| Statistica Sinica Preprint No: SS-2024-0307 | |
|---|---|
| **Title** | Estimation and Inference for High-dimensional Multi-response Growth Curve Model |
| **Manuscript ID** | SS-2024-0307 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202024.0307 |
| **Complete List of Authors** | Xin Zhou,<br>Yin Xia and<br>Lexin Li |
| **Corresponding Authors** | Yin Xia |
| **E-mails** | xiayin@fudan.edu.cn |

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



# Estimation and Inference for High-dimensional Multi-response Growth Curve Model

Xin Zhou[1], Yin Xia[2], and Lexin Li[1]

[1] *University of California at Berkeley, and* [2] *Fudan University*

*Abstract:* A growth curve model (GCM) aims to characterize how an outcome variable evolves, develops and grows as a function of time, along with other predictors. It provides a particularly useful framework to model growth trend in longitudinal data. However, the estimation and inference of GCM with a large number of response variables faces numerous challenges, and remains underdeveloped. In this article, we study the high-dimensional multivariate-response linear GCM, and develop the corresponding estimation and inference procedures. Our proposal involves several innovative components. Specifically, we introduce a Kronecker product structure, which allows us to effectively decompose a very large covariance matrix, and to pool the correlated samples to improve the estimation accuracy. We devise a highly non-trivial multi-step estimation approach to estimate the individual covariance components separately and effectively. We also develop rigorous statistical inference procedures to test both the global effects and the individual effects, and establish the size and power properties as well as the proper false discovery control. We demonstrate the effectiveness of the new method through both intensive simulations, and the analysis of a longitudinal neuroimaging data for Alzheimer's disease.

*Key words and phrases:* Hypothesis testing; Longitudinal data; Kronecker product; Magnetic resonance imaging; Mixed-effects model.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



## 1. Introduction

A growth curve model (GCM) describes how an outcome variable evolves, develops and grows as a function of time along with other related covariates, and is particularly useful for modeling the growth trends in longitudinal data analyses. In a GCM, each individual subject is assumed to have her own unique trajectory of change over time, which represents how the outcome evolves for each subject as time progresses. Such individual trajectories are modeled as functional curves, typically in the form of linear functions. The model involves both fixed-effects that represent population-level relation between the predictors and outcome, as well as random-effects that account for individual variability in the growth trajectories. These random-effects allow for individual differences in both the starting point (intercept) and the rate of change (slope) over time, and capture the deviations of each individual's trajectory from the average trajectory. GCM has been commonly used in a wide range of applications, e.g., biology, psychology, social science, neuroscience, among others (Wei and He, 2006; Newhill et al., 2012; Nocentini et al., 2013; Ordaz et al., 2013).

Our motivation arises from longitudinal studies of Alzheimer's disease (AD). AD is an irreversible neurodegenerative disorder and the leading form of dementia in elderly subjects. It is characterized by progressive impairment of cognitive and memory functions, then loss of independent living, and ultimately death. Its prevalence is rapidly growing as the worldwide population is aging, and it becomes imperative to understand,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



diagnose, and treat this disorder (Jagust, 2018). In recent years, a number of longitudinal AD studies are emerging to track and better understand the progression of AD. One example is Marcus et al. (2010), who collected T1-weighted magnetic resonance imaging (MRI) scans of 150 subjects aged between 60 to 96 years old. Each subject was scanned on two or more visits, separated by at least one year, for a total of 373 image scans. Each MRI scan was preprocessed, mapped to a common brain atlas, and summarized as a vector of region-wise brain volume measurements. Among those subjects, 64 were characterized as demented, and 72 were normal developing controls. A scientific question of central interest is to track the change of brain volumes of different brain regions across time, and understand the difference of developmental trajectories between the AD patients and normal controls. Another example is Yan et al. (2020), who collected blood plasma samples of 35 subjects over 70 years old and followed them up to 15 years. Each subject were sampled three times or more, for a total of 164 plasma samples. For each sample, the extracellular RNA (exRNA) sequence measurements were obtained. Among those subjects, 15 were confirmed AD patients by pathological analysis of their post mortem brains, and 9 were normal controls. A central scientific question is to track the change of exRNA expression levels of a set of AD-related genes over time, and differentiate their developmental trajectories between the two groups of subjects. Such questions are pivotal for our understanding of AD development, and have important diagnostic and therapeutic implications. For both examples, GCM provides a natural framework to address the scientific questions of interest, which translate to the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



estimation and inference of the corresponding fixed-effects parameters in the model. Meanwhile, it is crucial to take into account the spatial correlations between different brain regions or genes, the temporal correlations across different time points, as well as the individual subject variability.

In this article, we study a high-dimensional multivariate-response linear GCM, and develop the corresponding estimation and inference procedures. Our proposal directly addresses the questions in our motivating examples, where the brain regions or genes are modeled as the multivariate response variables. Although GCM is a classical model, the estimation and inference of a high-dimensional multi-response GCM faces numerous serious challenges, and remains underdeveloped. First, the dimension of the response variables can be large, resulting in a very large covariance matrix. Meanwhile, the number of subjects and the number of time points may be limited. To obtain a good covariance estimator, we adopt a Kronecker product structure for the covariance, which allows us to pool the correlated samples effectively to improve the estimation accuracy. Second, the covariance structure is particularly complex. Under the Kronecker product assumption, the covariance involves three key components, including a spatial covariance that accounts for the correlations among different regions or genes, a temporal covariance that accounts for the correlations across multiple time points, and a covariance matrix that reflects the random departure of individual subjects from the population-level intercept and slope. We devise a highly non-trivial multi-step approach to estimate the three covariance components separately and effectively. Last but not

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



least, we develop rigorous statistical inference procedures to test both the global effects and the individual effects, and we establish the size and power properties, as well as the proper false discovery control. We first recognize that our covariance estimator is not the usual sample covariance, because the means of the response variables in our setting are different for different subjects, regions and time points. Therefore, the asymptotics of the classical sample covariance estimator based on independent and identically distributed (i.i.d.) observations are not directly applicable. To overcome this issue, we introduce some level of sparsity, and show that our estimator performs asymptotically the same as the sample covariance matrix obtained by the i.i.d. centralized error terms. We then employ an advanced version of Hanson-Wright inequality (Chen et al., 2023), and derive the proper convergence rates of our covariance estimators, which in turn ensure the desired theoretical guarantees for our proposed tests.

Our proposal is built upon but is also clearly different from several lines of relevant research. The first line is classical GCM modeling, which often adopts the approach of linear multilevel analysis. There have been numerous proposals for GCM estimation (Kackar and Harville, 1981; Goldstein, 1986; Bryk and Raudenbush, 1992) and inference (Giesbrecht and Burns, 1985; Kenward and Roger, 1997; Carpenter et al., 1999; Li, 2015). However, most of the classical GCM solutions focus on a univariate response, or a low and fixed dimensional response scenario, and none tackles the global testing and multiple testing problems simultaneously. More recently, Wang et al. (2025) studies the multi-response GCM, but focus on estimation only, without considering inference.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



In contrast, we target both types of testing problems in a high-dimensional response scenario. The second line of related research is inference for high-dimensional linear mixed-effects model, since GCM can essentially be rewritten as a mixed-effects model. There have been a number of proposals toward this end (Ma et al., 2013; Bradic et al., 2020; Law and Ritov, 2023; Li et al., 2022). However, they mostly focus on testing a single fixed-effect coefficient, and none considers the global and multiple testing problems either. Moreover, they usually obtain an estimator of the fixed-effects through a proxy of the covariance matrix, which may result in efficiency loss. In comparison, we borrow the region and time information to obtain an appropriate estimator for the true covariance matrix, which improves the power of the subsequent inferences. The third line of related research is high-dimensional inference, including the global and multiple testing procedures (Cai et al., 2013, 2014; Liu, 2013; Xia et al., 2018), and the testing procedures involving a Kronecker product structure (Xia and Li, 2017, 2019; Chen and Liu, 2019; Chen et al., 2023). We also adopt the Kronecker product structure to facilitate our analysis. But we target a completely different problem from the existing solutions. As a result, we develop an utterly different set of inferential techniques to analyze the large and complex dependence structure among the test statistics, and then derive the normal approximations for the global null distribution and total false discoveries. In summary, our proposal addresses a particularly important class of scientific questions, and makes a useful addition to the toolbox of longitudinal data modeling.

We adopt the following notations in this article. For a vector $a = (a_1, \ldots, a_p)^\intercal$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



let $[a]_j$ denote its $j$th entry and $\|a\|_q = (\sum_{j=1}^{p} |a_j|^q)^{1/q}$. For a matrix $A$, let $[A]_{j_1,j_2}$ denote its $(j_1, j_2)$th entry. Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues of $A$, respectively. Let $\|A\|_{\max} = \max_{j,k} |A_{jk}|$, $\|A\|_{1,1} = \sum_{j,k} |A_{jk}|$, and $\|A\|_q = \sup_{\|a\|_q=1} \|Aa\|_q$ for $q \geq 1$. Let $\|A\|_0$ denote the number of nonzero entries in $A$, and let $\text{diag}(A)$ denote the diagonal matrix with its diagonal equal to the diagonal entries of $A$. For two positive sequences $a_n$ and $b_n$, $a_n \lesssim b_n$ means that there exists a constant $c > 0$, such that $a_n \leq c b_n$ for all $n$; $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$, and $a_n \ll b_n$ if $\limsup_{n\to\infty} a_n/b_n = 0$.

The rest of the article is organized as follows. Section 2 presents the model and introduces our proposed estimation and inference procedures. Section 3 establishes the theoretical properties. Section 4 reports the simulations, and Section 5 presents an application to a longitudinal AD study. Section 6 concludes the paper, and the Supplementary Material collects all technical proofs and additional numerical results.

## 2. Model Estimation and Inference

### 2.1 High-dimensional multi-response GCM

Suppose there are $N$ subjects, $R$ response variables, and for subject $i$, there are longitudinal data collected at $T_i$ time points, $i = 1, \ldots, N$. In this article, we only consider the scenario when $T_1 = \cdots = T_N = T$, and leave the scenario when the subjects have varying numbers of observations as future research. Let $y_{i,r,t} \in \mathbb{R}$ denote the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$r$th response variable, $g_{i,t} \in \mathbb{R}$ the time variable, $x_i \in \mathbb{R}^p$ the time-invariant predictor vector, and $z_{i,t} \in \mathbb{R}^q$ the time-variant predictor vector, for subject $i$ at time $t$, $i = 1, \ldots, N, r = 1, \ldots, R, t = 1, \ldots, T$. For our motivating examples, $y_{i,r,t}$ corresponds to the individual brain region or gene, $g_{i,t}$ is the age variable, $x_i$ collects the binary AD status, and other time-invariant covariates such as sex and education level, and $z_{i,t}$ collects the time-variant covariates such as the cognitive scores. We consider the following classical two-level GCM, at level 1,

$$\text{Level 1:} \quad y_{i,r,t} = \beta_{0,i,r} + \beta_{1,i,r} g_{i,t} + \xi_r^\intercal z_{i,t} + \epsilon_{i,r,t}, \tag{2.1}$$

where $\beta_{0,i,r}, \beta_{1,i,r} \in \mathbb{R}$ are the individual-level initial state and the growth rate of the mean growth curve for subject $i$ at region $r$, respectively, $\xi_r \in \mathbb{R}^q$ is the time-invariant fixed-effect of $z_{i,t}$, and $\epsilon_{i,r,t} \in \mathbb{R}$ is the random error, and at level 2,

$$\begin{aligned} \text{Level 2:} \quad \beta_{0,i,r} &= \mu_{0,r} + \gamma_{0,r}^\intercal x_i + \zeta_{0,i,r}, \\ \beta_{1,i,r} &= \mu_{1,r} + \gamma_{1,r}^\intercal x_i + \zeta_{1,i,r}, \end{aligned} \tag{2.2}$$

where $\mu_{0,r}, \mu_{1,r}$ are the population-level intercepts, and $\gamma_{0,r}, \gamma_{1,r} \in \mathbb{R}^p$ are the population-level slopes for the initial state and the growth rate of the mean growth curve of region $r$, respectively, and $\zeta_{0,i,r}, \zeta_{1,i,r} \in \mathbb{R}$ are the random errors.

We assume the random errors $\epsilon_{i,r,t}, \zeta_{0,i,r}, \zeta_{1,i,r}$ follow a mean zero normal distribution, that is,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$$
\begin{aligned}
&(\epsilon_{i,1,1},\ldots,\epsilon_{i,1,T},\ldots,\epsilon_{i,R,1},\ldots,\epsilon_{i,R,T})^{\mathsf{T}} \sim \text{Normal}(0,\Sigma_R\otimes\Sigma_T),\\
&(\zeta_{0,i,r},\zeta_{1,i,r})^{\mathsf{T}} \sim \text{Normal}(0,\Sigma_\zeta),
\end{aligned}
\tag{2.3}
$$

where $\otimes$ denotes the Kronecker product, $\Sigma_R \in \mathbb{R}^{R\times R}$ captures the spatial correlation among different regions or genes, $\Sigma_T \in \mathbb{R}^{T\times T}$ captures the temporal correlation across different time points, and $\Sigma_\zeta \in \mathbb{R}^{2\times 2}$ captures the random departure of individual subjects from the population-level intercept and slope. Here we introduce the Kronecker structure to simplify the covariance of the random error $\epsilon_{i,r,t}$. Such a structure has been often adopted in the literature (see, e.g., Yin and Li, 2012; Leng and Tang, 2012; Xia and Li, 2017). To make $\Sigma_R$ and $\Sigma_T$ identifiable, we assume that $\text{tr}(\Sigma_T) = T$, without loss of generality. Additionally, we assume that $\text{cov}(\beta_{d,i,r_1},\epsilon_{i,r_2,t}) = 0$ for $d = 0,1$ and $r_1, r_2 = 1,\ldots,R$, and $\text{cov}(\zeta_{d_1,i,r_1},\zeta_{d_2,i,r_2}) = 0$ for $d_1, d_2 = 0,1$ and $r_1 \neq r_2$, which are commonly imposed in the GCM literature (see, e.g., Hox and Stoel, 2005; Snijders and Bosker, 2011).

Plugging (2.2) into (2.1), we obtain that

$$
\begin{aligned}
y_{i,r,t} &= \mu_{0,r} + \mu_{1,r}g_{i,t} + \gamma_{0,r}^{\mathsf{T}}x_i + \gamma_{1,r}^{\mathsf{T}}g_{i,t}x_i + \xi_r^{\mathsf{T}}z_{i,t} + \zeta_{0,i,r} + \zeta_{1,i,r}g_{i,t} + \epsilon_{i,r,t}\\
&= \big(\mu_{0,r},\mu_{1,r},\gamma_{0,r}^{\mathsf{T}},\gamma_{1,r}^{\mathsf{T}},\xi_r^{\mathsf{T}}\big)\big(1,g_{i,t},x_i^{\mathsf{T}},g_{i,t}x_i^{\mathsf{T}},z_{i,t}^{\mathsf{T}}\big)^{\mathsf{T}} + \Big\{\big(\zeta_{0,i,r},\zeta_{1,i,r}\big)^{\mathsf{T}}\big(1,g_{i,t}\big) + \epsilon_{i,r,t}\Big\}
\end{aligned}
\tag{2.4}
$$

Write $y_r = (y_{1,r,1},\ldots,y_{1,r,T},\ldots,y_{N,r,1},\ldots,y_{N,r,T})^{\mathsf{T}} \in \mathbb{R}^{TN}$, $\beta^{(r)} = (\eta_r^{\mathsf{T}},\xi_r^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{2p+q+2}$, $\eta_r = (\mu_{0,r},\mu_{1,r},\gamma_{0,r}^{\mathsf{T}},\gamma_{1,r}^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{2p+2}$, $\epsilon_r = (\epsilon_{1,r,1},\ldots,\epsilon_{1,r,T},\ldots,\epsilon_{N,r,1},\ldots,\epsilon_{N,r,T})^{\mathsf{T}} \in \mathbb{R}^{TN}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$\zeta_r = (\zeta_{0,1,r}, \zeta_{1,1,r}, \ldots, \zeta_{0,N,r}, \zeta_{1,N,r})^\intercal \in \mathbb{R}^{2N}$, $r = 1, \ldots, R$, and

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \in \mathbb{R}^{TN \times (2p+q+2)}, \qquad G = \operatorname{diag}\left(\{G_i\}_{i=1}^N\right) = \begin{pmatrix} G_1 & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & G_N \end{pmatrix} \in \mathbb{R}^{TN \times 2N},$$

$$X_i = \begin{pmatrix} 1 & g_{i,1} & x_i^\intercal & g_{i,1}x_i^\intercal & z_{i,1}^\intercal \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & g_{i,T} & x_i^\intercal & g_{i,T}x_i^\intercal & z_{i,T}^\intercal \end{pmatrix} \in \mathbb{R}^{T \times (2p+q+2)}, \; G_i = \begin{pmatrix} 1 & g_{i,1} \\ \vdots & \vdots \\ 1 & g_{i,T} \end{pmatrix} \in \mathbb{R}^{T \times 2}, \; i = 1, \ldots, N.$$

Then, the model in (2.4) can be written in the matrix form as,

$$(y_1, \ldots, y_R) = X\left(\beta^{(1)}, \ldots, \beta^{(R)}\right) + (G\zeta_1 + \epsilon_1, \ldots, G\zeta_R + \epsilon_R), \tag{2.5}$$

and the covariance matrix of $G\zeta_r + \epsilon_r$ is of the form,

$$\Sigma^{(r)} = G(I_N \otimes \Sigma_\zeta)G^\intercal + \operatorname{diag}\left(\{[\Sigma_R]_{r,r}\Sigma_T\}_{i=1}^N\right) \in \mathbb{R}^{TN \times TN}, \quad r = 1, \ldots, R, \tag{2.6}$$

where $I_N \in \mathbb{R}^{N \times N}$ is the identity matrix, and $\operatorname{diag}\left(\{[\Sigma_R]_{r,r}\Sigma_T\}_{i=1}^N\right)$ is a block-diagonal matrix with all the blocks equal to the same matrix $[\Sigma_R]_{r,r}\Sigma_T \in \mathbb{R}^{T \times T}$.

Our goal is to estimate $\beta^{(r)}$ and $\Sigma^{(r)}$, then infer the parameters regarding the mean growth curves. Specifically, our inference aims at the population-level intercepts and slopes of the initial state and the growth rate of the mean growth curve of region $r$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



i.e., $\eta_r = (\mu_{0,r}, \mu_{1,r}, \gamma_{0,r}^\intercal, \gamma_{1,r}^\intercal)^\intercal$. We study the global test and see if the population-level growth curves are mean zero and unaffected by the predictors $x_i$ for all response variables $y_r$'s,

$$H_0 : (\eta_1, \ldots, \eta_R) = 0 \quad \text{versus} \quad H_1 : (\eta_1, \ldots, \eta_R) \neq 0. \tag{2.7}$$

We also study multiple individual tests, and aim to identify the nonzero population mean effects, and the subset of predictors $x_i$ that affect some response variable $y_r$,

$$H_{0,r,j} : [\eta_r]_j = 0 \quad \text{versus} \quad H_{1,r,j} : [\eta_r]_j \neq 0, \quad r = 1, \ldots, R, \ j = 1, \ldots, 2p + 2. \tag{2.8}$$

Meanwhile, we aim to control the false discovery rate (FDR) and the false discovery proportion (FDP). We remark that no existing literature simultaneously tackles the inference problems (2.7) and (2.8) in a high-dimensional multi-response GCM setting.

### 2.2 Parameter estimation

We recognize the key challenge of parameter estimation for our model is the covariance matrix $\Sigma^{(r)}$ in (2.6), for $r = 1, \ldots, R$, as it involves a large number of unknown parameters when the dimension of the response $R$ is large, while the number of subjects $N$ and the number of time points $T$ are often limited. In addition, it involves three covariance matrices, $\Sigma_R, \Sigma_T, \Sigma_\zeta$, which need to be decoupled from each other and be estimated separately. We next develop a novel five-step procedure to estimate $\Sigma^{(r)}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



In Step 1, we first estimate the off-diagonal elements of the spatial covariance matrix $\Sigma_R$. We comment that only the diagonal elements of $\Sigma_R$ are required in $\Sigma^{(r)}$ in (2.6), but the off-diagonal elements of $\Sigma_R$ turn out to be useful for the estimation of $\Sigma_T, \Sigma_\zeta$, and are also easier to estimate than the diagonal elements. Note that, for each individual $i$, the spatial variance averaging over the time points can be written as,

$$\Sigma_{1,i} = \frac{1}{T}\sum_{t=1}^{T} \text{var}\left\{(y_{i,1,t}, \ldots, y_{i,R,t})^\intercal\right\} = \frac{\text{tr}(G_i \Sigma_\zeta G_i^\intercal)}{T} I_R + \Sigma_R,$$

for $i = 1, \ldots, N$. A sample variance estimator of $\Sigma_{1,i}$ is,

$$\hat{\Sigma}_{1,i} = \frac{1}{T}\sum_{t=1}^{T} (\check{y}_{i,1,t}, \ldots, \check{y}_{i,R,t})^\intercal (\check{y}_{i,1,t}, \ldots, \check{y}_{i,R,t}) \in \mathbb{R}^{R\times R}. \tag{2.9}$$

where $\check{y}_{i,r,t} = y_{i,r,t} - \bar{y}_{r,t}$ is the centered response across subjects with $\bar{y}_{r,t} = N^{-1}\sum_{i=1}^{N} y_{i,r,t}$. Here the centering is with respect to subjects, as the subjects are independent, but different regions and time points are not. We use the same centering for other sample variance and covariance estimators later, and we show they achieve the desirable convergence rates. We also note that the first term in $\Sigma_{1,i}$ is diagonal, and we can pool both individual and time information to estimate the off-diagonal elements of $\Sigma_R$ as,

$$[\hat{\Sigma}_R]_{r_1,r_2} = \left[\hat{\Sigma}_1 - \text{diag}(\hat{\Sigma}_1)\right]_{r_1,r_2}, \text{ with } \hat{\Sigma}_1 = \frac{1}{N}\sum_{i=1}^{N} \hat{\Sigma}_{1,i}, \quad r_1, r_2 = 1, \ldots, R, r_1 \neq r_2. \tag{2.10}$$

In Step 2, we estimate the temporal covariance matrix $\Sigma_T$. Note that, for each

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



individual $i$, the temporal covariance between two regions can be written as,

$$\Sigma_{2,r_1,r_2} = \text{cov}\left\{(y_{i,r_1,1},\ldots,y_{i,r_1,T})^\intercal,(y_{i,r_2,1},\ldots,y_{i,r_2,T})^\intercal\right\} = [\Sigma_R]_{r_1,r_2}\Sigma_T,$$

for $i = 1,\ldots,N, r_1, r_2 = 1,\ldots,R, r_1 \neq r_2$. A sample covariance estimator of $\Sigma_{2,r_1,r_2}$ is,

$$\hat{\Sigma}_{2,r_1,r_2} = \frac{1}{N}\sum_{i=1}^{N}(\check{y}_{i,r_1,1},\ldots,\check{y}_{i,r_1,T})^\intercal(\check{y}_{i,r_2,1},\ldots,\check{y}_{i,r_2,T}) \in \mathbb{R}^{T\times T}, \tag{2.11}$$

where $\check{y}_{i,r,t}$ is as defined before. Therefore, we can pool both individual and region information to estimate $\Sigma_T$, based on the average of $\hat{\Sigma}_{2,r_1,r_2}$ across the pairs $(r_1, r_2)$, along with the estimator $[\hat{\Sigma}_R]_{r_1,r_2}$ from Step 1. In addition, to help achieve a desired convergence rate for our estimator of $\Sigma_T$, we propose to choose the set of pairs $(r_1, r_2)$ with the largest $K$ off-diagonal entries $[\hat{\Sigma}_R]_{r_1,r_2}$ among all $r_1, r_2 = 1,\ldots,R, r_1 < r_2$, and we denote this set as $\mathcal{S}$. Our study suggests that, as long as $K$ is of the same order of $R$, the corresponding estimator has a desired convergence rate, and thus we set $K = R$ in our implementation. That is, we estimate the temporal covariance matrix $\Sigma_T$ as,

$$\hat{\Sigma}_T = \frac{1}{R}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{\hat{\Sigma}_{2,r_1,r_2}}{[\hat{\Sigma}_R]_{r_1,r_2}}. \tag{2.12}$$

In Step 3, we estimate the covariance matrix $\Sigma_\zeta$. Note that, for each individual $i$, the temporal variance averaging over regions can be written as,

$$\Sigma_{3,i} = \frac{1}{R}\sum_{r=1}^{R}\text{var}\left\{(y_{i,r,1},\ldots,y_{i,r,T})^\intercal\right\} = G_i\Sigma_\zeta G_i^\intercal + \frac{\text{tr}(\Sigma_R)}{R}\Sigma_T,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



for $i = 1, \ldots, N$. A sample variance estimator of $\Sigma_{3,i}$ is,

$$\hat{\Sigma}_{3,i} = \frac{1}{R}\sum_{r=1}^{R}(\check{y}_{i,r,1}, \ldots, \check{y}_{i,r,T})^{\intercal}(\check{y}_{i,r,1}, \ldots, \check{y}_{i,r,T}) \in \mathbb{R}^{T\times T}. \tag{2.13}$$

Thus, to estimate $\Sigma_\zeta$, we need to estimate $\kappa = \mathrm{tr}(\Sigma_R)/R$, and plug in the estimate of $\Sigma_T$ from Step 2. Note that, as long as the number of time points $T \geq 3$, there exists a vector $u_i \in \mathbb{R}^T$, such that $\|u_i\|_2 = 1$, and $u_i^{\intercal}G_i = (0, 0)$, for $i = 1, \ldots, N$. Correspondingly, $u_i^{\intercal}\Sigma_{3,i}u_i = \kappa(u_i^{\intercal}\Sigma_T u_i)$. Therefore, we estimate $\kappa$ as,

$$\hat{\kappa} = \frac{\sum_{i=1}^{N} u_i^{\intercal}\hat{\Sigma}_{3,i}u_i}{\sum_{i=1}^{N} u_i^{\intercal}\hat{\Sigma}_T u_i}. \tag{2.14}$$

Similarly, as long as the two columns of $G_i$ are linearly independent for $i = 1, \ldots, N$, there exist vectors $v_{i,1}, v_{i,2} \in \mathbb{R}^T$, such that $v_{i,1}^{\intercal}G_i = (1, 0)$ and $v_{i,2}^{\intercal}G_i = (0, 1)$. Correspondingly, $v_{i,j_1}^{\intercal}\Sigma_{3,i}v_{i,j_2} = [\Sigma_\zeta]_{j_1,j_2} + \kappa(v_{i,j_1}^{\intercal}\Sigma_T v_{i,j_2})$, for $j_1, j_2 = 1, 2$. Therefore, we estimate $\Sigma_\zeta$ as,

$$[\hat{\Sigma}_\zeta]_{j_1,j_2} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(v_{i,j_1}^{\intercal}\hat{\Sigma}_{3,i}v_{i,j_2}\right) - \hat{\kappa}\left(v_{i,j_1}^{\intercal}\hat{\Sigma}_T v_{i,j_2}\right)\right], \quad j_1, j_2 = 1, 2. \tag{2.15}$$

In Step 4, we estimate the diagonal elements of the spatial covariance matrix $\Sigma_R$. Recall $\mathrm{tr}(\Sigma_{1,i}) = \mathrm{tr}(G_i\Sigma_\zeta G_i^{\intercal})R/T + \mathrm{tr}(\Sigma_R)$. Thus we estimate $(NT)^{-1}\sum_{i=1}^{N}\mathrm{tr}(G_i\Psi G_i^{\intercal})$ by $\mathrm{tr}(\hat{\Sigma}_1)/R - \hat{\kappa}$. Correspondingly, we estimate the diagonal elements of $\Sigma_R$ as

$$[\hat{\Sigma}_R]_{r,r} = \left[\mathrm{diag}(\hat{\Sigma}_1) - \mathrm{diag}\left\{\mathrm{tr}(\hat{\Sigma}_1)/R - \hat{\kappa}\right\}\right]_{r,r}, \quad r = 1, \ldots, R. \tag{2.16}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



**Algorithm 1** Covariance estimation procedure

1: Estimate the off-diagonal elements of the spatial covariance matrix $\Sigma_R$ via (2.10).
2: Estimate the temporal covariance matrix $\Sigma_T$ via (2.12).
3: Estimate the covariance matrix $\Sigma_\zeta$ via (2.15).
4: Estimate the diagonal elements of the spatial covariance matrix $\Sigma_R$ via (2.16).
5: Estimate $\Sigma^{(r)}$ by plugging the estimates $\hat{\Sigma}_T$, $\hat{\Sigma}_\zeta$, and $[\hat{\Sigma}_R]_{r,r}$ into (2.6).

Finally, in Step 5, we plug the estimates $\hat{\Sigma}_T$, $\hat{\Sigma}_\zeta$, and $[\hat{\Sigma}_R]_{r,r}$ into (2.6) to obtain an estimate $\hat{\Sigma}^{(r)}$ of $\Sigma^{(r)}$. We summarize our estimation procedure in Algorithm 1.

Once obtaining $\hat{\Sigma}^{(r)}$, we estimate $\beta^{(r)}$ via least-squares straightforwardly as,

$$\hat{\beta}^{(r)} = \left[X^\intercal \left\{\hat{\Sigma}^{(r)}\right\}^{-1} X\right]^{-1} X^\intercal \left(\hat{\Sigma}^{(r)}\right)^{-1} y_r, \quad r = 1, \ldots, R. \tag{2.17}$$

We make a few remarks regarding our estimation procedure.

First, we note that, to obtain a good estimator for $\Sigma^{(r)}$, we need to respectively estimate $\Sigma_R, \Sigma_T$, and $\Sigma_\zeta$. There are different options to estimate $\Sigma_T$. One is to pool $NR$ correlated samples together, e.g., through averaging over individual $\hat{\Sigma}_{3,i}$ in (2.13). However, $\Sigma_T$ and $\Sigma_\zeta$ are not easily separable if we go this way. Recognizing that the random departures are uncorrelated with each other across regions, we take another option, by first obtaining the time covariance across different regions, i.e., $[\Sigma_R]_{r_1,r_2}\Sigma_T$, then pooling $TN$ correlated samples to estimate the off-diagonal elements of $\Sigma_R$, then averaging over individuals and different region pairs to estimate $\Sigma_T$. Subsequently, we pool $NR$ samples to estimate $\Sigma_\zeta$, and pool $NT$ samples to estimate $[\Sigma_R]_{r,r}$. Such sample pooling steps are crucial to ensure an accurate estimation of $\Sigma_R, \Sigma_T$, and $\Sigma_\zeta$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



Second, we estimate $\Sigma_{1,i}$ and $\Sigma_{3,i}$ via (2.9) and (2.13), respectively. However, we note that, $\mathbb{E}(y_{i,r,t}) = [X]_{(i-1)T+t,\cdot}\beta^{(r)}$, where $[X]_{(i-1)T+t,\cdot}$ is the $\{(i-1)T+t\}$th row of $X$, and thus the means of $y_{i,r,t}$ are different for all $(i,r,t), i = 1,\ldots,N, r = 1,\ldots,R, t = 1,\ldots,T$. Consequently, both $\hat{\Sigma}_{1,i}$ and $\hat{\Sigma}_{3,i}$ are not the usual sample covariance matrices. We later develop a set of new tools to establish their convergence rates in Theorem 1.

Last but not least, in our motivating applications, the number of time-invariant and time-variant predictors $(p,q)$ are both relatively small compared to the product of the sample size and the number of time points $NT$. As such, we simply estimate $\beta^{(r)}$ using the least squares (2.17). Meanwhile, our method can be extended to more general scenarios in a straightforward fashion. For instance, if $\max(p,q)$ is large, we can apply the debiased Lasso estimator (Zhang and Zhang, 2014) by imposing certain sparsity structures on $\beta^{(r)}$'s. We also note that, one may replace the estimator in (2.17) with a best linear unbiased estimator. Nevertheless, it requires inverting a large $NTR \times NTR$ covariance matrix, with a much higher computational complexity of $O(NTR)^3$ and possibly reduced estimation accuracy. In contrast, our estimator in (2.17) involves inverting an $NT \times NT$ covariance matrix $R$ times, leading to a reduced computational complexity of $O(NT)^3R$ and an improved accuracy, especially when $R$ is large.

### 2.3 Hypothesis testing

We next develop a global testing procedure for the hypothesis in (2.7), then a multiple testing procedure for the hypotheses in (2.8).

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



**Algorithm 2** Global testing procedure

1: Compute the test statistics via (2.18).
2: Define the global test,

$$\Psi_\alpha = I\left(J \geq 2\log\tilde{p} - \log\log\tilde{p} + q_\alpha\right),$$

where $q_\alpha = -\log\pi - 2\log\{\log(1-\alpha)^{-1}\}$, $I(\cdot)$ is the indicator function, and $\alpha$ is the pre-specified significance level.
3: Reject the null hypothesis $H_0$ in (2.7) if $\Psi_\alpha = 1$.

For global testing, we consider the following test statistic,

$$J = \max_{r=1,\ldots,R, j=1,\ldots,2p+2} J_{r,j}^2, \quad \text{where } J_{r,j}^2 = \frac{\left(\left[\hat{\beta}^{(r)}\right]_j\right)^2}{\left[\left[X^\intercal\left\{\hat{\Sigma}^{(r)}\right\}^{-1}X\right]^{-1}\right]_{j,j}}. \tag{2.18}$$

We then propose a global testing procedure as summarized in Algorithm 2. The test is built on the asymptotic property of the test statistic $J$, as we establish in Section 3. Intuitively, with some mild dependence conditions, $\{J_{r,j} : r = 1,\ldots,R, j = 1,\ldots,2p+2\}$ are close to weakly dependent standard normal random variables under $H_0$. Therefore, the proposed test statistic $J$, which takes the form of the maximum of the square of $J_{r,j}$'s, should be close to $2\log\tilde{p}$ under the null hypothesis, where $\tilde{p} = (2p+2)R$.

For multiple testing, the key is to control the false discovery, and we propose a multiple testing procedure as summarized in Algorithm 3. Let $\tau$ denote the threshold value such that $H_{0,r,j}$ is rejected if $|J_{r,j}| \geq \tau$, $r = 1,\ldots,R, j = 1,\ldots,2p+2$. Let $\mathcal{H}_0 = \{(r,j) : \beta_j^{(r)} = 0, r = 1,\ldots,R, j = 1,\ldots,2p+2\}$ denote the set of true null hypotheses and let $\mathcal{H} = \{(r,j) : r = 1,\ldots,R, j = 1,\ldots,2p+2\}$. Then the FDP and

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



**Algorithm 3** Multiple testing procedure

1: Compute the individual test statistics $J_{r,j}$ via (2.18), for $r = 1, \ldots, R, j = 1, \ldots, 2p+2$.
2: Estimate the FDP by

$$\widehat{\mathrm{FDP}}(\tau) = \frac{2\{1-\Phi(\tau)\}\tilde{p}}{\sum_{(r,j)\in\mathcal{H}} I(|J_{r,j}| > \tau) \vee 1}.$$

3: Compute the threshold value

$$\hat{\tau} = \inf\left\{0 \le \tau \le t_{\tilde{p}} : \widehat{\mathrm{FDP}}(\tau) \le \alpha\right\}, \quad \text{where} \quad t_{\tilde{p}} = \left(2\log\tilde{p} - 2\log\log\tilde{p}\right)^{1/2}.$$

If $\hat{\tau}$ dos not exist, set $\hat{\tau} = (2\log\tilde{p})^{1/2}$.
4: Reject $H_{0,r,j}$ if $|J_{r,j}| \ge \hat{\tau}$, for $r = 1, \ldots, R, j = 1, \ldots, 2p+2$.

the FDR are, respectively,

$$\mathrm{FDP}(\tau) = \frac{\sum_{(r,j)\in\mathcal{H}_0} I(|J_{r,j}| > \tau)}{\sum_{(r,j)\in\mathcal{H}} I(|J_{r,j}| > \tau) \vee 1}, \qquad \mathrm{FDR} = \mathbb{E}\{\mathrm{FDP}(\tau)\}.$$

We aim to find a threshold $\tau$ so that we can reject as many true positives as possible while controlling the estimated FDP at the pre-specified level $\alpha$. We note that, since the set of true nulls $\mathcal{H}_0$ is unknown, we estimate $|\mathcal{H}_0|$ by $\tilde{p}$ under the belief that the nulls are dominant among the tests. Subsequently, we estimate the numerator in FDP by $2\{1-\Phi(\tau)\}\tilde{p}$ based on normal approximation. We also note that, the number of false discoveries may not be appropriately estimated if $\hat{\tau}$ exceeds a certain threshold (Xia et al., 2018). We thus set a range $[0, t_{\tilde{p}}]$ for selecting $\hat{\tau}$, and threshold it at $(2\log\tilde{p})^{1/2}$ if it is not attained in the range.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



## 3. Theoretical Properties

### 3.1 Regularity conditions

We now study the theoretical properties of the proposed estimation and inference methods. We clarify that, in our asymptotic setting, we let both the number of subjects $N$ and the number of response variables $R$ to diverge to infinity. We first present a set of regularity conditions, then examine these conditions in detail.

(C1) Suppose $c_1^{-1} \leq (NT)^{-1}\lambda_{\min}(X^\intercal X) \leq (NT)^{-1}\lambda_{\max}(X^\intercal X) \leq c_1$, $\|G\|_{\max} \leq c_1$, and $\min_{1\leq i\leq N} \lambda_{\min}(G_i^\intercal G_i) \geq c_1^{-1}$, for some constant $c_1 > 0$.

(C2) Suppose $c_2^{-1} \leq \lambda_{\min}(\Sigma_T) \leq \lambda_{\max}(\Sigma_T) \leq c_2$, $c_2^{-1} \leq \lambda_{\min}(\Sigma_R) \leq \lambda_{\max}(\Sigma_R) \leq c_2$, and $T\lambda_{\max}(\Sigma_\zeta) \leq c_2$, for some constant $c_2 > 0$.

(C3) Suppose there exist at least $R$ entries of $[\Sigma_R]_{r_1,r_2}$, $r_1 < r_2$, such that $|[\Sigma_R]_{r_1,r_2}| \geq c_3$ for some constant $c_3 > 0$.

(C4) Denote $\Sigma^{(r_1,r_2)} = \text{diag}\left(\{[\Sigma_R]_{r_1,r_2}\Sigma_T\}_{i=1}^N\right)$, for $r_1, r_2 = 1, \ldots, R, r_1 \neq r_2$, $\tilde{\Sigma}^{(r_1,r_2)} = \{X^\intercal(\Sigma^{(r_1,r_1)})^{-1}X\}^{-1}X^\intercal\{\Sigma^{(r_1,r_1)}\}^{-1}\Sigma^{(r_1,r_2)}[\{\Sigma^{(r_2,r_2)}\}^{-1}X]^{-1}[X^\intercal\{\Sigma^{(r_2,r_2)}\}^{-1}X]^{-1}$, and $D_r = \text{diag}[(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}]$, for $r = 1, \ldots, R$. Denote $\check{\Sigma}^{(r_1,r_2)} = D_{r_1}^{-1/2}\tilde{\Sigma}^{(r_1,r_2)}D_{r_2}^{-1/2}$. Suppose $\max_{r_1,r_2}\max_{j_1<j_2}\left|[\check{\Sigma}^{(r_1,r_2)}]_{j_1,j_2}\right| \leq c_4$, and $\max_{r_1\neq r_2}\max_{j_1=j_2}\left|[\check{\Sigma}^{(r_1,r_2)}]_{j_1,j_2}\right| \leq c_4$, for some constant $0 < c_4 < 1$.

(C5) Denote $\beta_{-1}^{(r)} \in \mathbb{R}^{2p+q+1}$ as the sub-vector of $\beta^{(r)}$ with the first entry $\mu_{0,r}$ removed, and $B = \left(\beta_{-1}^{(1)}, \ldots, \beta_{-1}^{(R)}\right) \in \mathbb{R}^{(2p+q+1)\times R}$. Let $s_B = \|B\|_0$, $c_B = \|B\|_{\max}$, and

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$c_R = \max_{1\le r\le R} \|\beta_{-1}^{(r)}\|_2^2$. Let $\theta_{N,T,R,B} = T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 TR^{-1}]$. Suppose $T(\log \tilde{p})\ \theta_{N,T,R,B} = o(1)$ as $N, R \to \infty$.

Condition (C1) imposes some mild bounded eigenvalue requirement on the design matrix $X$. In this article, we treat $X$ as deterministic, and similar conditions have been commonly imposed; see, e.g., Zhang and Zhang (2014). When the design matrix $X$ is random, we can simply replace this condition with the bounded eigenvalue requirement on the covariance of $X$. Moreover, the condition on $\{G_i^{\intercal} G_i, i = 1, \ldots, N\}$ is mild, as each of them is a $2 \times 2$ matrix. Condition (C2) is placed on the eigenvalues of $\Sigma_R$ and $\Sigma_T$, which is relatively mild too, as it requires that most of the variables are not highly correlated with each other across regions or over the time points. In addition, the condition on $T\lambda_{\max}(\Sigma_\zeta)$ ensures the bounded eigenvalue property of the covariance matrix $\Sigma^{(r)}$. Condition (C3) naturally holds for a general region-wise dependence structure, and it is only a sufficient condition for Theorem 1. Condition (C4) requires the correlations are bounded away from $-1$ and 1, and basically excludes the singular cases. Moreover, Conditions (C2) and (C4) are commonly assumed in the high-dimensional literature (e.g., Bickel et al., 2008; Yuan, 2010; Cai et al., 2014; Liu, 2013; Xia et al., 2015). Condition (C5) characterizes the mean variations across subjects, regions and time points. Recall that our covariance estimators in (2.9), (2.11) and (2.13) are not the usual sample covariances, because the means of the response variables are different for different subjects, regions and time points. Condition (C5) ensures that our covari-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



ance estimators perform asymptotically the same as the sample covariance estimators obtained by the i.i.d. centralized error terms. Condition (C5) also imposes some level of sparsity on $B$, in that it requires the number of nonzero entries in $B$ and their maximum magnitude are not too large. On the other hand, we do not impose any particular sparsity structures on the coefficient matrix $B$ nor the covariances $\Sigma_R$, $\Sigma_T$, $\Sigma_\zeta$. As such, we do not employ any penalized procedure such as Lasso in our parameter estimation.

For our motivating applications, these regularity conditions all seem reasonable. Specifically, for the data example in Section 5, we have $N = 56$ subjects, $R = 68$ responses, $T = 3$ time points, $p = 4$ time-invariant predictors, $q = 3$ time-variant predictors, and the time variable, age. Condition (C1) is reasonable, since most of those predictors are likely to be weakly correlated. Conditions (C2)) to (C4) are suitable too, since it is expected that the majority of the response variables are weakly correlated across regions and time points, and those structural-connected ones are moderately correlated across regions (Mechelli et al., 2005). Finally, Condition (C5) is acceptable, since the terms $T^2(\log \tilde{p})\{\log R/(NT)\}^{1/2}$ and $T^2(\log \tilde{p})\{\log T/(NR)\}^{1/2}$ in this condition are both reasonably small in our setting. Moreover, as $N$ and $R$ grow, (C5) assumes a decaying value of $c_R$, which is equivalent as assuming a decaying maximum magnitude $c_B$ in our application because both $p$ and $q$ are small. Since it is reasonable to expect that the difference of mean growth curves between the AD and healthy controls usually concentrate in a few regions and the magnitude of such difference is usually not too large, the terms involving $c_R$, $c_B$ and $s_B$ are reasonable as well.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



### 3.2 Estimation consistency, false discovery control and power

First, we establish the convergence rate of the covariance matrix estimator $\hat{\Sigma}^{(r)}$ from Algorithm 1, which is key for the subsequent size, power and error rate control analyses. For the asymptotics, we let $N$ and $R$ to diverge to infinity, while we allow $T, p, q$ to be either fixed or to diverge, and the relations of $(N, R, T, p, q)$ are regulated by (C5). That is, we have the desired convergence as long as $\theta_{N,T,R,B}$ in (C5) is of order $o(1)$.

**Theorem 1.** *Suppose Conditions (C1) to (C3) hold, and* $\theta_{N,T,R,B} = o(1)$. *Then,*

$$\|\hat{\Sigma}^{(r)} - \Sigma^{(r)}\|_{\max} = O_{\mathbb{P}}(\theta_{N,T,R,B}) = o_{\mathbb{P}}(1).$$

Theorem 1 implies the estimation consistency. Moreover, we note that, because the means of $y_{i,r,t}$ are different for all $\{(i, r, t), i = 1, \ldots N, r = 1, \ldots, R, t = 1, \ldots, T\}$, the asymptotics of the classical sample covariance estimator based on i.i.d. observations are not directly applicable to the sample covariance estimators such as $\hat{\Sigma}_{1,i}$, $\hat{\Sigma}_{2,r_1,r_2}$ and $\hat{\Sigma}_{3,i}$ in (2.9), (2.11) and (2.13). To obtain the desired convergence rate, we first show in Step 1 of the proof of Theorem 1 the mean variations of the response variables are negligible across different $\{(i, r, t)\}$. We then repeatedly apply an advanced version of Hanson-Wright inequality (Chen et al., 2023) to derive the convergence rates of $[\hat{\Sigma}_R]_{r_1,r_2}$, $\hat{\Sigma}_T$, $\hat{\Sigma}_\zeta$, and $[\hat{\Sigma}_R]_{r,r}$ in turn, which eventually leads to the convergence rate of $\hat{\Sigma}^{(r)}$.

Next, we derive the limiting distribution of the test statistic $J$ in (2.18) under the null hypothesis of the global test (2.7). We show that, $(J - 2\log\tilde{p} + \log\log\tilde{p})$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



weakly converges to a Gumbel random variable with cumulative distribution function, $\exp\{-\pi^{-1/2}\exp(-\phi/2)\}$, under the null. We establish the asymptotics when both the product $NT$ and $\tilde{p} = (2p+2)R$ diverge to infinity, which hold automatically when $N$ and $R$ diverge to infinity.

**Theorem 2.** *Suppose Conditions (C1) to (C5) hold. Then for any $\phi \in \mathbb{R}$,*

$$\mathbb{P}_{H_0}\left(J - 2\log\tilde{p} + \log\log\tilde{p} \le \phi\right) \to \exp\left\{-\pi^{-1/2}\exp(-\phi/2)\right\}, \text{ as } NT \text{ and } \tilde{p} \to \infty.$$

Next, we study the asymptotic power of the proposed global test in Algorithm 2. We consider the following class of regression coefficients $\{\beta^{(r)}, r = 1, \ldots, R\}$,

$$\mathcal{U}(c) = \left\{ \{\beta_j^{(r)}\}_{(r,j)\in\mathcal{H}} : \max_{r,j} \frac{|\beta_j^{(r)}|}{[\{X^\intercal(\Sigma^{(r)})^{-1}X\}^{-1}]_{j,j}^{1/2}} \ge c(\log\tilde{p})^{1/2} \right\}.$$

By the bounded eigenvalue conditions in (C1) and (C2), the denominator $\left[[X^\intercal\{\Sigma^{(r)}\}^{-1}X]^{-1}\right]_{j,j}^{1/2}$ is of order $(NT)^{-1/2}$. Therefore, the class $\mathcal{U}(c)$ includes all coefficient matrices that have one of the entries with a magnitude of order $\{\log\tilde{p}/(NT)\}^{1/2}$, where Condition (C5) can be satisfied. As a simple example, $\{\beta^{(r)}, r = 1, \ldots, R\}$ belongs to $\mathcal{U}(c)$ if it only has one nonzero entry, $\beta_{j_0}^{(r_0)}$ with $(r_0, j_0) \in \mathcal{H}$, and $|\beta_{j_0}^{(r_0)}| = C\{\log\tilde{p}/(NT)\}^{1/2}$ for some large enough constant $C > 0$. In such a setting, $s_B = 1$, $c_R$ and $c_B^2$ are both of order $\log\tilde{p}/(NT)$, and henceforth Condition (C5) is easily satisfied.

**Theorem 3.** *Suppose Conditions (C1) to (C3), and (C5) hold. Then,*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$$\inf_{\{\beta_j^{(r)}\}\in\mathcal{U}(2\sqrt{2})} \mathbb{P}(\Psi_\alpha = 1) \to 1, \ \textit{as } NT \textit{ and } \tilde{p} \to \infty.$$

We immediately see that, based on this theorem, if we set the constant $c$ as $2\sqrt{2}$, then our proposed global test enjoys a full power asymptotically.

Finally, we establish the error rate control of the proposed multiple testing procedure in Algorithm 3. We introduce one more regularity condition (C6), which ensures that $\hat{\tau}$ is attained in the range $[0, (2\log\tilde{p} - 2\log\log\tilde{p})^{1/2}]$. We note that Condition (C6) only requires a few entries of $\{\beta_j^{(r)}\}$ to have a magnitude of the order $\{\log\tilde{p}\}^{(1+\rho)/2}/(NT)^{1/2}$, and is thus mild. Moreover, when this condition is not satisfied, the asymptotic FDR control can still be obtained but more conservatively.

(C6) Denote $S_\rho = \left\{(r,j)\in\mathcal{H} : \left\{\beta_j^{(r)}\right\}^2 / \left[\{X^\intercal(\Sigma^{(r)})^{-1}X\}^{-1}\right]_{j,j} \geq (\log\tilde{p})^{1+\rho}\right\}$. Suppose $|S_\rho| \geq \{1/(\pi^{1/2}\alpha) + \delta\}(\log\tilde{p})^{1/2}$ for some $\rho > 0$ and $\delta > 0$.

**Theorem 4.** *Suppose Conditions (C1) to (C6) hold, and* $\tilde{p}_0 = |\mathcal{H}_0| \asymp \tilde{p}$. *Then,*

$$\lim_{(NT,\tilde{p})\to\infty} \frac{FDR}{\alpha\tilde{p}_0/\tilde{p}} = 1, \lim_{(NT,\tilde{p})\to\infty} \frac{FDP(\hat{\tau})}{\alpha\tilde{p}_0/\tilde{p}} = 1 \textit{ in probability.}$$

### 3.3 Extension to sub-Gaussian distribution

In our GCM model, we have assumed that the errors follow a Gaussian distribution as in (2.3). We now extend it to the sub-Gaussian distribution. More specifically, instead of assuming that $\Sigma_T^{-1/2}\epsilon_i\Sigma_R^{-1/2}$ and $\Sigma_\zeta^{-1/2}\zeta_{i,r}$ have i.i.d. Gaussian entries, we assume they have i.i.d. sub-Gaussian entries, and study the corresponding theoretical properties. We

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



begin with a regularity condition.

(C7) Suppose $\log \tilde{p} = o[\{N/\max(p,q)\}^{1/4}]$, $\|X\|_{\max} = O(1)$, $\mathbb{E}\{\exp(\nu\epsilon_{i,r,t}^2)\} \leq c_5$, and $\mathbb{E}\{\exp(\nu\zeta_{d,i,r}^2)\} \leq c_5$, for some constants $\nu, c_5 > 0$, $i = 1, \ldots, N$, $r = 1, \ldots, R$, $t = 1, \ldots, T$, and $d = 0, 1$.

Condition (C7) is mild, as both $p$ and $q$ are small in our targeted settings, and the bounded design matrix is also reasonable.

Next, we establish the theoretical properties under the sub-Gaussian scenario.

**Theorem 5.** *Suppose Condition (C7) holds.*

*(i) Suppose the same conditions in Theorem 2 hold. Then, for any $\phi \in \mathbb{R}$,*

$$\mathbb{P}_{H_0}\left(J - 2\log\tilde{p} + \log\log\tilde{p} \leq \phi\right) \to \exp\left\{-\pi^{-1/2}\exp(-\phi/2)\right\}, \text{ as } \tilde{p} \to \infty.$$

*(ii) Suppose the same conditions in Theorem 3 hold. Then,*

$$\inf_{\{\beta_j^{(r)}\} \in \mathcal{U}(2\sqrt{2})} \mathbb{P}(\Psi_\alpha = 1) \to 1, \text{ as } \tilde{p} \to \infty.$$

*(iii) Suppose the same conditions in Theorem 4 hold, and $(\log\tilde{p})^{7+\varepsilon} = O\{N/\max(p,q)\}$ for some small constant $\varepsilon > 0$. Then,*

$$\lim_{\tilde{p}\to\infty} \frac{FDR}{\alpha\tilde{p}_0/\tilde{p}} = 1, \lim_{\tilde{p}\to\infty} \frac{FDP(\hat{\tau})}{\alpha\tilde{p}_0/\tilde{p}} = 1 \text{ in probability.}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



We remark that, under the sub-Gaussian condition, our test statistic can be written as a sum of $2N + NT$ random variables with unequal variances. In comparison to the well established normal approximation where the test statistic can be expressed as the sum of i.i.d. sub-Gaussian random variables, our case is more challenging, as we need to perform truncations on those $2N+NT$ non-identically distributed random variables, which leads to a more complicated normal approximation on the sum of those variables. See the proof of Theorem 5 for more details.

## 4. Simulation Studies

### 4.1 Simulation setup

We study the finite-sample performance of the proposed method. We also compare with a simple alternative solution that fits one response variable at a time using the restricted maximum likelihood (REML) approach. That is, we obtain the coefficient estimates using REML, compute the test statistic $J_{r,j}$ following (2.18), and carry out the global and multiple testing procedures accordingly. We note that REML is widely employed in classical GCM modeling. Moreover, we have chosen not to numerically compare with the solutions based on high-dimensional linear mixed-effects model such as Li et al. (2022), for several reasons. First, this family of methods all assume i.i.d. random errors, ignore the spatial correlations, and target the inference for a single fixed-effect coefficient. Second, their estimation does not effectively pool the sample information

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



across times or regions as our method does. Finally, their test statistics rely on the cross-validated estimation of a huge proxy covariance matrix with dimension $RTN \times RTN$, and the subsequent $(2p+q+2)R$-dimensional debiasing procedure requires running Lasso $(2p+2)R$ times. As such, the involved computation is prohibitively expensive and extremely slow.

We simulate the model following the setup in (2.5) and (2.6). We set $N = \{100, 200\}$, $T = \{4, 8\}$, $R = \{50, 100\}$, $p = 10$, and $q = 2$. We draw $g_{i,t}$ randomly from Uniform$[0, 1]$, and generate each entry of $x_i$ and $z_{i,t}$ independently from Normal$(0, 1)$, for $i = 1, \ldots, N$, $t = 1, \ldots, T$. We randomly set 5% of the coefficients $\{\xi_r\}$ as non-zero, which equal 0.2 in the global testing case and 0.5 in the estimation evaluations and the multiple testing case.

We consider two structures for the temporal covariance matrix $\Sigma_T$, an autoregressive structure and a moving average structure. Specifically, we begin with $[\Sigma_T^{'}]_{t_1,t_2} = 0.4^{|t_1-t_2|}$ for $1 \le t_1, t_2 \le T$, and $[\Sigma_T^{'}]_{t_1,t_2} = 1/(|t_1 - t_2| + 1)$ for $|t_1 - t_2| \le 3$ and 0 otherwise. We next set $\Sigma_T^{''} = \Sigma_T^{'} \odot V_T$, to adopt different variances among the time points, where $\odot$ is the Hadamard product, and $V_T = u_T u_T^{\intercal}$, with $u_T = (1, 2, 3, 4)^{\intercal}$ for $T = 4$, and $u_T = (1, 2, 3, 4, 1, 2, 3, 4)^{\intercal}$ for $T = 8$. We then set $\Sigma_T = \{T/\mathrm{tr}(\Sigma_T^{''})\}\Sigma_T^{''}$, so that $\mathrm{tr}(\Sigma_T) = T$. We also consider two structures for the spatial covariance matrix $\Sigma_R$. Specifically, we begin with the precision matrix $\Omega_R^{'}$, and consider a hub graph where the nodes are evenly partitioned into disjoint groups with 5 nodes each while there exists one node connecting all the other nodes inside each group, and a small-world

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



graph, with one starting neighbor and 5% probability of rewiring. We set the diagonal values to one, and draw the non-zero entries of the off-diagonal values randomly from Uniform$[-0.6,-0.2]\cup[0.2,0.6]$. We next set $\Omega_R^{''}=(\Omega_R^{'}+\delta_R I_R)/(1+\delta_R)$, where $\delta_R=\max\{0.05,-\lambda_{\min}(\Omega_R^{'})\}$. We then set $\Sigma_R=\{R/\mathrm{tr}(\Omega_R^{''-1})\}\Omega_R^{''-1}$. Finally, we set the random departure covariance $\Sigma_\Psi=T^{-1}\left(\begin{smallmatrix}6&3\\3&9\end{smallmatrix}\right)$.

We also carry out additional simulations to evaluate the performance of our proposed method when the error terms in (2.3) follow a non-Gaussian distribution, when the Kronecker product structure in (2.3) does not hold, and when $R$ is much larger than $N$. We report those results in the Supplementary Material.

### 4.2 Estimation results

We first evaluate the empirical performance of the parameter estimation. We vary the proportion of the non-zero coefficients $\{[\eta_r]_j\}$, i.e., $\omega=1-|\mathcal{H}_0|/\{(2p+2)R\}=\{0.03,0.05\}$, and set the non-zero entries of $\{[\eta_r]_j\}$ equal to 0.5. We repeat the experiment 200 times. To evaluate the estimation accuracy of the covariance matrix $\Sigma^{(r)}$, we report the average and standard error of the bias criterion $\big\{[\hat{\Sigma}^{(r)}]_{b_1,b_2}-[\Sigma^{(r)}]_{b_1,b_2}:1\le r\le R,|b_1-b_2|\le T\big\}$, as $\Sigma^{(r)}$ and $\hat{\Sigma}^{(r)}$ are both block-diagonal matrices. To evaluate the estimation accuracy of the regression coefficient $\beta^{(r)}$, we report the average and standard error of the bias criterion $\big\{\hat{\beta}_j^{(r)}-\beta_j^{(r)}:1\le r\le R,1\le j\le 2p+2\big\}$. We report the results for $T=4$ in Table 1, and the results for $T=8$ in Table S1 of the Supplementary Material in the interest of space. We observe from these tables

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



Table 1: Parameter estimation: the bias and standard error based on 200 data replications for the autoregressive and moving average temporal structures with $T = 4$.

| Temporal structure | | | Autoregressive | | | | Moving average | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\omega = 0.03$ | | $\omega = 0.05$ | | $\omega = 0.03$ | | $\omega = 0.05$ | |
| $R$ | $N$ | accuracy | hub | small | hub | small | hub | small | hub | small |
| Bias and SE of covariance estimation | | | | | | | | | | |
| 50 | 100 | Bias | 0.0798 | 0.0653 | 0.1487 | 0.1411 | 0.0817 | 0.0657 | 0.1467 | 0.1455 |
| | | SE | 0.5043 | 0.3680 | 0.5206 | 0.3879 | 0.4886 | 0.3727 | 0.5054 | 0.3941 |
| | 200 | Bias | 0.0742 | 0.0669 | 0.1205 | 0.1151 | 0.0748 | 0.0664 | 0.1205 | 0.1151 |
| | | SE | 0.2502 | 0.2175 | 0.3255 | 0.2328 | 0.2587 | 0.2291 | 0.3229 | 0.2444 |
| 100 | 100 | Bias | 0.0935 | 0.0887 | 0.1540 | 0.1484 | 0.0960 | 0.0909 | 0.1554 | 0.1504 |
| | | SE | 0.4200 | 0.3787 | 0.4767 | 0.4531 | 0.4211 | 0.3839 | 0.4718 | 0.4520 |
| | 200 | Bias | 0.0831 | 0.0742 | 0.1342 | 0.1328 | 0.0838 | 0.0747 | 0.1336 | 0.1331 |
| | | SE | 0.3638 | 0.2244 | 0.2666 | 0.2401 | 0.3538 | 0.2363 | 0.2753 | 0.2520 |
| Bias and SE of coefficient estimation | | | | | | | | | | |
| 50 | 100 | Bias | 0.0002 | -0.0006 | 0.0005 | 0.0001 | 0.0003 | -0.0006 | 0.0005 | 0.0000 |
| | | SE | 0.1682 | 0.1688 | 0.1688 | 0.1686 | 0.1672 | 0.1682 | 0.1677 | 0.1675 |
| | 200 | Bias | -0.0002 | 0.0002 | 0.0001 | 0.0003 | -0.0002 | 0.0002 | 0.0001 | 0.0003 |
| | | SE | 0.1103 | 0.1101 | 0.1106 | 0.1106 | 0.1097 | 0.1094 | 0.1099 | 0.1099 |
| 100 | 100 | Bias | 0.0002 | 0.0000 | 0.0005 | -0.0003 | 0.0001 | 0.0000 | 0.0005 | -0.0003 |
| | | SE | 0.1688 | 0.1688 | 0.1689 | 0.1689 | 0.1677 | 0.1678 | 0.1678 | 0.1678 |
| | 200 | Bias | 0.0000 | -0.0001 | -0.0001 | -0.0001 | 0.0000 | -0.0001 | -0.0001 | -0.0001 |
| | | SE | 0.1108 | 0.1104 | 0.1104 | 0.1106 | 0.1102 | 0.1097 | 0.1097 | 0.1099 |

that, the bias and standard error of the covariance estimation are larger for a larger $\omega$, while those of the regression coefficient estimation do not change much. This matches our theoretical convergence rate in Theorem 1, as a larger $\omega$ represents a larger $s_B$ and $c_R$, thus a larger $\theta_{N,T,R,B}$. In addition, as the sample size increases, both the bias and the standard error of the coefficient estimation decrease. We also briefly comment that the biases associated with covariance estimation and coefficient estimation differ in magnitude. However, our intention here is not to directly compare these two types

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



Table 2: Global testing: the empirical size and power in percentage based on 2000 replications for the autoregressive and moving average temporal structures with $T = 4$ and 8, and the significance level $\alpha = 0.05$.

| | | | $T = 4$ | | | | $T = 8$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Empirical Size | | Empirical Power | | Empirical Size | | Empirical Power | |
| $R$ | $N$ | Method | hub | small | hub | small | hub | small | hub | small |
| Auto-regressive temporal structure | | | | | | | | | | |
| 50 | 100 | Proposed | 5.6 | 4.2 | 20.5 | 15.7 | 5.4 | 6.0 | 52.4 | 58.1 |
| | | REML | 9.5 | 9.8 | 25.1 | 23.6 | 11.1 | 9.0 | 49.6 | 53.0 |
| | 200 | Proposed | 4.3 | 5.0 | 58 | 55.1 | 5.0 | 5.9 | 99.4 | 99.9 |
| | | REML | 6.0 | 6.2 | 56.1 | 55.5 | 6.9 | 7.0 | 98.6 | 98.3 |
| 100 | 100 | Proposed | 4.3 | 4.4 | 17.7 | 17.3 | 5.8 | 5.0 | 74.5 | 65.3 |
| | | REML | 9.8 | 11.1 | 27.4 | 29 | 12.2 | 10.6 | 69.2 | 67.6 |
| | 200 | Proposed | 4.3 | 3.9 | 61.1 | 67.1 | 5.1 | 4.8 | 100.0 | 99.9 |
| | | REML | 7.0 | 7.6 | 65.4 | 68.3 | 7.2 | 7.0 | 99.7 | 99.4 |
| Moving average temporal structure | | | | | | | | | | |
| 50 | 100 | Proposed | 5.6 | 4.2 | 21.2 | 14.6 | 6.2 | 7.2 | 54.9 | 64.1 |
| | | REML | 9.7 | 9.2 | 24.6 | 23.4 | 10.8 | 10.0 | 47.8 | 51.7 |
| | 200 | Proposed | 4.7 | 4.7 | 58.3 | 54.9 | 5.1 | 6.1 | 99.5 | 99.8 |
| | | REML | 6.1 | 5.9 | 54.8 | 54.2 | 6.9 | 7.6 | 97.9 | 97.4 |
| 100 | 100 | Proposed | 4.7 | 4.0 | 17.7 | 16.4 | 6.7 | 5.7 | 79.0 | 68.3 |
| | | REML | 10.0 | 11.1 | 27.4 | 28.0 | 12.2 | 11.1 | 66.6 | 65.1 |
| | 200 | Proposed | 4.2 | 3.7 | 61.1 | 67.2 | 5.3 | 5.2 | 100.0 | 100.0 |
| | | REML | 7.0 | 7.9 | 63.7 | 66.4 | 7.6 | 7.1 | 99.4 | 99.0 |

of bias. Instead, we examine how each bias changes as $\omega$ increases. We see that the bias of covariance estimation increases as $\omega$ increases, whereas the bias of coefficient estimation remains relatively stable across different values of $\omega$.

### 4.3 Global testing results

We next evaluate the performance of the global testing procedure. To evaluate the size of the test, we set $\omega = 0$, whereas to evaluate the power of the test, we set

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



$\omega = 0.05$ and set the non-zero entries of $\{[\eta_r]_j\}$ equal to 0.2. We set the significance level $\alpha = 0.05$. Table 2 reports the empirical size and power, in percentages, based on 2000 data replications. We see that the empirical size is close to the significance level under all settings, especially for a larger sample size $N$. In contrast, the testing method based on REML has serious size inflation in most settings. For those cases where REML controls the size relatively well, our method achieves a better power. In addition, we also observe that the proposed method achieves a notable power gain when the sample size $N$ or the number of time points $T$ increases, which again agrees with our theoretical findings.

### 4.4 Multiple testing results

We next evaluate the performance of the multiple testing procedure. We adopt the same setting as in Section 4.2, and set the pre-specified FDR level $\alpha = 0.1$. We report the empirical FDR and power in Table 3 based on 200 data replications with $T = 4$, and in Table S2 of the Supplementary Material with $T = 8$. We see from these tables that, for the empirical FDR, the proposed method has FDR well under control across all settings, while the testing method based on REML has some FDR inflation for the cases with a small sample size and a large dimension, e.g., when $N = 100$ and $R = 100$. For the empirical power, the proposed method is in general more powerful than REML, especially when both methods have the FDR under control. We also see that the empirical power of our method increases and the power gain is more apparent

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



Table 3: Multiple testing: the empirical FDR and power in percentage based on 200 data replication for the autoregressive and moving average temporal structures with $T = 4$ and the FDR level $\alpha = 0.1$.

| Temporal structure | | | Autoregressive | | | | Moving average | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\omega = 0.03$ | | $\omega = 0.05$ | | $\omega = 0.03$ | | $\omega = 0.05$ | |
| $R$ | $N$ | method | hub | small | hub | small | hub | small | hub | small |
| Empirical FDR | | | | | | | | | | |
| 50 | 100 | Proposed | 6.82 | 9.23 | 7.47 | 7.07 | 7.06 | 7.02 | 7.18 | 7.01 |
| | | REML | 9.46 | 8.85 | 11.83 | 12.34 | 9.35 | 8.27 | 11.76 | 12.58 |
| | 200 | Proposed | 7.65 | 7.69 | 7.52 | 7.44 | 7.86 | 7.66 | 7.33 | 7.47 |
| | | REML | 10.42 | 10.67 | 10.99 | 11.18 | 10.39 | 10.51 | 10.78 | 11.23 |
| 100 | 100 | Proposed | 7.05 | 7.85 | 6.78 | 6.85 | 7.02 | 7.60 | 6.68 | 6.73 |
| | | REML | 12.48 | 13.18 | 11.74 | 11.96 | 12.53 | 13.31 | 11.78 | 11.89 |
| | 200 | Proposed | 7.17 | 7.99 | 6.85 | 6.56 | 7.33 | 8.07 | 6.79 | 6.68 |
| | | REML | 10.57 | 10.94 | 10.58 | 10.25 | 10.39 | 10.89 | 10.59 | 10.22 |
| Empirical Power | | | | | | | | | | |
| 50 | 100 | Proposed | 34.98 | 35.98 | 48.43 | 47.04 | 37.09 | 35.68 | 48.71 | 47.48 |
| | | REML | 36.55 | 34.42 | 49.78 | 48.84 | 35.89 | 33.32 | 49.54 | 48.55 |
| | 200 | Proposed | 91.52 | 91.67 | 90.68 | 92.54 | 92.21 | 92.15 | 91.41 | 93.01 |
| | | REML | 89.14 | 88.80 | 89.20 | 90.84 | 89.82 | 89.42 | 89.85 | 91.29 |
| 100 | 100 | Proposed | 39.76 | 40.09 | 47.05 | 46.87 | 40.29 | 40.54 | 47.49 | 47.36 |
| | | REML | 41.23 | 41.44 | 49.55 | 49.33 | 41.61 | 41.38 | 49.61 | 49.59 |
| | 200 | Proposed | 90.45 | 91.20 | 92.53 | 93.10 | 90.95 | 91.74 | 93.05 | 93.59 |
| | | REML | 88.70 | 88.34 | 91.32 | 91.60 | 89.31 | 89.15 | 91.89 | 92.06 |

when $N$ or $T$ increases.

## 5. Longitudinal Neuroimaging Analysis

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder characterized by progressive impairment of cognitive functions, then global incapacity and ultimately death. It is the leading form of dementia, and is currently affecting 5.8 million American adults aged 65 years or older. Its prevalence continues to grow, and is projected to

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



reach 13.8 million by 2050 (Alzheimer's Association, 2020). It is thus crucial to better understand, diagnose, and treat this disorder (Jagust, 2018). We analyze a dataset OASIS-2 from Open Access Series of Imaging Studies (`www.oasis-brains.org`). The data consists of a longitudinal collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, multiple T1-weighted MRI scans measuring brain gray matter volume were obtained. The subjects are all right-handed and include both men and women. Among them, 72 were characterized as nondemented throughout the study, and 64 were characterized as demented at the time of their initial visits and remained so for subsequent scans (Marcus et al., 2010). We process the data and only include in our data analysis the subjects with $T = 3$ time points and meeting the quality control criteria. This results in $N = 56$ subjects. We then further process the MRI images and parcellate the brain into $R = 68$ regions-of-interest (ROIs) using the Desikan-Killiany atlas (Desikan et al., 2006). For each subject, we also include the binary AD status, sex, education, and socioeconomic status as the time-invariant covariates with $p = 4$, and the mini-mental state examination score (Folstein et al., 1975), atlas scaling factor (Buckner et al., 2004), and estimated total intracranial volume (Buckner et al., 2004) as the time-variant covariates with $q = 3$.

We apply the proposed method with the pre-specified FDR level $\alpha = 0.05$. The residual plots and the QQ-plots in Section C.5 of the Supplementary Material suggest that the GCM model seems to be a reasonable choice for this data. Our test identifies

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



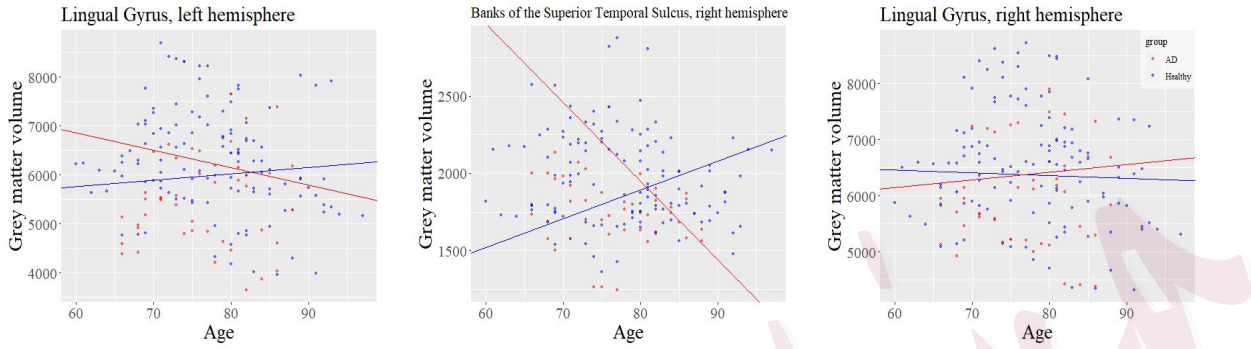


Figure 1: Scattershot for grey matter volume versus age in three identified brain regions. The short lines denote the individual growth curves of 20 randomly selected subjects from both the AD group and the healthy group. The long bold lines denote the overall growth curves of the two groups.

50 significant coefficients among the total of 680 coefficients. We focus on the ones associated with the binary AD status, and three brain regions are identified, i.e., the lingual gyrus of the left hemisphere, the lingual gyrus of the right hemisphere, and the banks of the superior temporal sulcus of the right hemisphere. Figure 1 plots the estimated mean growth curves for individual subjects and the overall trend in those identified regions. We see that, for the lingual gyrus of the left hemisphere and the banks of the superior temporal sulcus of the right hemisphere, both the intercept and slope coefficients differ significantly between the groups of subjects with AD and without, suggesting different starting values as well as different decaying rates. Meanwhile, for the lingual gyrus of the right hemisphere, only the slope coefficient differs significantly between the two groups, suggesting a different decaying rate for the AD patients. These identified brain regions also agree with the AD literature well. In particular, the lingual gyrus is located in the occipital lobe, primarily in the visual processing areas of the brain. Its

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



primary functions include visual processing, visual memory, and visual recognition. It is found associated with emotional processing and visual hallucinations under certain neurological conditions. The superior temporal sulcus is located within the temporal lobe. It plays a significant role in a variety of cognitive and perceptual functions, including processing of auditory and speech information, narrative comprehension, social cognition, among others. Yang et al. (2019) found that the cortical thickness of both lingual gyrus regions demonstrate significance between the AD patients and healthy controls. Guo et al. (2020) found that banks of the superior temporal sulcus is the highest beta amyloid affected region, where beta amyloid is one of the most prominent pathological proteins of AD (Jack et al., 2013).

## 6. Discussion

In this article, we have proposed a new set of estimation and inference procedures for the high-dimensional multi-response GCM. It fills an existing gap in the literature and helps address an important family of scientific questions studying the longitudinal growth trend. Meanwhile, the methodological innovations mainly lie in the proposed multi-step estimation approach and the establishment of the proper convergence rates of the covariance estimators. The former enables us to effectively utilize the data information, while the latter allows us to obtain the desired theoretical guarantees for the proposed tests.

There are several potential extensions of the current proposal. Due to our targeting

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



motivation examples, we consider a large number of response variables, but a relatively small number of predictor variables in our setting. It is possible to extend to high-dimensional predictors as well. Moreover, we focus on a linear type model, which we believe is a good starting point. It is also warranted to consider a nonlinear type GCM. Finally, in this article, we focus on the balanced data setting, where each subject has the longitudinal data collected at the same time points. Such a setting is commonly encountered in numerous applications. It also simplifies the methodology development, whereas the resulting solution is already highly nontrivial. Nevertheless, it is worthwhile to study the unbalanced data setting as well. Toward that end, we may impose a Kronecker covariance structure for each individual subject, and modify some regularity conditions. A full investigation of these extensions are beyond the scope of this article, and we leave them as future research.

## Supplementary Material

The Supplementary Material contains all technical proofs of the theoretical results and additional numerical results.

## Acknowledgments

We are grateful to the editor, associate editor, and the two reviewers for their insightful comments and suggestions, which have greatly improved this manuscript. Xia's research was partially supported by NSFC 12331009. Li's research was partially supported by

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



NSF CIF-2102227, NIH R01AG080043, and NIH UG3NS140730.

## References

Alzheimer's Association (2020). 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 16*(3), 391–460.

Bickel, P. J., E. Levina, et al. (2008). Regularized estimation of large covariance matrices. *Ann. Statist. 36*(1), 199–227.

Bradic, J., G. Claeskens, and T. Gueuning (2020). Fixed effects testing in high-dimensional linear mixed models. *J. Am. Statist. Assoc. 115*(532), 1835–1850.

Bryk, A. S. and S. W. Raudenbush (1992). *Hierarchical linear models: Applications and data analysis methods.* Sage Publications, Inc.

Buckner, R. L., D. Head, J. Parker, A. F. Fotenos, D. Marcus, J. C. Morris, and A. Z. Snyder (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage 23*(2), 724–738.

Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc. 108*(501), 265–277.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependency. *J. R. Statist. Soc. B 76*, 349–372.

Carpenter, J., H. Goldstein, and J. Rasbash (1999). A non-parametric bootstrap for multilevel models. *Mul-*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



*tilevel modelling newsletter 11*(1), 2–5.

Chen, X. and W. Liu (2019). Graph estimation for matrix-variate gaussian data. *Stat. Sin. 29*(1), 479–504.

Chen, X., D. Yang, Y. Xu, Y. Xia, D. Wang, and H. Shen (2023). Testing and support recovery of correlation structures for matrix-valued observations with an application to stock market data. *J. Econom. 232*(2), 544–564.

Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage 31*(3), 968–980.

Folstein, M. F., S. E. Folstein, and P. R. McHugh (1975). "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res. 12*(3), 189–198.

Giesbrecht, F. G. and J. C. Burns (1985). Two-stage analysis based on a mixed model: Large-sample asymptotic theory and small-sample simulation results. *Biometrics 41*(2), 477–486.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika 73*(1), 43–56.

Guo, T., S. M. Landau, W. J. Jagust, A. D. N. Initiative, et al. (2020). Detecting earlier stages of amyloid deposition using pet in cognitively normal elderly adults. *Neurology 94*(14), e1512–e1524.

Hox, J. and R. D. Stoel (2005). Multilevel and sem approaches to growth curve modeling. In *Encyclopedia of Statistics in Behavioral Science*, Volume 3, pp. 1296–1305. Citeseer.

Jack, C. R., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, et al. (2013). Tracking pathophysiological processes in alzheimer's disease:

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



an updated hypothetical model of dynamic biomarkers. *Lancet Neurol. 12*(2), 207–216.

Jagust, W. (2018, November). Imaging the evolution and pathophysiology of alzheimer disease. *Nat. Rev. Neurosci. 19*(11), 687–700.

Kackar, R. N. and D. A. Harville (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *COMMUN. STAT-THEOR. M. 10*(13), 1249–1261.

Kenward, M. G. and J. H. Roger (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics 53*(3), 983–997.

Law, M. and Y. Ritov (2023). Inference and estimation for random effects in high-dimensional linear mixed models. *Journal of the American Statistical Association 118*(543), 1682–1691.

Leng, C. and C. Y. Tang (2012). Sparse matrix graphical models. *J. Amer. Statist. Assoc. 107*(499), 1187–1200.

Li, S., T. T. Cai, and H. Li (2022). Inference for high-dimensional linear mixed-effects models: A quasi-likelihood approach. *J. Am. Statist. Assoc. 117*(540), 1835–1846.

Li, Z. (2015). Testing for random effects in growth curve models. *COMMUN. STAT-THEOR. M. 44*(3), 564–572.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist. 41*(6), 2948–2978.

Ma, S., Q. Song, and L. Wang (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli 19*(1), 252 – 274.

Marcus, D. S., A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner (2010, 12). Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *J. Cogn.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



*Neurosci. 22*(12), 2677–2684.

Mechelli, A., K. J. Friston, R. S. Frackowiak, and C. J. Price (2005). Structural covariance in the human cortex. *Journal of Neuroscience 25*(36), 8303–8310.

Newhill, C. E., S. M. Eack, and E. P. Mulvey (2012). A growth curve analysis of emotion dysregulation as a mediator for violence in individuals with and without borderline personality disorder. *J. Pers. Disord. 26*(3), 452–467.

Nocentini, A., E. Menesini, and C. Salmivalli (2013). Level and change of bullying behavior during high school: A multilevel growth curve analysis. *J. Adolesc. 36*(3), 495–505.

Ordaz, S. J., W. Foran, K. Velanova, and B. Luna (2013). Longitudinal growth curves of brain function underlying inhibitory control through adolescence. *J. Neurosci. 33*(46), 18109–18124.

Snijders, T. A. B. and R. J. Bosker (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE Publications.

Wang, L., X. Lyu, and L. Li (2025). High-dimensional response growth curve modeling for longitudinal neuroimaging analysis. *Computational Statistics & Data Analysis 212*, 108239.

Wei, Y. and X. He (2006). Conditional growth charts (with discussion). *Ann. Statist. 34*, 2069–2131.

Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika 102*(2), 247–266.

Xia, Y., T. Cai, and T. T. Cai (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *J. Am. Statist. Assoc. 113*(521), 328–339.

Xia, Y. and L. Li (2017). Hypothesis testing of matrix graph model with application to brain connectivity

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



analysis. *Biometrics 73*(3), 780–791.

Xia, Y. and L. Li (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Stat. Sin. 29*(1), 303–328.

Yan, Z., Z. Zhou, Q. Wu, Z. B. Chen, E. H. Koo, and S. Zhong (2020). Presymptomatic increase of an extracellular rna in blood plasma associates with the development of alzheimer's disease. *Curr. Biol. 30*(10), 1771–1782.e3.

Yang, H., H. Xu, Q. Li, Y. Jin, W. Jiang, J. Wang, Y. Wu, W. Li, C. Yang, X. Li, et al. (2019). Study of brain morphology change in alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *Gen. Psychiatry 32*(2), e100005.

Yin, J. and H. Li (2012). Model selection and estimation in the matrix normal graphical model. *J. Multivar. Anal. 107*, 119 – 140.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res. 11*(Aug), 2261–2286.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc. B 76*(1), 217–242.

Xin Zhou, Division of Biostatistics, University of California at Berkeley, Berkeley, California, U.S.A.

E-mail: xinzhou@berkeley.edu

Yin Xia, Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, P.R. China.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0307



E-mail: xiayin@fudan.edu.cn

Lexin Li, Division of Biostatistics, University of California at Berkeley, Berkeley, California, U.S.A.

E-mail: lexinli@berkeley.edu

Statistica Sinica