A functional mixed effect model for scalar on function regression with application to a fMRI pain study

Wanying Ma, Bowen Liu, Luo Xiao, Martin A.Lindquist

September 11, 2018

1 Introduction

Recently, there has been a rapid emerge in the area of brain imaging study[9, 16, 23, 6]. Scientific interest has raised in the potential relationship between the continuously observed brain activity trajectories with one's subjective reaction to certain experiment conditions[8], results of the subject-level difference in the observed trajectories in the reaction, and thus how the brain activity interacts the psychological/physiological process. Equipped with instruments such as fMRI or EEG machine[17], a functional brain activity curve can be noninvasively observed on the equally spaced time grid $t_1, \ldots, t_p$ over time, that is, $t_{i+1} - t_i = t_{j+1} - t_j$, $\forall 1 \leq i, j \leq p$. Functional magnetic resonance imaging, or functional MRI(fMRI) measures the brain activity noninvasively by detecting the change in the blood flow associated with the neuron. During the entire course of fMRI, a sequence of magnetic resonance imaging(MRI) are performed, with a number of volume elements, or voxels, placed evenly over the brain[9]. At the mean time, subjects can be asked for a set of tasks in the brain imaging experiment. Therefore, fMRI provides people with both spatial and temporal resolution for brain mapping function in response for the experimental task, and facilitate people with a non-surgery technique for the neuron activity research purpose[1]. Researchers have gained lots of insights in the connections between brain activity function and psychological/physiological reaction through modeling fMRI data. In this work, we mainly focus on exploring the effects of subject-level brain

activity function on self-reporting response in the frame work of functional data analysis.

Functional data analysis plays an important role in dealing with densely observed data in the observational studies and clinical trials in the past decades[4, 3, 12, 14, 20, 21, 22, 10, 5, 19]. Monitored by the development of technology, one can observe full trajectories with repeated measurements during a time interval or at several discrete time points[2, 19, 4, 18]. The methods and techniques in the development of functional data analysis have been extensitively overviewed [11, 14, 17]. Functional principal component analysis (FPCA) as a main tool in functional data analysis for dimension reduction and identifying the main directions in the covariance function has a long history[13, 7, 15, 22, 3, 23];

There has been a vast amount of applications using functional data analysis. Functional linear regressionFew methods has been developed to make good use of fully observed fMRI trajectory in stead of summarized statistics in the past decades[1, 8]. These immense data reduction leads to an obscure picture on the brain activity change over time, and thus leaves the connection between subject's experimental response not well explained. The development of functional data analysis Functional data analysis has great application in the area of physical activity dataMartin[8] incorporated the functional data analysis settings into the standard mediation model(SME), established a linear functional standard mediation model(lfSME), being able to identifiably explain the causal effects from the parameters in lfSME under certain conditions.

The remainder of this paper is organized as follows. Section 2 will explain the experiment for aggregating the data, explicitly describe steps on developing the functional linear mixed model and test procedures. Corresponding simulation study under different cases will be given in Section 3. The proposed methodology to a real data application on fMRI dataset will be given in Section 4.

2 Data and Method

2.1 Data from a fMRI pain study

The following description of the fMRI data is mostly based on Martin's recent JASA paper [8] on the functional data analysis. In that experiment, a heat

stimuli was applied at one of two different levels (high and low) to each of 20 subjects at each repetition for 18 seconds. In 14 seconds later than the stimulus each subject was asked to report the subjective pain rating and responded in a few seconds. During the entire course of the experiment, each subject's brain activity was measured by fMRI method. Therefore, functional MRI data was sampled from 21 diverse classic pain-responsive brain regions. Each sample regiment had 23 equidistant temporal measurements made every 2 seconds, totally covering a 46-second brain activity from the application of the heat stimuli to the subjective pain report. The same experiment was conducted on each subject multiple times but the number of the repetitions is unbalanced design ranging from 39 to 48.

Based on the description above, the matrix $M_{I\times J}$ denotes the design matrix of Brain activation data sampled at a particular voxel on 20 interviewees with 39-48 replication each during 46-second fMRI experiment in 23 sampling time points.

$$M_{I\times J} = \begin{pmatrix} M_{1,1}(t_1) & M_{1,1}(t_2) & \cdots & M_{1,1}(t_{23}) \\ M_{1,2}(t_1) & M_{1,2}(t_2) & \cdots & M_{1,2}(t_{23}) \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,J_n}(t_1) & M_{n,J_n}(t_2) & \cdots & M_{n,J_n}(t_{23}) \end{pmatrix}$$

Without the consideration on the subjective random error in the population, denote y_{ij} as the subjective rating at ith subject in jth replication, Z_{ij} as the indicator variable of thermal stimuli level at ith subject in jth replication whose value 1 for the high level and 0 for the low level, δ and γ as the parameters of the general linear model, $M_{ij}(t)$ as the temporal functional covariate of the brain activation at a particular voxel in fMRI data set, $\beta(t)$ as the corresponding functional curve and ϵ_{ij} as the white error. The mean mixed model is described as $y_{ij} = \delta + \gamma Z_{ij} + \int \beta(t)M_{ij}(t)\,\mathrm{d}t + \epsilon_{ij}$, where ϵ_{ij} is i.i.d. across i and j and $\epsilon_{ij} \sim N(0, \sigma^2)$.

However, based on our previous knowledge on that fMRI data across experimental interviewees, the individual random effect accounts for the causal effect on the thermal pain rating across different subjects. In order to explain the subjective error drawn from the population we consider, the individual-level random error terms are affiliated to each fixed parameter and parameter curve. Therefore, the extended functional mixed effect model is formed as the following.

The underlying distribution on y_{ij} based on the model we formulate is

of our interest, but the integration on the temporal random effect curve is the obstacle in the routine of functional regression analysis. On approach in fitting the functional mixed model is to assemble the sampling distribution and the conditional distribution $y_{ij}|\beta_i$ which maintains a simple form and relatively easy to be estimated.

Based on the assumption above on $E(\beta_i) = 0$ and $COV[\beta_i(s), \beta_i(t)] = K(s,t)$, the complex structure of the functional covariance results in great computational burden in maximum likelihood method and large bias in method of moment. In order to simplify the model, the assumption on the functional covariance should be made.

Luo: I started working from here.

3 Functional mixed effects model for scalar on function regression with repeated outcomes

To model subject-specific random effect of a functional predictor, we propose a novel functional mixed effect model extending the scalar on function linear regression for repeated outcomes. The proposed model is

$$Y_{ij} = \alpha + Z_{ij}(\gamma + \gamma_i) + \int \{\beta(t) + \beta_i(t)\} X_{ij}(t)\, \mathrm{d}t + \epsilon_{ij}, \tag{1}$$

where α is the population intercept, γ is the population effect of Z_{ij}, γ_i is the random subject-specific effect of Z_{ij} for subject i, $\beta(\cdot)$ is the population effect of the functional predictor $M_{ij}(t)$, $\beta_i(\cdot)$ is the random subject-specific effect of the functional predictor, and ϵ_{ij}s are independently and identically distributed (i.i.d.) random errors with distribution $\mathcal{N}(0, \sigma_\epsilon^2)$. We assume that γ_is are i.i.d. with distribution $\mathcal{N}(0, \sigma_\gamma^2)$, $\beta_i(\cdot)$s are i.i.d. random functions following a Gaussian process over $\mathcal{T}$ with mean function $\mathbb{E}\{\beta_i(t)\} = 0$ and covariance function $\mathrm{cov}\{\beta_i(s), \beta_i(t)\} = \mathcal{C}(s,t)$, and all random terms are independent across subjects and from each other.

3.1 Model estimation

The key idea is to reduce model (1) to a linear mixed effects model using the functional principal component analysis (fPCA) of the functional predictor.

Specifically, assume that $X_{ij}(\cdot)$ are independent random functions from a Gaussian process with a mean function $\mathbb{E}\{X_{ij}(t)\} = \mu(t)$ and covariance function $\text{cov}\{X_{ij}(s), X_{ij}(t)\} = \mathcal{K}(s,t)$. The observed functional predictor might be contaminated with measurement errors. Thus, we consider the model for the functional predictor

$$W_{ijk} = X_{ij}(t_k) + e_{ijk}, \tag{2}$$

where W_{ijk} is the observation at time t_k for subject i during visit j and e_{ijk}s are measurement errors that are independent across i, j and k and are independent from the true random functions $X_{ij}(\cdot)$s. Note that model (2) has a two-level nested structure, and a more complicated fPCA method such as the multilevel fPCA [3] may be employed. For simplicity, we assume that $X_{ij}(\cdot)$ are independent across subjects and visits. By Meyer's theorem, $\mathcal{K}(s,t)$ can be decomposed as $\sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are non-increasing eigenvalues with associated eigenfunctions $\phi_k(\cdot)$s that satisfy $\int_{\mathcal{T}} \phi_k(s)\phi_\ell(s)ds = 1_{\{k=\ell\}}$. Here $1_{\{\cdot\}}$ is 1 if the statement inside the bracket is true and 0 otherwise. Then by the Karhunen-Loeve expansion, $X_{ij}(\cdot)$ can be written as a linear combination of the eigenfunctions, i.e., $X_{ij}(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ijk}\phi_k(t)$, where ξ_{ijk}s are independent random scores with $\xi_{ijk} \sim \mathcal{N}(0, \lambda_k)$. The fPCA can be conducted using the fast covariance estimation (FACE) method [21], which is based on penalized splines and is implemented in the R function "fpca.face" in the R package *refund*. Then, we obtain estimates of the eigenfunctions, $\hat{\phi}_k(\cdot)$, and estimates of the eigenvalues, $\hat{\lambda}_k$. The random scores ξ_{ijk} can also be predicted, denoted by $\hat{\xi}_{ijk}$; see [21] for more details.

For model identifiability, we assume that $\beta(\cdot)$ can also be written as a linear combination of the eigenfunctions so that $\beta(t) = \sum_{k=1}^{\infty} \theta_k \phi_k(t)$, where θ_ks are associated scalar coefficients to be determined. Similarly, let $\beta_i(\cdot) = \sum_{k=1}^{\infty} \theta_{ik}\phi_k(t)$, where θ_{ik}s are independent subject-specific random coefficients with distribution $\mathcal{N}(0, \tau_k^2)$, where $\tau_k^2 \geq 0$ are to be determined as well. Note that the induced covariance function $\mathcal{C}(s,t)$ equals $\sum_{k\geq 1} \tau_k^2 \phi_k(s)\phi_k(t)$. It follows that model (1) can be rewritten as

$$Y_{ij} = \alpha + Z_{ij}(\gamma + \gamma_i) + \sum_{k=1}^{\infty} \xi_{ijk}(\theta_k + \theta_{ik}) + \epsilon_{ij}. \tag{3}$$

Model (3) has infinitely many parameters and hence can be not be fitted, a well known problem for scalar on function regression. We follow the standard

approach by truncating the number of eigenfunctions for approximating the functional predictor, so that the associated scores and parameters for $\beta(\cdot)$ and $\beta_i(\cdot)$ are all finite dimensional. Specifically, let K be the number of eigenfunctions to be selected. Then an approximate and identifiable model is

$$Y_{ij} = \alpha + Z_{ij}(\gamma + \gamma_i) + \sum_{k=1}^{K} \xi_{ijk}(\theta_k + \theta_{ik}) + \epsilon_{ij}. \tag{4}$$

Model (4) is a linear mixed effects model and can be easily fitted by standard software.

In practice, the number of eigenfunctions K is selected by AIC [22]. Denote the selected number by $\hat{K}$. Then a practical model for (4) is

$$Y_{ij} = \alpha + Z_{ij}(\gamma + \gamma_i) + \sum_{k=1}^{\hat{K}} \hat{\xi}_{ijk}(\theta_k + \theta_{ik}) + \epsilon_{ij}. \tag{5}$$

Denote the corresponding estimates of θ_k by $\hat{\theta}_k$ and the prediction of θ_{ik} by $\hat{\theta}_{ik}$. Then, $\hat{\beta}(t) = \sum_{k=1}^{\hat{K}} \hat{\theta}_k \hat{\phi}_k(t)$ and $\hat{\beta}_i(t) = \sum_{k=1}^{\hat{K}} \hat{\theta}_{ik} \hat{\phi}_k(t)$.

4 Test of random functional effect

The interest is to assess if the functional effect is subject specific or the same across subjects. In other words, if $\beta_i(t) = 0$ for all i and $t \in \mathcal{T}$ in model (1) or $\beta_i(t) \neq 0$ for some i at some $t \in \mathcal{T}$. Because $\beta_i(\cdot)$s are random coefficient functions, the test can be formulated in terms of its covariance function. The null hypothesis is $H_0 : \mathcal{C}(s,t) = 0$ for all $(s,t) \in \mathcal{T}^2$ and the alternative hypothesis is $H_a : \mathcal{C}(s,t) \neq 0$ for some $(s,t) \in \mathcal{T}^2$. Under H_0, $\beta_i(t) = 0$ for all i and $t \in \mathcal{T}$ and model (1) reduces to a standard scalar on function linear regression model. Testing zeroness of a covariance function is nonstandard and our ideas are as follows. First note that since $\mathcal{C}(s,t) = \sum_{k \geq 1} \tau_k^2 \phi_k(s) \phi_k(t)$, an equivalent test is $H_0' : \tau_k^2 = 0$ for all k against $H_a' : \tau_k^2 > 0$ for at least one k. Following [10], we further simplify the test by making the assumption that

$$\tau_k^2 = \tau^2 \text{ for all } k, \tag{6}$$

and consider the corresponding test $\tilde{H}_0 : \tau^2 = 0$ against $\tilde{H}_a : \tau^2 \neq 0$. Note that under H_0, $\tilde{H}_0$ still holds. While H_a is much more general than $\tilde{H}_a$,

simultaneous test of zeroness of multiple variance components is challenging. While existing tests can be implemented, as will be illustrated in the simulation studies, these tests do not even maintain proper sizes under the null hypothesis. On the contrary, the proposed test of testing zeroness of one variance component maintains proper size and has good power.

Another issue is that standard testing procedure such as LRT/RLRT is not applicable to model (4) because the model has multiple additive random slopes. In the next subsection, we transform (4) into an equivalent mixed effect model, which has only one random slope term and can be easily tested.

4.1 Equivalent model formulation of (4)

Under the assumption (6), the random effects and random errors are independent from each other and satisfy the following distribution assumptions:

$$\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2),\ \theta_{ik} \sim \mathcal{N}(0, \tau^2), \epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{7}$$

The goal of the equivalent model formulation is to convert a set of homoscedastic random subject-specific slopes in (4) into one simple random slope, so that the test on homoscedastic random slopes can be conducted using standard software.

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$, $\mathbf{A}_i = (\xi_{ijk})_{jk} \in \mathbb{R}^{J_i \times K}$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iJ_i})^\mathsf{T} \in \mathbb{R}^{J_i}$. Also let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\mathsf{T} \in \mathbb{R}^K$ and $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iK})^\mathsf{T} \in \mathbb{R}^K$. Then model (4) can be written in the matrix form

$$\mathbf{Y}_i = \delta \mathbf{1}_{J_i} + \mathbf{Z}_i(\gamma + \gamma_i) + \mathbf{A}_i(\boldsymbol{\theta} + \boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_i.$$

Let $\boldsymbol{\Delta}_i = \begin{pmatrix} \mathbf{1}_{J_i} & \mathbf{Z}_i & \mathbf{A}_i \end{pmatrix} \in \mathbb{R}^{J_i \times (2+K)}$ and $\boldsymbol{\alpha} = (\delta, \gamma, \boldsymbol{\theta}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{2+K}$. It follows that

$$\mathbf{Y}_i = \boldsymbol{\Delta}_i \boldsymbol{\alpha} + \gamma_i \mathbf{Z}_i + \mathbf{A}_i \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i. \tag{8}$$

Let $\mathbf{U}_i \mathbf{D}_i^{\frac{1}{2}} \mathbf{V}_i^\mathsf{T}$ be the singular value decomposition of $\mathbf{A}_i$, where $\mathbf{U}_i^\mathsf{T}\mathbf{U}_i = \mathbf{I}_{J_i}$, $\mathbf{V}_i^\mathsf{T}\mathbf{V}_i = \mathbf{I}_{J_i}$, and $\mathbf{D}_i = \mathrm{diag}(d_{i1}, \ldots, d_{iJ_i})$ is a diagonal matrix of the singular values of $\mathbf{A}_i$. Let $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iJ_i})^\mathsf{T} = \mathbf{U}_i^\mathsf{T}\mathbf{Y}_i \in \mathbb{R}^{J_i}$, $\tilde{\boldsymbol{\theta}}_i = (\tilde{\theta}_{i1}, \ldots, \tilde{\theta}_{iJ_i})^\mathsf{T} = \mathbf{V}_i^\mathsf{T}\boldsymbol{\theta}_i$, and $\tilde{\boldsymbol{\epsilon}}_i = (\tilde{\epsilon}_{i1}, \ldots, \tilde{\epsilon}_{iJ_i})^\mathsf{T} = \mathbf{U}_i^\mathsf{T}\boldsymbol{\epsilon}_i$. Then a left multiplication of (8) by $\mathbf{U}_i^\mathsf{T}$ gives

$$\tilde{\mathbf{Y}}_i = (\mathbf{U}_i^\mathsf{T}\boldsymbol{\Delta}_i)\boldsymbol{\alpha} + (\mathbf{U}_i^\mathsf{T}\mathbf{Z}_i)\gamma_i + \mathbf{D}_i^{\frac{1}{2}}\tilde{\boldsymbol{\theta}}_i + \tilde{\boldsymbol{\epsilon}}_i,$$

or equivalently,

$$\tilde{Y}_{ij} = (\mathbf{U}_{ij}^\mathsf{T}\boldsymbol{\Delta}_i)\boldsymbol{\alpha} + (\mathbf{U}_{ij}^\mathsf{T}\mathbf{Z}_i)\gamma_i + \sqrt{d_{ij}}\tilde{\theta}_{ij} + \tilde{\epsilon}_{ij}, \tag{9}$$

where $\mathbf{U}_{ij}$ is the jth column of $\mathbf{U}_i$. The specification (7) now becomes $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$, $\tilde{\theta}_{ij} \sim \mathcal{N}(0, \tau^2)$, $\tilde{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and the random terms are independent across i and j, and are independent from each other. Model (9) can be fitted by standard mixed model using the R function *lmer* and then the test of $\tau^2 = 0$ can be conducted by the LRT/RLRT test.

5 A simulation study

In this section we conduct simulations to illustrate the performance of our proposed functional mixed model and demonstrate that our proposed test maintains size and has good power.

5.1 Simulation setting

Let $\mathcal{T} = [0, 1]$. Each simulated data has I subjects with each subject having J visits. We generate the response Y_{ij} from model (4) with $K = 3$, $\alpha = 0.5$, $\gamma = 2$, $\gamma_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $Z_{ij} \overset{i.i.d.}{\sim}$ Bernoulli(0.5), $\xi_{ijk} \overset{i.i.d.}{\sim} \mathcal{N}(0, \lambda_k)$, $\theta = 2$, $\theta_{ik} \overset{i.i.d.}{\sim} \mathcal{N}(0, \tau_i^2)$, and $\epsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Here, $\lambda_k = 0.5^k, k = 1, \ldots, K$. Then the noisy functional data is generated from model (2) with $X_{ij}(t) = \sum_{k=1}^{K} \xi_{ijk}\phi_k(t)$ and $e_{ijk} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_W^2)$. Here, $\phi_1(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2(t) = \sqrt{2}\cos(4\pi t)$,$\phi_3(t) = \sqrt{2}\sin(4\pi t)$ and σ_W^2 is chosen so that the signal to noise ratio in the functional data $r = \sigma_W^{-2} \int_\tau \mathcal{K}(t, t)\mathrm{d}t$ equals either 0 or 3. Note that $r = 0$ corresponds to smooth functional data without noises and $r = 3$ corresponds to noisy functional data. Finally homogeneous τ_i^2 and heteroscedastic τ_i^2 are considered for variance of interest. That is, we consider two different scenarios for generating random scores θ_{ik}:(1)Homogeneous variance: $\tau_1^2 = \cdots = \tau_K^2 \equiv \tau^2$;(2) $\tau_i^2 = \frac{1}{2^{i-1}}\tau^2, i = 1, \ldots, K$.

Hereby, we simulate data using a factorial design with four factors: the number of subject I, the number of visits per subject J, the signal to noise ratio r in the functional data, and homogeneity of random score θ_{ik}. A total of 24 different model conditions are used: $\{(I, J, r) : I \in \{20, 50, 200\}, J \in \{20, 50\}, r \in \{0, 3\}\}$ with random scores are being homogenous or heteroscedastic. Under each model condition, 20000 datasets are simulated.

To test $\tilde{H}_0 : \tau^2 = 0$ versus $\tilde{H}_a : \tau^2 > 0$, $\tau^2 = 0$ is used to generate response under the null hypothesis, and different nonzero τ^2s (assuming either homogeneity or heteroscedasity) under the alternative hypothesis are used to

generate power curves. Testing method using Bonferroni correction following Mcleans paper is compared with our proposed testing method.

5.2 Results on test

Luo: I stopped here.

Table 1: Sizes of test at the 5% level compared with Bonferroni testing.

Model condition	Equal variance		bonferroni	
	$r = 0$	$r = 3$	$r = 0$	$r = 3$
I = 20, J = 20	0.05	0.05	0.049	0.050
I = 20, J = 50	0.051	0.05	0.048	0.049
I = 50, J = 20	0.051	0.051	0.051	0.049
I = 50, J = 50	0.05	0.050	0.049	0.048
I = 200, J = 20	0.049	0.049	0.053	0.050
I = 200, J = 50	0.051	0.051	0.047	0.048

Table 1 provides with the size test results based on 20000 times simulation under full combinations of the simulation design $I \in \{20, 50, 200\}, J \in \{20, 50\}\}$ with $\hat{M}_{I\times J}$ being true and noise added, and interested testing target θ_{ik} are truly homogeneous or heteroscedastic. As we can see, the test p-value under each model condition is nearly 0.05, which demonstrate that both our test procedure and bonferroni testing can well control false negative rate around 0.05, and maintain stable performance over change on sample size I, visit repeats J, and noise added to the functional corariates.

In Figure 1 power curves using both true and noisy $\hat{M}_{I\times J}$ across full combinations of subject number I and visits J are illustrated by our testing method and Bonferroni testing. Power curve is generated at each given value of common variance τ^2. One can find that as subject number and visit increase, power of test will be correspondingly increased, and is able to converge to 1 eventually under each model condiction. Test of using true $\hat{M}_{I\times J}$ is always more powerful than the case when measurement error are considered, i.e. noise added $\hat{M}_{I\times J}$. If we fix the number of subjects and visits, one can easily find that our proposed method gives more powerful testing results than bonferroni's correction, as the power given by our method using noisy $\hat{M}_{I\times J}$ is always greater than Bonferroni testing using true functional data

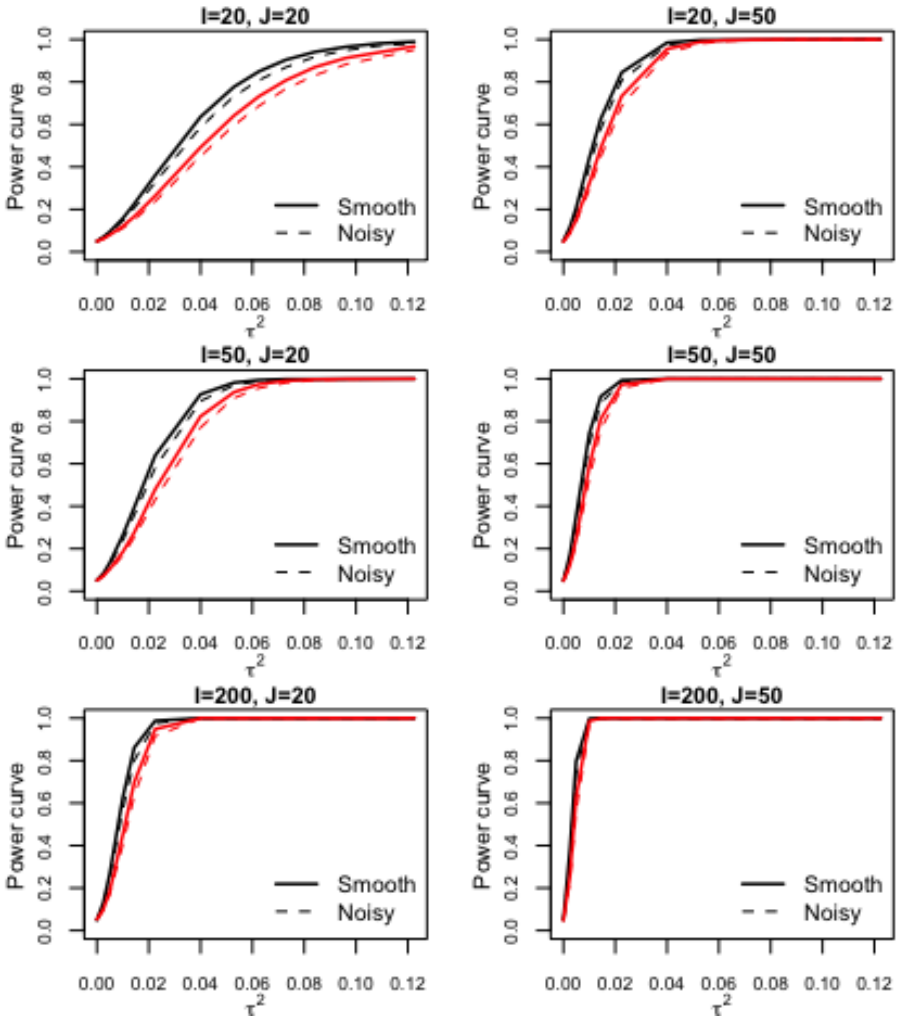


Figure 1: Power of test ($\tau_k^2 \equiv \tau^2$)at the 5% significance level under 12 model conditions. (1) In each panel, solid line represents that true functional data are used; dashed line represents the functional data with measurement error are used. (2) Black line denotes the power curve given by our proposed testing method assuming homogeneity of random scores; red line denotes the power curve generated by Bonferroni's correction testing method.

$\hat{M}_{I\times J}$ when τ^2 is not too large. If we fix only one of subject number and visit number, that is, if we look at Figure 1 either horizontally or vertically, power curves always converge to 1 faster when the sample size is getting

more. Along with this, increasing the number of visits benefits more than increasing the number of subjects in terms of stronger power. In the case of 20 subjects with 50 visits per subject, power magnitude of our test achieves at 1 when τ^2 is around 0.04, whereas it comes to 0.06 when our testing gets approximately 100% control of false positive error.

Figure 2 illustrates the behavior of power when intrinsically τ_k^2 increases. It shows the similar pattens with random coefficients are generated by assuming $\tau_k^2 \equiv \tau^2$ for $k \in 1, \ldots, 3$ under each of the conditions. That is, power of test will be increased as subject number and visit increase, and power of test will converge to 1 when τ^2 is truly large (non-zero) or sample size is large enough. Under each model condition, test of using true $\hat{M}_{I\times J}$ is more powerful than using measurement error added $\hat{M}_{I\times J}$. Our testing method still outperforms Bonferroni's correction when the ramdom coefficients are truly heteroskedastic. While under heteroskedastic variance, it requires larger variance τ^2 for detecting 100% type II error compared with homogeneous variance case.

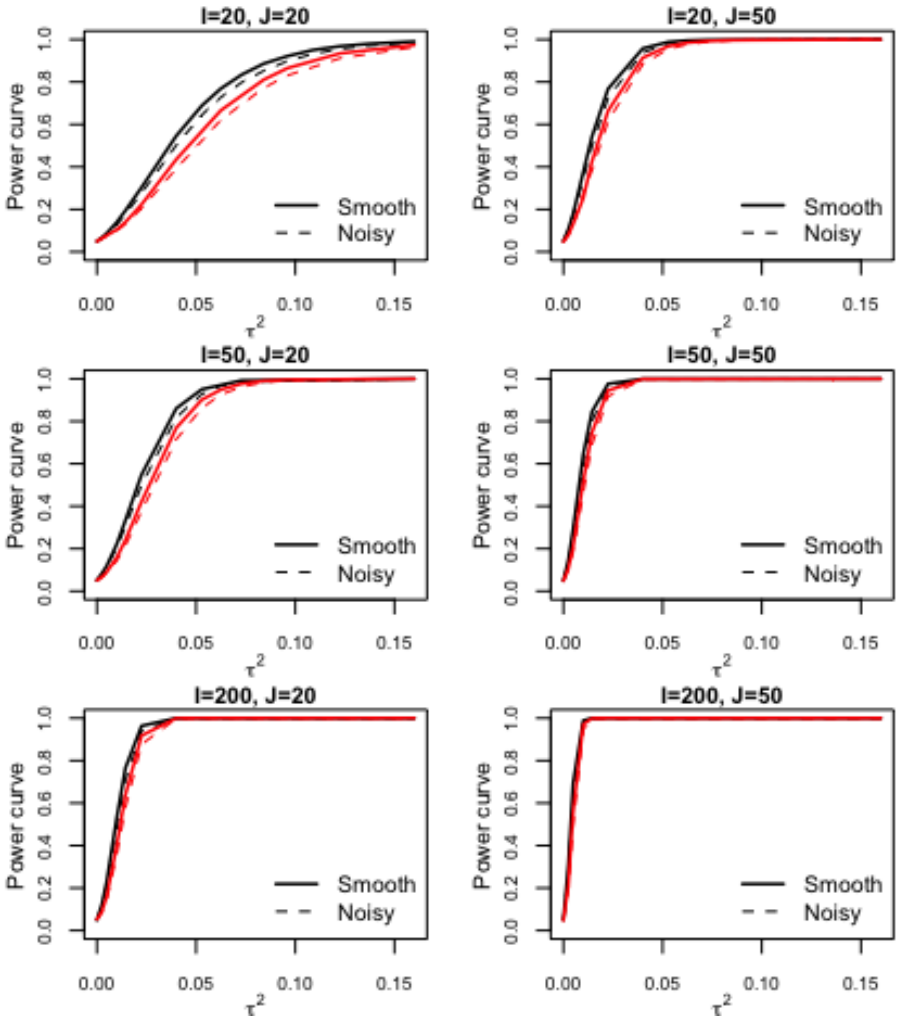


Figure 2: Power of test ($\tau_k^2 = \frac{1}{2^{k-1}}\tau^2$) at the 5% significance level under 12 model conditions.(1) In each panel, solid line represents that true functional data are used; dashed line represents the functional data with measurement error are used. (2) Black line denotes the power curve given by our proposed testing method assuming homogeneity of random scores; red line denotes the power curve generated by Bonferroni's correction testing method.

5.3 Result on estimation

Here, in order to evaluate subject-specific derivation $\beta_i(t)$ from the population effect $\beta(t)$ under heteroskedastic variance model, and compare the estimates for population effect $\beta(t)$ from heteroskedastic variance model and no functional random effect model, we consider multiple model conditions.

We generate the response Y_{ij} from model (4) with $K = 3$, $\delta = 0.5$, $\gamma = 2$, $\gamma_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $Z_{ij} \overset{i.i.d.}{\sim}$ Bernoulli(0.5), $\xi_{ijk} \overset{i.i.d.}{\sim} \mathcal{N}(0,\lambda_k)$, $\theta = 2$, $\theta_{ik} \overset{i.i.d.}{\sim} \mathcal{N}(0,\tau_i^2)$, $i = 1,\ldots,K$, and $\epsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Here, $\lambda_k = 0.5^k, k = 1,\ldots,K$. Then the noisy functional data is generated from model (2) with $X_{ij}(t) = \sum_{k=1}^{K} \xi_{ijk}\phi_k(t)$ and $e_{ijk} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_W^2)$. Here, $\phi_1(t) = \sqrt{2}\sin(2\pi t)$, $\phi_2(t) = \sqrt{2}\cos(4\pi t)$,$\phi_3(t) = \sqrt{2}\sin(4\pi t)$, and σ_W^2 is chosen so that the signal to noise ratio in the functional data $r = \sigma_W^{-2}\int_\tau r(t,t)\mathrm{d}t$ equals either 0 or 3. Note that $r = 0$ corresponds to smooth functional data without noises and $r = 3$ corresponds to noisy functional data. Finally homogeneous τ_i^2 and heteroscedastic τ_i^2 are considered for variance of interest, to be specific, we take $\tau_1^2 = \cdots = \tau_K^2 = \tau^2, \tau^2 \in \{0.02, 0.04, 0.08\}$, and $\tau_i^2 = \frac{1}{2^{i-1}}\tau^2, i = 1,\ldots,K, \tau^2 \in \{0.02, 0.04, 0.08\}$ separately.

A full combination of the number of subject I, the number of visits per subject, J, the signal to noise ratio r and variance of interest $\tau_i^2 (i = 1,\ldots,K)$ in the functional data are taken into model conditions consideration. Hence, datasets are constructed under a total of 72 different model conditions: $\big\{(I, J, r, (\tau_i^2, i = 1,\ldots,K), \tau^2) : I \in \{20, 50, 200\}, J \in \{20, 50\}, r \in \{0, 3\}, (\tau_k^2, k = 1,\ldots,K) \in \{(\tau_1^2 = \cdots = \tau_K^2 = \tau^2), (\tau_k^2 = \frac{1}{2^{k-1}}\tau^2, k = 1,\ldots,K)\}, \tau^2 \in \{0.02, 0.04, 0.08\}\big\}$. Under each model condition, 1000 datasets are simulated.

ISE= $\int(\beta(t) - \hat{\beta}(t))^2\mathrm{dt}$ is used to compare estimate performance for population functional effect $\beta(t)$ from heteroskedastic variance model and no functional random effect model. MISE = $\frac{1}{I}\sum_{i=1}^{I}\int(\beta_i(t) - \hat{\beta}_i(t))^2\mathrm{dt}$ is used for evaluating the subject-specific functional random effect $\beta_i(t)$ estimated by heteroscedastic variance model. Note that $\beta(t) = \sum_{k=1}^{K}\theta\phi_k(t)$, $\beta_i(t) = \sum_{k=1}^{K}\theta_{ik}\phi_k(t), \hat{\beta}(t) = \sum_{k=1}^{K}\hat{\theta}_k\hat{\phi}_k(t)$ and $\hat{\beta}_i(t) = \sum_{k=1}^{K}\hat{\theta}_{ik}\hat{\phi}_k(t)$, MSE = $\frac{1}{I}\sum_{i=1}^{I}\frac{1}{J_i}\sum_{j=1}^{J_i}(y_{ij} - \hat{y}_{ij})^2$ is used to compare predictions from heteroskedastic variance model and no functional random effect model.

Table 2 and Table 3 summarize the ISE for fixed effect $\hat{\beta}(t)$, MSE for response $\hat{Y}_{ij}$, and MISE for subject-level random effect $\hat{\beta}_i(t)$ across a total of 72 model conditions. In each scenario, compared with the model without

Table 2: Estimation for heteroskedastic model with homogeneous τ_i^2

Model	$\tau_i^2 \equiv \tau^2 = 0.02$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
I = 20, J = 20	0	0.0295	0.0295	0	0.9304	0.8655	0	0.0662
I = 20, J = 20	3	0.0351	0.0351	3	1.0096	0.9430	3	0.0682
I = 20, J = 50	0	0.0165	0.0164	0	0.9928	0.9412	0	0.0479
I = 20, J = 50	3	0.0187	0.0187	3	1.0750	1.0228	3	0.0493
I = 50, J = 20	0	0.0147	0.0147	0	0.9401	0.8813	0	0.0581
I = 50, J = 20	3	0.0170	0.0170	3	1.0180	0.9578	3	0.0596
I = 50, J = 50	0	0.0099	0.0098	0	0.9929	0.9410	0	0.0440
I = 50, J = 50	3	0.0108	0.0108	3	1.0755	1.0237	3	0.0453
I = 200, J = 20	0	0.0079	0.0080	0	0.9395	0.8829	0	0.0527
I = 200, J = 20	3	0.0086	0.0086	3	1.0190	0.9626	3	0.0536
I = 200, J = 50	0	0.0067	0.0067	0	0.9945	0.9419	0	0.0414
I = 200, J = 50	3	0.0070	0.0070	3	1.0775	1.0249	3	0.0426
Model	$\tau_i^2 \equiv \tau^2 = 0.04$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
I = 20, J = 20	0	0.0334	0.0331	0	0.9608	0.8585	0	0.1067
I = 20, J = 20	3	0.0389	0.0386	3	1.0401	0.9370	3	0.1102
I = 20, J = 50	0	0.0200	0.0196	0	1.0253	0.9342	0	0.0723
I = 20, J = 50	3	0.0221	0.0219	3	1.1075	1.0162	3	0.0755
I = 50, J = 20	0	0.0160	0.0159	0	0.9712	0.8703	0	0.0975
I = 50, J = 20	3	0.0183	0.0182	3	1.0492	0.9473	3	0.1004
I = 50, J = 50	0	0.0112	0.0110	0	1.0260	0.9326	0	0.0668
I = 50, J = 50	3	0.0121	0.0120	3	1.1085	1.0152	3	0.0696
I = 200, J = 20	0	0.0083	0.0083	0	0.9715	0.8683	0	0.0905
I = 200, J = 20	3	0.0090	0.0090	3	1.0510	0.9481	3	0.0928
I = 200, J = 50	0	0.0071	0.0070	0	1.0280	0.9328	0	0.0634
I = 200, J = 50	3	0.0074	0.0074	3	1.1111	1.0156	3	0.0663

Model	$\tau_i^2 \equiv \tau^2 = 0.08$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
I = 20, J = 20	0	0.0410	0.0399	0	1.0215	0.8454	0	0.1677
I = 20, J = 20	3	0.0465	0.0454	3	1.101	0.9242	3	0.1738
I = 20, J = 50	0	0.0266	0.0258	0	1.0899	0.9258	0	0.1020
I = 20, J = 50	3	0.0289	0.0281	3	1.1723	1.0079	3	0.1075
I = 50, J = 20	0	0.0187	0.0183	0	1.0335	0.8524	0	0.1546
I = 50, J = 20	3	0.0209	0.0205	3	1.1116	0.9298	3	0.1608
I = 50, J = 50	0	0.0137	0.0133	0	1.0919	0.9234	0	0.0933
I = 50, J = 50	3	0.0147	0.0143	3	1.1745	1.0060	3	0.0986
I = 200, J = 20	0	0.0091	0.0090	0	1.0353	0.8482	0	0.1454
I = 200, J = 20	3	0.0098	0.0096	3	1.1150	0.9277	3	0.1507
I = 200, J = 50	0	0.0078	0.0077	0	1.0952	0.9235	0	0.0889
I = 200, J = 50	3	0.0081	0.0080	3	1.1782	1.0063	3	0.0943

functional random effects $\beta_i(t)$, the heteroskedastic model provides slightly smaller ISE for estimating the fixed effect $\beta(t)$, and better precision on estimating the response Y_{ij} in terms of smaller MSE. Analysis using the true functional matrix ($r = 0$) always outperforms modeling with biased functional matrix (r=3). As the sample size increases, both model have better performance regards smaller ISE for fixed effect estimation, smaller MSE for response prediction, and smaller MISE for subject-specific random effect estimation. In both cases of hetero-variance and homo-variance, for the heteroskedastic model, one can find that as the variance τ^2 increases, even though the estimation errors are inflated for the fixed effect $\beta(t)$ and subject-level random effect $\beta_i(t)$, prediction accuracy for the response Y_{ij} keeps increasing, whereas model without functional random effect keeps working worse. Possible reasons could be that heteroskdastic model fit the data naturally better when the random coefficient underlying the data has relatively large variance, and model without functional random effect will underfit data.

Table 3: Estimation for heteroskedastic model with hetero-variance

Model	$\tau^2 = 0.02, \tau_i^2 = \frac{1}{2^{i-1}}\tau^2$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
$I = 20, J = 20$	0	0.0279	0.0279	0	0.9234	0.8662	0	0.0443
$I = 20, J = 20$	3	0.0335	0.0335	3	1.0024	0.9433	3	0.0462
$I = 20, J = 50$	0	0.0151	0.0151	0	0.9844	0.9428	0	0.0293
$I = 20, J = 50$	3	0.0173	0.0173	3	1.0665	1.0241	3	0.0303
$I = 50, J = 20$	0	0.0142	0.0142	0	0.9325	0.8828	0	0.0360
$I = 50, J = 20$	3	0.0165	0.0165	3	1.0105	0.9589	3	0.0372
$I = 50, J = 50$	0	0.0093	0.0093	0	0.9843	0.9441	0	0.0262
$I = 50, J = 50$	3	0.0103	0.0102	3	1.0671	1.0266	3	0.0269
$I = 200, J = 20$	0	0.0078	0.0078	0	0.9315	0.8863	0	0.0309
$I = 200, J = 20$	3	0.0085	0.0084	3	1.0110	0.9656	3	0.0314
$I = 200, J = 50$	0	0.0066	0.0066	0	0.9859	0.9464	0	0.0241
$I = 200, J = 50$	3	0.0069	0.0069	3	1.0690	1.0292	3	0.0248
Model	$\tau^2 = 0.04, \tau_i^2 = \frac{1}{2^{i-1}}\tau^2$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
$I = 20, J = 20$	0	0.0302	0.0301	0	0.9465	0.8608	0	0.0670
$I = 20, J = 20$	3	0.0359	0.0357	3	1.0254	0.9383	3	0.0698
$I = 20, J = 50$	0	0.0171	0.0169	0	1.0085	0.9384	0	0.0442
$I = 20, J = 50$	3	0.0193	0.0192	3	1.0908	1.0198	3	0.0458
$I = 50, J = 20$	0	0.0151	0.0149	0	0.9559	0.8750	0	0.0582
$I = 50, J = 20$	3	0.0174	0.0173	3	1.0340	0.9513	3	0.0602
$I = 50, J = 50$	0	0.0101	0.0100	0	1.0088	0.9386	0	0.0403
$I = 50, J = 50$	3	0.0110	0.0109	3	1.0915	1.0210	3	0.0416
$I = 200, J = 20$	0	0.0081	0.0080	0	0.9555	0.8763	0	0.0525
$I = 200, J = 20$	3	0.0087	0.0087	3	1.0351	0.9555	3	0.0537
$I = 200, J = 50$	0	0.0068	0.0068	0	1.0110	0.9400	0	0.0377
$I = 200, J = 50$	3	0.0072	0.0071	3	1.094	1.0227	3	0.0392

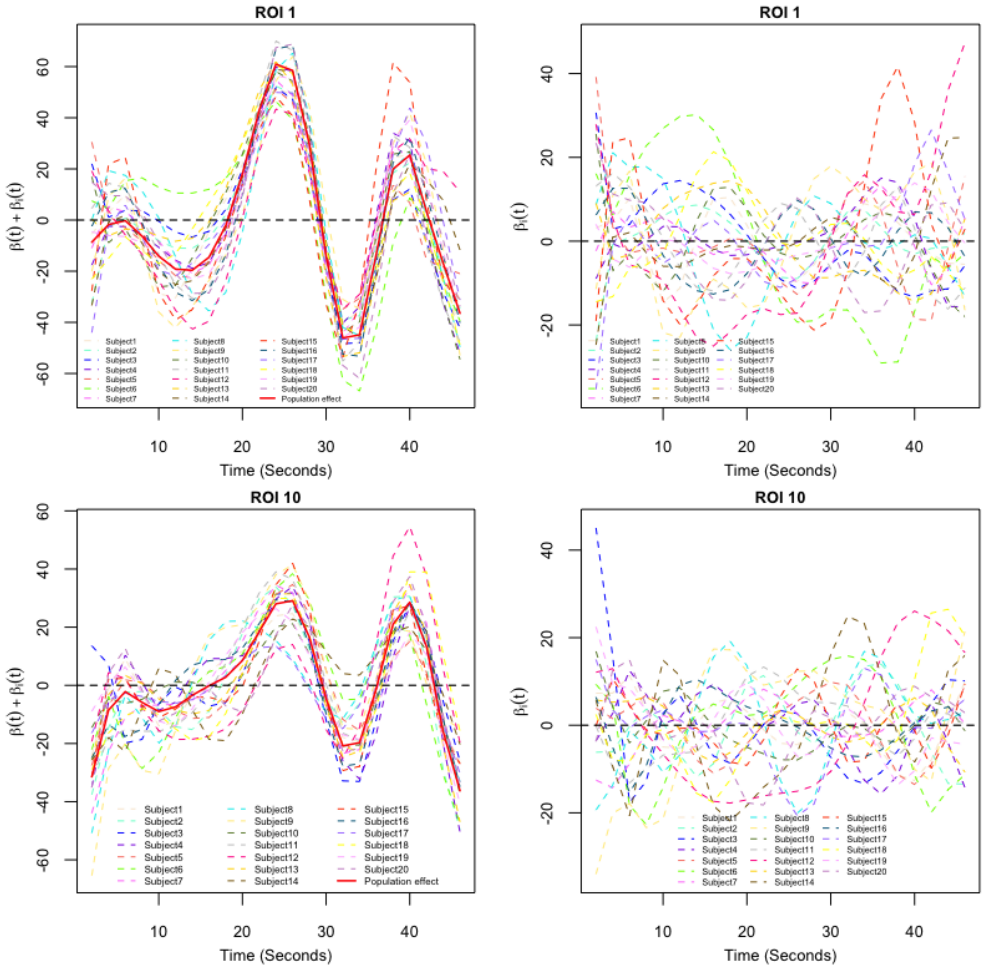
Model	$\tau^2 = 0.08, \tau_i^2 = \frac{1}{2^{i-1}}\tau^2$							
	ISE for $\hat{\beta}(t)$			MSE for $\hat{Y}_{ij}$			MISE for $\hat{\beta}_i(t)$	
	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	Without $\beta_i(t)$	With $\beta_i(t)$	r	With $\beta_i(t)$
I = 20, J = 20	0	0.0350	0.0341	0	0.9925	0.8513	0	0.1017
I = 20, J = 20	3	0.0406	0.0398	3	1.0715	0.9290	3	0.1060
I = 20, J = 50	0	0.0212	0.0206	0	1.0568	0.9326	0	0.0648
I = 20, J = 50	3	0.0233	0.0229	3	1.1392	1.0143	3	0.0675
I = 50, J = 20	0	0.0167	0.0163	0	1.0027	0.8630	0	0.0917
I = 50, J = 20	3	0.0190	0.0187	3	1.0810	0.9394	3	0.0952
I = 50, J = 50	0	0.0116	0.0113	0	1.0577	0.9319	0	0.0591
I = 50, J = 50	3	0.0125	0.0122	3	1.1405	1.0143	3	0.0616
I = 200, J = 20	0	0.0085	0.0084	0	1.0035	0.8624	0	0.0849
I = 200, J = 20	3	0.0092	0.0090	3	1.0832	0.9414	3	0.0876
I = 200, J = 50	0	0.0073	0.0072	0	1.0612	0.9325	0	0.0556
I = 200, J = 50	3	0.0076	0.0075	3	1.1443	1.0152	3	0.0582

6 Real Data Application

In this section, we conducted a real data application using data from fMRI painful stimulus study we introduced in section 2. During this painful study, a total number of 20 subjects are asked for the subjective painful rating after corresponding thermal stimuli is given. Following the thermal stimuli, fMRI machine records the brain activity every 2 seconds from 21 regions of interest (ROI) simultaneously at one's brain for a entire course of 46 seconds. Then the subjective painful rating regarding to the hot/warm stimuli is reported by the individual. For each individual, different amount of visits ranging in [39,48] are taken. We treat the continuously observed brain neural imaging data as a functional covariate, thermal stimuli as a binary categorical covariate, and subjective rating as scalar response. Then we applied our proposed functional mixed effect model extending the scalar on function linear regression for this repeated outcomes.

Test procedures on 21 ROIs show that there are about half which have significant subjective-specific functional random effect at level of 0.05. We picked 3 out of the most significant ROIs (ROI 1, ROI 10, and ROI 19) to visualize difference among $\beta_i(t)$ and compare modeling them with no func-

tional random effect and with heteroscedastic variance on subject-specific brain activity. Figure 3 shows the



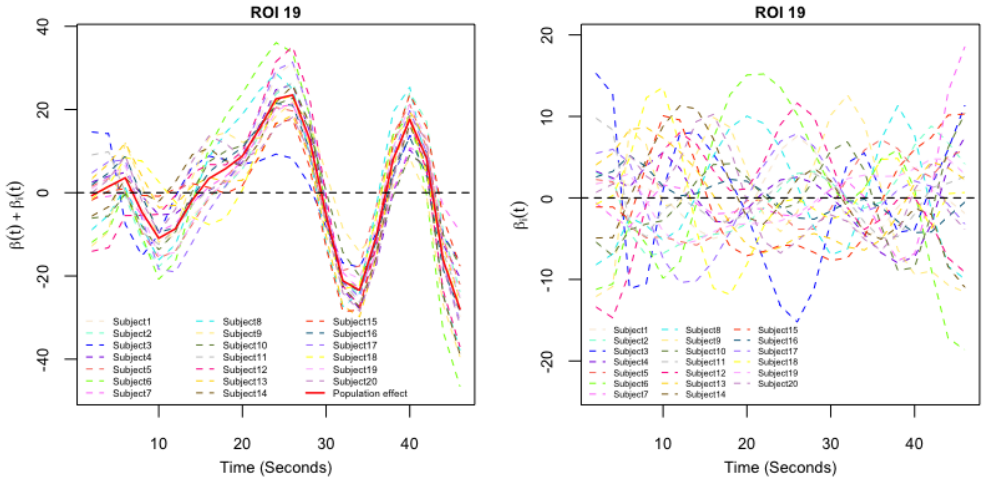


Figure 3: Estimates of subject-specific functional mediation effects and subject deviation from the population functional mediation for ROI 1, 10 and 19. In the left panel, the red solid line represents the population functional mediation effect $\beta(t)$; each of the dashed curve is subject-specific functional mediation effect $\beta(t) + \beta_i(t)$; In the right panel, each of the dashed curve is subject-specific deviation $\beta_i(t)$ from the population functional mediation effect $\beta(t)$. Each subject has the same color across all ROIs.

Table 4: RMSE comparison of two models for ROI 1, 10, and 19

ROI	RMSE	
	No random effect	Heteroscedastic
ROI 1	79.236	73.67
ROI 10	80.649	74.64
ROI 19	80.39	76.67

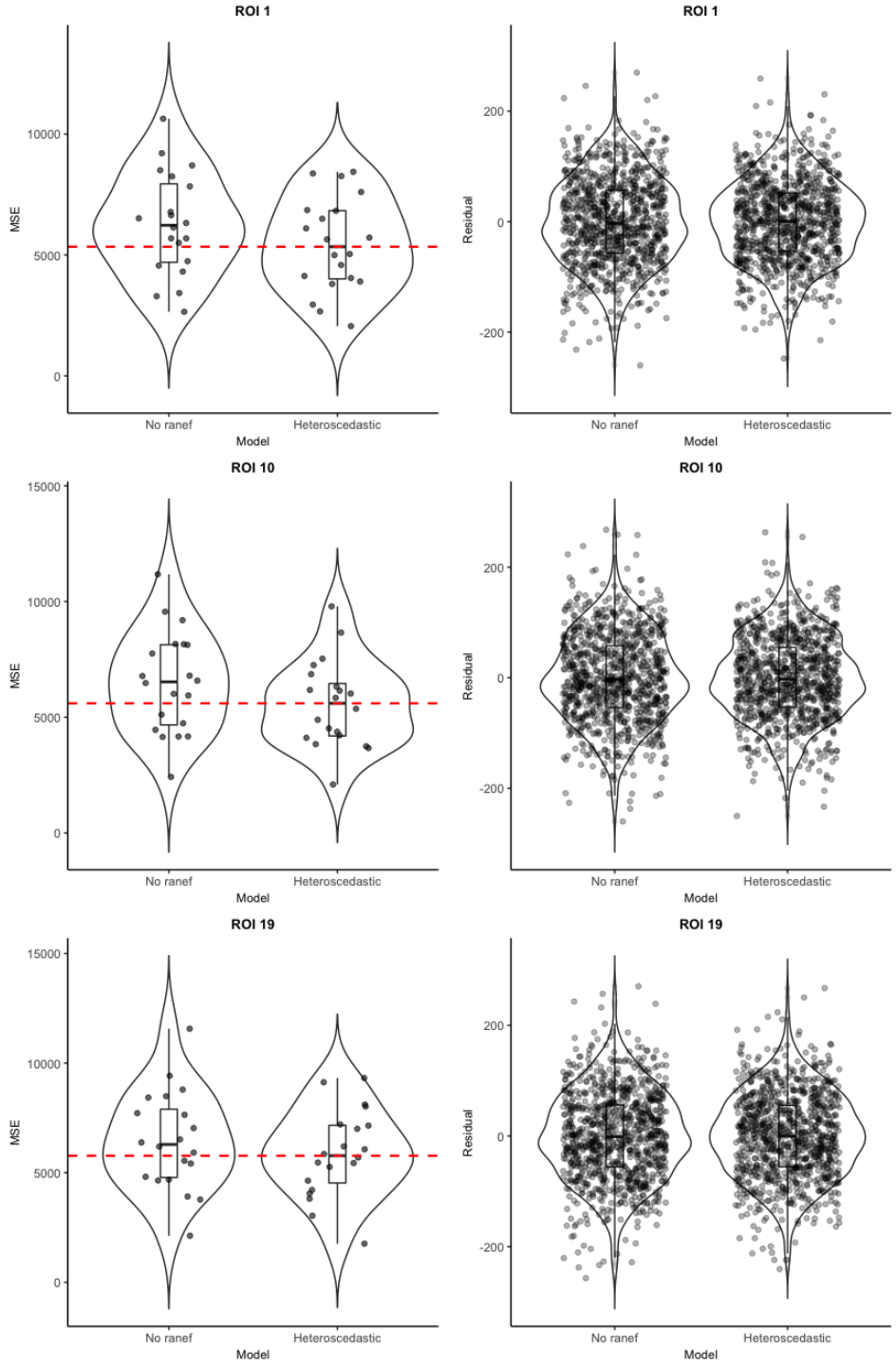

ROI 1
MSE
10000
5000
0
No ranef
Heteroscedastic
Model
ROI 1
Residual
200
0
-200
No ranef
Heteroscedastic
Model
ROI 10
MSE
15000
10000
5000
0
No ranef
Heteroscedastic
Model
ROI 10
Residual
200
0
-200
No ranef
Heteroscedastic
Model
ROI 19
MSE
15000
10000
5000
0
No ranef
Heteroscedastic
Model
ROI 19
Residual
200
0
-200
No ranef
Heteroscedastic
Model

Figure 4: Comparison of model fitting between model without subject-level random effect and model with heteroscedastic subject-level random effect for ROI 1, 10 and 19. Left panel shows the violin plot for MSE of each of 20 subjects in two models; right panel shows the violin plot for 943 residuals in the two models. The red dashed curve represents the median MSE given by Model with heteroscedastic variance

References

[1] Jonathan D Cohen, Nathaniel Daw, Barbara Engelhardt, Uri Hasson, Kai Li, Yael Niv, Kenneth A Norman, Jonathan Pillow, Peter J Ramadge, Nicholas B Turk-Browne, et al. Computational approaches to fmri analysis. *Nature neuroscience*, 20(3):304, 2017.

[2] Ciprian M Crainiceanu, Ana-Maria Staicu, and Chong-Zhi Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

[3] Chong-Zhi Di, Ciprian M Crainiceanu, Brian S Caffo, and Naresh M Punjabi. Multilevel functional principal component analysis. *The annals of applied statistics*, 3(1):458, 2009.

[4] Jeff Goldsmith, Vadim Zipunnikov, and Jennifer Schrack. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2):344–353, 2015.

[5] Hui Huang, Yehua Li, and Yongtao Guan. Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*, 109(508):1412–1424, 2014.

[6] Lei Huang, Philip T Reiss, Luo Xiao, Vadim Zipunnikov, Martin A Lindquist, and Ciprian M Crainiceanu. Two-way principal component analysis for matrix-variate data, with an application to functional magnetic resonance imaging data. *Biostatistics*, 18(2):214–229, 2017.

[7] Alois Kneip. Nonparametric estimation of common regressors for similar curve data. *The Annals of Statistics*, pages 1386–1427, 1994.

[8] Martin A Lindquist. Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500):1297–1309, 2012.

[9] Martin A Lindquist et al. The statistical analysis of fmri data. *Statistical science*, 23(4):439–464, 2008.

[10] Mathew W McLean, Giles Hooker, and David Ruppert. Restricted likelihood ratio tests for linearity in scalar-on-function regression. *Statistics and Computing*, 25(5):997–1008, 2015.

[11] James Ramsay. Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*, 2005.

[12] James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

[13] James O Ramsay and CJ Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 539–572, 1991.

[14] James O Ramsay and Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2007.

[15] Joan G Staniswalis and J Jack Lee. Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 93(444):1403–1418, 1998.

[16] Tor D Wager, Matthew L Davidson, Brent L Hughes, Martin A Lindquist, and Kevin N Ochsner. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59(6):1037–1050, 2008.

[17] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[18] Luo Xiao, Bing He, Annemarie Koster, Paolo Caserotti, Brittney Lange-Maia, Nancy W Glynn, Tamara B Harris, and Ciprian M Crainiceanu. Movement prediction using accelerometers in a human population. *Biometrics*, 72(2):513–524, 2016.

[19] Luo Xiao, Lei Huang, Jennifer A Schrack, Luigi Ferrucci, Vadim Zipunnikov, and Ciprian M Crainiceanu. Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics*, 16(2):352–367, 2014.

[20] Luo Xiao, Cai Li, William Checkley, and Ciprian Crainiceanu. Fast covariance estimation for sparse functional data. *Statistics and computing*, 28(3):511–522, 2018.

[21] Luo Xiao, Vadim Zipunnikov, David Ruppert, and Ciprian Crainiceanu. Fast covariance estimation for high-dimensional functional data. *Statistics and computing*, 26(1-2):409–421, 2016.

[22] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

[23] Vadim Zipunnikov, Brian Caffo, David M Yousem, Christos Davatzikos, Brian S Schwartz, and Ciprian Crainiceanu. Multilevel functional principal component analysis for high-dimensional data. *Journal of Computational and Graphical Statistics*, 20(4):852–873, 2011.