FSEM: Functional Structural Equation Models for Twin Functional Data

August 19, 2017

Abstract

The aim of this paper is to develop a novel class of functional structural equation models (FSEMs) for dissecting functional genetic and environmental effects on twin functional data, while characterizing the varying association between functional data and covariates of interest. We propose a three-stage estimation procedure to estimate varying coefficient functions for various covariates (e.g., gender) as well as three covariance operators for the genetic and environmental effects. We develop an inference procedure based on weighted likelihood ratio statistics to test the genetic/environmental effect at either a fixed location or a compact region. We also systematically carry out the theoretical analysis of the estimated varying functions, the weighted likelihood ratio statistics, and the estimated covariance operators. We conduct extensive Monte Carlo simulations to examine the finite-sample performance of the estimation and inference procedures. We apply the proposed FSEM to quantify the degree of genetic and environmental effects on twin white-matter tracts obtained from the UNC early brain development study.

Keywords Covariance function; Functional structural equation model; Genetic and environmental effects; Weighted likelihood ratio test.

1 Introduction

This paper is primarily motivated by the development of statistical methods for dissecting genetic and environmental contributions to functional data, such as brain structure and function, observed from twin pairs. The twin study has been widely used to quantify the difference between the similarity of monozygotic twins (MZ), who have the same genetic materials, and that of dizygotic twins (DZ), who share about 50% of their genes (Feng et al., 2009; Neale et al., 1989; Haseman and Elston, 1970). Such difference between MZ and DZ allows us to disentangle genetic factors from environmental factors on a known phenotype, such as total brain volume. So far, a substantial amount of heritability has been found in different brain volumes (Thompson et al., 2001; Posthuma et al., 2002; Peper et al., 2007), cortical thickness in sensorimotor cortex, middle frontal cortex and anterior temporal cortex (Panizzon et al., 2009), and white matter microstructures (Brouwer et al., 2010; Chiang et al., 2011; Lee et al., 2015).

The aim of this paper is to develop a new functional data analysis (FDA) framework, called functional structural equation model (FSEM), to dissect functional genetic and environmental effects on twin functional data, such as cortical thickness or white matter bundles. Specifically, for $i = 1, \ldots, n$ and $j = 1, 2$, we observe a $p \times 1$ vector of clinical variables, denoted by $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$, and functional data, denoted by $\{y_{ij}(v) : v \in V_0\}$, from the j-th subject of the i-th twin pair, where v is a grid point in $V_0 = \{v_1, v_2, \ldots, v_{N_G}\}$, which is a set of grid points in a common compact space, denoted by V. The n twin pairs consist of n_1 MZ twin pairs for $i = 1, \ldots, n_1$ and $j = 1, 2$, n_2 DZ twin pairs for $i = n_1 + 1, \ldots, n_1 + n_2$ and $j = 1, 2$, and n_3 twin individuals for $i = n_1 + n_2 + 1, \ldots, n = n_1 + n_2 + n_3$ and $j = 1$. As an illustration, Figure 2 (a) and (b) present the plots of two diffusion properties including fractional anisotropy (FA) and mean diffusivity (MD) measured at 152 grid points along the genu tract of the corpus callosum from 40 randomly selected infants in a clinical study of neurodevelopment with 356 neonates, who have at least a twin sibling. Figure 2 (c) and (d) are the FA curves of two randomly selected MZ twin pairs and two randomly selected DZ twin pairs. We are particularly interested in delineating the genetic and environmental variability of these functional FA and MD data and their association with a set of covariates of interest, such

as age and gender. See Section 5 for detailed data analysis on this data set.

The development of FDA for twin functional data represents several major statistical challenges. First, conventional analyses of twin functional data include two steps: the commonly used Gaussian kernel or spline for smoothing functional data and then independently fitting a statistical model, such as structural equation model, at each grid point. Such methods for smoothing raw functional data can change the covariance structure of twin functional data, which is primarily associated with genetic and environmental factors; thus, it can introduce substantial bias in estimating these factors and dramatically increase the Type I and II errors as demonstrated in (Li et al., 2012). Second, although researchers have modeled independent functional data with levels of hierarchies (Morris and Carroll, 2006; Bathia et al., 2010; Ramsay and Silverman, 2005; Li et al., 2013; Zhang and Chen, 2007; Zhu et al., 2012b, 2011; Guo, 2002; Scheipl et al., 2015; Chen and Müller, 2012; Zhu et al., 2011), their associated inference methods primarily make statistical inferences on the mean structure of functional data and are not directly applicable to twin data, which requires a careful analysis of the covariance structure of functional data. Some popular statistical methods include nonparametric mixed-effects models (Gu and Ma, 2005; Wang, 1998a,b; Wood, 2006, 2013; Zhang et al., 1998) and varying coefficient models (Fan and Zhang, 2008; Zhu et al., 2012a; Wu and Zhang, 2006; Huang et al., 2002). For instance, Morris and Carroll (2006) developed general functional mixed effects models with multiple levels of random effect functions as well as curve-to-curve deviations. See comprehensive reviews in Shi and Choi (2011); Hsing and Eubank (2015); Horváth and Kokoszka (2012) and references therein. Third, according to the best of our knowledge, little has been done on the analysis of functional data from twin studies. Recently, Lei et al. (2015) introduced a functional data model for genetically related individuals and developed its associated estimation procedure, but they have not rigorously investigated several key statistical inference procedures, such as asymptotic properties of estimates and test statistics. Moreover, their functional data model does not incorporate the functional environmental factors, so it does not allow us to dissect genetic and environmental effects on functional data, which are scientifically critical for twin studies.

We develop a functional structural equation modeling (FSEM) framework to dissect

functional genetic and environmental effects on twin functional responses and their association with a set of covariates. Our FSEM can be regarded as a novel extension of standard structural equation models for twin scalar responses (Neale et al., 1989; Haseman and Elston, 1970; Feng et al., 2009; Wang et al., 2011; Rabe-Hesketh et al., 2008), nonparametric mixed-effects models (Gu and Ma, 2005; Wang, 1998a,b; Wood, 2006, 2013; Zhang et al., 1998), and varying coefficient models (Fan and Zhang, 2008; Zhu et al., 2012a; Wu and Zhang, 2006; Huang et al., 2002). Our major contributions of this paper are as follows:

- FSEM not only disentangles functional genetic and environmental effects on twin functional data, but also characterizes the varying association between functional data and covariates of interest.

- We propose an estimation procedure based on weighted maximum likelihood functions, which explicitly incorporate the spatial smoothness in the mean and covariance structure of functional data. We develop an inference procedure based on weighted likelihood ratio statistics to test the genetic/environmental effect either at a fixed location or over a compact region.

- We use simulations and a real data analysis to show that FSEM can substantially boost the detection power for dissecting functional genetic and environmental effects on twin functional responses.

- We provide a comprehensive theoretical analysis of FSEM. Formally, we establish the weak convergence of the estimated varying association functions and the estimated genetic and environmental covariance operators, and the asymptotic distribution of the local and global test statistics.

- We have developed the FSEM package by using both matlab and python and will release it through the website "https://www.nitrc.org/" and our group website. Our package includes a Graphical User Interface (GUI) to pack the code, also freely downloadable from the same website.

The rest of this paper is organized as follows. In Section 2, we introduce FSEM and its associated estimation and inference procedures. We construct weighted likelihood ratio

statistics to test both local and global genetic effects on functional data. In Section 3, we establish the theoretical properties of our estimation and inference procedures. In Section 4, we carry out simulations to examine the finite sample performance of FSEM. In Section 5, we use FSEM to dissect the functional genetic and environmental effects on the genu tract of the corpus callosum in a clinical twin study of neurodevelopment.

2 Methods

2.1 Functional Structural Equation Model

We start with a simple extension of standard ACDE model (Haseman and Elston, 1970) for twin scalar responses as follows:

$$y_{ij}(v) = x_{ij}^T \boldsymbol{\beta}(v) + a_{ij}(v) + d_{ij}(v) + c_i(v) + e_{ij}(v) \tag{1}$$

for $i = 1, \ldots, n$ and $j = 1, m_i$, in which $m_i = 2$ for $i \leq n_1 + n_2$ and 1 for $i > n_1 + n_2$, where $\boldsymbol{\beta}(v) = (\beta_1(v), \ldots, \beta_p(v))^T$ is a $p \times 1$ vector of coefficient functions. Moreover, $a_{ij}(v) \sim N(0, \sigma_a(v)^2)$, $c_i(v) \sim N(0, \sigma_c(v)^2)$, $d_{ij}(v) \sim N(0, \sigma_d(v)^2)$, and $e_{ij}(v) \sim N(0, \sigma_e(v)^2)$ represent the additive genetic, common environmental, dominant genetic, and unique environmental effects, respectively. It is also assumed that

$$\text{cov}(a_{i1}(v), a_{i2}(v)) = \sigma_a(v)^2 (\text{MZ})_i + 0.5 \times \sigma_a(v)^2 (\text{DZ})_i, \tag{2}$$

$$\text{cov}(d_{i1}(v), d_{i2}(v)) = \sigma_d(v)^2 (\text{MZ})_i + 0.25 \times \sigma_d(v)^2 (\text{DZ})_i, \tag{3}$$

where $(\text{MZ})_i$ (or $(\text{DZ})_i$) is an indicator function of the event that the $i-$th pair of twin subjects is MZ (or DZ). See Figure 3 for the diagram of ACDE model for twin data. Due to an identifiability issue (Wang et al., 2011), it is common to consider two simpler models. One is the ACE model, which only includes additive genetic, common environmental, and unique environmental effects. The other is the ADE model, which includes additive and dominant genetic, and unique environmental effects.

A conceptual issue associated with (1) comes from the difficulty in characterizing the relationship between $(a_{i1}(v), d_{i1}(v))$ and $(a_{i2}(v), d_{i2}(v))$ and in dissecting functional genetic and environmental effects. Specifically, for MZ twins, $a_{i1}(v) = a_{i2}(v)$ for all $v \in V$, whereas for DZ twins, $\text{corr}(a_{i1}(v), a_{i2}(v)) = 0.5$ holds for all $v \in V$. The next question

is how to define a bivariate process that satisfies both conditions as v varies in V, while preserving spatial smoothness in V. Similar comments hold for $d_{i1}(v)$ and $d_{i2}(v)$. To deal with such issue, we introduce a reparametrization of ACE model for twin functional data, which reduces to the standard ACE model at each grid point.

We introduce a new functional ACE model as a special case of FSEM as follows:

$$\begin{aligned} y_{ij}(v) &= x_{ij}^T \boldsymbol{\beta}(v) + r_{ij}(v), \\ r_{ij}(v) &= \sqrt{0.5}(\mathrm{DZ})_i a_{ij}(v) + \{(\mathrm{MZ})_i + \sqrt{0.5}(\mathrm{DZ})_i\} a_i(v) + c_i(v) + e_{ij}(v), \end{aligned} \tag{4}$$

in which $a_{ij}(v)$ and $a_i(v)$ are introduced to represent within-curve and between-curve functional additive genetic effects on the i-th twin pair. It is also assumed that

$$e_{ij}(v) = e_{ij,G}(v) + e_{ij,L}(v), \tag{5}$$

where $e_{ij,L}(v)$ are measurement errors representing local variability and $e_{ij,G}(s)$ are stochastic processes representing unique functional environmental effects.

It is assumed that $a_{ij}(v)$, $a_i(v)$, $c_i(v)$, $e_{ij,G}(v)$, and $e_{ij,L}(v)$ are mutually independent, and they are independent and identical copies of $\mathrm{GP}(\mathbf{0}, \Sigma_a)$, $\mathrm{GP}(\mathbf{0}, \Sigma_a)$, $\mathrm{GP}(\mathbf{0}, \Sigma_c)$, $\mathrm{GP}(\mathbf{0}, \Sigma_{e,G})$, and $\mathrm{GP}(\mathbf{0}, \Sigma_{e,L})$, respectively, where $\mathrm{GP}(\mathbf{0}, \Sigma)$ represents a Gaussian process with mean $\mathbf{0}$ and covariance function $\Sigma(v, v')$. Moreover, $e_{ij,L}(v)$ and $e_{ij,L}(v')$ are assumed to be independent for $v \neq v'$, that is, $\Sigma_{e,L}(v, v') = 0$ for $v \neq v'$. We also denote $\Sigma_e(v, v') = \Sigma_{e,G}(v, v') + \Sigma_{e,L}(v, v')$. The functional ACE model reduces to standard ACE model at each grid point. For $v \neq v'$, we have

$$\begin{aligned} \mathrm{cov}(y_{ij}(v), y_{ij}(v')) &= \Sigma_a(v, v') + \Sigma_c(v, v') + \Sigma_{e,G}(v, v') + \Sigma_{e,L}(v, v')\mathbf{1}(v = v'), \\ \mathrm{cov}(y_{i1}(v), y_{i2}(v')) &= \{(\mathrm{MZ})_i + 0.5 \times (\mathrm{DZ})_i\}\Sigma_a(v, v') + \Sigma_c(v, v'). \end{aligned} \tag{6}$$

To the best of our knowledge, (4) is the first FDA framework of its kind for twin functional data and differs significantly from other models in the existing literature. Most FDA models focus on the mean structure of functional data, while they do not dissect genetic and environmental effects on functional data. See Wang et al. (2016) and Morris (2015) for comprehensive reviews of various FDA models. Moreover, most existing structural equation models developed for twin scalar responses are not directly applicable to twin functional data. Due to these major differences, we are facing many challenges in

accurately estimating covariance functions $\Sigma_a(v,v')$, $\Sigma_c(v,v')$, and $\Sigma_{e,G}(v,v')$ and making statistical inference on genetic and environmental effects across all $v \in V$.

2.2 Estimation Procedure

Under model (4), the primary interest is to estimate $\boldsymbol{\beta}(\cdot)$, $\Sigma_a(\cdot,\cdot)$, $\Sigma_c(\cdot,\cdot)$, $\Sigma_{e,G}(\cdot,\cdot)$, and $\{\Sigma_{e,L}(v,v) : v \in V_0\}$. The estimation procedure consists of three steps and the key idea of each step is given as follows.

- Step (I): calculate the maximum likelihood estimate of $\boldsymbol{\theta}(v) = (\boldsymbol{\sigma}^2(v)^T, \boldsymbol{\beta}(v)^T)^T$ at each grid point $v \in V_0$, denoted as $\widehat{\boldsymbol{\theta}}(v) = (\widehat{\boldsymbol{\sigma}}^2(v)^T, \widehat{\boldsymbol{\beta}}(v)^T)^T$, where $\boldsymbol{\sigma}^2(v) = (\sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))^T$;

- Step (II): update the estimate of $\boldsymbol{\sigma}^2(v) = (\sigma_a^2(v), \sigma_c^2(v), \sigma_e^2(v))^T$ at all $v \in V$ by using the weighted likelihood function of the estimated residuals $\widehat{r}_{ij}(v') = y_{ij}(v') - x_{ij}^T\widehat{\boldsymbol{\beta}}(v')$ for a set of grid points $v' \in V_0$ near v, denote it as $\widehat{\boldsymbol{\sigma}}_K^2(v)$;

- Step (III): calculate the estimates of covariance operators $\Sigma_a(\cdot,\cdot)$, $\Sigma_c(\cdot,\cdot)$, and $\Sigma_{e,G}(\cdot,\cdot)$ by using the local constant regression technique (Fan and Gijbels, 1996) based on (6).

The key idea of Step (I) is to make statistical inference at each $v \in V_0$ by using data observed at the grid point v, which is the standard pixel-wise method. In Step (I), we calculate the maximum likelihood estimate of $\widehat{\boldsymbol{\theta}}(v)$ based on the log-likelihood function at the grid point $v \in V_0$, denoted as $\mathcal{L}_n(\boldsymbol{\theta}(v); \boldsymbol{Y}(v))$, where $\boldsymbol{Y}(v)$ are imaging measures for all subjects at v. We will use the maximum likelihood estimates $\widehat{\boldsymbol{\beta}}(v)$ for $v \in V_0$ in Steps (II) and (III). If needed, we can employ smoothing methods (e.g., local linear) to obtain an estimate of $\boldsymbol{\beta}(v)$ for all $v \in V$, we still denote it as $\widehat{\boldsymbol{\beta}}(v)$ (Fan and Gijbels, 1996; Zhu et al., 2012b).

The key idea of Step (II) is to make statistical inference across all $v \in V$ by using data observed at those grid points close to v. In Step (II), we construct a weighted log-likelihood function at each v, denoted as $\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$, in order to obtain an updated estimate of $\boldsymbol{\sigma}^2(v)$, where $\widehat{\boldsymbol{R}}$ represents all $\widehat{r}_{ij}(v')$'s for $v' \in V_0$. Specifically, $\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ is given by $m^{-1}\sum_{v' \in V_0} K_{h_1}(v'-v)\mathcal{L}_n(\boldsymbol{\sigma}^2(v), \widehat{\boldsymbol{\beta}}(v'); \boldsymbol{Y}(v'))$, where $K(v)$ is a kernel function

and $K_{h_1}(v) = h_1^{-1}K(v/h_1)$ is the rescaled kernel function with a bandwidth h_1. We select the bandwidth h_1 by using 5-fold cross-validation.

The key idea of Step (III) is to apply the local constant regression technique (Fan and Gijbels, 1996) to estimate $\Sigma_a(v,v')$, $\Sigma_c(v,v')$, and $\Sigma_{e,G}(v,v')$. By using the covariance structures in (6), we minimize an objective function $W_n(v,v')$ given by

$$\begin{aligned}
&\frac{1}{2n-n_3}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{v_0\neq v_0'\in V_0}\left[\widehat{U}_{ij}(v_0,v_0')-\Sigma_a(v,v')-\Sigma_c(v,v')-\Sigma_{e,G}(v,v')\right]^2K_{h_2}^2(v_0,v;v_0',v')\\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{v_0\neq v_0'\in V_0}\left[\widehat{U}_{i}(v_0,v_0')-\Sigma_a(v,v')-\Sigma_c(v,v')\right]^2K_{h_2}^2(v_0,v;v_0',v')\\
&+\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\sum_{v_0\neq v_0'\in V_0}\left[\widehat{U}_{i}(v_0,v_0')-0.5\Sigma_a(v,v')-\Sigma_c(v,v')\right]^2K_{h_2}^2(v_0,v;v_0',v'),
\end{aligned}$$

where $K_{h_2}^2(v_0,v;v_0',v') = K_{h_2}(v_0-v)K_{h_2}(v_0'-v')$, $\widehat{U}_{ij}(v_0,v_0') = \widehat{r}_{ij}(v_0)\widehat{r}_{ij}(v_0')$, and $\widehat{U}_i(v_0,v_0') = [\widehat{r}_{i1}(v_0)\widehat{r}_{i2}(v_0')+\widehat{r}_{i1}(v_0')\widehat{r}_{i2}(v_0)]/2$. With some calculations, we have

$$\begin{aligned}
\widehat{\Sigma}_a(v,v') &= [2\widehat{S}_{w1}(v,v')-2\widehat{S}_{w2}(v,v')]/[w(v,v')],\\
\widehat{\Sigma}_c(v,v') &= [-\widehat{S}_{w1}(v,v')+2\widehat{S}_{w2}(v,v')]/[w(v,v')],\\
\widehat{\Sigma}_{e,G}(v,v') &= [\widehat{S}_{w0}(v,v')-\widehat{S}_{w1}(v,v')]/[w(v,v')],
\end{aligned}\tag{7}$$

where $w(v,v')$, $\widehat{S}_{w0}(v,v')$, $\widehat{S}_{w1}(v,v')$, and $\widehat{S}_{w2}(v,v')$ are, respectively, given by

$$\begin{aligned}
w(v,v') &= \frac{1}{N_G(N_G-1)}\sum_{v_0\neq v_0'\in V_0}K_{h_2}^2(v_0,v;v_0',v'),\\
\widehat{S}_{w0}(v,v') &= \frac{1}{2n-n_3}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{N_G(N_G-1)}\sum_{v_0\neq v_0'\in V_0}\widehat{U}_{ij}(v_0,v_0')K_{h_2}^2(v_0,v;v_0',v'),\\
\widehat{S}_{w1}(v,v') &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_G(N_G-1)}\sum_{v_0\neq v_0'\in V_0}\widehat{U}_{i}(v_0,v_0')K_{h_2}^2(v_0,v;v_0',v'),\\
\widehat{S}_{w2}(v,v') &= \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}\frac{1}{N_G(N_G-1)}\sum_{v_0\neq v_0'\in V_0}\widehat{U}_{i}(v_0,v_0')K_{h_2}^2(v_0,v;v_0',v').
\end{aligned}$$

Furthermore, if we set $\Sigma_a(v,v') = 0$ for $v,v'\in V$, then $\widehat{\Sigma}_c(v,v')$ and $\widehat{\Sigma}_{e,G}(v,v')$, which minimize $W_n(v,v')$, are, respectively, given by

$$\begin{aligned}
\widehat{\Sigma}_c(v,v') &= [\widehat{S}_{w1}(v,v')+\widehat{S}_{w2}(v,v')]/[2w(v,v')],\\
\widehat{\Sigma}_{e,G}(v,v') &= [2\widehat{S}_{w0}(v,v')-\widehat{S}_{w1}(v,v')-\widehat{S}_{w2}(v,v')]/[2w(v,v')].
\end{aligned}\tag{8}$$

We also select the bandwidth h_2 by using 5-fold cross-validation.

2.3 Inference Procedure

We propose an efficient inference procedure to test the functional genetic or environmental effects at either a fixed location or a compact region. For the sake of space, we focus on the genetic effects from now on. We need to introduce some notation. We use subscript $*$ to denote the true parameter values, say $\boldsymbol{\theta}_*(v_k)$. Let $\boldsymbol{y}_{ik} = \boldsymbol{y}_i(v_k) = (y_{i1}(v_k), y_{i2}(v_k))^T$ and $\boldsymbol{Y}_k = \boldsymbol{Y}(v_k) = (\boldsymbol{y}_{1k}, \cdots, \boldsymbol{y}_{nk})$. We denote $\mathcal{L}_n^{(M)}(\boldsymbol{\theta}(v_k); \boldsymbol{Y}_k)$ as $\sum_{i=1}^{n_1} \log\left[f^{(M)}(\boldsymbol{y}_{ik}; \boldsymbol{\theta}(v_k))\right]$ for all MZ twins, where $f^{(M)}(\boldsymbol{y}_{ik}; \boldsymbol{\theta}(v_k))$ is the density of $\boldsymbol{y}_{ik}$ for the i-th MZ twin pair at the grid point v_k. We define $\mathcal{L}_{n,K}^{(M)}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ as

$$\mathcal{L}_{n,K}^{(M)}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}}) = \frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \mathcal{L}_n^{(M)}(\boldsymbol{\sigma}^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{Y}_k).$$

Moreover, we introduce similar notation for DZ twin pairs and twin individuals by replacing the superscript (M) by (D) and (I), respectively.

2.3.1 Local Test

To test the genetic effect at a fixed location, we formulate it as follows:

$$H_0 : \sigma_{a*}^2(v) = 0 \quad \text{v.s.} \quad H_1 : \sigma_{a*}^2(v) > 0. \tag{9}$$

We consider a likelihood ratio test statistic and a weighted likelihood ratio test statistic at v given by

$$\begin{aligned} \mathrm{LR}_n(v) &:= -2\left[\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_0(v); \boldsymbol{Y}(v)) - \mathcal{L}_n(\widehat{\boldsymbol{\theta}}(v); \boldsymbol{Y}(v))\right], \ v \in V_0, \\ \mathrm{WLR}_n(v) &:= -2\left[\mathcal{L}_{n,K}(\widehat{\boldsymbol{\sigma}}_{0,K}^2(v); \widehat{\boldsymbol{R}}) - \mathcal{L}_{n,K}(\widehat{\boldsymbol{\sigma}}_K^2(v); \widehat{\boldsymbol{R}})\right], \ v \in V. \end{aligned}$$

where $\widehat{\boldsymbol{\theta}}_0(v)$ and $\widehat{\boldsymbol{\theta}}(v)$ are the maximum likelihood estimates of $\boldsymbol{\theta}(v)$ under H_0 and H_1, respectively. Moreover, $\widehat{\boldsymbol{\sigma}}_{0,K}^2(v)$ and $\widehat{\boldsymbol{\sigma}}_K^2(v)$ are the maximum weighted likelihood estimates of $\mathcal{L}_{n,K}(\boldsymbol{\sigma}_{0,K}^2(v); \widehat{\boldsymbol{R}})$ under H_0 and H_1, respectively. Note that for $\mathrm{LR}_n(v)$, v has to be in V_0 since no data is available for $v \in V/V_0$. In contrast, our newly proposed weighted likelihood ratio test statistic does not require $v \in V_0$, since data can be borrowed from nearby grid points of v in V_0. The asymptotic distribution of $\mathrm{LR}_n(v)$ can be similarly derived based on the results of Self and Liang (1987). A special case is when $\sigma_{c*}^2(v) > 0$ and $\sigma_{e*}^2(v) > 0$, the asymptotic distribution of $\mathrm{LR}_n(v)$ is $0.5\chi_0^2 + 0.5\chi_1^2$. The test $\mathrm{LR}_n(v)$ is similar to the test for random effects in regression models (Wood, 2013).

A challenging issue is to derive the asymptotic distribution of $\mathrm{WLR}_n(v)$. We only give a brief derivation for $\sigma^2_{c*}(v) > 0$ and $\sigma^2_{e*}(v) > 0$. We use $\Lambda_{\mathcal{E}}$ to denote the approximation cone at a given vertex $\boldsymbol{\sigma}^2(v)$ for a given set $\mathcal{E}$ (Andrews, 2001; Self and Liang, 1987). In this case, the null and alternative domains of test (9) are given by $\mathcal{E}_0 = \{0\} \times (0, \infty)^2$ and $\mathcal{E}_1 = (0, \infty)^3$, respectively. We approximate $\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ by a quadratic function of $\boldsymbol{\sigma}^2_*(v)$ and then calculate the approximation cones of $\mathcal{E}_1$ and $\mathcal{E}_0$, denoted as $\Lambda_{\mathcal{E}_1}$ and $\Lambda_{\mathcal{E}_0}$, respectively (Self and Liang, 1987; Andrews, 2001; Zhu and Zhang, 2004). The approximation cone of $\mathcal{E}_0$ and $\mathcal{E}_1$ at $\boldsymbol{\sigma}^2(v)$ are $\Lambda_{\mathcal{E}_0} = \{0\} \times R^2$ and $\Lambda_{\mathcal{E}_1} = (0, \infty) \times R^2$, respectively.

We have a quadratic approximation to $\mathrm{WLR}_n(v)$ as follows:

$$\begin{aligned}
\mathrm{WLR}_n(v) &= 2 \sup_{\boldsymbol{\sigma}^2(v) \in \mathcal{E}_1} \mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}}) - 2 \sup_{\boldsymbol{\sigma}^2(v) \in \mathcal{E}_0} \mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}}) \\
&= \sup_{\lambda \in \Lambda_{\mathcal{E}_1}} \left[2\lambda^T \mathcal{J}_{n,K}(\boldsymbol{\sigma}^2_*(v); \widehat{\boldsymbol{R}}) - \lambda^T \mathcal{I}_{n,K}(\boldsymbol{\sigma}^2_*(v); \widehat{\boldsymbol{R}}) \lambda \right] \\
&\quad - \sup_{\lambda \in \Lambda_{\mathcal{E}_0}} \left[2\lambda^T \mathcal{J}_{n,K}(\boldsymbol{\sigma}^2_*(v); \widehat{\boldsymbol{R}}) - \lambda^T \mathcal{I}_{n,K}(\boldsymbol{\sigma}^2_*(v); \widehat{\boldsymbol{R}}) \lambda \right] + o_p(1),
\end{aligned} \tag{10}$$

where $\mathcal{J}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ and $-\mathcal{I}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ are the first-order and second-order derivatives of the log-likelihood function of all data $\mathcal{L}_{n,K}(\boldsymbol{\sigma}^2(v); \widehat{\boldsymbol{R}})$ with respect to $\boldsymbol{\sigma}^2(v)$. Subsequently, based on (10), we are able to derive the asymptotic distribution of $\mathrm{WLR}_n(v)$ presented in Section 3.

We propose an efficient resampling method to approximate the empirical distribution of $\mathrm{WLR}_n(v)$ as follows:

Step 1: We generate independent and identically distributed random samples, $\{\xi_{i,g} : i = 1, \ldots, n\}$, from standard normal distribution $N(0, 1)$. Here g represents a replication number.

Step 2: We calculate

$$\begin{aligned}
\sqrt{n}\mathcal{J}^{(g)}_{n,K}(\boldsymbol{\sigma}^2_*(v); \widehat{\boldsymbol{R}}) &= \frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \sum_{i=1}^{n_1} \xi_{i,g} \dot{\ell}^{(M)}_{\sigma}(\boldsymbol{\sigma}^2_*(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) \\
&+ \frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \sum_{i=n_1+1}^{n_1+n_2} \xi_{i,g} \dot{\ell}^{(D)}_{\sigma}(\boldsymbol{\sigma}^2_*(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik}) \\
&+ \frac{1}{N_G} \sum_{k=1}^{N_G} K_{h_1}(v_k - v) \sum_{i=n_1+n_2+1}^{n} \xi_{i,g} \dot{\ell}^{(I)}_{\sigma}(\boldsymbol{\sigma}^2_*(v), \widehat{\boldsymbol{\beta}}(v_k); y_{ik}),
\end{aligned}$$

where $\dot{\ell}_{\sigma}^{(S)}(\boldsymbol{\sigma}_*^2(v), \widehat{\boldsymbol{\beta}}(v_k); \boldsymbol{y}_{ik})$ is the first order derivative of log-likelihood for the i-th twin pair for $S = M, D$, and I. It is important to note that $\mathcal{J}_{n,K}^{(g)}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}})$ converges weakly to the same stochastic process as $\mathcal{J}_{n,K}(\boldsymbol{\sigma}_*(v); \widehat{\boldsymbol{R}})$ by using the conditional central limit theorem (Theorem 10.2 of Pollard (1990)).

Step 3: The third step is to calculate the weighted likelihood ratio

$$\begin{aligned}\mathrm{WLR}_n^{(g)}(v) = &\sup_{\lambda \in \Lambda_{\mathcal{E}_1}} \left[2\lambda^T \mathcal{J}_{n,K}^{(g)}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}}) - \lambda^T \mathcal{I}_{n,K}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}})\lambda\right] \\ &- \sup_{\lambda \in \Lambda_{\mathcal{E}_0}} \left[2\lambda^T \mathcal{J}_{n,K}^{(g)}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}}) - \lambda^T \mathcal{I}_{n,K}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}})\lambda\right] + o_p(1).\end{aligned}$$

Step 4: We repeat the above three steps G times and obtain realization, $\{\mathrm{WLR}_n^{(g)}(v) : g = 1, 2, \ldots, G\}$. It can be shown that the empirical distribution of $\mathrm{WLR}_n^{(g)}(v)$ converges to the asymptotic distribution of $\mathrm{WLR}_n(v)$. Therefore, the empirical distribution of this realization forms the basis for the calculation of critical values in hypothesis testing as well as power analysis.

We approximate p-value for the testing problem (9) by using

$$p \approx \widehat{p}^G = \frac{1}{G}\sum_{g=1}^{G} \mathbf{1}(\mathrm{WLR}_n^{(g)}(v) > \mathrm{WLR}_n(v)), \tag{11}$$

where $\mathbf{1}(\cdot)$ is an indicator function of an event. Note that the true $\boldsymbol{\sigma}_*^2(v)$ are involved in all the calculations, and it is replaced by the $\widehat{\boldsymbol{\sigma}}_{0,K}^2(v)$ in practice, since $\widehat{\boldsymbol{\sigma}}_{0,K}^2(v)$ is a consistent estimator of $\boldsymbol{\sigma}_*^2(v)$ under the hypothesis H_0.

The above resampling approach is quite efficient since the optimization in Step 3 is fast. Consider the eigen-decomposition of $\mathcal{I}_{n,K}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}})$ as $P^T D P$, where P is an orthogonal matrix and D is a diagonal matrix with $d_1 \geq d_2 \geq d_3 \geq 0$ as its diagonal elements. Furthermore, since $P[(0, \infty) \times R^2] = (0, \infty) \times R^2$ and $P[\{0\} \times R^2] = \{0\} \times R^2$ (under different orthogonal bases), we have

$$\begin{aligned}\mathrm{WLR}_n^{(g)}(v) &= \sup_{\lambda \in \Lambda_{\mathcal{E}_1}} \left[2\lambda^T P \mathcal{J}_{n,K}^{g}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}}) - \lambda^T D\lambda\right] \\ &\quad - \sup_{\lambda \in \Lambda_{\mathcal{E}_0}} \left[2\lambda^T P \mathcal{J}_{n,K}^{g}(\boldsymbol{\sigma}_*^2(v); \widehat{\boldsymbol{R}}) - \lambda^T D\lambda\right] + o_p(1) \\ &= \frac{[P\mathcal{J}_{n,K}^{g}]_1^2}{d_1} \mathbf{1}([P\mathcal{J}_{n,K}^{g}]_1 \geq 0),\end{aligned}$$

where $[P\mathcal{J}_{n,K}^{g}]_1$ denotes the first entry of $P\mathcal{J}_{n,K}^{g}$.

2.3.2 Global Test

We propose a global tractwise test procedure to the following hypotheses

$$H_0 : \int_{v\in V} \Sigma_{a*}(v,v)dv = 0 \quad \text{v.s.} \quad H_1 : \int_{v\in V} \Sigma_{a*}(v,v)dv > 0. \tag{12}$$

Note that the tractwise test procedure developed here is not limited to the entire tract V and can be adapted to any subinterval of V. We consider the global test statistics as follows:

$$D_i(v,v') = [\widehat{r}_{i1}(v) - \widehat{r}_{i2}(v)]\,[\widehat{r}_{i1}(v') - \widehat{r}_{i2}(v')]\,,$$
$$T_n = \sup_{w:\|w\|_2=1} \frac{\sum_{i\in DZ}\int_{v\in V}\int_{v'\in V} w(v)D_i(v,v')w(v')dvdv'}{\sum_{i\in MZ}\int_{v\in V}\int_{v'\in V} w(v)D_i(v,v')w(v')dvdv'},$$

where $\|w\|_2^2 = \int_{v\in V} w(v)^2dv$. The bigger value of T_n indicates the stronger evidence of global genetic effect. Under the nulll hypothesis of no global genetic effect, $(\widehat{r}_{i1}(\cdot), \widehat{r}_{i2}(\cdot))$ are independently and identically distributed across MZ and DZ twin pairs. The p-value of the test statistic T can be obtained through a permutation procedure over MZ and DZ twin pairs. To be specific, for each permutation $g = 1, \ldots, G$, we randomly select n_1 twin pairs as MZ twin pairs from $n_1 + n_2$ MZ and DZ twin pairs and compute the new test statistic as $T_n^{(g)}$. Then the p-value can be approximated by $\sum_{g=1}^{G} \mathbf{1}(T_n^{(g)} > T_n)/G$.

3 Asymptotic Properties

In this section, we systematically investigate the asymptotic properties (e.g., convergence rate) of various estimates and test statistics developed in Section 2. Without loss of generality, we set $V = [0,1]$. The kernel function $K(v)$ is a symmetric density with compact support $[-1,1]$. Let $u_r(K) = \int v^r K(v)dv$ and $v_r(K) = \int v^r K^2(v)dv$, where r is any nonnegative integer. We assume that $O_p(1)$ and $o_p(1)$ hold uniformly across all v in either V or V_0 throughout the paper. Moreover, the sample sizes n_1 and n_2 diverge to infinity such that $n_1/n \to \alpha_1$, $n_2/n \to \alpha_2$, and $n_3/n \to \alpha_3$, where $\alpha_1 > 0, \alpha_2 > 0$, and $\alpha_3 \geq 0$ such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$. We use $\to_d$ to denote convergence in distribution and use $\Rightarrow_d$ to denote weak convergence for stochastic processes. $N(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2. We state the following theorems, whose assumptions can be found in the Appendix and detailed proofs in supplemental materials.

Theorem 1. *Suppose that Assumptions (C1)-(C4) and (C7) hold. As* $n \to \infty$, *the following results hold:*

(i) We have $\sqrt{n}\left\{\widehat{\boldsymbol{\theta}}(v_k) - \boldsymbol{\theta}_*(v_k)\right\} \to_d \widehat{\boldsymbol{\lambda}}(v_k)$ *and*

$$\sup_{v_k \in V_0} \sqrt{n}\left\{\widehat{\boldsymbol{\theta}}(v_k) - \boldsymbol{\theta}_*(v_k)\right\} = O_p(\log(1 + N_G)),$$

where $\widehat{\boldsymbol{\lambda}}(v_k) = \arg\min_{\boldsymbol{\lambda} \in [0,\infty) \times R^{(p+2)}} \{\boldsymbol{\lambda} - Z(v_k)\}^T \mathcal{I}(v_k) \{\boldsymbol{\lambda} - Z(v_k)\}$, *in which* $Z(v_k) = \mathcal{I}(v_k)^{-1} \mathcal{J}(v_k)$. *Moreover,* $\mathcal{J}(v_k)$ *denotes a random score vector and* $\mathcal{I}(v_k)$ *is the average Fisher information matrix at* $\boldsymbol{\theta}_*(v_k)$.

(ii) If $\sigma^2_{c*}(v_k) > 0$ *and* $\sigma^2_{e*}(v_k) > 0$, *then under local alternatives* $H_n : \sigma^2_a(v_k) = h(v_k)/\sqrt{n}$, *we have*

$$\sqrt{n}\{\widehat{\sigma}^2_a(v_k) - \sigma^2_{a*}(v_k)\} \xrightarrow{H_n}_d \{H\mathcal{I}(v_k)^{-1}H^T\}^{1/2} \times N(\widetilde{h}(v_k), 1)\mathbf{1}(N(\widetilde{h}(v_k), 1) \geq 0),$$
$$LR_n(v_k) \xrightarrow{H_n}_d N(\widetilde{h}(v_k), 1)^2 \mathbf{1}\left(N(\widetilde{h}(v_k), 1) \geq 0\right),$$

where $\widetilde{h}(v_k) = \left\{H\mathcal{I}^{-1}(v_k)H^T\right\}^{-1/2} h(v_k)$ *and* $H = [1, \mathbf{0}^T_{p+2}] \in R^{1 \times (p+3)}$.

Theorem 1 (i) delineates the uniform convergence rate of $\widehat{\boldsymbol{\theta}}(v_k)$ for all $v_k \in V_0$. Theorem 1 (ii) gives the asymptotic distribution of $\widehat{\sigma}^2_a(v_k)$ and $\mathrm{LR}_n(v_k)$ under a sequence of local alternatives when both $\sigma^2_{c*}(v_k)$ and $\sigma^2_{e*}(v_k)$ are greater than zero. When $h(v_k)$ is equal to zero, Theorem 1 (ii) reduces to the well-known results in (Self and Liang, 1987; Andrews, 2001). Theorem 1 (ii) characterizes the local power of $\mathrm{LR}_n(v_k)$ as $h(v_k) > 0$.

Theorem 2. *Suppose that Assumptions (C1)-(C6) and (C7b) hold. As* $n \to \infty$, *the following results hold:*

(i) $\sup_{v \in V} \sqrt{n}\left\{\widehat{\boldsymbol{\sigma}}^2_K(v) - \boldsymbol{\sigma}^2_*(v)\right\} = O_p(1)$ *and* $\sqrt{n}\left\{\widehat{\boldsymbol{\sigma}}^2_K(v) - \boldsymbol{\sigma}^2_*(v)\right\} \Rightarrow_d \widehat{\lambda}_K(v)$, *where* $\widehat{\boldsymbol{\lambda}}_K(v) = \arg\min_{\boldsymbol{\lambda} \in [0,\infty) \times R^2} \{\boldsymbol{\lambda} - Z_K(v)\}^T \mathcal{I}_K(v) \{\boldsymbol{\lambda} - Z_K(v)\}$, *in which* $Z_K(v) = \mathcal{I}_K(v)^{-1} \mathcal{J}_K(v)$. *Moreover,* $\mathcal{J}_K(v)$ *denotes a random weighted score vector and* $\mathcal{I}_K(v)$ *is the average Fisher information matrix at* $\boldsymbol{\sigma}^2_*(v)$.

(ii) If $\sigma^2_{c*}(v) > 0$ *and* $\sigma^2_{e*}(v) > 0$, *then under local alternatives* $H_n : \sigma^2_a(v) = h(v)/\sqrt{n}$, *we*

have

$$\sqrt{n}\{\widehat{\sigma}^2_{a,K}(v) - \sigma^2_{a*}(v)\} \overset{H_n}{\Longrightarrow}_d \left\{H_K\mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T\right\}^{1/2} \times N(\widetilde{h}_K(v),1)\mathbf{1}\left(N(\widetilde{h}_K(v),1)\geq 0\right),$$

$$WLR_n(v) \overset{H_n}{\Longrightarrow}_d \frac{H_K\mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T}{H_K\mathcal{I}_K^{-1}(v)H_K^T} \times N(\widetilde{h}_K(v),1)^2\mathbf{1}\left(N(\widetilde{h}_K(v),1)\geq 0\right),$$

where $H_K = [1,0,0]$ *and* $\widetilde{h}_K(v) = \left\{H_K\mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T\right\}^{-1/2}h(v)$. *Moreover,* $\mathcal{I}_{1,K}(v)$ *is the covariance matrix of* $\mathcal{J}_K(v)$ *and* $\mathcal{I}_K(v) - \mathcal{I}_{1,K}(v)$ *is positive semi-definite.*

Theorem 2 (i) delineates the uniform convergence rate of $\widehat{\boldsymbol{\sigma}}^2_K(v)$ for all $v \in V$. Theorem 2 (ii) shows that when $\min(\sigma^2_{c*}(v), \sigma^2_{e*}(v)) > 0$ and $h(v) = 0$, $\widehat{\sigma}^2_{a,K}(v)$ is asymptotically a scaled truncated standard normal and the corresponding weighted likelihood ratio is a scaled $0.5\chi^2_1 + 0.5\chi^2_0$ as well with a scale parameter smaller than one. Since $\mathcal{I}_{1,K}(v) < \mathcal{I}_K(v)$ as long as $\sigma^2_{e*}(v) > 0$, the ratio of $H_K\mathcal{I}_K^{-1}(v)\mathcal{I}_{1,K}(v)\mathcal{I}_K^{-1}(v)H_K^T$ to $H_K\mathcal{I}_K^{-1}(v)H_K^T$ is smaller than one. When $h(v) > 0$, Theorem 2 (ii) characterizes the asymptotic local power of $\mathrm{WLR}_n(v)$. An important implication of Theorem 2 (ii) is that $\mathrm{WLR}_n(v)$ is statistically more powerful than $\mathrm{LR}_n(v)$, since $\widetilde{h}(v)$ is smaller than $\widetilde{h}_K(v)$ due to the fact that $\mathcal{I}_K(v) - \mathcal{I}_{1,K}(v)$ is nonnegative definite.

Theorem 3. *Suppose that Assumptions (C1)-(C8) hold. The following results hold:*

$$\sqrt{n}\left\{\widehat{\Sigma}_a(u,v) - \Sigma_a(u,v)\right\} \to_d N(0, W_a(u,v,u,v)),$$
$$\sqrt{n}\left\{\widehat{\Sigma}_c(u,v) - \Sigma_c(u,v)\right\} \to_d N(0, W_c(u,v,u,v)),$$
$$\sqrt{n}\left\{\widehat{\Sigma}_{e,G}(u,v) - \Sigma_{e,G}(u,v)\right\} \to_d N(0, W_{e,G}(u,v,u,v)),$$

where $W_a(u,v,u,v)$, $W_c(u,v,u,v)$, *and* $W_{e,G}(u,v,u,v)$ *are given in supplemental document.*

Theorem 3 shows the asymptotic normality of local constant covariance kernel estimators under ultra high dense situations. The ultra high dense condition is $n = o(N_G h_2)$, which means that the number of tract points with observations is extremely large compared to the sample size.

4 Simulation Studies

In this section, we present two simulation studies to evaluate the finite sample performance of the proposed estimation and inference procedures. We run all our simulations on our department linux cluster with configuration: DELL R815 Quad Processor AMD Opteron 16 core 2.3 GHz machines with 512GB RAM each running 64Bit Fedora Core 20. Our code is in Matlab. We use Matlab 2013a on the linux cluster. Each replication takes around 2.5 seconds. We define three signal to noise ratios (SNRs) as follows:

$$\begin{aligned}
\mathrm{SNR}_\beta &= \int \beta(v)dv \Big/ \left(\int \Sigma_a(v,v) + \Sigma_c(v,v) + \Sigma_e(v,v)dv \right), \\
\mathrm{SNR}_a &= \int \Sigma_a(v,v)dv \Big/ \int \Sigma_e(v,v)dv, \quad \mathrm{SNR}_c = \int \Sigma_c(v,v)dv \Big/ \int \Sigma_e(v,v)dv,
\end{aligned}$$

where $\Sigma_e(v,v) = \sigma_e^2(v) = \Sigma_{e,G}(v,v) + \Sigma_{e,L}(v,v)$. We design our simulation based on the SNRs obtained from normalized real data. Specifically, in real data, $(\mathrm{SNR}_\beta, \mathrm{SNR}_a, \mathrm{SNR}_c)$ is given by $(93.5, 1.5, 0.6)$ for FA, whereas it is equal to $(335.3, 1.5, 0.6)$ for MD. Although SNR_β is very large in real data, we reduce it in order to make it comparable with SNR_a and SNR_c. Table 1 summarizes $(\mathrm{SNR}_\beta, \mathrm{SNR}_a, \mathrm{SNR}_c)$ under different simulation settings. The functional forms of $\beta(v)$s are taken from Zhu et al. (2012a) and the eigenfunctions of the covariance kernels $\Sigma_a(v,v')$ and $\Sigma_c(v,v')$ are commonly used in the literature (Yao, 2007a; Yao et al., 2005; Yao, 2007b; Yao et al., 2005; Zhu et al., 2012a). Without loss of generality, the true $\Sigma_{e,G}(v,v')$ is set to be a zero function.

Example 1. The first set of simulations is designed to evaluate the estimation procedure. The data were generated from model (4), in which $x_{ij} = (1, x_{ij1}, x_{ij2})^T$ and

$$(x_{ij1}, x_{ij2})^T \sim N((0,0)^T, \mathrm{diag}(1-2^{-0.5}, 1-2^{-0.5}) + 2^{-0.5}(1,1)^{\otimes 2}).$$

Moreover, we set

$$a_{ij}(v_k) = \sum_{l=1}^{N^a} \xi_{ijl}\psi_l^a(v_k), \; a_i(v_k) = \sum_{l=1}^{N^a} \xi_{il}\psi_l^a(v_k), \text{ and } c_i(v_k) = \sum_{l=1}^{N^c} \zeta_{il}\psi_l^c(v_k),$$

where $\xi_{ijl} \sim N(0, \lambda_l^a)$, $\xi_{il} \sim N(0, \lambda_l^a)$, and $\zeta_{il} \sim N(0, \lambda_l^c)$, in which we set $N^a = N^c = 2$ and $(\lambda_1^a, \lambda_2^a) = (\lambda_1^c, \lambda_2^c) = c(1, 0.5)$ with $c = 0.05$ and 0.1. Moreover, $v_1, \ldots, v_{N_G}$ are equally spaced on $[0,1]$ and $N_G = 100$. Furthermore, we set the coefficient functions and

eigenfunctions as follows:

$$\beta_1(v) = v^2, \quad \beta_2(v) = (1-v)^2, \quad \beta_3(v) = 4v(1-v) - 0.4;$$
$$\psi_1^a = \psi_2^c = \sqrt{2}\sin(2\pi v), \quad \psi_2^a = \psi_1^c = \sqrt{2}\cos(2\pi v).$$

We set $n = 150$ (300), among which there are 50 (100) MZ twin pairs, 50(100) DZ twin pairs, and 50(100) single twin individuals. We run our simulations over all combinations of $n = (150, 300)$ and $(c, \Sigma_e) = (0.05, 0.05), (0.1, 0.1), (0.05, 0.1)$, and $(0.1, 0.2)$. We report the mean squared errors of parameter estimates over 400 replications.

As shown in Tables 2-4, the finite-sample performance of our estimators improves as sample size increases, whereas it decreases as SNR_β roughly drops from 1.5 ($c = 0.05$) to 0.75 ($c = 0.1$). The mean square errors of $\widehat{\sigma}^2_{a,K}(\cdot)$ and $\widehat{\sigma}^2_{c,K}(\cdot)$ are at the same scale since $\text{SNR}_a = \text{SNR}_c$ in all our settings of Example 1. Furthermore, Tables 2 and 3 show that the mean squared errors of maximum weighted likelihood estimators (MWLEs) are roughly half of those of the maximum likelihood estimators (MLEs). Figures 4 and 5 present the estimation results for $n = 300$, $c = 0.1$, and $\Sigma_e = 0.2$. This example shows that our estimates are quite accurate even for moderate sample sizes and are better than the maximum likelihood estimates.

Example 2. The second set of simulations is designed to evaluate Type I and II error rates of $\text{WLR}_n(v)$, and T_n. We aim to show that the inference performance of our procedure is good even when SNR_β is very small. The simulated data were generated in the same way as Example 1 except that we set $(\lambda_1^c, \lambda_2^c) = (0.5, 1)$, $\Sigma_e = 1, 2$ and $(\lambda_1^a, \lambda_2^a) = c(0.5, 1)$, where c (effect size) is a scalar specified below. We set $c = 0$ to assess the Type I error rate of various test statistics. We set $c = 0.1, 0.2, 0.3, 0.4$ in order to examine Type II error rate of $\text{WLR}_n(v)$ and set $c = 0.01, 0.02, 0.03, 0.04$ in order to examine Type II error rate of T_n at different effect sizes. For each simulation, the significance levels were set at $\alpha = 0.01$ and 0.05, respectively, and 400 replications were used to estimate the rejection rates.

Figures 6 and 7 present the Type I and II error rates along the entire tract of both $\text{LR}_n(v)$ and $\text{WLR}_n(v)$ for $\Sigma_e = 1$ and $\Sigma_e = 2$, respectively. As shown in Figures 6 and 7, the Type I error rates for both test statistics are quite accurate, whereas $\text{WLR}_n(v)$ is

statistically more powerful than $\mathrm{LR}_n(v)$ along the entire tract. This result is consistent with our theoretical power calculation in Theorem 2. Figures 8 presents the power function of T_n. They show that the rejection rates for the global test statistic based on the permutation method are accurate for moderate sample sizes ($n = 150$ or 300) at both significance levels $\alpha = 0.01$ and 0.05. Finally, the statistical power for rejecting the null hypothesis increases with sample size and SNR_a (determined by the effect size, and Σ_e), which is consistent with our expectation.

5 Real Data Analysis

The data set consists of 356 healthy twin neonates with 190 males and 166 females from the neonatal project as part of the UNC Early Brain Development Studies between 2004 and 2014. There are 129 twin pairs (48 MZ twin pairs and 81 DZ twin pairs) and 98 unrelated "singleton" twins - a single unpaired twin subject, in which a usable scan was not obtained from the co-twin. The gestational ages of these infants range from 257 to 401 days, and their mean gestational age is 289 days with standard deviation 18 days. The Diffussion Tensor Imaging (DTI) and T1-weighted images were acquired for each subject. For the DTIs, the imaging parameters were as follows: the six nonlinear directions at the b-value of 1000 s/mm^2 with a reference scan ($b = 0$), the isotropic voxel resolution = 2 mm, and the in-plane filed of view = 286×192 mm^2. A total of five repetitions were acquired to improve the signal-to-noise ratio of the DTIs.

The DTI data were processed by two key steps including a weighted least squared fitting method (Basser et al., 1994; Goodlett et al., 2009) to construct the diffusion tensors and a DTI atlas building pipeline (Goodlett et al., 2009; Zhu et al., 2011) to register DTIs from multiple subjects to create a study specific unbiased DTI atlas, to track fiber tracts in the atlas space and to propagate them back into each subject's native space by using registration information. Subsequently, diffusion tensors (DTs) and their scalar diffusion properties were calculated at each location along each individual fiber tract by using DTs in neighboring voxels close to the fiber tract. Figure 2 (a) and (b) display the diffusion properties (FA and MD) along the fiber bundle of the genu of the corpus callosum (GCC), which is an area of white matter in the brain. The GCC is the

anterior end of the corpus callosum, and is bent downward and backward in front of the septum pellucidum; diminishing rapidly in thickness, it is prolonged backward under the name of the rostrum, which is connected below with the lamina terminals. It was found that neonatal microstructural development of GCC was positively correlated with age and callosal thickness. Furthermore, twin DTI studies have reported high heritability of regional FA in the genu of corpus callosum (Pfefferbaum et al., 2001; Lee et al., 2015).

The aims of this analysis are to compare the diffusion properties including FA and MD along the GCC between male and female groups, to delineate the development of fiber diffusion properties across time, which is addressed by including the gestational age at MRI scanning as a covariate and, most importantly, to test the genetic effects along the GCC fiber tract. FA and MD, repectively, measure the inhomogeneous extent of local barriers to water diffusion and the averaged magnitude of local water diffusion. We fit model (4) to the FA and MD values from all 356 subjects, in which $x_{ij} = (1, \text{Gender}, \text{Age})^T$ and the number of grid points $N_G = 152$.

Figure 9 presents the estimated coefficient functions corresponding to 1, Gender and Age associated with FA and MD (blue solid lines in all panels of Figure 9). The intercept functions [panels (a) and (d) in Figure 9] describe the overall trend of FA and MD. The gender coefficients for FA and MD in Figure 9 (b) and (e) are negative at most of the grid points, which may indicate that compared with female infants, male infants have relatively smaller magnitudes of local water diffusivity along the genu of the corpus callosum. The gestational age coefficients for FA [panel (c) of Figure 9] are positive at most grid points, indicating that FA measures increase with age in both male and female infants, whereas those corresponding to MD [panel (f) of Figure 9] are negative at most grid points. This may indicate a negative correlation between the magnitudes of local water diffusivity and gestational age along the genu of the corpus callosum. These results are consistent with those in Zhu et al. (2012b). For FA, Figures 10 and 11 indicate that there are large genetic and environmental effects in the middle range of the genu tract. For MD, we can see large genetic and environmental effects near both ends of this tract. The p values of the global genetic effect tests on FA and MD measures are much smaller than 0.001 for T_n, indicating a significant global genetic effect.

Finally, we use 5-fold cross-validation to evaluate the estimation error of FSEM and

that of standard pixel-wise method (MLE). Specifically, we used the training set to estimate all the parameters and the genetic and environmental covariance operators and their corresponding basis functions. Then, for each subject in the test set, these basis functions were used to estimate the genetic and common environmental components. Figure 12 presents the estimated residual processes of four randomly selected subjects. Furthermore, the estimation error of FSEM and that of standard pixel-wise method (MLE) are, respectively, equal to 0.001 and 0.087.

6 Conclusions

We have developed a novel class of functional structural equation models (FSEMs) to dissect functional genetic and environmental effects on twin functional data. We have developed a three-stage estimation procedure to estimate varying coefficient functions for various covariates (e.g., gender) as well as three covariance operators for the genetic and environmental effects. We have developed an inference procedure based on weighted likelihood ratio statistics to test the genetic/environmental effect either at a fixed location or over a compact region. We have established the asymptotic properties of the estimated varying functions, the weighted likelihood ratio statistics, and the estimated covariance operators.

Many important issues need to be addressed in future research. First, although we focus on modeling twin functional response, it is interesting to extend FSEM to functional data obtained from family studies. Second, although it is assumed that both varying coefficient functions and covariance operators are smooth functions, it is possible to replace such smoothness condition by some piecewise smoothness condition (Li et al., 2012, 2011; Polzehl and Spokoiny, 2000). Finally, it is scientifically interesting to extend FSEM to carry out whole-genome analysis of twin functional data.

7 Appendix A: Assumptions

Before we present all the assumptions, we first introduce some notation. We set $\boldsymbol{x}_i = (x_{i1}^T, x_{i2}^T)^T$ and $\boldsymbol{\eta}_i(v) = (\eta_{i1}(v), \eta_{i2}(v))^T$, where $\eta_{ij}(v) = \sqrt{0.5}(\mathrm{DZ})_i a_{ij}(v) + [(\mathrm{MZ})_i +$

$\sqrt{0.5}(\mathrm{DZ})_i]a_i(v)+c_i(v)+e_{i,G}(v)$ for $1 \le i \le n_1+n_2$. We set $\boldsymbol{x}_i = (x_{i1})$ and $\boldsymbol{\eta}_i(v) = (\eta_{i1}(v))$ for $n_1+n_2+1 \le i \le n$. Denote $\sigma_1^2(v) = \sigma_a^2(v)+\sigma_c^2(v)+\sigma_e^2(v)$, $\sigma_2^2(v) = \sigma_a^2(v)+\sigma_c^2(v)$ and $\sigma_3^2(v) = 0.5\sigma_a^2(v)+\sigma_c^2(v)$. Without loss of generality, we assume that the parameter space for $\boldsymbol{\beta}(v)$ and $\boldsymbol{\sigma}^2(v)$ are R^p and $[0,\infty)^3$ respectively. Furthermore, we denote $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ as minimum and maximum eigenvalues of marix A.

Assumption (C0). $\boldsymbol{\beta}_*(v) \in \mathcal{B}$ where $\mathcal{B}$ is a open subset of R^p and $\boldsymbol{\sigma}_*^2(v) \in \mathcal{E}$ where $\mathcal{E}$ is a bounded subset of $[0,\infty)^3$. The $\boldsymbol{\Theta} = \mathcal{E} \times \mathcal{B}$ can be approximated by cone at $\boldsymbol{\theta}_*(v) = (\boldsymbol{\sigma}_*^2(v), \boldsymbol{\beta}_*(v))$ for each $v \in V$.

Assumption (C1). The number of parameters p is finite. n_1, n_2 and N_G increase to infinity. The weighted likelihood bandwidth h_1 and the covariance kernel smoothing bandwidth h_2 converges to zero. $n_1/n \to \alpha_1$, $n_2/n \to \alpha_2$, $n_3/n \to \alpha_3$ where $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 \ge 0$ and $\alpha_1+\alpha_2+\alpha_3 = 1$. Furthermore,

Assumption (C2). For MZ twin pairs, the covariates $\boldsymbol{x}_i = (x_{i1}, x_{i2})$ are independently and identically distributed with $E\boldsymbol{x}_i = \boldsymbol{\mu}_x^{(M)}$ and $\|\boldsymbol{x}_i\|_\infty < \infty$. Moreover, $E\left[\boldsymbol{x}_i\boldsymbol{x}_i^T\right] = \Omega_x^{(M)}$ is invertible. Similar assumptions for DZ twin pairs and single twins.

Assumption (C3). $a_{ij}(v)$, $a_i(v)$, $c_i(v)$, $e_{ij,G}(v)$ and $e_{ij,L}(v)$ are mutually independent copies of $GP(0,\Sigma_a)$, $GP(0,\Sigma_a)$, $GP(0,\Sigma_c)$, $GP(0,\Sigma_{e,G})$ and $GP(0,\Sigma_{e,L})$, respectively. Furthermore, $e_{ij,L}(v)$ and $e_{ij,L}(v')$ are independent for $v \ne v' \in V$. Moreover, $\Sigma_a(v,v')$, $\Sigma_c(v,v')$ and $\Sigma_{e,G}(v,v')$ have continuous second-order partial derivative with respect to $(v,v') \in V^2$ and $\Sigma_e(v,v) = \Sigma_{e,G}(v,v)+\Sigma_{e,L}(v,v)$ is Lipschitz continuous with a Lipschitz constant C_e with $\inf_{v\in V}\Sigma_e(v,v) > 0$.

Assumption (C4). Denote the Fisher information matrix at v as $\mathcal{I}(v)$,

$$0 < \inf_{v\in V}\lambda_{\min}(\mathcal{I}(v)) < \sup_{v\in V}\lambda_{\max}(\mathcal{I}(v)) < \infty.$$

Assumption (C5). Each component of $\{\boldsymbol{\eta}(v) : v \in V\}$, $\{\boldsymbol{\eta}(v)\boldsymbol{\eta}(v') : (v,v') \in V^2\}$, and $\{\boldsymbol{x}\boldsymbol{\eta}^T(v) : v \in V\}$ are Donsker classes.

Assumption (C6). The tract points in V_0 are randomly generated from a density function $\pi(v)$. Moreover, $\pi(v) > 0$ for all $v \in V$ and $\pi(v)$ has continuous second-order derivative with bounded support $V = [0,1]$.

Assumption (C7a). $\log^8(1+N_G)/n \to 0$.

Assumption (C7b). $\log^8(1+N_G)/n \to 0$, $nh_1^2 \to 0$, and $|\log(h_1)|/(N_G h_1) \to 0$.

Assumption (C8). $nh_2^4 \to 0$ and $n/(N_G h_2) \to 0$.

REMARKS. Assumption (C1) requires that n_1 and n_2 diverge to infinity at the same rate as n, whereas n_3 can be zero. Assumption (C2) is a relatively weak condition on the covariate vector, and the boundness is not essential. The Gaussian process requirement in Assumption (C3) is not necessary but for simplicity. Sub-Gaussian requirement is enough for most theoretical arguments. In this case, we have to assume that the Kullback-Leibler distance between the true underlying density and the proposed density family is zero in order to ensure the consistency of quasi-maximum likelihood estimators. The smoothness condition on the three covariance operators are quite general. Assumption (C4) is necessary for establishing the uniform consistency of related estimators and is satisfied as long as $\inf_{v\in V}\Sigma_e(v,v) > 0$ and $\sup_{v\in V}[\Sigma_a(v,v)+\Sigma_c(v,v)+\Sigma_e(v,v)] < \infty$ for Gaussian process. Assumption (C5) avoids specific smoothness condition on the sample path $\boldsymbol{\eta}(v)$, which are commonly assumed in the literature (Hall et al., 2006; Zhu et al., 2012b). Assumption (C6) is a weak condition on the random grid points. In many neuroimaging applications, N_G is quite large compared to sample size n and for such a large N_G, a regular grid of tract points is fairly well approximated by a simple uniform distribution in a compact region V. For notational simplicity, we only state the theoretical results for the random grid points. Assumptions (C7a) and (C7b) are fairly weak conditions on (N_G, n, h_1). Assumption (C8) is a ultra-dense condition, under which we can achieve $\sqrt{n}$ convergence rate for local constant estimators.

8 Appendix B: Simulations

8.1 Heritability

For better comparison, we rewrite Dr. Wang's (2011) functional mixed effects model for longitudinal family data as follows:

$$y_{ij}(v) = \mu(v) + \alpha_i + \mathbf{c}_{ij}(v)^T\beta + \eta_{ij}(v) + \epsilon_{ij}(v) \tag{13}$$

for $j = 1, \ldots, m_i$, where v is time in Wang (2011), $\mathbf{c}_{ij}(v)$ is a time-varying coefficient, α_i is a random family-specific shared environmental effect, $\eta_{ij}(v)$ is a random subject-specific genetic effect, and $\epsilon_{ij}(v)$ is a residual measurement error. Furthermore, in Wang (2001), she modeled $\eta_{ij}(v)$ as $\eta_{ij}(v) = B(v)^T \boldsymbol{b}_{ij}$, where $B(v)$ is a vector of spline basis and $\boldsymbol{b}_{ij}$ is the corresponding vector of subject-specific polygenic coefficients. Moreover, let $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^T, \ldots, \boldsymbol{b}_{im_i}^T)^T$, it is assumed that $\text{Cov}(\boldsymbol{b}_i, \boldsymbol{b}_i) = K_i \otimes \Omega$, where Ω is the unknown covariance matrix of the polygenic effects basis, $\otimes$ denotes the Kronecker and K_i is the known kinship coefficient matrix of the $i-$th family. Compared with (4), model (13) assumes that α_i and β are independent of v. Figure (1) shows that the method in Wang's (2011) is not able to correctly estimate the heritability curve when the common environment effect is not constant along the tract.

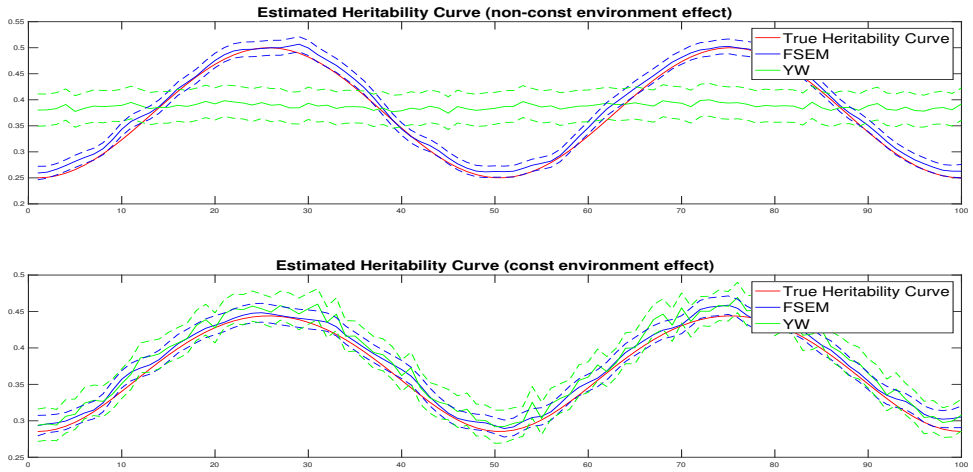

Figure 1: Estimated Heritability Curve when $n = 300, c = 0.1$, and $\Sigma_e = 0.2$.

References

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*(3), 683–734.

Basser, P. J., J. Mattiello, and D. LeBihan (1994). Estimation of the effective self-diffusion

tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B 103*(3), 247–254.

Bathia, N., Q. Yao, and F. Ziegelmann (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics 38*, 3352–3386.

Brouwer, R. M., R. C. Mandl, J. S. Peper, G. C. M. van Baal, R. S. Kahn, D. I. Boomsma, and H. E. H. Pol (2010). Heritability of dti and mtr in nine-year-old children. *NeuroImage 53*(3), 1085–1092.

Chen, K. and H.-G. Müller (2012). Modeling repeated functional observations. *Journal of the American Statistical Association 107*(500), 1599–1609.

Chiang, M.-C., K. L. McMahon, G. I. de Zubicaray, N. G. Martin, I. Hickie, A. W. Toga, M. J. Wright, and P. M. Thompson (2011). Genetics of white matter development: a dti study of 705 twins and their siblings aged 12 to 29. *NeuroImage 54*(3), 2308–2317.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*, Volume 66. CRC Press.

Fan, J. and W. Zhang (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface 2*.

Fang, Y. and Y. Wang (2009). Testing for familial aggregation of functional traits. *Statistics in medicine 28*(29), 3611–3625.

Feng, R., G. Zhou, M. Zhang, and H. Zhang (2009). Analysis of twin data using sas. *Biometrics 65*(2), 584–589.

Goodlett, C. B., P. T. Fletcher, J. H. Gilmore, and G. Gerig (2009). Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage 45*(1), S133–S142.

Gu, C. and P. Ma (2005). Optimal smoothing in nonparametric mixed-effect models. *Annals of Statistics 33*, 1357–1379.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*(1), 121–128.

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

Haseman, J. K. and R. C. Elston (1970). The estimation of genetic variance from twin data. *Behavior Genetics 1*, 11–19.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Huang, J. Z., C. O. Wu, and L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika 89*(1), 111–128.

Lee, S. J., R. J. Steiner, S. Luo, M. C. Neale, M. Styner, H. Zhu, and J. H. Gilmore (2015). Quantitative tract-based white matter heritability in twin neonates. *NeuroImage 111*, 123–135.

Lei, E., F. Yao, N. Heckman, and K. Meyer (2015). Functional data model for genetically related individuals with application to cow growth. *Journal of Computational and Graphical Statistics 24*, 756–770.

Li, Y., J. Gilmore, J. Wang, M. Styner, W. L. Lin, and H. T. Zhu (2012). Twinmarm: Two-stage multiscale adaptive regression methods of twin neuroimaging data. *IEEE Transactions on Medical Imaging 31*, 1100–1112.

Li, Y., N. Wang, and R. J. Carroll (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association 108*(504), 1284–1294.

Li, Y., H. Zhu, D. Shen, W. Lin, J. H. Gilmore, and J. G. Ibrahim (2011). Multiscale adaptive regression models for neuroimaging data. *Journal of the Royal Statistical Society: Series B 73*, 559–578.

Morris, J. (2015). Functional regression. *Annual Review of Statistics and Its Application 2*, 321–359.

Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B Stat. Methodol. 68*, 179–199.

Neale, M. C., A. C. Heath, J. K. Hewitt, L. J. Eaves, and D. W. Fulker (1989). Fitting genetic models with lisrel: Hypothesis testing. *Behavior Genetics 19*, 37–49.

Panizzon, M. S., C. Fennema-Notestine, L. T. Eyler, T. L. Jernigan, E. Prom-Wormley, M. Neale, K. Jacobson, M. J. Lyons, M. D. Grant, C. E. Franz, et al. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex 19*, 2728–35.

Peper, J. S., R. M. Brouwer, D. I. Boomsma, R. S. Kahn, H. Pol, and E. Hilleke (2007). Genetic influences on human brain structure: a review of brain imaging studies in twins. *Human Brain Mapping 28*(6), 464–473.

Pfefferbaum, A., E. V. Sullivan, and D. Carmelli (2001). Genetic regulation of regional microstructure of the corpus callosum in late life. *Neuroreport 12*(8), 1677–1681.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. IMS.

Polzehl, J. and V. G. Spokoiny (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Statist. Soc. B 62*, 335–354.

Posthuma, D., E. J. De Geus, W. F. Baaré, H. E. H. Pol, R. S. Kahn, and D. I. Boomsma (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience 5*(2), 83–84.

Rabe-Hesketh, S., A. Skrondal, and H. Gjessing (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics 64*(1), 280–288.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. New York: Springer.

Scheipl, F., A. Staicu, and S. Greven (2015). Functional additive mixed models. *Journal of Computational and Graphic Statistics 24*, 477–501.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*(398), 605–610.

Shi, J. Q. and T. Choi (2011). *Gaussian Process Regression Analysis for Functional Data*. CRC Press.

Thompson, P. M., T. D. Cannon, K. L. Narr, T. Van Erp, V.-P. Poutanen, M. Huttunen, J. Lönnqvist, C.-G. Standertskjöld-Nordenstam, J. Kaprio, M. Khaledy, et al. (2001). Genetic influences on brain structure. *Nature Neuroscience 4*(12), 1253–1258.

Wang, J. L., J. M. Chiou, and H. G. Mueller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*, 257–295.

Wang, X., X. Guo, M. He, and H. Zhang (2011). Statistical inference in mixed models and analysis of twin and family data. *Biometrics 67*, 987–995.

Wang, Y. (1998a). Mixed effects smoothing spline analysis of variance. *Journal of the royal statistical society: Series B (statistical methodology) 60*(1), 159–174.

Wang, Y. (1998b). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association 93*(441), 341–348.

Wang, Y. (2011). Flexible estimation of covariance function by penalized spline with application to longitudinal family data. *Statistics in medicine 30*(15), 1883–1897.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics 62*(4), 1025–1036.

Wood, S. N. (2013). A simple test for random effects in regression models. *Biometrika 100*(4), 1005–1010.

Wu, H. L. and J. T. Zhang (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, New Jersey.: John Wiley & Sons, Inc.

Yao, F. (2007a). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis 98*(1), 40–56.

Yao, F. (2007b). Functional principal component analysis for longitudinal and survival data. *Statistica Sinica 17*, 965–983.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Yao, F., H. G. Muller, and J. L. Wang (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*, 2873–2903.

Zhang, D., X. Lin, J. Raz, and M. Sowers (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association 93*(442), 710–719.

Zhang, J. and J. Chen (2007). Statistical inference for functional data. *The Annals of Statistics 35*, 1052–1079.

Zhu, H., P. Brown, and J. Morris (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Asssociation 106*, 1167–1179.

Zhu, H., L. Kong, R. Li, M. Styner, G. Gerig, W. Lin, and J. H. Gilmore (2011). Fadtts: functional analysis of diffusion tensor tract statistics. *NeuroImage 56*, 1412–1425.

Zhu, H., R. Li, and L. Kong (2012a). Multivariate varying coefficient model for functional responses. *Annals of Statistics 40*(5), 2634–2666.

Zhu, H. and H. Zhang (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B 66*(1), 3–16.

Zhu, H. T., R. Z. Li, and L. L. Kong (2012b). Multivariate varying coefficient model for functional responses. *Annals of Statistics 40*, 2634–2666.

Table 1: SNRs of all simulation settings

		SNR_β	SNR_a	SNR_c
$c = 0.05$	$\Sigma_e = 0.05$	1.7	1.5	1.5
	$\Sigma_e = 0.1$	1.4	0.7	0.7
$c = 0.1$	$\Sigma_e = 0.1$	0.85	1.5	1.5
	$\Sigma_e = 0.2$	0.7	0.7	0.7

Table 2: Mean squared errors $\times 1000$ (standard errors $\times 1000$ in brackets) of $\widehat{\sigma}_a^2(\cdot)$, $\widehat{\sigma}_c^2(\cdot)$ and $\widehat{\sigma}_e^2(\cdot)$ for all simulation settings

			$\text{MSE}(\widehat{\sigma}_a^2(\cdot))$	$\text{MSE}(\widehat{\sigma}_c^2(\cdot))$	$\text{MSE}(\widehat{\sigma}_e^2(\cdot))$
$n = 150$	$c = 0.05$	$\Sigma_e = 0.05$	1.40(0.01)	1.35(0.02)	0.10(0.00)
		$\Sigma_e = 0.1$	2.79(0.02)	2.23(0.02)	0.35(0.00)
	$c = 0.1$	$\Sigma_e = 0.1$	5.57(0.06)	5.11(0.07)	0.39(0.00)
		$\Sigma_e = 0.2$	11.14(0.09)	8.92(0.09)	1.40(0.01)
$n = 300$	$c = 0.05$	$\Sigma_e = 0.05$	0.76(0.01)	0.73(0.01)	0.05(0.00)
		$\Sigma_e = 0.1$	1.78(0.01)	1.38(0.01)	0.18(0.00)
	$c = 0.1$	$\Sigma_e = 0.1$	3.03(0.03)	2.91(0.05)	0.20(0.00)
		$\Sigma_e = 0.2$	7.11(0.06)	5.52(0.06)	0.73(0.01)

Table 3: Mean squared errors $\times 1000$ (standard errors $\times 1000$ in brackets) of $\widehat{\sigma}_{a,K}^2(\cdot)$, $\widehat{\sigma}_{c,K}^2(\cdot)$ and $\widehat{\sigma}_{e,K}^2(\cdot)$ for all simulation settings

			$\text{MSE}(\widehat{\sigma}_{a,K}^2(\cdot))$	$\text{MSE}(\widehat{\sigma}_{c,K}^2(\cdot))$	$\text{MSE}(\widehat{\sigma}_{e,K}^2(\cdot))$
$n = 150$	$c = 0.05$	$\Sigma_e = 0.05$	0.61(0.01)	0.76(0.02)	0.03(0.00)
		$\Sigma_e = 0.1$	1.31(0.02)	1.21(0.02)	0.11(0.00)
	$c = 0.1$	$\Sigma_e = 0.1$	2.46(0.06)	2.80(0.07)	0.11(0.00)
		$\Sigma_e = 0.2$	5.24(0.09)	4.82(0.09)	0.45(0.01)
$n = 300$	$c = 0.05$	$\Sigma_e = 0.05$	0.35(0.01)	0.42(0.01)	0.02(0.00)
		$\Sigma_e = 0.1$	0.84(0.01)	0.72(0.01)	0.07(0.00)
	$c = 0.1$	$\Sigma_e = 0.1$	1.41(0.03)	1.68(0.05)	0.07(0.00)
		$\Sigma_e = 0.2$	3.37(0.05)	2.88(0.06)	0.27(0.00)

Table 4: Mean squared errors $\times 1000$ (standard errors $\times 1000$ in brackets) of MSE$(\widehat{\beta}(\cdot))$, MSE$(\widehat{\Sigma}_a(\cdot,\cdot))$, MSE$(\widehat{\Sigma}_c(\cdot,\cdot))$ and MSE$(\widehat{\Sigma}_{e,G}(\cdot,\cdot))$ for all simulation settings

			MSE$(\widehat{\beta}(\cdot))$	MSE$(\widehat{\Sigma}_a(\cdot,\cdot))$	MSE$(\widehat{\Sigma}_c(\cdot,\cdot))$	MSE$(\widehat{\Sigma}_{e,G}(\cdot,\cdot))$
$n = 150$	$c = 0.05$	$\Sigma_e = 0.05$	1.02(0.02)	4.37(0.18)	2.51(0.10)	0.32(0.01)
		$\Sigma_e = 0.1$	1.48(0.02)	4.43(0.18)	2.55(0.10)	0.33(0.01)
	$c = 0.1$	$\Sigma_e = 0.1$	2.05(0.04)	16.50(0.65)	9.75(0.38)	1.22(0.05)
		$\Sigma_e = 0.2$	2.95(0.04)	17.70(0.74)	10.21(0.42)	1.31(0.05)
$n = 300$	$c = 0.05$	$\Sigma_e = 0.05$	0.50(0.01)	2.16(0.09)	1.28(0.05)	0.16(0.01)
		$\Sigma_e = 0.1$	0.75(0.01)	2.29(0.09)	1.33(0.06)	0.17(0.01)
	$c = 0.1$	$\Sigma_e = 0.1$	1.00(0.02)	8.64(0.37)	5.12(0.22)	0.64(0.03)
		$\Sigma_e = 0.2$	1.49(0.02)	9.14(0.37)	5.32(0.22)	0.67(0.03)

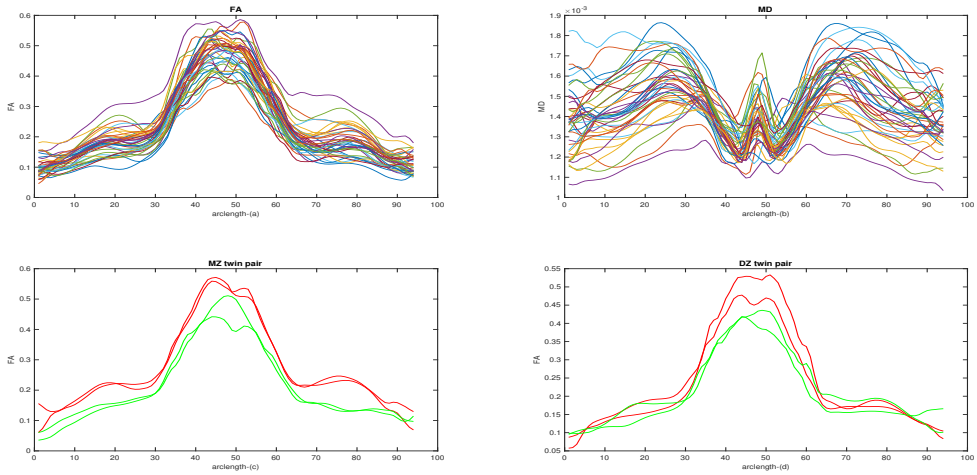

Figure 2: Representative twin neuroimaging data: [(a), (b)] are fractional anisotropy (FA) and mean diffusivity (MD) along the splenium tract of the corpus callosum from 40 randomly selected twin pairs; and [(c),(d)] are FA measures of two randomly selected MZ twin pairs and two randomly selected DZ twin pairs.

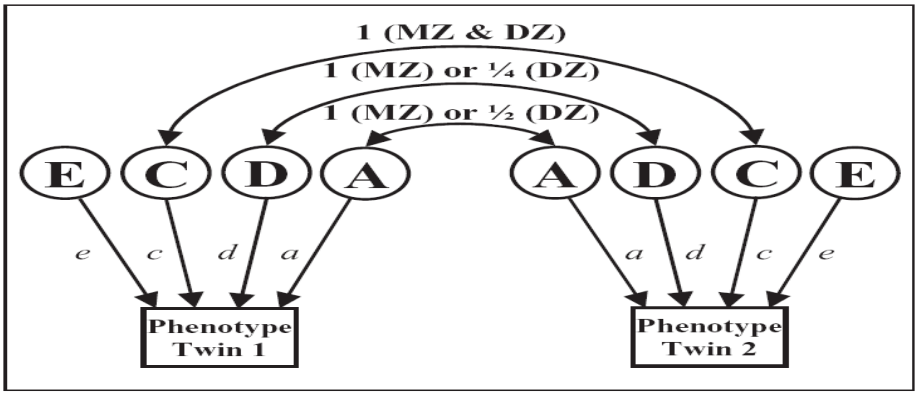

Figure 3: Diagram for the structural equation model for twin data. The correlation of additive effects (a1, a2) is 1 for MZ twin and 0.5 for DZ twin. The correlation of dominant effects (d1, d2) is 1 for MZ twin and 0.25 for DZ twin. The twin share the same common environmental effect (c). Residual effects (e1, e2) for twin are not correlated.

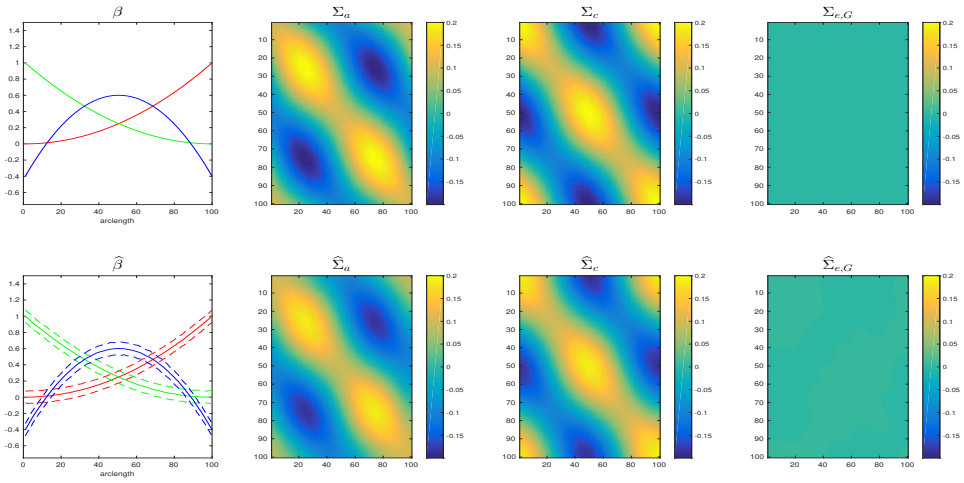

Figure 4: Simulation results for estimation performance at $n = 300, c = 0.1$, and $\Sigma_e = 0.2$: [(a), (b), (c)] are true coefficient functions, genetic and environmental covariance functions; [(d), (e), (f)] are the means of 400 corresponding estimators.

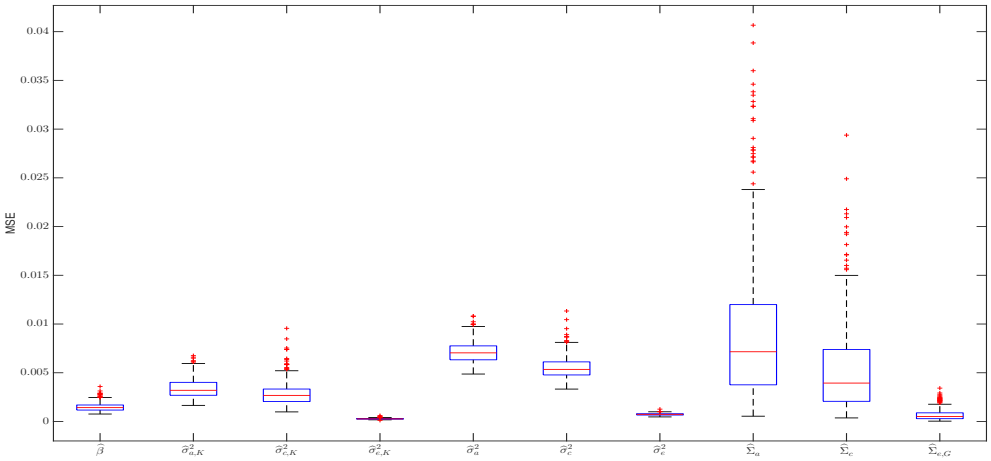

Figure 5: Box plot for mean squared errors of related estimators $n = 300, c = 0.1$, and $\Sigma_e = 0.2$.

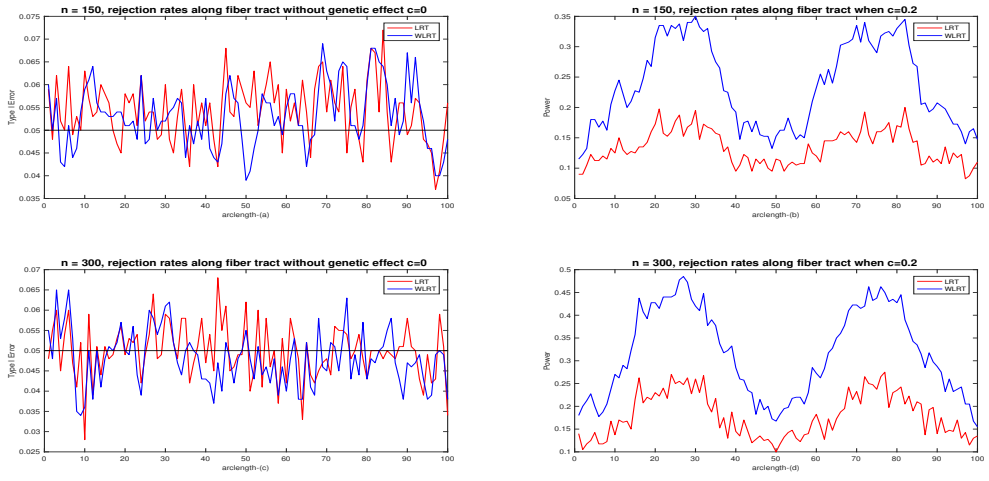

Figure 6: Inference Performance ($\Sigma_e = 1$): [(a),(c)] are rejection rates (type I error) of the two test statistics along fiber tract when $c = 0$ for $n = 150$ and $n = 300$; [(b),(d)] are rejection rates (power) of the two test statistics along fiber tract when $c = 0.2$ for $n = 150$ and $n = 300$.

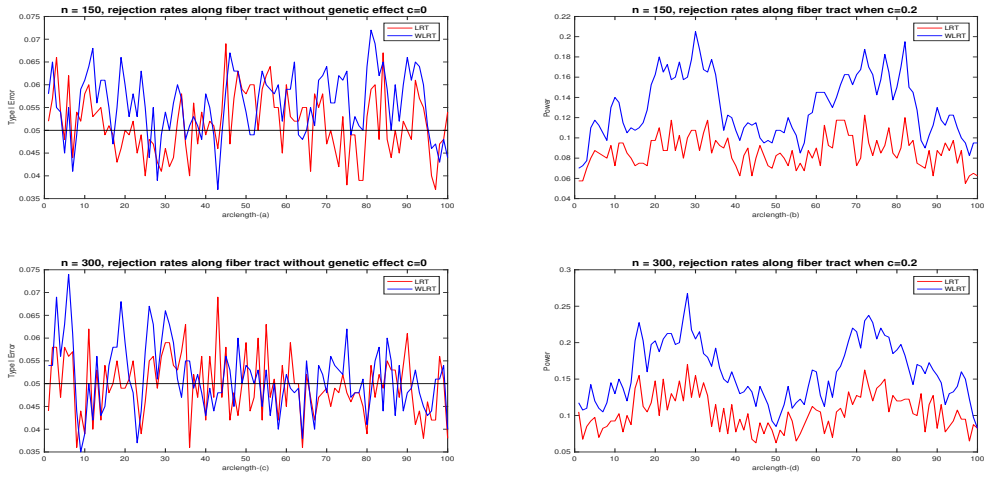

Figure 7: Inference Performance ($\Sigma_e = 2$): [(a),(c)] are rejection rates (type I error) of the two test statistics along fiber tract when $c = 0$ for $n = 150$ and $n = 300$; [(b),(d)] are rejection rates (power) of the two test statistics along fiber tract when $c = 0.2$ for $n = 150$ and $n = 300$.

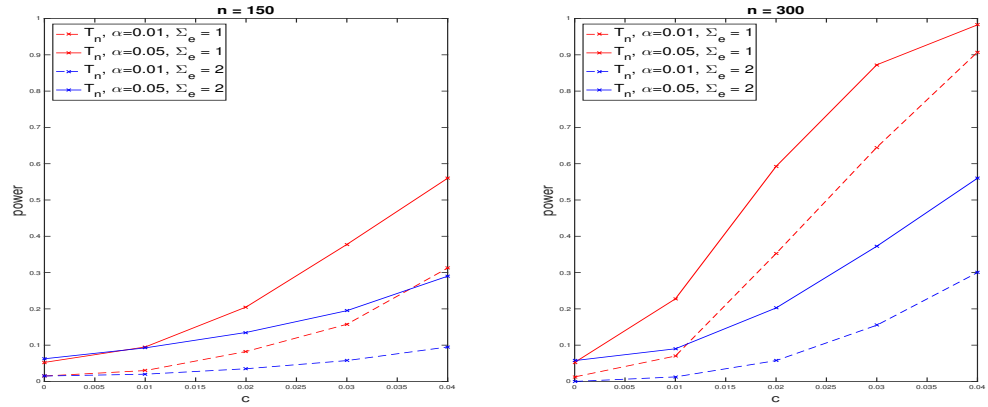

Figure 8: Plot of power curves ($\Sigma_e = 1, 2$). Rejection rates of T_n based on the permutation method are calculated at five different values of c $(0, 0.01, 0.02, 0.03, 0.04)$ for sample size of $n = 150, 300$ twin pairs (including singletons) at 5% (solid) and 1% (dashed) significance levels.

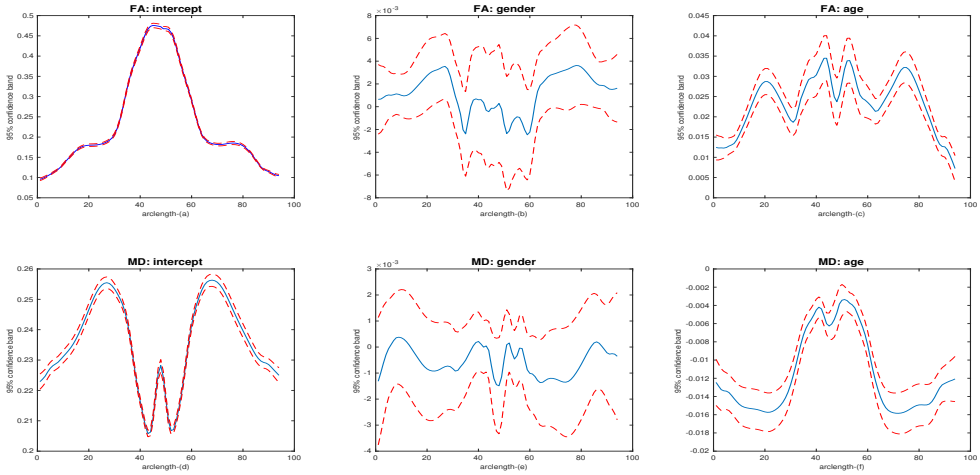

Figure 9: Plot of estimated cofficient functions of intercept [(a),(d)], gender [(b),(e)] and age [(c),(f)] and their pointwise 95% confidence bands. The first three panels [(a),(b),(c)] are for FA and the last three panels [(d),(e),(f)] are for MD. The blue solid curves are the estimated coefficient functions, and the red dashed curves are the confidence bands.

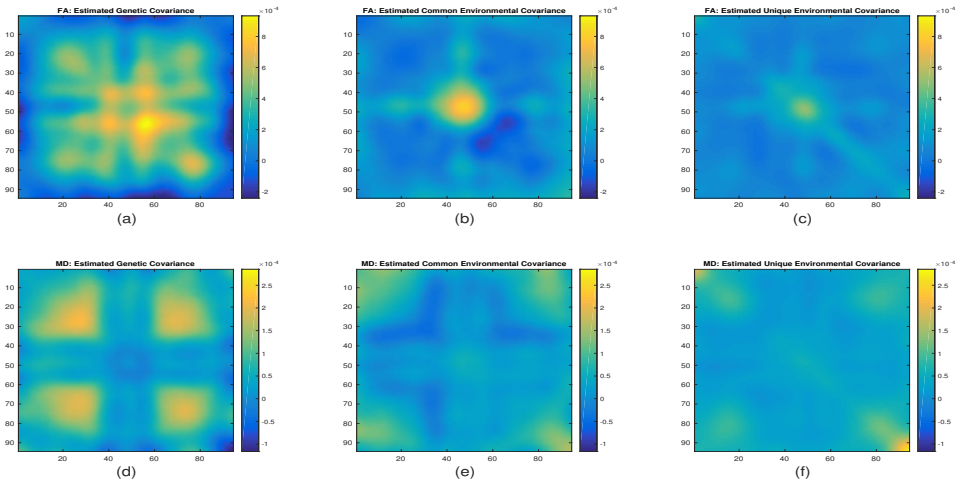

Figure 10: Plot of estimated covariance structure for genetic effect [(a),(d)], common environmental effect [(b),(e)] and unique environmental effect [(c),(f)]. The first three panels [(a),(b),(c)] are for FA and the last three [(d),(e),(f)] are for MD.

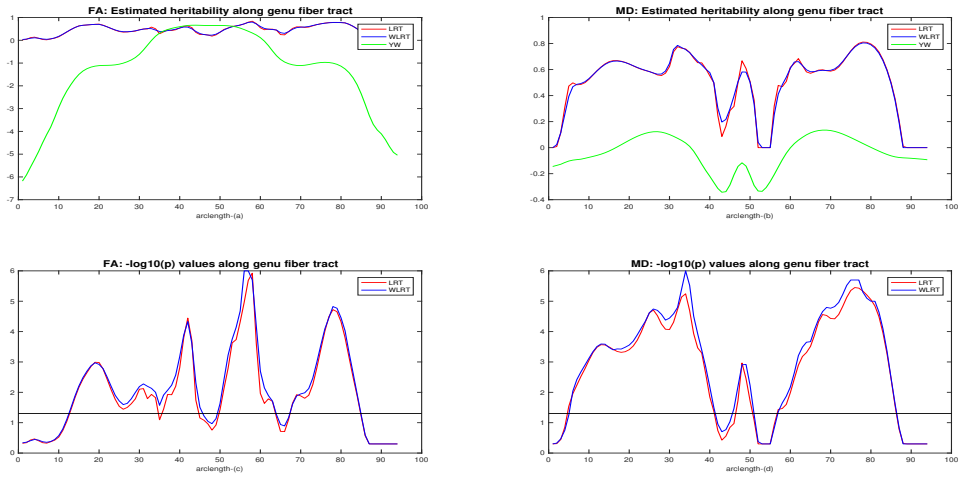

Figure 11: Plot of estimated heritability along fiber tract [(a),(b)] and the corresponding $-\log_{10}(p)$ values of tests for genetic variance existence. Panels [(a),(c)] are for FA and panels [(b),(d)] are for MD.

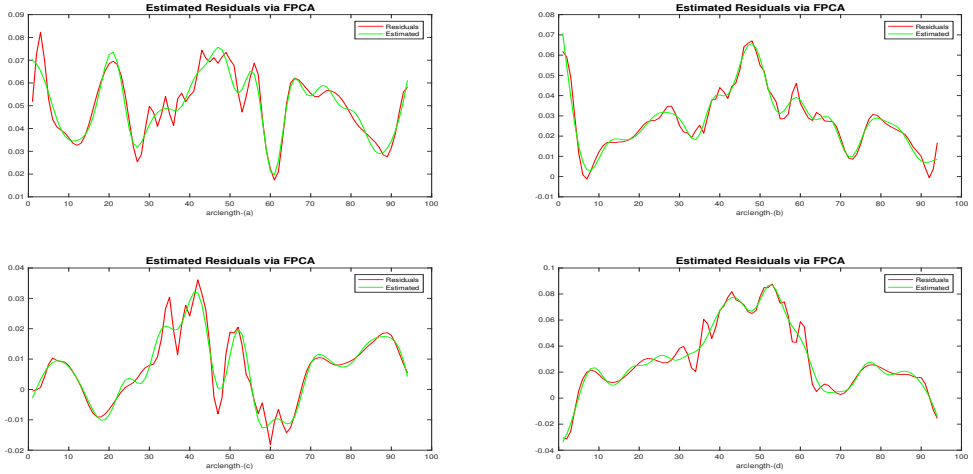

Figure 12: Plot of residual processes after removing $x_{ij}^T\hat{\boldsymbol{\beta}}(v)$ in FSEM versus the estimated ones based on FSEM from four randomly selected subjects.