Journal of Multivariate Analysis 156 (2017) 70–88

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

High-dimensional tests for functional networks of brain anatomic regions

Jichun Xie [a,*], Jian Kang [b]

[a] Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27705, United States
[b] Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:
Received 23 March 2015
Available online 7 February 2017

Keywords:
High dimensionality
Hypothesis testing
Brain network
Sparsity
fMRI study

ABSTRACT

Exploring resting-state brain functional connectivity of autism spectrum disorders (ASD) using functional magnetic resonance imaging (fMRI) data has become a popular topic over the past few years. The data in a standard brain template consist of over 170,000 voxel specific points in time for each human subject. Such an ultra-high dimensionality makes the voxel-level functional connectivity analysis (involving four billion voxel pairs) both statistically and computationally inefficient. In this work, we introduce a new framework to identify the functional brain network at the anatomical region level for each individual. We propose two pairwise tests to detect region dependence, and one multiple testing procedure to identify global structures of the network. The limiting null distribution of each test statistic is derived. It is also shown that the tests are rate optimal when the alternative block networks are sparse. The numerical studies show that the proposed tests are valid and powerful. We apply our method to a resting-state fMRI study on autism and identify patient-unique and control-unique hub regions. These findings are biologically meaningful and consistent with the existing literature.

1. Introduction

The functional brain network refers to the coherence of the brain activities among multiple spatially distinct brain regions. It plays an important role in information processing and mental representations [12,44], and could be altered by the status of one's disease. Recent work [18,31,46] showed that patients with neurological diseases (such as Alzheimer's disease and the Autism Spectrum Disorder) have different functional networks compared to the controls. As a result, reliable and efficient inference on functional brain networks will benefit the study of these diseases. The goal of this research is to infer the functional networks of the brain region using neuro-imaging data.

Recent advances in neuro-imaging technologies provide great opportunities for researchers to study functional brain networks based on massive neuro-imaging data, which are generated using various imaging modalities such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG). In a neuro-imaging experiment, the scanner records the brain signals at multiple points in time at each location (or voxel) in the three-dimensional brain, leading to a four-dimensional imaging data structure. In a typical fMRI study, the number of voxels can be up to 200,000 and the number of imaging scans over time ranges from 100 to 1000. In the light of the brain function and the neuroanatomy, the human brain can be partitioned into about 100–200 anatomical regions and each region contains

* Corresponding author.
E-mail addresses: jichun.xie@duke.edu (J. Xie), jiankang@umich.edu (J. Kang).

http://dx.doi.org/10.1016/j.jmva.2017.01.011
0047-259X/

approximately 200–4000 voxels. Such high dimensionality and complexity of the data pose great challenges on the inference of the whole brain network.

Due to the ultra-high dimensionality of voxel numbers (up to 200,000), direct inference on the network of voxels is extremely resource intensive and computationally expensive. More importantly, the primary interest focuses on the network inference at the region level rather than at the voxel level. To this end, [3] examines the functional connectivity of a particular brain region, called seed region, by correlating the seed region brain signals against the brain signals from all other regions. Although this method yields a clear view of the functional connectivities between one region of interest (the seed region) and other regions [10,19], it fails to examine the functional network on a whole brain scale. Alternatively, [50] proposed to form meshes around a seed voxel by regressing p functionally nearest neighbor voxels on the seed voxel, where the number of regressors p is determined by minimizing Akaike's final prediction error [1]. Then two voxels are considered as functionally connected if one serves as a functional predictor for the other. The number of all connected voxel pairs between two anatomic regions is treated as the dependence measure between these two regions. Although this method successfully provides a functional network among anatomic regions, no inference results are provided on what level of connectivities should be regarded as significant. Another commonly used method [28,29] is to summarize one statistic (such as the largest principal component of voxel signals) in each region and then study the dependence between these statistics. Commonly used measures of dependence include the covariance matrix or the Gaussian Graphical model; see [29,37,46,51]. Since only one statistic is summarized in each region, the dependence among these summarized statistics sometimes fails to characterize the functional connectivities between brain regions.

In this article, we propose a new method to estimate the region level functional connectivity for each individual. Instead of summarizing one statistic in each region, we utilize multiple statistics such that the region wise brain connectivity information can be adequately captured. These statistics can be viewed as functional components of the region. The correlation matrix between the components in two regions is used to measure the dependence between these two regions. We assume that two regions are functionally connected if and only if at least one pair of components is correlated between these two regions.

We then concatenate these functional components region by region. No region level functional connectivity implies that the covariance matrix (or equivalently its inverse) of the concatenated components has a block-diagonal structure. Numerous examples in the existing literature confirm this is a reasonable assumption; see [11,29,42]. Thus, to construct a functional network of brain anatomic regions, we test whether the correlation matrix of two regions has a block diagonal structure.

Previous literature for testing a high-dimensional covariance/correlation matrix include testing whether the covariance matrix is proportional to the identity matrix [9,15,17,32,34,43], and testing whether two covariance or correlation matrices are equal [13,14,33,34]. To the best of our knowledge, no existing methods have been published to address whether high-dimensional covariance matrix has block-diagonal structure. However, ideas from published work can be borrowed to construct test statistics for our problem. There are mainly two types of existing test statistics: one is a chi-square statistic which is based on the sum of squared sample covariances, and the other is the extreme statistic based on the largest absolute self-standardized sample covariance. In general, the chi-square statistical test performs better when the alternative network is dense and the extreme statistical test performs better when the alternative network is sparse. In imaging studies, the network of functional components is usually sparse. Therefore, we will use an extreme type of statistical test.

This paper is organized as follows. In Section 2, we introduce the notation and define the testing hypotheses. Section 3 presents two procedures to control type I error for each hypothesis and a multiple testing procedure to control the family-wise error rate. We will discuss their theoretical properties in Section 4. One of the procedures involves estimating sparse regression coefficients. The estimators are discussed in Section 5. We will demonstrate the numerical performances of our procedures in Section 6. In Section 7, we apply the proposed procedures on resting-state fMRI data from subjects with and without autism spectrum disorder (ASD), and compare the functional networks of anatomic regions between cases and controls. The neuro-imaging results reflect the clinical characteristics of ASD. These methods are further discussed in Section 8.

2. Model and hypotheses

In fMRI studies, blood-oxygen-level dependent (BOLD) signals are collected at a large number of voxel locations for n scans. The standard preprocessing steps applied to BOLD signals include motion correction, slice-timing correction, normalization, and de-trending or de-meaning procedures [23,35,52], and then the signals are clustered based on their voxel locations mapping to the existing anatomic regions. After clustering, the signals are summarized into functional components to reduce the dimension of voxels and eliminate the redundancy of high coherent signals. One way to summarize the functional components is to perform principal component analysis (PCA) in region s to extract the first q_s principal components. Alternatively, independent component analysis (ICA) can be performed to extract q_s independent components. The choice of summarizing method depends on the distribution of the processed signals; see [2,41].

For each patient, assume that q_s functional components are summarized in region s. Each functional component is of length n, containing replications of signals across n scans. Let $X_{k,s,i}$ be the kth scan of the ith component in the sth brain region, after removing the temporal-correlation between the scans. These components can be treated as independent across scans.

Denote by $\boldsymbol{X}_{k,s} = (X_{k,s,1}, \ldots, X_{k,s,q_s})^\top$ the vector of functional components in region s of scan k, and by $\boldsymbol{\Upsilon}_{st} = \mathrm{corr}(\boldsymbol{X}_{k,s}, \boldsymbol{X}_{k,t})$ the correlation matrix between region s and region t. To test whether region s and region t are functionally connected, we set up the hypotheses:

$$\mathcal{H}_{0,st} : \boldsymbol{\Upsilon}_{st} = \mathbf{0} \quad \text{vs.} \quad \mathcal{H}_{1,st} : \boldsymbol{\Upsilon}_{st} \neq \mathbf{0}. \tag{1}$$

A rejection of $\mathcal{H}_{0,st}$ implies that regions s and region t have significant functional connectivity. The goal is to test $\mathcal{H}_{0,st}$ with controlled type I error, and also to perform multiple testing on $\mathcal{H}_{0,st}$ simultaneously to control the family-wise error rate.

The difficulty of this testing problem lies in the large number of parameters and relatively small number of replications. First, the number of summarized functional components in each region may increase with the number n of scans. Second, the number of total region pairs $p(p-1)/2$ usually exceeds n. Therefore, we need to address the high-dimensional challenges in testing a large number of them simultaneously.

3. Testing procedures

To test $\mathcal{H}_{0,st}$, we propose two procedures to fit different distribution assumptions on the functional components. Therefore, neither of them can universally outperform the other. We further develop a multiple testing procedure to control the family-wise error rate (FWER) for testing $\{\mathcal{H}_{0,st} : 1 \leq s < t \leq p\}$ simultaneously.

3.1. Test I: Marginal dependence testing

The first procedure is based on the Pearson correlation between the components in two regions. Denote by the pairwise correlation $\rho_{st,ij} = \mathrm{corr}(X_{k,s,i}, X_{k,t,j})$. Then the null hypothesis $\mathcal{H}_{st,0} : \boldsymbol{\Upsilon}_{st} = \mathbf{0}$ is equivalent to $\mathcal{H}_{st,0} : \max_{1\leq i\leq q_s, 1\leq j\leq q_t} |\rho_{st,ij}| = 0$. A straightforward approach is to check whether the sample correlation between two regions is close to zero. Denote the Pearson correlation between the ith component in region s and the jth component in region t by $\hat{\rho}_{st,ij}$, i.e.,

$$\hat{\rho}_{st,ij} = \hat{\sigma}_{st,ij} / \left(\hat{\sigma}_{ss,ii}\hat{\sigma}_{tt,jj}\right)^{1/2},$$

where

$$\bar{X}_{s,i} = \frac{1}{n}\sum_{k=1}^{n} X_{k,s,i}, \quad \bar{X}_{t,j} = \frac{1}{n}\sum_{k=1}^{n} X_{k,t,j}, \quad \hat{\sigma}_{st,ij} = \frac{1}{n}\sum_{k=1}^{n}(X_{k,s,i} - \bar{X}_{s,i})(X_{k,t,j} - \bar{X}_{t,j}).$$

The test statistic is defined as

$$T_{st}^{(1)} = n \max_{i,j} \hat{\rho}_{st,ij}^2 - 2\ln(q_s q_t) + \ln\{\ln(q_s q_t)\}. \tag{2}$$

With mild conditions (details in Section 4), under $\mathcal{H}_{0,st}$, $T_{st}^{(1)}$ asymptotically follows the Gumbel distribution

$$F(x) = \exp\{-\pi^{1/2}\exp(-x/2)\}. \tag{3}$$

To control type I error at level α, we reject $\mathcal{H}_{0,st}$ if $T_{st}^{(1)}$ exceeds the $(1-\alpha)$th quantile of $F(x)$, i.e., $T_{st}^{(1)} > q_\alpha$, with

$$q_\alpha = -\ln(\pi) - 2\ln[\ln\{1/(1-\alpha)\}]. \tag{4}$$

We would like to point out the intrinsic differences between our test statistic and the statistics proposed by [14] and [13]. First, [14] proposes to test whether two covariance matrices are equal, and [13] proposes to test whether two correlation matrices are equal, and whether a correlation matrix is equal to the identity matrix. These two papers address different problems, and therefore their methods cannot be applied directly to our paper. Second, both [14] and [13] used self-standardized sample covariance (or correlation) as the test statistics, while our paper standardizes the sample covariance by its null variance. Our statistic not only has optimal theoretical properties under weaker conditions, but also performs better numerically.

3.2. Test II: Local conditional dependence testing

The alternative testing procedure is based on the Pearson correlation between the residuals of local neighborhood selection in two regions. In region s, we regress other $q_s - 1$ variables on $X_{k,s,i}$,

$$X_{k,s,i} = \alpha_{s,i} + \boldsymbol{X}_{k,s,-i}^\top \boldsymbol{\beta}_{s,i} + \varepsilon_{k,s,i}, \tag{5}$$

where $\boldsymbol{X}_{k,s,-i}$ is the vector of $\boldsymbol{X}_{k,s}$ by removing the ith component. In region t with $t \neq s$, we build up a similar regression model, viz.

$$X_{k,t,j} = \alpha_{t,j} + \boldsymbol{X}_{k,t,-j}^\top \boldsymbol{\beta}_{t,j} + \varepsilon_{k,t,l}. \tag{6}$$

Let $\rho_{\varepsilon,st,ij} = \mathrm{corr}(\varepsilon_{k,s,i}, \varepsilon_{k,t,j})$ be the correlation of the error terms in two models. Clearly, the null hypothesis $\mathcal{H}_{0,st}$ is equivalent to

$$\mathcal{H}_{0,st}: \max_{i,j} \rho_{\varepsilon,st,ij} = 0.$$

We therefore develop a testing procedure to test if the correlations $\rho_{\varepsilon,st,ij}$ are all zero. If the coefficients $\boldsymbol{\beta}_{s,i}$ and $\boldsymbol{\beta}_{t,j}$ in models (5) and (6) were known, we would know the value of each realization of the errors $\varepsilon_{k,s,i}$ and $\varepsilon_{k,t,j}$, and center them as $\tilde{\varepsilon}_{k,v,\ell} = \varepsilon_{k,v,\ell} - \bar{\varepsilon}_{v,\ell}$ with $\bar{\varepsilon}_{v,\ell} = \sum_{k=1}^{n} \varepsilon_{k,v,\ell}/n$, $(v,\ell) = (s,i)$ or $(v,\ell) = (t,j)$. Based on models (5) and (6), the centered realization of the error $\tilde{\varepsilon}_{k,v,\ell}$ could be expressed as

$$\tilde{\varepsilon}_{k,v,\ell} = (X_{k,v,\ell} - \bar{X}_{v,\ell}) - (\boldsymbol{X}_{k,v,-l} - \bar{\boldsymbol{X}}_{v,-l})^{\top} \boldsymbol{\beta}_{v,\ell}, \quad (v,\ell) = (s,i) \text{ or } (v,\ell) = (t,j). \tag{7}$$

Consequently, the Pearson correlation between $\tilde{\varepsilon}_{k,s,i}$ and $\tilde{\varepsilon}_{k,t,j}$ would be

$$\tilde{\rho}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \frac{\tilde{\sigma}_{\varepsilon,st,ij}}{\left(\tilde{\sigma}_{\varepsilon,ss,ii} \tilde{\sigma}_{\varepsilon,tt,jj}\right)^{1/2}},$$

where

$$\tilde{\sigma}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\varepsilon}_{k,s,i} \tilde{\varepsilon}_{k,t,j}, \quad \tilde{\sigma}_{\varepsilon,ss,ii} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\varepsilon}_{k,s,i}^2, \quad \tilde{\sigma}_{\varepsilon,tt,jj} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\varepsilon}_{k,t,j}^2.$$

Unfortunately in practice, the coefficients in (5) and (6) are unknown. However, the coefficients can be well estimated by existing methods, such as Lasso or Dantzig selector. Suppose "good" coefficient estimators $\hat{\boldsymbol{\beta}}_{s,i}$ and $\hat{\boldsymbol{\beta}}_{t,j}$ exist. (This notion is made more precise in Section 4, where we also address how to obtain "good" coefficient estimators.) Then the centered error term $\tilde{\varepsilon}_{k,v,\ell}$ can be estimated by

$$\hat{\varepsilon}_{k,v,\ell} = (X_{k,v,\ell} - \bar{X}_{v,\ell}) - (\boldsymbol{X}_{k,v,-l} - \bar{\boldsymbol{X}}_{v,-l})^{\top} \hat{\boldsymbol{\beta}}_{v,\ell}, \quad (v,\ell) = (s,i) \text{ or } (v,\ell) = (t,j). \tag{8}$$

Consequently, we calculate the Pearson correlation based on $\hat{\varepsilon}_{k,s,i}$ and $\hat{\varepsilon}_{k,t,j}$,

$$\hat{\rho}_{\varepsilon,st,ij} = \hat{\sigma}_{\varepsilon,st,ij} / \left(\hat{\sigma}_{\varepsilon,ss,ii} \hat{\sigma}_{tt,jj}\right)^{1/2},$$

where

$$\hat{\sigma}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \hat{\varepsilon}_{k,s,i} \hat{\varepsilon}_{k,t,j}, \quad \hat{\sigma}_{\varepsilon,ss,ii} = \frac{1}{n} \sum_{k=1}^{n} \hat{\varepsilon}_{k,s,i}^2, \quad \hat{\sigma}_{\varepsilon,tt,jj} = \frac{1}{n} \sum_{k=1}^{n} \hat{\varepsilon}_{k,t,j}^2.$$

As for Test I, we construct the test statistic as follows:

$$T_{st}^{(2)} = n \max_{i,j} \hat{\rho}_{\varepsilon,st,ij}^2 - 2\ln(q_s q_t) + \ln\ln(q_s q_t).$$

Under certain conditions (discussed in Section 4) and $\mathcal{H}_{0,st}$, $T_{st}^{(2)}$ also follows the Gumbel distribution defined in (3). Therefore, to control type I error at level α, we reject $\mathcal{H}_{0,st}$ if $T_{st}^{(2)} > q_\alpha$, where q_α is the $(1-\alpha)$th quantile of F.

3.3. Family-wise error rate control

Considering the standard space of the brain [39] and the commonly used brain atlas (the Automated Anatomical Labeling [49] regions), the number of region pairs in the whole brain is over 4000, which is much larger than the number of scans (typically of the order of 200). This provides motivation to develop a solution to correct for multiplicity when constructing the functional connectivity network of the whole brain. We propose procedure (9) to test $\{\mathcal{H}_{0,st}: 1 \le s < t \le p\}$ simultaneously and control the family-wise error rate (FWER). The procedure can involve either $\widetilde{T}_{st}^{(1)}$ or $\widetilde{T}_{st}^{(2)}$, depending on the assumption of the dependence structure of local voxels. Our findings show that to control FWER at level α, we only need to adopt a higher threshold. The adjusted testing procedure is as follows:

$$\text{Reject } \mathcal{H}_{0,st} \quad \leftrightarrow \quad \forall_{1\le s<t\le p} \quad T_{st}^{(b)} > 2\ln\{p(p-1)/2\} + q_\alpha, \tag{9}$$

for $b = 1, 2$. The threshold depends on the desired FWER α, and the total number of region pairs $p(p-1)/2$.

4. Theory

In this section, we derive the null distributions of the statistics on which Tests I and II are based. We also examine the power and the optimality properties of the proposed tests. Furthermore, we prove that the multiple testing procedure (9) is able to control FWER.

For the remainder of the paper, unless otherwise stated, we use the following notation: for a vector $\mathbf{a} = (a_1, \ldots, a_p)^\top \in \mathbb{R}^p$, denote by $|\mathbf{a}|_2 = (\sum_{j=1}^p a_j^2)^{1/2}$ its Euclidean norm; for a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p\times q}$, define the spectral norm $\|\mathbf{A}\|_2 = \sum_{|\mathbf{x}|_2=1} |\mathbf{A}\mathbf{x}|_2$ and the Frobenius norm $\|\mathbf{A}\|_F = (\sum_{ij} a_{ij}^2)^{1/2}$; for a finite set $\mathcal{A} = \{a_1, \ldots, a_s\}$, $\mathrm{card}(\mathcal{A}) = s$ counts the number of elements in $\mathcal{A}$; for two real number sequences (a_n) and (b_n), write $a_n = O(b_n)$ if $|a_n| \le C|b_n|$ hold for a certain positive constant C when n is sufficiently large, write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$, and write $a_n \asymp b_n$ if $c|b_n| \le |a_n| \le C|b_n|$, for some positive constants c and C when n is sufficiently large.

Also assume the number of variables in all regions are comparable, i.e., $q_1 \asymp \cdots \asymp q_p$. Let $q_0 = \max(q_1, \ldots, q_p)$, and assume $\boldsymbol{X}_{1,v}, \ldots, \boldsymbol{X}_{n,v}$ are independently and identically distributed for each region v.

4.1. Asymptotic properties of Test I

Denote by $\Upsilon_{vv} = (\rho_{vv,ij})_{q_v\times q_v}$ the correlation matrix of $\boldsymbol{X}_{k,v}$. For $X_{k,v,i}$, denote by $r_{v,i}^{(1)}$ the number of other components in region v non-negligibly correlated with $X_{k,v,i}$, viz.

$$r_{v,i}^{(1)} = \mathrm{card}\{j : |\rho_{vv,ij}| \ge (\ln q_0)^{-1-\alpha_0},\ j \ne i\},$$

where α_0 is a positive constant. For a positive constant $\rho_0 < 1$, define

$$\mathcal{D}_v^{(1)} = \{i : |\rho_{vv,ij}| > \rho_0 \text{ for some } j \ne i\}.$$

Thus, $\mathcal{D}_v^{(1)}$ contains index i such that $X_{k,v,i}$ is highly correlated to at least one other component in region v.

We need the following conditions:

(C1.1) For region $v = s, t$, there exists a subset $\mathcal{M}_v \subset \{1, \ldots, q_v\}$ with $\mathrm{card}(\mathcal{M}_v) = o(q_v)$ and a constant $\alpha_0 > 0$ such that for all $\gamma > 0$, $\max_{i\in\mathcal{M}_v^c} r_{v,i}^{(1)} = o(q_v^\gamma)$. Moreover, assume there exists a constant $0 \le \rho_0 < 1$ such that $\mathrm{card}\{\mathcal{D}_v^{(1)}\} = o(q_v)$.

Condition (C1.1) constraints the sparsity level of non-negligible and large signals. It specifies that for each region v, for almost all components i within the region, the count of non-negligible $|\rho_{vv,ij}|$ is of a smaller order of q_v^γ. The condition is weaker than the commonly seen condition which imposes a constant upper bound on the largest eigenvalue of $\boldsymbol{\Sigma}_{vv}$. In fact, if $\lambda_{\max}(\boldsymbol{\Sigma}_{vv}) = o\{q_v^\gamma/(\ln q_0)^{1+\alpha_0}\}$, $\max_{1\le i\le q_v} r_{v,i}^{(1)} = o(q_v^\gamma)$. Additionally, Condition (C1.1) also requires the number of components that are very highly correlated with at least one other component to be small. This condition can be easily satisfied if all the correlations $\rho_{vv,ij}$ are bounded by ρ_0.

(C1.2) Sub-Gaussian type tails: For region $v = s, t$, suppose that $\ln(q_v) = o(n^{1/5})$. There exist some constants $\eta > 0$ and $K > 0$ such that

$$\max_{1\le i\le q_v} \mathrm{E}\left[\exp\{\eta(X_{k,v,i} - \mu_{v,i})^2/\sigma_{vv,ii}\}\right] \le K.$$

(C1.2*) Polynomial-type tails: For region $v = s, t$, suppose that for some $\gamma_1, c_1 > 0$, $q_0 \le c_1 n^{\gamma_1+1/2}$, and for some $\epsilon > 0$,

$$\max_{1\le i\le q_v} \mathrm{E}|(X_{k,v,i} - \mu_{xi})/\sigma_{vv,ii}^{1/2}|^{4\gamma_1+4+\epsilon} \le K.$$

Conditions (C1.2) and C(1.2*) impose constraints on the tail of the distribution of $X_{k,v,i}$, and the corresponding order of q_v. They fit a wide range of distributions. For example, Gaussian distributions satisfy Condition (C1.2), and Pareto distributions *Pareto*(α) (which are heavy tailed) with α sufficiently large satisfy Condition (C1.2*).

Theorem 1. *Suppose that Conditions* (C1.1) *and* (C1.2) *or* (C1.2*) *hold. Then under* $\mathcal{H}_{0,st}$, *as* $n, q_0 \to \infty$, *the distribution* $T_{st}^{(1)}$ *converges point-wise to the Gumbel distribution* F *defined in* (3).

When Condition (C1.1) is not satisfied, i.e., the correlation matrices Υ_{ss} and Υ_{tt} are arbitrary, it is difficult to derive the limiting null distribution of $T_{st}^{(1)}$. However, Test I can still control the type I error.

Proposition 1. *Under Condition* (C1.2) *or* (C1.2*) *and the null* $\mathcal{H}_{0,st}$, *for* $0 < \alpha < 1$, $\Pr(T_{st}^{(1)} \ge q_\alpha) \le \ln\{1/(1-\alpha)\}$, *where* q_α *is defined in* (4).

When the desired type I error α is small, $\ln\{1/(1-\alpha)\} \approx \alpha$. Therefore, Test I can still control type I error close to the desired level. When there comes a rare circumstance that a larger type I error is desired for the test, we can define $\alpha' = 1 - e^{-\alpha}$ and reject $\mathcal{H}_{0,st}$ when $\widetilde{T}_{st}^{(1)} \ge q_{\alpha'}$. Since $\alpha = \ln\{1/(1-\alpha')\}$, Test I is always an asymptotically valid test, for arbitrary correlation matrices Υ_{ss} and Υ_{tt}. However, the power will be reduced when we threshold $T_{st}^{(1)}$ at the higher level $q_{\alpha'}$.

We now turn to the power analysis of Test I. To test the correlation between region s and region t, we define the following class of correlation matrices:

$$\mathcal{U}_{st}^{(1)}(c) = \left\{\Upsilon_{st} : \max_{i,j} \rho_{st,ij}^2 \ge c\ \ln(d_{st}/n)\right\}.$$

We also set

$$\kappa_{st} = \sup_{1\le i\le q_s, 1\le j\le q_t} \frac{\sigma_{ss,ii}\sigma_{tt,jj}}{\theta_{st,ij}}.$$

We first show that κ_{st} is bounded. By Hölder's inequality,

$$\left[\mathrm{E}\{(X_{s,i}-\mu_{s,i})(X_{t,j}-\mu_{t,j})\}\right]^2 \le \sigma_{ss,ii}\sigma_{tt,jj}.$$

Therefore,

$$\mathrm{E}\{(X_{s,i}-\mu_{s,i})^2(X_{t,j}-\mu_{t,j})^2\} - \sigma_{ss,ii}\sigma_{tt,jj} \le \theta_{st,ij} \le \mathrm{E}\{(X_{s,i}-\mu_{s,i})^2(X_{t,j}-\mu_{t,j})^2\} + \sigma_{ss,ii}\sigma_{tt,jj}.$$

Because

$$\mathrm{E}\{(X_{s,i}-\mu_{s,i})^2(X_{t,j}-\mu_{t,j})^2\} \le \left[\mathrm{E}\{(X_{s,i}-\mu_{s,i})^4\}\mathrm{E}\{(X_{t,j}-\mu_{t,j})^4\}\right]^{1/2},$$

and by Condition (C1.2) (or (C1.2*)),

$$\mathrm{E}\{(X_{s,i}-\mu_{s,i})^4\} \le K\sigma_{ss,ii}^2, \quad \mathrm{E}\{(X_{t,j}-\mu_{t,j})^4\} \le K\sigma_{tt,jj}^2,$$

$$(K-1)\sigma_{ss,ii}\sigma_{tt,jj} \le \theta_{st,ij} \le (K+1)\sigma_{ss,ii}\sigma_{tt,jj}.$$

Thus $\kappa_{st} \le K+1$.

Theorem 2. *Suppose that Condition* (C1.2) *or* (C1.2*) *holds. Then as n and q_0 both go to infinity,*

$$\inf_{\Upsilon_{st}\in\mathcal{U}_{st}^{(1)}\{4(1+\kappa_{st})\}} \Pr(T_{st}^{(1)} > q_\alpha) \to 1.$$

To distinguish the alternative from the null, Test I requires only one entry in the correlation matrix Υ_{st} larger than $\{4(1+\kappa_{st})\ln(d_{st}/n)\}^{1/2}$. The rate is optimal in terms of the following minimax argument. Denote by $\mathcal{F}_{st}^{(1)}$ the collection of distributions satisfying (C1.2) or (C1.2*), and by $\mathcal{T}_{st,\alpha}^{(1)}$ the collection of all α-level tests over $\mathcal{F}_{st}^{(1)}$, i.e.,

$$\forall_{\Phi_{st,\alpha}\in\mathcal{T}_{st,\alpha}^{(1)}} \quad \Pr(\Phi_{st,\alpha}=1) \le \alpha.$$

Theorem 3. *Suppose Condition* (C1.2) *or* (C1.2*) *holds. Let α and β be any positive numbers with $\alpha+\beta<1$. There exists a positive constant c_0 such that for all large n and q_0,*

$$\inf_{\Upsilon_{st}\in\mathcal{U}_{st}^{(1)}(c_0)} \sup_{T_{st,\alpha}\in\mathcal{T}_{st,\alpha}^{(1)}} \Pr(T_{st,\alpha}=1) \le 1-\beta.$$

For any hypothesis testing, if the null is never rejected, the type I error $\alpha_0=0$ and the type II error $\beta_0=1$; if the null is always rejected, the type I error $\alpha_0=1$ and the type II error $\beta_0=0$. These are trivial testing methods, and for both tests we have $\alpha_0+\beta_0=1$. For non-trivial testing methods, we expect $\alpha_0+\beta_0<1$. Theorem 3 shows that, when $\max_{1\le i\le q_s,1\le j\le q_t}\rho_{st,ij}^2 \le c_0\ln d_{st}/n$ for some sufficiently small c_0, the sum of the type I error α_0 and the type II error β_0 cannot be bounded away from 1. Therefore, the lower bound order $\max_{1\le i\le q_s,1\le j\le q_t}\rho_{st,ij}^2 \ge c\ln d_{st}/n$ cannot be improved. On the other hand, Theorem 2 shows that when $\max_{1\le i\le q_s,1\le j\le q_t}\rho_{st,ij}^2 \ge 4(1+\kappa_{st})\ln(d_{st}/n)$, the power of Test I will approach 1. Therefore, Test I enjoys a certain optimality.

In Theorems 2 and 3, the difference between the null and the alternative is measured by $|\Upsilon_{st}|_\infty^2 = \max_{1\le i\le q_s,1\le j\le q_t}\rho_{ij,st}^2$. Another commonly used measure is the Frobenius norm $\|\Upsilon_{st}\|_F$. Denote by r_{st} the count of the nonzero entries in Υ_{st}, i.e.,

$$r_{st} = \sum_{i=1}^{q_s}\sum_{j=1}^{q_t}\mathbf{1}(\rho_{st,ij}\ne 0).$$

Consider the following class of matrices:

$$\mathcal{V}_{st}^{(1)}(c) = \left\{\Upsilon_{st} : \|\Upsilon_{st}\|_F^2 \ge cr_{st}\ln(d_{st}/n)\right\}.$$

We now show that Test I also enjoys the rate optimality property measured by Frobenius norm.

Corollary 1. *Suppose that Condition* (C1.2) *or* (C1.2*) *holds. Then for a sufficiently large c, as n and q_0 both go to infinity,*

$$\inf_{\Upsilon_{st}\in\mathcal{V}_{st}^{(1)}\{4(1+\kappa_{st})\}} \Pr(T_{st}^{(1)} > q_\alpha) \to 1.$$

If $\|\Upsilon_{st}\|_F^2 \ge 4(1+\kappa_{st})r_{st}\ln(d_{st})$, $\max_{1\le i\le q_s,1\le t\le q_t}\rho_{st,ij}^2 \ge 4(1+\kappa_{st})\ln(d_{st})$. Therefore, Corollary 1 can be derived from Theorem 2.

Theorem 4. *Suppose that Condition* (C1.2) *or* (C1.2*) *holds. Assume that* $r_{st} \le q_0^{\gamma_2}$ *for some* $0 < \gamma_2 < 1/2$. *Let* α, β *be any positive number with* $\alpha + \beta < 1$. *There exists a positive constant* c_0 *such that for all large n and* q_0,

$$\inf_{\Sigma_{st} \in \mathcal{V}_{st}^{(1)}(c_0)} \sup_{T_{st,\alpha} \in \mathcal{T}_{st,\alpha}^{(1)}} \Pr(\Phi_{st,\alpha} = 1) \le 1 - \beta.$$

In this theorem, we assume that $r_{st} \le q_0^{\gamma_2}$. The assumption is quite reasonable for the brain network, because if the connections of the functional components exist between two brain regions, they are usually sparse.

Theorem 4 shows that, when $\Upsilon_{st} \in \mathcal{V}_{st}^1(c_0)$ for some sufficiently small c_0, the type I error α_0 and type II error β_0 cannot be bounded away from 1. Therefore, the minimax rate $\|\Upsilon_{st}\|_F^2 \ge c r_{st} \ln d_{st}$ cannot be improved. Theorem 4 and Corollary 1 show that Test I also performs optimally over $\mathcal{V}_{st}^{(1)}(c)$.

4.2. Asymptotic properties of Test II

For Test II, the conditions required for achieving its asymptotic property are different from what is required for Test I.

Recall that $\varepsilon_{k,s,i}$ and $\varepsilon_{k,t,j}$ are the error term of regressing all other components on one component within the region, as defined in (5) and (6), and $\sigma_{\varepsilon,st,ij} = \mathrm{cov}(\varepsilon_{k,s,i}, \varepsilon_{k,t,j})$. Let $\Upsilon_{\varepsilon,st} = (\rho_{\varepsilon,st,ij})$ be the correlation matrix between $\boldsymbol{\varepsilon}_{k,s} = (\varepsilon_{k,s,1}, \ldots, \varepsilon_{k,s,q_s})^\top$ and $\boldsymbol{\varepsilon}_{k,t} = (\varepsilon_{k,t,1}, \ldots, \varepsilon_{k,t,q_t})^\top$. Then

$$\rho_{\varepsilon,st,ij} = \frac{\sigma_{\varepsilon,st,ij}}{(\sigma_{\varepsilon,ss,ii}\sigma_{\varepsilon,tt,jj})^{1/2}},$$

where $\sigma_{\varepsilon,st,ij} = \mathrm{cov}(\varepsilon_{k,s,i}, \varepsilon_{k,t,j})$, $\sigma_{\varepsilon,ss,ii} = \mathrm{var}(\varepsilon_{k,s,i})$ and $\sigma_{\varepsilon,tt,jj} = \mathrm{var}(\varepsilon_{k,t,j})$.

For $\varepsilon_{k,s,i}$, denote by $r_{v,i}^{(2)}$ the number of other $\varepsilon_{k,s,j}$ that are non-negligibly correlated ($>(\ln q_0)^{-1-\alpha_0}$) with it, viz.

$$r_{v,i}^{(2)} = \mathrm{card}\{j : |\rho_{\varepsilon,vv,ij}| \ge (\ln q_0)^{-1-\alpha_0},\ j \ne i\}.$$

For a positive constant $\rho_0 < 1$, define the set that $\varepsilon_{k,v,i}$ is highly correlated with at least one $\varepsilon_{k,v,j}$ as

$$\mathcal{D}_v^{(2)} = \{i : |\rho_{\varepsilon,vv,ij}| > \rho_0 \text{ for some } j \ne i\}.$$

We need the following conditions:

(C2.1) For regions $v = s, t$, there exists a subset $\mathcal{M}_v \in \{1, \ldots, q_v\}$ with $\mathrm{card}(\mathcal{M}_v) = o(q_v)$ and a constant $\alpha_0 > 0$ such that all $\gamma > 0$, $\max_{1 \le i \le p, i \in \mathcal{M}_v} r_{v,i}^{(2)} = o(q_v^\gamma)$. Moreover, assume there exists a constant $0 \le \rho_0 < 1$ such that $\mathrm{card}\{\mathcal{D}_v\} = o(q_0)$.

Condition (C2.1) parallels Condition (C1.1). It imposes conditions on within region correlation $\Upsilon_{\varepsilon,vv}$. Suppose $\boldsymbol{X}_{k,v}$ follows a multivariate Gaussian distribution with the inverse covariance matrix $\Omega_{vv} = (\omega_{vv,ij})$. Because $\rho_{\varepsilon,vv,ij} = \omega_{vv,ij}/(\omega_{vv,ii}\omega_{vv,jj})^{1/2}$ [2], Condition (C2.1) holds under many cases when the inverse covariance matrix is sparse and bounded; see [26,28,38]. Because covariance matrices and inverse covariance matrices are different, some data only satisfy one of these two conditions. Consequently, the corresponding procedure should be applied.

(C2.2) For region $v = s, t$, the variable $\boldsymbol{X}_{k,v} \sim \mathcal{N}(\boldsymbol{\mu}_v, \Sigma_{vv})$, with $\lambda_{\max}(\Sigma_{vv}) \le c_0$, where $\lambda_{\max}$ is the maximum eigenvalue operator. Also assume $\ln q_0 = o(n^{1/5})$.

In general, the theoretical properties of Test II hold for many non-Gaussian distributions as well. However, only under the Gaussian distribution assumption, $\rho_{\varepsilon,vv,ij}$ has an interpretation of conditional dependence such that

$$\rho_{\varepsilon,vv,ij} = 0 \quad \leftrightarrow \quad X_{k,v,i} \perp\!\!\!\perp X_{k,v,j} \mid \{X_{k,v,\ell} : \ell \ne i, j\}.$$

Condition (C2.2) makes Condition (C2.1) a natural assumption on the conditional dependency. Since $\sigma_{vv,ii} \le \lambda_{\max}(\Sigma_{vv})$ and $\sigma_{vv,ii}\omega_{vv,ii} \ge 1$, this condition also implies that $\mathrm{var}(\varepsilon_{k,s,i}) = 1/\omega_{vv,ii} \le c_0$.

(C2.3) Recall the definition of $\tilde{\varepsilon}_{k,v,\ell}$ and $\hat{\varepsilon}_{k,v,\ell}$ in (7) and (8). Under the cases (i) $s \ne t$ and (ii) $s = t$ and $i = j$, with probability tending to 1,

$$\max_{i,j} \left| \frac{1}{n}\sum_{k=1}^n \hat{\varepsilon}_{k,s,i}\hat{\varepsilon}_{k,t,j} - \frac{1}{n}\sum_{k=1}^n \tilde{\varepsilon}_{k,s,i}\tilde{\varepsilon}_{k,t,j} \right| \le C(\ln q_0)^{-1-\alpha_0}. \tag{10}$$

Note that $\hat{\varepsilon}_{k,v,i}$ is the centered residual and $\tilde{\varepsilon}_{k,v,i}$ is the centered random error. The term $|\sum_{k=1}^n \hat{\varepsilon}_{k,s,i}\hat{\varepsilon}_{k,t,j} - \sum_{k=1}^n \tilde{\varepsilon}_{k,s,i}\tilde{\varepsilon}_{k,t,j}|/n$ is determined by the difference between $\boldsymbol{\beta}_{v,i}$ and its estimator $\hat{\boldsymbol{\beta}}_{v,i}$. We will specify in Section 5 some estimation methods and corresponding sufficient conditions under which Condition (C2.3) will hold.

We derive the null distribution of $T_{st}^{(2)}$ in the next theorem.

Theorem 5. *Suppose that Conditions* (C2.1), (C2.2) *and* (C2.3) *hold. Then under* $\mathcal{H}_0$, *as* $n, q_0 \to \infty$, $T_{st}^{(2)}$ *converges point-wise to the Gumbel distribution F in* (3).

The derivation of the limiting null distribution of $T_{st}^{(2)}$ calls for Condition (C2.1). When it is not satisfied, we can still control type I error based on the following proposition.

Proposition 2. *Under Conditions* (C2.2) *and* (C2.3) *and the null* $\mathcal{H}_{0,st}$, $\Pr(T_{st}^{(2)} \geq q_\alpha) \leq \ln\{1/(1-\alpha)\}$, *where* $q_\alpha = -\ln(\pi) - 2\ln\ln\{1/(1-\alpha)\}$ *is the* $(1-\alpha)$*th quantile of* F *defined in* (3).

The power analysis of Test II parallels to that of Test I. Let

$$r_{\varepsilon,st} = \sum_{i=1}^{q_s}\sum_{j=1}^{q_t} \mathbf{1}(\rho_{\varepsilon,st,ij} \neq 0).$$

Define the following two classes of matrices:

$$\mathcal{U}_{st}^{(2)}(c) = \left\{ \Upsilon_{\varepsilon,st} : \max_{1\leq i\leq q_s, 1\leq j\leq q_t} \rho_{\varepsilon,st,ij}^2 \geq c\ln d_{st}/n \right\};$$
$$\mathcal{V}_{st}^{(2)}(c) = \left\{ \Upsilon_{\varepsilon,st} : \|\Upsilon_{\varepsilon,st}\|_F^2 \geq c r_{\varepsilon,st}\ln(d_{st}/n) \right\}.$$

We have the following theorem.

Theorem 6. *Suppose that Conditions* (C2.2) *and* (C2.3) *hold. Then*

$$\lim_{n,q_0\to\infty} \inf_{\mathbf{R}_{st}\in\mathcal{U}_{st}^{(2)}(c_1)} \Pr(T_{st}^{(2)} \geq q_\alpha) = 1, \quad \text{and} \quad \lim_{n,p\to\infty} \inf_{\mathbf{R}_{st}\in\mathcal{V}_{st}^{(2)}(c_2)} \Pr(T_{st}^{(2)} \geq q_\alpha) = 1,$$

for some $c_2 \geq c_1$.

Denote by $\mathcal{F}_{st}^{(2)}$ the collection of distributions satisfying (C2.2), and by $\mathcal{T}_{st,\alpha}^{(2)}$ the collection of all α-level test over $\mathcal{F}_{st}^{(2)}$. As for Test I, Test II enjoys a certain rate optimality in its power.

Theorem 7. *Suppose Condition* (C2.2) *holds. Let* α, β *be any positive number with* $\alpha + \beta < 1$, *There exists a positive constant* c_3 *such that for all large* n *and* q_0,

$$\inf_{\mathbf{R}_{\varepsilon,st}\in\mathcal{U}_{st}^{(2)}(c_3)} \sup_{\Phi_{st,\alpha}\in\mathcal{T}_{st,\alpha}^{(2)}} \Pr(\Phi_{st,\alpha} = 1) \leq 1-\beta$$

and

$$\inf_{\mathbf{R}_{\varepsilon,st}\in\mathcal{V}_{st}^{(2)}(c_3)} \sup_{\Phi_{st,\alpha}\in\mathcal{T}_{st,\alpha}^{(2)}} \Pr(\Phi_{st,\alpha} = 1) \leq 1-\beta.$$

4.3. Asymptotic properties of the multiple testing procedure

The properties of the multiple testing procedure (9) are based on the limiting null distribution of each test statistic. In light of Theorems 1 and 5, we have the following results.

Theorem 8. *Consider the multiple testing procedure* (9). *If Conditions* (C1.1) *and* (C1.2) *or* (C1.2*) *hold, the procedure* (9) *with* $T_{st}^{(1)}$ *controls the family-wise error rate at level* α. *If Conditions* (C2.1) *and* (C2.2) *hold, the procedure with* $T_{st}^{(2)}$ *controls the family-wise error rate at level* α.

5. Estimation of $\widehat{\beta}_{v,i}$

Test II depends on the estimators of the regression model. Estimating regression coefficients has been investigated extensively in the past several decades; methods include the Dantzig selector [16], the Lasso [47], the SCAD [22], the adaptive Lasso [53], the Scaled-Lasso [45], the Square-root Lasso [7], and so on. In this paper, we focus on the Dantzig selector and Lasso, and determine when they yield good estimators for the proposed testing procedures. In particular, we discuss the sufficient conditions for Condition (C2.3) to hold.

Before we discuss the estimating methods, we introduce the following notation. For region v and component i, let

$$\mathbf{b}_{v,i} = \frac{1}{n}\sum_{k=1}^{n} (\boldsymbol{X}_{k,v,-i} - \bar{\boldsymbol{X}}_{v,-i})^\top (X_{k,v,i} - \bar{X}_{v,i})$$

be the sample covariance between this component and other components in the region. Denote by

$$\hat{\boldsymbol{\Sigma}}_{vv,-i,-i} = \frac{1}{n}\sum_{k=1}^{n} (\boldsymbol{X}_{k,v,-i} - \bar{\boldsymbol{X}}_{v,-i})(\boldsymbol{X}_{k,v,-j} - \bar{\boldsymbol{X}}_{v,-j})^\top$$

the sample covariance matrix without component i, and let $\mathbf{D}_{v,i} = \text{diag}(\hat{\Sigma}_{vv,-i,-i})$. For the following methods, the tuning parameters are

$$\lambda_{v,i}(\delta) = \delta(\hat{\sigma}_{vv,ii} \ln q_v/n)^{1/2}.$$

Dantzig Selector: For each $v \in \{1, \ldots, p\}$ and $i \in \{1, \ldots, q_v\}$, the Dantzig selector estimators are obtained by

$$\hat{\boldsymbol{\beta}}_{v,i}(\delta) = \arg\min |\boldsymbol{\alpha}|_1, \quad \text{subject to } |\mathbf{D}_{v,i}^{-1/2} \hat{\Sigma}_{-i,-i}\boldsymbol{\alpha} - \mathbf{D}_{v,i}^{-1/2}\mathbf{b}_{v,i}|_\infty \le \lambda_{v,i}(\delta). \tag{11}$$

Lasso: For each $v \in \{1, \ldots, p\}$ and $i \in \{1, \ldots, q_v\}$, the Lasso estimators are obtained by

$$\hat{\boldsymbol{\beta}}_{v,i}(\delta) = \mathbf{D}_{v,i}^{-1/2}\hat{\boldsymbol{\alpha}}_{v,i}(\delta), \tag{12}$$

where

$$\hat{\boldsymbol{\alpha}}_{v,i}(\delta) = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^{p-1}} \left[\frac{1}{2n}\sum_{k=1}^{n} \{X_{k,v,i} - \bar{X}_{v,i} - (\boldsymbol{X}_{k,v,-i} - \bar{\boldsymbol{X}}_{v,-i})\mathbf{D}_{v,i}^{-1/2}\boldsymbol{\alpha}\}^2 + \lambda_{v,i}(\delta)|\boldsymbol{\alpha}|_1 \right].$$

We now demonstrate that under certain conditions, the methods yield good estimators that satisfy the need for testing. Define the error bound by $a_{v,1}$ and $a_{v,2}$

$$a_{v,1} = \max_{1\le i\le q_v} |\hat{\boldsymbol{\beta}}_{v,i} - \boldsymbol{\beta}_{v,i}|_1, \quad a_{v,2} = \max_{1\le i\le q_v} |\hat{\boldsymbol{\beta}}_{v,i} - \boldsymbol{\beta}_{v,i}|_2. \tag{13}$$

Proposition 3. *Suppose that Condition* (C2.2) *holds. Consider the Dantzig selector estimator* $\hat{\boldsymbol{\beta}}_{v,i}(2)$ *in* (11). *If* $\max_{1\le i\le q_v}|\boldsymbol{\beta}_{v,i}|_0 = o[n(\ln q_0)^{-3-2\alpha_0}\{\lambda_{\min}(\Sigma)\}^2]$, *then Condition* (C2.3) *holds.*

Proposition 4. *Suppose that Condition* (C2.2) *holds. Consider the Lasso estimator* $\hat{\boldsymbol{\beta}}_{v,i}(2.02)$ *in* (12). *If* $\max_{1\le i\le q_v}|\boldsymbol{\beta}_{v,i}|_0 = o[n(\ln q_0)^{-3-2\alpha_0}\{\lambda_{\min}(\Sigma)\}^2]$, *Condition* (C2.3) *holds.*

In fact, Proposition 3 holds for any Dantzig selector estimator $\hat{\boldsymbol{\beta}}_{v,i}(\delta)$ with $\delta \ge 2$; and Proposition 4 holds for any Lasso estimator $\hat{\boldsymbol{\beta}}_{v,i}(\delta)$ with $\delta > 2$. For computational simplicity, we chose $\delta = 2.02$. The numeric studies indicate that such choice works well in testing.

6. Simulation studies

In this section, we evaluate the performance of the our methods via two simulation studies: one is focused on the size and power of the proposed tests for two regions, the other illustrates how to identity the functional brain network using the proposed tests under family-wise error rate controls.

6.1. Size and power

We simulate a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a normal distribution with mean zero and covariance $\Sigma_{11,22}$. For each $k \in \{1, \ldots, n\}$,

$$\boldsymbol{X}_k \sim \mathcal{N}(\mathbf{0}_{q_1+q_2}, \Sigma_{11,22}) \quad \text{with} \quad \Sigma_{11,22} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{pmatrix},$$

where $\boldsymbol{X}_k = (\boldsymbol{X}_{k,1}^\top, \boldsymbol{X}_{k,2}^\top)^\top$ and $\boldsymbol{X}_{k,s}$ is of dimension q_s for $s = 1, 2$. For comparison, we also consider two alternative test procedures for $\mathcal{H}_{0,12}$ in (1):

Test III: A simple test based on the Pearson correlation coefficient between the principal component scores. Specifically, denote by $\mathbf{Z}_s$ the first principal component score of data $(\boldsymbol{X}_{1,s}^\top, \ldots, \boldsymbol{X}_{n,s}^\top)^\top$. We compute the sample correlation between $\mathbf{Z}_1$ and $\mathbf{Z}_2$, denoted $\widehat{\rho}_{st}$. The Fisher's Z transformation is then taken to obtain the testing statistics $T_{st}^{(3)}$ for this simple approach, which is given by

$$T_{12}^{(3)} = \frac{1}{2}\ln\left(\frac{1+\widehat{\rho}_{12}}{1-\widehat{\rho}_{12}}\right).$$

Using the results by [27], it is straightforward to show that $\sqrt{n-3}\,T_{12}^{(3)} \rightsquigarrow N(0,1)$ under $\mathcal{H}_{0,12}$ in (1). This implies that we reject $\mathcal{H}_{0,12}$ if $\sqrt{n-3}\,|T_{12}^{(3)}| > z_{\alpha/2}$, where z_α is the $1-\alpha$ normal quantile.

Test IV: An alternative test procedure that extends the idea by [13]. The test statistic is

$$T_{12}^{(4)} = n\max_{i,j}\tilde{\rho}_{12,ij}^2 - 2\ln(q_1q_2) + \ln\ln(q_1q_2)$$

where $\tilde{\rho}_{12,ij} = \hat{\sigma}_{12,ij}/(\hat{\theta}_{12,ij})^{1/2}$ and

$$\hat{\theta}_{12,ij} = \frac{1}{n}\sum_{k=1}^{n}\left\{(X_{k,1,i} - \bar{X}_{1,i})(X_{k,2,j} - \bar{X}_{2,j}) - \hat{\sigma}_{12,ij}\right\}^2.$$

Note that $\hat{\theta}_{12,ij}$ is an estimator of $\theta_{12,ij} = \mathrm{var}\{(X_{k,1,i} - \mu_{1,i})(X_{k,2,j} - \mu_{2,j})\}$, where $\mu_{1,i}$ and $\mu_{1,j}$ are the expectations of $X_{k,1,i}$ and $X_{k,2,j}$, respectively. Although $T_{12}^{(4)}$ may also work for our problem, Test I statistic $T_{12}^{(1)}$ requires weaker conditions for its asymptotic properties to hold. More specifically, to derive the asymptotic null distribution of $T_{12}^{(4)}$, we need an additional moment bound condition; see Condition (C2) in [13]. Test I statistic $T_{12}^{(1)}$ does not need this additional condition, and therefore can work for more general distributions.

To define different model specifications on $\Sigma_{11,22}$, we introduce a few auxiliary matrices. Let $\mathbf{A}_d = (a_{ij})_{d\times d}$ where $a_{ii} = 1$ and $a_{ij} \sim 0.5 \times$ Bernoulli(0.5) for $10(k-1)+1 \le i \ne j \le 10k$, where $k = 1, \ldots, [d/10]$ and $a_{ij} = 0$ otherwise. Let $\mathbf{B}_d = (b_{ij})_{d\times d}$ where $b_{ii} = 1$, $b_{i,i+1} = b_{i-1,i} = 0.5$ and $b_{i,j} = 0$ for $|i-j| > 3$.

Let $\Lambda_d = (\lambda_{ij})_{d\times d}$ with $\lambda_{ii} \sim \mathcal{U}(0.5, 2.5)$ and $\lambda_{ij} = 0$ for $i \ne j$. Now, we define four different models for Σ_{11} and Σ_{22}.

- Model 1 (Independent Cases): $\Sigma_{ss} = \Lambda_{q_s}$, for $s = 1, 2$.
- Model 2 (Block Sparse Covariance Matrices): $\Sigma_{ss} = \Lambda_{q_s}^{1/2}(\mathbf{A}_{q_s} + \delta_i \mathbf{I}_{q_s})/(1+\delta_i)\Lambda_{q_s}^{1/2}$, for $s = 1, 2$, where $\delta_i = |\lambda_{\min}(\mathbf{A}_{q_s})| + 0.05$.
- Model 3 (Block Sparse Precision Matrices): $\Sigma_{ss} = \Lambda_{q_s}^{1/2}(\mathbf{A}_{q_s}^{-1} + \delta_i^* \mathbf{I}_{q_s})/(1+\delta_i^*)\Lambda_{q_s}^{1/2}$, for $s = 1, 2$, where $\delta_i^* = |\lambda_{\min}(\mathbf{A}_{q_s}^{-1})| + 0.05$.
- Model 4 (Binded Sparse Covariance Matrices): $\Sigma_{ss} = \Lambda_{q_s}^{1/2}(\mathbf{B}_{q_s} + \tau_s \mathbf{I}_{q_s})/(1+\tau_s)\Lambda_{q_s}^{1/2}$, for $s = 1, 2$, where $\tau_s = |\lambda_{\min}(\mathbf{B}_{q_s})| + 0.05$.
- Model 5 (Binded Sparse Precision Matrices): $\Sigma_{ss} = \Lambda_{q_s}^{1/2}(\mathbf{B}_{q_s}^{-1} + \tau_s^* \mathbf{I}_{q_s})/(1+\tau_s^*)\Lambda_{q_s}^{1/2}$, for $s = 1, 2$, where $\tau_s^* = |\lambda_{\min}(\mathbf{B}_{q_s}^{-1})| + 0.05$.
- Model 6 (Heavy Tail distribution): Instead of a normal distribution, we assume that $\boldsymbol{X}_k$ follows a multivariate t distribution with zero mean, 5 degrees of freedom, and the covariance structure is the same as Model 1.

To simulate the empirical size, we assume $\Sigma_{12} = \mathbf{0}_{q_1\times q_2}$. To evaluate the empirical power, let $\Sigma_{12} = (\sigma_{ij})_{q_1\times q_2}$ with $\sigma_{ij} \sim s_{ij} \times$ Bernoulli$[5/(q_1 q_2)]$ with $s_{ij} \sim \mathcal{N}(4\sqrt{\ln(q_1 q_2)/n}, 0.5)$. The sample size is taken to be $n = 80$ and 150, while the dimension (q_1, q_2) varies over (50, 50), (100, 150), (200, 200) and (250, 300). The nominal significant level for all the tests is set at $\alpha = 0.05$. The empirical sizes and powers for the six models, reported in Tables 1 and 2, are estimated from 5000 replications.

Obviously when the covariance matrix of each region is sparse, Test I controls the type I error better, while when the precision matrix is sparse, Test II controls the type I error better. This shows the influence of Conditions (C1.1) and (C2.1) when deriving the limiting null distribution. On the other hand, the simulation also shows that without these two conditions, there is very little inflation in the type I error. The power analysis shows the similar pattern. In general, Test I/II has a larger power when the covariance/precision matrix is sparse. Both Tests I and II achieve a much larger power than Test III (the Pearson correlation test on the first PC scores), although the empirical sizes of Test III are comparable to the proposed tests. Test IV has a similar performance as Test I in terms of empirical size, where the power is slightly smaller than Test 1 for Models 1–5. For Model 6 (heavy tail distribution case), Tests I, III and IV can still control the size, while Test II has an inflated size. For Model 6, Tests I and II still maintain a high power, while Test IV can hardly detect any signal. In order for Test IV to work, additional constraints are required on the high order moments, and it is possible that these conditions do not hold for heavy-tailed distributions.

6.2. Network identifications

In this section, we perform the simulation studies to illustrate the performance of our proposed testing procedure with the family-wise error rate control on the network identifications. We simulate a region level brain network according to the Erdős–Rényi model [21]. We set the number of regions $p = 90$, and the probability of any two brain regions being functional connected as 0.01. The simulated brain network is shown in Fig. B.1 of the Online Supplement.

For every two connected brain regions s and t on the simulated network, we consider four models that we discussed in Section 6.1 for the specifications of Σ_{ss} and Σ_{tt}. Similar to the simulation studies for evaluating the empirical power, we set $\Sigma_{st} = (\sigma_{ij})_{q_s\times q_t}$ with $\sigma_{ij} \sim s_{ij} \times$ Bernoulli$(10/d_{st})$ with $s_{ij} \sim \mathcal{N}(4\sqrt{\ln(d_{st})/n}, 1)$. We set $n = 150$ and simulate the fMRI time series based on a normal model, i.e., $\boldsymbol{X}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_{q\times q})$, for each $k \in \{1, \ldots, n\}$, where $q = q_1 + \cdots + q_p$ and

$$\Sigma_{q\times q} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1p} \\ \Sigma_{21} & \Sigma_{12} & \ldots & \Sigma_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \Sigma_{p1} & \Sigma_{p2} & \ldots & \Sigma_{pp} \end{pmatrix}.$$

Table 1
Empirical size of Tests I, II, III and IV for different sample sizes and models ($\times 10^{-2}$).

Model	Test	(q_1, q_2)				
		(30, 30)	(50, 50)	(100, 150)	(200, 200)	(300, 250)
		$n = 80$				
1	I	4.50	4.46	4.54	5.14	6.16
	II	4.58	4.48	4.70	5.70	5.44
	III	6.48	6.26	3.38	5.34	7.60
	IV	3.98	4.56	5.02	6.04	4.08
2	I	4.20	4.60	4.52	6.04	6.06
	II	2.88	4.06	4.08	3.86	2.88
	III	6.46	4.58	8.88	7.34	6.32
	IV	4.32	4.40	4.62	3.04	3.06
3	I	3.44	4.02	4.50	4.98	3.20
	II	4.56	3.94	5.02	5.76	5.74
	III	8.26	3.36	7.40	6.38	3.48
	IV	4.24	3.42	3.56	3.88	4.18
4	I	4.80	4.82	5.12	5.22	6.02
	II	1.92	2.28	3.04	2.16	3.12
	III	4.42	3.36	6.56	4.78	3.20
	IV	3.60	3.98	4.84	4.98	5.82
5	I	0.88	1.02	1.06	1.90	1.90
	II	4.52	4.60	4.32	6.28	6.14
	III	4.52	4.28	5.38	4.36	6.40
	IV	1.08	1.04	1.42	2.14	2.10
6	I	3.38	2.62	1.02	0.82	1.20
	II	6.20	6.82	9.64	13.78	7.98
	III	5.58	4.38	5.58	5.58	4.80
	IV	1.98	2.36	1.38	1.60	1.58
		$n = 150$				
1	I	4.94	4.10	5.04	4.62	4.84
	II	4.76	4.34	4.78	5.18	5.36
	III	8.80	4.04	6.44	5.56	5.76
	IV	4.82	4.12	5.00	4.72	4.72
2	I	5.08	4.62	4.48	4.88	4.74
	II	4.02	4.68	4.40	4.70	4.24
	III	5.86	7.46	3.30	4.04	5.02
	IV	4.98	4.32	4.26	4.72	4.64
3	I	4.94	4.68	4.50	4.86	4.60
	II	5.34	4.68	4.26	5.12	5.04
	III	2.76	8.80	4.74	5.22	3.98
	IV	4.64	4.88	4.80	4.96	4.84
4	I	5.02	4.78	4.96	4.92	5.10
	II	2.62	2.46	3.62	3.42	3.78
	III	2.92	5.74	6.50	5.52	4.00
	IV	4.82	5.12	5.06	4.86	4.88
5	I	1.96	1.92	1.96	2.18	3.10
	II	5.62	4.46	4.04	4.92	4.94
	III	3.38	5.92	3.90	5.42	2.34
	IV	2.04	2.12	1.98	2.24	4.08
6	I	4.00	2.82	0.64	1.62	0.80
	II	7.18	7.20	5.40	9.98	7.42
	III	4.00	4.98	4.00	4.24	4.02
	IV	2.00	2.22	2.82	1.62	1.96

Table 3 reports the accuracy of the network identification and the performance for multiple testing. Denote E_{st} as the indicator of the true connectivity between region s and region t, and $\hat{E}_{a,st}$ as the indicator of the estimated connectivity at the ath iteration for each $1 \le s < t \le p$ and $a \in \{1, \ldots, 5000\}$. The NETTPR is defined as the percentage of exactly identifying the correct network, the FWER is the empirical family-wise error rate which is the frequency of having one or mode false discoveries of the functional connectivity over the brain network, and the FDP is the false discovery proportion among the entire detections. Mathematically,

$$\text{NETTPR} = \frac{1}{5000}\sum_{a=1}^{5000} \mathbf{1}(\forall_{1\le s<t\le p} \quad \hat{E}_{a,st} = E_{st}),$$

Table 2
Empirical power of Tests I, II, III and IV for different sample sizes and models ($\times 10^{-2}$).

Model	Test	(q_1, q_2)				
		(30, 30)	(50, 50)	(100, 150)	(200, 200)	(300, 250)
		$n = 80$				
1	I	88.58	85.00	60.20	55.44	54.74
	II	88.46	85.46	60.36	55.84	54.04
	III	11.32	6.26	7.06	8.66	6.18
	IV	86.48	83.92	56.12	54.98	45.18
2	I	88.04	80.20	59.78	55.08	55.10
	II	69.72	64.10	49.70	44.72	43.94
	III	6.46	4.00	7.00	5.72	7.28
	IV	85.02	78.12	49.98	54.04	50.90
3	I	69.88	65.50	50.24	44.40	44.36
	II	87.46	80.40	59.30	54.94	55.90
	III	3.84	3.36	7.80	4.50	3.96
	IV	66.28	62.40	48.22	42.20	42.12
4	I	90.24	95.42	63.40	56.08	64.32
	II	56.82	59.16	43.98	42.18	42.84
	III	8.02	8.52	10.12	5.96	8.64
	IV	89.24	90.22	61.22	52.08	62.20
5	I	80.82	75.14	44.30	35.00	34.78
	II	89.94	85.36	54.30	49.90	44.96
	III	8.12	5.30	6.52	6.68	7.60
	IV	78.28	72.10	45.20	36.20	35.80
6	I	45.60	18.04	100.00	100.00	100.00
	II	99.82	99.14	100.00	100.00	100.00
	III	7.8	6.22	8.64	9.42	9.04
	IV	0.24	0.62	0.04	0.02	0.42
		$n = 150$				
1	I	98.82	98.08	96.66	89.24	85.22
	II	98.96	98.04	96.98	87.78	85.04
	III	13.82	4.04	8.82	7.52	9.48
	IV	98.12	94.38	98.84	88.18	84.10
2	I	99.14	97.86	97.02	87.62	84.46
	II	86.98	75.92	73.30	55.58	55.18
	III	8.10	11.48	6.26	5.02	3.64
	IV	94.36	94.78	94.08	84.28	88.64
3	I	90.06	87.74	76.38	54.88	55.48
	II	94.58	94.70	92.48	84.80	79.94
	III	3.80	9.26	4.26	5.84	3.04
	IV	84.06	83.42	68.22	74.32	30.16
4	I	95.26	92.56	88.68	74.92	85.42
	II	85.40	67.54	64.48	58.32	59.26
	III	9.34	10.14	9.24	6.56	6.08
	IV	98.12	98.54	76.86	72.00	65.12
5	I	84.74	79.74	56.00	44.96	45.40
	II	95.10	89.96	78.44	55.24	53.32
	III	7.94	9.08	5.26	3.62	2.34
	IV	74.74	86.74	63.22	45.44	47.12
6	I	98.24	79.62	90.04	100.00	100.00
	II	98.98	99.98	95.22	100.00	100.00
	III	8.84	8.60	9.20	9.8	9.40
	IV	0.62	0.04	0.24	0.02	0.04

$$\text{FWER} = \frac{1}{5000}\sum_{a=1}^{5000}\mathbf{1}(\exists_{s<t} \quad \hat{E}_{a,st} = 1, E_{st} = 0),$$

$$\text{FDP} = \frac{\sum_{a=1}^{5000}\sum_{1\le s<t\le p}\mathbf{1}(\hat{E}_{a,st} = 1, E_{st} = 0)}{\sum_{a=1}^{5000}\sum_{1\le s<t\le p}\mathbf{1}(\hat{E}_{a,st} = 1)}.$$

Table 3 shows the similar pattern as Tables 1 and 2. When the covariance matrix is the identity matrix, Test I performs better than Test II since the optimization step of Test II introduces extra errors. In addition, Test I is computationally much faster

Table 3
Accuracy of the network identification for Tests I and II.

	Test I			Test II		
	NETTPR	FWER	FDP	NETTPR	FWER	FDP
Model 1	0.72	0.02	0.08	0.60	0.02	0.08
Model 2	0.64	0.02	0.04	0.56	0.08	0.02
Model 3	0.24	0.10	0.06	0.68	0.04	0.12
Model 4	0.66	0.04	0.02	0.36	0.16	0.08
Model 5	0.18	0.12	0.07	0.70	0.02	0.06
Model 6	0.38	0.04	0.04	0.64	0.12	0.16

than Test II. Therefore we recommend Test I when the covariance matrix is the identity matrix or sparse, and Test II when the precision matrix is sparse and its inverse is not sparse.

7. Application

In this section, we demonstrate our method via an analysis of the resting-state fMRI data that are collected in the autism brain imaging data exchange (ABIDE) study [20]. The major goal of the ABIDE is to explore the association of brain activity with the autism spectrum disorder (ASD), which is a widely recognized disease due to its high prevalence and substantial heterogeneity in children [6]. The ABIDE study collected 20 resting-state fMRI data sets from 17 different sites consisting of 1112 individuals with 539 ASDs and 573 age-matched typical controls (TCs). The resting-state fMRI is a popular non-invasive imaging technique that measures the blood oxygen level to reflect the resting brain activity. For each subject, the fMRI signal was recorded for each voxel in the brain over multiple points in time (multiple scans). The different sites in the ABIDE consortium produced a different number of fMRI scans ranging from 72 to 310. Several regular imaging preprocessing steps [20,30], e.g., motion corrections, slice-timing correction, spatial smoothing, have been applied to the fMRI data, which were registered into the MNI space (image size: $91 \times 109 \times 91(2\ \text{mm}^3)$) consisting of 228,483 voxels. We concentrate on the network identification for over 90 regions in the brain, with regions defined according to the AAL system.

We take a whitening transformation of original fMRI signals using the AR(1) model [52] to remove the temporal correlations. The de-trending and de-meaning procedures are also applied for original fMRI signals. We perform the principal component analysis (PCA) to summarize the voxel-level fMRI time series into a relatively small number of principal component signals within each region. The number of signals is chosen according to the criterion of the cumulative variance contribution being larger than 90%. The mean number of the principal components over 90 regions is 18 and ranges from 6 to 36. We apply the proposed methods to identify the resting state brain network for each subject. The network for a group of subjects is defined by including the connections for regions i and j if they are connected with over 85% of subject-level networks. The ASD patient and control networks include 445 connections and 502 connections respectively, where numbers of unique connections are 31 and 88. The number of connections shared by both groups is 441. The control network is denser than the ASD patient network.

Fig. 1 shows the unique connections for the ASD patient network and the healthy control network. In the ASD patient network, there are two "hub" brain regions that have at least four unique connections to other regions in the brain. They are the medial part of the superior frontal gyrus (SFGmed-R) and Gyrus rectus (REC). These regions were demonstrated in the previous publications [5,25,40,48] to be strongly associated with autism. Our results suggest that ASD patients have active region level functional connectivities between these three regions, while the controls do not have these connectivities. On the other hand, in the healthy control network, there are three "hub" regions that have at least seven connections. They are the dorsolateral part of right superior frontal gyrus (SFGdor-R), the left middle frontal gyrus (MFG-L) and the right middle frontal gyrus (MFG-R). Our results suggest that autism patients break the greatest number of the connectivities in these three regions. The brain functions of these regions are consistent with the autism clinical symptom. For example, the superior frontal gyrus is known for being involved in self-awareness, in coordination with the action of the sensory system [24].

8. Discussion

The novel contributions of our work include the following: (1) we propose a new framework to identify the region level functional brain network using formal statistical testing procedures, which makes full use of the massive voxel-level brain signals and incorporate the brain anatomy into the analysis, producing neurologically more meaningful interpretations; (2) we establish the statistical theory of the proposed testing procedures, which provides a solid foundation for making valid inference on the functional brain network; (3) the proposed method is computationally very efficient and the computational algorithm is straightforward to implement in parallel mode; and (4) although the development of our proposed approach is motivated by the analysis of brain imaging data, it is a general method for network construction and can be readily applied to other problems, such as identification of gene networks and social networks.

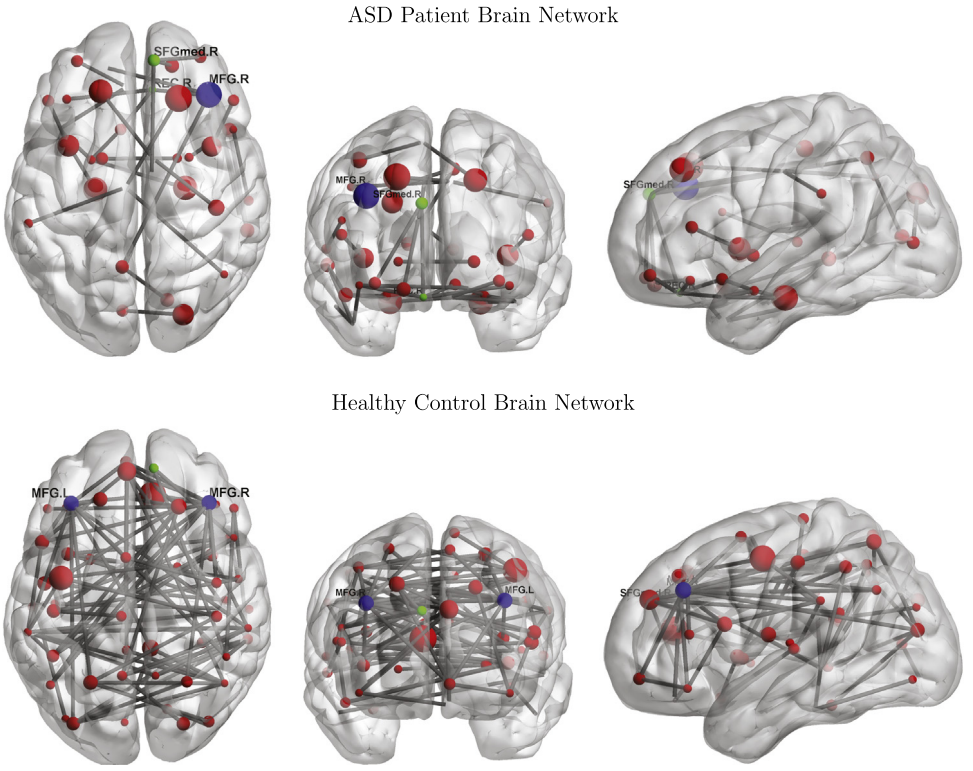

Fig. 1. Identified region level resting state brain networks for ASD patient group and healthy control group.

Acknowledgments

Jichun Xie's research was partially supported by the NIH grant UL1TR001117. Jian Kang's research was partially supported by the NIH grant 1R01MH105561. The authors thank the autism brain imaging data exchange (ABIDE) study [20] that shares the resting-state fMRI data. Detailed input from the Editor-in-Chief, Christian Genest, and his editorial committee is gratefully acknowledged.

Appendix A. Proof of the main theorems

Without loss of generality, in this section, we assume $\mathrm{E}(X_{k,s,i}) = \mathrm{E}(X_{k,t,j}) = 0$, and $\mathrm{var}(X_{k,s,i}) = \mathrm{var}(X_{k,t,j}) = 1$ unless otherwise stated. Due to space limitations, we list the proofs of some theorems (Theorems 2, 4, 5, and Propositions 3 and 4) here. Theorem 6 follows similar arguments as Theorem 2, and the proof of Theorem 7 follows along similar lines as that of Theorem 4. The proof of Theorem 1 is relatively long and the main techniques follow the proof of Theorem 1 in [14]; thus it is relegated to the Online Supplement.

In addition, to simplify the notation in the proofs, we denote by $d_{st} = q_s q_t$ the total number of entries in the covariance matrix $\boldsymbol{\Upsilon}_{st}$. We also define $c(d_{st}, \alpha) = 2\ln(d_{st}) - \ln\ln(d_{st}) + q_\alpha$, where q_α is the $(1-\alpha)$th quantile of the null distribution F.

To prove Theorem 2, we need Lemmas 1 and 2.

Lemma 1. *Recall that* $\theta_{1,st,ij} = \sigma_{ss,ii}\sigma_{tt,jj}$ *and* $\hat{\theta}_{1,st,ij} = \hat{\sigma}_{ss,ii}\hat{\sigma}_{tt,jj}$. *Under Condition* (C1.2) *or* (C1.2*) *and the null* $\mathcal{H}_{0,st}$, *there exists some constant* $C > 0$, *such that as* $n, q_0 \to \infty$,

$$\Pr\left\{\max_{i,j}\left|1 - \frac{\hat{\theta}_{1,st,ij}}{\theta_{1,st,ij}}\right| \geq C\frac{1}{(\ln q_0)^2}\right\} = O(q_0^{-1} + n^{-\epsilon/4}). \tag{A.1}$$

Lemma 2. *Recall that* $\theta_{st,ij} = \mathrm{var}\{(X_{k,s,i} - \mu_{s,i})(X_{k,t,j} - \mu_{t,j})\}$. *Under Condition* (C1.2) *or* (C1.2*), *there exists a constant* $C > 0$ *such that*

$$\Pr\left\{\max_{(i,j)\in\mathcal{A}}\frac{(\tilde{\sigma}_{st,ij} - \sigma_{st,ij})^2}{\theta_{st,ij}/n} \geq x^2\right\} \leq C|\mathcal{A}|\{1 - \Phi(x)\} + O(q_0^{-M} + n^{-\epsilon/8}) \tag{A.2}$$

uniformly for $0 \le x \le (8 \ln q_0)^{1/2}$ *and* $\mathcal{A} \subseteq \{(i,j) : 1 \le i \le q_s,\ 1 \le j \le q_t\}$. *Under* $\mathcal{H}_{0,st}$, (A.2) *also holds when substituting* $\theta_{st,ij}$ *to* $\theta_{1,st,ij}$.

Proof of Theorem 2. Define

$$T_{st,2} = \max_{i,j} \frac{n\hat{\sigma}^2_{st,ij}}{\theta_{1,st,ij}}, \quad T_{st,3} = \max_{i,j} \frac{n\sigma^2_{st,ij}}{\theta_{1,st,ij}},$$
$$T_{st,4} = \max_{ij} \frac{n(\hat{\sigma}_{st,ij} - \sigma_{st,ij})^2}{\theta_{1,st,ij}}, \quad T_{st,5} = \max_{ij} \frac{n(\hat{\sigma}_{st,ij} - \sigma_{st,ij})^2}{\theta_{st,ij}}.$$

By Lemma 1,

$$\Pr(T^{(1)}_{st} > q_\alpha) \ge \Pr[T_{st,2} \ge c(d_{st}, \alpha)\{1 + o(1)\}].$$

Since $T_{st,3} \le 2T_{st,4} + 2T_{st,2}$ and $T_{st,3} \ge 4(1 + \kappa_{st}) \ln d_{st}$, we get

$$\begin{aligned}\Pr[T_{st,2} \ge c(d_{st}, \alpha)\{1 + o(1)\}] &\ge \Pr[T_{st,3} - 2T_{st,4} \ge 2c(d_{st}, \alpha)\{1 + o(1)\}] \\ &= \Pr[T_{st,4} \le T_{st,3}/2 - c(d_{st}, \alpha)\{1 + o(1)\}] \\ &= \Pr[T_{st,4} \le \{2\kappa_{st} \ln d_{st} + \ln(\ln d_{st}) - q_\alpha\}\{1 - o(1)\}].\end{aligned}$$

Because $T_{st,5} \ge T_{st,4}/\kappa_{st}$. It follows that

$$\Pr[T_{st,4} \le \{2\kappa_{st} \ln d_{st} + \ln(\ln d_{st}) - q_\alpha\}\{1 + o(1)\}] \ge \Pr[T_{st,5} \le \{2 \ln d_{st} + (1/\kappa_{st}) \ln(\ln d_{st}) - (1/\kappa_{st})q_\alpha\}\{1 - o(1)\}].$$

By Lemma 2,

$$\Pr[T_{st,5} \le \{2 \ln d_{st} + (1/\kappa_{st}) \ln \ln d_{st} - (1/\kappa_{st})q_\alpha\}\{1 - o(1)\}] \to 1. \quad \square$$

Proof of Theorem 4. It suffices to prove the conclusion for variables with a normal distribution satisfying Condition (C2) and (C2*). Let $\min(q_s, q_t) = q^*(s,t)$. Denote by $\mathcal{M}(s,t) = \{\mathcal{S} :\ \mathcal{S} \subseteq \{1, \ldots, q^*\},\ \mathrm{card}(\mathcal{S}) = r_{st}\}$ the set of all the subsets of $\{1, \ldots, q^*\}$ with cardinality r_{st}. Let $\hat{m}$ be a random subset of $\{1, \ldots, q^*\}$, which is uniformly distributed on $\mathcal{M}$. Consider the following covariance matrix of $(\boldsymbol{X}_s, \boldsymbol{X}_t)^\top$:

$$\boldsymbol{\Sigma}^*_{\hat{m}} = \begin{pmatrix} \mathbf{I}_{q_s \times q_s} & \boldsymbol{\Sigma}^*_{st,\hat{m}} \\ \boldsymbol{\Sigma}^{*\top}_{st,\hat{m}} & \mathbf{I}_{q_t \times q_t} \end{pmatrix}, \quad \text{and } \boldsymbol{\Sigma}^*_{st,\hat{m}} = (\sigma_{st,ij})_{q_s \times q_t},$$

with

$$\sigma_{st,i_1 i_1} = \rho = c\{\ln(d_{st}/n)\}^{1/2},\ \sigma_{st,i_2 i_2} = \sigma_{st,ij} = 0$$

for all $i_1 \in \mathcal{M}(s,t),\ i_2 \in \mathcal{M}(s,t)^c,\ j \ne i$. Here c is a positive constant which will be specified later. Without loss of generality, suppose $q_s \le q_t$. Let us reorder the variables $\boldsymbol{X} = (X_{s,1}, X_{t,1}, \ldots, X_{s,q_s}, X_{t,q_s}, \ldots, X_{t,q_t})^\top$. Then the covariance matrix of $\boldsymbol{X}$ is $\boldsymbol{\Sigma}_{\hat{m}} = \mathrm{diag}(A(i), \ldots, A(i), \mathbf{I}_{q_t - q_s})$, with

$$A(i) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ if } i \in \hat{m} \quad \text{and} \quad A(i) = \mathbf{I}_2 \text{ if } i \in \hat{m}^c.$$

It is easy to see that the precision matrix is $\boldsymbol{\Omega}_{\hat{m}} = \mathrm{diag}(B(i), \ldots, B(i), \mathbf{I}_{q_t - q_s})$, with

$$A(i) = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \text{ if } i \in \hat{m} \quad \text{and} \quad A(i) = \mathbf{I}_2 \text{ if } i \in \hat{m}^c.$$

We construct a class of $\boldsymbol{\Sigma}$: $\mathcal{Q} = \{\boldsymbol{\Sigma}_{\hat{m}}, \hat{m} \in \mathcal{M}(s,t)\}$. Let $\boldsymbol{\Sigma}_0 = \mathbf{I}$ and $\boldsymbol{\Sigma}_1$ be uniformly distributed on $\mathcal{Q}$. Let μ_ρ be the distribution of $\boldsymbol{\Sigma}_1$. It is a measure on $\{\Delta \in \mathcal{S}(r_{st}, s, t) :\ \|\Delta\|_F^2 = r_{st}\rho^2\}$. For $a \in \{0, 1\}$, let $dP_a(\boldsymbol{X})$ be the likelihood function given $\boldsymbol{\Sigma}_a$. Define

$$L_{\mu_\rho}(\boldsymbol{X}) = \mathrm{E}_{\mu_\rho}\left\{\frac{dP_1(\boldsymbol{X})}{dP_0(\boldsymbol{X})}\right\},$$

where E_{μ_ρ} is the expectation on $\boldsymbol{\Sigma}_{\hat{m}}$. By the arguments in Section 7.1 of Baraud [4], it suffices to show that $\mathrm{E}_0(L^2_{\mu_\rho}) \le 1 + o(1)$.

We have

$$L_{\mu_\rho} = \mathrm{E}_{\hat{m}}\left[\prod_{k=1}^n \frac{1}{|\boldsymbol{\Sigma}_{\hat{m}}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{X}_k^\top(\boldsymbol{\Omega}_{\hat{m}} - \mathbf{I})\boldsymbol{X}_k\right\}\right].$$

Let E_0 be the expectation on $\boldsymbol{X}_k$ with $\mathcal{N}(0, \mathbf{I})$ distribution. Then

$$\mathrm{E}_0(L^2_{\mu_\rho}) = \mathrm{E}_0\left[\frac{1}{\binom{q^*}{r_{st}}} \sum_{m \in \mathcal{M}}\left[\prod_{k=1}^n \frac{1}{|\boldsymbol{\Sigma}_m|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{X}_k^\top(\boldsymbol{\Omega}_m - \mathbf{I})\boldsymbol{X}_k\right\}\right]\right]^2$$

$$= \frac{1}{\binom{q^*}{r_{st}}^2} \sum_{m,m' \in \mathcal{M}} \mathrm{E}_0 \left[\prod_{k=1}^{n} \frac{1}{|\Sigma_m|^{1/2}} \frac{1}{|\Sigma_{m'}|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{X}_k^\top (\Omega_m + \Omega_{m'} - 2\mathbf{I}) \boldsymbol{X}_k \right\} \right].$$

Set $\Omega_m + \Omega_{m'} - 2\mathbf{I} = (a_{s_1,s_2,i,j})$, with $s_1, s_2 \in \{s, t\}$, $i \in \{1, \ldots, q_{s_1}\}$, and $j \in \{1, \ldots, q_{s_2}\}$. If $i \in m \cap m'$, $a_{ss,ii} = a_{tt,ii} = 2\rho^2/(1-\rho^2)$, $a_{st,ii} = -2\rho/(1-\rho^2)$. If $i \in m \Delta m'$, $a_{ss,ii} = a_{tt,ii} = 1/(1-\rho^2) - 1$, $a_{st,ii} = -\rho/(1-\rho^2)$. Otherwise, $a_{s_1,s_2,i,j} = 0$. Now let $t = |m \cap m'|$. By simple calculations, we have

$$\begin{aligned}
\mathrm{E}_0(L^2_{\mu_\rho}) &= \frac{1}{\binom{q^*}{r_{st}}^2} (1-\rho^2)^{-nr_{st}} \sum_{t=0}^{r_{st}} \binom{q^*}{r_{st}} \binom{r_{st}}{t} \binom{q^* - r_{st}}{r_{st} - t} 1^{tn} (1-\rho^2)^{(2r_{st}-t)n/2} \\
&= \binom{q^*}{r_{st}}^{-1} \sum_{t=1}^{r_{st}} \binom{r_{st}}{t} \binom{q^* - r_{st}}{r_{st} - t} (1-\rho^2)^{-tn/2} \\
&\le q^{*r_{st}} \frac{(q^* - r_{st})!}{q^*!} \sum_{t=0}^{r_{st}} \binom{r_{st}}{t} \left(\frac{s}{q^*}\right)^t \left(\frac{1}{1-\rho^2}\right)^{tn/2} \\
&= \{1 + o(1)\} \left\{ 1 + \frac{r_{st}}{q^*(1-\rho^2)^{n/2}} \right\}^{r_{st}} \\
&\le \exp\{r_{st} \ln(1 + r_{st} q^{*c^2-1})\}\{1 + o(1)\} \\
&\le \exp(r_{st}^2 q^{*c^2-1})\{1 + o(1)\}.
\end{aligned}$$

For sufficiently small c^2, $\mathrm{E}_0(L^2_{\mu_\rho}) = 1 + o(1)$, and hence the theorem is proved. □

Proof of Theorem 5. Define

$$T_{st} = n \max_{ij} \rho_{\varepsilon,st}, \quad \hat{T}_{st} = \max_{i,j} \frac{n(\hat{\sigma}_{\varepsilon,st,ij} - \sigma_{\varepsilon,st,ij})^2}{\theta_{\varepsilon,st,ij}}$$

$$\tilde{T}_{st} = \max_{ij} \frac{n(\tilde{\sigma}_{\varepsilon,st,ij} - \sigma_{\varepsilon,st,ij})^2}{\theta_{\varepsilon,st,ij}}, \quad \breve{T}_{st} = \max_{i,j} \frac{n(\breve{\sigma}_{\varepsilon,st,ij} - \sigma_{\varepsilon,st,ij})^2}{\theta_{\varepsilon,st,ij}},$$

where

$$\hat{\sigma}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \hat{\varepsilon}_{k,s,i} \hat{\varepsilon}_{k,t,j}, \quad \tilde{\sigma}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\varepsilon}_{k,s,i} \tilde{\varepsilon}_{k,t,j}, \quad \breve{\sigma}_{\varepsilon,st,ij} = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,s,i} \varepsilon_{k,t,j}.$$

By Condition (2.3) and $\max_i |\tilde{\sigma}_{\varepsilon,ss,ii} - \sigma_{\varepsilon,ss,ii}| = O_P\{(\ln q_0)^{-1-\alpha_0}\}$,

$$\begin{aligned}
|\hat{\theta}_{\varepsilon,st,ij} - \theta_{\varepsilon,st,ij}| &\le |\hat{\sigma}_{\varepsilon,ss,ii}\hat{\sigma}_{\varepsilon,tt,jj} - \sigma_{\varepsilon,ss,ii}\sigma_{\varepsilon,tt,jj}| \\
&\le O_P\left\{ \max(|\hat{\sigma}_{\varepsilon,ss,ii} - \sigma_{\varepsilon,ss,ii}|, |\hat{\sigma}_{\varepsilon,tt,jj} - \sigma_{\varepsilon,tt,jj}|) \right\} = O_P\{(\ln q_0)^{-1-\alpha_0}\}.
\end{aligned}$$

By (C2.2), $\theta_{\varepsilon,st,ij} \ge 1/c_0^2$. Thus with probability tending to 1,

$$\begin{aligned}
&|T_{st} - \hat{T}_{st}| \le C\hat{T}_{st}(\ln q_0)^{-1-\alpha_0} \\
&|\hat{T}_{st} - \tilde{T}_{st}| \le C(\ln q_0)^{-1-\alpha_0} \\
&|\breve{T}_{st} - \tilde{T}_{st}| \le Cn(\max_{1\le i\le q_s} \bar{\varepsilon}^4_{s,i} + \max_{1\le j\le q_t} \bar{\varepsilon}^4_{t,j}) + Cn^{1/2}\breve{T}^{1/2}_{st}(\max_{1\le i\le q_s} \bar{\varepsilon}^2_{s,i} + \max_{1\le j\le q_t} \bar{\varepsilon}^2_{t,j}).
\end{aligned}$$

The second inequality above stems from Condition (C2.3). Note that

$$\max_{1\le i\le q_s} |\bar{\varepsilon}_{s,i}| + \max_{1\le t\le q_t} |\bar{\varepsilon}_{t,j}| = O_P\{(\ln q_0/n)^{1/2}\}.$$

Thus, it suffices to show that for any $x \in \mathbb{R}$,

$$\Pr\{\breve{T}_{st} \le 2\ln d_{st} - 2\ln(\ln d_{st}) + x\} \to \exp\left(-\frac{1}{\pi^{1/2}} e^{-x/2} \right).$$

The rest of the proof is similar to the proof of Theorem 1. □

Proof of Proposition 3. We first decompose $\hat{\sigma}_{\varepsilon,st,ij}$ as follows:

$$\frac{1}{n} \sum_{k=1}^{n} \hat{\varepsilon}_{k,s,i} \hat{\varepsilon}_{k,t,j} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\varepsilon}_{k,s,i} \tilde{\varepsilon}_{k,t,j} - A_{1,s,t,i,j} - A_{2,s,t,i,j} + A_{3,s,t,i,j},$$

where

$$A_{1,s,t,i,j} = \frac{1}{n}\sum_{k=1}^{n} \tilde{\varepsilon}_{k,s,i}(\boldsymbol{X}_{k,t,-j} - \bar{\boldsymbol{X}}_{t,-j})^{\top}(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j})$$

$$A_{2,s,t,i,j} = \frac{1}{n}\sum_{k=1}^{n} \tilde{\varepsilon}_{k,t,j}(\boldsymbol{X}_{k,s,-i} - \bar{\boldsymbol{X}}_{s,-i})^{\top}(\hat{\boldsymbol{\beta}}_{s,i} - \boldsymbol{\beta}_{s,i})$$

$$A_{3,s,t,i,j} = (\hat{\boldsymbol{\beta}}_{s,i} - \boldsymbol{\beta}_{s,i})^{\top}\hat{\boldsymbol{\Sigma}}_{st,-i,-j}(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j}).$$

We bound each term in order. First note that, for all $s, t \in \{1, \ldots, p\}$,

$$|A_{1,s,t,i,j}| \le \left|\frac{1}{n}\sum_{k=1}^{n} \tilde{\epsilon}_{k,s,i}(\boldsymbol{X}_{k,t,-j} - \bar{\boldsymbol{X}}_{k,t,-j}) - \mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,t,-j})\right|_{\infty} \left|\hat{\boldsymbol{\beta}}_{t,j} - \beta_{t,j}\right|_1 + \left|\mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,s,-j}^{\top})(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j})\right|. \tag{A.3}$$

Now for any $M > 0$, there exists a sufficiently large $C > 0$ such that

$$\Pr\left[\max_{1\le i\le q_s, 1\le j\le q_t}\left|\frac{1}{n}\sum_{k=1}^{n}\tilde{\varepsilon}_{k,s,i}(X_{k,t,-j} - \bar{X}_{t,-j}) - \mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,t,-j})\right|_{\infty} \ge C\{\ln(d_{st}/n)\}^{1/2}\right] = O(q_0^{-M}).$$

Next, recall the definition of $a_{v,1}$ and $a_{v,2}$ in (13). When $s = t$ and $i = j$, $\mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,s,-i}) = \boldsymbol{0}$. Therefore

$$\max_{1\le i\le q_s}\left|A_{1,s,s,i,i}\right| = O_P\left\{a_{s,1}(\ln q_s/n)^{1/2}\right\}.$$

When $s \ne t$, under $\mathcal{H}_{0,st}$, $\mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,t,-j}) = \boldsymbol{0}$. Therefore

$$\max_{1\le i\le q_s, 1\le j\le q_t}\left|A_{1,s,t,i,j}\right| = O_P\left\{a_{t,1}(\ln d_{st}/n)^{1/2}\right\}.$$

When $s \ne t$ and under $\mathcal{H}_{1,st}$,

$$\left|\mathrm{cov}(\tilde{\varepsilon}_{k,s,i}, \boldsymbol{X}_{k,s,-j}^{\top})(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j})\right| \le \left\{\mathrm{var}(\tilde{\varepsilon}_{k,s,i})\right\}^{1/2}\left\{(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j})^{\top}\boldsymbol{\Sigma}_{tt,-j,-j}(\hat{\boldsymbol{\beta}}_{t,j} - \boldsymbol{\beta}_{t,j})\right\}^{1/2} \le c_0 a_{t,2}.$$

Therefore,

$$\max_{1\le i\le q_s, 1\le j\le q_t}\left|A_{1,s,t,i,j}\right| = O_P[a_{t,1}\{\ln(d_{st}/n)\}^{1/2} + a_{t,2}].$$

We can derive bounds for $A_{2,s,t,i,j}$ similarly. Next, we bound $A_{3,s,t,i,j}$. First,

$$A_{3,s,t,i,j} = (\hat{\boldsymbol{\beta}}_{k,s,i} - \boldsymbol{\beta}_{k,s,i})^{\top}(\hat{\boldsymbol{\Sigma}}_{st,-i,-j} - \boldsymbol{\Sigma}_{st,-i,-j})(\hat{\boldsymbol{\beta}}_{k,t,j} - \boldsymbol{\beta}_{k,t,j}) + (\hat{\boldsymbol{\beta}}_{k,s,i} - \boldsymbol{\beta}_{k,s,i})^{\top}\boldsymbol{\Sigma}_{st,-i,-j}(\hat{\boldsymbol{\beta}}_{k,t,j} - \boldsymbol{\beta}_{k,t,j}).$$

It is easy to show that for any $M > 0$, there exists a sufficiently large $C > 0$ such that

$$\Pr\left[\max_{1\le i\le q_s, 1\le j\le q_t}|\hat{\sigma}_{st,ij} - \sigma_{st,ij}| \ge C\{\ln(d_{st}/n)\}^{1/2}\right] = O(q_0^{-M}).$$

When $s \ne t$, one has $\boldsymbol{\Sigma}_{st,-i,-j} = \boldsymbol{0}$ under $\mathcal{H}_{0,st}$ while $\|\boldsymbol{\Sigma}_{st,-i,-j}\|_2 \le c_0$ under $\mathcal{H}_{1,st}$. By the inequality

$$\left|(\hat{\boldsymbol{\beta}}_{k,s,i} - \boldsymbol{\beta}_{k,s,i})^{\top}(\hat{\boldsymbol{\Sigma}}_{st,-i,-j} - \boldsymbol{\Sigma}_{st,-i,-j})(\hat{\boldsymbol{\beta}}_{k,t,j} - \boldsymbol{\beta}_{k,t,j})\right| \le \left|\hat{\boldsymbol{\Sigma}}_{st,-i,-j} - \boldsymbol{\Sigma}_{st,-i,-j}\right|_{\infty}|\hat{\boldsymbol{\beta}}_{k,s,i} - \boldsymbol{\beta}_{k,s,i}|_1|\hat{\boldsymbol{\beta}}_{k,t,j} - \boldsymbol{\beta}_{k,t,j}|_1, \tag{A.4}$$

we have under $\mathcal{H}_{0,st}$,

$$\max_{1\le i\le q_s, 1\le j\le q_t}\left|A_{3,s,t,i,j}\right| = O_P\left[a_{s,1}a_{t,1}\{\ln(d_{st}/n)\}^{1/2}\right];$$

and under $\mathcal{H}_{1,st}$,

$$\max_{1\le i\le q_s, 1\le j\le q_t}\left|A_{3,s,t,i,j}\right| = O_P\left[a_{s,1}a_{t,1}\{\ln(d_{st}/n)\}^{1/2} + a_{s,2}a_{t,2}\right].$$

When $s = t$, we can show by a similar argument that under $\mathcal{H}_{0,st}$,

$$\max_{1\le i\le q_s, 1\le j\le q_t}|A_{3,s,s,i,j}| = O_P\left\{a_{s,1}^2(\ln q_s/n)^{1/2}\right\};$$

and under $\mathcal{H}_{1,st}$,

$$\max_{1\le i\le q_s, 1\le j\le q_t}|A_{3,s,s,i,j}| = O_P\left\{a_{s,1}^2(\ln q_s/n)^{1/2} + a_{s,2}^2\right\}.$$

Therefore, when $s \neq t$, under $\mathcal{H}_{0,st}$

$$\frac{1}{n}\sum_{k=1}^{n}\hat{\varepsilon}_{k,s,i}\hat{\varepsilon}_{k,t,j} = \frac{1}{n}\sum_{k=1}^{n}\tilde{\varepsilon}_{k,s,i}\tilde{\varepsilon}_{k,t,j} + O_P\left\{(a_{s,1}a_{t,1} + a_{s,1} + a_{t,1})\left(\frac{\ln d_{st}}{n}\right)^{1/2}\right\}; \tag{A.5}$$

and under $\mathcal{H}_{1,st}$,

$$\frac{1}{n}\sum_{k=1}^{n}\hat{\varepsilon}_{k,s,i}\hat{\varepsilon}_{k,t,j} = \frac{1}{n}\sum_{k=1}^{n}\tilde{\varepsilon}_{k,s,i}\tilde{\varepsilon}_{k,t,j} + O_P\left\{(a_{s,1}a_{t,1} + a_{s,1} + a_{t,1})\left(\frac{\ln d_{st}}{n}\right)^{1/2} + (a_{s,2}a_{t,2} + a_{s,2} + a_{t,2})\right\}. \tag{A.6}$$

When $s = t$ and $i = j$, under $\mathcal{H}_{0,st}$,

$$\frac{1}{n}\sum_{k=1}^{n}\hat{\varepsilon}_{k,s,i}^2 = \frac{1}{n}\sum_{k=1}^{n}\tilde{\varepsilon}_{k,s,i}^2 + O_P\left\{(a_{s,1}^2 + a_{s,1})\left(\frac{\ln q_s}{n}\right)^{1/2}\right\}; \tag{A.7}$$

and under $\mathcal{H}_{1,st}$,

$$\frac{1}{n}\sum_{k=1}^{n}\hat{\varepsilon}_{k,s,i}^2 = \frac{1}{n}\sum_{k=1}^{n}\tilde{\varepsilon}_{k,s,i}^2 + O_P\left\{(a_{s,1}^2 + a_{s,1})\left(\frac{\ln q_s}{n}\right)^{1/2} + a_{s,2}^2\right\}. \tag{A.8}$$

It then suffices to show that for each $v \in \{1, \ldots, p\}$, $a_{v,2} = O_P\{(\ln q_0)^{-1-\alpha_0}\}$ and $a_{v,1} = O_P\{n(\ln q_0)^{-2-\alpha_0}\}$.
From the proof of Proposition 4.1 in [36], p. 2975, with probability tending to 1,

$$|\mathbf{D}_{v,i}^{-1/2}\hat{\boldsymbol{\Sigma}}_{vv,-i,-i}\hat{\boldsymbol{\beta}}_{v,i} - \mathbf{D}_{v,i}^{-1/2}\mathbf{b}_{v,i}|_\infty \leq \lambda_{v,i}(2).$$

It then follows that

$$|\mathbf{D}_{v,i}^{-1/2}\hat{\boldsymbol{\Sigma}}_{vv,-i,-i}(\hat{\boldsymbol{\beta}}_{v,i} - \boldsymbol{\beta}_{v,i})|_\infty \leq 2\lambda_{v,i}(2).$$

From

$$\max_{1\leq i\leq q_v}|\boldsymbol{\beta}_{v,i}|_0 = o\left\{\lambda_{\min}(\boldsymbol{\Sigma})(n/\ln q_0)^{1/2}\right\}$$

and the inequality

$$\boldsymbol{\delta}^\top\hat{\boldsymbol{\Sigma}}_{vv,-i,-i}\boldsymbol{\delta} \geq \lambda_{\min}(\boldsymbol{\Sigma}_{-i,-i})|\boldsymbol{\delta}|_2^2 - O_P\{(\ln q_0/n)^{1/2}\}|\boldsymbol{\delta}|_1,$$

we can further see that the restricted eigenvalue assumption RE$(s, s, 1)$ in [8], p. 1711, holds with $\kappa(s, s, 1) \geq c\lambda_{\min}(\boldsymbol{\Sigma})^{1/2}$. From the proof of Theorem 7.1 in [8],

$$a_{v,1} = O_P\left\{\max_{1\leq i\leq q_v}|\boldsymbol{\beta}_{v,i}|_0(\ln q_v/n)^{1/2}\right\}, \quad a_{v,2} = O_P\left[\left\{\max_{1\leq i\leq q_v}|\boldsymbol{\beta}_{v,i}|_0(\ln q_n/n)\right\}^{1/2}\{\lambda_{\min}(\boldsymbol{\Sigma})\}^{-1}\right]$$

Therefore, we can conclude. □

Proof of Proposition 4. From the proof of Proposition 4.2 in [36], we have with probability tending to 1,

$$|\mathbf{D}_{v,i}^{-1/2}\hat{\boldsymbol{\Sigma}}_{vv,-i,-i}\mathbf{D}_{v,i}^{-1/2}(\hat{\boldsymbol{\alpha}}_{v,i} - \mathbf{D}_{v,i}^{1/2}\boldsymbol{\beta}_{v,i})|_\infty \leq 2\lambda_{v,i}(\delta).$$

Then by (A.5), (A.6), (A.7), (A.8), and the proof of Theorem 7.2 in [8], we get that Condition (2.3) holds for $\boldsymbol{\beta}_{v,i}(\delta)$ with $\delta > 2$. □

Appendix B. Supplementary data

The supplementary material includes the proofs of theorems and lemmas that are not included in Appendix A, and the simulated network (Fig. B.1) on 90 regions using Erdős–Rényi model discussed in Section 6.2.

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2017.01.011.

References

[1] H. Akaike, Fitting autoregressive models for prediction, Ann. Inst. Statist. Math. 21 (1969) 243–247.
[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 2003.
[3] J.R. Andrews-Hanna, A.Z. Snyder, J.L. Vincent, C. Lustig, D. Head, M.E. Raichle, R.L. Buckner, Disruption of large-scale brain systems in advanced aging, Neuron 56 (2007) 924–935.
[4] Y. Baraud, Non asymptotic minimax rates of testing, Bernoulli 8 (2002) 577–606.
[5] S. Baron-Cohen, H.A. Ring, S. Wheelwright, E.T. Bullmore, M.J. Brammer, A. Simmons, S.C. Williams, Social intelligence in the normal and autistic brain: An fMRI study, Eur. J. Neurosci. 11 (1999) 1891–1898.

[6] M.L. Bauman, T.L. Kemper, The Neurobiology of Autism, JHU Press, 2005.
[7] A. Belloni, V. Chernozhukov, L. Wang, Square-root lasso: Pivotal recovery of sparse signals via conic programming, Biometrika 98 (2011) 791–806.
[8] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, Ann. Statist. 37 (2009) 1705–1732.
[9] M. Birke, D. Holder, A note on testing the covariance matrix for large dimension, Statist. Probab. Lett. 74 (2005) 281–289.
[10] B. Biswal, F. Zerrin Yetkin, V.M. Haughton, J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI, Magn. Reson. Med. 34 (1995) 537–541.
[11] F.D. Bowman, L. Zhang, G. Derado, S. Chen, Determining functional connectivity using fMRI data with diffusion-based anatomical weighting, NeuroImage 62 (2012) 1769–1779.
[12] E. Bullmore, O. Sporns, Complex brain networks: Graph theoretical analysis of structural and functional systems, Nature Rev. Neurosci. 10 (2009) 186–198.
[13] T. Cai, W. Liu, Large-scale multiple testing of correlations, J. Amer. Statist. Assoc. 111 (2016) 229–240.
[14] T. Cai, W. Liu, Y. Xia, Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings, J. Amer. Statist. Assoc. 108 (2013) 265–277.
[15] T. Cai, Z. Ma, Optimal hypothesis testing for high dimensional covariance matrices, Bernoulli 19 (2013) 2359–2388.
[16] E. Candès, T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, Ann. Statist. 35 (2007) 2313–2351.
[17] S. Chen, L. Zhang, P. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (2010) 810–819.
[18] V.L. Cherkassky, R.K. Kana, T.A. Keller, M.A. Just, Functional connectivity in a baseline resting-state network in autism, Neuroreport 17 (2006) 1687–1690.
[19] D. Cordes, V.M. Haughton, K. Arfanakis, G.J. Wendt, P.A. Turski, C.H. Moritz, M.A. Quigley, M.E. Meyerand, Mapping functionally related regions of brain with functional connectivity MR Imaging, Am. J. Neuroradiol. 21 (9) (2000) 1636–1644.
[20] A. Di Martino, C. Yan, Q. Li, E. Denio, F. Castellanos, K. Alaerts, J. Anderson, M. Assaf, S. Bookheimer, M. Dapretto, et al., The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism, Mol. Psychiatry 19 (2014) 659–667.
[21] P. Erdős, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17–61.
[22] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96–456 (2001) 1348–1360.
[23] O. Friman, C.-F. Westin, Resampling fMRI time series, NeuroImage 25 (2005) 859–867.
[24] I.I. Goldberg, M. Harel, R. Malach, When the brain loses its self: Prefrontal inactivation during sensorimotor processing, Neuron 50 (2) (2006) 329–339.
[25] A.Y. Hardan, R.R. Girgis, J. Adams, A.R. Gilbert, M.S. Keshavan, N.J. Minshew, Abnormal brain size effect on the thalamus in autism, Psychiatry Res. Neuro-Imaging 147 (2) (2006) 145–151.
[26] J. Honorio, D. Samaras, N. Paragios, R. Goldstein, L. Ortiz, Sparse and locally constant Gaussian graphical models, Adv. Neural Inf. Process. Syst. (2009) 745–753.
[27] H. Hotelling, New light on the correlation coefficient and its transforms, J. Roy. Statist. Soc. Ser. B 15 (1953) 193–232.
[28] S. Huang, J. Li, L. Sun, A. Fleisher, T. Wu, K. Chen, E. Reiman, Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation, Neuroimage 50–3 (2010) 935–949.
[29] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman, J. Ye, Learning brain connectivity of alzheimer's disease from neuro-imaging data, in: NIPS, vol. 22, 2009, pp. 808–816.
[30] S.A. Huettel, A.W. Song, G. McCarthy, Functional Magnetic Resonance Imaging, Vol. 1, Sinauer Associates Sunderland, MA, 2004.
[31] H. Koshino, P.A. Carpenter, N.J. Minshew, V.L. Cherkassky, T.A. Keller, M.A. Just, Functional connectivity in an fMRI working memory task in high-functioning autism, Neuroimage 24 (2005) 810–821.
[32] O. Ledoit, M. Wolf, Some hypothesis test for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (2002) 1081–1102.
[33] J. Li, S. Chen, Two sample tests for high-dimensional covariance matrices, Ann. Statist. 40 (2012) 908–940.
[34] M. Li, Y. Qin, Hypothesis testing for high-dimensional covariance matrices, J. Multivariate Anal. 128 (2014) 108–119.
[35] M. Lindquist, The statistical analysis of fMRI data, Statist. Sci. 23 (2008) 439–463.
[36] W. Liu, Gaussian graphical model estimation with false discovery rate control, Ann. Statist. 41 (2013) 2948–2978.
[37] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, H. Benali, Partial correlation for functional brain interactivity investigation in functional MRI, Neuroimage 32 (2006) 228–237.
[38] R. Mazumder, T. Hastie, The graphical lasso: New insights and alternatives, Electron. J. Stat. 6 (2012) 2125–2149.
[39] J.C. Mazziotta, A.W. Toga, A. Evans, P. Fox, J. Lancaster, A probabilistic atlas of the human brain: Theory and rationale for its development the international consortium for brain mapping (iCBM), Neuroimage 2 (2PA) (1995) 89–101.
[40] A.L. Oblak, T.T. Gibbs, G.J. Blatt, Reduced gaba receptors and benzodiazepine binding sites in the posterior cingulate cortex and fusiform gyrus in autism, Brain Res. 1380 (2011) 218–228.
[41] J. Richard, M. Yuan, Independent component analysis via nonparametric maximum likelihood estimation, Ann. Statist. 40 (2012) 2973–3002.
[42] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: Uses and interpretations, Neuroimage 52 (2010) 1059–1069.
[43] J. Schott, A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, Comput. Statist. Data Anal. 51 (2007) 6535–6542.
[44] O. Sporns, D.R. Chialvo, M. Kaiser, C.C. Hilgetag, Organization development and function of complex brain networks, Trends Cogn. Sci. 8 (2004) 418–425.
[45] T. Sun, C. Zhang, Scaled sparse linear regression, Biometrika 99 (2012) 879–898.
[46] K. Supekar, V. Menon, D. Rubin, M. Musen, M.D. Greicius, Network analysis of intrinsic functional brain connectivity in Alzheimer's disease, PLoS Comput. Biol. 4 (6) (2008) e1000100.
[47] R.J. Tibshirani, Regression shrinkage and selection via the lasso, J. Roy. Statist. Soc. Ser. B 58 (1996) 267–288.
[48] K.D. Tsatsanis, B.P. Rourke, A. Klin, F.R. Volkmar, D. Cicchetti, R.T. Schultz, Reduced thalamic volume in high-functioning individuals with autism, Biol. Psychiatry 53 (2003) 121–129.
[49] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, Neuroimage 15 (1) (2002) 273–289.
[50] B. Velioglu, E. Aksan, I. Onal, O. Firat, M. Ozay, F. Yarman Vural, Functional networks of anatomic brain regions, in: 2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing. http://dx.doi.org/10.1109/ICCI-CC.2014.6921441.
[51] Y. Weiss, W.T. Freeman, Correctness of belief propagation in Gaussian graphical models of arbitrary topology, Neural Comput. 13 (2001) 2173–2200.
[52] K.J. Worsley, C. Liao, J. Aston, V. Petre, G. Duncan, F. Morales, A. Evans, A general statistical analysis for fMRI data, Neuroimage 15 (2002) 1–15.
[53] H. Zou, The adaptive lasso and its oracle properties, J. Amer. Statist. Assoc. 101 (2006) 1418–1429.