

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

August 22, 2018

1 Introduction

Recently, there has been a rapid emerge in the area of brain imaging study[5, 9]. Scientific interest has raised in the potential relationship between the continuously observed brain activity trajectories with one's subjective reaction to certain experiment conditions[4], results of the subject-level difference in the observed trajectories in the reaction, and thus how the brain activity interacts the psychological/physiological process. By means of the instrument such as fMRI or EEG machine[10], a functional brain activity curve can be noninvasively observed on the equally spaced time grid t_1, \dots, 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[5]. 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 has been playing a key role in dealing with the type of richness data[3, 7, 8, 11, 12, 13, 6]. Few methods has been developed to make good use of fully observed fMRI trajectory in stead of summarized statistics in the past decades[1, 4]. 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. Martin[4] 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 [4] 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 ma-

trix 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 i th subject in j th replication, Z_{ij} as the indicator variable of thermal stimuli level at i th subject in j th 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 $y_{ij} = \delta + \gamma Z_{ij} + \int \beta(t) M_{ij}(t) dt + \epsilon_{ij}$, 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

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

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_i s 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 $\text{cov}\{\beta_i(s), \beta_i(t)\} = \mathcal{C}(s, t)$, 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

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

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 [2] 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 $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ 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 [12], 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 [12] 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_k s 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

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

Model (3) has infinitely many parameters and hence can 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

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

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 [13]. De-

note the selected number by \hat{K} . Then a practical model for (4) is

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

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 [6], we further simplify the test by making the assumption that

$$\tau_k^2 = \tau^2 \text{ for all } k, \quad (6)$$

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:

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

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 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ_i})^\top \in \mathbb{R}^{J_i}$, $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iJ_i})^\top \in \mathbb{R}^{J_i}$, $\mathbf{A}_i = (\xi_{ijk})_{jk} \in \mathbb{R}^{J_i \times K}$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iJ_i})^\top \in \mathbb{R}^{J_i}$. Also let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^\top \in \mathbb{R}^K$ and $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{iK})^\top \in \mathbb{R}^K$. 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 $\boldsymbol{\Delta}_i = (\mathbf{1}_{J_i} \quad \mathbf{Z}_i \quad \mathbf{A}_i) \in \mathbb{R}^{J_i \times (2+K)}$ and $\boldsymbol{\alpha} = (\delta, \gamma, \boldsymbol{\theta}^\top)^\top \in \mathbb{R}^{2+K}$. It follows that

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

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

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

or equivalently,

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

where \mathbf{U}_{ij} is the j th 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 $K = 3$, $\alpha = 0.5$, $\gamma = 2$, $\gamma_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $Z_{ij} \stackrel{i.i.d.}{\sim} \text{Bernoulli}(0.5)$, $\xi_{ijk} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \lambda_k)$, $\theta = 2$, $\theta_{ik} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \tau^2)$, and $\epsilon_{ij} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Here, $\lambda_k = 0.5^k$, $k = 1, \dots, K$. Then the noisy functional data is generated from model (2) with $X_{ij}(t) = \sum_{k=1}^K \xi_{ijk} \phi_k(t)$ and $e_{ijk} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_W^2)$. 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 $r = \sigma_W^{-2} \int_{\mathcal{T}} \mathcal{K}(t, t) dt$ equals either 0 or 3. Note that $r = 0$ corresponds to smooth functional data without noises and $r = 3$ corresponds to noisy functional data.

We simulate data using a factorial design with three factors: the number of subject I , the number of visits per subject J , and the signal to noise ratio r in the functional data. A total of 12 different model conditions are used: $\{(I, J, r) : I \in \{20, 50, 200\}, J \in \{20, 50\}, r \in \{0, 3\}\}$. 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^2 s under the alternative hypothesis are used to generate power curves.

5.2 Results on test

Luo: I stopped here.

Table 1 provides with the size test results based on 20000 times simulation under full combinations of the simulation design $I \in \{20, 50, 200\}$, $J \in \{20, 50\}$, $K \in \{2, 4, 8\}$ with $\hat{M}_{I \times J}$ being true and noise added. As we can see, the test p-value under each model condition is nearly 0.05, which demonstrate that our test procedure can well control false negative rate around 0.05, and maintain stable performance over change on sample size I , visit repeats J and K .

Table 1: Sizes of test at the 5% level under different model conditions.

Model condition	$r = 0$	$r = 3$
I = 20, J = 20	0.05	0.05
I = 20, J = 50	0.051	0.05
I = 50, J = 20	0.051	0.051
I = 50, J = 50	0.05	0.050
I = 200, J = 20	0.049	0.049
I = 200, J = 50	0.051	0.051

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 shown. As subject number and visit increase, power of test will be correspondingly increased, and is able to converge to 1 eventually. Under each model condition, test of using true $\hat{M}_{I \times J}$ is more powerful than measurement error considered, i.e. noise added $\hat{M}_{I \times J}$.

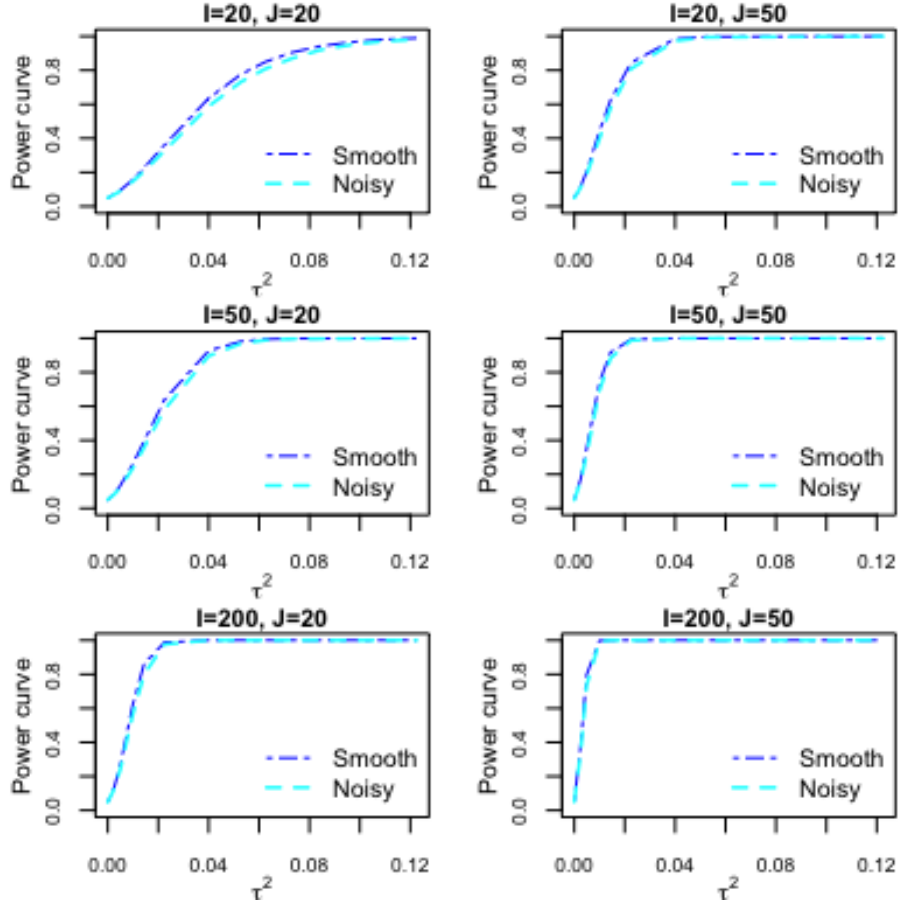


Figure 1: Power of test ($\tau_k^2 = \tau^2$) at the 5% significance level under different model conditions.

5.3 Size Testing To Be Added

In this section, to compare with our testing procedure, we use the pdIdent model fitting structure in lme approach to fit model under the homoscedasticity assumption, and we would like to show that the size test is too conservative using this method.

Another method is to fit model under the heteroscedasticity assumption, and then perform multiple test using Bonferroni or benjamini & Hochberg method. And we would like to show that the size test is also too conservative using this method.

5.4 Power Testing To Be Added

In the preceding simulation settings, we have shown that our testing procedure by assuming $\tau_k^2 = \tau^2$ for all k has good power property when the data is generated following $\tau_k^2 = \tau^2$. In this section, we use heteroscedastic variances τ_k^2 over $k \in \{1, \dots, K\}$ to construct the datasets. Specifically, let $\tau_k^2 = \frac{1}{2^{k-1}}\tau^2, k = 1, \dots, 3$, and different nonzero τ_k^2 s under the alternative hypothesis are used to generate power curves.

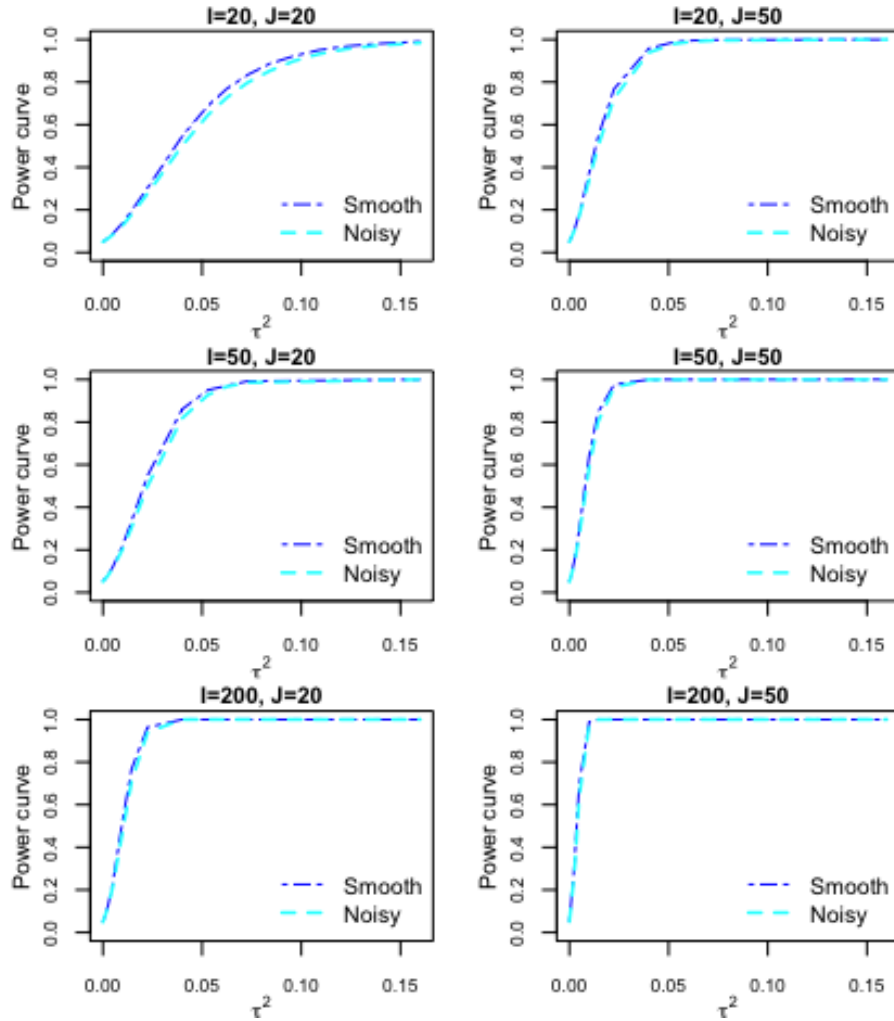


Figure 2: Power of test ($\tau_k^2 = \frac{1}{2^{k-1}}\tau^2$) at the 5% significance level under different model conditions.

Figure 2 illustrates the behavior of power when τ_k^2 increases. It shows the similar patterns with data is generated by assuming $\tau_k^2 = \tau^2$ for $k \in 1, \dots, 3$ 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. Under each model condition, test of using true $\hat{M}_{I \times J}$ is more powerful than using measurement error added $\hat{M}_{I \times J}$. While under heterogeneous variance, it takes longer for the test of power to increase to 1 compared with using homogeneous variance to construct dataset.

5.5 Estimation To Be Added

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

We generate the response Y_{ij} from model (4) with $K = 3$, $\delta = 0.5$, $\gamma = 2$, $\gamma_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $Z_{ij} \stackrel{i.i.d.}{\sim} \text{Bernoulli}(0.5)$, $\xi_{ijk} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \lambda_k)$, $\theta = 2$, $\theta_{ik} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \tau_i^2)$, $i = 1, \dots, K$, and $\epsilon_{ij} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Here, $\lambda_k = 0.5^k$, $k = 1, \dots, K$. Then the noisy functional data is generated from model (2) with $X_{ij}(t) = \sum_{k=1}^K \xi_{ijk} \phi_k(t)$ and $e_{ijk} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_W^2)$. 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 $r = \sigma_W^{-2} \int_{\tau} r(t, t) dt$ 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 $\tau_1^2 = \dots = \tau_K^2 = \tau^2$, $\tau^2 \in \{0.02, 0.04, 0.08\}$, and $\tau_i^2 = \frac{1}{2^{i-1}}\tau^2$, $i = 1, \dots, K$, $\tau^2 \in \{0.02, 0.04, 0.08\}$ 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 τ_i^2 ($i = 1, \dots, K$) in the functional data are taken into model conditions consideration. Hence, datasets are constructed under a total of 72 different model conditions: $\{(I, J, r, (\tau_i^2, i = 1, \dots, K), \tau^2) : I \in \{20, 50, 200\}, J \in \{20, 50\}, r \in \{0, 3\}, (\tau_k^2, k = 1, \dots, K) \in \{(\tau_1^2 = \dots = \tau_K^2 = \tau^2), (\tau_k^2 = \frac{1}{2^{k-1}}\tau^2, k = 1, \dots, K)\}, \tau^2 \in \{0.02, 0.04, 0.08\}\}$. Under each model condition, 20000 datasets are simulated.

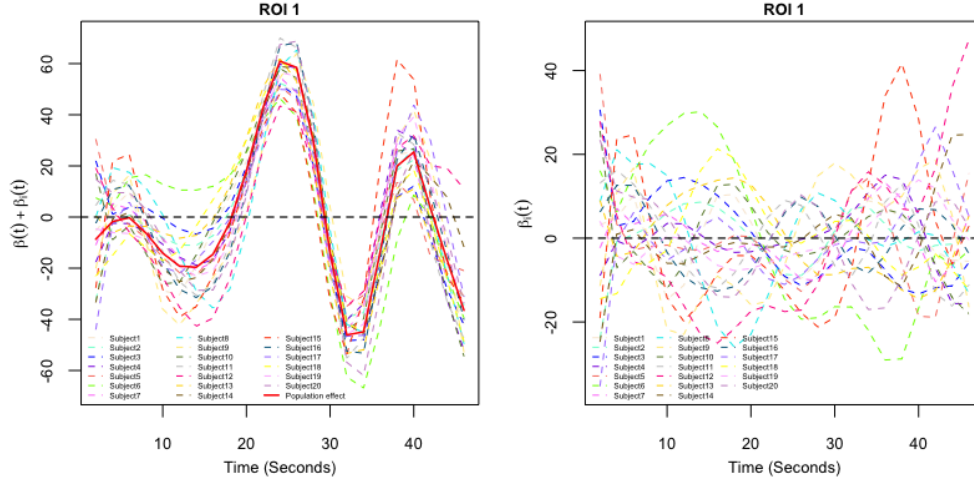
ISE = $\int (\beta(t) - \hat{\beta}(t))^2 dt$ is used to compare estimate performance for population functional effect $\beta(t)$ from heteroscedastic variance model and no

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

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



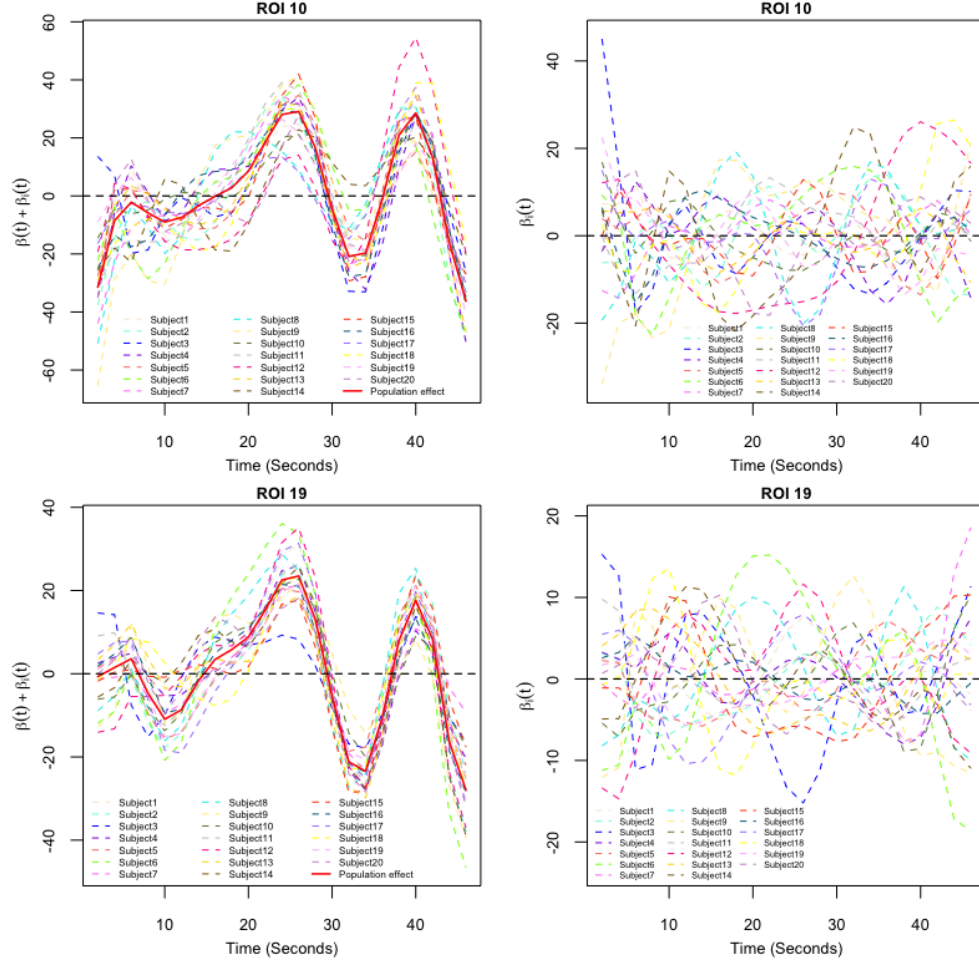
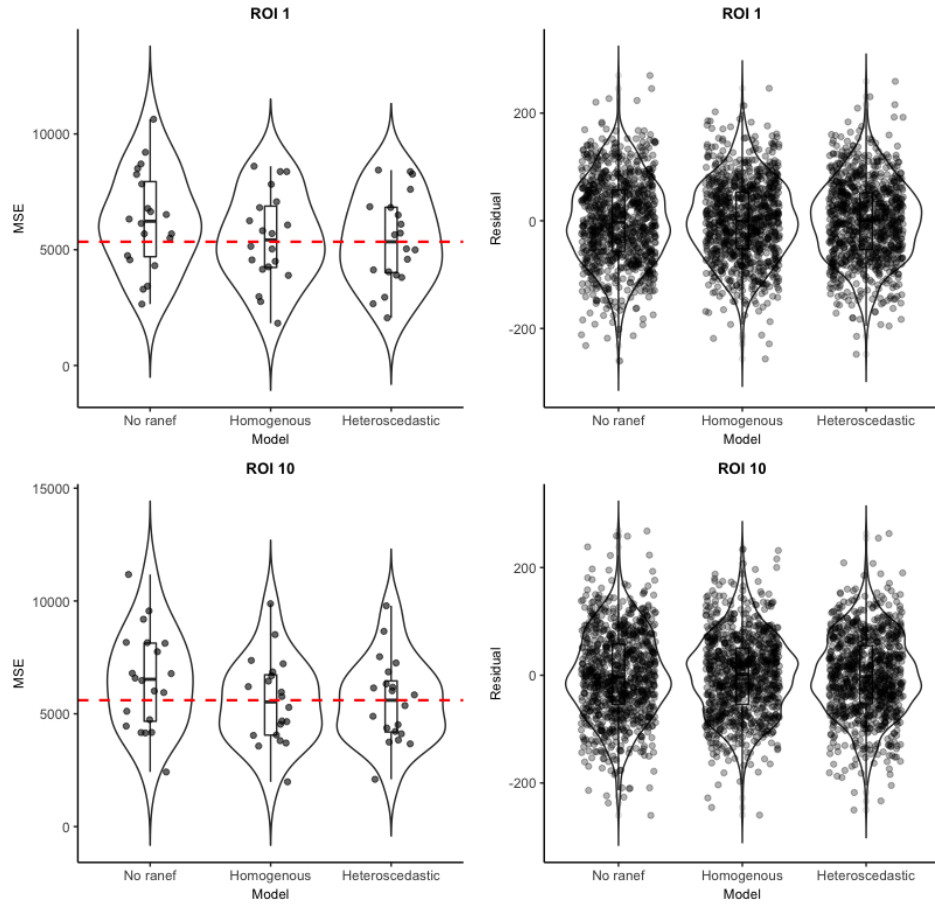


Figure 3: Estimates of subject-specific functional mediation effects and subject deviation from the population functional mediation for ROI 1, 10 and 19. In the left panel, the red solid line represents the population functional mediation effect $\beta(t)$; each of the dashed curve is subject-specific 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 2: RMSE comparison of three models for ROI 1, 10, and 19

ROI	RMSE		
	No random effect	Homogeneous	Heteroscedastic
ROI 1	79.236	74.181	73.67
ROI 10	80.649	74.59	74.64
ROI 19	80.39	76.37	76.67



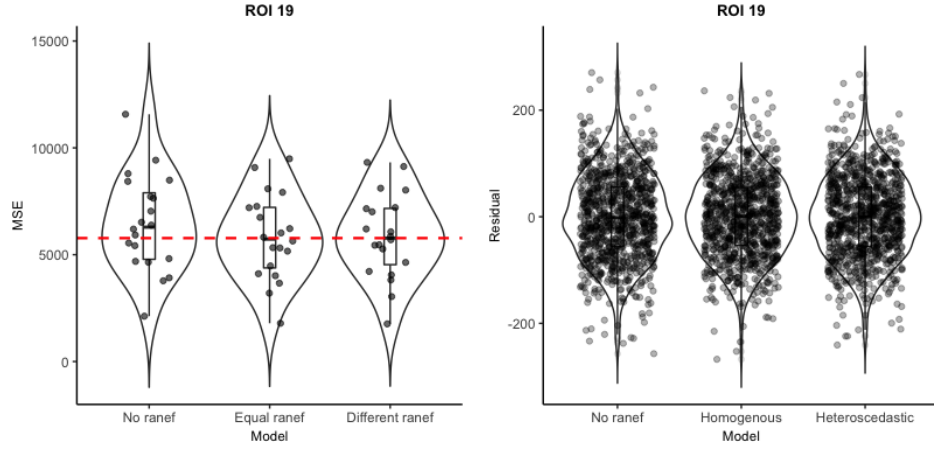
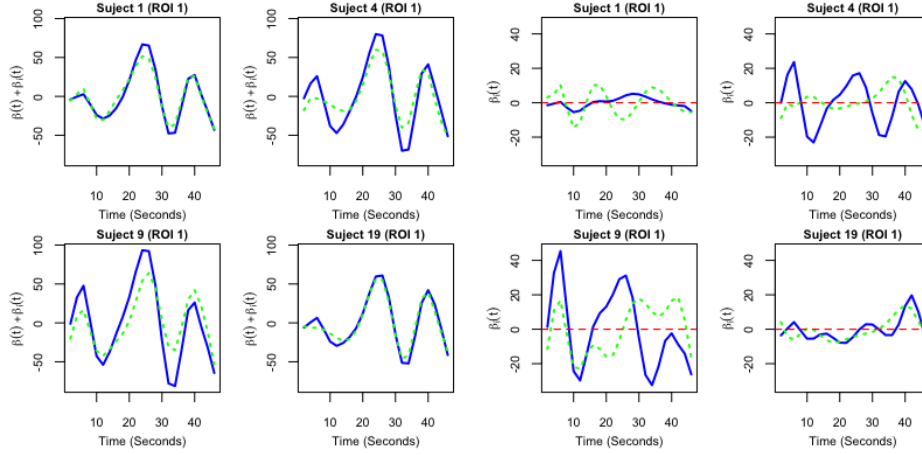


Figure 4: Comparison of model fitting among three models for ROI 1, 10 and 19. Left panel shows the violin plot for MSE of each of 20 subjects in three models; right panel shows the violin plot for 943 residuals in three models. The red dashed curve represents the median MSE given by Model with heteroscedastic variance



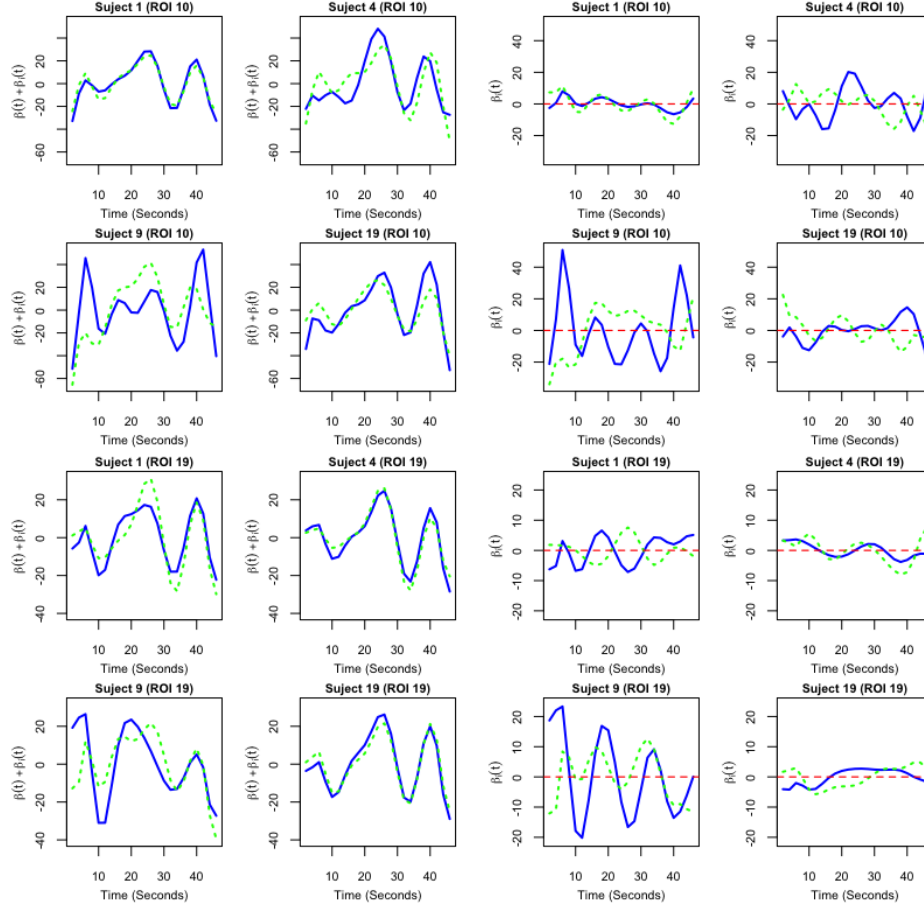


Figure 5: Comparison of estimates of subject-specific functional mediation effects and subject deviation from the population functional mediation of subject1, subject4, subject9, and subject 19 for ROI 1, 10 and 19. In the left panel, the blue solid line represents subject-specific functional mediation effect $\beta(t) + \beta_i(t)$ given by model with heteroscedastic variance, the green dash line represents $\beta(t) + \beta_i(t)$ given by model with homogeneous variance; In the right panel, the blue solid line represents subject-specific functional mediation effect deviation from the population effect $\beta_i(t)$ given by model with heteroscedastic variance, the green dash line represents $\beta_i(t)$ given by model with our method

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] 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.
- [3] Jeff Goldsmith, Vadim Zipunnikov, and Jennifer Schrack. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2):344–353, 2015.
- [4] Martin A Lindquist. Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500):1297–1309, 2012.
- [5] Martin A Lindquist et al. The statistical analysis of fmri data. *Statistical science*, 23(4):439–464, 2008.
- [6] 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.
- [7] James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.
- [8] James O Ramsay and Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Springer, 2007.
- [9] 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.
- [10] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.
- [11] Luo Xiao, Cai Li, William Checkley, and Ciprian Crainiceanu. Fast covariance estimation for sparse functional data. *Statistics and computing*, 28(3):511–522, 2018.

- [12] 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.
- [13] 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.

Supplementary Material

6.1 Simulation Setting

We generate response Y_{ij} based on the following model

$$Y_{ij} = \beta_0 + X_{ij}(\beta + \beta_i) + \mathbf{Z}_{ij}^\top(b + \mathbf{B}_i) + \epsilon_{ij},$$

$$\text{where } \begin{cases} \beta_0 = 0.5 \\ \beta = 2 \\ b = 2 \\ X_{ij} \stackrel{i.i.d.}{\sim} \text{Bernolli}(0.5) \\ \beta_i \stackrel{i.i.d.}{\sim} \text{Normal}(0, 1) \\ \mathbf{Z}_{ij}^\top = (Z_{ij1}, Z_{ij2}, \dots, Z_{ijK}) \stackrel{i.i.d.}{\sim} \text{MVN}(\mathbf{0}, \mathbf{I}) \\ Z_{ijk} \stackrel{i.i.d.}{\sim} \text{Normal}(0, 1) \\ \mathbf{B}_i = (b_1, b_2, \dots, b_K)^\top \stackrel{i.i.d.}{\sim} \text{MVN}(\mathbf{0}, \tau^2 \mathbf{I}) \\ b_{ik} \stackrel{i.i.d.}{\sim} \text{Normal}(0, \tau^2) \\ \epsilon_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(0, 1), i \in \{1, \dots, I\}, j \in \{1, \dots, J\} \end{cases}.$$

Note that β_i is a scalar random effect and \mathbf{B}_i is a K -dimensional random effects. Testing \mathbf{B}_i , that is, whether $\tau^2 = 0$ is of our interest. For both size test and power test, we have total sample size $I \in \{20, 50, 200\}$, and repeated visits $J \in \{20, 50\}$. According to the factorial design, we finally have six model designs $\{(I, J)\} \in \{(20, 20), (20, 50), (50, 20), (50, 50), (200, 20), (200, 50)\}$. Under each of the six model designs, dimension K of the covariate \mathbf{Z}_{ij}^\top varies in $\{2, 4, 8\}$. For test of size, we have data generated using $\tau^2 = 0$, and for test of power, we generate data using multiple different non-zero τ^2 as $\{0.05^2, 0.07^2, 0.1^2, 0.15^2, 0.2^2\}$. Thus for test of size, considering different settings of sample size I , repeated visit J and dimension K , for all of above $3 \times 2 \times 3 = 18$ model settings, we conduct the test $H_0: \tau^2 = 0$ v.s. $H_a: \tau^2 > 0$ through 'exactRLR' based on 20000 times simulation; for test of power, with data generated using multiple different non-zero τ^2 , 3 different setting for sample size I , 2 different setting for repeated visit J and 3 different settings for dimension K , and thus under each of above $18 \times 5 = 90$ model settings, we conduct the test $H_0: \tau^2 = 0$ v.s. $H_a: \tau^2 > 0$ through 'exactRLR' based on 20000 times simulation. For size test, simulated results showing around 5% type I error rate will be ideal, and for power test, as τ^2 increases, an

incremental power curve is expected. We use parallel computing to short the computation time.

6.2 Simulation Results for Size Test

Table 3: Table of size test under 18 model conditions by SVD decomposition transformation

Model	I=20,J=20	I=20,I=50	I=50,J=20	I=50,J=50	I=200,J=20	I=200,J=50
$K = 2$	0.052	0.048	0.050	0.050	0.048	0.052
$K = 4$	0.050	0.050	0.050	0.049	0.051	0.049
$K = 8$	0.050	0.052	0.048	0.049	0.051	0.049

Table 1 provides with the size test results based on 20000 times simulation under full combinations of the simulation design $I \in \{20, 50, 200\}$, $J \in \{20, 50\}$, $K \in \{2, 4, 8\}$. As we can see, the test p-values fall in range $[0.048, 0.052]$, which demonstrate that our test procedure can well control type I error around 0.05, and maintain stable performance over change on sample size I , visit repeats J and K .

6.3 Simulation Results for Power Test

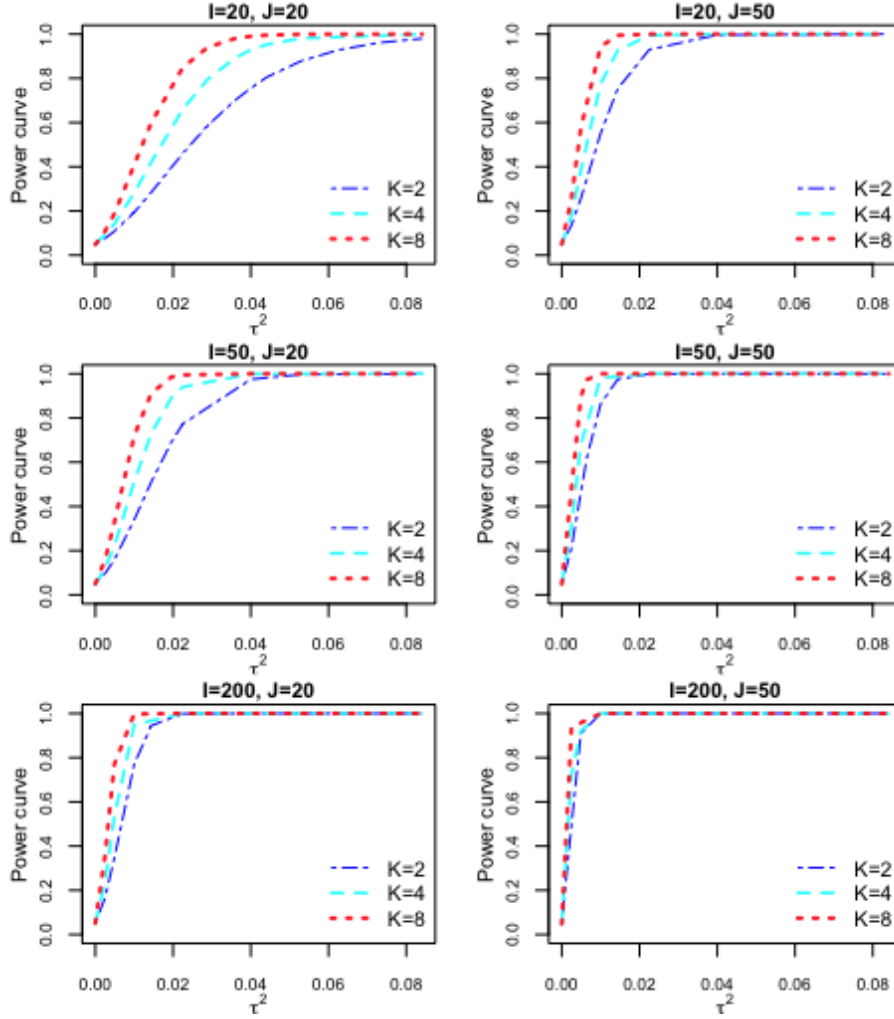


Figure 6: Power curve under different sample size I , repeated visit J , and K

Figure 6 illustrates the pattern of power as the real variance τ^2 increases from 0 up to 0.21^2 under 18 different combinations of sample size I , repeated visit J , and K . Under all cases, a monotone increasing power curve with range from approximately 0.05 to 1 is observed. As one would expect, increase on the dimension K of \mathbf{B}_i can lead to increasing test power when sample size I and visit J are fixed. As sample size I increases, power of test

will also increase and reach 1 much faster. Increasing visit J can also increase test power. Actually the speed of test power converges to 1 is more sensitive to the affect of visit J than sample size I , and thus power will reach 1 faster when J is increased compared with I is increased. For example, comparing the model condition $\{I = 20, J = 20, K = 8\}$ with $\{I = 20, J = 50, K = 8\}$ and $\{I = 50, J = 20, K = 8\}$ respectively, we can see that power reaches over 0.9 at $\tau^2 = 0.1^2$ and is increased to about 1 at $\tau^2 = 0.12^2$ when J is changed from 20 to 50 and I stays 20, while power is around 0.7 at $\tau^2 = 0.1^2$ when I is changed from 20 to 50 and J stay 20. Similar situations can be observed in Figure 6.