



## Statistical Inferences for Complex Dependence of Multimodal Imaging Data

Jinyuan Chang, Jing He, Jian Kang & Mingcong Wu

**To cite this article:** Jinyuan Chang, Jing He, Jian Kang & Mingcong Wu (2024) Statistical Inferences for Complex Dependence of Multimodal Imaging Data, Journal of the American Statistical Association, 119:546, 1486-1499, DOI: [10.1080/01621459.2023.2200610](https://doi.org/10.1080/01621459.2023.2200610)

**To link to this article:** <https://doi.org/10.1080/01621459.2023.2200610>



View supplementary material [↗](#)



Published online: 26 May 2023.



Submit your article to this journal [↗](#)



Article views: 1284



View related articles [↗](#)



View Crossmark data [↗](#)



# Statistical Inferences for Complex Dependence of Multimodal Imaging Data

Jinyuan Chang<sup>a,b</sup> , Jing He<sup>a</sup>, Jian Kang<sup>c</sup> , and Mingcong Wu<sup>a</sup>

<sup>a</sup>Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China; <sup>b</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; <sup>c</sup>Department of Biostatistics, University of Michigan, Ann Arbor, MI

## ABSTRACT

Statistical analysis of multimodal imaging data is a challenging task, since the data involves high-dimensionality, strong spatial correlations and complex data structures. In this article, we propose rigorous statistical testing procedures for making inferences on the complex dependence of multimodal imaging data. Motivated by the analysis of multi-task fMRI data in the Human Connectome Project (HCP) study, we particularly address three hypothesis testing problems: (a) testing independence among imaging modalities over brain regions, (b) testing independence between brain regions within imaging modalities, and (c) testing independence between brain regions across different modalities. Considering a general form for all the three tests, we develop a global testing procedure and a multiple testing procedure controlling the false discovery rate. We study theoretical properties of the proposed tests and develop a computationally efficient distributed algorithm. The proposed methods and theory are general and relevant for many statistical problems of testing independence structure among the components of high-dimensional random vectors with arbitrary dependence structures. We also illustrate our proposed methods via extensive simulations and analysis of five task fMRI contrast maps in the HCP study. Supplementary materials for this article are available online.

## ARTICLE HISTORY

Received June 2021  
Accepted March 2023

## KEYWORDS

FDR control;  
High-dimensional inference;  
Independence test;  
Multimodal neuroimaging

## 1. Introduction

There is an exponential increase in neuroimaging studies during the past decades. Recent advances in different neuroimaging techniques—for example, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Electroencephalography (EEG)—have generated a large amount of multimodal imaging data, which directly or indirectly measure the brain function and structure from different aspects in unprecedented detail. It provides an important tool for researchers to have a better understanding of human brain pathology and etiology of neuropsychiatry diseases (Liu et al. 2015). It is challenging to analyze the multimodal imaging data, which involve high-dimensionality, strong spatial correlations, and complex data structures, as the data type, range and scale can be quite different for different modalities.

In this article, we aim at making statistical inferences on the complex dependence of multimodal imaging data. Our motivation is the analysis of multiple task fMRI data from the Human Connectome Project (Van Essen et al. 2012, HCP). The main goal of HCP is to understand the patterns of structural and functional connectivity in the healthy adult human brain. In addition to collecting structural imaging data and the resting state fMRI data, HCP collects behavioral measures including cognitive, sensory and emotional processes and different types of task fMRI data to understand the relationships between individual differences in the neurobiological substrates of cognitive

processing and functional connectivity (Barch et al. 2013). The fMRI tasks include working memory tasks, language processing tasks, and social cognition tasks. In the analysis of multimodal task fMRI data, we are interested in learning whether the brain activation patterns are independent between different fMRI tasks for each region of interests. For example, it has been shown that inferior frontal gyrus is related to cognitive control (Drobyshevsky, Baumann, and Schneider 2006), language processing (Binder et al. 2011) and social cognition (Castelli et al. 2000), while it remains unknown whether these three types of brain activity are dependent with each other in inferior frontal gyrus. Another interesting topic is to study the brain activation dependence between regions for each type of task fMRI data. It is also of interest to learn the region pair dependence across different modalities (Xia et al. 2020). To develop formal statistical inference procedures to address these questions, we focus on analyzing the five statistical parametric maps (Friston et al. 2011) or contrast maps that are derived from the task fMRI data. However, our proposed methods and theory can be applied for general multimodal imaging data analysis.

Different statistical approaches have been proposed for analysis of the multi-task fMRI data such as the joint independent component analysis (Calhoun et al. 2006) and the joint sparse representation analysis (Ramezani et al. 2014), which focused on estimating the joint brain activation patterns rather than testing. Some statistical testing procedures have been developed for multimodal neuroimaging data. For instance, a voxelwise

testing procedure of multimodal MRI (Naylor et al. 2014) has been proposed for localizing brain activity rather than testing the association between modalities. A simultaneous covariance inferences procedure (Xia et al. 2020) has been developed for multimodal PET analysis. Although this method is general and can be applied for different types of neuroimaging data analysis, it only focuses on region pair independence across two modalities and cannot be directly applied to address all the scientific questions of our interests.

Specifically, in the imaging data,  $J$  types of brain images with different modalities are collected for each subject. Let  $j$  ( $j = 1, \dots, J$ ) index the image type. All the images are registered to the same standard brain template with  $V$  voxels in total. Let  $\mathcal{R} = \{1, \dots, V\}$  denote the set of indices for the voxels in the whole brain template, and the voxels are indexed by  $v \in \mathcal{R}$ . The whole brain  $\mathcal{R}$  is partitioned into  $G$  predefined brain regions, denoted as  $\mathcal{R}_1, \dots, \mathcal{R}_G$ , such that  $\mathcal{R} = \cup_{g=1}^G \mathcal{R}_g$  and  $\mathcal{R}_g \cap \mathcal{R}_{g'} = \emptyset$  for  $g \neq g'$ . Different imaging modalities may focus on measuring different types of functional brain activity, thus, the voxels of interests within the same region is the image type specific. Let  $Y_j(v)$  represent the image signal of type  $j$  at voxel  $v \in \mathcal{R}$ . For each  $g = 1, \dots, G$ , let  $Y_{jg}$  be a vector consisting of  $Y_j(v)$ 's with the voxel  $v$ 's of interests in region  $\mathcal{R}_g$  for image type  $j$ . Suppose that the dataset consists of  $n$  independent subjects. We assume all the images are mutually independent over subjects. In neuroimaging studies, different types of hypothesis testing problems have been considered according to the specific study aims. In particular, it is of great interest to test whether the different imaging modalities are independent within each brain region. The null hypothesis can be formulated as

- (a) For each region  $g = 1, \dots, G$ ,

$$H_{0,g}^{(a)} : Y_{1,g}, \dots, Y_{J,g} \text{ are mutually independent.} \quad (1.1)$$

Another type of problems has drawn much attention in neuroimaging studies is to test the independence in imaging measurements between different regions either within the same modalities or across different modalities. The corresponding null hypotheses are given as follows.

- (b) For each pair  $(g, g')$  with  $g \neq g'$ ,

$$H_{0,g,g'}^{(b)} : Y_{jg} \text{ and } Y_{jg'} \text{ are independent} \quad (1.2)$$

for each modality  $j = 1, \dots, J$ .

- (c) For each pair  $(g, g')$  with  $g, g' = 1, \dots, G$ ,

$$H_{0,g,g'}^{(c)} : Y_{jg} \text{ and } Y_{j'g'} \text{ are independent} \quad (1.3)$$

for all modality pairs  $j \neq j'$ .

As we will mention later, testing for hypotheses (a)–(c) all can be formulated in a unified framework that testing independence structure among the components of a random vector, which is so fundamental that are relevant for many statistical problems. In the application of the complex multimodal imaging data with large sample sizes, the dimension of random vectors can be extremely large.

Testing the independence between two random vectors  $X \in \mathbb{R}^{p_x}$  and  $Y \in \mathbb{R}^{p_y}$  can be viewed as a special case of the testing problems (a) with  $J = 2$ , (b) with  $J = 1$  and (c) with  $J = 2$ . When

$p_x = p_y = 1$ , the classical tests target on testing whether the Pearson correlation coefficient (Pearson 1920), Spearman's rho coefficient (Spearman 1904), Kendall's tau coefficient (Kendall 1938), Blum-Kiefer-Rosenblatt's R (Blum, Kiefer, and Rosenblatt 1961) and the maximal information coefficient (Reshef et al. 2011) equal to zero. When  $p_x, p_y > 1$ , under the normality assumption, Wilks (1935) proposed the test based on the sample correlations and Hotelling (1936) proposed the test based on the canonical correlations. To get rid of the normality assumption, some nonparametric tests based on nonlinear dependence metrics have been proposed, for example, the distance covariance/correlation (Székely, Rizzo, and Bakirov 2007), the Hilbert-Schmidt independence criterion (Gretton et al. 2008, HSIC) and the sign covariance (Bergsma and Dassios 2014). See also Heller, Heller, and Gorfine (2013) and Gieser and Randles (1997) for tests based on the ranks of the pairwise distances between the samples and quadrant statistic, respectively, and Taskinen, Oja, and Randles (2005) for extensions of Kendall's tau and Spearman's rho statistics in multivariate cases. More recently, Shi, Drton, and Han (2022) proposed a distribution-free test combined the distance covariance with the center-outward ranks and signs (Hallin et al. 2021), and Deb and Sen (2023) developed a more general framework for the multivariate rank-based test using optimal transportation. See also Berrett, Kontoyiannis, and Samworth (2021) for the permutation test based on  $U$ -statistic. When  $p_x$  and  $p_y$  diverge with the sample size  $n$ , Gao et al. (2021) considered the test based on the bias-corrected distance correlation, Zhu et al. (2020) proposed tests based on an aggregation of marginal sample distance covariance and HSIC, and Chakraborty and Zhang (2021) suggested the test based on some newly proposed dependence metrics. Zhu et al. (2017) used the squared projection covariance to characterize the dependence between two random vectors with arbitrary dimensions, where such measure equals to zero if and only if the two random vectors are independent.

A more general problem of interest is to test the mutual independence of  $k$  random vectors  $X^{(1)}, \dots, X^{(k)}$ , which is identical to our testing problem (a) with fixed  $g$ . When  $X^{(1)}, \dots, X^{(k)}$  are univariate, the methods in Blum, Kiefer, and Rosenblatt (1961) and Nagao (1973) can be applied for the scenario with fixed  $k$ . For normally distributed  $(X^{(1)}, \dots, X^{(k)})^\top$ , Bai et al. (2009), Ledoit and Wolf (2002) and Schott (2005) considered the cases with  $k/n \rightarrow c$  for some constant  $c > 0$ , where all these methods aim at testing whether the covariance matrix of  $(X^{(1)}, \dots, X^{(k)})^\top$  is an identity matrix. Han, Chen, and Liu (2017) considered two families of distribution-free test statistics, which allows  $k$  to diverge exponentially fast in  $n$ . Leung and Drton (2018) and Yao, Zhang, and Shao (2018) proposed tests based on pairwise rank correlations and pairwise distance covariance, respectively. When  $X^{(1)}, \dots, X^{(k)}$  are multivariate, Pfister et al. (2018) and Jin and Matteson (2018) proposed tests based on  $k$ -variate HSIC and generalized distance covariance, respectively, with fixed  $k$ . See also Chakraborty and Zhang (2019) for the test based on the high order distance covariance.

Both global and multiple tests for hypotheses (a)–(c) are of great interest. However, there is lack of formal statistical inference procedure to address these questions by jointly analyzing the complex multimodal imaging data. Our setting is more complicated than those considered in the aforementioned

methods. For the three global null hypotheses (i)  $H_{0,g}^{(a)}$  holds for all  $g = 1, \dots, G$ , (ii)  $H_{0,g,g'}^{(b)}$  holds for all pair  $(g, g')$  with  $g \neq g'$ , and (iii)  $H_{0,g,g'}^{(c)}$  holds for all pair  $(g, g')$  with  $g, g' = 1, \dots, G$ , for (a)–(c), respectively, the aforementioned methods cannot be applied directly. Although some of these methods can be used to derive the  $p$ -value of each  $H_{0,g}^{(a)}$  in the hypothesis (a), the validity of the associated multiple testing procedure is still unknown. Meanwhile, none of these methods can address the multiple tests for hypotheses (b) and (c). To fill this gap, in this work, we propose rigorous statistical testing procedures.

For a vector  $X$  and an index set  $\mathcal{S}$ , denote by  $X_{\mathcal{S}}$  the subvector of  $X$  that includes all components of  $X$  indexed by  $\mathcal{S}$ . As we will state in Section 2.1, the testing problems (a)–(c) can be all covered by a general form  $H_{0,q} : X_{\mathcal{S}_1}$  is independent of  $X_{\mathcal{S}_2}$  for any  $(\mathcal{S}_1, \mathcal{S}_2) \in \mathcal{K}_q$ , where  $q = 1, \dots, Q$ , and  $\mathcal{K}_q$  is a set consisting of all  $(\mathcal{S}_1, \mathcal{S}_2)$  of interest with  $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$ . For each  $(\mathcal{S}_1, \mathcal{S}_2) \in \mathcal{K}_q$ , we apply the squared projection covariance proposed in Zhu et al. (2017) to characterize the pairwise dependency between  $X_{\mathcal{S}_1}$  and  $X_{\mathcal{S}_2}$ , and propose a computationally efficient method to estimate the squared projection covariance rather than using the  $U$ -statistic type estimate considered in Zhu et al. (2017). For the global testing problem such that  $H_{0,q}$  holds for all  $q = 1, \dots, Q$ , we first propose an  $L$ -statistic type test statistic based on the estimates of the squared projection covariances, and then approximate its null-distribution by the distribution of some functional of a normal random vector, which allows (i) arbitrary dependency among the components of  $X$ , and (ii)  $d = \sum_{q=1}^Q |\mathcal{K}_q|$  diverging exponentially fast in the sample size  $n$ . For the multiple testing problem that identifies which  $H_{0,q}$ 's are not true, we also use an  $L$ -statistic as the marginal test statistic for each  $H_{0,q}$ . Since these marginal test statistics are not pivotal and the dependency among them is complicated, to establish the theoretical guarantee of the associated FDR control procedure is quite challenging. In most FDR control works, the limiting distributions of the marginal test statistics usually have explicit forms (e.g., normal distribution,  $t$  distribution, Chi-square distribution) and thus the  $p$ -values of the marginal hypotheses can be obtained easily. However, in our setting, the limiting distributions of the proposed marginal test statistics do not have explicit forms (or even do not exist) in general. We need to approximate their null-distributions by the Gaussian approximation technique (Chernozhukov, Chetverikov, and Kato 2013, Chang, Chen, and Wu in press) and then obtain the associated  $p$ -values, where the approximation error due to Gaussian approximation plays a key role in our theoretical analysis of the FDR procedure, for which we need to derive the uniform nonasymptotic error bounds for the approximations over  $q = 1, \dots, Q$ . Some recent development of the Gaussian approximation technique can be found in, for example, Deng and Zhang (2020), Chernozhukov et al. (2022), and Lopes (2022).

Our proposed method and theory make several novel contributions. Motivated by neuroimaging applications, we develop a general and unified approach for analyzing the independence structure among the components of an ultra high-dimensional random vector. Our method provides a powerful tool to study complex dependence structure in multimodal imaging data. It can construct the brain dependence networks by aggregating

the local test results across region pairs or modality pairs. To implement the proposed test procedures, we develop a computationally efficient distributed algorithm which enables the scalability of our method to analyze large-scale high-resolution imaging data. To the best of our knowledge, our theoretical analysis is the first attempt in the literature of FDR control that uses Gaussian approximation technique to derive the  $p$ -values of the marginal hypotheses which rely on complicated test statistics, and also establish the theoretical guarantee of the associated FDR control procedure.

The rest of this article is organized as follows. The methodology is given in Section 2 which includes the procedures for the global test and the multiple test with FDR control, and the distributed algorithm for the implementation of our proposed procedures. Section 3 studies the finite sample performance of our proposed methods. The analysis of HCP data is presented in Section 4. Section 5 concludes this article by some brief discussion. All technical proofs are relegated to the supplementary material. At the end of this section, we introduce some notation that is used throughout the article. For any positive integers  $N$  and  $a$ , we write  $[N] := \{1, \dots, N\}$  and  $[N] + a := \{1 + a, \dots, N + a\}$ . The notation  $I(\cdot)$  denotes the indicator function and  $\mathbf{1}_n$  denotes the  $n$ -dimensional vector with all the elements being 1. For a  $v_1 \times v_2$  matrix  $A = (a_{ij})_{v_1 \times v_2}$ , define  $|A|_{\infty} = \max_{i \in [v_1], j \in [v_2]} |a_{ij}|$  and write  $A_{\mathcal{V}_1, \mathcal{V}_2} = (a_{ij})_{i \in \mathcal{V}_1, j \in \mathcal{V}_2}$  with  $\mathcal{V}_1 \subset [v_1]$  and  $\mathcal{V}_2 \subset [v_2]$ . For a  $v \times v$  matrix  $B$ , set  $\text{diag}(B)$  to be the  $v \times v$  matrix only having the main diagonal of  $B$ . For a  $k$ -dimensional vector  $X = (X_1, \dots, X_k)^{\top}$  and an index set  $\mathcal{V} \subset [k]$ , let  $X_{\mathcal{V}}$  be the subvector of  $X$  consisting of components indexed by  $\mathcal{V}$ . For any set  $\mathcal{S}$ , let  $|\mathcal{S}|$  denote its cardinality. For two nonzero vectors  $X, Y \in \mathbb{R}^k$ , we write  $\angle(X, Y) = \arccos\{X^{\top}Y / (|X|_2|Y|_2)\}$  for the angle between them, where  $|\cdot|_2$  denotes the vector  $l_2$ -norm. If  $X = 0$  or  $Y = 0$ , we define  $\angle(X, Y) = 0$ . For two sequences of positive numbers  $\{a_n\}$  and  $\{b_n\}$ , we write  $a_n \lesssim b_n$  or  $b_n \gtrsim a_n$  if  $\limsup_{n \rightarrow \infty} a_n/b_n \leq c$  for some positive constant  $c$ , and  $a_n \ll b_n$  or  $b_n \gg a_n$  if  $\lim_{n \rightarrow \infty} a_n/b_n = 0$ .

## 2. Methodology

### 2.1. Preliminaries

Let  $X \sim F$  be a generic  $p$ -dimensional random vector. The null hypotheses (a)–(c) defined as (1.1)–(1.3) in Section 1 all can be formulated as the following general form:

$$H_{0,q} : X_{\mathcal{S}_1} \text{ is independent of } X_{\mathcal{S}_2} \text{ for any } (\mathcal{S}_1, \mathcal{S}_2) \in \mathcal{K}_q, \quad (2.1)$$

where  $q \in [Q]$  and  $\mathcal{K}_q$  is a set consisting of all  $(\mathcal{S}_1, \mathcal{S}_2)$  of interest with  $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$ . Here we allow that  $\mathcal{K}_q$  may be different for different  $q$ .

Write  $X = (Y_{1,1}^{\top}, \dots, Y_{1,G}^{\top}, \dots, Y_{J,1}^{\top}, \dots, Y_{J,G}^{\top})^{\top}$  with  $Y_{j,g}$  defined above (1.1). The null hypothesis  $H_{0,g}^{(a)}$  of the testing problem (a) given in (1.1) can be reformulated as (2.1) with  $Q = G$ ,  $q = g$  and  $\mathcal{K}_q = \{(\mathcal{S}_{q,j}, \mathcal{S}_{q,-j}) : j \in [J]\}$  such that  $X_{\mathcal{S}_{q,j}} = Y_{j,g}$ ,  $X_{\mathcal{S}_{q,-j}} = (Y_{1,g}^{\top}, \dots, Y_{j-1,g}^{\top}, Y_{j+1,g}^{\top}, \dots, Y_{J,g}^{\top})^{\top}$  and  $|\mathcal{K}_q| = J$ . Write  $\mathcal{G}_1 = \{(g, g') : g, g' \in [G] \text{ with } g < g'\}$  and  $Q = |\mathcal{G}_1|$ . There exists a bijective mapping  $\psi_1 : \mathcal{G}_1 \rightarrow [Q]$ . For given  $(g, g') \in \mathcal{G}_1$ , let  $q = \psi_1(g, g')$  and the associated null hypothesis  $H_{0,g,g'}^{(b)}$  of the testing problem (b) given in (1.2)

can be stated as (2.1) with  $\mathcal{K}_q = \{(\mathcal{S}_{qj,1}, \mathcal{S}_{qj,2}) : j \in [J]\}$  such that  $X_{\mathcal{S}_{qj,1}} = Y_{jg}$ ,  $X_{\mathcal{S}_{qj,2}} = Y_{jg'}$  and  $|\mathcal{K}_q| = J$ . Let  $\mathcal{G}_2 = \{(g, g') : g, g' \in [G] \text{ with } g \leq g'\}$  and  $Q = |\mathcal{G}_2|$ . We can also construct a bijective mapping  $\psi_2 : \mathcal{G}_2 \rightarrow [Q]$ . For given  $(g, g') \in \mathcal{G}_2$ , letting  $q = \psi_2(g, g')$ , the associated null hypothesis  $H_{0,g,g'}^{(c)}$  of the testing problem (c) given in (1.3) can be stated as (2.1) with  $\mathcal{K}_q = \{(\mathcal{S}_{qj,1}, \mathcal{S}_{qj',2}) : j \neq j'\}$  such that  $X_{\mathcal{S}_{qj,1}} = Y_{jg}$ ,  $X_{\mathcal{S}_{qj',2}} = Y_{j',g'}$  and  $|\mathcal{K}_q| = J(J-1)I(g \neq g') + 2^{-1}J(J-1)I(g = g')$ .

To characterize the dependence between two random vectors  $X_{S_1}$  and  $X_{S_2}$ , we consider using the squared projection covariance (Zhu et al. 2017) defined as

$$\begin{aligned} \text{Pcov}^2(X_{S_1}, X_{S_2}) &= \int_{|\alpha|_2=1} \int_{|\beta|_2=1} \int_{(u,v) \in \mathbb{R}^2} \text{cov}^2\{I(\alpha^\top X_{S_1} \leq u), I(\beta^\top X_{S_2} \leq v)\} \\ &\quad \times dF_{\alpha^\top X_{S_1}, \beta^\top X_{S_2}}(u, v) d\alpha d\beta, \end{aligned} \quad (2.2)$$

where  $F_{\alpha^\top X_{S_1}, \beta^\top X_{S_2}}(\cdot, \cdot)$  is the joint distribution function of  $(\alpha^\top X_{S_1}, \beta^\top X_{S_2})$ . Such defined  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  has two advantages: (i)  $\text{Pcov}^2(X_{S_1}, X_{S_2}) \geq 0$  and  $\text{Pcov}^2(X_{S_1}, X_{S_2}) = 0$  if and only if  $X_{S_1}$  is independent of  $X_{S_2}$ , and (ii)  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  does not require  $X_{S_1}$  and  $X_{S_2}$  to share the same dimensionality and allows arbitrary dimensions of  $X_{S_1}$  and  $X_{S_2}$ . The second advantage is useful in high-dimensional problems where the dimensions of  $X_{S_1}$  and  $X_{S_2}$  can be quite large in practice. As shown in Zhu et al. (2017),  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  defined as (2.2) has the following explicit form:

$$\begin{aligned} \text{Pcov}^2(X_{S_1}, X_{S_2}) &= \mathbb{E}\{\angle(X_{S_1}^{(1)} - X_{S_1}^{(3)}, X_{S_1}^{(4)} - X_{S_1}^{(3)}) \cdot \angle(X_{S_2}^{(1)} - X_{S_2}^{(3)}, X_{S_2}^{(4)} - X_{S_2}^{(3)})\} \\ &\quad + \mathbb{E}\{\angle(X_{S_1}^{(1)} - X_{S_1}^{(3)}, X_{S_1}^{(4)} - X_{S_1}^{(3)}) \cdot \angle(X_{S_2}^{(2)} - X_{S_2}^{(3)}, X_{S_2}^{(5)} - X_{S_2}^{(3)})\} \\ &\quad - 2\mathbb{E}\{\angle(X_{S_1}^{(1)} - X_{S_1}^{(3)}, X_{S_1}^{(4)} - X_{S_1}^{(3)}) \cdot \angle(X_{S_2}^{(2)} - X_{S_2}^{(3)}, X_{S_2}^{(4)} - X_{S_2}^{(3)})\}, \end{aligned} \quad (2.3)$$

where  $(X_{S_1}^{(1)}, X_{S_2}^{(1)}), \dots, (X_{S_1}^{(5)}, X_{S_2}^{(5)})$  are five independent copies of  $(X_{S_1}, X_{S_2})$ . Based on (2.3), for given independent and identically distributed observations  $\{X_1, \dots, X_n\} \sim F$ , Zhu et al. (2017) suggest to estimate  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  by

$$\begin{aligned} \widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2}) &= \frac{1}{n^3} \sum_{i,k,l=1}^n \angle(X_{i,S_1} - X_{k,S_1}, X_{l,S_1} - X_{k,S_1}) \\ &\quad \cdot \angle(X_{i,S_2} - X_{k,S_2}, X_{l,S_2} - X_{k,S_2}) \\ &\quad + \frac{1}{n^5} \sum_{i,j,k,l,r=1}^n \angle(X_{i,S_1} - X_{k,S_1}, X_{l,S_1} - X_{k,S_1}) \\ &\quad \cdot \angle(X_{j,S_2} - X_{k,S_2}, X_{r,S_2} - X_{k,S_2}) \\ &\quad - \frac{2}{n^4} \sum_{i,j,k,l=1}^n \angle(X_{i,S_1} - X_{k,S_1}, X_{l,S_1} - X_{k,S_1}) \\ &\quad \cdot \angle(X_{j,S_2} - X_{k,S_2}, X_{l,S_2} - X_{k,S_2}). \end{aligned} \quad (2.4)$$

When the dimensions of  $X_{S_1}$  and  $X_{S_2}$  are quite large, to calculate the  $U$ -statistic type estimate (2.4) requires high computing cost and large storage space even with moderately large

sample size  $n$ , since all observations need to be read from files at once and saved as a very large matrix. In our motivating HCP task fMRI data, the sample size  $n = 922$  and the number of voxels in the standard 2mm MNI brain template  $V = 185,405$ . In this dataset, we have five different fMRI contrast maps (2back-0back, task-rest, story-math, mental, random) and 116 predefined brain regions from the automated anatomical labeling (AAL) (Tzourio-Mazoyer et al. 2002). If we are interested in the testing problem (a) given in (1.1) for the first 90 brain regions and 2 modalities “story-math” and “mental”, we can formulate this problem as (2.1) with  $Q = 90$  and  $|\mathcal{K}_q| = 1$  for each  $q = 1, \dots, 90$ . On a server with two Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, the computation time of the associated  $\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})$  given in (2.4) for  $H_{0,q}$  using all 922 observations ranges from 185.65 sec to 1768.08 sec for  $q = 1, \dots, 90$ .

To improve the computational efficiency, we consider to divide the  $n$  observations  $\{X_1, \dots, X_n\}$  into several moving subgroups and compute the associated squared projection covariance within these subgroups. Let  $B \geq 5$  denote the length of each subgroup, and write  $M = n - B + 1$ . For any  $m \in [M]$ , let  $\mathcal{B}_m = \{X_m, \dots, X_{m+B-1}\}$  be the  $m$ th moving subgroup. For given  $q \in [Q]$  and  $(S_1, S_2) \in \mathcal{K}_q$ , we estimate  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  based on  $\mathcal{B}_m$  in the same manner as (2.4) but with replacing  $\{X_1, \dots, X_n\}$  by  $\mathcal{B}_m$ , and then denote the associated estimate by  $U_{m,q,S_1,S_2}$ . Gathering the estimates in all moving subgroups together, we finally estimate  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  by

$$\bar{U}_{q,S_1,S_2} = \frac{1}{M} \sum_{m=1}^M U_{m,q,S_1,S_2} \quad (2.5)$$

for any  $q \in [Q]$  and  $(S_1, S_2) \in \mathcal{K}_q$ . By doing so, we just need to read  $B$  observations of  $\mathcal{B}_m$  at one time to calculate  $U_{m,q,S_1,S_2}$ , which can largely reduce the computing cost. The computation time of such method with  $B = 5$  using all 922 observations ranges between 2.72 sec and 33.30 sec for  $q = 1, \dots, 90$ , which economizes 98% computing cost in comparison to that of Zhu et al. (2017). As we will discuss in Section 2.4 later, our method is convenient to develop a distributed algorithm for (2.5) to further enhance the computational efficiency. This is another advantage of our proposed method. Our theoretical analysis allows both fixed  $B$  and diverging  $B$ . Extensive numerical studies indicate that it works quite well even when we select  $B = 5$ .

Our proposed estimate  $\bar{U}_{q,S_1,S_2}$  in (2.5) also has another convenience in solving the hypothesis testing problem (2.1). For the  $U$ -statistic type estimate  $\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})$  given in (2.4), Zhu et al. (2017) shows that under null hypothesis  $n\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})$  converges in distribution to the weighted infinite sum of independent Chi-square distributions. If we construct the test statistic for the null hypothesis  $H_{0,q}$  specified in (2.1) based on  $\{\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})\}_{(S_1,S_2) \in \mathcal{K}_q}$ , we will encounter two obstacles. First, if  $|\mathcal{K}_q| > 1$ , how to characterize the dependency among different  $\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})$ 's is extremely difficult (if not impossible) since their limiting distributions are too complicated. Second, to do inference based on  $\widehat{\text{Pcov}^2}(X_{S_1}, X_{S_2})$  usually needs to involve the random permutation procedure which requires very high computing cost. Notice that  $\bar{U}_{q,S_1,S_2}$  is consistent to  $\text{Pcov}^2(X_{S_1}, X_{S_2})$  for any  $q \in [Q]$  and

$(S_1, S_2) \in \mathcal{K}_q$ . Then  $\max_{(S_1, S_2) \in \mathcal{K}_q} \sqrt{M} \bar{U}_{q, S_1, S_2} / \hat{\sigma}_{q, S_1, S_2}$  provides a natural test statistic for the null hypothesis  $H_{0,q}$  specified in (2.1), where  $\hat{\sigma}_{q, S_1, S_2}$  is the estimation of the standard deviation of  $\sqrt{M} \bar{U}_{q, S_1, S_2}$ . Since  $\sqrt{M} \bar{U}_{q, S_1, S_2} / \hat{\sigma}_{q, S_1, S_2} \rightarrow_d \mathcal{N}(0, 1)$  for any  $(S_1, S_2) \in \mathcal{K}_q$ , using the Gaussian approximation technique, we know the null-distribution of  $\max_{(S_1, S_2) \in \mathcal{K}_q} \sqrt{M} \bar{U}_{q, S_1, S_2} / \hat{\sigma}_{q, S_1, S_2}$  can be always approximated by the distribution of the largest component in a  $|\mathcal{K}_q|$ -dimensional normal distributed random vector with mean zero. Hence, the inference of the test statistic based on  $\{\bar{U}_{q, S_1, S_2}\}_{(S_1, S_2) \in \mathcal{K}_q}$  can be implemented easily and efficiently. See Sections 2.2–2.4 for details.

## 2.2. A Global Testing Procedure

In this part, we first consider how to test the following global null hypothesis:

$$H_0 : H_{0,q} \text{ given in (2.1) holds for all } q \in [Q]. \quad (2.6)$$

Due to  $\mathbb{E}(\bar{U}_{q, S_1, S_2}) = \text{Pcov}^2(X_{S_1}, X_{S_2})$ , large values of  $\bar{U}_{q, S_1, S_2}$  provide evidence against the marginal null hypothesis that  $X_{S_1}$  is independent of  $X_{S_2}$  for given  $(S_1, S_2) \in \mathcal{K}_q$ . Denote by  $\sigma_{q, S_1, S_2}$  the standard deviation of  $\sqrt{M} \bar{U}_{q, S_1, S_2}$  and let  $\hat{\sigma}_{q, S_1, S_2}$  be its consistent estimate which will be given below (2.10). Then we define the studentized statistic

$$T_{q, S_1, S_2} = \frac{\sqrt{M} \bar{U}_{q, S_1, S_2}}{\hat{\sigma}_{q, S_1, S_2}} \quad (2.7)$$

for any  $q \in [Q]$  and  $(S_1, S_2) \in \mathcal{K}_q$ . To test the global null hypothesis (2.6), we use the information of the largest  $L$  values among  $\{T_{q, S_1, S_2}\}_{q \in [Q], (S_1, S_2) \in \mathcal{K}_q}$  to calculate the global test statistic. More specifically, for each  $q \in [Q]$ , we first stack  $\{T_{q, S_1, S_2}\}_{(S_1, S_2) \in \mathcal{K}_q}$  into a  $|\mathcal{K}_q|$ -dimensional vector  $T_n^{(q)} = (T_{n,1}^{(q)}, \dots, T_{n,|\mathcal{K}_q|}^{(q)})^\top$ . We then obtain a  $d$ -dimensional vector  $T_n = (T_{n,1}, \dots, T_{n,d})^\top = (T_n^{(1)}, \dots, T_n^{(Q)})^\top$  with  $d = \sum_{q=1}^Q |\mathcal{K}_q|$ . Given a pre-determined integer  $L$ , the global test statistic for (2.6) is given by

$$W_{n,L} = \max_{1 \leq j_1 < \dots < j_L \leq d} \sum_{s=1}^L T_{n,j_s}, \quad (2.8)$$

which is an  $L$ -statistic. Such kind of statistic that combines several largest signals together has been also used in Zhang (2015) and Fan, Shao, and Zhou (2018) for solving other problems. Based on  $W_{n,L}$ , we formally reject the global null hypothesis  $H_0$  in (2.6) at the significance level  $\alpha \in (0, 1)$  if

$$W_{n,L} > \text{cv}_{L,\alpha}, \quad (2.9)$$

where  $\text{cv}_{L,\alpha}$  is the critical value satisfying  $\mathbb{P}_{H_0}(W_{n,L} > \text{cv}_{L,\alpha}) \rightarrow \alpha$  as  $n \rightarrow \infty$ .

To determine  $\text{cv}_{L,\alpha}$ , we need to investigate the null-distribution of  $W_{n,L}$ . When  $L = 1$ ,  $W_{n,1} = \max_{j \in [d]} T_{n,j}$  is a maximum-type test statistic. Due to each  $T_{n,j} \rightarrow_d \mathcal{N}(0, 1)$  under  $H_0$ , we may consider to approximate the null-distribution of  $W_{n,1}$  by some extreme-value distribution. However, such approximation relies heavily on the structural assumptions on

the asymptotic covariance matrix of  $T_n = (T_{n,1}, \dots, T_{n,d})^\top$  which are quite restrictive and difficult to be verified in practice. On the other hand, even such approximation works, its convergence rate is usually slow which will usually cause size distortion. Taking the extreme-value distribution of type I as an example, the convergence rate is just of order  $O(\log(\log n)/\log n)$ . In comparison to  $W_{n,1}$ ,  $W_{n,L}$  with  $L > 1$  gathers more information which will lead to power-enhancement under alternatives.

To the best of our knowledge, the null-distribution of  $W_{n,L}$  with  $L > 1$  and  $d \gg n$  has been rarely studied in the literature. Note that  $\bar{U}_{q, S_1, S_2}$  in (2.7) is the average of the dependent sequence  $\{U_{m,q, S_1, S_2}\}_{m=1}^M$ . Deriving the null-distribution of  $W_{n,L}$  is challenging. In this article, we propose a novel procedure to approximate the null-distribution of  $W_{n,L}$  with general  $L \geq 1$  that can diverge with  $n$ . Our procedure does not need to impose any structural assumption on the asymptotic covariance matrix of  $T_n$  and allows arbitrary dependency among the components of  $T_n$ .

For each  $m \in [M]$ , we stack  $\{U_{m,q, S_1, S_2}\}_{q \in [Q], (S_1, S_2) \in \mathcal{K}_q}$  into a  $d$ -dimensional vector, denoted by  $U_m = (U_{m,1}, \dots, U_{m,d})^\top$ , with the same order as  $T_n = (T_{n,1}, \dots, T_{n,d})^\top$ . Write  $\Sigma_n = \text{cov}(\sqrt{M} \bar{U})$  with  $\bar{U} = M^{-1} \sum_{m=1}^M U_m$ . Then  $\{\sigma_{q, S_1, S_2}^2\}_{q \in [Q], (S_1, S_2) \in \mathcal{K}_q}$  are the  $d$  elements in the main diagonal of  $\Sigma_n$ . Let  $R_n = D_n^{-1/2} \Sigma_n D_n^{-1/2}$  with  $D_n = \text{diag}(\Sigma_n)$ . For a  $d$ -dimensional vector  $Z = (Z_1, \dots, Z_d)^\top$ , we sort its components in descending order as  $Z_{(1)} \geq \dots \geq Z_{(d)}$  and define a function

$$f_L(Z) = \sum_{j=1}^L Z_{(j)}.$$

Then  $W_{n,L} = f_L(T_n)$ . Proposition 1(i) shows that the critical value  $\text{cv}_{L,\alpha}$  in (2.9) can be selected as the  $(1 - \alpha)$ -quantile of the distribution of  $f_L(\xi)$ , where  $\xi \sim \mathcal{N}(0, R_n)$ .

In practice, we need to estimate  $R_n$ . Since  $\{U_m\}_{m=1}^M$  is a  $d$ -dimensional  $(B - 1)$ -dependent process, then  $\Sigma_n = M^{-1} \sum_{|m_1 - m_2| < B} \text{cov}(U_{m_1}, U_{m_2})$ , which indicates that

$$\hat{\Sigma}_n = \sum_{j=-B+1}^{B-1} \hat{H}_j, \quad (2.10)$$

with  $\hat{H}_j = M^{-1} \sum_{m=j+1}^M (U_m - \bar{U})(U_{m-j} - \bar{U})^\top$  for  $j \geq 0$  and  $\hat{H}_j = M^{-1} \sum_{m=-j+1}^M (U_{m+j} - \bar{U})(U_m - \bar{U})^\top$  for  $j < 0$ , provides an estimate of  $\Sigma_n$ . Let  $\hat{D}_n = \text{diag}(\hat{\Sigma}_n)$  which gives the estimates of  $\{\sigma_{q, S_1, S_2}^2\}_{q \in [Q], (S_1, S_2) \in \mathcal{K}_q}$ . Write  $\hat{R}_n = \hat{D}_n^{-1/2} \hat{\Sigma}_n \hat{D}_n^{-1/2}$  and  $\hat{F}(t) = \mathbb{P}\{f_L(\hat{\xi}) \leq t | X_1, \dots, X_n\}$  with  $\hat{\xi} | X_1, \dots, X_n \sim \mathcal{N}(0, \hat{R}_n)$ . Proposition 1(ii) shows that  $\text{cv}_{L,\alpha}$  can be approximated by

$$\hat{\text{cv}}_{L,\alpha} = \inf \{t \in \mathbb{R} : \hat{F}(t) \geq 1 - \alpha\},$$

the  $(1 - \alpha)$ -quantile of the conditional distribution of  $f_L(\hat{\xi})$  given  $\{X_1, \dots, X_n\}$ . Practically, we can draw  $d$ -dimensional random vectors  $\hat{\xi}_1, \dots, \hat{\xi}_N$  independently from  $\mathcal{N}(0, \hat{R}_n)$  for some large integer  $N$ , and use the  $\lfloor N\alpha \rfloor$ th largest value of  $f_L(\hat{\xi}_1), \dots, f_L(\hat{\xi}_N)$  to approximate  $\hat{\text{cv}}_{L,\alpha}$ . Such proposed test has several advantages: (i) it can handle the cases with  $d \gg n$ ; (ii) it allows arbitrary dependence structure of  $\Sigma_n$ ; and (iii) its implementation is

quite simple which only needs to generate several  $d$ -dimensional normal random vectors. Write  $\mathcal{F} = \{v = (v_1, \dots, v_d)^\top : v_j \in \{0, 1\} \text{ for each } j \in [d], \text{ and } |v|_0 = L\}$  and  $D_n = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ .

**Proposition 1.** Assume there exists a universal constant  $c_1 > 0$  such that  $\min_{k \in [d]} \sigma_k^2 \geq c_1$  and  $\min_{v \in \mathcal{F}} v^\top R_n v \geq c_1$ . Let  $\xi \sim \mathcal{N}(0, R_n)$  and  $\hat{\xi} \sim \mathcal{N}(0, \hat{R}_n)$ . Under the global null hypothesis  $H_0$  given in (2.6), we have that (i)  $\sup_{z \in \mathbb{R}} |\mathbb{P}(W_{n,L} > z) - \mathbb{P}\{f_L(\xi) > z\}| = o(1)$  provided that  $B^4 L^{13} (\log d)^7 = o(n)$  and  $B^4 L^{-2} (\log B)^3 (\log d)^{-2} = o(n)$ , and (ii)  $\sup_{z \in \mathbb{R}} |\mathbb{P}(W_{n,L} > z) - \mathbb{P}\{f_L(\hat{\xi}) > z | X_1, \dots, X_n\}| = o_p(1)$  provided that  $B^4 L^{13} (\log d)^7 = o(n)$  and  $B^5 L^8 (\log d)^4 \log(dB) = o(n)$ .

The conditions  $\min_{k \in [d]} \sigma_k^2 \geq c_1$  and  $\min_{v \in \mathcal{F}} v^\top R_n v \geq c_1$  require that the long-run variances of the sequences  $\{U_{m,k}\}_{m=1}^M$  and  $\{v^\top U_m\}_{m=1}^M$  are uniformly bounded away from zero over  $k \in [d]$  and  $v \in \mathcal{F}$ , which are mild in practice and will be used when we apply Nazarov's inequality to bound the probability of a normal vector taking values in a small region. When we select  $B$ , the length of each subgroup, as a fixed constant, Proposition 1 holds if  $L^{13} (\log d)^7 = o(n)$ , which requires the pre-determined integer  $L$  involved in the  $L$ -statistic  $W_{n,L}$  defined as (2.8) should satisfy  $L = o(n^{1/13})$ . Furthermore, if  $L$  is also selected as a fixed constant, Proposition 1 holds provided that  $\log d = o(n^{1/7})$ , which allows  $d = \sum_{q=1}^Q |\mathcal{K}_q|$  growing exponentially with the sample size  $n$ . Given Proposition 1(i), in order to prove Proposition 1(ii), it suffices to show  $\delta_n := \sup_{z \in \mathbb{R}} |\mathbb{P}\{f_L(\xi) > z\} - \mathbb{P}\{f_L(\hat{\xi}) > z | X_1, \dots, X_n\}| = o_p(1)$ . As shown in (S.15) of the supplementary material,  $\delta_n \lesssim \Delta_n^{1/3} \{1 \vee \log(d^L / \Delta_n)\}^{2/3}$  with  $\Delta_n = \max_{v_1, v_2 \in \mathcal{F}} |v_1^\top (R_n - \hat{R}_n) v_2|$ . Due to  $\Delta_n \leq L^2 \|\hat{R}_n - R_n\|_\infty$ , then  $\delta_n = o_p(1)$  provided that  $\|\hat{R}_n - R_n\|_\infty = o_p[L^{-4} (\log d)^{-2}]$ . For our suggested  $\hat{R}_n$  specified below (2.10), such requirement can be easily satisfied for any  $d$ -dimensional correlation matrix  $R_n$  as long as  $B^5 L^8 (\log d)^4 \log(dB) = o(n)$  without imposing any structural assumption on  $R_n$ . This advantage of Gaussian approximation has also been found in, for example, Chernozhukov, Chetverikov, and Kato (2013), Chang et al. (2017), and Chang, Chen, and Wu (in press). The next theorems state the theoretical guarantee of the proposed global test.

**Theorem 1.** Under the conditions of Proposition 1, if  $B^4 L^{13} (\log d)^7 = o(n)$  and  $B^5 L^8 (\log d)^4 \log(dB) = o(n)$ , then  $\mathbb{P}_{H_0}(W_{n,L} > \hat{c}_{v_{L,\alpha}}) \rightarrow \alpha$  as  $n \rightarrow \infty$ .

Theorem 1 shows that the proposed global test has correct size control. The restrictions on  $(B, L, d, n)$  for Theorem 1 are identical to those for Proposition 1(ii). If we select  $B$  and  $L$  as fixed constants, the proposed global test is valid provided that  $\log d = o(n^{1/7})$ . Write  $\mathbb{E}(U_m) = (\mu_1, \dots, \mu_d)^\top$ . Theorem 2 indicates that the proposed global test is a consistent test under certain local alternatives.

**Theorem 2.** Let  $\rho = \max_{v \in \mathcal{F}} v^\top R_n v$  and  $\lambda(|\mathcal{F}|, \alpha) = (2 \log |\mathcal{F}|)^{1/2} + \{2 \log(1/\alpha)\}^{1/2}$ . Under the conditions of Proposition 1, if  $\max_{1 \leq j_1 < \dots < j_L \leq d} \sum_{l=1}^L \sigma_{j_l}^{-1} \mu_{j_l} \geq \rho^{1/2} (1 + \epsilon_n) n^{-1/2} \lambda(|\mathcal{F}|, \alpha)$  under the alternative hypothesis  $H_1$  for some  $\epsilon_n > 0$  satisfying  $(\log |\mathcal{F}|)^{-1/2} \ll \epsilon_n \ll 1$ , then  $\mathbb{P}_{H_1}(W_{n,L} >$

$\hat{c}_{v_{L,\alpha}}) \rightarrow 1$  as  $n \rightarrow \infty$ , provided that  $B^5 L^4 \log(dB) (\log |\mathcal{F}|)^2 = o(n)$  and  $B^3 L^2 \{\log(dB)\}^2 \log |\mathcal{F}| = o(n)$ .

Recall  $d = \sum_{q=1}^Q |\mathcal{K}_q|$  and  $U_{m,1}, \dots, U_{m,d}$ , the  $d$  components of  $U_m$ , are a permutation of the  $d$  elements in the set  $\{U_{m,q,S_1,S_2} : q \in [Q], (S_1, S_2) \in \mathcal{K}_q\}$ . For any  $j \in [d]$ , there exists  $q' \in [Q]$  and  $(S'_1, S'_2) \in \mathcal{K}_{q'}$  such that  $\mu_j = \mathbb{E}(U_{m,q',S'_1,S'_2}) = \text{Pcov}^2(X_{S'_1}, X_{S'_2})$ . Hence, the global null hypothesis  $H_0$  in (2.6) is equivalent to  $\mathbb{E}(U_m) = (\mu_1, \dots, \mu_d)^\top = 0$ , and the test statistic  $W_{n,L}$  given in (2.8) targets on testing whether such equivalence holds or not based on the obtained data  $U_1, \dots, U_M$ , which essentially formulates the original hypothesis testing problem as a  $d$ -dimensional mean vector testing problem. The condition  $\max_{1 \leq j_1 < \dots < j_L \leq d} \sum_{l=1}^L \sigma_{j_l}^{-1} \mu_{j_l} \geq \rho^{1/2} (1 + \epsilon_n) n^{-1/2} \lambda(|\mathcal{F}|, \alpha)$  states the restriction on mean vector  $(\mu_1, \dots, \mu_d)^\top$  under which the proposed global test has power approaching one. Recall  $\mathcal{F} = \{v = (v_1, \dots, v_d)^\top : v_j \in \{0, 1\} \text{ for each } j \in [d], \text{ and } |v|_0 = L\}$ . For fixed  $L$ , the lower bound  $\rho^{1/2} (1 + \epsilon_n) n^{-1/2} \lambda(|\mathcal{F}|, \alpha)$  has order  $n^{-1/2} (\log d)^{1/2}$ , which is the minimax optimal separation rate of any tests for  $d$ -dimensional mean vector (Cai, Liu, and Xia 2014). In this case, our test shares the minimax optimal property. Due to  $|\mathcal{F}| = \binom{d}{L} \leq d^L$ ,  $B^5 L^6 \log(dB) (\log d)^2 = o(n)$  provides a sufficient condition for the restrictions on  $(B, L, d, n)$  for Theorem 2. If we select  $B$ , the length of each subgroup, as a fixed constant, Theorem 2 holds if  $L^6 (\log d)^3 = o(n)$ , which requires the pre-determined integer  $L$  involved in the  $L$ -statistic  $W_{n,L}$  should satisfy  $L = o(n^{1/6})$ . Furthermore, if  $L$  is also selected as a fixed constant, Theorem 2 holds provided that  $\log d = o(n^{1/3})$ , which allows  $d$  growing exponentially with the sample size  $n$ . Together with the discussion below Theorem 1, we conclude that when  $\log d = o(n^{1/7})$  the proposed global test with fixed  $B$  and  $L$  can control Type I error at the prescribed significance level and also have power approaching one under certain local alternatives.

### 2.3. A Multiple Testing Procedure with FDR Control

Section 2.2 considers the global test whether all  $H_{0,1}, \dots, H_{0,Q}$  given in (2.1) hold simultaneously or not. When the global null hypothesis  $H_0$  is rejected, we are interested in identifying which  $H_{0,q}$ 's are not true. Note that  $T_{q,S_1,S_2}$  defined in (2.7) provides a valid test statistic for the marginal null hypothesis  $H_{0,q,S_1,S_2} : X_{S_1}$  is independent of  $X_{S_2}$  for given  $(S_1, S_2) \in \mathcal{K}_q$ . For  $T_n^{(q)} = (T_{n,1}^{(q)}, \dots, T_{n,|\mathcal{K}_q|}^{(q)})^\top$  defined below (2.7), we can formulate it as follows:

$$T_n^{(q)} = \sqrt{M} \{\hat{D}_n^{(q)}\}^{-1/2} \bar{U}^{(q)},$$

where  $\bar{U}^{(q)} = M^{-1} \sum_{m=1}^M U_m^{(q)}$ ,  $U_m^{(q)}$ , and  $\hat{D}_n^{(q)}$  are, respectively, the subvector of  $U_m$  and the submatrix of  $\hat{D}_n$  including the associated components related to  $\{T_{q,S_1,S_2}\}_{(S_1,S_2) \in \mathcal{K}_q}$ . Let  $\Sigma_n^{(q)} = \text{cov}(\sqrt{M} \bar{U}^{(q)})$  and define its estimate  $\hat{\Sigma}_n^{(q)}$  in the same manner as  $\hat{\Sigma}_n$  given in (2.10) but with replacing  $\{U_m\}_{m=1}^M$  by  $\{U_m^{(q)}\}_{m=1}^M$ . The test statistic for  $H_{0,q}$  can be selected as

$$W_{n,L}^{(q)} = \max_{1 \leq j_1 < \dots < j_L \leq |\mathcal{K}_q|} \sum_{s=1}^L T_{n,j_s}^{(q)}. \quad (2.11)$$

If  $|\mathcal{K}_q| = 1$ , we can only select  $L = 1$ . Then  $W_{n,L}^{(q)}$  can be also formulated as  $W_{n,L}^{(q)} = f_L(T_n^{(q)})$  for  $f_L(\cdot)$  defined in Section 2.2.

We reject  $H_{0,q}$  when  $W_{n,L}^{(q)}$  takes some large values.

Let  $\mathcal{H}_0$  be the set of all true null hypotheses  $H_{0,q}$  with  $Q_0 = |\mathcal{H}_0|$ . For each  $q \in [Q]$ , generate  $\hat{\xi}_1^{(q)}, \dots, \hat{\xi}_N^{(q)}$  independently from  $\mathcal{N}(0, \hat{R}_n^{(q)})$ , where  $\hat{R}_n^{(q)} = \{\hat{D}_n^{(q)}\}^{-1/2} \hat{\Sigma}_n^{(q)} \{\hat{D}_n^{(q)}\}^{-1/2}$ . Identical to Proposition 1, we can also show that

$$\max_{q \in \mathcal{H}_0} \sup_{z \in \mathbb{R}} |\mathbb{P}\{W_{n,L}^{(q)} > z\} - \mathbb{P}\{f_L(\hat{\xi}^{(q)}) > z \mid X_1, \dots, X_n\}| = o_p(1). \quad (2.12)$$

Then the  $p$ -value of the  $q$ th hypothesis test  $H_{0,q}$  is given by  $\text{pv}_L^{(q)} = \mathbb{P}\{f_L(\hat{\xi}^{(q)}) \geq W_{n,L}^{(q)} \mid X_1, \dots, X_n\}$  with  $\hat{\xi}^{(q)} \sim \mathcal{N}(0, \hat{R}_n^{(q)})$ , and its normal quantile transformation has the form  $V_{n,L}^{(q)} = \Phi^{-1}(1 - \text{pv}_L^{(q)})$ . For implementation,  $\text{pv}_L^{(q)}$  can be approximated by  $\hat{\text{pv}}_L^{(q)} = N^{-1} \sum_{j=1}^N I\{f_L(\hat{\xi}_j^{(q)}) \geq W_{n,L}^{(q)}\}$  for some large integer  $N$ . For any  $t \in \mathbb{R}$ , the *False Discovery Proportion* (FDP) is defined as

$$\text{FDP}(t) = \frac{\sum_{q \in \mathcal{H}_0} I\{V_{n,L}^{(q)} \geq t\}}{1 \vee \sum_{q=1}^Q I\{V_{n,L}^{(q)} \geq t\}}$$

and the *False Discovery Rate* (FDR) is given by  $\text{FDR}(t) = \mathbb{E}\{\text{FDP}(t)\}$ .

Given a prescribed level  $\alpha \in (0, 1)$ , the key object here is to find the smallest  $\hat{t}$  such that  $\text{FDR}(\hat{t}) \leq \alpha$ . To do this, we first consider  $\text{FDP}(t)$ . Since the true null hypotheses set  $\mathcal{H}_0$  is unknown, we need to estimate the numerator of  $\text{FDP}(t)$ . Based on (2.12), we have  $\mathbb{P}\{V_{n,L}^{(q)} \geq t\} = 1 - \Phi(t) + o(1)$  for any  $q \in \mathcal{H}_0$ . An ideal estimate for  $\text{FDP}(t)$  can be selected as

$$\widetilde{\text{FDP}}(t) = \frac{Q_0\{1 - \Phi(t)\}}{1 \vee \sum_{q=1}^Q I\{V_{n,L}^{(q)} \geq t\}}.$$

Since  $Q_0$  is unknown,  $\widetilde{\text{FDP}}(t)$  is infeasible in practice and we can only estimate  $\text{FDP}(t)$  via a more conservative way:

$$\widehat{\text{FDP}}(t) = \frac{Q\{1 - \Phi(t)\}}{1 \vee \sum_{q=1}^Q I\{V_{n,L}^{(q)} \geq t\}}. \quad (2.13)$$

For given  $\alpha \in (0, 1)$ , we choose

$$\hat{t} = \inf \{0 < t \leq (2 \log Q - 2 \log \log Q)^{1/2} : \widehat{\text{FDP}}(t) \leq \alpha\}. \quad (2.14)$$

If  $\hat{t}$  defined in (2.14) does not exist, let  $\hat{t} = (2 \log Q)^{1/2}$ . We then reject all  $H_{0,q}$ 's with  $V_{n,L}^{(q)} \geq \hat{t}$ .

The dependency among the test statistics  $\{W_{n,L}^{(q)}\}_{q \in [Q]}$  plays a key role in analyzing the theoretical property of the proposed multiple testing procedure. Since the limiting distribution of  $W_{n,L}^{(q)}$  is not pivotal and does not admit explicit form, how to characterize the dependency among  $\{W_{n,L}^{(q)}\}_{q \in [Q]}$  is nontrivial. Denote by  $F_q(\cdot)$  the distribution function of  $W_{n,L}^{(q)}$ . Note that  $\eta^{(q)} = \Phi^{-1}\{F_q(W_{n,L}^{(q)})\}$  follows the standard normal distribution. We can measure the dependency between  $W_{n,L}^{(q)}$  and  $W_{n,L}^{(q')}$

via the correlation between their transformations  $\eta^{(q)}$  and  $\eta^{(q')}$ , where the latter one is essential in our theoretical analysis. It is obvious that the independence between  $W_{n,L}^{(q)}$  and  $W_{n,L}^{(q')}$  is equivalent to  $\text{corr}\{\eta^{(q)}, \eta^{(q')}\} = 0$ . Define the set

$$\mathcal{A}_q(\gamma) = \{q' \in [Q] : q' \neq q, |\text{corr}\{\eta^{(q)}, \eta^{(q')}\}| \geq (\log Q)^{-2-\gamma}\}$$

for  $\gamma > 0$  and  $q \in [Q]$ . For the  $q$ th test statistic  $W_{n,L}^{(q)}$ , we can divide the other  $Q - 1$  test statistics  $\{W_{n,L}^{(q')}\}_{q' \in [Q] \setminus \{q\}}$  into two parts: (i) the ones not belonging to  $\mathcal{A}_q(\gamma)$  have quite weak dependence with  $W_{n,L}^{(q)}$ , and (ii) the ones belonging to  $\mathcal{A}_q(\gamma)$  have relatively strong dependence with  $W_{n,L}^{(q)}$ . To construct the theoretical guarantee of the proposed multiple testing procedure, it is a common practice to analyze these two parts separately with different technical tools. See also Liu (2013), Chang, Shao, and Zhou (2016), and Xia et al. (2020). The next theorem indicates that our multiple testing procedure can control both FDR and FDP at the level  $\alpha Q_0/Q$  asymptotically.

**Theorem 3.** Let the conditions in Proposition 1 hold. Assume that  $\max_{1 \leq q \neq q' \leq Q} |\text{corr}\{\eta^{(q)}, \eta^{(q')}\}| \leq r$  for some constant  $r \in (0, 1)$ , and  $\max_{q \in [Q]} |\mathcal{A}_q(\gamma)| = o(Q^\nu)$  for some constants  $\gamma > 0$  and  $0 < \nu < (1 - r)/(1 + r)$ . If  $Q^{12} B^5 L^4 \log(nBK_{\max}) \max\{(L \log K_{\max})^4, (\log n)^4\} = o(n)$  and  $Q^{12} B^4 L^{13} (\log K_{\max})^7 = o(n)$  with  $K_{\max} = \max_{q \in \mathcal{H}_0} |\mathcal{K}_q|$ , then  $\limsup_{n, Q \rightarrow \infty} \text{FDR}(\hat{t}) \leq \alpha Q_0/Q$  and  $\lim_{n, Q \rightarrow \infty} \mathbb{P}\{\text{FDP}(\hat{t}) \leq \alpha Q_0/Q + \varepsilon\} = 1$  for any  $\varepsilon > 0$ .

The conditions on  $\text{corr}\{\eta^{(q)}, \eta^{(q')}\}$  and  $\mathcal{A}_q(\gamma)$  are commonly used in the literature, which are applied to bound the variance of  $\sum_{q \in \mathcal{H}_0} I\{V_{n,L}^{(q)} \geq t\}$  appeared in the numerator of  $\text{FDP}(t)$ . See, for example, Liu (2013), Chang, Shao, and Zhou (2016), and Xia et al. (2020). Specifically, the condition  $\max_{1 \leq q \neq q' \leq Q} |\text{corr}\{\eta^{(q)}, \eta^{(q')}\}| \leq r$  for some constant  $r \in (0, 1)$  imposes a constraint on the magnitudes of the correlations between different  $\eta^{(q)}$  and  $\eta^{(q')}$ . The condition  $\max_{q \in [Q]} |\mathcal{A}_q(\gamma)| = o(Q^\nu)$  controls the number of pairs  $\{\eta^{(q)}, \eta^{(q')}\}$  those are highly correlated. Recall  $B$  and  $L$  are, respectively, the length of each subgroup and the pre-determined integer involved in the  $L$ -statistic  $W_{n,L}^{(q)}$  defined as (2.11). When  $B$  and  $L$  are selected as fixed constants, Theorem 3 holds if  $Q^{12} (\log K_{\max})^7 = o(n)$  and  $Q^{12} \log(nK_{\max}) \max\{(\log K_{\max})^4, (\log n)^4\} = o(n)$ . Under the mild condition  $K_{\max} \lesssim n^\kappa$  for some sufficiently large constant  $\kappa > 0$  in practice, the restrictions can be simplified as  $Q^{12} (\log n)^7 = o(n)$ , which requires  $Q$ , the number of multiple tests, to satisfy  $Q = o\{n^{1/12} (\log n)^{-7/12}\}$ .

## 2.4. A Distributed Algorithm

Generating multivariate normal random vectors is a key step in the implementation of our proposed tests. For example, in the global test (2.6), we need to generate  $\hat{\xi}_1, \dots, \hat{\xi}_N$  independently from  $\mathcal{N}(0, \hat{R}_n)$  to determine the critical value  $\hat{c}_{V_{L,\alpha}}$ , where  $\hat{R}_n$  is a  $d \times d$  matrix with  $d = \sum_{q=1}^Q |\mathcal{K}_q|$ . The standard approach for generating an  $s$ -dimensional random vector  $\zeta \sim \mathcal{N}(0, \Xi)$  includes three steps: (i) perform the Cholesky decomposition

for the  $s \times s$  matrix  $\Xi = A^\top A$ ; (ii) generate  $s$  independent  $\mathcal{N}(0, 1)$  random variables  $z = (z_1, \dots, z_s)^\top$ ; (iii) perform the transformation  $\zeta = A^\top z$ . Computationally this is an  $s^3$ -hard problem. In practice,  $Q$  and  $|K_q|$ 's can be large, which causes the generation of a  $d$ -dimensional random vector followed  $\mathcal{N}(0, \hat{R}_n)$  requires very high computing cost. To overcome this challenge and further enhance the computational efficiency of our proposed methods, we propose in this section a distributed algorithm that can avoid generating high-dimensional normal distributed random vectors directly.

We use the global test (2.6) as the example to illustrate our idea. We first divide the full dataset into  $K$  blocks with  $n_k$  observations in the  $k$ th block  $\mathcal{I}_k$ . Given  $B \geq 5$ , we can obtain  $M_k = n_k - B + 1$  moving subgroups in the  $k$ th block  $\mathcal{I}_k$  and calculate  $U_{m,q,S_1,S_2}^{(k),\text{dist}}$  in the same manner as  $U_{m,q,S_1,S_2}$  appeared in (2.5) but with replacing  $\{X_m, \dots, X_{m+B-1}\}$  by the  $m$ th subgroup in the  $k$ th block  $\mathcal{I}_k$  for  $m \in [M_k]$ ,  $(S_1, S_2) \in K_q$  and  $q \in [Q]$ . For given  $k \in [K]$  and  $m \in [M_k]$ , we stack  $\{U_{m,q,S_1,S_2}^{(k),\text{dist}}\}_{q \in [Q], (S_1, S_2) \in K_q}$  into a  $d$ -dimensional vector  $U_m^{(k),\text{dist}} = (U_{m,1}^{(k),\text{dist}}, \dots, U_{m,d}^{(k),\text{dist}})^\top$ . Write  $\bar{U}^{(k),\text{dist}} = (\bar{U}_1^{(k),\text{dist}}, \dots, \bar{U}_d^{(k),\text{dist}})^\top = M_k^{-1} \sum_{m=1}^{M_k} U_m^{(k),\text{dist}}$ . Then we define the distributed analogue of the global test statistic  $W_{n,L}$  specified in (2.8) as follows:

$$W_{n,L}^{\text{dist}} = \max_{1 \leq j_1 < \dots < j_L \leq d} \sum_{s=1}^L \frac{n_{\text{dist}}^{-1/2} \sum_{k=1}^K M_k \bar{U}_{j_s}^{(k),\text{dist}}}{[n_{\text{dist}}^{-1} \sum_{k=1}^K M_k \{\hat{\sigma}_{j_s}^{(k),\text{dist}}\}^2]^{1/2}}, \quad (2.15)$$

where  $n_{\text{dist}} = \sum_{k=1}^K M_k$  and  $\hat{\sigma}_j^{(k),\text{dist}}$  is the estimate for the standard deviation of  $\sqrt{M_k} \bar{U}_j^{(k),\text{dist}}$ . Define  $\hat{\Sigma}^{(k),\text{dist}}$  in the same manner as  $\hat{\Sigma}_n$  given in (2.10) but with replacing  $\{U_m\}_{m=1}^{M_k}$  by  $\{U_m^{(k),\text{dist}}\}_{m=1}^{M_k}$ . We can select  $\{\hat{\sigma}_j^{(k),\text{dist}}\}^2$  as the associated element in the main diagonal of  $\hat{\Sigma}^{(k),\text{dist}}$ .

Identical to Proposition 1, we can also show that the null-distribution of  $W_{n,L}^{\text{dist}}$  can be approximated by the distribution of  $f_L(\xi)$  with  $f_L(\cdot)$  defined in Section 2.2 and  $\xi \sim \mathcal{N}(0, \check{R}_n)$  for some  $d \times d$  matrix  $\check{R}_n$ . To avoid generating  $d$ -dimensional normal distributed random vector  $\xi$ , we consider to generate  $d$ -dimensional  $\check{\xi}$  as follows:

$$\check{\xi} = (\tilde{D}_n^{\text{dist}})^{-1/2} \sum_{k=1}^K \varepsilon_k n_{\text{dist}}^{-1/2} M_k \{\bar{U}^{(k),\text{dist}} - \bar{U}^{\text{dist}}\} \quad (2.16)$$

with  $\tilde{D}_n^{\text{dist}} = \text{diag}[n_{\text{dist}}^{-1} \sum_{k=1}^K M_k \{\hat{\sigma}_1^{(k),\text{dist}}\}^2, \dots, n_{\text{dist}}^{-1} \sum_{k=1}^K M_k \{\hat{\sigma}_d^{(k),\text{dist}}\}^2]$ , where  $\{\varepsilon_k\}_{k=1}^K$  are independent Rademacher random variables with  $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$ ,  $\bar{U}^{\text{dist}} = (\bar{U}_1^{\text{dist}}, \dots, \bar{U}_d^{\text{dist}})^\top = n_{\text{dist}}^{-1} \sum_{k=1}^K M_k \bar{U}^{(k),\text{dist}}$ , and  $\hat{\sigma}_j^{(k),\text{dist}}$  is the estimate for the standard deviation of  $\sqrt{M_k} \{\bar{U}_j^{(k),\text{dist}} - \bar{U}_j^{\text{dist}}\}$  that can be determined in the same manner as  $\hat{\sigma}_j^{(k),\text{dist}}$  in (2.15) but with replacing the data  $\{U_m^{(k),\text{dist}}\}_{m=1}^{M_k}$  by  $\{\varepsilon_k U_m^{(k),\text{dist}} - \varepsilon_k \bar{U}^{\text{dist}}\}_{m=1}^{M_k}$ . In comparison to generating  $\xi \sim \mathcal{N}(0, \check{R}_n)$ , the newly method (2.16) only requires to generate  $K$  independent Rademacher random variables  $\varepsilon_1, \dots, \varepsilon_K$  which can largely reduce the computational cost. Let  $R_n^{\text{dist}} =$

$(D_n^{\text{dist}})^{-1/2} \Sigma_n^{\text{dist}} (D_n^{\text{dist}})^{-1/2}$  with  $\Sigma_n^{\text{dist}} = \text{cov}(n_{\text{dist}}^{1/2} \bar{U}^{\text{dist}})$  and  $D_n^{\text{dist}} = \text{diag}(\Sigma_n^{\text{dist}}) = \text{diag}\{(\sigma_1^{\text{dist}})^2, \dots, (\sigma_d^{\text{dist}})^2\}$ . Proposition 2 provides the theoretical guarantee for such method.

**Proposition 2.** Write  $\sigma_j^{(k),\text{dist}} = \text{var}^{1/2}(M_k^{-1/2} \sum_{m=1}^{M_k} U_{mj}^{(k),\text{dist}})$  for any  $k \in [K]$  and  $j \in [d]$ . Let  $\min_{k \in [K]} n_k \asymp \max_{k \in [K]} n_k$ . Assume there exists a universal constant  $c_2 > 0$  such that  $\min_{v \in \mathcal{F}} v^\top R_n^{\text{dist}} v \geq c_2$  and  $\min_{k \in [K]} \min_{j \in [d]} \sigma_j^{(k),\text{dist}} \geq c_2$ . Under the global null hypothesis  $H_0$  given in (2.6), we have  $\sup_{z \in \mathbb{R}} |\mathbb{P}(W_{n,L}^{\text{dist}} > z) - \mathbb{P}\{f_L(\check{\xi}) > z | X_1, \dots, X_n\}| = o_p(1)$  provided that  $B^5 L^8 K \log(dBK) (\log d)^4 = o(n)$ ,  $B^3 L^{13} \{\log(dK)\}^3 (\log d)^7 = o(K)$  and  $B^4 K^{3/2} L^{-2} \{\log(dBK)\}^3 (\log d)^{-2} = o(n)$ .

When we select  $B$  as a fixed constant, Proposition 2 holds provided that  $L^8 K \log(dK) (\log d)^4 = o(n)$ ,  $L^{13} \{\log(dK)\}^3 (\log d)^7 = o(K)$  and  $K^{3/2} L^{-2} \{\log(dK)\}^3 (\log d)^{-2} = o(n)$ , which requires the pre-determined integer  $L$  should satisfy  $L \ll \min\{n^{1/8}, K^{1/13}\}$ . Furthermore, if  $L$  is also selected as a fixed constant, Proposition 2 holds provided that  $\{\log(dK)\}^3 (\log d)^7 = o(K)$  and  $K^{3/2} \{\log(dK)\}^3 (\log d)^{-2} = o(n)$ . Additionally, with restricting  $K$ , the number of dataset blocks involved in the distributed algorithm, satisfying  $K = o(d)$ , Proposition 2 is valid for any  $K$  such that  $(\log d)^{10} \ll K \ll \min\{d, n^{2/3} (\log d)^{-2/3}\}$ , which allows  $\log d = o(n^{1/16})$  that  $d = \sum_{q=1}^Q |K_q|$  growing exponentially with the sample size  $n$ . To obtain a guideline for the choices of  $K$  in practice, we have performed simulations to assess the statistical properties of the proposed tests for different  $K$ . The simulation settings in Section 3 have similar sample size and image resolutions as the real applications. For example,  $V = 200^2$  and  $V = 500^2$  correspond to the number of voxels in the 3mm and 2mm MNI standard brain templates respectively. For the dataset with a large sample size, for example,  $n \geq 900$ , the proposed testing method performs very well for a wide range of  $K$  ( $K = 20, 30, 40$ ). For the dataset with a moderate or small sample size, that is,  $n \leq 600$ , a relatively small  $K$  (e.g.,  $K = 20$ ) may lead to a slightly better power compared to other choices while the empirical size still can be controlled at around 0.05. Please refer to Section 3 for more details. Besides considering the statistical properties, we may choose  $K$  taking into account the available computational resources including the number of processors, memory storage and computing speed. A smaller  $K$  may require less processors but more local memory per processors leading to a longer computational time. A larger  $K$  may need more processors but less local memory which can reduce the total computational time.

Let  $\hat{F}^{\text{dist}}(t) = \mathbb{P}\{f_L(\check{\xi}) \leq t | X_1, \dots, X_n\}$  for  $\check{\xi}$  defined as (2.16). We can reject the global null hypothesis  $H_0$  in (2.6) at the significance level  $\alpha \in (0, 1)$  if  $W_{n,L}^{\text{dist}} > \hat{c}_{L,\alpha}^{\text{dist}}$  with  $\hat{c}_{L,\alpha}^{\text{dist}} = \inf\{t \in \mathbb{R} : \hat{F}^{\text{dist}}(t) \geq 1 - \alpha\}$ . In practice, for some sufficiently large  $N$ , letting  $\check{\xi}_1, \dots, \check{\xi}_N$  be independent random vectors followed (2.16), we can approximate  $\hat{c}_{L,\alpha}^{\text{dist}}$  by the  $[N\alpha]$ th largest value of  $f_L(\check{\xi}_1), \dots, f_L(\check{\xi}_N)$ .

In the multiple test, the distributed analogue of the  $p$ -value  $\text{pv}_L^{(q)}$  specified below (2.12) can be similarly defined. For given  $q \in [Q]$ ,  $k \in [K]$  and  $m \in [M_k]$ , we

stack  $\{U_{m,q,S_1,S_2}^{(k),\text{dist}}\}_{(S_1,S_2) \in \mathcal{K}_q}$  into a  $|\mathcal{K}_q|$ -dimensional vector  $U_{m,(q)}^{(k),\text{dist}} = (U_{m,q,1}^{(k),\text{dist}}, \dots, U_{m,q,|\mathcal{K}_q|}^{(k),\text{dist}})^\top$ . For the  $q$ th hypothesis  $H_{0,q}$ , the test statistic  $W_{n,L}^{(q),\text{dist}}$  can be defined in the same manner as  $W_{n,L}^{\text{dist}}$  in (2.15) but with replacing  $\{U_m^{(k),\text{dist}}\}_{m=1}^{M_k}$  by  $\{U_{m,(q)}^{(k),\text{dist}}\}_{m=1}^{M_k}$ . Write  $\bar{U}_{(q)}^{(k),\text{dist}} = M_k^{-1} \sum_{m=1}^{M_k} U_{m,(q)}^{(k),\text{dist}}$  and  $\bar{U}_{(q)}^{\text{dist}} = n_{\text{dist}}^{-1} \sum_{k=1}^K M_k \bar{U}_{(q)}^{(k),\text{dist}}$ . Then the  $|\mathcal{K}_q|$ -dimensional  $\xi_{(q)}^{(q)}$  can be generated in the same manner as  $\xi$  in (2.16) but with replacing  $\bar{U}^{(k),\text{dist}}$  and  $\bar{U}^{\text{dist}}$  by  $\bar{U}_{(q)}^{(k),\text{dist}}$  and  $\bar{U}_{(q)}^{\text{dist}}$ , respectively. The distributed analogue of the  $p$ -value of  $H_{0,q}$  is then defined as  $\text{pv}_L^{(q),\text{dist}} = \mathbb{P}\{f_L(\xi_{(q)}^{(q)}) \geq W_{n,L}^{(q),\text{dist}} | X_1, \dots, X_n\}$ . For implementation, we generate independent random vectors  $\xi_1^{(q)}, \dots, \xi_N^{(q)}$  for some sufficiently large  $N$ , and  $\text{pv}_L^{(q),\text{dist}}$  can be approximated by  $\hat{\text{pv}}_L^{(q),\text{dist}} = N^{-1} \sum_{j=1}^N I\{f_L(\xi_j^{(q)}) \geq W_{n,L}^{(q),\text{dist}}\}$ .

### 3. Simulations

In this section, we perform simulations to evaluate the finite-sample performance of the proposed tests. We consider the following data generating process for the multimodal imaging data. Assume there are  $J$  different imaging modalities for each subject. Let  $\mathcal{R} = \{1, \dots, V\}$  denote the set of indices of the two-dimensional (2D) image, which is divided as  $G$  predefined brain regions  $\mathcal{R}_1, \dots, \mathcal{R}_G$  by the  $k$ -means clustering method, and  $v \in \mathcal{R}$  denote the index of the voxels. The coordinates of the voxel  $v$  are denoted by  $h_v \in \mathbb{R}^2$ , and the coordinates of the center in the region  $\mathcal{R}_g$  are denoted by  $\bar{h}_g$ . For each  $i \in [n]$ , let  $Y_j^{(i)}(v)$  be the image signal of type  $j$  at voxel  $v$  for the subject  $i$ . Assume  $Y_j^{(i)}(v) = b_j^{(i)}(v) + \sum_{j'=1}^J d_{jj'}^{(i)}(v) \alpha_{jj'}^{(i)}(v) + \epsilon_j^{(i)}(v)$ , where  $b_j^{(i)}(v) = \sum_{g=1}^G \beta_{jg}^{(i)} I(v \in \mathcal{R}_g)$  and  $d_{jj'}^{(i)}(v) = \sum_{g=1}^G \delta_{jj',g}^{(i)} I(v \in \mathcal{R}_g)$  with  $\delta_{jj',g}^{(i)} = \delta_{j',j,g}^{(i)} \geq 0$  for  $j < j'$ . Here  $d_{jj'}^{(i)}(v)$ 's are regression coefficients, and  $b_j^{(i)}(v)$  and  $\alpha_{jj'}^{(i)}(v)$ 's are some random variables. For each  $i \in [n]$ , write  $\beta^{(i)} = \{\beta_1^{(i),\top}, \dots, \beta_J^{(i),\top}\}^\top$  with  $\beta_j^{(i)} = \{\beta_{j,1}^{(i)}, \dots, \beta_{j,G}^{(i)}\}^\top$ . We generate  $\beta^{(i)}$  independently from the normal distribution  $\mathcal{N}(0, E)$ , where  $E = (E^{jj'})_{j,j' \in [J]}$  with  $E^{jj'} = (e_{gg'}^{jj'})_{g,g' \in [G]}$ . For any  $i \in [n]$  and  $(j, j')$  such that  $j \leq j'$ ,  $\{\alpha_{jj'}^{(i)}(v)\}_{v \in \mathcal{R}}$  is independently generated from a Gaussian process  $\mathcal{GP}(0, \kappa_\alpha)$  with the covariance function  $\kappa_\alpha(v, v') = \sum_{g=1}^G \exp\{-0.001(|h_v - \bar{h}_g|_2^2 + |h_{v'} - \bar{h}_g|_2^2) - 10|h_v - h_{v'}|_2^2\} I(v, v' \in \mathcal{R}_g)$ . For  $j > j'$ , let  $\alpha_{jj'}^{(i)}(v) = \alpha_{j'j}^{(i)}(v)$ . We will consider several scenarios for the regression coefficients  $d_{jj'}^{(i)}(v)$ 's and the covariance matrix  $E$  of  $\beta^{(i)}$  to mimic the null hypothesis and the alternative hypothesis, which will be specified later. Given the regression coefficients  $d_{jj'}^{(i)}(v)$ 's and the generated  $b_j^{(i)}(v)$ 's and  $\alpha_{jj'}^{(i)}(v)$ 's,  $\epsilon_j^{(i)}(v)$ 's are independently generated from  $\mathcal{N}(0, \sigma^2)$ , where  $\sigma^2$  is chosen such that the R-squared is 0.95.

In our simulations, we set  $J = 3$ ,  $G \in \{16, 25\}$  and  $V \in \{100^2, 200^2, 500^2\}$ . We consider the sample size  $n \in \{300, 600, 900\}$ , and divide the observations equally to  $K = 30$  blocks and put each block on a processor to implement the distributed computing procedures presented in Section 2.4 with

$N = 5000$ . All simulation results are based on 1000 replications. We have also tried  $K = 20$  and 40 in the simulation which lead to similar results as  $K = 30$ . We only present the results based on  $K = 30$ . The results based on  $K = 20$  and 40 can be found in the supplementary material.

We consider both the global and multiple tests for the null hypotheses (a)–(c) given in (1.1)–(1.3), respectively. To evaluate the empirical sizes, we set  $\delta_{jj',g} = 0$  and  $e_{gg'}^{jj'} = I(j = j')I(g = g')$  for any  $j, j' \in [J]$  and  $g, g' \in [G]$ . To evaluate the empirical powers, we consider the following three scenarios (M1)–(M3), which corresponds to the testing problems (a)–(c), respectively.

- (M1) Let  $e_{gg'}^{jj'} = I(j = j')I(g = g')$  and  $\delta_{jj',g} = \{I(j = j') + 6I(|j - j'| = 1)\}I(1 \leq g \leq 4)$ .
- (M2) Let  $e_{gg'}^{jj'} = I(j = j')\{I(g = g') + 0.4I(|g - g'| = 1) - 0.6I(|g - g'| = 1)I(j = 2)\}$  and  $\delta_{jj',g} = 0$ .
- (M3) Let  $e_{gg'}^{jj'} = I(j = j')I(g = g') + I(g' = g + 1)\{0.4I(j = 1, j' = 2) - 0.4I(j = 1, j' = 3) + 0.2I(j = 2, j' = 3)\}$  for  $j \leq j'$ , and  $e_{gg'}^{jj'} = e_{g',g}^{j',j}$  for  $j > j'$ , and also let  $\delta_{jj',g} = 0$ .

For the global tests, we set the significance level  $\alpha = 5\%$ . When  $G = 16$ ,  $d = 48, 360$ , and 768 in the global tests of (a)–(c), respectively. When  $G = 25$ ,  $d = 75, 900$ , and 1875 in the global tests of (a)–(c), respectively. See the definition of  $d$  above (2.8). The empirical sizes and powers of the proposed global tests are reported in Table 1 with  $L = 1, 3$  and 5. For  $L = 1$ , the proposed tests have good size control around 5% and the power of the proposed tests increases quickly to 1 in all the three testing problems as the sample size  $n$  grows from 300 to 900. In comparison to the proposed tests with  $L = 1$ , we find that the proposed tests with  $L > 1$  still have good size control around 5% when  $d$  is small (testing problem (a)) and they are a little bit conservative when  $d$  is large (testing problems (b) and (c)), but the associated powers grow as  $L$  increases.

The empirical FDRs and powers of the proposed multiple tests are reported in Table 2. Due to the number of modalities  $J = 3$  in our setting, we know  $|\mathcal{K}_q| = 3$  in (2.1) for all  $q \in [Q]$  in the multiple tests for (a)–(b) and we can only consider the proposed test statistics with  $L \leq 3$ . For the testing problems (a)–(c),  $(Q, Q_0) = (16, 12), (120, 105)$ , and  $(136, 121)$  when  $G = 16$ , and  $(Q, Q_0) = (25, 21), (300, 276)$ , and  $(325, 301)$  when  $G = 25$ , respectively. Given the significance level  $\alpha = 5\%$ , we can obtain  $\hat{t}_r$  defined as (2.14) in the  $r$ th simulation replication. For each  $q \in [Q]$ , denote by  $V_{n,L,r}^{(q)}$  the normal quantile transformation of the  $p$ -value for  $H_{0,q}$  in the  $r$ th simulation replication. See its definition below (2.12). We define the empirical power of the proposed multiple test as  $\frac{1}{1000} \sum_{r=1}^{1000} \frac{1}{Q-Q_0} \sum_{q \notin \mathcal{H}_0} I\{V_{n,L,r}^{(q)} > \hat{t}_r\}$ . Theorem 3 implies that the FDR should be controlled by  $\alpha Q_0/Q$ , which equals to 3.75%, 4.38%, and 4.45% when  $G = 16$ , and 4.20%, 4.60%, and 4.63% when  $G = 25$  in the testing problems (a)–(c), respectively. Table 2 shows that the proposed multiple tests have good performances for FDR control, and the empirical powers improve considerably when the sample size  $n$  increases from 300 to 900.

As we mentioned in Section 1, some existing works can be also used to derive the  $p$ -value of each  $H_{0,g}^{(a)}$  in testing problem (a). Based on the  $p$ -values obtained by these methods, we can

**Table 1.** Empirical sizes and powers of the proposed global tests with  $K = 30$  for the testing problems (a)–(c) given in (1.1)–(1.3), respectively.

	$G$	$n$	$V$	Problem (a)			Problem (b)			Problem (c)		
				$L = 1$	$L = 3$	$L = 5$	$L = 1$	$L = 3$	$L = 5$	$L = 1$	$L = 3$	$L = 5$
Global sizes	16	300	$100^2$	5.70	5.70	3.90	6.00	3.50	2.10	5.10	4.00	2.90
			$200^2$	6.30	4.80	3.80	6.50	4.40	2.20	5.20	4.00	3.10
			$500^2$	5.20	4.50	4.40	6.20	3.50	1.70	7.60	5.70	3.70
		600	$100^2$	4.40	4.40	4.00	3.90	3.30	2.30	3.90	2.90	2.30
			$200^2$	4.70	5.10	4.40	4.30	3.00	1.90	4.80	2.50	1.80
			$500^2$	5.20	4.90	3.70	5.10	3.70	2.30	5.10	2.90	2.50
	25	300	$100^2$	4.90	4.40	4.60	2.90	2.90	2.60	4.70	2.80	2.30
			$200^2$	6.00	5.30	5.20	4.10	2.30	2.00	2.90	2.40	2.00
			$500^2$	5.20	5.10	4.90	4.40	2.30	2.50	3.70	2.30	1.90
		600	$100^2$	6.20	4.80	3.20	5.40	3.10	1.50	6.20	3.50	2.10
			$200^2$	7.40	5.50	4.40	7.20	3.50	1.60	7.50	4.30	3.10
			$500^2$	7.50	6.10	5.60	6.20	3.00	1.90	7.20	5.10	3.20
	900	300	$100^2$	7.20	5.30	5.80	4.70	3.10	2.20	3.90	3.00	2.50
			$200^2$	4.90	4.70	3.80	4.60	2.80	1.50	3.60	2.40	1.20
			$500^2$	4.50	4.70	4.40	4.70	2.80	2.30	4.90	2.80	1.00
		600	$100^2$	2.90	3.80	4.20	3.40	2.50	1.70	3.30	1.90	1.00
			$200^2$	3.90	3.90	3.70	3.00	2.10	1.50	3.70	2.20	1.30
			$500^2$	4.50	4.40	3.80	6.10	2.70	2.10	2.40	1.80	1.50
Global powers	16	300	$100^2$	100.00	100.00	100.00	93.30	99.40	99.90	88.90	98.20	99.30
			$200^2$	100.00	100.00	100.00	92.70	99.10	99.40	86.90	96.30	97.70
			$500^2$	100.00	100.00	100.00	94.40	99.00	99.60	86.20	97.70	99.40
		600	$100^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$200^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$500^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	25	300	$100^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$200^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$500^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		600	$100^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$200^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$500^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	900	300	$100^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$200^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$500^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		600	$100^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$200^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
			$500^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

NOTE: All numbers reported in the table are multiplied by 100.

consider the associated multiple testing procedure in the same manner as ours given in Section 2.3. Here we consider three methods: (i) the  $k$ -variate HSIC-based test (dHSIC) in Pfister et al. (2018), (ii) the generalized distance covariance based test (GdCov) in Jin and Matteson (2018), and (iii) the test based on the high order distance covariance in Chakraborty and Zhang (2019). The dHSIC test and GdCov test are implemented by calling the R-functions `dhsic.test` and `mdm_test`, respectively, in the R-packages `dHSIC` and `EDMeasure`. There are the three test statistics in Chakraborty and Zhang (2019). Here, we choose their bias-corrected Studentized test statistic (JdCov) with  $c = 2$  as suggested by the authors. The code for JdCov test is available in the supplementary material of Chakraborty and Zhang (2019). Table 2 also includes the empirical FDRs and powers of the multiple testing procedures based on these three competitors in testing problem (a), which shows that our proposed method has similar performance as these three methods. However, these three competitors cannot be applied to address the multiple tests in the testing problems (b) and (c). Since the implementation of the JdCov test relies on a bootstrap procedure which requires high computing cost, the results of

JdCov in Table 2 are only based on 100 replications, and its associated results for  $V = 500^2$  are omitted. Different from the linear dependency considered here among  $Y_j^{(i)}(v)$ 's, we have also considered a simulation setting with nonlinear dependency imposed among  $Y_j^{(i)}$ 's for the testing problem (a). See Section A.2 in the supplementary material. It shows that (i) the proposed method significantly outperforms the three competitors; (ii) the proposed method is robust to  $V$ ; and (iii) the three competitors suffer obvious power loss with the increase of  $V$ .

#### 4. Analysis of HCP Data

We performed the analysis of the motivating HCP task fMRI data using the proposed testing procedures. All the data were collected from the HCP-1200 release (Van Essen et al. 2013). The HCP minimally preprocessed pipeline (Glasser et al. 2013) was adopted to preprocess the fMRI time series data. In particular, the pipeline includes motion correction, distortion correction, brain-boundary based linear registration, nonlinear registration to the standard MNI 2mm space and the intensity normalization. Our analysis focused on 922 subjects with three fMRI

**Table 2.** Empirical FDRs and powers of the proposed multiple tests with  $K = 30$ , and three competitors (JdCov, dHSIC, GdCov) for the testing problems (a)–(c) given in (1.1)–(1.3), respectively.

	$G$	$n$	$V$	Problem (a)					Problem (b)		Problem (c)		
				$L = 1$	$L = 3$	JdCov	dHSIC	GdCov	$L = 1$	$L = 3$	$L = 1$	$L = 3$	
Empirical FDRs	16	300	100 <sup>2</sup>	1.96	2.19	2.80	1.83	3.83	4.09	2.43	4.11	2.32	
			200 <sup>2</sup>	2.44	2.09	3.53	1.89	3.26	3.75	2.55	4.50	2.57	
			500 <sup>2</sup>	1.93	2.46	NA	1.81	3.22	3.10	2.52	4.08	1.46	
		600	100 <sup>2</sup>	2.37	2.42	2.73	1.76	2.23	3.75	5.11	2.99	3.32	
			200 <sup>2</sup>	2.77	2.67	2.33	2.04	1.75	3.44	4.98	2.12	2.73	
			500 <sup>2</sup>	2.31	2.57	NA	2.09	1.96	3.58	4.95	2.35	2.98	
		900	100 <sup>2</sup>	2.27	2.86	1.80	1.60	1.95	4.00	5.47	3.16	3.39	
			200 <sup>2</sup>	2.43	2.67	2.13	2.04	2.88	3.88	5.28	3.21	3.33	
			500 <sup>2</sup>	2.37	2.52	NA	1.93	2.49	3.83	5.39	3.59	3.59	
	25	300	100 <sup>2</sup>	2.67	2.31	2.20	1.54	3.88	4.31	3.88	4.29	1.79	
			200 <sup>2</sup>	2.55	2.49	1.60	1.58	3.38	3.83	4.08	4.55	2.00	
			500 <sup>2</sup>	2.40	2.65	NA	1.53	3.81	3.91	4.02	5.70	3.38	
		600	100 <sup>2</sup>	2.69	2.79	1.60	1.56	3.97	4.47	5.74	3.72	3.92	
			200 <sup>2</sup>	2.59	2.71	3.80	1.69	3.74	4.00	5.53	3.65	3.83	
			500 <sup>2</sup>	2.37	2.62	NA	1.74	3.89	4.13	5.79	3.69	4.14	
		900	100 <sup>2</sup>	2.72	3.30	2.13	1.80	3.43	4.40	6.47	3.13	4.16	
			200 <sup>2</sup>	3.07	3.09	3.13	1.83	4.18	4.33	6.53	3.25	4.22	
			500 <sup>2</sup>	2.45	2.83	NA	1.87	3.77	4.29	6.37	3.38	4.30	
	Empirical powers	16	300	100 <sup>2</sup>	99.78	98.80	100.00	100.00	100.00	18.57	41.80	14.63	22.96
				200 <sup>2</sup>	99.88	99.70	100.00	100.00	100.00	19.10	42.93	14.07	22.03
				500 <sup>2</sup>	99.93	99.73	NA	100.00	100.00	18.87	42.48	14.12	23.73
			600	100 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	87.21	96.95	72.99	91.04
				200 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	87.00	96.82	68.83	88.44
				500 <sup>2</sup>	100.00	100.00	NA	100.00	100.00	86.83	96.99	68.66	89.43
900			100 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	99.44	99.89	97.63	99.20	
			200 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	99.36	99.85	97.63	99.25	
			500 <sup>2</sup>	100.00	100.00	NA	100.00	100.00	99.45	99.91	97.90	99.36	
25		300	100 <sup>2</sup>	99.95	99.78	100.00	100.00	100.00	12.67	41.88	8.93	15.28	
			200 <sup>2</sup>	99.93	99.58	100.00	100.00	100.00	12.80	40.80	8.92	15.30	
			500 <sup>2</sup>	99.90	99.50	NA	100.00	100.00	12.38	40.90	9.03	15.23	
		600	100 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	86.85	96.00	78.22	91.86	
			200 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	87.39	95.77	78.43	91.77	
			500 <sup>2</sup>	100.00	100.00	NA	100.00	100.00	87.20	95.76	78.99	90.88	
		900	100 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	99.00	99.79	98.34	99.42	
			200 <sup>2</sup>	100.00	100.00	100.00	100.00	100.00	99.07	99.76	98.51	99.47	
			500 <sup>2</sup>	100.00	100.00	NA	100.00	100.00	99.05	99.81	97.85	99.32	

NOTE: All numbers reported in the table are multiplied by 100.

task domains: working memory (n-back), language processing (story and math) and social cognition. For the details of the tasks, please refer to Barch et al. (2013). To generate the subject-specific contrast maps from the fMRI time series data, the fixed-effects models were fitted to estimate the average brain activation over time using FSL (Jenkinson et al. 2012). For some tasks, multiple contrast maps were created beyond the standard experimental versus control condition. We analyzed five contrast maps with two (2back-0back, task-rest) for the working memory task, one (story-math) for language processing task, and two (mental, random) for the social cognition task. Our analysis indicates that the language processing, working memory and social cognition brain activities are strongly dependent with statistical significance in most parts of the inferior frontal gyrus. For more details, please refer to Section B in the supplementary material.

We considered three testing problems (a)–(c), defined as (1.1)–(1.3) respectively, in the analysis of the HCP fMRI contrast maps. We performed the distributed computing procedure by splitting the first 900 subjects equally on 30 processors and

putting the rest 22 observations on the 31st processor, and calculated the critical values by the method given in Section 2.4.

*Testing independence among modalities over regions:* In the testing problem (a), we tested the independence of the brain activation from the three fMRI tasks, that is, working memory (2back-0back), language processing (story-math) and social cognition (mental) over the 90 AAL regions. The global test, in which  $d = 270$ , rejected the null hypothesis of the three types of brain activation being mutually independent in all 90 regions with a  $p$ -value  $< 0.0001$ . Of note, we took different values for  $L$ : 1, 5, 10, and 15; and the associated  $p$ -values were all less than 0.0001. The multiple tests were conducted for the same null hypothesis in each of the 90 AAL regions with the FDR controlled at 0.001 and  $L = 1$ . Each test was rejected if the corresponding  $p$ -value  $\leq 1.35 \times 10^{-3}$ . We identified 30 out of 90 AAL regions where the three types of fMRI task activities are significantly dependent. This result provides some new insights about relationships between brain functions and human behaviors.

*Testing independence between regions within modalities:* In the testing problem (b), we mainly concentrated on assessing the independence of fMRI task activity between regions in the default mode network (Raichle 2015, DMN). In our analysis, we adopted the parcellation of brain by Power et al. (2011) consisting of 264 regions among which the DMN has 58 regions. The DMN was developed for the brain activity in experimental participants at wakeful rest, and the DMN also can be active in social cognition and working memory tasks (Spreng 2012). In this problem, we tested the independence of brain regions in the three fMRI tasks, that is, working memory (2back-0back), language processing (story-math) and social cognition (mental) over the 58 DMN regions. The global test, in which  $d = 4959$ , rejected the null hypothesis of independence of brain activities between all the region pairs in DMN for the three fMRI task contrasts with a  $p$ -value  $< 0.0001$  for  $L = 1, 5, 10$ , and 15. Furthermore, we are interested in identifying brain regions which are statistically significantly correlated in the working memory task ( $j = 1$ ), the language processing task ( $j = 2$ ) and the social cognition task ( $j = 3$ ), respectively. For each  $j \in \{1, 2, 3\}$ , we consider the multiple tests for null hypotheses  $H_{0,j,g,g'}: Y_{j,g}$  and  $Y_{j,g'}$  are independent. Due to  $g < g'$ , there are  $\binom{58}{2} = 1653$  null hypotheses. The multiple tests were conducted with the FDR controlled at 0.001 and  $L = 1$ . Each test was rejected if the corresponding  $p$ -value  $\leq 5.91 \times 10^{-5}$  for the working memory task (2back-0back), the language processing (story-math) task and the social cognition (mental) task, respectively. There are 627, 227, and 494 region pairs identified for strong dependence of brain activity in the working memory task (2back-0back), language processing (story-math) task and the social cognition (mental) task, respectively. Each testing result can be represented as an undirected graph, where the regions in the DMN denote the nodes and each significant region pair has an edge. Given the graphs are very dense, in Figure 1, we present the degree for each region, that is, the number of other regions in the DMN that are significantly dependent on this region. The brain activation dependence among the default mode network on the working memory task is much larger than that on the language processing task. The medial prefrontal cortex has relatively strong co-activation interactions with other regions on all three tasks. On the social cognition task, the left angular gyrus has weak co-activation interactions with other regions compared to the posterior cingulate cortex and the precuneus.

*Testing independence between regions across different modalities:* In the testing problem (c), we consider to test the independence of brain activity between the 90 AAL regions across

the working memory (2back-0back) task and the language processing (story-math) task. The global test, in which  $d = 8100$ , rejected the null hypothesis of the two task fMRI activations being independent for across all 90 regions under the significance level  $\alpha = 0.05$  with  $p$ -values  $< 0.0001$  for  $L = 1, 5, 10$ , and 15. To identify brain regions where the working memory activity ( $j = 1$ ) and the language processing activity ( $j' = 2$ ) are statistically significantly correlated with each other, we consider the multiple tests for null hypotheses  $H_{0,g,g'}: Y_{1,g}$  and  $Y_{2,g'}$  are independent. Due to  $g, g' = 1, \dots, 90$ , there are  $90 \times 90 = 8100$  null hypotheses. The multiple tests were conducted with the FDR controlled at 0.001 and  $L = 1$ . Each test was rejected if the corresponding  $p$ -value  $\leq 1.10 \times 10^{-5}$ . Our testing results indicate that the two types of task fMRI activations are significantly dependent with each other in 14 brain regions ( $p$ -value  $< 0.0001$ ) and across 23 different region pairs ( $p$ -value  $< 0.0001$ ). The results imply some new findings for the multi task fMRI activity. For example, the inferior frontal gyrus is well known for language comprehension and production (Tyler et al. 2011), but also related to working memory (Nissim et al. 2017). We found that working memory activity is strongly associated with the language processing activity in the left opercular part and the left triangular part of the inferior frontal gyrus. In addition, working memory activity in the left triangular part of the inferior frontal gyrus is strongly associated with the language processing activity in the right triangular part of inferior frontal gyrus. Figure 2 shows the whole bipartite graph indicating the dependence in brain activation between the working memory task (2back-0back) and the language processing (story-math) task.

## 5. Discussion

In this article, we propose a novel method for testing independence structure among the components of the ultra high-dimensional random vector. For a general form of null hypothesis, we develop a global testing procedure and a multiple testing procedure with FDR control using Gaussian approximations. To substantially improve the computational efficiency, we develop a distributed algorithm that is scalable to ultra high-dimensional data. We perform rigorous theoretical analyses and extensive simulations to demonstrate that the tests have good properties for a wide range of dimensions and sample sizes. In the analysis of task fMRI data in the HCP study, we apply the proposed method to test three types of complex dependences for five different contrast maps at group level. We obtain significant

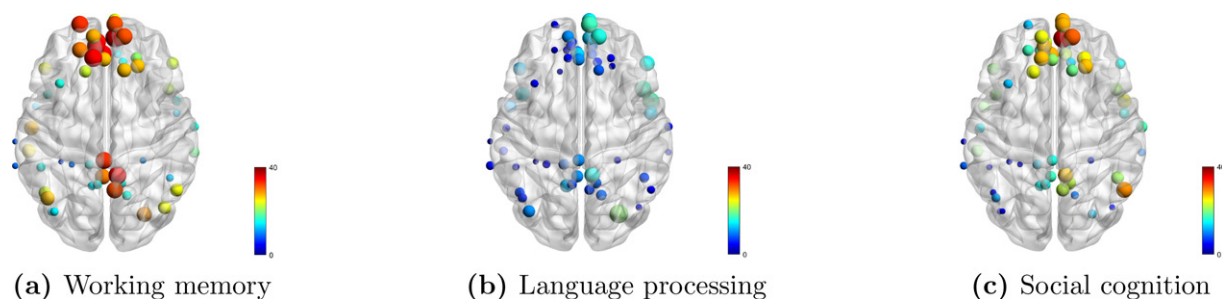
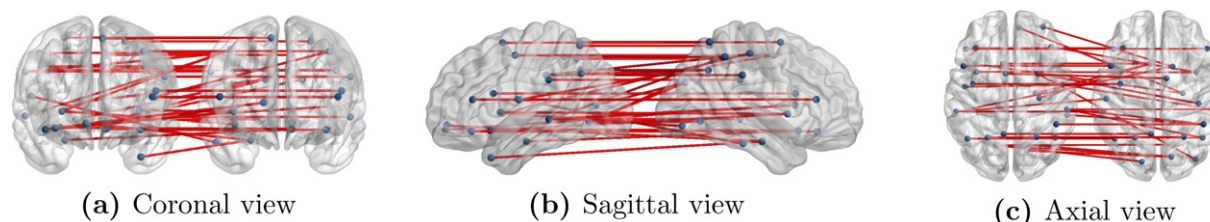


Figure 1. The degrees of nodes (regions) in the default mode network. The sizes of the nodes are proportional to the degrees of nodes.



**Figure 2.** Bipartite graph for the testing result on the independence between regions across the working memory (2back-0back) task (left) and the language processing (story-math) task (right). The edge between two nodes represents significantly correlation and the corresponding  $p$ -values are all  $< 0.0001$ .

evidences for some scientifically meaningful findings, providing insights on associations among brain activities in different regions under different cognitive tasks. We highlight several advantages of our methods. Compared to existing tests that may be sensitive to the dependence structure of data, our proposed test is particularly powerful for analysis of multimodal imaging data which appear to be noisy, spatially dependent, high-dimensional and complexly associated with anatomical structures and modalities. The scalable computation algorithm is another key contribution of the proposed test compared to other methods. To the best of our knowledge, we are among the first to propose the computational feasible independence tests with theoretical guarantees on high-dimensional voxel-level data ( $\approx 200,000$  voxels) in neuroimaging studies. In addition to imaging applications, the proposed test can be considered as a general tool for testing independence for many other statistical problems. For example, a multi-omics study may collect SNP genotypes, gene expressions, DNA copy number alternations, DNA methylation changes, along with other genetic information. It is also of great interest to test the independence among multimodal genomic data. Our proposed method provides a useful tool to address this problem.

## Supplementary Materials

The supplementary material contains additional simulation results, additional details in analysis of HCP data and proofs of the main results in the article.

## Acknowledgments

We are grateful to the editor, the associate editor and three referees for their constructive comments and suggestions.

## Funding

Chang, He and Wu were supported in part by the National Natural Science Foundation of China (grant nos. 71991472, 72125008, 11871401, and 11701466). Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics. Kang was supported in part by NIH R01DA048993, NIH R01MH105561, NIH R01GM124061, and NSF IIS2123777.

## ORCID

Jinyuan Chang <http://orcid.org/0000-0001-7933-4449>  
Jian Kang <http://orcid.org/0000-0002-5643-2668>

## References

- Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009), "Corrections to LRT on Large-Dimensional Covariance Matrix by RMT," *Annals of Statistics*, 37, 3822–3840. [1487]
- Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., et al. (2013), "Function in the Human Connectome: Task-fMRI and Individual Differences in Behavior," *NeuroImage*, 80, 169–189. [1486,1496]
- Bergsma, W., and Dassios, A. (2014), "A Consistent Test of Independence based on a Sign Covariance Related to Kendall's Tau," *Bernoulli*, 20, 1006–1028. [1487]
- Berrett, T. B., Kontoyiannis, I., and Samworth, R. J. (2021), "Optimal Rates for Independence Testing via U-statistic Permutation Tests," *Annals of Statistics*, 49, 2457–2490. [1487]
- Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., et al. (2011), "Mapping Anterior Temporal Lobe Language Areas with fMRI: A Multicenter Normative Study," *Neuroimage*, 54, 1465–1475. [1486]
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961), "Distribution Free Tests of Independence based on the Sample Distribution Function," *Annals of Mathematical Statistics*, 32, 485–498. [1487]
- Cai, T. T., Liu, W., and Xia, Y. (2014), "Two-Sample Test of High Dimensional Means under Dependence," *Journal of the Royal Statistical Society, Series B*, 76, 349–372. [1491]
- Calhoun, V. D., Adali, T., Kiehl, K. A., Astur, R., Pekar, J. J., and Pearlson, G. D. (2006), "A Method for Multitask fMRI Data Fusion Applied to Schizophrenia," *Human Brain Mapping*, 27, 598–610. [1486]
- Castelli, F., Happé, F., Frith, U., and Frith, C. (2000), "Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns," *Neuroimage*, 12, 314–325. [1486]
- Chakraborty, S., and Zhang, X. (2019), "Distance Metrics for Measuring Joint Dependence with Application to Causal Inference," *Journal of the American Statistical Association*, 114, 1638–1650. [1487,1495]
- (2021), "A New Framework for Distance and Kernel-based Metrics in High Dimensions," *Electronic Journal of Statistics*, 15, 5455–5522. [1487]
- Chang, J., Chen, X., and Wu, M. (in press), "Central Limit Theorems for High Dimensional Dependent Data," *Bernoulli*. [1488,1491]
- Chang, J., Shao, Q.-M., and Zhou, W.-X. (2016), "Cramér-Type Moderate Deviations for Studentized Two-Sample  $u$ -statistics with Applications," *Annals of Statistics*, 44, 1931–1956. [1492]
- Chang, J., Zheng, C., Zhou, W.-X., and Zhou, W. (2017), "Simulation-based Hypothesis Testing of High Dimensional Means under Covariance Heterogeneity," *Biometrics*, 73, 1300–1310. [1491]
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High Dimensional Random Vectors," *Annals of Statistics*, 41, 2786–2819. [1488,1491]
- Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2022), "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics*, 50, 2562–2586. [1488]
- Deb, N., and Sen, B. (2023), "Multivariate Rank-based Distribution-Free Nonparametric Testing Using Measure Transportation," *Journal of the American Statistical Association*, 118, 192–207. [1487]

- Deng, H., and Zhang, C.-H. (2020), "Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors," *Annals of Statistics*, 48, 3643–3671. [1488]
- Drobyshevsky, A., Baumann, S. B., and Schneider, W. (2006), "A Rapid fMRI Task Battery for Mapping of Visual, Motor, Cognitive, and Emotional Function," *Neuroimage*, 31, 732–744. [1486]
- Fan, J., Shao, Q.-M., and Zhou, W.-X. (2018), "Are Discoveries Spurious? Distributions of Maximum Spurious Correlations and their Applications," *Annals of Statistics*, 46, 989–1017. [1490]
- Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (2011), *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, London: Academic Press. [1486]
- Gao, L., Fan, Y., Lv, J., and Shao, Q.-M. (2021), "Asymptotic Distributions of High-Dimensional Distance Correlation Inference," *Annals of Statistics*, 49, 1999–2020. [1487]
- Gieser, P. W., and Randles, R. H. (1997), "A Nonparametric Test of Independence between Two Vectors," *Journal of the American Statistical Association*, 92, 561–567. [1487]
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013), "The Minimal Preprocessing Pipelines for the Human Connectome Project," *Neuroimage*, 80, 105–124. [1495]
- Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008), "A Kernel Statistical Test of Independence," in *Advances in Neural Information Processing Systems*, eds. J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, pp. 585–592, Red Hook, NY: Curran Associates, Inc. [1487]
- Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2021), "Distribution and Quantile Functions, Ranks and Signs in Dimension  $d$ : A Measure Transportation Approach," *Annals of Statistics*, 49, 1139–1165. [1487]
- Han, F., Chen, S., and Liu, H. (2017), "Distribution-Free Tests of Independence in High Dimensions," *Biometrika*, 104, 813–828. [1487]
- Heller, R., Heller, Y., and Gorfine, M. (2013), "A Consistent Multivariate Test of Association based on Ranks of Distances," *Biometrika*, 100, 503–510. [1487]
- Hotelling, H. (1936), "Relations between Two Sets of Variates," *Biometrika*, 28, 321–377. [1487]
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012), "FSL," *Neuroimage*, 62, 782–790. [1496]
- Jin, Z., and Matteson, D. S. (2018), "Generalizing Distance Covariance to Measure and Test Multivariate Mutual Dependence via Complete and Incomplete V-Statistics," *Journal of Multivariate Analysis*, 168, 304–322. [1487,1495]
- Kendall, M. G. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30, 81–93. [1487]
- Ledoit, O., and Wolf, M. (2002), "Some Hypothesis Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size," *Annals of Statistics*, 30, 1081–1102. [1487]
- Leung, D., and Drton, M. (2018), "Testing Independence in High Dimensions with Sums of Rank Correlations," *Annals of Statistics*, 46, 280–307. [1487]
- Liu, S., Cai, W., Liu, S., Zhang, F., Fulham, M., Feng, D., et al. (2015), "Multimodal Neuroimaging Computing: A Review of the Applications in Neuropsychiatric Disorders," *Brain Informatics*, 2, 167–180. [1486]
- Liu, W. (2013), "Gaussian Graphical Model Estimation with False Discovery Rate Control," *Annals of Statistics*, 41, 2948–2978. [1492]
- Lopes, M. E. (2022), "Central Limit Theorem and Bootstrap Approximation in High Dimensions: Near  $1/\sqrt{n}$  Rates via Implicit Smoothing," *Annals of Statistics*, 50, 2492–2513. [1488]
- Nagao, H. (1973), "On Some Test Criteria for Covariance Matrix," *Annals of Statistics*, 1, 700–709. [1487]
- Naylor, M. G., Cardenas, V. A., Tosun, D., Schuff, N., Weiner, M., and Schwartzman, A. (2014), "Voxelwise Multivariate Analysis of Multimodality Magnetic Resonance Imaging," *Human Brain Mapping*, 35, 831–846. [1487]
- Nissim, N. R., O'Shea, A. M., Bryant, V., Porges, E. C., Cohen, R., and Woods, A. J. (2017), "Frontal Structural Neural Correlates of Working Memory Performance in Older Adults," *Frontiers in Aging Neuroscience*, 8. [1497]
- Pearson, K. (1920), "Notes on the History of Correlation," *Biometrika*, 13, 25–45. [1487]
- Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. (2018), "Kernel-based Tests for Joint Independence," *Journal of the Royal Statistical Society, Series B*, 80, 5–31. [1487,1495]
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011), "Functional Network Organization of the Human Brain," *Neuron*, 72, 665–678. [1497]
- Raichle, M. E. (2015), "The Brain's Default Mode Network," *Annual Review of Neuroscience*, 38, 433–447. [1497]
- Ramezani, M., Marble, K., Trang, H., Johnsrude, I. S., and Abolmaesumi, P. (2014), "Joint Sparse Representation of Brain Activity Patterns in Multi-Task fMRI Data," *IEEE Transactions on Medical Imaging*, 34, 2–12. [1486]
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011), "Detecting Novel Associations in Large Data Sets," *Science*, 334, 1518–1524. [1487]
- Schott, J. R. (2005), "Testing for Complete Independence in High Dimensions," *Biometrika*, 92, 951–956. [1487]
- Shi, H., Drton, M., and Han, F. (2022), "Distribution-Free Consistent Independence Tests via Center-Outward Ranks and Signs," *Journal of the American Statistical Association*, 117, 395–410. [1487]
- Spearman, C. (1904), "The Proof and Measurement of Association between Two Things," *American Journal of Psychology*, 15, 72–101. [1487]
- Spreng, R. N. (2012), "The Fallacy of a 'Task-Negative' Network," *Frontiers in Psychology*, 3. [1497]
- Székel, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *Annals of Statistics*, 35, 2769–2794. [1487]
- Taskinen, S., Oja, H., and Randles, R. H. (2005), "Multivariate Nonparametric Tests of Independence," *Journal of the American Statistical Association*, 100, 916–925. [1487]
- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., et al. (2011), "Left Inferior Frontal Cortex and Syntax: Function, Structure and Behaviour in Patients with Left Hemisphere Damage," *Brain*, 134, 415–431. [1497]
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002), "Automated Anatomical Labeling of Activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *Neuroimage*, 15, 273–289. [1489]
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., and Ugurbil, K. (2013), "The WU-Minn Human Connectome Project: An Overview," *Neuroimage*, 80, 62–79. [1495]
- Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E., Bucholz, R., et al. (2012), "The Human Connectome Project: A Data Acquisition Perspective," *Neuroimage*, 62, 2222–2231. [1486]
- Wilks, S. S. (1935), "On the Independence of  $k$  Sets of Normally Distributed Statistical Variables," *Econometrica*, 3, 309–326. [1487]
- Xia, Y., Li, L., Lockhart, S. N., and Jagust, W. J. (2020), "Simultaneous Covariance Inference for Multimodal Integrative Analysis," *Journal of the American Statistical Association*, 115, 1279–1291. [1486,1487,1492]
- Yao, S., Zhang, X., and Shao, X. (2018), "Testing Mutual Independence in High Dimension via Distance Covariance," *Journal of the Royal Statistical Society, Series B*, 80, 455–480. [1487]
- Zhang, X. (2015), "Testing High Dimensional Mean under Sparsity," arXiv:1509.08444. [1490]
- Zhu, C., Zhang, X., Yao, S., and Shao, X. (2020), "Distance-based and RKHS-based Dependence Metrics in High Dimension," *Annals of Statistics*, 48, 3366–3394. [1487]
- Zhu, L., Xu, K., Li, R., and Zhong, W. (2017), "Projection Correlation between Two Random Vectors," *Biometrika*, 104, 829–843. [1487,1488,1489]