

## RESEARCH ARTICLE

# Multivariate partial linear varying coefficients model for gene-environment interactions with multiple longitudinal traits

Honglang Wang<sup>1</sup> | Jingyi Zhang<sup>2,3</sup> | Kelly L. Klump<sup>4</sup> | Sybil Alexandra Burt<sup>4</sup> | Yuehua Cui<sup>2</sup> 

<sup>1</sup>Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA

<sup>2</sup>Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, USA

<sup>3</sup>Amazon Lab126, Sunnyvale, California, USA

<sup>4</sup>Department of Psychology, Michigan State University, East Lansing, Michigan, USA

## Correspondence

Yuehua Cui, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA.  
Email: [cuiy@msu.edu](mailto:cuiy@msu.edu)

## Funding information

National Institutes of Health, Grant/Award Numbers: R01GM131398, R01MH082054, R21HG010073

Correlated phenotypes often share common genetic determinants. Thus, a multi-trait analysis can potentially increase association power and help in understanding pleiotropic effect. When multiple traits are jointly measured over time, the correlation information between multivariate longitudinal responses can help to gain power in association analysis, and the longitudinal traits can provide insights on the dynamic gene effect over time. In this work, we propose a multivariate partially linear varying coefficients model to identify genetic variants with their effects potentially modified by environmental factors. We derive a testing framework to jointly test the association of genetic factors and illustrated with a bivariate phenotypic trait, while taking the time varying genetic effects into account. We extend the quadratic inference functions to deal with the longitudinal correlations and used penalized splines for the approximation of nonparametric coefficient functions. Theoretical results such as consistency and asymptotic normality of the estimates are established. The performance of the testing procedure is evaluated through Monte Carlo simulation studies. The utility of the method is demonstrated with a real data set from the Twin Study of Hormones and Behavior across the menstrual cycle project, in which single nucleotide polymorphisms associated with emotional eating behavior are identified.

## KEYWORDS

gene-environment interaction, longitudinal traits, multi-trait analysis, partial linear model, quadratic inference function

## 1 | INTRODUCTION

Cross-sectional disease traits have been the primary focus in genetic association studies. Given the improved power to identify disease genes with phenotypic data measured over time, longitudinal designs are becoming popular in genetic

Honglang Wang and Jingyi Zhang are contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

association studies.<sup>1-4</sup> Most statistical methods developed so far focus on a single outcome of interest. When multiple outcomes are measured over time, for example, multiple measures of heart function in a longitudinal study of cardiac function, methods focusing on just a single outcome over time may not provide a complete picture of cardiac function.

In genetics, the phenomenon that a single gene or locus influences more than one trait is known as pleiotropy.<sup>5,6</sup> Genetic pleiotropy plays a crucial role in many complex diseases. One of the most well-known examples is the phenylketonuria (PKU) disease.<sup>7</sup> The conventional approach to identify genetic pleiotropic effects on multiple traits is to test the association between a gene and each trait individually and then determine whether the genetic effect is significantly associated with more than one trait. The disadvantages of this approach, such as the inflation in the family wise Type I error and incomplete information in individual tests compared to a combined analysis for multiple traits, have been discussed in some studies.<sup>5</sup> Therefore, a joint genetic association test on multiple traits is more desirable to control the family wise Type I error and enhance the power of tests.

In real life, timing is a very important factor in the development of a disease. Genetic effects on a disease trait vary during the life span of an individual. The function of a gene depends largely on when it turns on and off, which could show a temporal pattern. In order to capture the dynamic effect of a gene on a disease trait over time, it is natural to model the dynamic effect as a potential (nonlinear) function over time. Considering multiple longitudinal traits, we proposed the following partially linear varying coefficients model,

$$y_{li}(t_{ij}) = \beta_{0l}(t_{ij}) + \beta_{1l}(t_{ij})G_i + \alpha_l^T \mathbf{Z}_{ij} + \epsilon_{li}(t_{ij}), \quad l = 1, \dots, L, \quad i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad (1)$$

where  $y_{lij} = y_{li}(t_{ij})$  is the response variable which measures the  $l$ th phenotype on the  $i$ th subject at the  $j$ th time point;  $\mathbf{Z}_{ij}$  is a  $p$ -dimensional vector of covariates, which can be either time dependent or independent;  $G_i$  denotes the time-invariant genetic variable within subject;  $\beta_{0l}(\cdot)$  and  $\beta_{1l}(\cdot)$  are unknown functions; and the stacked error vector  $\epsilon_i = (\epsilon_{1i}^T, \dots, \epsilon_{Li}^T)^T$  with  $\epsilon_{li} = (\epsilon_{li1}, \dots, \epsilon_{lin_i})^T$  is assumed to have mean zero and covariance  $\Sigma_i$ . Models for multivariate longitudinal traits are necessarily complex, because they must consider different types of correlations for each independent subject: correlation between measurements for the same trait at different time points, correlation between measurements at the same time point on different traits, and correlation between measurements at different time points and on different traits. With the stacked error vector  $\epsilon_i$ , its covariance matrix  $\Sigma_i$  carries all of these correlations.

If we use a time-varying environmental factor  $X_{ij}$  instead of  $t_{ij}$  in the model, that is,

$$y_{lij} = \beta_{0l}(X_{ij}) + \beta_{1l}(X_{ij})G_i + \alpha_l^T \mathbf{Z}_{ij} + \epsilon_{lij},$$

then the model can be used for jointly modeling nonlinear gene-environment (G×E) interactions for multiple longitudinal traits. In the model, one can assess the influence of  $X$  on  $G$  to affect multiple responses  $Y$ . Models for nonlinear G×E interactions have been studied.<sup>8,9</sup>

Qu and Li applied the method of quadratic inference functions (QIF) to the varying coefficients models for longitudinal data.<sup>10</sup> One important advantage is that the QIF method only requires correct specification of the mean structure and does not require any likelihood or approximation of the likelihood in hypothesis testing. In addition, when the working correlation structure is misspecified, QIF is more efficient than the generalized estimation equation (GEE) approach. Another advantage of the QIF approach is that the inference function has an asymptotic form, which provides a model selection criteria similar to AIC and BIC. It also allows us to test whether coefficients are significantly time-varying based on the asymptotic results.

Rochon analyzed bivariate longitudinal data for discrete and continuous outcomes by using generalized estimating equations, which did not utilize the nice property of the QIF.<sup>11</sup> Cho applied QIF for multivariate longitudinal data with generalized linear models, which is not adequate to consider nonlinear effects as in varying coefficient models.<sup>12</sup> Using random effects for modeling longitudinal data is another very popular and common way.<sup>13</sup> Proudfoot and his coauthors modeled the longitudinal data using random effects then combines multiple outcomes together again similar to the generalized estimating equations.<sup>14</sup> Recently, Zhao and his coauthors proposed a joint penalized quasi-likelihood modeling based on splines for multivariate longitudinal data using random effects with applications to HIV-1 RNA load levels and CD4 cell counts.<sup>15</sup> Hector and Song investigated a distributed quadratic inference function framework to jointly estimate regression parameters from multiple heterogeneous data sets with correlated responses.<sup>16</sup>

With the nice property of QIF to deal with complicated correlated data focusing on a univariate longitudinal response,<sup>10</sup> in this article, we consider to generalize it to partial linear varying coefficient models with multivariate longitudinal responses.<sup>17,18</sup> The purpose of this article is to develop a powerful joint testing procedure using QIF for model

(1). If the correlation between the longitudinal outcomes is reasonably high, we aim to show that the joint test has higher power than the marginal tests to detect the signal like the genetic effect in model (1). We first use splines to approximate the nonparametric functions in the model,<sup>19</sup> followed by penalized estimation to avoid over fitting. Then we develop a 2-step testing procedure to have a joint test for the interaction effect on multiple outcomes based on the QIF approach, followed by separate test of marginal effect on each outcome if the overall null is rejected. In cross-sectional studies, Wu et al.<sup>20</sup> developed a multivariate partially linear varying coefficient model to detect G×E interactions with multiple traits. Their method can select genetic variants with pleiotropic effects incorporating either the homogeneity (ie, pleiotropy) or heterogeneity (ie, no pleiotropy) assumptions. However, their approach cannot provide uncertainty quantification for the selected variables. Generalizing their method to multivariate longitudinal data is worth further studying.

This article is organized as follows. We state our proposed model in Section 2.1, and generalize the QIF method to the multivariate longitudinal responses in Section 2.2. Estimation procedure and asymptotical properties of estimators are provided in Section 2.3. A theorem for the general goodness-of-fit test via QIF is established in Section 2.4, based on which we propose a 2-step testing procedure. We assess the finite sample performance of the proposed procedure with Monte Carlo simulation in Section 3 and illustrate the proposed methodology by the analysis of an emotional eating behavior study in Section 4. Conclusions and discussion are made in Section 5. Proofs are rendered to Appendix.

## 2 | STATISTICAL METHODS

### 2.1 | A joint multivariate partial linear model

In multivariate longitudinal studies, suppose  $y_{lij}$  is the  $l$ th continuous outcome collected on the  $i$ th observation at time point  $t_{ij}$ , where  $l = 1, \dots, L, i = 1, \dots, N, j = 1, \dots, n_i$ . The joint partially linear varying coefficient models are defined as

$$y_{lij} = y_{li}(t_{ij}) = \beta_{0l}(t_{ij}) + \beta_{1l}(t_{ij})G_i + \alpha_l^T \mathbf{Z}_{ij} + \epsilon_{lij},$$

where  $G_i$  is the single nucleotide polymorphism (SNP) variable which does not depend on time and other types of measurement;  $\mathbf{Z}_{ij}$  is a  $p$ -dimensional covariate vector, which can be either time-dependent or time-independent; to accommodate the correlation between multiple responses, we stack the error terms  $\epsilon_{lij}$  together into a long vector

$$\epsilon_i = \begin{pmatrix} \epsilon_{1i} \\ \vdots \\ \epsilon_{Li} \end{pmatrix} \text{ where } \epsilon_{li} = \begin{pmatrix} \epsilon_{li1} \\ \vdots \\ \epsilon_{lin_i} \end{pmatrix}.$$

We assume  $\epsilon_i$  mean 0 with covariance  $\Sigma$ , which carries three different association information: the within-subject correlation across different time points, the between-subject correlation at the same time point and the between-subject correlation across different time points;  $\beta_{0l}(\cdot)$  and  $\beta_{1l}(\cdot)$  are unknown nonparametric smooth functions, representing the main time effect and time dependent genetic effect respectively. To illustrate the idea, in the following we demonstrate the methods assuming  $L=2$ . For the situation where there are more than two traits ( $L > 2$ ), the technique can be easily extended.

### 2.2 | Quadratic inference function

To construct the objective function using the QIF approach, we first approximate the unknown functions  $\beta_{01}, \beta_{11}, \beta_{02},$  and  $\beta_{12}$  by a  $q$ -degree truncated power spline basis, that is,

$$\beta_{sl}(t) \approx \mathbf{B}_{sl}(t)^T \boldsymbol{\gamma}_{sl}, \text{ for } s = 0, 1 \text{ and } l = 1, 2, \quad (2)$$

where  $\mathbf{B}_{sl}(t) = (1, t, t^2, \dots, t^{q_{sl}}, (t - \kappa_{sl,1})_+^{q_{sl}}, \dots, (t - \kappa_{sl,K_{sl}})_+^{q_{sl}})^T$  is a truncated power spline basis with degree  $q_{sl}$  and  $K_{sl}$  knots  $\kappa_{sl,1}, \dots, \kappa_{sl,K_{sl}}$ .  $\boldsymbol{\gamma}_{sl}$  is a  $(q_{sl} + K_{sl} + 1)$ -dimensional vector of spline coefficients.

Under the GEE framework, we solve

$$\sum_{i=1}^N \dot{\boldsymbol{\mu}}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0, \quad (3)$$

where  $\mathbf{y}_i = (\mathbf{y}_{1i}^T, \mathbf{y}_{2i}^T)^T$ ,  $\mathbf{y}_{li} = (y_{li1}, \dots, y_{lin_i})^T$ ;  $\boldsymbol{\mu}_i = E(\mathbf{y}_i)$  is the mean function and  $\dot{\boldsymbol{\mu}}_i$  is the first derivative of  $\boldsymbol{\mu}_i$  with respect to the parameters;  $\mathbf{V}_i$  is the covariance matrix of  $\mathbf{y}_i$  and can be decomposed as  $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\rho}) \mathbf{A}_i^{1/2}$  with  $\mathbf{A}_i$  being a diagonal matrix of marginal variances and  $\mathbf{R}(\boldsymbol{\rho})$  being a working correlation matrix with nuisance parameters  $\boldsymbol{\rho}$ . To avoid the estimation of  $\boldsymbol{\rho}$ , QIF approach considers the inverse of the correlation matrix  $\mathbf{R}$  as a linear combination of several known basis matrices in a form

$$\mathbf{R}^{-1} \approx a_1 \mathbf{M}_1 + a_2 \mathbf{M}_2 + \dots + a_h \mathbf{M}_h, \quad (4)$$

where  $\mathbf{M}_1$  is the identity matrix and  $\mathbf{M}_2, \dots, \mathbf{M}_h$  are symmetric basis matrices. As discussed in the existing literature,<sup>10,12</sup> the choice of the basis for the inverse of the correlation matrix plays an important role. Suppose  $\boldsymbol{\Gamma}$  as the within-subject correlation structure and  $\omega$  as between-subject correlation coefficient, that is, the working correlation structure can be expressed as the Kronecker product (tensor product)  $\mathbf{R} = \boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}$  with  $\boldsymbol{\Omega}$  as the  $2 \times 2$  symmetric matrix with 1 on the diagonal and  $\omega$  elsewhere. The inverse of the Kronecker product is  $\mathbf{R}^{-1} = \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}^{-1} = (\gamma_0 \mathbf{I} + \gamma_1 \mathbf{W}) \otimes \boldsymbol{\Gamma}^{-1}$  with  $\mathbf{W}$  as a  $2 \times 2$  symmetric matrix with 0 on the diagonal and 1 elsewhere, and  $\mathbf{I}$  as the identity matrix with compatible dimension. So if the basis matrix for the inverse of the within-subject correlation  $\boldsymbol{\Gamma}^{-1}$  is given by  $\mathbf{U}_1 (= \mathbf{I}), \mathbf{U}_2, \dots, \mathbf{U}_k$ , then we have the bases  $\mathbf{M}'_s$  as  $\{\mathbf{I} \otimes \mathbf{U}_j, \mathbf{W} \otimes \mathbf{U}_j : 1 \leq j \leq k\}$ . For exchangeable working correlation, we can set  $k = 2$  and  $\mathbf{U}_2$  has 0 on the diagonal and 1 elsewhere. If the working correlation is AR(1), we can set  $k = 2$  and  $\mathbf{U}_2$  to have 1 on its two subdiagonals and 0 elsewhere. Following QIF approach, we define the estimation function as

$$\bar{\mathbf{g}}_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \mathbf{g}_i(\boldsymbol{\theta}) = \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N \dot{\boldsymbol{\mu}}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_1 \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \sum_{i=1}^N \dot{\boldsymbol{\mu}}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_h \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \end{bmatrix}. \quad (5)$$

Using the spline approximation, the mean function  $\boldsymbol{\mu}_i$  can be written as

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mu}_{1i}(\boldsymbol{\theta}) \\ \boldsymbol{\mu}_{2i}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{1i1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{\mu}_{1in_i}(\boldsymbol{\theta}) \\ \boldsymbol{\mu}_{2i1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{\mu}_{2in_i}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{01}^T(t_{i1})\boldsymbol{\gamma}_{01} + \mathbf{B}_{11}^T(t_{i1})\boldsymbol{\gamma}_{11}G_i + \boldsymbol{\alpha}_1^T \mathbf{Z}_{i1} \\ \vdots \\ \mathbf{B}_{01}^T(t_{in_i})\boldsymbol{\gamma}_{01} + \mathbf{B}_{11}^T(t_{in_i})\boldsymbol{\gamma}_{11}G_i + \boldsymbol{\alpha}_1^T \mathbf{Z}_{in_i} \\ \mathbf{B}_{02}^T(t_{i1})\boldsymbol{\gamma}_{02} + \mathbf{B}_{12}^T(t_{i1})\boldsymbol{\gamma}_{12}G_i + \boldsymbol{\alpha}_2^T \mathbf{Z}_{i1} \\ \vdots \\ \mathbf{B}_{02}^T(t_{in_i})\boldsymbol{\gamma}_{02} + \mathbf{B}_{12}^T(t_{in_i})\boldsymbol{\gamma}_{12}G_i + \boldsymbol{\alpha}_2^T \mathbf{Z}_{in_i} \end{bmatrix},$$

and the first derivative of  $\boldsymbol{\mu}_i$  is given as,

$$\dot{\boldsymbol{\mu}}_i = \begin{bmatrix} \mathbf{B}_{01}^T(t_{i1}) & \mathbf{B}_{11}^T(t_{i1})G_i & \mathbf{Z}_{i1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{B}_{01}^T(t_{in_i}) & \mathbf{B}_{11}^T(t_{in_i})G_i & \mathbf{Z}_{in_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{B}_{02}^T(t_{i1}) & \mathbf{B}_{12}^T(t_{i1})G_i & \mathbf{Z}_{i1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{B}_{02}^T(t_{in_i}) & \mathbf{B}_{12}^T(t_{in_i})G_i & \mathbf{Z}_{in_i} \end{bmatrix},$$

where  $\boldsymbol{\theta} = (\boldsymbol{\gamma}_{01}^T, \boldsymbol{\gamma}_{11}^T, \boldsymbol{\alpha}_1^T, \boldsymbol{\gamma}_{02}^T, \boldsymbol{\gamma}_{12}^T, \boldsymbol{\alpha}_2^T)^T$ .

Setting each component in (5) to be zero will result in more equations than unknown parameters. Following the idea of generalized method of moments,<sup>21</sup> the QIF method is defined as

$$Q_N(\boldsymbol{\theta}) = N \bar{\mathbf{g}}_N^T \bar{\mathbf{C}}_N^{-1} \bar{\mathbf{g}}_N, \quad (6)$$

where  $\bar{\mathbf{C}}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{g}_i \mathbf{g}_i^T$  is a consistent estimator for  $\text{var}(\mathbf{g}_i)$ . Minimizing the objective function (6) provides the estimation of the parameters.

### 2.3 | Estimation procedure via penalized QIF

The estimation of the parameters can be obtained through minimizing the objective function, that is,

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta}).$$

To avoid over-fitting, we can define a penalized QIF in a form

$$N^{-1} Q_N(\boldsymbol{\theta}) + \lambda \boldsymbol{\theta}^T \mathbf{D} \boldsymbol{\theta}, \quad (7)$$

where  $\mathbf{D}$  is a diagonal matrix with 1 if the corresponding parameter is the spline coefficient associated with knots, and 0 otherwise. Minimizing the penalized QIF provides

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} (N^{-1} Q_N(\boldsymbol{\theta}) + \lambda \boldsymbol{\theta}^T \mathbf{D} \boldsymbol{\theta}). \quad (8)$$

To estimate the tuning parameter  $\lambda$ , we can extend the generalized cross-validation<sup>10,22,23</sup> to the penalized QIF and define the generalized cross-validation statistic as

$$\text{GCV}(\lambda) = \frac{N^{-1} Q_N}{(1 - N^{-1} df)^2}$$

with the effective degree of freedom

$$df = \text{tr} [(\ddot{Q}_N + 2N\lambda D)^{-1} \ddot{Q}_N],$$

where  $\ddot{Q}_N$  is the second derivative of  $Q_N$ . The optimized tuning parameter  $\lambda$  is given as

$$\hat{\lambda} = \arg \min_{\lambda} \text{GCV}(\lambda).$$

To establish the asymptotic properties for the penalized quadratic inference function estimators with fixed knots, we assume  $\boldsymbol{\theta}_0$  to be the parameter satisfying  $E_{\boldsymbol{\theta}_0}(\mathbf{g}_i) = 0$ . Similar theoretical results are provided in Qu and Li.<sup>10</sup> Following their idea and extend those results to the estimators in our model, we get the strong consistency of the resulting estimators in Theorem 1. The  $\sqrt{N}$ -consistency and asymptotic normality of the estimators are given in Theorem 2.

**Theorem 1.** Suppose conditions (A1)-(A6) in the Appendix hold and the smoothing parameter  $\lambda_N = o(1)$ , then the estimator  $\hat{\boldsymbol{\theta}}$ , which is obtained by minimizing the penalized quadratic function in (7), exists and converges to  $\boldsymbol{\theta}_0$  almost surely.

**Theorem 2.** Suppose conditions (A1)-(A6) in the Appendix hold and the smoothing parameter  $\lambda_N = o(N^{-1/2})$ , then the estimator  $\hat{\boldsymbol{\theta}}$  obtained by minimizing the penalized quadratic function in (7) is asymptotically normally distributed with the limiting distribution,

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbf{G}_0^T \mathbf{C}_0^{-1} \mathbf{G}_0)^{-1}),$$

where the calculation of  $\mathbf{G}_0$  defined in (A6) and  $\mathbf{C}_0$  defined in (A5) can be found in the Appendix.

## 2.4 | A two-step hypothesis testing procedure

Compared to GEE, an advantage of the QIF approach is that QIF provides a goodness-of-fit test without estimating the second moment parameters. Suppose that the  $d$ -dimensional parameter vector  $\boldsymbol{\gamma}$  is partitioned into  $(\boldsymbol{\psi}, \boldsymbol{\zeta})$ , where  $\boldsymbol{\psi}$  is the parameter of interest with dimension  $d_1$ , and  $\boldsymbol{\zeta}$  is a nuisance parameter with dimension  $d_2 = d - d_1$ . If we are interested in testing

$$H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0,$$

then the test statistic

$$Q_N(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) - Q_N(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}),$$

follows an asymptotically chi-square distribution with  $d_1$  degrees of freedom as from Qu and her coauthors work cited below.<sup>24</sup>

**Theorem 3.** *Suppose that all required regularity conditions are satisfied and  $\boldsymbol{\psi}$  has dimension  $d_1$ . Under the null hypothesis,  $Q_N(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) - Q_N(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}})$  is asymptotically chi-square distributed with  $d_1$  degrees of freedom, where*

$$\tilde{\boldsymbol{\zeta}} = \arg \min Q_N(\boldsymbol{\psi}_0, \boldsymbol{\zeta}), \quad (\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) = \arg \min Q_N(\boldsymbol{\psi}, \boldsymbol{\zeta}). \quad (9)$$

In Model (1), it is of interest to test whether the genetic effects on multiple traits are significant or not. Based on Theorem 3, we develop a 2-step testing procedure for testing the significance of the varying coefficient functions. In the first step, the joint test is performed to see whether a genetic factor has a significant effect on at least one longitudinal trait. If the testing result in the first step is significant, we then further conduct the marginal test in the second step to assess if the genetic effect is significant on both traits or just one trait. The first step is a joint test of significance followed by a marginal test to assess individual significance. For associated multiple traits with reasonably strong correlation, the joint test is more powerful than the marginal tests, which is empirically verified in our simulation studies.

### 2.4.1 | Step 1: Joint test

First, we are interested in testing whether the genetic factor  $G$  has an effect on at least one longitudinal trait. The hypothesis is stated as

$$H_0 : \beta_{11}(\cdot) = \beta_{12}(\cdot) = 0 \text{ v.s. } H_1 : \beta_{11}(\cdot) \neq 0 \text{ or } \beta_{12}(\cdot) \neq 0.$$

This can be handled through the truncated power spline approximation of the nonparametric functions stated in (2). In particular, testing this hypothesis is equivalent to test the following null hypothesis

$$H_0 : \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \mathbf{0}.$$

According to Theorem 3, we can construct a test statistic

$$T_N = Q_N(\tilde{\boldsymbol{\theta}}) - Q_N(\hat{\boldsymbol{\theta}}),$$

where

$$\tilde{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \mathbf{0}} Q_N \{ \boldsymbol{\gamma}_{01}, \boldsymbol{\gamma}_{11}, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_{02}, \boldsymbol{\gamma}_{12}, \boldsymbol{\alpha}_2 | \mathbf{y}_1, \mathbf{y}_2, \mathbf{G}, \mathbf{Z} \},$$

and

$$\hat{\boldsymbol{\theta}} = \arg \min Q_N \{ \boldsymbol{\gamma}_{01}, \boldsymbol{\gamma}_{11}, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_{02}, \boldsymbol{\gamma}_{12}, \boldsymbol{\alpha}_2 | \mathbf{y}_1, \mathbf{y}_2, \mathbf{G}, \mathbf{Z} \}.$$

The test statistic  $T_N$  has an asymptotic  $\chi^2$  distribution with the degrees of freedom equal the number of constraints under  $H_0$ , according to Theorem 3.

## 2.4.2 | Step 2: Marginal tests

From the joint test, if there exists a significant genetic effect on at least one longitudinal trait, then we can further test the marginal effects, that is,

$$H_{0l} : \beta_{1l}(\cdot) = 0 \text{ v.s. } H_1 : \beta_{1l}(\cdot) \neq 0, \quad l = 1, 2.$$

Based on (2), this is equivalent to test  $H_{01} : \gamma_{11} = \mathbf{0}$  and  $H_{02} : \gamma_{12} = \mathbf{0}$ , separately.

For testing  $H_{01} : \gamma_{11} = \mathbf{0}$ , we use test statistic  $T_{N1} = Q_N(\tilde{\gamma}_{01}, \mathbf{0}, \tilde{\alpha}_1) - Q_N(\hat{\gamma}_{01}, \hat{\gamma}_{11}, \hat{\alpha}_1)$ , where

$$(\tilde{\gamma}_{01}, \mathbf{0}, \tilde{\alpha}_1) = \arg \min_{\gamma_{11}=\mathbf{0}} Q_N \{ \gamma_{01}, \gamma_{11}, \alpha_1 | \mathbf{y}_1, \mathbf{G}, \mathbf{Z} \},$$

and

$$(\hat{\gamma}_{01}, \hat{\gamma}_{11}, \hat{\alpha}_1) = \arg \min Q_N \{ \gamma_{01}, \gamma_{11}, \alpha_1 | \mathbf{y}_1, \mathbf{G}, \mathbf{Z} \}.$$

Similarly, we can construct a test statistic  $T_{N2} = Q_N(\tilde{\gamma}_{02}, \mathbf{0}, \tilde{\alpha}_2) - Q_N(\hat{\gamma}_{02}, \hat{\gamma}_{12}, \hat{\alpha}_2)$  for testing  $H_{02} : \gamma_{12} = \mathbf{0}$ , where

$$(\tilde{\gamma}_{02}, \mathbf{0}, \tilde{\alpha}_2) = \arg \min_{\gamma_{12}=\mathbf{0}} Q_N \{ \gamma_{02}, \gamma_{12}, \alpha_2 | \mathbf{y}_2, \mathbf{G}, \mathbf{Z} \},$$

and

$$(\hat{\gamma}_{02}, \hat{\gamma}_{12}, \hat{\alpha}_2) = \arg \min Q_N \{ \gamma_{02}, \gamma_{12}, \alpha_2 | \mathbf{y}_2, \mathbf{G}, \mathbf{Z} \}.$$

The asymptotic distribution of the test statistics  $T_{N1}$  and  $T_{N2}$  can be obtained from Theorem 3.

## 3 | SIMULATION STUDIES

### 3.1 | Simulation setup

In this section, the finite sample performance of the proposed method is evaluated through Monte Carlo simulation studies. Two continuous longitudinal responses are generated from the models

$$\begin{cases} y_{1ij} = y_{1i}(t_{ij}) = \beta_{01}(t_{ij}) + \beta_{11}(t_{ij})G_i + \alpha_1 Z_i + \epsilon_{1ij}, \\ y_{2ij} = y_{2i}(t_{ij}) = \beta_{02}(t_{ij}) + \beta_{12}(t_{ij})G_i + \alpha_2 Z_i + \epsilon_{2ij}, \end{cases}$$

where  $\beta_{01}(t_{ij}) = 0.5 \cos(2\pi t_{ij})$ ,  $\beta_{11} = \sin(\pi(t_{ij} - 0.2))$ ,  $\beta_{02}(t_{ij}) = \sin(\pi t_{ij}) - 0.5$ ,  $\beta_{12}(t_{ij}) = \cos(\pi t_{ij} - 0.8)$ ,  $\alpha_1 = 0.2$  and  $\alpha_2 = 0.3$ . We generate the same number of time points  $n$  for each individual  $\mathbf{t}_i = (t_{i1}, \dots, t_{in})$  from a uniform distribution  $U(0, 1)$ . The time independent predictor variable  $Z_i$  is also generated from  $U(0, 1)$ . We set the minor allele frequency (MAF) for  $G_i$  as  $p_A$  and assume Hardy-Weinberg equilibrium. Three different SNP genotypes  $AA$ ,  $Aa$ , and  $aa$  are simulated from a multinomial distribution with frequencies  $p_A^2$ ,  $2p_A(1 - p_A)$  and  $(1 - p_A)^2$ , respectively. In this simulation study, we vary  $p_A \in \{0.1, 0.3, 0.5\}$  to investigate the effect of minor allele frequency. Variable  $G$  takes value  $\{0, 1, 2\}$  corresponding to genotypes  $\{aa, Aa, AA\}$ , following an additive model. We assume  $\epsilon_{1ij}$  and  $\epsilon_{2ij}$  are jointly normally distributed as

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 \Sigma_{11} & \sigma_1 \sigma_2 \Sigma_{12} \\ \sigma_1 \sigma_2 \Sigma_{12} & \sigma_2^2 \Sigma_{22} \end{pmatrix} \right].$$



We set the marginal variances  $\sigma_1^2 = \sigma_2^2 = 0.1$ . The true correlation structure of  $\Sigma_{11}$  and  $\Sigma_{22}$  are both exchangeable with the structure

$$\begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}$$

with  $\rho_1 = \rho_2 = 0.5$ . And for  $\Sigma_{12}$  we choose

$$\begin{pmatrix} \tau & \rho_{12} & \rho_{12} & \cdots & \rho_{12} \\ \rho_{12} & \tau & \rho_{12} & \cdots & \rho_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{12} & \rho_{12} & \rho_{12} & \cdots & \tau \end{pmatrix}$$

with  $\rho_{12} = 0.2$  as the between-subject correlation across different time points. We vary the between-subject correlation at the same time point  $\tau = \text{corr}(\epsilon_{1ij}, \epsilon_{2ij})$  to investigate the power gain for the joint test.

We draw 1000 data sets with sample size  $N = 200,500$  and time points  $n_i = n = 10$ , in order to compare the performances of our proposed method under different sample sizes. We set  $\mathbf{M}_1$  to be the identity matrix and  $\mathbf{M}_2$  to be 1 on subdiagonals and 0 elsewhere, that is, AR(1) working correlation. An important issue for the model selection is to decide whether the spline model (2) is adequate for further penalization by (8). In the following simulations, we use quartic splines with the number of knots taken to be the largest integer not greater than  $0.6 \times N^{1/5}$  as suggested in Tian and his colleagues work.<sup>25</sup>

### 3.2 | Estimation performance

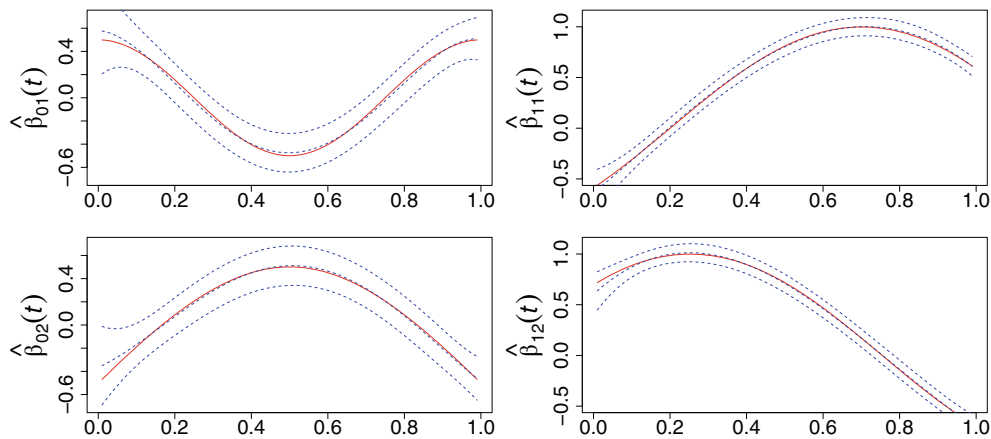
We use the asymptotic normality in Theorem 2 to construct the Wald type confidence interval for parameters  $\alpha_1$  and  $\alpha_2$ . Table 1 summarizes the empirical coverage probability (CP) in percentage and the average length (AL) of the confidence intervals at 95% confidence level based on 1000 simulation replicates. As we can see from the table, the CPs are close to the nominal level 95%. When the sample size gets larger, the ALs are shorter and the CPs are closer to 95%.

Next, we consider the estimation performance of the nonparametric functions  $\beta_{sl}(t)$  for  $s = 0, 1, l = 1, 2$ . In Figure 1, the plots are from the case with sample size  $N = 200$  and  $p_A = 0.1$  (for other situations, please refer to the Appendix). For each function, the red solid line is the true function, and the three blue dashed lines correspond to the average of the estimated functions from 1000 simulation replicates in the middle and the 95% pointwise confidence bands with the standard error calculated from the standard deviation of 1000 replicates. The estimation is quite accurate with low sample size and MAF. As the MAF or sample increases, the estimation performance improves (see Appendix). We also use the asymptotic normality in Theorem 2 to construct the pointwise confidence intervals. Table 2 summarized the empirical coverage probability (CP) in percentage and the average length (AL) (in parentheses) of the CIs for  $\beta_{sl}(t)$  at  $t = 0.2, 0.4, 0.6$ , and  $0.8$  for sample size  $N = 500$  and  $N = 1000$ . The CPs are all close to the nominal level 95% and the ALs are shorter under a larger sample size.

**TABLE 1** Empirical coverage probability (%) and average length of confidence intervals for  $\alpha_l, l = 1, 2$

	$N = 200$		$N = 500$	
	CP	AL	CP	AL
$\alpha_1$	93.2	0.078	94.8	0.050
$\alpha_2$	93.7	0.078	95.6	0.050





**FIGURE 1** The estimation of nonparametric functions  $\beta_{sl}(\cdot)$  for  $s = 0, 1, l = 1, 2$  with  $N = 200$  and  $p_A = 0.1$ . In each panel, the red solid line is the true function, and the three blue dashed lines correspond to the estimated function in the middle and the 95% pointwise confidence bands

**TABLE 2** Empirical coverage probability (%) and average length of pointwise confidence intervals (in parentheses) for  $\beta_{sl}(t)$  at  $t = 0.2, 0.4, 0.6$ , and  $0.8$

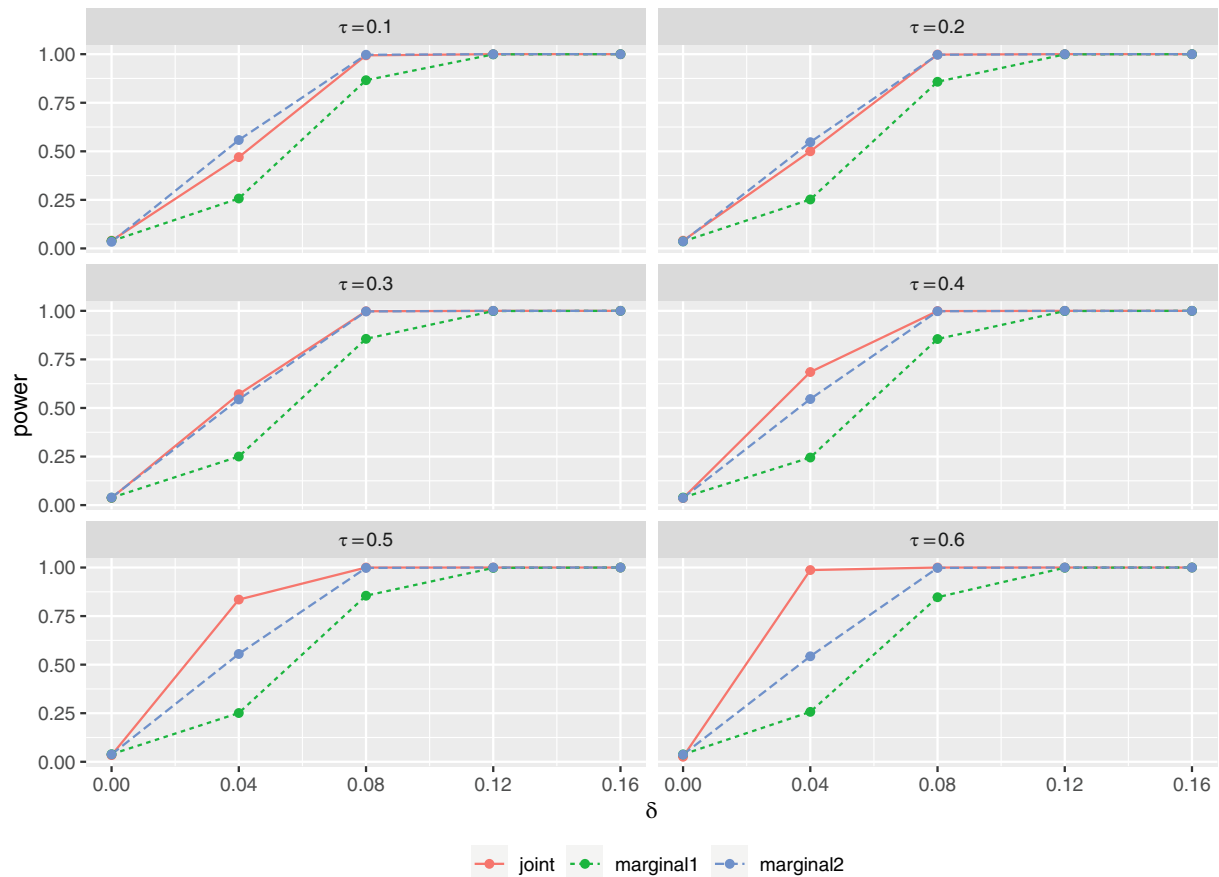
$t$	$N$	Intercept $\beta_{01}(t)$	Slope $\beta_{11}(t)$	Intercept $\beta_{02}(t)$	Slope $\beta_{12}(t)$
0.2	500	92.2 (0.115)	92.6 (0.073)	92.9 (0.115)	92.7 (0.073)
	1000	91.7 (0.082)	93.3 (0.052)	92.9 (0.082)	92.9 (0.052)
0.4	500	90.5 (0.109)	91.9 (0.069)	92.4 (0.108)	92.8 (0.068)
	1000	92.3 (0.077)	93.7 (0.049)	94.4 (0.077)	93.1 (0.049)
0.6	500	91.0 (0.111)	91.0 (0.070)	92.3 (0.110)	93.1 (0.070)
	1000	94.0 (0.079)	93.8 (0.050)	95.1 (0.079)	94.4 (0.050)
0.8	500	91.5 (0.113)	93.1 (0.072)	92.9 (0.113)	94.1 (0.071)
	1000	91.5 (0.081)	94.2 (0.051)	94.8 (0.080)	94.6 (0.051)

### 3.3 | Testing performance

We propose a two-step hypothesis testing procedure to detect the genetic effects on multiple traits. With the joint test, higher power is expected with correlated traits than the marginal tests. We would like to evaluate how much we can gain in power when the correlation between multiple traits increases. This is done by varying the correlation coefficient  $\tau$  at the same time point for the two simulated traits.

We evaluate the performance of the joint test under the null hypothesis  $H_0 : \beta_{11}(\cdot) = \beta_{12}(\cdot) = 0$ . Power is evaluated under a sequence of alternative models with different values of  $\delta$ , which is denoted by  $H_1 : \beta_{11}(\cdot) = \delta\beta_{11}(\cdot)$  and  $\beta_{12}(\cdot) = \delta\beta_{12}(\cdot)$ . The performance of the marginal tests for the nonparametric functions corresponding to different traits is evaluated under the two null hypotheses  $H_{01} : \beta_{11}(\cdot) = 0$  and  $H_{02} : \beta_{12}(\cdot) = 0$  respectively. For each test, power is evaluated under a sequence of alternative models, denoted by  $H_{al} : \beta_{1l}(\cdot) = \delta\beta_{1l}(\cdot)$ ,  $l = 1, 2$ , correspondingly.

Figure 2 shows the power comparison between joint test and marginal tests under different correlation coefficient  $\tau$  varying from 0.1 through 0.6. Each panel corresponds to the results with one  $\tau$  value and displays the comparison of the three power curves, with the empirical size (when  $\delta = 0$ ) and power at different  $\delta (> 0)$  at the significance level 0.05 and sample size  $N = 200$ . Similar pattern can be observed for larger sample size  $N = 500$ . As expected, the Type I error is closer to 0.05 and the power increase as the signal  $\delta$  increases for every power curve. When  $\tau$  is small (low correlation), we do not see much power gain of the joint test compared to the marginal tests. As  $\tau$  increases (correlation between the two traits increases), we observe higher power of the joint test (starting at  $\tau = 0.4$ ). This shows that the joint test is more powerful than the marginal tests for moderate or high correlation between traits. We also conducted simulations to



**FIGURE 2** The power comparison between the joint test and marginal tests under different correlation coefficient  $\tau$  from 0.1 to 0.6 with sample size  $N = 200$ . The exact empirical sizes are given in Table 3

**TABLE 3** Empirical size for the joint and marginal tests under different correlation coefficient  $\tau$  from 0.1 to 0.6 with sample size  $N = 200$

	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$
Joint	0.040	0.040	0.037	0.036	0.034	0.026
Marginal 1	0.038	0.036	0.038	0.039	0.039	0.038
Marginal 2	0.034	0.036	0.039	0.037	0.038	0.037

evaluate the impact of between-subject correlation across different time points on the testing power. We observed similar results as the one by varying the between-subject correlation at the same time point. Due to space limit, the results were rendered in the supplemental file.

MAF also plays a major role for the inference performance of an association test in general. For the proposed method, the power increases as the MAF  $p_A$  increases from 0.1 to 0.5. This is in align with the general conception. In particular, there is a big power improvement as  $p_A$  increases from 0.1 to 0.3 as shown in Figure 3.

## 4 | REAL DATA APPLICATION

We applied the proposed multivariate partially linear varying coefficients model and the two-step hypothesis testing procedure to the Twin Study of Hormones and Behavior across the Menstrual Cycle project<sup>26</sup> from the Michigan State University Twin Registry (MSUTR).<sup>27-29</sup> The goal of the study was to examine associations between changes in estradiol and progesterone levels and emotional eating across the menstrual cycle. Emotional eating was measured with the Dutch

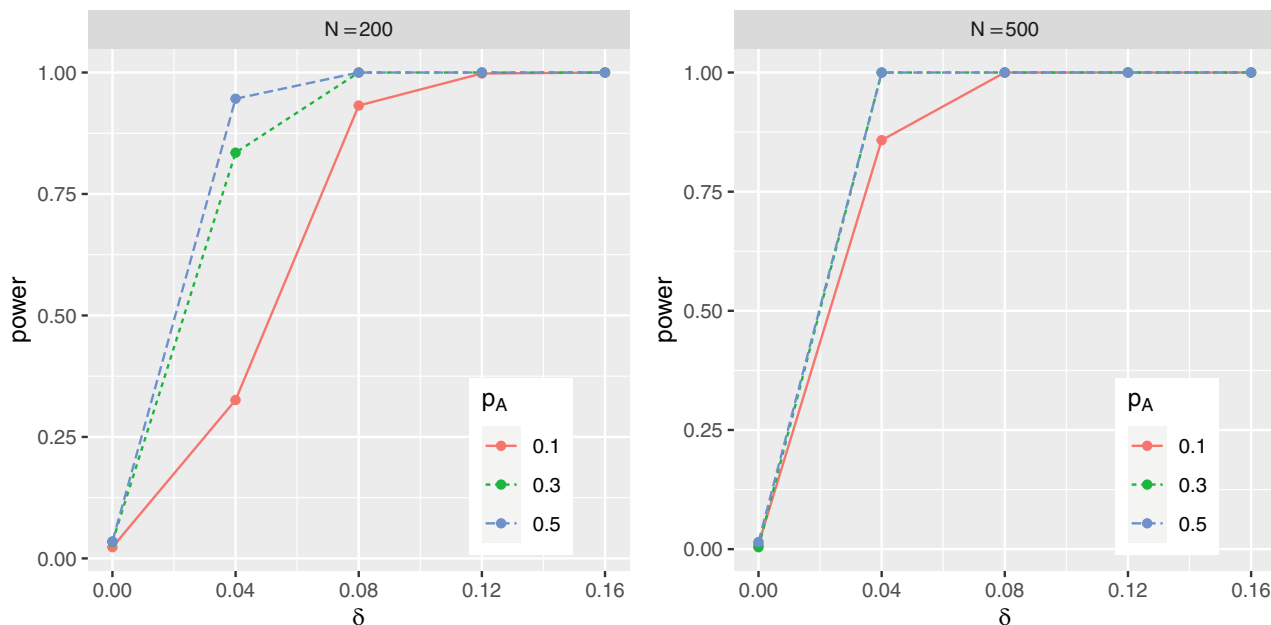


FIGURE 3 The power comparison of the joint test under different minor allele frequencies ( $P_A = 0.1, 0.3, 0.5$ ) and different sample sizes ( $N = 200, 500$ )

Eating Behavior Questionnaire (DEBQ) and negative affect was measured with the Negative Affect scale from the Positive and Negative Affect Schedule (PANAS). The DEBQ assesses the tendency to eat in response to negative emotions while PANAS is used to measure negative emotional states like sadness and anxiety.

In this study, we wanted to examine how genes respond to the hormone change (eg, estrogen) to affect emotional eating measured by DEBQ and PANAS. Since body mass index (BMI) is an important covariate for the study, we included it in the linear component of the model. Although the original study contains twins data, we only included one of the twins in each family in the analysis to make the samples independent. Measurements for each participant were collected for 45 consecutive days, which then were grouped into eight menstrual cycle phases, that is, ovulatory phase (1), transition ovulatory to midluteal (2), midluteal phase (3), transition midluteal to premenstrual (4), premenstrual phase including the first day of menstrual cycle (5), remaining days of menstrual cycle, part of follicular phase (6), follicular phase (7) and transition follicular to ovulatory phase (8). They were grouped into these phases based on profiles of changes in estrogen and progesterone across the cycle.<sup>30</sup> Data that belong to the same phase were averaged to get a phase-level measure. All individuals were aligned according to the 8 phases for further analysis.

To demonstrate the utility of the method, here we focused on a candidate gene, nuclear receptor coactivator 7 (NCoA7). This gene codes for an estrogen receptor-associated protein which plays an important role in the cellular response to estrogen. After removing SNPs with MAF < 0.05, we had 12 SNPs measured on 327 participants for further analysis.

We consider the partially linear varying coefficient model with the two longitudinal traits, namely DEBQ and PANAS, with the form

$$y_{ij}^D = \beta_0^D(X_{ij}) + \beta_1^D(X_{ij})G_i + \alpha^D Z_{ij} + \epsilon_{ij}^D,$$

$$y_{ij}^P = \beta_0^P(X_{ij}) + \beta_1^P(X_{ij})G_i + \alpha^P Z_{ij} + \epsilon_{ij}^P.$$

For the  $i$ th individual measured at menstrual cycle phase  $j$ , the two longitudinal traits are denoted as  $y_{ij}^D$  and  $y_{ij}^P$  for DEBQ and PANAS, respectively. One phase dependent covariate BMI is denoted as  $Z_{ij}$ .  $X_{ij}$  refers to the hormone estradiol level, which is standardized to range between 0 and 1 by  $\Phi((E - \bar{X}_E)/S_E)$  where  $E$  is the original estrogen level,  $\bar{X}_E$  and  $S_E$  are the sample mean and standard deviation of  $E$ , and  $\Phi$  is the cumulative distribution function of a standard normal.  $G_i$  represents the SNP variable and the 12 SNPs were analyzed separately.

We aimed to test if an SNP is associated with the two traits with its effect modified by the estrogen hormone level, that is,

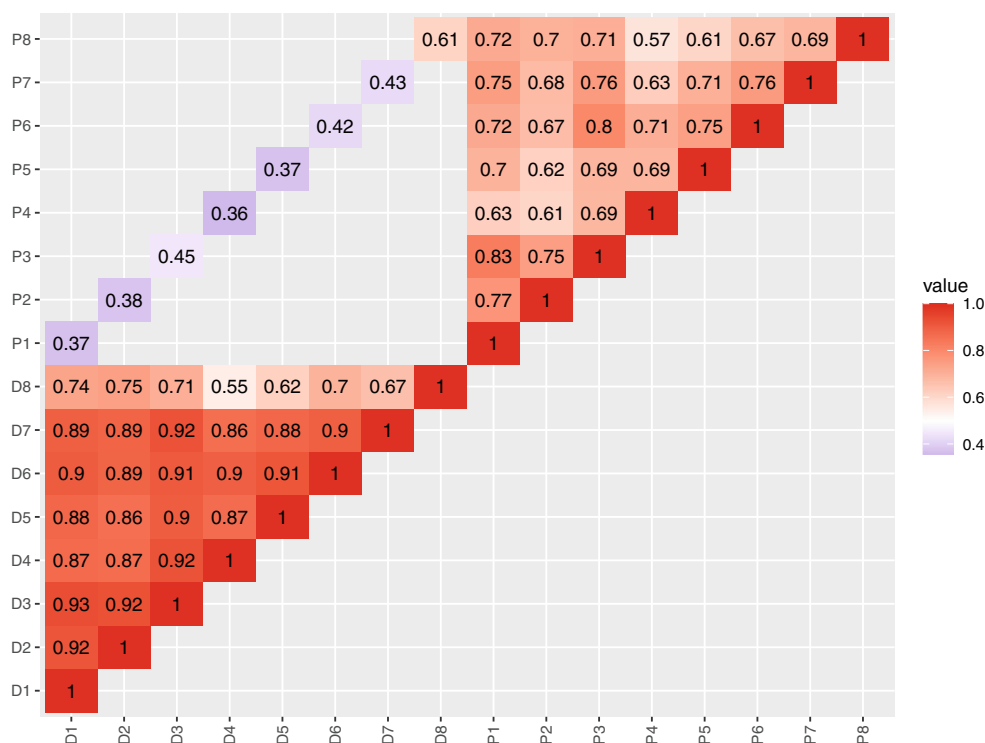
$$H_0 : \beta_1^D(\cdot) = \beta_1^P(\cdot) = 0 \quad v.s. \quad H_1 : \beta_1^D(\cdot) \neq 0 \text{ or } \beta_1^P(\cdot) \neq 0.$$

We applied the quadratic splines and the exchangeable working correlation structure for this real data analysis. After the Bonferroni correction for the 12 SNPs, we found three SNPs, rs584032, rs6911452, and rs9401855, are significant with the joint test. Table 4 lists the results. The joint test results are all more significant than the marginal tests. The between-trait correlations between DEBQ and PANAS at the same cycle phase are shown in Figure 4, which shows a quite strong correlation at different phases ranging from 0.36 to 0.61. This explains why the joint test shows stronger significance than the marginal tests. Figure 4 shows the detailed correlation information about the three components: within-trait correlation, between-trait correlation at the same time points and across different time points.

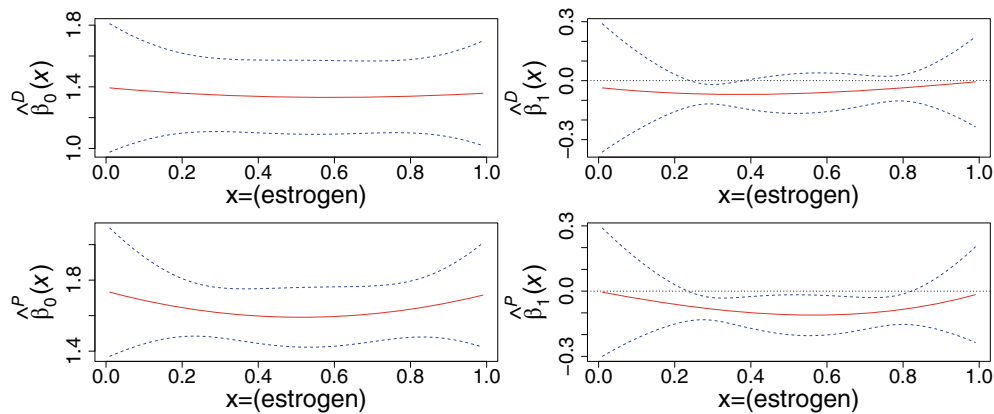
Figure 5 shows the estimated nonparametric coefficient functions for both responses DEBQ and PANAS, with SNP rs9401855 as an example. The point-wise 95% confidence bands cover a large part of the zero line for the DEBQ, which is consistent with its *P*-value .0545 from the marginal test. And for PANAS, in the central region, the zero line is outside of the 95% point-wise confidence bands, which is also consistent with the marginal *P*-value .0011. The result shows that this SNP interacts with estrogen hormone and only affects PANAS, but not DEBQ. The negative coefficients show that

**TABLE 4** The test results of the 3 significant SNPs with their rs numbers, the alleles (minor allele shows with bold font), the MAF, and the *P*-values for the joint test (denoted as  $P_{joint}$ ) and the two marginal tests (denoted as  $P_{DEBQ}$  and  $P_{PANAS}$ )

SNP	Alleles	MAF	$P_{joint}$	$P_{DEBQ}$	$P_{PANAS}$
rs584032	T/A	0.176	2.390e-4	3.472e-2	1.937e-3
rs6911452	A/G	0.089	8.225e-4	3.491e-3	4.627e-3
rs9401855	A/G	0.129	3.758e-4	5.449e-2	1.108e-3



**FIGURE 4** The correlation information including within-trait correlation, between-trait correlation at the same and across different cycle phases. The x-axis and y-axis represent the 8 cycle phases for the two variables DEBQ and PANAS, respectively



**FIGURE 5** The estimated intercept and slope functions for DEBQ and PANAS from the joint model (red solid curve) and their point-wise 95% confidence bands (dashed curve)

estrogen hormone negatively impacts PANAS. Individuals carrying the GG genotype are more likely to experience negative affective states such as sadness and anxiety, compared to those carrying one or no G allele. From the slightly quadratic effect curve, it can be seen that the negative impact peaks around phase 5-6, that is, the Premenstrual phase including the first day of menstrual cycle (5) to Remaining days of menstrual cycle, part of follicular phase (6), while less negative impact is observed at the beginning and the end of the eight phased cycle (ie, during the ovulatory phase).

## 5 | DISCUSSION

Joint analysis of multiple correlated traits can potentially improve the power to identify genetic variants associated with complex traits. However, association analysis focusing on multiple longitudinal traits has not been well studied. Method on G×E interaction with multiple traits under a longitudinal design is even rarer. In this article, we proposed a joint multivariate varying coefficient modeling approach to accommodate correlated longitudinal traits and proposed a testing procedure to identify genetic variants associated with multiple longitudinal traits with their effects modified by some environmental factors. By modeling the environmental effect with a nonparametric function, one can estimate the dynamic changing effect of *G* on *Y* over the changing values of *X*. The nonparametric function is flexible in the sense that the function is determined by the data without assuming a parametric structure. Both simulation and real data analysis demonstrate the utility of the proposed method.

One difficulty in jointly modeling multiple longitudinal traits is to model the complex correlation structure. For each subject, we should consider correlation between measurements for the same trait at different time points, correlation between measurements at the same time point on different traits, and correlation between measurements at different time points and on different traits. We applied the QIF approach in estimation and testing procedures. There are several advantages for the QIF approach. First, the QIF approach only requires correct specification of the mean structure and does not require any joint likelihood in hypothesis testing. Second, it avoids estimating the nuisance correlation structure parameters by assuming that the inverse of the working correlation matrix can be approximated by a linear combination of several known basis matrices. Third, when the working correlation structure is misspecified, the QIF is more efficient than the GEE approach. Fourth, the inference function of the QIF approach has an explicit asymptotic form, which provides a model selection criteria and allows us to test whether coefficients are significant or time varying based on the asymptotic results. It is worth mentioning that missing completely at random (MCAR) is assumed in this work, which is a common assumption under the QIF framework when dealing with missing data.<sup>31</sup>

In the real application, we investigated the association of SNPs in a candidate gene with two longitudinal traits DEBQ and PANAS. Although the data were regrouped into eight phases, they still carry the temporal information and can be treated like longitudinal data. The results show that three SNPs passed the Bonferroni threshold with the joint test and the *P*-values of the joint test are smaller than the individual marginal test. This shows the relative advantage of the joint test. As shown in the simulation study, the joint test can achieve power gain when the traits are correlated. Therefore, it is essential to assess the correlations between traits when fitting multiple traits jointly and conducting joint testing.

Our method was demonstrated with two traits. The method can be extended to multiple longitudinal traits with  $L > 2$ , although the computational cost might increase. In addition, our method is not restricted to a longitudinal study. It also applies to other studies where multiple traits can be measured over a linear scale. For example, in a pharmacogenetic study, multiple drug responses (eg, blood pressure and heart rate) can be measured over different dosage of a drug treatment. The proposed model can be fitted to assess how genes respond to the increasing dosage levels to affect the drug responses. For another example, in a brain imaging genetic study, brain activities in different brain regions can be measured over a spatial scale and can be treated as multiple traits. One can fit the proposed model to understand how genes affect brain activities over a spatial scale.

## ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their insightful comments that greatly improved the presentation of the manuscript. Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health (NIH) under award number R21HG010073, by the National Institute of General Medical Sciences of the NIH under award number R01GM131398 and by the National Institute of Mental Health of the NIH under award number R01MH082054. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## DATA AVAILABILITY STATEMENT

Research data used for real data analysis are from a different group and are not shared. R code used to implement the method can be downloaded at <https://github.com/Honglang/MPLVC>.

## ORCID

Yuehua Cui  <https://orcid.org/0000-0001-8099-1753>

## REFERENCES

1. Sitlani CM, Rice KM, Lumley T, et al. Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Stat Med*. 2015;34(1):118-130.
2. Macgregor S, Knott SA, White I, Visscher PM. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics*. 2005;171(3):1365-1376.
3. Furlotte NA, Eskin E, Eyheramendy S. Genome-wide association mapping with longitudinal data. *Genet Epidemiol*. 2012;36(5):463-471.
4. Xu Z, Shen X, Pan W, Alzheimer's Disease Neuroimaging Initiative. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS One*. 2014;9(8):e102312.
5. Wang W, Feng Z, Bull SB, Wang Z. A 2-step strategy for detecting pleiotropic effects on multiple longitudinal traits. *Front Genet*. 2014;5:357.
6. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med*. 2016;8(1):78.
7. Lobo I. Pleiotropy: one gene can affect multiple traits; 2008.
8. Ma S, Yang L, Romero R, Cui Y. Varying coefficient model for gene-environment interaction: a non-linear look. *Bioinformatics*. 2011;27(15):2119-2126.
9. Liu X, Cui Y, Li R. Partial linear varying multi-index coefficient model for integrative gene-environment interactions. *Stat Sin*. 2016;26:1037.
10. Qu A, Li R. Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*. 2006;62(2):379-391.
11. Rochon J. Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics*. 1996;52(2):740.
12. Cho H. The analysis of multivariate longitudinal data using multivariate marginal models. *J Multivar Anal*. 2016;143:481-491.
13. Fieuws S, Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Stat Med*. 2004;23(20):3093-3104.
14. Proudfoot J, Faig W, Natarajan L, Xu R. A joint marginal-conditional model for multivariate longitudinal data. *Stat Med*. 2018;37(5):813-828.
15. Zhao L, Chen T, Novitsky V, Wang R. Joint penalized spline modeling of multivariate longitudinal data, with application to HIV-1 RNA load levels and CD4 cell counts. *Biometrics*. 2021;77(3):1061-1074.
16. Hector EC, Song PXX. Joint integrative analysis of multiple data sources with correlated vector outcomes; 2020. arXiv preprint arXiv:2011.14996.
17. Bandyopadhyay S, Ganguli B, Chatterjee A. A review of multivariate longitudinal data analysis. *Stat Methods Med Res*. 2011;20(4):299-330.
18. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res*. 2014;23(1):42-59.
19. Ruppert D, Carroll RJ. Theory & methods: spatially-adaptive penalties for spline fitting. *Aust N Z J Stat*. 2000;42(2):205-223.
20. Wu C, Cui Y, Ma S. Integrative analysis of gene-Environment interactions under a multi-response partially linear varying coefficient model. *Stat Med*. 2014;33(28):4988-4998.

21. Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica*. 1982;50(4):1029-1054.
22. Ruppert D. Selecting the number of knots for penalized splines. *J Comput Graph Stat*. 2002;11(4):735-757.
23. Bai Y, Fung WK, Zhu ZY. Penalized quadratic inference functions for single-index models with longitudinal data. *J Multivar Anal*. 2009;100(1):152-161.
24. Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika*. 2000;87(4):823-836.
25. Tian R, Xue L, Liu C. Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *J Multivar Anal*. 2014;132:94-110.
26. Klump KL, Keel PK, Racine SE, et al. The interactive effects of estrogen and progesterone on changes in emotional eating across the menstrual cycle. *J Abnorm Psychol*. 2013;122(1):131.
27. Klump KL, Burt SA. The Michigan State University Twin Registry (MSUTR): genetic, environmental and neurobiological influences on behavior across development. *Twin Res Hum Genet*. 2006;9(6):971-977.
28. Burt SA, Klump KL. The Michigan state university twin registry (MSUTR): an update. *Twin Res Hum Genet*. 2013;16(1):344-350.
29. Burt SA, Klump KL. The Michigan State University Twin Registry (MSUTR): 15 years of twin and family research. *Twin Res Human Genet Offic J Int Soc Twin Stud*. 2019;22(6):741.
30. Klump KL, Racine SE, Hildebrandt B, et al. Ovarian hormone influences on dysregulated eating: a comparison of associations between women with versus without binge episodes. *Clin Psychol Sci*. 2014;2(5):545-559.
31. Song PXX, Jiang Z, Park E, Qu A. Quadratic inference functions in marginal models for longitudinal data. *Stat Med*. 2009;28(29):3683-3696.

## SUPPORTING INFORMATION

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

**How to cite this article:** Wang H, Zhang J, Klump KL, Alexandra Burt S, Cui Y. Multivariate partial linear varying coefficients model for gene-environment interactions with multiple longitudinal traits. *Statistics in Medicine*. 2022;41(19):3643-3660. doi: 10.1002/sim.9440

## APPENDIX. PROOFS OF THEOREMS

To establish the asymptotic properties for the estimator of  $\theta$ , we need the following regularity conditions.

- (A1)  $\{n_i\}$  is a bounded sequence of integers.
- (A2) The parameter space  $\Omega_\theta$  is compact and  $\theta_0$  is an interior point of  $\Omega_\theta$ .
- (A3) The parameter  $\theta$  is identified, that is, there is a unique  $\theta_0 \in \Omega_\theta$  such that the first moment assumption  $E[g_i(\theta_0)] = 0$  holds for  $i = 1, \dots, N$ , and  $E[g_i(\theta)]$  is continuous.
- (A4)  $E[g(\theta)]$  is continuous in  $\theta$ .
- (A5)  $\bar{C}_N(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N g_i(\hat{\theta}) g_i(\hat{\theta})^T$  converges almost surely to  $C_0$ , which is a constant and invertible matrix.
- (A6) The first derivative of  $\bar{g}_N$  exists and is continuous.  $\frac{\partial \bar{g}_N}{\partial \theta}(\hat{\theta})$  converges in probability to  $G_0$  if  $\hat{\theta}$  converges in probability to  $\theta_0$ .

*Proof of Theorem 1.*  $\hat{\theta}$  exists because (7) has zero as a lower bound and the global minimum exists. To prove the consistency, first, the estimator  $\hat{\theta}$  is obtained by minimizing (7), then we have

$$\frac{1}{N} Q_N(\hat{\theta}) + \lambda_N \hat{\theta}^T D \hat{\theta} \leq \frac{1}{N} Q_N(\theta_0) + \lambda_N \theta_0^T D \theta_0. \quad (\text{A1})$$

Since

$$\frac{1}{N} Q_N(\theta_0) = \bar{g}_N^T(\theta_0) \bar{C}_N^{-1}(\theta_0) \bar{g}_N(\theta_0) = o(1),$$

by the strong law of large number and (A5), and  $\lambda_N = o(1)$ ,

$$\frac{1}{N} Q_N(\theta_0) + \lambda_N \theta_0^T D \theta_0 \xrightarrow{a.s.} 0.$$



Thus, we can obtain from (A1) that

$$\frac{1}{N}Q_N(\hat{\theta}) = \bar{g}_N^T(\hat{\theta})\bar{C}_N^{-1}(\hat{\theta})\bar{g}_N(\hat{\theta}) \xrightarrow{a.s.} 0. \quad (\text{A2})$$

Since the parameter space  $\Omega_\theta$  is compact, by Glivenko-Cantelli theorem,

$$\sup_{\theta \in \Omega_\theta} |\bar{g}_N(\theta) - E[g(\theta)]| \xrightarrow{a.s.} 0.$$

Hence, by (A5) and the continuity mapping theorem,

$$\left| \bar{g}_N^T(\hat{\theta})\bar{C}_N^{-1}(\hat{\theta})\bar{g}_N(\hat{\theta}) - E[g(\hat{\theta})]^T \mathbf{C}_0^{-1} E[g(\hat{\theta})] \right| \xrightarrow{a.s.} 0.$$

Combined with (A2), we get

$$E[g(\hat{\theta})]^T \mathbf{C}_0^{-1} E[g(\hat{\theta})] \xrightarrow{a.s.} 0. \quad (\text{A3})$$

Suppose  $\hat{\theta}$  is not a strong consistent estimator of  $\theta$ , then there exists a neighborhood of the true parameter  $\theta_0$ , say  $U$ , such that  $\hat{\theta} \in U^c$ . Since  $E[g(\theta)]^T \mathbf{C}_0^{-1} E[g(\theta)]$  is a continuous function and  $U^c$  is compact, there exists a point  $\theta^* \in U^c$  such that

$$E[g(\theta^*)]^T \mathbf{C}_0^{-1} E[g(\theta^*)],$$

achieve its minimum in  $U^c$ . By the identification of  $\theta$  in (A3), there is a unique  $\theta_0 \in \Omega_\theta$  satisfying  $E[g(\theta_0)] = 0$ , and we have

$$E[g(\theta^*)]^T \mathbf{C}_0^{-1} E[g(\theta^*)] > 0,$$

which contradicts (A3). Hence,  $\hat{\theta}$  is a consistent estimator of  $\theta$ . ■

*Proof of Theorem 2.* The estimate of  $\theta$  satisfies

$$\frac{1}{N} \frac{\partial Q_N}{\partial \theta}(\hat{\theta}) + 2\lambda_N D \hat{\theta} = 0.$$

By Taylor's expansion, we obtain

$$\frac{1}{N} \frac{\partial Q_N}{\partial \theta}(\theta_0) + 2\lambda_N D \theta_0 + \left( \frac{1}{N} \frac{\partial^2 Q_N}{\partial \theta^2}(\tilde{\theta}) + 2\lambda_N D \right) (\hat{\theta} - \theta_0) = 0,$$

where  $\tilde{\theta}$  is some value between  $\hat{\theta}$  and  $\theta_0$ . Thus, we can have

$$\hat{\theta} - \theta_0 = - \left( \frac{1}{N} \frac{\partial^2 Q_N}{\partial \theta^2}(\tilde{\theta}) + 2\lambda_N D \right)^{-1} \left( \frac{1}{N} \frac{\partial Q_N}{\partial \theta}(\theta_0) + 2\lambda_N D \theta_0 \right). \quad (\text{A4})$$

Since  $\hat{\theta}$  converges to  $\theta_0$  in probability and  $\tilde{\theta}$  is between  $\hat{\theta}$  and  $\theta_0$ , by (A5) and (A6) we can get

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 Q_N}{\partial \theta^2}(\tilde{\theta}) &= 2 \frac{\partial \bar{g}_N^T}{\partial \theta}(\tilde{\theta}) \bar{C}_N^{-1}(\tilde{\theta}) \frac{\partial \bar{g}_N}{\partial \theta}(\tilde{\theta}) + o_p(1) \\ &\xrightarrow{p} 2 \mathbf{G}_0^T \mathbf{C}_0^{-1} \mathbf{G}_0. \end{aligned}$$

When  $\lambda_N = o(N^{-1/2})$ ,

$$\left( \frac{1}{N} \frac{\partial^2 Q_N}{\partial \boldsymbol{\theta}^2}(\tilde{\boldsymbol{\theta}}) + 2\lambda_N D \right)^{-1} = \frac{1}{2} (\mathbf{G}_0^T \mathbf{C}_0^{-1} \mathbf{G}_0)^{-1} + o_p(N^{-1/2}).$$

Similarly, since

$$\frac{1}{N} \frac{\partial Q_N}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \frac{\partial \bar{\mathbf{g}}_N}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) \bar{\mathbf{C}}_N^{-1}(\boldsymbol{\theta}_0) \bar{\mathbf{g}}_N(\boldsymbol{\theta}_0),$$

and  $\lambda_N = o(N^{-1/2})$ , we have

$$\frac{1}{N} \frac{\partial Q_N}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) + 2\lambda_N D \boldsymbol{\theta}_0 = \mathbf{G}_0^T \mathbf{C}_0^{-1} \bar{\mathbf{g}}_N(\boldsymbol{\theta}_0) + o(N^{-1/2}).$$

Therefore, (A4) can be written as

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\sqrt{N}(\mathbf{G}_0^T \mathbf{C}_0^{-1} \mathbf{G}_0)^{-1} \mathbf{G}_0^T \mathbf{C}_0^{-1} \bar{\mathbf{g}}_N(\boldsymbol{\theta}_0) + o_p(1). \tag{A5}$$

By Central Limit Theorem,

$$\sqrt{N} \bar{\mathbf{g}}_N(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{C}_0). \tag{A6}$$

Using (A5) and (A6), we obtain

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbf{G}_0^T \mathbf{C}_0^{-1} \mathbf{G}_0)^{-1}).$$

■

*Proof of Theorem 3.* By Taylor's expansion,

$$Q(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0) - Q(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) = \begin{pmatrix} \boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}} \\ \boldsymbol{\zeta}_0 - \hat{\boldsymbol{\zeta}} \end{pmatrix}^T \dot{Q}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) + \frac{1}{2} \begin{pmatrix} \boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}} \\ \boldsymbol{\zeta}_0 - \hat{\boldsymbol{\zeta}} \end{pmatrix}^T \ddot{Q}(\boldsymbol{\psi}^*, \boldsymbol{\zeta}^*) \begin{pmatrix} \boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}} \\ \boldsymbol{\zeta}_0 - \hat{\boldsymbol{\zeta}} \end{pmatrix},$$

where  $(\boldsymbol{\psi}^*, \boldsymbol{\zeta}^*)$  is some value between  $(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0)$  and  $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}})$ . We can also obtain from Taylor's expansion that

$$Q(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0) - Q(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) = (\boldsymbol{\zeta}_0 - \tilde{\boldsymbol{\zeta}})^T \dot{Q}_{\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) + \frac{1}{2} (\boldsymbol{\zeta}_0 - \tilde{\boldsymbol{\zeta}})^T \ddot{Q}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \boldsymbol{\zeta}^{**}) (\boldsymbol{\zeta}_0 - \tilde{\boldsymbol{\zeta}}),$$

where  $\boldsymbol{\zeta}^{**}$  is between  $\boldsymbol{\zeta}_0$  and  $\tilde{\boldsymbol{\zeta}}$ . From conditions in (9), we have

$$\dot{Q}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) = 0 \text{ and } \dot{Q}(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) = 0.$$

Hence

$$Q(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) - Q(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) = \frac{1}{2} \begin{pmatrix} \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0 \\ \hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 \end{pmatrix}^T \ddot{Q}(\boldsymbol{\psi}^*, \boldsymbol{\zeta}^*) \begin{pmatrix} \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0 \\ \hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \mathbf{0} \\ \tilde{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 \end{pmatrix}^T \ddot{Q}(\boldsymbol{\psi}_0, \boldsymbol{\zeta}^{**}) \begin{pmatrix} \mathbf{0} \\ \tilde{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 \end{pmatrix}.$$

If we expand  $\dot{Q}_{\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}})$  about  $\boldsymbol{\zeta}_0$ , and  $\dot{Q}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}})$  about  $(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0)$ , we obtain

$$\begin{aligned} 0 &= \dot{Q}_{\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \tilde{\boldsymbol{\zeta}}) = \dot{Q}_{\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0) + \ddot{Q}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\tilde{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) + O_p(N^{-\frac{1}{2}}), \\ 0 &= \dot{Q}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\zeta}}) = \dot{Q}_{\boldsymbol{\zeta}}(\boldsymbol{\psi}_0, \boldsymbol{\zeta}_0) + \ddot{Q}_{\boldsymbol{\zeta}\boldsymbol{\psi}}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) + \ddot{Q}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) + O_p(N^{-\frac{1}{2}}). \end{aligned}$$

The above two equations give us

$$\tilde{\zeta} - \zeta_0 = \ddot{Q}_{\zeta\zeta}^{-1} \ddot{Q}_{\zeta\psi} (\hat{\psi} - \psi) + (\hat{\zeta} - \zeta_0) + O_p(N^{-\frac{1}{2}}),$$

which can be written as

$$\begin{pmatrix} \mathbf{0} \\ \tilde{\zeta} - \zeta_0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \ddot{Q}_{\zeta\zeta}^{-1} \ddot{Q}_{\zeta\psi} & I \end{pmatrix} \begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix}.$$

Then  $Q(\psi_0, \tilde{\zeta}) - Q(\hat{\psi}, \hat{\zeta})$  can be written as

$$\begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix}^T \left\{ \ddot{Q}(\psi^*, \zeta^*) - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \ddot{Q}_{\zeta\zeta}^{-1} \ddot{Q}_{\zeta\psi} & I \end{pmatrix}^T \ddot{Q}(\psi_0, \zeta^{**}) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \ddot{Q}_{\zeta\zeta}^{-1} \ddot{Q}_{\zeta\psi} & I \end{pmatrix} \right\} \begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix} + O_p(N^{\frac{1}{2}}),$$

which is asymptotically equivalent to

$$\begin{aligned} & \begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix}^T \left\{ \begin{pmatrix} J_{\psi\psi} & J_{\psi\zeta} \\ J_{\zeta\psi} & J_{\zeta\zeta} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ J_{\zeta\zeta}^{-1} J_{\zeta\psi} & I \end{pmatrix}^T \ddot{Q}(\psi_0, \zeta^{**}) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ J_{\zeta\zeta}^{-1} J_{\zeta\psi} & I \end{pmatrix} \right\} \begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix} \\ & = (\hat{\psi} - \psi_0)^T (J_{\psi\psi} - J_{\psi\zeta} J_{\zeta\zeta}^{-1} J_{\zeta\psi}) (\hat{\psi} - \psi_0). \end{aligned}$$

By theorem 3.2 in Hansen,<sup>21</sup>

$$\begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\zeta} - \zeta_0 \end{pmatrix} \xrightarrow{d} N_d \left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} J_{\psi\psi} & J_{\psi\zeta} \\ J_{\zeta\psi} & J_{\zeta\zeta} \end{pmatrix}^{-1} \right\}.$$

Therefore,

$$\hat{\psi} - \psi_0 \xrightarrow{d} N_{d1} \left( \mathbf{0}, (J_{\psi\psi} - J_{\psi\zeta} J_{\zeta\zeta}^{-1} J_{\zeta\psi})^{-1} \right),$$

thus,  $Q(\psi_0, \tilde{\zeta}) - Q(\hat{\psi}, \hat{\zeta})$  follows  $\chi_{d_1}^2$  asymptotically. ■