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

Jin Tang, Yehua Li, and Yongtao Guan

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 misspecified. 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. Supplementary materials for this article are available online.

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
\[
g(\mu_{k,i}(t)) = X_{k,i}(t)\beta + \theta_k(t),
\]
where $g(\cdot)$ is a known link function, $X_{k,i}(\cdot)$ is a $p$-dim covariate vector that may be time dependent, $\beta$ 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:
\[
H_0 : \theta_1 = \cdots = \theta_q \quad \text{versus} \quad H_1 : \text{not all } \theta_k \text{'s are the same.}
\]
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 follow-up period and the associated covariate vector $X_{k,i}$ 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 $X_{k,i}$. To answer this question, we will test hypotheses in the form of (2). We do not require any parametric form for $\theta_k$'s to maintain modeling flexibility.

The generalized likelihood ratio (GLR) test proposed by Fan, Zhang, and Zhang (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 articles 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 article, 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, Carroll, and Lin (2005). However, most of these articles 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 articles 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 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 \( \hat{b}_k \)'s, which cannot replace a rigorous test as we are about to propose.

The rest of the article 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 \( T \). Let \( T_{k,i} = (T_{k,i1}, \ldots, T_{k,i,m_{k,i}})^T \) 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 \( Y_{k,i} = (Y_{k,i1}, \ldots, Y_{k,i,m_{k,i}})^T \), \( \mu_{k,i} = (\mu_{k,i1}, \ldots, \mu_{k,i,m_{k,i}})^T \), \( \beta_{k,i} = (\beta_{k,i1}, \ldots, \beta_{k,i,m_{k,i}})^T \), where \( Y_{k,i} = \beta_{k,i} + \mu_{k,i} + \epsilon_{k,i} \), \( \epsilon_{k,i} \) are independent and identically distributed errors with \( \epsilon_{k,i} \sim N(0, \Sigma) \). Note that the conditional variance of the response is \( \text{var}(Y_{k,i}|\mu_{k,i}) = \sigma^2(\mu_{k,i}) \), 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 \( \hat{b}_R \) and \( \hat{b}_G(t) \) as the estimators under the reduced model and \( \beta_F \) and \( \theta_F(t), k = 1, \ldots, q \), as the estimators under the full model. Our estimation procedures under both models are based on profile-kernel estimating equations.

We first consider estimation under the reduced model, where 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

\[
\theta(T_{k,ij}) \approx \theta(t) + \frac{\partial \theta(t)}{\partial t}(T_{k,ij} - t) = \alpha_0 + \alpha_1(T_{k,ij} - t)/h.
\]

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 \( U_{k,i}(t) = [U_{k,i1}(t), \ldots, U_{k,i,m_{k,i}}]^T \) with \( U_{k,i}(t) = [1, (T_{k,ij} - t)/h]^T \). For a given \( \beta, \theta(t) \) is estimated by solving the following local linear estimating equation regarding \( \alpha = (\alpha_0, \alpha_1)^T \),

\[
\sum_{k=1}^q \sum_{i=1}^{m_{k,i}} U_{k,i}(t)^T \Delta_{k,i}(t)K_h(T_{k,ij} - t)
\times \{Y_{k,i} - \mu_{k,i}(t)\} = 0,
\]

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

\[
\sum_{k=1}^q \sum_{i=1}^{m_{k,i}} \left[X_{k,ij}^T + \frac{\partial \hat{\theta}_R(T_{k,ij}; \beta)}{\partial \beta}\right] \Delta_{k,i}(T_{k,ij})W_{k,i}^{-1}
\times \{Y_{k,i} - g^{-1}[X_{k,ij}^T \beta + \hat{\theta}_R(T_{k,ij}; \beta)]\} = 0.
\]

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

As shown by Lin and Carroll (2001), the most efficient estimators within the class defined by (3) are obtained by setting \( W_{k,i} = \text{diag}([\omega(\mu_{k,i,j})])_{j=1}^{m_{k,i}}, \) 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, Müller, and Wang (2005), Hall, Müller, and Wang (2006), Hall, Müller, and Yao (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 \( b_j(t) \) using the \( k \)-th treatment group only. For a given \( \beta \), the profile local linear estimator for \( \theta_F(t) \) is given by \( \hat{\theta}_F,k(t; \beta) = \hat{\theta}_R, \) where \( \hat{\theta}_R = (\hat{\alpha}_0, \hat{\alpha}_1)^T \) is the solution of

\[
\sum_{i=1}^{m_{k,i}} U_{k,i}(t)^T \Delta_{k,i}(t)K_h(T_{k,ij} - t)
\times \{Y_{k,i} - \mu_{k,i}(t)\} = 0.
\]

To obtain \( \hat{\beta}_F \), we will again solve an estimating equation by pooling all treatment groups together

\[
\sum_{k=1}^q \sum_{i=1}^{m_{k,i}} \left[X_{k,ij}^T + \frac{\partial \hat{\theta}_F,k(T_{k,ij}; \beta)}{\partial \beta}\right] \Delta_{k,i}(T_{k,ij})W_{k,i}^{-1}
\times \{Y_{k,i} - g^{-1}[X_{k,ij}^T \beta + \hat{\theta}_F,k(T_{k,ij}; \beta)]\} = 0.
\]

The nonparametric components are then updated as \( \hat{\theta}_F,k(t) = \hat{\theta}_F,k(t; \hat{\beta}_F) \), for \( k = 1, \ldots, q \).

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 \( \beta \) and \( \theta(t) \) under both the full and reduced models. Denote the true parameters as \( \beta_0 \) and \( \theta_{0,t}\), \( k = 1, \ldots, q \). Under the reduced model, \( \theta_{0,t} = \cdots = \theta_{0,t} \equiv \theta_0 \). For ease of exposition, we assume that the number of observations per subject is a constant in our theoretical derivations, that is \( m_{k,i} = ...
where \( C.2 \) The kernel density function \( \mathbb{R} \) is an unknown correlation parameter. Many authors modeled the variance function \( \sigma^2(\cdot) \) as a nonparametric function but the correlation function \( \mathbb{R}(s, t; \tau) \) as a member of a parametric family, such as the autoregressive (AR) or autoregressive and moving-average (ARMA) correlations (see, Fan, Huang, and Li 2007, and Fan and Wu 2008). The within-subject covariance matrix is

\[
\Sigma_{k,i} = \mathbf{S}_{k,i} \mathbf{R}_{k,i}(\tau) \mathbf{S}_{k,i},
\]

where \( \mathbf{S}_{k,i} = \text{diag}(\sigma(\mu_k)) \) and \( \mathbf{R}_{k,i}(\tau) = \{\mathbb{R}(T_{k,ij}, T_{k,ij}; \tau)\}_{i,j=1}^{n} \). We allow the covariance to depend on the mean and the parameters \( \mathbf{B} \) and \( \theta(\cdot) \), since this is usually the case for generalized longitudinal data, for example binary data.

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

(C.1) Assume that the true mean functions \( \theta_0(\cdot) \), \( k = 1, \ldots, q \), are smooth and twice continuously differentiable. Using the shorthand notation \( \mu_k(X, T) = g^{-1}(X^T(T)\mathbf{B}_0 + \theta_0(T)) \), \( \omega_k(X, T) = \omega(\mu_k(X, T)) \) and \( \mu^{(1)}_k(X, T) = \mu^{(1)}(X^T(T)\mathbf{B}_0 + \theta_0(T)) \), define

\[
B_{ik}(t) = E[[\mu^{(1)}_k(X, T)]]^2 \omega^{-1}_k(X, T)|T = t]f(t),
\]

\[
B_1(t) = \sum_{k=1}^q \rho_k B_{ik}(t),
\]

\[
\mu_{X,k}(t) = E[\mu^{(1)}_k(X, T)]^2 \omega^{-1}_k(X, T)|T = t]f(t)/B_{ik}(t),
\]

\[
\mu_{X}(t) = B^{-1}_1(t) \sum_{k=1}^q \rho_k E[[\mu^{(1)}_k(X, T)]]^2 \omega^{-1}_k
\]

\[
\times (X, T)|T = t]f(t).
\]

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, that is \( \int t K(t) dt = 0 \) and \( \int t^2 K(t) dt = 1 \).

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

Let \( \tilde{X}_{ik} \) be an \( m \times p \) matrix with the \( j \)th row being \( X_{k,ij} - \mu_{X,k}(T_{k,ij}), \) where \( \mu_{X,k}(\cdot) \) is defined in condition (C.1). For ease of exposition, we often suppress the index \( i \), and write \( \Sigma_k, \Delta_k, \mathcal{V}_k, \) and \( \tilde{X}_k \) as a generic version of these quantities for a typical subject receiving the \( k \)th treatment. When the null hypothesis in (2) is true, \( \Sigma_k = \Sigma, \Delta_k = \Delta, \mathcal{V}_k = \mathcal{V}, \) and \( \tilde{X}_k = \tilde{X} \) for \( k = 1, \ldots, q \). 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

\[
\hat{\beta}_R - \beta_0 = D_R^{-1} \mathcal{E}_R + o_p(n^{-1/2}),
\]

where \( D_R = E(\tilde{X}^T \Delta \mathcal{V}^{-1/2} \tilde{X}) \) and \( \mathcal{E}_R = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \mathcal{V}_k^{-1} \mathcal{E}_{k,i} \).

The nonparametric estimator has the following asymptotic expansion

\[
\hat{\beta}_R(t) - \theta_0(t) = \frac{1}{2} \theta_0^{(2)}(t)h^2 + U_R(t) - \mu_\tilde{X}_k(t)(\hat{\beta}_R - \beta_0) + o_p(n^{-1/2}),
\]

where \( U_R(t) = (mm_B(t))^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=1}^{m_k} \mu^{(1)}_{k,i} \omega_k(t) \epsilon_{k,i} \).

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, Carroll, and Lin 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),

\[
\hat{\beta}_F - \beta_0 = D_F^{-1} \mathcal{E}_F + o_p(n^{-1/2}),
\]

where \( D_F = \sum_{k=1}^q \rho_k E(\tilde{X}^T_k \Delta_k \mathcal{V}_k^{-1} \tilde{X}_k) \) and \( \mathcal{E}_F = n^{-1} \sum_{k=1}^q \sum_{i=1}^{n_k} \mathcal{V}_k^{-1} \mathcal{E}_{k,i} \).

The nonparametric estimator \( \hat{\theta}_F(t) \) has the following asymptotic expansion,

\[
\hat{\theta}_F(t) - \theta_0(t) = \frac{1}{2} \theta_0^{(2)}(t)h^2 + U_{F,k}(t) - \mu_{\tilde{X}_k}(t)(\hat{\beta}_F - \beta_0) + o_p(n^{-1/2}).
\]

for \( k = 1, \ldots, q \) where \( U_{F,k}(t) = (mB_{ik}(t))^{-1} \sum_{i=1}^{n_k} \sum_{j=1}^{m_k} \mu^{(1)}_{k,j} \omega_k(t) \epsilon_{k,j} + B_{ik}(t) \) is defined in condition (C.1).

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

\[
\frac{\partial Q(\mu, Y)}{\partial \mu} = V(\mu)^{-1}(Y - \mu),
\]

(13)

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

\[
\ell(\theta, \beta) = \sum_{k=1}^{q} \sum_{j=1}^{m} Q[g^{-1}(X_{k,j}\beta + \theta_k(T_{k,i})), Y_{k,i}].
\]

(14)

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

\[
\lambda_n(H_0) = \sum_{k=1}^{q} \sum_{j=1}^{m} \left( Q[g^{-1}(X_{k,j}\hat{\beta}_F + \hat{\theta}_F(T_{k,i})), Y_{k,i}] - Q[g^{-1}(X_{k,j}\hat{\beta}_R + \hat{\theta}_R(T_{k,i})), Y_{k,i}] \right).
\]

(15)

Under the null hypothesis, let \( \Sigma \) and \( V \) be generic copies of \( \Sigma_{k,i} \) and \( V_{k,i} \) in a typical subject. These matrices are random because the within-subject correlation might depend on the subject-specific time vector \( T \) and the conditional variance might depend on the conditional mean vector \( \mu = g^{-1}(X\beta_0 + \theta_0(T)) \). Denote \( \sigma_{jk}, \nu^j, \) and \( \Delta_{jk} \) as the \((j, \ell)\)th entry of \( \Sigma \), \( V^{-1} \) and \( \Delta \), respectively, and they are considered random as well. Define

\[
B_2(t) = \frac{1}{m} \sum_{j=1}^{m} E(\Delta_{jj}^2 \nu^j | T_j = t) f(t),
\]

\[
B_3(t) = \frac{1}{m} \sum_{j=1}^{m} E[\sigma_{jj} \Delta_{jj}^2 (\nu^j)^2 | T_j = t] f(t),
\]

\[
B_4(t) = \frac{1}{m} \sum_{j=1}^{m} E \left( \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \nu^{j_1} \nu^{j_2} \sigma_{jj_1 j_2} \Delta_{jj_1}^2 \Delta_{jj_2}^2 | T_j = t \right) f(t),
\]

\[
B_5(t) = \frac{1}{m} \sum_{j=1}^{m} E \left( \sum_{j=1}^{m} \sigma_{jj} \nu^j \Delta_{jj}^2 | T_j = t \right) f(t).
\]

(16)

Denote \( \nu_k = \int K^2(t) dt \) and \( K \) as the convolution of the kernel function \( K \), so that \( K * K(t) = \int K(s)K(t - s) ds \). 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 \( B_2(t) - B_3(t) \) defined in (16) exist and are Lipschitz continuous in \( t \). In addition, assume that \( nh^2 \to 0 \), then under the null hypothesis in (2)

\[
\sigma_n^{-1}(\lambda_n(H_0) - \mu_n - d_n) \rightarrow N(0, 1),
\]

where \( d_n = \sigma_\nu(h^{-1/2}) \),

\[
\mu_n = \frac{q - 1}{h} \left\{ K(0) \frac{B_3(T)}{B_1(T) f(T)} - \nu_k \frac{B_5(T)}{B_1(T) f(T)} \right\} + O_p(1),
\]

\[
\sigma_n^2 = \frac{q - 1}{h} \left\{ \frac{B_3(T)}{B_1(T) f(T)} \int K^2(t) dt + \frac{B_2^2(T)}{2B_1^2(T) f(T)} \int (K * K)^2(t) dt \right\} - \frac{2B_5(T)}{B_1^2(T) f(T)} \left( \int K(t)K(t) dt \right) + O_p(1).
\]

Remark 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 \( \Sigma_{k,i} \) and the local expansion of the link function \( \Delta_{k,i} \) are random and subject specific; 3) we allow the working covariances \( V_{k,i} \) and \( V_{k,i} \) to be misspecified. To be more specific on the second challenge, \( \Sigma_{k,i} \) and \( \Delta_{k,i} \) depend on the mean vector \( \mu_{k,i} \) and the within-subject correlation may depend on the subject specific time vector \( T_{k,i} \). As a result, our proof is more involved when calculating the asymptotic mean and variance of the test statistic.

Remark 2. For nonparametric hypothesis testing in independent data, Fan, Zhang, and Zhang (2001) established a property called the Wilks phenomenon for the GLR test, that is 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 (coefficient) of \( \epsilon \) depends on the mean, and if \( V, W, \) and \( \Sigma \) are different, \( B_3(t) - B_2(t) \) and thus, the asymptotic distribution of \( \lambda_n(H_0) \) in Theorem 1 depend on the true parameters \( \beta_0 \) 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 \( \beta_0 \) and \( \theta_0(t) \). Recall that the true covariance matrix \( \Sigma \) has the structure as in (8) and denote \( S_d = \Sigma_d^\tau \) 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 \( \Sigma = \Sigma_d \) and \( V = \Sigma C(\tau)S \), where \( C(\tau) \) is a working correlation matrix that may be different from the true correlation matrix \( R(\tau) \).
When \( \Sigma = SR(\tau)S, W = \Sigma_d, \) and \( V = SC(\tau)S, \) one can show \( B_{j}(t) = B_{j}(t)B_{1}(t) \) for \( j = 2, \ldots, 5, \) where

\[
B_{2}(t) = \frac{1}{m} \sum_{j=1}^{m} E[(C^{-1}(\tau)_{j}][T_{j} = t], \quad B_{3}(t) = 1,
\]

\[
B_{4}(t) = \frac{1}{m} \sum_{j=1}^{m} \left[ (C^{-1}(\tau)R(\tau)C^{-1}(\tau)_{jj}][T_{j} = t]. \right.
\]

\[
B_{5}(t) = \frac{1}{m} \sum_{j=1}^{m} \left[ (R(\tau)C^{-1}(\tau)_{jj}][T_{j} = t] \right.
\]

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

**Corollary 2.** Under the setting of Theorem 1, suppose \( \mathcal{W} = \Sigma_d, \) \( \mathcal{V} = SC(\tau)S, \) and \( \Sigma = SR(\tau)S, \) then the asymptotic distribution of \( \lambda_n(H_0) \) under the null hypothesis in (2) is

\[
\sigma_{nt}^{-1}\{\lambda_n(H_0) - \mu_n - d_n\} \xrightarrow{d} \text{Normal}(0, 1),
\]

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

\[
\mu_n = \frac{q - 1}{h} \int \left[ (K(0)B_{5}(T) - \frac{v}{2} B_{3}(T))/f(T) \right] + O_P(1),
\]

\[
\sigma_{nt}^{2} = \frac{q - 1}{h} \int \left[ B_{4}(T) + B_{2}(T)/f(T) \right] \int K^{2}(t)dt + \frac{B_{2}^{2}(T)}{2f(T)} \int (K * K)^{2}(t)dt - 2B_{3}(T)B_{2}(T)/f(T) \int (K(t)K * K(t))dt + O_P(1).
\]

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

From the expressions of \( B_{j}(t), j = 2, \ldots, 5, \) it is easy to see that they do not depend on \( \beta_0 \) 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 \( R(\tau) \) 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 dependence is assumed for both estimation and hypothesis testing.

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

\[
\sigma_{nt}^{-1}\{\lambda_n(H_0) - \mu_n - d_n\} \xrightarrow{d} \text{Normal}(0, 1),
\]

where \( d_n = o_p(h^{-1/2}), \) \( \mu_n = (q - 1)T|T|^{-1}(K(0) - \frac{v}{2}), \)

\[
\sigma_{nt}^{2} = 2(q - 1)T|T|^{-1} \int (K(t) - \frac{1}{2}K * K(t))^{2}dt, \quad \text{and} \quad |T| \text{is the length of the time domain.}
\]

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 \( \beta_0, \theta_0(t) \) or the true correlation structure \( R(\tau). \) 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

\[
H_{1n}: \theta_k(t) > \theta_0(t) + G_{kn}(t),
\]

where \( G_{kn}(t) \) are twice continuously differentiable smooth functions with \( \sup_{t \in T} G_{kn}(t) \rightarrow 0 \) 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 in the supplementary material.

**Theorem 2.** Suppose that assumptions (C.1) - (C.3) and the local alternative (18) hold, \( m^h \rightarrow 0, \) 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} E[G_{kn}^{T}(T_{i})\Delta_{k, i}V_{-1}^{1}\Delta_{k, i}G_{kn}(T_{i})] \quad \text{and} \quad \sigma_{1n}^{2} = \sigma_{nt}^{2} + \sum_{k=1}^{q} \sum_{i=1}^{n} E[G_{kn}^{T}(T_{i})\Delta_{k, i}V_{-1}^{1}\Delta_{k, i}G_{kn}(T_{i})] \quad \text{and} \quad \mu_{n} \quad \text{and} \quad \sigma_{nt}^{2} \quad \text{are as defined in Theorem 1.}

An approximate level-\( \alpha \) test based on the GQLR test statistic is \( \varphi_n = \left[ \lambda_n(H_{1n}) - \mu_n - \mu_{1n} \right] \rightarrow N(0, 1), \)

where \( \sigma_{1n}^{2} \) and \( \mu_n, \sigma_n \) are defined in Theorem 1.

Then the test statistic has the following limiting distribution

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

where \( \sigma_{nt}^{2} = \sigma_{nt}^{2} + \sum_{k=1}^{q} \sum_{i=1}^{n} E[G_{kn}^{T}(T_{i})\Delta_{k, i}V_{-1}^{1}\Delta_{k, i}G_{kn}(T_{i})] \quad \text{and} \quad \mu_n \quad \text{and} \quad \sigma_{nt}^{2} \quad \text{are as defined in Theorem 1.}

Following Ingster (1993a,b,c) and Fan, Zhang, and Zhang (2001), define the minimax rate of a test \( \varphi \) with a Type II error \( \beta(\alpha, \varphi) \) as \( \varphi_n \) such that

(a) for any \( \varphi > \varphi_n, \alpha > 0, \beta > 0, \) there exists a constant \( c \) such that \( \beta(\alpha, \varphi) \leq \beta + o(1); \)

(b) for any sequence \( \varphi_n \) such that \( \varphi_n \rightarrow \infty \) for any \( c > 0 \) such that \( \beta(\alpha, \varphi) = \alpha + o(1) \) and \( \lim \inf_{\varphi_n \rightarrow \infty} \beta(\alpha, \varphi) > \beta. \)
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 \( g_n(h) = n^{-4/9} \) 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 (1993a,b,c).

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 \( e(t) \) 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, Müller, and Wang 2005; Fan, Huang, and Li 2007; Li 2011). In what follows, we assume that the variance, denoted as \( \hat{\sigma}^2(\mu) \), is a smooth function of the mean value, and estimate this function using a local linear smoother

\[
\hat{\sigma}^2(\mu) = e^T \left\{ \sum_{k=1}^{q} \sum_{i=1}^{n_k} \sum_{j=1}^{m_{k,i}} U_{k,ij}(\mu) U_{k,ij}^T(\mu) K_k(\tilde{\mu}_{k,ij} - \mu) \right\}^{-1} \\
\times \left\{ \sum_{k=1}^{q} \sum_{i=1}^{n_k} \sum_{j=1}^{m_{k,i}} U_{k,ij}(\mu) K_k(\tilde{\mu}_{k,ij} - \mu) e_{k,ij}^2 \right\},
\]

where \( e = (1, 0)^T \), \( U_{k,ij}(\mu) = \{1, (\tilde{\mu}_{k,ij} - \mu)/h\}^T \), \( \tilde{\mu}_{k,ij} \) are pilot estimators of \( \mu_{k,ij} \) from the full model by setting \( V_{k,i} \) to be identity matrices and \( \tilde{e}_{k,ij} \) 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 \( \hat{\sigma}^2(\mu) \) 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 \( V_{k,i} = S_k C_{k,i}(\tau) S_k \), where \( C_{k,i}(\tau) \) is the working correlation matrix for the \((k, i)\)th subject with a correlation parameter \( \tau \), then \( \tau \) can be estimated by the quasi-maximum likelihood (QMLE) method of Fan, Huang, and Li (2007). The estimator \( \hat{\tau} \) is the maximizer of

\[
-\frac{1}{2} \sum_{k=1}^{q} \sum_{i=1}^{n_k} \left\{ \log|C_{k,i}(\tau)| + \hat{\sigma}_{k,i}^2 C_{k,i}^{-1}(\tau) \hat{S}_{k,i}^{-1} \hat{S}_{k,i} \right\},
\]

where \( \hat{S}_{k,i} = \text{diag}(\hat{\sigma}(\mu_{k,i,j}))_{j=1}^{m_{k,i}} \). When the working correlation model is correctly specified, Fan and Wu (2008) showed that \( \hat{\tau} \) 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 \( B_1(t) \cdot B_2(t) \) 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

![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. (a) Exchangeable working correlation. (b) Working independence.](image-url)
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, for example 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 \( Y_{k,ij}^* = g^{-1}(\mu^*_{k,ij}) + \epsilon^*_{k,ij} \), where \( \mu^*_{k,ij} = X_{k,ij}^T \hat{\beta}_R + \hat{\theta}_R(T_{k,ij}) \).
ξk,ij = ξ̂k,ij, ξ̂k,ij’s are the residuals from the full model and ̂ξk,ij’s are independent random perturbation factors with mean 0 and variance 1.

(iii) Calculate the GQLR test statistic λn(Ho) from the bootstrap sample {Y∗ k,i, Xk,i, Tk,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 λn(Ho). The estimated p-value is the percentage of λn(Ho) that are greater than λn(Ho).

Davidson and Flachaire (2008) advocate to generate the perturbation factor ̂ξk,ij in Step (ii) from a simple Rademacher distribution, that is P(ξ = 1) = P(ξ = −1) = 1/2, which is also what we use in our numerical studies. Put X∗ = (Xk,i, Tk,i), i = 1, . . . , mk, k = 1, . . . , q). The following theorem shows the consistency of the bootstrap procedure above. It states that, conditioning on X∗, λn(Ho) from the working independent bootstrap procedure above has the same asymptotic normal distribution as λn(Ho) in Corollary 2. The proof of the theorem is provided in the supplementary material.

**Theorem 4.** Under the same assumptions as Corollary 2,

\[
P[σ^2_{ns}(λn(Ho) − μns − dns) < x \mid X] \to Φ(x) \quad \text{in probability for any } x,
\]

where σns, μns, and dns 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 ̂hopt be the minimizer of

\[
CV(h) = \sum_{k=1}^q \sum_{i=1}^{mk} Q[g^{-1}[Xk,i ̂βh−(k,i) + ̂θh−(k,i)(Tk,i)], Yk,i],
\]

where Q(·) is the quasi-likelihood function and ̂βh−(k,i) and ̂θh−(k,i) are the full model estimators with bandwidth h by removing the (k, i)th subject from the dataset. 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 ̂h = ̂hopt × 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:

\[
Y_{k,ij} = X_{k,ij}β + θ_{k}(Tk,ij) + ε_{k,ij}, \quad k = 1, \ldots, 4,
\]

\[
i = 1, \ldots, 50, \quad j = 1, \ldots, 4,
\]

where Tk,ij are independent random variables following a uniform distribution on [0, 1] and Xk,ij = Tk,ij + Uk,ij, Uk,ij ~ i.i.d. Uniform(−1, 1). Under this setting, the marginal density of (Xk,ij, Tk,ij) is the same for all (k, i, j) and E(Xk,ij | Tk,ij, Tk,l) = E(Xk,ij | Tk,ij) for j ≠ l. We assume that the variance of εk,ij depends on the conditional mean and is given as follows

\[
σ^2(μ) = \text{var}(Y|X, T) = 0.5μ^2 + 0.3, \quad \text{where } μ = E(Y|X, T).
\]

We also assume an ARMA(1,1) correlation structure for the errors within the same subject, that is corr(ε(s), ε(t)) = 1 for s = t and γ exp(−|s − t|/v) for s ≠ t, and set the correlation parameters at γ = 0.6 and v = 1.

To examine the Wilks phenomenon of the GQLR test, we first assume that the null hypothesis is true, that is θ1(t) = θ2(t) = ··· = θp(t) = 0. Then we examine the Wilks statistic:

\[
λ = \sum_{h=1}^q \sum_{i=1}^{mk} Q[g^{-1}[Xk,i ̂βh−(k,i) + ̂θh−(k,i)(Tk,i)], Yk,i] - \sum_{h=1}^q \sum_{i=1}^{mk} Q[g^{-1}[Xk,i ̂βh−(k,i), Yk,i] - \sum_{h=1}^q \sum_{i=1}^{mk} Q[g^{-1}[Xk,i ̂βh−(k,i), Yk,i]
\]

This test statistic has an asymptotic χ2 distribution with q−1 degrees of freedom.
\[ \theta_i(t) = \theta_0(t), \] 
and consider the following three scenarios:

- **Scenario I:** \( \beta = 1, \theta_0(t) = \cos(2\pi t); \)
- **Scenario II:** \( \beta = -1, \theta_0(t) = \cos(2\pi t); \)
- **Scenario III:** \( \beta = 1, \theta_0(t) = \sin(2\pi t). \)

We simulate 200 datasets 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 dataset, we first fit the full model in a pilot analysis setting \( \forall k_i's \) 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

\[
Q(\mu, Y) = -\frac{1}{2}(Y - \mu)^T V(\mu)^{-1}(Y - \mu),
\]

where \( V(\mu) = S(\mu)C(\tau)S(\mu), S(\mu) = \text{diag}(\sigma(\mu_j)), \) and \( C(\tau) \) is a working correlation matrix. For the working correlation \( C(\tau), \) we consider both a misspecified 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, Huang, and Li (2007). The bandwidth for the local linear estimators is chosen using the method described in Section 4.3 on a pilot dataset 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 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 misspecified 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 confirm 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 \( \beta \) 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 \( C(\tau) \). 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 misspecified 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 Misspecification

To better understand the behavior of the GQLR test when the variance is misspecified, we consider the same simulation as above but misspecify 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 misspecified 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 \( \theta_i(t) = \theta_0(t) - 2\phi G(t), \theta_i(t) = \theta_0(t) - \phi G(t), \theta_i(t) = \theta_0(t) + \phi G(t), \) and \( \theta_i(t) = \theta_0(t) + 2\phi G(t). \) 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 \( \sigma_n^2, \sigma_n^2, \mu_n, \) and \( \mu_n \) 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 datasets. 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 \( \hat{Q}_{1-\alpha} \). For each \( 1 - \alpha \), the lower end of the vertical bar is the 2.5% percentile of the bootstrap quantile \( \hat{Q}_{1-\alpha} \), and the upper end of the bar is the 97.5% percentile of \( \hat{Q}_{1-\alpha} \). 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 dataset 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, Hearon, and Otto 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 12 50-min weekly cognitive behavioral therapy sessions during the first 12 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 \((1 = \text{female} / 0 = \text{male})\) and the highest level of education completed \((1 = \text{High School or Higher} \text{ and } 0 = \text{otherwise});\) the time variable was day with the range from day 0 to day 195. The response variable was urine toxicology testing result \((1 = \text{positive} / 0 = \text{negative}).\)

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.

| 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   |
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

\[
Q_{\text{binary}}(\mu, Y) = \sum_{j=1}^{m} Y_j \log(\mu_j/(1 - \mu_j)) + \log(1 - \mu_j).
\]

(23)

The \( p \)-value of the test is 0.044 based on 1000 bootstrap replications. 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 misspecified, 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, Carroll, and Lin (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.

Supplementary Materials

The online supplementary material contains technical proofs for the theoretical results and additional simulation results.

Acknowledgments

We thank the Associate Editor and an anonymous referee for their helpful and constructive comments, which lead to significant improvement in this article. This article is based on the first author’s Ph.D. dissertation under Li’s supervision.

Funding

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.

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 of Systematic Reviews, 10, CD004147. [745]

Beck, A. T. (2005), “The Current State of Cognitive Therapy: A 40-year Retrospective,” Archives of General Psychiatry, 62, 953–959. [745]

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. [736]

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,” Clinical Psychology Review, 26, 17–31. [745]
