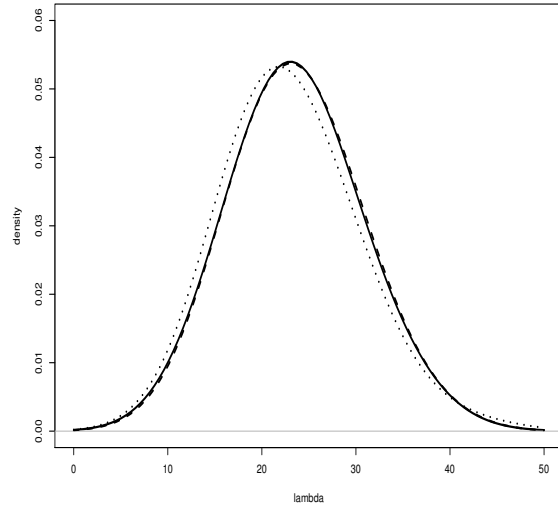
- Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96, 1045-1056.
- Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society, Series B*, 68, 69-88.
- Li, R., and Liang, H. (2008), Variable selection in semiparametric regression modeling, *The 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. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, *Annals of Statistics*, 38, 3321-3351.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics*, 21, 1, 255-285.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models (2nd ed.)*, New York: Chapman & Hall.
- McHugh, P. K., Hearon, B. A., Otto, M. W. (2010). Cognitive behavioral therapy for substance use disorders, *Psychiatr Clin North Am.*, **33**, 511-525.
- Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B*, 68, 179-199.
- Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *Journal of the Royal Statistical Society*, 53, 233-243.
- Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data, *Journal of the American Statistical Association*, **100**, 469, 147-57.
- Xia, Y. and Li, W. K. (2002) Asymptotic behavior of bandwidth selected by cross-validation method under dependence. *Journal Multivariate Analysis*, 83, 265-287.
- Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577-590.
- Zhou, L., Huang, J., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association*, 105, 390-400.

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

Table 1: Estimation results of the regression coefficients in the opioid dependence treatment data.



(a) Exchangeable working correlation.



(b) Working independence.

Figure 1: Simulation study: the empirical distribution of  $\lambda_n(H_0)$  when the variance function is consistently estimated using a local linear estimator. Panel (a): an exchangeable correlation structure is assumed for the test, where the correlation parameter is estimated by the method described in Section 4.1. Panel (b): working independence is assumed for both estimation and test. In both panels, the solid, dashed and dotted curves correspond to Scenarios I to III respectively.

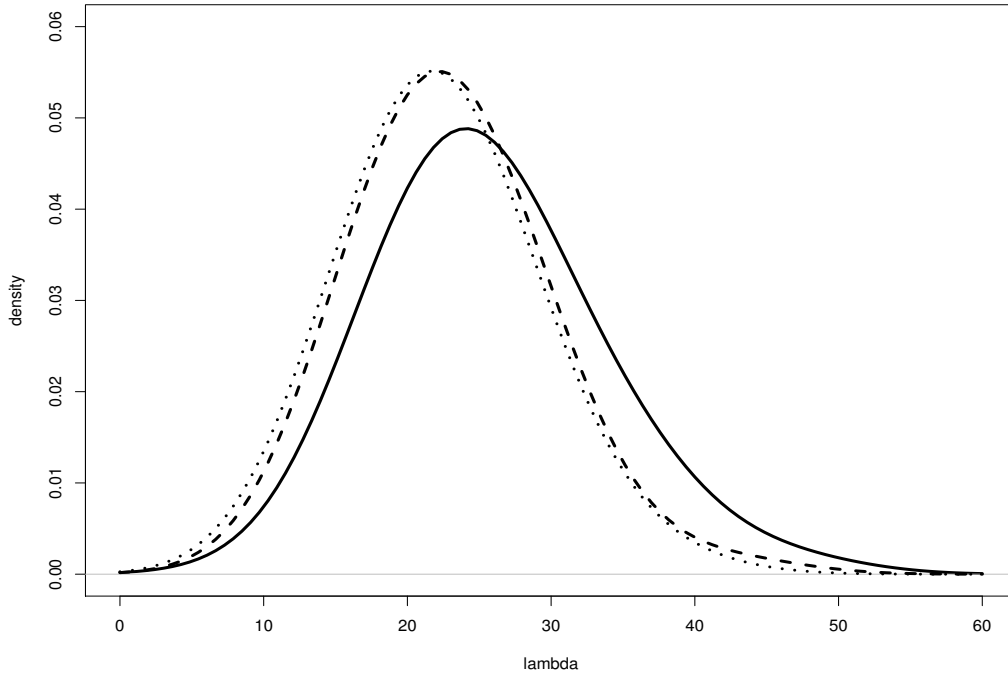


Figure 2: Simulation study: the empirical distributions of  $\lambda_n(H_0)$  when the variance function is misspecified as a constant and the correlation is working independence. The solid line is the density of  $\lambda_n(H_0)$  under Scenario I, the dashed line is for Scenario II and the dotted line is for Scenario III.



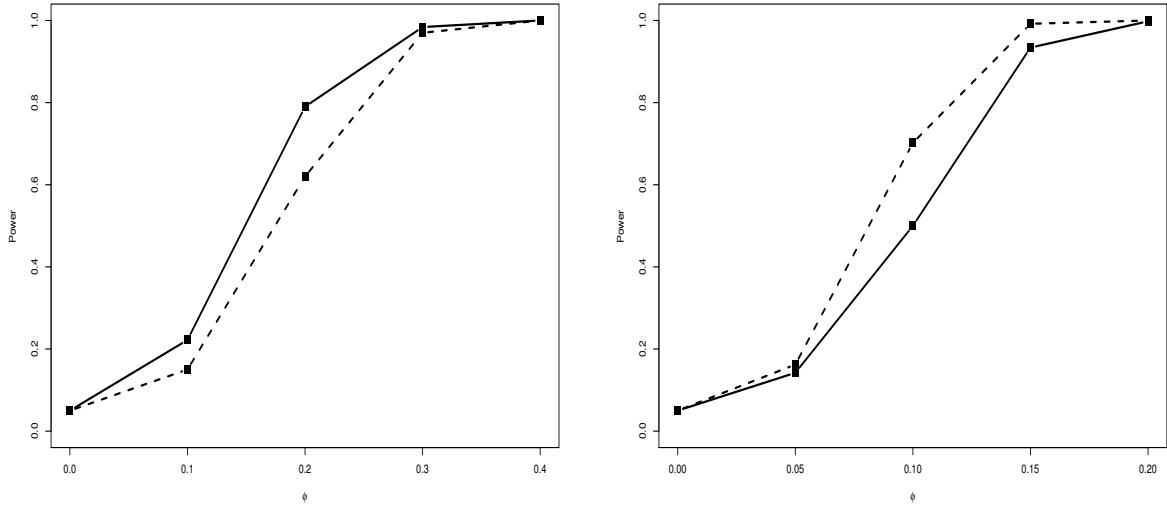


Figure 3: Simulation study: the power of the GQLR test under the local alternative  $\theta_1(t) = \theta_0(t) - 2\phi G(t)$ ,  $\theta_2(t) = \theta_0(t) - \phi G(t)$ ,  $\theta_3(t) = \theta_0(t) + \phi G(t)$  and  $\theta_4(t) = \theta_0(t) + 2\phi G(t)$ . The left panel contains the power curves for  $G(t) = \sin(t)$  and the right panel contains the power curves for  $G = \sin(3\pi t)$ . In each panel, the solid curve is the power under working independence and the dotted curve is the power when the true correlation is used.

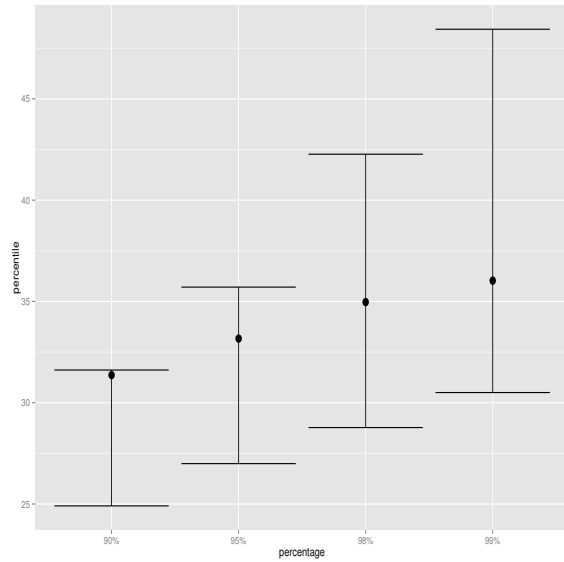
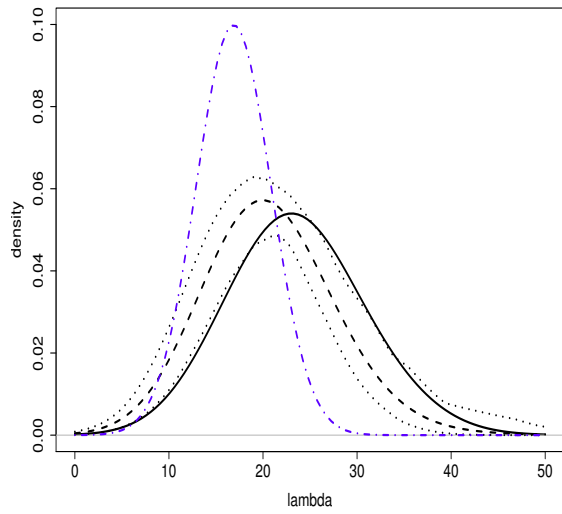


Figure 4: Simulation study: performance of wild bootstrap under Scenario I. Left panel: the solid curve is the true density of  $\lambda_n(H_0)$  under working independence; the dashed curve is the mean of the bootstrap densities; the two dotted curves are the pointwise 1% and 99% percentiles of the bootstrap densities; the dash-dot curve is the theoretical normal distribution. Right panel: the dark dots are the true quantiles of  $\lambda_n(H_0)$ , and the vertical bars are the percentile bands of bootstrap estimator of the quantiles.

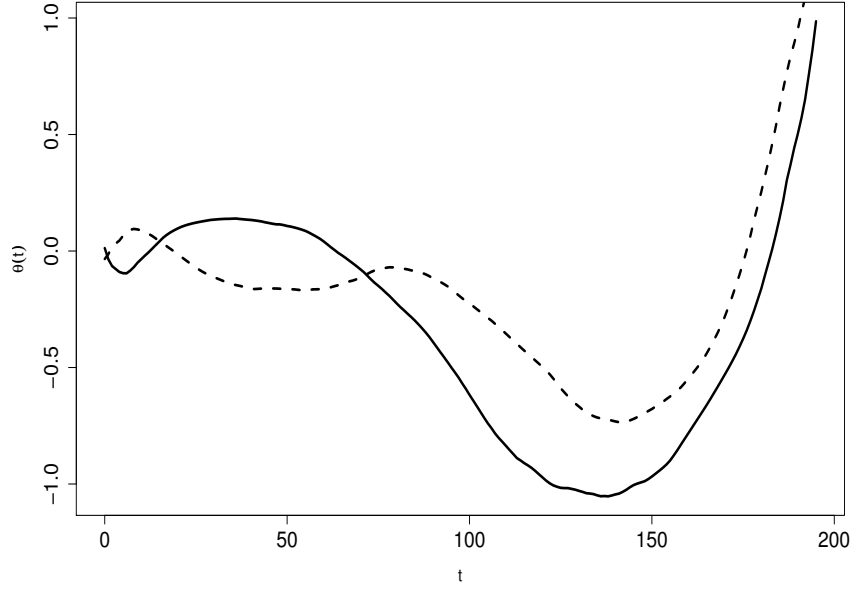


Figure 5: The estimated time effect  $\theta_k(t)$ : PM+CBT treatment (solid line); PM treatment: (dotted line)

# Online Supplement For

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

## Web Appendix: Technical Proofs and Additional Simulation Results

### W.1 Proof of Theorem 1

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

$$\begin{aligned}\frac{\partial \mathcal{Q}}{\partial \mathbf{x}}\{g^{-1}(\mathbf{x}), \mathbf{y}\} &= \Delta(\mathbf{x})\mathbf{V}^{-1}\{g^{-1}(\mathbf{x})\}\{\mathbf{y} - g^{-1}(\mathbf{x})\}, \\ \frac{\partial^2 \mathcal{Q}}{\partial \mathbf{x} \partial \mathbf{x}^\top}\{g^{-1}(\mathbf{x}), \mathbf{y}\} &= -\Delta(\mathbf{x})\mathbf{V}^{-1}\{g^{-1}(\mathbf{x})\}\Delta(\mathbf{x}) + \sum_{j=1}^m \{y_j - g^{-1}(x_j)\}\mathcal{D}_j(\mathbf{x}),\end{aligned}$$

where  $\Delta(\mathbf{x}) = \text{diag}\{\frac{dg^{-1}}{dx}(x_j)\}_{j=1}^m$ ,  $\mathcal{D}_j = \partial(\mathbf{V}^{j\bullet}\Delta)/\partial \mathbf{x}$ , and  $\mathbf{V}^{j\bullet}$  is the  $j$ -th row of  $\mathbf{V}^{-1}$ . Denote  $\boldsymbol{\eta}_{0k,i} = \mathbf{X}_{k,i}\boldsymbol{\beta}_0 + \theta_0(\mathbf{T}_{k,i})$ ,  $\boldsymbol{\mu}_{0k,i} = g^{-1}(\boldsymbol{\eta}_{0k,i})$  and  $\boldsymbol{\varepsilon}_{k,i} = \mathbf{Y}_{k,i} - \boldsymbol{\mu}_{0k,i}$ . By taking a Taylor's expansion at  $\boldsymbol{\eta}_{0k,i}$ , we have

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

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

$$\lambda_n(H_0) = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + o_p(1), \quad (\text{W.1})$$

where

$$J_1 = \sum_k \boldsymbol{\varepsilon}_k^\top \mathbf{V}_k^{-1} \Delta_k \{\widehat{\theta}_{F,k}(\mathbf{T}_k) - \widehat{\theta}_R(\mathbf{T}_k)\}, \quad J_2 = \sum_k \boldsymbol{\varepsilon}_k^\top \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k (\widehat{\boldsymbol{\beta}}_F - \widehat{\boldsymbol{\beta}}_R),$$

$$\begin{aligned}
J_3 &= \sum_k (\widehat{\beta}_R - \beta_0)^\top \mathbf{X}_k^\top \Delta_k \mathbf{V}_K^{-1} \Delta_k \{\widehat{\theta}_R(\mathbf{T}_k) - \theta_0(\mathbf{T}_k)\} \\
&\quad - (\widehat{\beta}_F - \beta_0)^\top \mathbf{X}_k^\top \Delta_k \mathbf{V}_K^{-1} \Delta_k \{\widehat{\theta}_{F,k}(\mathbf{T}_k) - \theta_0(\mathbf{T}_k)\}, \\
J_4 &= \frac{1}{2} \sum_k \|\widehat{\theta}_R(\mathbf{T}_k) - \theta_0(\mathbf{T}_k)\|_{\Delta_k \mathbf{V}_K^{-1} \Delta_k}^2 - \|\widehat{\theta}_{F,k}(\mathbf{T}_k) - \theta_0(\mathbf{T}_k)\|_{\Delta_k \mathbf{V}_K^{-1} \Delta_k}^2, \\
J_5 &= \frac{1}{2} \sum_k \|\widehat{\beta}_R - \beta_0\|_{\mathbf{X}_k^\top \Delta_k \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k}^2 - \|\widehat{\beta}_F - \beta_0\|_{\mathbf{X}_k^\top \Delta_k \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k}^2, \\
J_6 &= \frac{1}{2} \sum_k \sum_{i=1}^{n_k} \|\mathbf{X}_{k,i}(\widehat{\beta}_F - \beta_0) + \widehat{\theta}_{F,k}(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2 \\
&\quad - \|\mathbf{X}_{k,i}(\widehat{\beta}_R - \beta_0) + \widehat{\theta}_R(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2.
\end{aligned}$$

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

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

**Proof:** (i) By the asymptotic expansions in Propositions 1 and 2,  $(\widehat{\beta}_F - \widehat{\beta}_R) = o_p(n^{-1/2})$ . Therefore  $J_2 = (\sum_k \varepsilon_k^\top \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k)(\widehat{\beta}_F - \widehat{\beta}_R) = O_p(n^{1/2}) \times o_p(n^{-1/2}) = o_p(1)$ . Similarly,  $J_5 = \sum_k (\widehat{\beta}_R - \beta_0) \mathbf{X}_k^\top \Delta_k \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k (\widehat{\beta}_R - \widehat{\beta}_F) + (\widehat{\beta}_R - \widehat{\beta}_F) \mathbf{X}_k^\top \Delta_k \mathbf{V}_k^{-1} \Delta_k \mathbf{X}_k (\widehat{\beta}_F - \beta_0) = o_p(1)$ .

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

$$(\widehat{\beta}_R - \beta_0)^\top \sum_{k=1}^q \sum_{i=1}^{n_k} \mathbf{X}_{k,i}^\top \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} \{\widehat{\theta}_R(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\} = O_p(1 + n^{1/2}h^2)$$

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

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

$$\begin{aligned}
J_{61} &= \frac{1}{2} \sum_k \sum_{i=1}^{n_k} \|\mathbf{X}_{k,i}(\widehat{\beta}_F - \beta_0)\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2 - \|\mathbf{X}_{k,i}(\widehat{\beta}_R - \beta_0)\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2 \\
J_{62} &= \frac{1}{2} \sum_k \sum_{i=1}^{n_k} \|\widehat{\theta}_{F,k}(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2 - \|\widehat{\theta}_R(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\|_{\sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij}}^2 \\
J_{63} &= \sum_k \sum_{i=1}^{n_k} (\widehat{\beta}_F - \beta_0)^\top \mathbf{X}_{k,i}^\top \left( \sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij} \right) \{\widehat{\theta}_{F,k}(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\} \\
&\quad - (\widehat{\beta}_R - \beta_0)^\top \mathbf{X}_{k,i}^\top \left( \sum_{j=1}^m \varepsilon_{k,ij} \mathcal{D}_{k,ij} \right) \{\widehat{\theta}_R(\mathbf{T}_{k,i}) - \theta_0(\mathbf{T}_{k,i})\}
\end{aligned}$$

It can easily show that  $J_{61} = O_p(n^{-1/2})$ . By Proposition 1,  $\widehat{\theta}_R(t) - \theta_0(t) = O_p(h^2 + n^{-1/2}h^{-1/2})$ . Under the null hypothesis, Proposition 2 provides the same convergence rate for  $\widehat{\theta}_{F,k}(t) - \theta_0(t)$ . Since  $\text{corr}(\varepsilon_{k,ij}, \varepsilon_{k',i'j'}) \neq 0$  only if  $k = k'$  and  $i = i'$ , by tedious moment calculation  $EJ_{62}^2 = O\{n \times (h^2 + n^{-1/2}h^{-1/2})^4\}$ . 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.6).  $\diamond$

**LEMMA W.2** *Suppose all assumptions in Theorem 1 hold, then*

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

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

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

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

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

$$\begin{aligned} R_1 &= \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j,j',\ell=1}^m \left(\frac{1}{n_k} - \frac{1}{n}\right) \varepsilon_{k,ij} \varepsilon_{k,i'j'} \nu_{k,i}^{j\ell} \mu_{k,i\ell}^{(1)} \frac{\mu_{k,ij'}^{(1)} \omega_{k,i}^{j'j'}}{mB_1(T_{k,i\ell})} K_h(T_{k,ij'} - T_{k,i\ell}), \\ R_2 &= \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{i' \neq i} \sum_{j,j',\ell=1}^m \left(\frac{1}{n_k} - \frac{1}{n}\right) \varepsilon_{k,ij} \varepsilon_{k,i'j'} \nu_{k,i}^{j\ell} \mu_{k,i\ell}^{(1)} \frac{\mu_{k,i'j'}^{(1)} \omega_{k,i'}^{j'j'}}{mB_1(T_{k,i\ell})} K_h(T_{k,i'j'} - T_{k,i\ell}), \\ R_3 &= - \sum_k \sum_{k' \neq k} \sum_{i=1}^{n_k} \sum_{i'=1}^{n_{k'}} \sum_{j,j',\ell=1}^m \frac{1}{nm} \varepsilon_{k,ij} \varepsilon_{k',i'j'} \nu_{k,i}^{j\ell} \mu_{k,i\ell}^{(1)} \frac{\mu_{k',i'j'}^{(1)} \omega_{k',i'}^{j'j'}}{B_1(T_{k,i\ell})} K_h(T_{k',i'j'} - T_{k,i\ell}). \end{aligned}$$

By straightforward calculation,

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

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

By similar calculations,

$$\begin{aligned}
J_4 &= \frac{1}{2} \sum_k \sum_i \left\{ (\widehat{\theta}_R - \theta_0)^\top (\mathbf{T}_{k,i}) \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} (\widehat{\theta}_R - \widehat{\theta}_{F,k}) (\mathbf{T}_{k,i}) \right. \\
&\quad \left. + (\widehat{\theta}_R - \widehat{\theta}_{F,k})^\top (\mathbf{T}_{k,i}) \Delta_{k,i} \mathbf{V}_{k,i}^{-1} \Delta_{k,i} (\widehat{\theta}_{F,k} - \theta_0) (\mathbf{T}_{k,i}) \right\} \\
&= \frac{1}{2} \sum_k \sum_i \sum_{\ell_1, \ell_2=1}^m \mu_{k, \ell_1}^{(1)} \nu_{k,i}^{\ell_1 \ell_2} \mu_{k, \ell_2}^{(1)} \left\{ \frac{\theta_0^{(2)}(T_{k, \ell_1}) h^2}{2} - \boldsymbol{\mu}_X^\top(T_{k, \ell_1}) (\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \right. \\
&\quad \left. + \frac{1}{nm B_1(T_{k, \ell_1})} \sum_{k_1=1}^q \sum_{i_1=1}^{n_{k_1}} \sum_{j_1=1}^m \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} K_h(T_{k_1, i_1 j_1} - T_{k, \ell_1}) \varepsilon_{k_1, i_1 j_1} \right\} \\
&\quad \times \left\{ \frac{1}{nm B_1(T_{k, \ell_2})} \sum_{k_2=1}^q \sum_{i_2=1}^{n_{k_2}} \sum_{j_2=1}^m \mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2} K_h(T_{k_2, i_2 j_2} - T_{k, \ell_2}) \varepsilon_{k_2, i_2 j_2} \right. \\
&\quad \left. - \frac{1}{n_k m B_1(T_{k, \ell_2})} \sum_{i_2=1}^{n_k} \sum_{j_2=1}^m \mu_{k, i_2 j_2}^{(1)} \omega_{k, i_2}^{j_2 j_2} K_h(T_{k, i_2 j_2} - T_{k, \ell_2}) \varepsilon_{k, i_2 j_2} \right\} \\
&+ \frac{1}{2} \sum_k \sum_i \sum_{\ell_1, \ell_2=1}^m \mu_{k, \ell_1}^{(1)} \nu_{k,i}^{\ell_1 \ell_2} \mu_{k, \ell_2}^{(1)} \left\{ \frac{\theta_0^{(2)}(T_{k, \ell_1}) h^2}{2} - \boldsymbol{\mu}_X^\top(T_{k, \ell_1}) (\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0) \right. \\
&\quad \left. + \frac{1}{n_k m B_1(T_{k, \ell_1})} \sum_{i_1=1}^{n_k} \sum_{j_1=1}^m \mu_{k, i_1 j_1}^{(1)} \omega_{k, i_1}^{j_1 j_1} K_h(T_{k, i_1 j_1} - T_{k, \ell_1}) \varepsilon_{k, i_1 j_1} \right\} \\
&\quad \times \left\{ \frac{1}{nm B_1(T_{k, \ell_2})} \sum_{k_2=1}^q \sum_{i_2=1}^{n_{k_2}} \sum_{j_2=1}^m \mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2} K_h(T_{k_2, i_2 j_2} - T_{k, \ell_2}) \varepsilon_{k_2, i_2 j_2} \right. \\
&\quad \left. - \frac{1}{n_k m B_1(T_{k, \ell_2})} \sum_{i_2=1}^{n_k} \sum_{j_2=1}^m \mu_{k, i_2 j_2}^{(1)} \omega_{k, i_2}^{j_2 j_2} K_h(T_{k, i_2 j_2} - T_{k, \ell_2}) \varepsilon_{k, i_2 j_2} \right\} + o_p(h^{-1/2}).
\end{aligned}$$

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

$$\begin{aligned}
R_4 &= \frac{1}{2} \sum_k \sum_i \sum_{\ell_1, \ell_2=1}^m \mu_{k, \ell_1}^{(1)} \nu_{k,i}^{\ell_1 \ell_2} \mu_{k, \ell_2}^{(1)} \left[ \sum_{k_1=1}^q \left\{ \frac{1}{n^2} I(k \neq k_1) + \left( \frac{1}{n^2} - \frac{1}{n_k^2} \right) I(k_1 = k) \right\} \right. \\
&\quad \times \sum_{i_1=1}^{n_{k_1}} \sum_{j_1=1}^m \sum_{j_2=1}^m \frac{\mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_1, i_1 j_2}^{(1)} \omega_{k_1, i_1}^{j_2 j_2}}{m^2 B_1(T_{k, \ell_1}) B_1(T_{k, \ell_2})} \\
&\quad \left. \times K_h(T_{k_1, i_1 j_1} - T_{k, \ell_1}) K_h(T_{k_1, i_1 j_2} - T_{k, \ell_2}) \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_1, i_1 j_2} \right] \\
&= \frac{1}{2} \sum_{k_1=1}^q \left\{ \sum_{k \neq k_1} \rho_{k_1} \rho_k + (\rho_{k_1}^2 - 1) \right\} \frac{1}{n_{k_1} m} \sum_{i_1=1}^{n_{k_1}} \\
&\quad \left\{ \sum_{j_1=1}^m \sum_{j_2 \neq j_1} \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_1, i_1 j_2}^{(1)} \omega_{k_1, i_1}^{j_2 j_2} \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_1, i_1 j_2} \frac{B_2(T_{k_1, i_1 j_2})}{B_1^2(T_{k_1, i_1 j_2})} K_h * K_h(T_{k_1, i_1 j_1} - T_{k_1, i_1 j_2}) \right\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{j_1=1}^m (\mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1})^2 \varepsilon_{k_1, i_1 j_1}^2 \int K_h^2(T_{k_1, i_1 j_1} - u) \frac{B_2(u)}{B_1^2(u)} du \Big\} + O_p(1) \\
& = \frac{1-q}{2h} \mathbb{E} \left\{ m^{-1} \sum_{j=1}^m \Delta_{jj}^2 (\omega^{jj})^2 \sigma_{jj} B_2(T_j) B_1^{-2}(T_j) \nu_K \right\} + O_p(1) \\
& = \frac{1-q}{2h} \nu_K \mathbb{E} \{ B_3(T) B_2(T) B_1^{-2}(T) f^{-1}(T) \} + O_p(1), \\
R_5 & = \frac{1}{2} \sum_k \sum_i \sum_{\ell_1, \ell_2=1}^m \frac{\mu_{k, i \ell_1}^{(1)} \nu_{k, i}^{\ell_1 \ell_2} \mu_{k, i \ell_2}^{(1)}}{B_1(T_{k, i \ell_1}) B_1(T_{k, i \ell_2})} \left[ \sum_{k_1=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\} \right. \\
& \quad \times \frac{1}{m^2} \sum_{i_1=1}^{n_{k_1}} \sum_{i_2 \neq i_1}^{n_{k_1}} \sum_{j_1, j_2=1}^m \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_1, i_2 j_2}^{(1)} \omega_{k_1, i_2}^{j_2 j_2} \\
& \quad \left. \times K_h(T_{k_1, i_1 j_1} - T_{k, i \ell_1}) K_h(T_{k_1, i_2 j_2} - T_{k, i \ell_2}) \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_1, i_2 j_2} \right] \\
& = \frac{1}{2} \sum_{k_1} (\rho_{k_1} - 1) \frac{1}{n_{k_1} m} \sum_{i_1=1}^{n_{k_1}} \sum_{i_2 \neq i_1}^{n_{k_1}} \sum_{j_1=1}^m \sum_{j_2=1}^m \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_1, i_2 j_2}^{(1)} \omega_{k_1, i_2}^{j_2 j_2} \frac{B_2(T_{k_1, i_2 j_2})}{B_1^2(T_{k_1, i_2 j_2})} \\
& \quad \times K_h * K_h(T_{k_1, i_1 j_1} - T_{k_1, i_2 j_2}) \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_1, i_2 j_2} + O_p(1), \\
R_6 & = \frac{1}{2} \sum_k \sum_i \sum_{\ell_1, \ell_2=1}^m \frac{\mu_{k, i \ell_1}^{(1)} \nu_{k, i}^{\ell_1 \ell_2} \mu_{k, i \ell_2}^{(1)}}{B_1(T_{k, i \ell_1}) B_1(T_{k, i \ell_2})} \left[ \sum_{k_1=1}^q \sum_{k_2 \neq k_1} \left\{ \frac{1}{n^2} + \frac{1}{n n_{k_1}} I(k_1 = k) - \frac{1}{n n_{k_2}} I(k_2 = k) \right\} \right. \\
& \quad \times \frac{1}{m^2} \sum_{i_1=1}^{n_{k_1}} \sum_{i_2=1}^{n_{k_2}} \sum_{j_1, j_2=1}^m \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2} \\
& \quad \left. \times K_h(T_{k_1, i_1 j_1} - T_{k, i \ell_1}) K_h(T_{k_2, i_2 j_2} - T_{k, i \ell_2}) \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_2, i_2 j_2} \right] \\
& = \frac{1}{2 n m} \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, j_2=1}^m \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2} \frac{B_2(T_{k_2, i_2 j_2})}{B_1^2(T_{k_2, i_2 j_2})} \\
& \quad \times K_h * K_h(T_{k_1, i_1 j_1} - T_{k_2, i_2 j_2}) \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_2, i_2 j_2} + O_p(1),
\end{aligned}$$

where  $B_1(t)$  is as defined in Condition (C.4) and  $B_2(t)$  and  $B_3$  are defined in (16).

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

We now collect the mean components in  $J_1 + J_4$  as

$$\begin{aligned}
\mu_n = R_1 + R_4 & = \frac{q-1}{mh} \mathbb{E} \left\{ K(0) \sum_j \sum_{j'} \sigma_{jj'} \nu^{jj'} \Delta_{j'j'}^2 \omega^{j'j'} B_1^{-1}(T_{j'}) \right. \\
& \quad \left. - \frac{\nu_K}{2} \sum_{j=1}^m \Delta_{jj}^2 (\omega^{jj})^2 \sigma_{jj} B_2(T_j) B_1^{-2}(T_j) \right\} + O_p(1) \\
& = \frac{q-1}{h} \mathbb{E} \left\{ K(0) \frac{B_5(T)}{B_1(T) f(T)} - \frac{\nu_K}{2} \frac{B_2(T) B_3(T)}{B_1^2(T) f(T)} \right\} + O_p(1). \quad (\text{W.2})
\end{aligned}$$



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

$$\begin{aligned}
W_n &= \sum_{k_1=1}^q (1 - \rho_{k_1}) \frac{1}{n_{k_1} m} \sum_{i_1=1}^{n_{k_1}} \sum_{i_2 \neq i_1}^m \sum_{j_1=1}^m \sum_{j_2=1}^m \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_1, i_2 j_2} \\
&\quad \times \left\{ \sum_{\ell=1}^m \nu_{k_1, i_1}^{j_1 \ell} \mu_{k_1, i_1 \ell}^{(1)} \frac{\mu_{k_1, i_2 j_2}^{(1)} \omega_{k_1, i_2}^{j_2 j_2}}{B_1(T_{k_1, i_1 \ell})} K_h(T_{k_1, i_2 j_2} - T_{k_1, i_1 \ell}) \right. \\
&\quad \left. - \frac{1}{2} \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_1, i_2 j_2}^{(1)} \omega_{k_1, i_2}^{j_2 j_2} \frac{B_2(T_{k_1, i_2 j_2})}{B_1^2(T_{k_1, i_2 j_2})} K_h * K_h(T_{k_1, i_1 j_1} - T_{k_1, i_2 j_2}) \right\} \\
&- \sum_{k_1=1}^q \sum_{k_2 \neq k_1} \frac{1}{n m} \sum_{i_1=1}^{n_{k_1}} \sum_{i_2=1}^{n_{k_2}} \sum_{j_1=1}^m \sum_{j_2=1}^m \varepsilon_{k_1, i_1 j_1} \varepsilon_{k_2, i_2 j_2} \\
&\quad \times \left\{ \sum_{\ell=1}^m \nu_{k_1, i_1}^{j_1 \ell} \mu_{k_1, i_1 \ell}^{(1)} \frac{\mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2}}{B_1(T_{k_1, i_1 \ell})} K_h(T_{k_2, i_2 j_2} - T_{k_1, i_1 \ell}) \right. \\
&\quad \left. - \frac{1}{2} \mu_{k_1, i_1 j_1}^{(1)} \omega_{k_1, i_1}^{j_1 j_1} \mu_{k_2, i_2 j_2}^{(1)} \omega_{k_2, i_2}^{j_2 j_2} \frac{B_2(T_{k_2, i_2 j_2})}{B_1^2(T_{k_2, i_2 j_2})} K_h * K_h(T_{k_1, i_1 j_1} - T_{k_2, i_2 j_2}) \right\}.
\end{aligned} \tag{W.3}$$

We can see that

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

where

$$\begin{aligned}
V_w &= \text{E} \left[ \frac{1}{m^2} \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \sum_{j_4=1}^m \varepsilon_{i_1 j_1} \varepsilon_{i_2 j_2} \varepsilon_{i_1 j_3} \varepsilon_{i_2 j_4} \left\{ \sum_{\ell_1=1}^m \nu_{i_1}^{j_1 \ell_1} \mu_{i_1 \ell_1}^{(1)} \frac{\mu_{i_2 j_2}^{(1)} \omega_{i_2}^{j_2 j_2}}{B_1(T_{i_1 \ell_1})} K_h(T_{i_2 j_2} - T_{i_1 \ell_1}) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \mu_{i_1 j_1}^{(1)} \omega_{i_1}^{j_1 j_1} \mu_{i_2 j_2}^{(1)} \omega_{i_2}^{j_2 j_2} \frac{B_2(T_{i_2 j_2})}{B_1^2(T_{i_2 j_2})} K_h * K_h(T_{i_1 j_1} - T_{i_2 j_2}) \right\} \right. \\
&\quad \times \left\{ \sum_{\ell_2=1}^m \nu_{i_1}^{j_3 \ell_2} \mu_{i_1 \ell_2}^{(1)} \frac{\mu_{i_2 j_4}^{(1)} \omega_{i_2}^{j_4 j_4}}{B_1(T_{i_1 \ell_2})} K_h(T_{i_2 j_4} - T_{i_1 \ell_2}) \right. \\
&\quad \left. \left. - \frac{1}{2} \mu_{i_1 j_3}^{(1)} \omega_{i_1}^{j_3 j_3} \mu_{i_2 j_4}^{(1)} \omega_{i_2}^{j_4 j_4} \frac{B_2(T_{i_2 j_4})}{B_1^2(T_{i_2 j_4})} K_h * K_h(T_{i_1 j_3} - T_{i_2 j_4}) \right\} \right] \\
&+ \text{E} \left[ \frac{1}{m^2} \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m \sum_{j_4=1}^m \varepsilon_{i_1 j_1} \varepsilon_{i_2 j_2} \varepsilon_{i_2 j_3} \varepsilon_{i_1 j_4} \left\{ \sum_{\ell_1=1}^m \nu_{i_1}^{j_1 \ell_1} \mu_{i_1 \ell_1}^{(1)} \frac{\mu_{i_2 j_2}^{(1)} \omega_{i_2}^{j_2 j_2}}{B_1(T_{i_1 \ell_1})} K_h(T_{i_2 j_2} - T_{i_1 \ell_1}) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \mu_{i_1 j_1}^{(1)} \omega_{i_1}^{j_1 j_1} \mu_{i_2 j_2}^{(1)} \omega_{i_2}^{j_2 j_2} \frac{B_2(T_{i_2 j_2})}{B_1^2(T_{i_2 j_2})} K_h * K_h(T_{i_1 j_1} - T_{i_2 j_2}) \right\} \right. \\
&\quad \times \left\{ \sum_{\ell_2=1}^m \nu_{i_2}^{j_3 \ell_2} \mu_{i_2 \ell_2}^{(1)} \frac{\mu_{i_1 j_4}^{(1)} \omega_{i_1}^{j_4 j_4}}{B_1(T_{i_2 \ell_2})} K_h(T_{i_1 j_4} - T_{i_2 \ell_2}) \right. \\
&\quad \left. \left. - \frac{1}{2} \mu_{i_2 j_3}^{(1)} \omega_{i_2}^{j_3 j_3} \mu_{i_1 j_4}^{(1)} \omega_{i_1}^{j_4 j_4} \frac{B_2(T_{i_1 j_4})}{B_1^2(T_{i_1 j_4})} K_h * K_h(T_{i_2 j_3} - T_{i_1 j_4}) \right\} \right]
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{2}\mu_{i_2j_3}^{(1)}\omega_{i_2}^{j_3j_3}\mu_{i_1j_4}^{(1)}\omega_{i_1}^{j_4j_4}\frac{B_2(T_{i_1j_4})}{B_1^2(T_{i_1j_4})}K_h * K_h(T_{i_2j_3} - T_{i_1j_4})\Big\} \\
= & \frac{1}{h}\mathbb{E}\left\{\frac{B_3(T)B_4(T) + B_5^2(T)}{B_1^2(T)f(T)}\int K^2(t)dt + \frac{B_2^2(T)B_3^2(T)}{2B_1^4(T)f(T)}\int (K * K)^2(t)dt\right. \\
& \left. - 2\frac{B_2(T)B_3(T)B_5(T)}{B_1^3(T)f(T)}\int K(t)K * K(t)dt\right\} + O(1).
\end{aligned}$$

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

## W.2 Proof of Theorem 2

**LEMMA W.3** *Suppose assumptions (C.1) – (C.6) and the local alternative described in (18) and (19) hold,  $\widehat{\beta}_R$  is still root- $n$  consistent to  $\beta_0$ , and  $\widehat{\beta}_F - \widehat{\beta}_R = o_p(n^{-1/2})$ . The nonparametric estimator  $\widehat{\theta}_R(t)$  has the same asymptotic expansion as in (10).*

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

$$\begin{aligned}
\widehat{\theta}_R(t; \beta) - \theta_0(t) &= \frac{1}{mnB_1(t)}\sum_{k=1}^q\sum_{i=1}^{n_k}\sum_{j=1}^m\mu_{k,ij}^{(1)}\omega_{k,i}^{jj}K_h(T_{k,ij} - t)\{\varepsilon_{k,ij} + \mu_{k,ij}^{(1)}G_{kn}(T_{k,ij})\} \\
&+ \frac{h^2}{2}\theta_0^{(2)}(t) - \boldsymbol{\mu}_X^T(\beta - \beta_0) + o_p(n^{-1/2} + \|\beta - \beta_0\|). \tag{W.4}
\end{aligned}$$

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

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

Therefore, if  $\widehat{\beta}_R - \beta_0 = O_p(n^{-1/2})$ , the expansion of  $\widehat{\theta}_R(t)$  follows directly from (W.4) and the leading terms are identical to those in (10).

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

$$\widehat{\beta}_R - \beta_0 = \mathbf{D}_\dagger^{-1}\mathcal{E}_{n\dagger} + o_p(n^{1/2}),$$

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

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

◇

**Proof of Theorem 2:** The test statistic has a similar decomposition as (W.1)

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

where

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

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

Under the local alternative,  $G_{kn}(T) = O_p\{(nh)^{-1/2}\}$ , one can show  $\boldsymbol{\mu}_{X,k}(t) - \boldsymbol{\mu}_k(t) = O\{(nh)^{-1/2}\}$ . By Proposition 2 and Lemma W.3,

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

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

$o_p(h^{-1/2})$ , where

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

More detailed calculation shows

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

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

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

Therefore  $\lambda_n(H_{1n}) = \mu_n + W_n + R_1^\dagger + R_2^\dagger + o_p(h^{-1/2})$ , where  $\mu_n$  and  $W_n$  are as defined in Theorem 1. By the assumption in (19),  $R_2^\dagger = \mu_{1n} + O_p(n^{-1/2}h^{-1}) = \mu_{1n} + o_p(h^{-1/2})$ . Since  $R_1^\dagger$  is a linear combination of  $\varepsilon_{k,ij}$  and  $W_n$  only consists of quadratic terms, it is easy to

see that  $R_1^\dagger$  and  $W_n$  are uncorrelated and hence asymptotically independent. Therefore,  $\mathbb{E}\{\lambda_n(H_{1n})\} = \mu_n + \mu_{1n} + o_p(h^{-1/2})$  and  $\text{var}\{\lambda_n(H_{1n})\} = \text{var}(W_n) + \text{var}(R_1^\dagger) + o_p(h^{-1}) = \sigma_{1n}^2$ . The asymptotic normality of  $\lambda_n(H_{1n})$  follows from those of  $W_n$  and  $R_1^\dagger$ .  $\diamond$

### W.3 Proof of Theorem 3

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

$$\beta(\alpha, \mathbf{G}_n) = \Phi\{\sigma_{1n}^{-1}(z_\alpha \sigma_n - \mu_{1n})\},$$

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

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

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

### W.4 Proof of Theorem 4

By Proposition 2,  $\widehat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$  and  $\sup_{t \in \mathcal{T}} |\widehat{\theta}_{F,k}(t) - \theta_0(t)| = O_p\{h^2 + (\log n/nh)^{1/2}\}$ , the residuals of the full model satisfy  $\widehat{\varepsilon}_{k,ij} - \varepsilon_{k,ij} = O_p\{h^2 + (\log n/nh)^{1/2}\}$  and  $\mathbb{E}(\widehat{\varepsilon}_{k,ij}^2 | \mathcal{X}) = \sigma^2(\mu_{k,ij}) + O_p\{h^2 + (\log n/nh)^{1/2}\}$  uniformly for all  $k, i$  and  $j$ . Since the perturbation factors  $\xi_{k,ij}$  are of mean 0 and variance 1, it is easy to see that  $\mathbb{E}(\varepsilon_{k,ij}^* | \mathcal{X}) = 0$ ,  $\text{var}(\varepsilon_{k,ij}^* | \mathcal{X}) = \mathbb{E}(\widehat{\varepsilon}_{k,ij}^2 | \mathcal{X}) = \sigma^2(\mu_{k,ij}) + O_p\{h^2 + (\log n/nh)^{1/2}\}$  uniformly for all  $k, i$  and  $j$  and  $\text{corr}(\varepsilon_{k,ij}^*, \varepsilon_{k',i',j'}^*) = 0$  for  $(k, i, j) \neq (k', i', j')$ . As a result, conditioning on  $\mathcal{X}$  the bootstrap sample  $\{(Y_{k,ij}^*, \mathbf{X}_{k,ij}, T_{k,ij})\}$  are mutually independent and satisfy model

(1) and the null hypothesis  $H_0$  with the true parameters  $\beta_0^* = \widehat{\theta}_R$ ,  $\theta_0^*(\cdot) = \widehat{\theta}_R(\cdot)$  and  $\text{var}(\varepsilon^*|\mathcal{X}) = \sigma^2(\mu^*) + O_p\{h^2 + (\log n/nh)^{1/2}\}$ .

Using the same arguments as for Theorem 1, we can show  $\lambda_n^*(H_0) = R_1^* + R_4^* + W_n^* + o_p(h^{-1/2})$  where  $R_1^*$ ,  $R_4^*$  and  $W_n^*$  are the same as  $R_1$ ,  $R_4$  and  $W_n$  in Lemma W.2 except to replace  $\varepsilon_{k,ij}$  with  $\varepsilon_{k,ij}^*$ . By similar calculation as Lemma W.2,  $R_1^* + R_4^* = \mu_n + O_p(1)$  and  $\text{var}(W_n^*|\mathcal{X}) = \sigma_n^2 \times \{1 + o_p(1)\}$ , where the functions  $B_1(t) - B_5(t)$  are evaluated at  $\beta_0^* = \widehat{\theta}_R$ ,  $\theta_0^*(\cdot) = \widehat{\theta}_R(\cdot)$  and  $\text{var}(\varepsilon^*|\mathcal{X}) = \sigma^2(\mu^*) + O_p\{h^2 + (\log n/nh)^{1/2}\}$ . Because  $\varepsilon_{k,ij}^*$ 's are uncorrelated,  $\mu_n$  and  $\sigma_n$  simplifies to  $\mu_{n^*}$  and  $\sigma_{n^*}$  in Corollary 2, which do not depend on the true model parameters due to the Wilks phenomenon. 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 Corollary 2 for every event defined on  $\mathcal{X}$ . The conclusion of the theorem follows.

## W.5 Additional simulation results for binary longitudinal data

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

$$\text{logit}(p_{k,ij}) = X_{k,ij}\beta + \theta_k(T_{k,ij}).$$

We generate  $T_{k,ij}$  and  $X_{k,ij}$  in a similar way as in Simulation 1, and assume an exchangeable within-cluster correlation such that  $\text{corr}(Y_{k,ij}, Y_{k,ij'}) = \rho_{jj'} = 0.3$  for  $j \neq j'$ . To generate binary responses with the desired mean and correlation structures, we use a truncated Bahadur representation (Bahadur, 1961) ignoring expansions of order three and higher. Specifically,  $\mathbf{Y}_{k,i}$  is generated from the following joint distribution

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

where  $\widetilde{y}_j = (y_j - p_j) / \sqrt{p_j(1 - p_j)}$  is a standardized version of  $y_j$ .

To verify the Wilks results in Corollary 2, we study the empirical distribution of the test statistic under the null hypothesis. We generated 300 datasets from each of the following

three scenarios

$$\text{Scenario IV : } \beta = 0.5, \theta_0(t) = \frac{1}{2} \sin\left(\frac{3}{4}\pi t\right);$$

$$\text{Scenario V : } \beta = -1, \theta_0(t) = \frac{1}{2} \sin\left(\frac{3}{4}\pi t\right);$$

$$\text{Scenario VI : } \beta = 0.5, \theta_0(t) = \sin(\pi t) - 0.5.$$

For binary responses, it is natural to use a binary quasi-likelihood for the test. In such a quasi-likelihood, it is difficult to incorporate the within-subject correlation. More importantly, our previous simulation shows that incorporating correlation into the proposed test does not always increase its power. Therefore, we focus on a working independence GQLR test using the quasi-likelihood in (23).

As our theory shows, when working independence is used for both estimation and hypothesis test, the distribution of  $\lambda_n(H_0)$  does not depend on the true correlation structure, hence is identical to the case when the data are independent. Based on this result, we advocate a bootstrap procedure by simulating independent responses. To support this claim, we also compare the empirical distribution of  $\lambda_n(H_0)$  to a case where the responses are truly independent. We generate 300 datasets from a setting that is identical to scenario VI except that  $Y_{k,ij}$ 's are independent.

The empirical distributions of  $\lambda_n(H_0)$  under Scenarios IV - VI with correlated responses and that under Scenario VI with independent responses are shown in Figure W.1. As we can see, the four distributions are almost identical. The mean and standard deviation of  $\lambda_n(H_0)$  under different scenarios are displayed in Table W.1. We perform Kolmogorov-Smirnov tests to test if  $\lambda_n(H_0)$  from different scenarios have the same distribution and none of these tests is significant. We conclude that the distributions of  $\lambda_n(H_0)$  under the three Scenarios with dependent responses are the same as the distribution under Scenario VI with independent response variables. These results also corroborate our theory and the proposed bootstrap procedure.

	Correlated data			Indepadent data
	Scenario IV	Scenario V	Scenario VI	Senario VI
Mean	5.87	5.99	5.68	5.57
SD	2.57	2.44	2.44	2.49

Table W.1: Additional simulation: mean and standard deviation (SD) of  $\lambda_n(H_0)$  for binary longitudinal data under Scenarios IV - VI and those of  $\lambda_n(H_0)$  under Scenario VI and with independent responses.

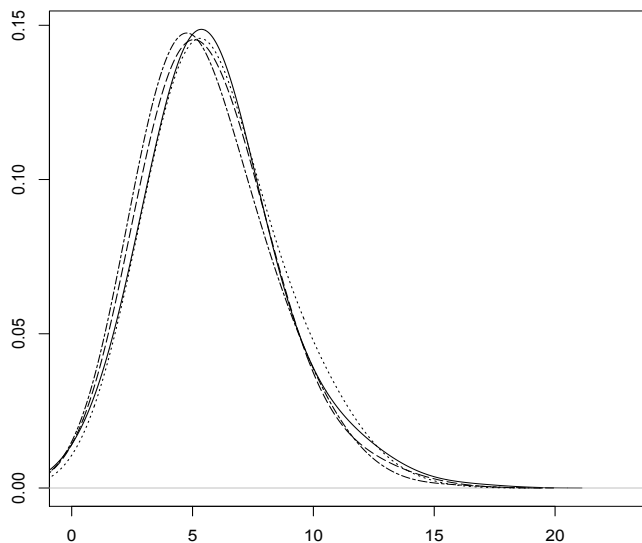


Figure W.1: Additional simulation: empirical distributions of  $\lambda_n(H_0)$  for correlated binary data (Scenario IV: solid line; Scenario V: dotted line; Scenario VI: dashed line) and independent binary data under parameter specification in Scenario VI (dotdash line). The test statistics are based on the binary quasi-likelihood (23).