Biostatistics (2021) **22**, 2, *pp.* 402–420
doi:10.1093/biostatistics/kxz037
Advance Access publication on October 21, 2019

Paired test of matrix graphs and brain connectivity analysis

YUTING YE[†]

Department of Biostatistics and Epidemiology, University of California at Berkeley, 2121 Berkeley Way, Berkeley, CA 94720-7360, USA

YIN XIA[†]

Department of Statistics, School of Management, Fudan University, 220 Handan Rd, Wu Jiao Chang, Yangpu, Shanghai 200433, China

LEXIN LI[*]

Department of Biostatistics and Epidemiology, University of California at Berkeley, 2121 Berkeley Way, Berkeley, CA 94720-7360, USA

lexinli@berkeley.edu

Summary

Inferring brain connectivity network and quantifying the significance of interactions between brain regions are of paramount importance in neuroscience. Although there have recently emerged some tests for graph inference based on independent samples, there is no readily available solution to test the change of brain network for paired and correlated samples. In this article, we develop a paired test of matrix graphs to infer brain connectivity network when the groups of samples are correlated. The proposed test statistic is both bias corrected and variance corrected, and achieves a small estimation error rate. The subsequent multiple testing procedure built on this test statistic is guaranteed to asymptotically control the false discovery rate at the pre-specified level. Both the methodology and theory of the new test are considerably different from the two independent samples framework, owing to the strong correlations of measurements on the same subjects before and after the stimulus activity. We illustrate the efficacy of our proposal through simulations and an analysis of an Alzheimer's Disease Neuroimaging Initiative dataset.

Keywords: Brain connectivity analysis; Gaussian graphical model; Matrix variate normal distribution; Multiple testing; Partial correlation; Variance correction.

1. Introduction

Brain functional connectivity reveals the intrinsic functional architecture of brains by measuring correlations in neurophysiological recordings of brain activities (Varoquaux and Craddock, 2013). Numerous studies have found that functional connectivity alters for individuals with neurological disorders, such as

[†]The first two authors share the co-first authorship.
[*]To whom correspondence should be addressed.

Alzheimer's diseases (AD) and autism spectrum disorder (Hedden *and others*, 2009; Rudie *and others*, 2013), or after experiencing stimulus activities such as stress or therapy (Peck *and others*, 2004; van Marle *and others*, 2010). The brain connectivity network is believed to hold crucial insight to help understand the pathologies of neurological disorders and to develop targeting treatment (Fox and Greicius, 2010; Quaedflieg *and others*, 2015).

Brain functional connectivity is commonly encoded as a network, or graph, with nodes representing brain regions, and links representing interactions and correlations between regions. Among multiple correlation measures, partial correlation is a well-accepted and frequently used metric, and correspondingly, the connectivity network is portrayed by a partial correlation matrix (Ryali *and others*, 2012; Chen *and others*, 2013). Current mainstream imaging modalities to study functional connectivity include electroencephalography (EEG), electrocorticography (ECoG), and resting-state functional magnetic resonance imaging (fMRI). After proper preprocessing, the resulting imaging data for each subject is summarized in the form of a location by time matrix, from which a partial correlation matrix is constructed to characterize brain connectivity.

A central problem in connectivity analysis is inference. Unlike network *estimation* (Ahn *and others*, 2015; Chen *and others*, 2015; Kang *and others*, 2016a; Qiu *and others*, 2016; Wang *and others*, 2016), network *inference* aims to directly quantify the statistical significance of individual links or their differences, meanwhile explicitly controlling for the false discovery. Recently there have been proposals of partial correlation matrix-based network inference for vector-valued data following a normal distribution (Liu *and others*, 2013; Xia *and others*, 2015), or matrix-valued data following a matrix normal distribution (Chen and Liu, 2019; Xia and Li, 2017, 2019). For brain connectivity analysis, the data obtained from EEG, ECoG, or fMRI are of a matrix form, and the primary scientific interest is on the spatial but not the temporal correlation patterns of the brain. Directly applying the tests for vector-valued data to infer the spatial patterns ignores the temporal correlations among the columns of the matrix data, and is to result in distorted test size and false discovery rate (FDR) (Xia and Li, 2017). Whitening can alleviate this problem, and in effect transforms the matrix data back to the vector case (Narayan *and others*, 2015). However, it does not utilize the data efficiently, would result in loss of power, and is also computationally intensive (Xia and Li, 2019). Alternatively, Chen and Liu (2019) and Xia and Li (2017) directly tackled inference of the matrix-valued data under the one-sample testing scenario, and Xia and Li (2019) tackled the two-sample scenario where the two groups of samples are independent.

In addition to inference about brain network alternation across independent subject groups, it is of equal interest and importance to infer the change of brain network of the same group of subjects before and after a "stimulus" activity, which could be a treatment, a disease conversion, or a different experimental condition. For instance, Peck *and others* (2004) studied brain connectivity activities in auditory and motor cortices of aphasic patients before and after a therapy. Gianaros *and others* (2008), van Marle *and others* (2010), and Quaedflieg *and others* (2015) studied amygdala-centered connectivity patterns in healthy subjects before and after the experimentally induced stress. Cai *and others* (2015) studied alterations in brain functional networks in patients with primary angle-closure glaucoma before and after the surgery. Kang *and others* (2016b) studied brain connectivity activities in left and right inferior frontal gyrus areas of the same subjects under different sleeping conditions. Ficek *and others* (2018) studied changes of functional connectivity before and after a language intervention therapy. In Section 5, we aim to identify the connectivity patterns that differ before and after a patient converted to AD. The two-sample test of Xia and Li (2019) does not directly apply to those studies, because of the strong correlations of brain measurements on the *same* subjects before and after the stimulus. For instance, a positive correlation before and after the stimulus would reduce the variance of the partial correlation difference between the two groups, causing the two-sample test to overestimate the variance and resulting in a low test power. On the contrary, a negative correlation would inflate the variance, causing the two-sample test to underestimate the variance and resulting in an inflated false discovery.

In this article, we develop a paired test of matrix graphs to infer brain connectivity network when the groups of samples are correlated, such as in the scenario of before and after the stimulus. The key of our proposal is an innovative variance correction procedure that incorporates the spatial and temporal dependency between the paired samples. The proposed test statistic is both bias corrected and variance corrected, and is shown to achieve a sufficiently small estimation error rate asymptotically. This in turn ensures that the subsequent multiple testing procedure built on this test statistic can asymptotically control the FDR at the pre-specified level. Our proposal extends the two-sample test of Xia and Li (2019), but is considerably different. This extension is far from trivial, and the theoretical investigation of the paired test is much more involved, as one needs to carefully evaluate both within-sample and between-sample correlations. To our knowledge, there is no existing graph inference procedure for paired matrix samples, and our proposal offers a timely response to an important problem of both scientific and methodological interest.

The rest of the article is organized as follows. Section 2 presents the formulation of the hypothesis testing problem, the proposed test statistic, the variance correction procedure for the paired samples, and the multiple testing procedure. Section 3 studies the corresponding asymptotic properties. Section 4 examines the empirical performance of the proposed test through simulations, and Section 5 analyzes a real fMRI dataset. The Supplementary material available at *Biostatistics* online collects all the technical assumptions, proofs, and additional numerical results.

2. PAIRED TEST

2.1. *Problem formulation*

Let $\boldsymbol{X}^{(t)}$ denotes the $p \times q$ matrix observed at time point $t, t = 1, 2$. In brain connectivity analysis, $\boldsymbol{X}^{(t)}$ denotes the spatial–temporal imaging data before ($t = 1$) and after ($t = 2$) a stimulus activity or conversion, and each $\boldsymbol{X}^{(t)}$ corresponds to p brain regions and the time course data of each region is of length q. We assume $\{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}\}$ follows a matrix normal distribution, i.e.,

$$\begin{pmatrix} \mathrm{vec}\{\boldsymbol{X}^{(1)}\} \\ \mathrm{vec}\{\boldsymbol{X}^{(2)}\} \end{pmatrix} \sim \mathrm{Normal}\left(\boldsymbol{0}_{2pq}, \boldsymbol{\Sigma}\right), \text{ with } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{S_1} \otimes \boldsymbol{\Sigma}_{T_1} & \boldsymbol{\Sigma}_{S_{1,2}} \otimes \boldsymbol{\Sigma}_{T_{1,2}} \\ \boldsymbol{\Sigma}_{S_{1,2}}^{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{T_{1,2}}^{\mathrm{T}} & \boldsymbol{\Sigma}_{S_2} \otimes \boldsymbol{\Sigma}_{T_2} \end{pmatrix}. \tag{2.1}$$

Without loss of generality, the mean is assumed to be zero, $\otimes$ is the Kronecker product, and $\mathrm{vec}(\cdot)$ is the operator that stacks the columns of a matrix into a vector. Furthermore, $\boldsymbol{\Sigma}_{S_t} \in \mathbb{R}^{p\times p}$ denotes the covariance matrix of the spatial regions, $\boldsymbol{\Sigma}_{T_t} \in \mathbb{R}^{q\times q}$ denotes the temporal covariance matrix of the time course data, at $t = 1, 2$, respectively, and $\boldsymbol{\Sigma}_{S_{1,2}}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ denote the between-sample spatial and temporal covariance, respectively. When $\boldsymbol{\Sigma}_{S_{1,2}} \otimes \boldsymbol{\Sigma}_{T_{1,2}} = \boldsymbol{0}$, (2.1) reduces to the independent two-sample setting of Xia and Li (2019). We remark that, the matrix normal distribution has been frequently adopted in numerous applications involving matrix-valued data (Yin and Li, 2012; Leng and Tang, 2012), and is also scientifically plausible in neuroimaging analysis (Smith *and others*, 2004; Friston *and others*, 2007). Moreover, Aston *and others* (2017) developed a test to check if the data conform with the Kronecker product structure. In Section 4.2, we further carry out sensitivity analysis, and show that our proposed test works reasonably well even when the data deviate from the matrix normal distribution (2.1).

Let $\boldsymbol{\Omega}_{S_t} = \boldsymbol{\Sigma}_{S_t}^{-1} = (\omega_{S_t,i,j})_{i,j=1}^{p}$ denote the spatial precision matrix, $\boldsymbol{D}_{S_t}$ denotes the diagonal matrix of $\boldsymbol{\Omega}_{S_t}$ and $\boldsymbol{R}_{S_t} = \boldsymbol{D}_{S_t}^{-1/2}\boldsymbol{\Omega}_{S_t}\boldsymbol{D}_{S_t}^{-1/2} = (\rho_{S_t,i,j})_{i,j=1}^{p}$ denote the spatial partial correlation matrix. In brain connectivity analysis, the primary interest is to infer the connectivity network characterized by the spatial partial correlation matrix. The temporal covariance or precision matrix is of little interest in this context and is to

be treated as a nuisance parameter. Consequently, we formulate our inference problem as simultaneously testing

$$H_{0,i,j} : \rho_{S_1,i,j} = \rho_{S_2,i,j} \text{ versus } H_{1,i,j} : \rho_{S_1,i,j} \neq \rho_{S_2,i,j}, \quad \text{for } 1 \leq i < j \leq p. \tag{2.2}$$

We next derive the test statistic and the associated variance correction to account for the correlations of the paired samples.

2.2. *Test statistic*

Consider n pairs of samples $\left\{\boldsymbol{X}_k^{(1)}, \boldsymbol{X}_k^{(2)}\right\}_{k=1}^{n}$ from the joint distribution (2.1). To construct the test statistic for (2.2), we first remove the temporal correlations by the linear transformation $\boldsymbol{Y}_k^{(t)} = \boldsymbol{X}_k^{(t)} \boldsymbol{\Sigma}_{T_t}^{-1/2}, k = 1, \ldots, n$, $t = 1, 2$, and

$$\begin{pmatrix} \text{vec}\{\boldsymbol{Y}^{(1)}\} \\ \text{vec}\{\boldsymbol{Y}^{(2)}\} \end{pmatrix} \sim \text{Normal}\left(\boldsymbol{0}_{2pq}, \begin{pmatrix} \boldsymbol{\Sigma}_{S_1} \otimes \boldsymbol{I}_q & \boldsymbol{\Sigma}_{S_{1,2}} \otimes \boldsymbol{P}_{T_{1,2}} \\ \boldsymbol{\Sigma}_{S_{1,2}}^{\mathrm{T}} \otimes \boldsymbol{P}_{T_{1,2}}^{\mathrm{T}} & \boldsymbol{\Sigma}_{S_2} \otimes \boldsymbol{I}_q \end{pmatrix}\right), \tag{2.3}$$

where $\boldsymbol{P}_{T_{1,2}} = \boldsymbol{\Sigma}_{T_1}^{-1/2} \boldsymbol{\Sigma}_{T_{1,2}} \boldsymbol{\Sigma}_{T_2}^{-1/2}$ denotes the between-sample temporal covariance matrix of the transformed samples. Clearly, for the independent case, $\boldsymbol{\Sigma}_{S_{1,2}} \otimes \boldsymbol{P}_{T_{1,2}} = 0$. In practice, $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ are generally unknown. There are multiple ways to estimate $\boldsymbol{\Sigma}_{T_t}$, or equivalently, $\boldsymbol{\Omega}_{T_t} = \boldsymbol{\Sigma}_{T_t}^{-1}$. Examples include the sample covariance estimator, the banded covariance estimator (Bickel and Levina, 2008), the adaptive thresholding estimator (Cai and Liu, 2011) for $\boldsymbol{\Sigma}_{T_t}$, or the Clime estimator (Cai *and others*, 2011) for $\boldsymbol{\Omega}_{T_t}$. We adopt the banded estimator in this article, given its competitive performance in both the one-sample test and the independent two-sample test under the matrix normal distribution (Xia and Li, 2017, 2019). In the following, we first derive the test statistic with known $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{P}_{T_{1,2}}$, which helps simplify the notations considerably. We then extend it by plugging in an estimated $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{P}_{T_{1,2}}$. Accordingly, we will add the superscript (d) in the resulting statistics to represent this scenario when $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{P}_{T_{1,2}}$ are estimated given the data. In Section 3, we show that the test statistics under the known $\boldsymbol{\Sigma}_{T_t}$, $\boldsymbol{P}_{T_{1,2}}$ and the estimated $\boldsymbol{\Sigma}_{T_t}$, $\boldsymbol{P}_{T_{1,2}}$ have the same asymptotic property. Consequently, they lead to the same multiple testing procedure with the guaranteed asymptotic control of false discovery.

The construction of our test statistic is based on the fact that, under the normal distribution, the precision matrix can be described through the regression model (Anderson, 2003),

$$Y_{k,i,l}^{(t)} = \boldsymbol{Y}_{k,-i,l}^{(t)\mathrm{T}} \boldsymbol{\beta}_i^{(t)} + \epsilon_{k,i,l}^{(t)}, \quad 1 \leq i \leq p,\ 1 \leq l \leq q,\ 1 \leq k \leq n,\ t = 1, 2, \tag{2.4}$$

where the error term $\epsilon_{k,i,l}^{(t)} \sim N\left(0, \sigma_{S_t,i,i} - \boldsymbol{\Sigma}_{S_t,i,-i} \boldsymbol{\Sigma}_{S_t,-i,-i}^{-1} \boldsymbol{\Sigma}_{S_t,-i,i}\right)$ and is independent of $\boldsymbol{Y}_{k,-i,l}^{(t)}$, and the subscript $-i$ means the ith entry is removed from a vector, or the ith row or column removed from a matrix. The regression coefficient $\boldsymbol{\beta}_i^{(t)}$ can be estimated using Lasso or other methods, as long as the estimator $\hat{\boldsymbol{\beta}}_i^{(t)}$ satisfies the regularity condition (A5) or (A6) in the Supplementary material available at *Biostatistics* online. See Xia and Li (2017, 2019) and Section S4 of the Supplementary material available at *Biostatistics* online for a more detailed discussion on estimation of $\boldsymbol{\beta}_i^{(t)}$ and the associated tuning procedure. Moreover, the error term satisfies that $r_{i,j}^{(t)} = \text{cov}\left\{\epsilon_{k,i,l}^{(t)}, \epsilon_{k,j,l}^{(t)}\right\} = \omega_{S_t,i,j}/(\omega_{S_t,i,i}\omega_{S_t,j,j})$. Therefore the element $\omega_{S_t,i,j}$ of the spatial precision matrix $\boldsymbol{R}_{S_t}$, and in turn, the element $\rho_{S_t,i,j}$ of the spatial partial correlation matrix $\boldsymbol{R}_{S_t}$

can be represented in terms of $r_{i,j}^{(t)}$. Following Xia and Li (2017, 2019), a bias-corrected estimator of $r_{i,j}^{(t)}$ is obtained from fitting the regression model (2.4),

$$\hat{r}_{i,j}^{(t)} = \begin{cases} -\tilde{r}_{i,j}^{(t)} - \tilde{r}_{i,i}^{(t)}\hat{\beta}_{i,j}^{(t)} - \tilde{r}_{j,j}^{(t)}\hat{\beta}_{j-1,i}^{(t)}, & \text{when } 1 \le i < j \le p \\ \tilde{r}_{i,i}^{(t)}, & \text{when } 1 \le i = j \le p, \end{cases}$$

where $\tilde{r}_{i,j}^{(t)} = (nq)^{-1}\sum_{k=1}^{n}\sum_{l=1}^{q}\hat{\epsilon}_{k,i,l}^{(t)}\hat{\epsilon}_{k,j,l}^{(t)}$ is the sample covariance between the residuals, $\hat{\epsilon}_{k,i,l}^{(t)} = Y_{k,i,l}^{(t)} - \bar{Y}_{i,l}^{(t)} - (\boldsymbol{Y}_{k,-i,l}^{(t)} - \bar{\boldsymbol{Y}}_{\cdot,-i,l}^{(t)})^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{i}^{(t)}$, $\bar{Y}_{i,l}^{(t)} = n^{-1}\sum_{k=1}^{n} Y_{k,i,l}^{(t)}$, and $\bar{\boldsymbol{Y}}_{\cdot,-i,l}^{(t)} = n^{-1}\sum_{k=1}^{n}\boldsymbol{Y}_{k,-i,l}^{(t)}$. Based on the estimator $\hat{r}_{i,j}^{(t)}$, we further obtain a bias-corrected estimator of the element $\rho_{S_t,i,j}$ of the spatial partial correlation matrix $\boldsymbol{R}_{S_t}$ as

$$\hat{\rho}_{S_t,i,j} = \hat{r}_{i,j}^{(t)}/\{\hat{r}_{i,i}^{(t)}\hat{r}_{j,j}^{(t)}\}^{1/2}, \quad 1 \le i < j \le p,\ t = 1, 2.$$

We then construct our test statistic for the pair of hypotheses (2.2) as

$$W_{i,j} = \frac{(\hat{\rho}_{S_1,i,j} - \hat{\rho}_{S_2,i,j})}{\hat{\Theta}_{i,j}^{1/2}}, \quad 1 \le i < j \le p,$$

where $\hat{\Theta}_{i,j}$ is an estimator of var $\left(\hat{\rho}_{S_1,i,j} - \hat{\rho}_{S_2,i,j}\right)$. We next develop such an estimator that incorporates the between-sample dependency of the paired samples.

2.3. *Variance correction*

We first recognize that the expression for the variance term var $\left(\hat{\rho}_{S_1,i,j} - \hat{\rho}_{S_2,i,j}\right)$ is quite involved. To alleviate this issue, we introduce an intermediate term, $\tilde{U}_{i,j}^{(t)} = \left\{r_{i,j}^{(t)} - U_{i,j}^{(t)}\right\} / \left\{r_{i,i}^{(t)}r_{j,j}^{(t)}\right\}^{\frac{1}{2}}$, where $U_{i,j}^{(t)} = (nq)^{-1}\sum_{k=1}^{n}\sum_{l=1}^{q}\left[\epsilon_{k,i,l}^{(t)}\epsilon_{k,j,l}^{(t)} - E\left\{\epsilon_{k,i,l}^{(t)}\epsilon_{k,j,l}^{(t)}\right\}\right]$. Lemma S.2.1 in the Supplementary material available at *Biostatistics* online implies that the difference between $\hat{\rho}_{S_t,i,j}$ and $\tilde{U}_{i,j}^{(t)}$ is negligible. Consequently, we estimate var $\left(\hat{\rho}_{S_1,i,j} - \hat{\rho}_{S_2,i,j}\right)$ by developing an estimator for

$$\Theta_{i,j} = \text{var}\left\{\tilde{U}_{i,j}^{(1)} - \tilde{U}_{i,j}^{(2)}\right\}.$$

For the independent two-sample setting, $\Theta_{i,j} = \theta_{i,j}^{(1)} + \theta_{i,j}^{(2)}$, where $\theta_{i,j}^{(t)} = \text{var}\left\{\tilde{U}_{i,j}^{(t)}\right\}, t = 1, 2$. Based further on the observation that $\text{var}\left\{\tilde{U}_{i,j}^{(t)}\right\} = \text{var}\left[\epsilon_{k,i,l}^{(t)}\epsilon_{k,j,l}^{(t)}/\{r_{i,i}^{(t)}r_{j,j}^{(t)}\}^{1/2}\right]/(nq) = \left[1 + \{\beta_{i,j}^{(t)}\}^2 r_{i,i}^{(t)}/r_{j,j}^{(t)}\right]/(nq)$, we estimate $\theta_{i,j}^{(t)}$ by

$$\hat{\theta}_{i,j}^{(t)} = \frac{1}{nq}\left[1 + \left\{\hat{\beta}_{i,j}^{(t)}\right\}^2 \hat{r}_{i,i}^{(t)}/\hat{r}_{j,j}^{(t)}\right], \quad 1 \le i < j \le p,\ t = 1, 2. \tag{2.5}$$

For the paired samples, however, it is crucial to account for the between-sample spatial–temporal dependency as presented in $\boldsymbol{\Sigma}_{S_{1,2}}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ when estimating $\Theta_{i,j}$. Next, we derive such an estimator of $\Theta_{i,j}$. Later in Section 3, we show that this estimator is accurate, in the sense that its scaled version achieves an $o_p(1/\log p)$ convergence rate. This error rate is essential for the subsequent asymptotic false discovery control in multiple testing.

The next proposition gives an explicit expression of $\Theta_{i,j}$ under the dependent setting. Its proof is given in the Supplementary material available at *Biostatistics* online. The key is the separable spatial and temporal dependence structures between the paired samples, and the decoupling of $\rho^{(1,2)}_{i,j;l_1,l_2} = \text{corr}\left\{\epsilon^{(1)}_{k,i,l_1}, \epsilon^{(2)}_{k,j,l_2}\right\}$ as $\rho^{(1,2)}_{i,j;l_1,l_2} = \rho_{S_{1,2},i,j}\boldsymbol{P}_{T_{1,2},l_1,l_2}$, where $\rho_{S_{1,2},i,j} = \sqrt{r^{(1)}_{i,i}r^{(2)}_{j,j}}\ \boldsymbol{\Omega}_{S_1,i,\cdot}\boldsymbol{\Sigma}_{S_{1,2}}\boldsymbol{\Omega}_{S_2,\cdot,j}$ accounts for the spatial correlation, and $\boldsymbol{P}_{T_{1,2},l_1,l_2}$ captures the temporal dependency. Here $\boldsymbol{\Omega}_{S_1,i,\cdot}$ denotes the ith row of the matrix $\boldsymbol{\Omega}_{S_1}$, and $\boldsymbol{\Omega}_{S_2,\cdot,j}$ denotes the jth column of $\boldsymbol{\Omega}_{S_2}$.

PROPOSITION 2.1 Under the data distribution (2.3), we have,

$$\Theta_{i,j} = \theta^{(1)}_{i,j} + \theta^{(2)}_{i,j} - \frac{2}{nq^2}\left(\rho_{S_{1,2},i,i}\rho_{S_{1,2},j,j} + \rho_{S_{1,2},i,j}\rho_{S_{1,2},j,i}\right)\|\boldsymbol{P}_{T_{1,2}}\|_F^2, \tag{2.6}$$

for $1 \le i < j \le p$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Define $\varrho^{(1,2)}_{i,j} = \rho_{S_{1,2},i,j} \cdot \text{tr}(\boldsymbol{P}_{T_{1,2}})/q$, which is the correlation coefficient $\rho_{S_{1,2},i,j}$ scaled by the term $\text{tr}(\boldsymbol{P}_{T_{1,2}})/q$, and $\text{tr}(\cdot)$ denotes the matrix trace. We observe that

$$\mathbb{E}\left\{\frac{1}{nq}\sum_{k=1}^{n}\sum_{l=1}^{q}\epsilon^{(1)}_{k,i,l}\epsilon^{(2)}_{k,j,l}\right\} = \sqrt{r^{(1)}_{i,i}r^{(2)}_{j,j}}\ \rho_{S_{1,2},i,j}\ \text{tr}(\boldsymbol{P}_{T_{1,2}})/q = \sqrt{r^{(1)}_{i,i}r^{(2)}_{j,j}}\ \varrho^{(1,2)}_{i,j}.$$

Therefore, we can estimate $\varrho^{(1,2)}_{i,j}$ by

$$\hat{\varrho}^{(1,2)}_{i,j} = \hat{\text{cov}}(\epsilon^{(1)}_{\cdot,i,\cdot}, \epsilon^{(2)}_{\cdot,j,\cdot})/\sqrt{\hat{r}^{(t)}_{i,i}\hat{r}^{(t)}_{j,j}}, \quad \text{and} \quad \hat{\text{cov}}(\epsilon^{(1)}_{\cdot,i,\cdot}, \epsilon^{(2)}_{\cdot,j,\cdot}) = \frac{1}{nq}\sum_{k=1}^{n}\sum_{l=1}^{q}\hat{\epsilon}^{(1)}_{k,i,l}\hat{\epsilon}^{(2)}_{k,j,l}. \tag{2.7}$$

Correspondingly, when $\boldsymbol{\Sigma}_{T_t}$, $\boldsymbol{\Sigma}_{T_{1,2}}$ and thus $\boldsymbol{P}_{T_{1,2}}$ are known, we can estimate $\Theta_{i,j}$ by

$$\hat{\Theta}_{i,j} = \hat{\theta}^{(1)}_{i,j} + \hat{\theta}^{(2)}_{i,j} - \frac{2}{nq}\left\{\hat{\varrho}^{(1,2)}_{i,i}\hat{\varrho}^{(1,2)}_{j,j} + \hat{\varrho}^{(1,2)}_{i,j}\hat{\varrho}^{(1,2)}_{j,i}\right\}\frac{q\|\boldsymbol{P}_{T_{1,2}}\|_F^2}{\text{tr}(\boldsymbol{P}_{T_{1,2}})^2}. \tag{2.8}$$

We show in Section 3 that $\hat{\Theta}_{i,j}$ in (2.8) provides an accurate estimation of $\Theta_{i,j}$, with an error rate of order $o_p(1/\log p)$, when $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ are known.

When $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ are unknown, we first estimate $\boldsymbol{P}_{T_{1,2}}$ by

$$\hat{\boldsymbol{P}}^{(d)}_{T_{1,2}} = \frac{1}{np}\sum_{k=1}^{n}\sum_{i=1}^{p}\left[\left\{\boldsymbol{Y}^{(1,d)}_{k,i,\cdot} - \frac{1}{np}\sum_{k=1}^{n}\sum_{i=1}^{p}\boldsymbol{Y}^{(1,d)}_{k,i,\cdot}\right\}^{\mathrm{T}}\left\{\boldsymbol{Y}^{(2,d)}_{k,i,\cdot} - \frac{1}{np}\sum_{k=1}^{n}\sum_{i=1}^{p}\boldsymbol{Y}^{(2,d)}_{k,i,\cdot}\right\}\right], \tag{2.9}$$

where $\boldsymbol{Y}^{(t,d)}_{k,i,\cdot}$ is the ith row of $\boldsymbol{Y}^{(t,d)}_{k} = \boldsymbol{X}^{(t)}_{k}\hat{\boldsymbol{\Sigma}}^{-1/2}_{T_t}$, and $\hat{\boldsymbol{\Sigma}}_{T_t}$ is an estimator of $\boldsymbol{\Sigma}_{T_t}$. We then plug (2.9) into (2.8). Again we show in Section 3 that this estimator also provides an accurate estimation of $\Theta_{i,j}$, with an error rate of order $o_p(1/\log p)$, when $\boldsymbol{\Sigma}_{T_t}$ and $\boldsymbol{\Sigma}_{T_{1,2}}$ are unknown.

We make a few remarks about our proposed variance correction. First, a crucial component of our method is to pool data information of the p-dimensional spatial and q-dimensional temporal measurements of n subjects in our estimations. The data pooling is possible due to the facts that $\mathbb{E}\{\boldsymbol{Y}^{(1)} \cdot (\boldsymbol{Y}^{(2)})^{\mathrm{T}}\} = \text{tr}(\boldsymbol{P}_{T_{1,2}}) \cdot \boldsymbol{\Sigma}_{S_{1,2}}$ and $\mathbb{E}\{(\boldsymbol{Y}^{(1)})^{\mathrm{T}} \cdot (\boldsymbol{Y}^{(2)})\} = \text{tr}(\boldsymbol{\Sigma}_{S_{1,2}}) \cdot \boldsymbol{P}_{T_{1,2}}$. Consequently, we can pool the columns of $\boldsymbol{Y}^{(t)}$

to estimate $\mathbf{\Sigma}_{S_{1,2}}$, and the rows of $\boldsymbol{Y}^{(t)}$ to estimate $\boldsymbol{P}_{T_{1,2}}$, up to a constant. More specifically, when $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are known, we pool $2nq$ samples to estimate the within-sample variance as in (2.5), and the between-sample spatial dependency as in (2.7) and (2.8). When $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are unknown, we also pool $2np$ samples to obtain the estimates $\hat{\mathbf{\Sigma}}_{T_t}^{-1/2}, t = 1, 2$, and estimate the temporal dependency between the before-stimulus scan and the after-stimulus scan as in (2.9). Such data pooling is the main difference between our method and a naive solution, which estimates the dependency between the paired samples by the usual sample covariance, namely, estimating $\operatorname{cov}\left\{\epsilon_{k,i,l_1}^{(1)}, \epsilon_{k,j,l_2}^{(2)}\right\}$ by $n^{-1}\sum_{k=1}^{n}\hat{\epsilon}_{k,i,l_1}^{(1)}\hat{\epsilon}_{k,j,l_2}^{(2)}$, for each $1 \le i < j \le p, 1 \le l_1, l_2 \le q$. Note that, the latter approach only uses n observations to estimate the dependence structure without any data pooling, and as a result, it cannot guarantee the estimation error rate required to ensure the performance of the test.

Second, we note that the spatial and temporal covariances $\mathbf{\Sigma}_{S_{1,2}}$ and $\boldsymbol{P}_{T_{1,2}}$ are only identifiable up to a constant. However, this does not affect our test statistic, nor our variance estimation. This is because, when replacing $(\mathbf{\Sigma}_{S_{1,2}}, \boldsymbol{P}_{T_{1,2}})$ with $(c\mathbf{\Sigma}_{S_{1,2}}, \boldsymbol{P}_{T_{1,2}}/c)$, where c is any positive factor, the terms $\varrho_{i,j}^{(1,2)}$ and $\|\boldsymbol{P}_{T_{1,2}}\|_F^2/\operatorname{tr}(\boldsymbol{P}_{T_{1,2}})^2$ remain the same, in which the factor c is canceled.

Third, Chen and Liu (2018) developed a variance correction method for matrix-valued data, but for a single group of samples. In contrast, we perform variance correction for two stages of samples from the same population. We first separate the spatial and temporal structures, so that the resulting test statistics do not require variance correction within each sample. Our variance correction differs from that of Chen and Liu (2018) considerably. On the other hand, if the temporal covariance between two stages has some particular structure, e.g., if it is sparse, then the method of Chen and Liu (2018) may be applied to our procedure, by thresholding $\hat{\boldsymbol{P}}_{T_{1,2}}^{(d)}$ in (2.9) accordingly. In this article, however, we do not impose any structural condition on the temporal dependence, and thus we use the general sample covariance estimator in (2.9) instead.

2.4. *Multiple testing*

We next develop a multiple testing procedure for $H_{0,i,j} : \rho_{S_1,i,j} = \rho_{S_2,i,j}, 1 \le i < j \le p$, so to identify spatial locations with their conditional dependence changed before and after the stimulus. With a total of $p(p-1)/2$ simultaneous tests, the key is to control false discovery. Let h be the rejection threshold value such that $H_{0,i,j}$ is rejected if $|W_{i,j}| \ge h$, and $\mathcal{H}_0 := \{(i,j) : \rho_{S_1,i,j} = \rho_{S_2,i,j}, 1 \le i < j \le p\}$ be the set of true nulls. Then the false discovery proportion (FDP) and the FDR are computed as

$$\mathrm{FDP}(h) = \frac{\sum_{(i,j)\in\mathcal{H}_0} I(|W_{i,j}| \ge h)}{\sum_{1\le i<j\le p} I(|W_{i,j}| \ge h) \vee 1}, \quad \mathrm{FDR}(h) = \mathbb{E}\{\mathrm{FDP}(h)\}.$$

Our multiple testing procedure is based on the test statistic $W_{i,j}$ derived in Section 2.2, with the corrected variance estimates $\hat{\Theta}_{i,j}$ derived in Section 2.3. The rest of the procedure is similar to that of the two-sample independent test of Xia and Li (2019). We thus only outline the main steps here. First, we compute the paired-test statistics $W_{i,j}$ in (2.5) for all $1 \le i < j \le p$. Next we estimate the FDP by

$$\hat{\mathrm{FDP}}(h) = \frac{2\{1-\Phi(h)\}(p^2-p)/2}{\sum_{1\le i<j\le p} I(|W_{i,j}| \ge h)1},$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Here, we conservatively estimate $|\mathcal{H}_0|$ by $(p^2-p)/2$, as it is at maximum $(p^2-p)/2$ and is close to $(p^2-p)/2$ when $\boldsymbol{R}_{S_1} - \boldsymbol{R}_{S_2}$ is sparse. Next,

we compute the rejection threshold value $\hat{h}_\alpha$ under a given significance level α as

$$\hat{h}_\alpha = \inf\left\{0 \le h \le 2(\log p)^{1/2} : \hat{FDP}(h) \le \alpha\right\}. \tag{2.10}$$

If $\hat{h}_\alpha$ does not exist, we set $\hat{h}_\alpha = 2(\log p)^{1/2}$. Finally, we reject $H_{0,i,j}$ if and only if $|W_{i,j}| \ge \hat{h}_\alpha$ for each $1 \le i < j \le p$. In Section 3, we show that the above multiple testing procedure can control FDR at the pre-specified level asymptotically.

3. Theory

We study in this section the asymptotic properties of the proposed testing procedure. In the interest of space, we present all the regularity conditions (A1)–(A7) in the Supplementary material available at *Biostatistics* online. We first show that the corrected variance estimator of $\Theta_{i,j}$ we develop in Section 2.3 achieves the estimation error rate of $o_p(1/\log p)$. We then show that, based on such an error rate, the subsequent multiple testing procedure can control the false discovery asymptotically.

When $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are known, our variance estimator is $\hat{\Theta}_{i,j}$ as given in (2.8). The next proposition establishes its error rate.

Proposition 3.1 Suppose (A1), (A3) and (A5) hold. Then we have

$$\max_{i,j}|nq(\hat{\Theta}_{i,j} - \Theta_{i,j})| = o_p(1/\log p).$$

When $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are unknown, we denote our variance estimator as $\hat{\Theta}^{(d)}_{i,j}$, which is obtained by plugging the estimator $\hat{\boldsymbol{P}}^{(d)}_{T_{1,2}}$ in (2.9) into (2.8). The next proposition establishes its error rate.

Proposition 3.2 Suppose (A1), (A3), (A6), and (A7) hold, then we have

$$\max_{i,j}|nq(\hat{\Theta}^{(d)}_{i,j} - \Theta_{i,j})| = o_p(1/\log p).$$

The above two propositions show that, the variance estimation error is bounded by the same error rate asymptotically, when $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are unknown and when they are known.

The next theorem shows that, for the dependent samples, as long as the majority of the regression residuals are not highly correlated with each other under the null hypothesis, then the FDR can be controlled asymptotically at the pre-specified level α following the multiple testing procedure outlined in Section 2.4.

Theorem 3.1 Let $\ell_0 = |\mathcal{H}_0|$ and $\ell = (p^2 - p)/2$. Suppose $\ell_0 \ge \tilde{c}_0 p^2$ for some constant $\tilde{c}_0 > 0$, and $p \le \tilde{c}_1(nq)^{\tilde{c}_2}$ for some $\tilde{c}_1, \tilde{c}_2 > 0$. Let $\hat{h}_\alpha$ denote the threshold value in (2.10). Then, when (A1)-(A5) hold and $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are known, or when (A1)–(A4), (A6) and (A7) hold and $\mathbf{\Sigma}_{T_t}$ and $\mathbf{\Sigma}_{T_{1,2}}$ are unknown, we have

$$\frac{\mathrm{FDR}(\hat{h}_\alpha)}{\alpha\ell_0/\ell} \to 1, \quad \frac{\mathrm{FDP}(\hat{h}_\alpha)}{\alpha\ell_0/\ell} \xrightarrow{p} 1, \quad \text{as } (nq, p) \to \infty.$$

In addition to false discovery control, the asymptotic power analysis is another interesting problem. It relies on the specific structure of the connectivity network. In Section 4, we conduct extensive simulations to study the power of our test under numerous network structures, and we leave the theoretical power analysis as future research.

4. SIMULATIONS

4.1. *Empirical FDR and power with and without variance correction*

We conduct numerous simulations to study the finite sample performance of our proposed variance-corrected testing procedure. We also compare with the two-sample test of Xia and Li (2019), which ignores the correlation before and after the stimulus and does not correct the variance accordingly. In all the simulations, we use Lasso to estimate the regression coefficient $\beta_i^{(t)}$, and use the banded covariance approach to estimate $\mathbf{\Sigma}_{T_t}$. We set the FDR level at $\alpha = 1\%$.

We examine a set of spatial and temporal dimensions, $(p, q) \in \{(200, 50), (200, 200), (800, 200)\}$, while we fix the sample size at $n = 15$. We consider two temporal covariance structures: an autoregressive structure, where $\mathbf{\Sigma}_{T_t} = (\sigma_{T_t,i,j})$, $\sigma_{T_t,i,j} = 0.4^{|i-j|}$ if $t = 1$, and $\sigma_{T_t,i,j} = 0.5^{|i-j|}$ if $t = 2$, $1 \le i,j \le p$, and a moving average structure, where $\mathbf{\Sigma}_{T_t} = (\sigma_{T_t,i,j})$, $\sigma_{T_t,i,j} = 1/(|i-j|+1)$ for $|i-j| < 3$ if $t = 1$, and $\sigma_{T_t,i,j} = 1/(|i-j|+1)$ for $|i-j| \le 4$ if $t = 2$. We also consider three spatial covariance structures: a banded graph, with bandwidth equal to 3 (Zhao *and others*, 2012), a hub graph, with rows and columns evenly partitioned into 20 disjoint groups, and a small-world graph, with 5 starting neighbors and 5% probability of rewiring (van Wieringen and Peeters, 2016). We first generate $\mathbf{\Omega}_{S_1}$ according to one of the above spatial structures, then construct $\mathbf{\Omega}_{S_2}$ by randomly eliminating 50% of the edges of $\mathbf{\Omega}_{S_1}$.

Moreover, we consider two settings of correlation patterns before and after the stimulus. In Setting I, we set $\mathbf{\Sigma}_{S_{1,2}} = \gamma \mathbf{\Sigma}_{S_1}$, where γ is the overall correlation level and $|\gamma| \le 1$. Since γ plays its role through γ^2, its sign does not matter, and we choose $\gamma \in \{0, 0.2, 0.4, 0.6\}$. When $\gamma = 0$, it reduces to the two-sample independent case, whereas a larger value of γ implies a stronger before-and-after stimulus correlation. We next set $\boldsymbol{P}_{T_{1,2}}$ as a diagonal matrix with $\boldsymbol{P}_{T_{1,2},i,i} = -1$ if $i \equiv k \pmod{15}, k \in \{1, 3, 5\}$, and 1 otherwise. Here for three positive integers a, b and c, $a \equiv b \pmod{c}$ means that, when divided by c, a and b have the same remainder that is non-negative and smaller than c. In this setting, it follows that

$$(\rho^{(1,2)}_{S_{1,2},i,i}\rho^{(1,2)}_{S_{1,2},j,j} + \rho^{(1,2)}_{S_{1,2},i,j}\rho^{(1,2)}_{S_{1,2},j,i}) = \gamma^2\sqrt{r^{(1)}_{i,i}r^{(2)}_{i,i}r^{(1)}_{j,j}r^{(2)}_{j,j}}(\omega_{S_2,i,i}\omega_{S_2,j,j} + \omega_{S_2,i,j}\omega_{S_2,j,i}) > 0,$$

as long as $\gamma > 0$, where we utilize the facts that $\Sigma_{S_{1,2}} = \gamma\Sigma_{S_1} = \gamma\Omega^{-1}_{S_1}$, and Ω_{S_2} is a positive definitive matrix. Correspondingly, $\Theta_{i,j}$ is smaller than that of the independent case, and the test statistic $W_{i,j}$ would be larger than that without variance correction in its absolute value. For this setting, the two-sample test without variance correction is to yield a smaller power, as it is more conservative in rejecting the null hypothesis in this setting.

In Setting II, we set $\boldsymbol{P}_{T_{1,2}}$ in the same way, but set $\mathbf{\Sigma}_{S_{1,2},i,j} = \gamma \cdot \mathbf{\Sigma}_{S_1,i,j}(-1)^{i+j}$, if $i \ne j$, and $\gamma \cdot \mathbf{\Sigma}_{S_1,i,i}(1 - 2 \cdot \mathbf{1}[i \equiv k \pmod 7, k \in \{1, 3, 5\}])$ if $i = j$, where $\mathbf{1}(\cdot)$ is the indicator function. In this setting, we no longer have a simplified expression for $\rho^{(1,2)}_{S_{1,2},i,i}\rho^{(1,2)}_{S_{1,2},j,j} + \rho^{(1,2)}_{S_{1,2},i,j}\rho^{(1,2)}_{S_{1,2},j,i}$, but empirically, we have observed that this term is negative for about half of (i, j) pairs regardless of the choice of the spatial structure and the dimension p. For those pairs, $\Theta_{i,j}$ is larger than that of the independent case, and the test statistic $W_{i,j}$ would be smaller than that without variance correction in its absolute value. For this setting, the two-sample test without variance correction is to yield an overestimated FDR in this setting.

Tables 1 and 2 report the empirical FDR and the empirical power, both in percentage, out of 100 data replications for the two settings, respectively. We make the following observations.

For Setting I, when $(p, q) = (200, 50)$, the test with variance correction controls the FDR around the anticipated level of $\alpha = 1\%$, whereas the test without variance correction yields a much lower FDR than the significance level. Moreover, as the correlation strength γ increases, the power of the test with variance correction improves considerably compared to the test without correction. Similar qualitative patterns are observed for $(p, q) = (200, 200)$ and $(p, q) = (800, 200)$.

Table 1. The empirical FDR (with standard error in the parenthesis) and the empirical power, both in percentage, for Setting I

Temporal structure		Moving average						Autoregressive					
Spatial structure		Banded		Hub		Small		Banded		Hub		Small	
Variance correction		✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
(p, q)	γ	Empirical FDR (SE) (in %)											
(200, 50)	0.0	0.7(0.5)	0.7(0.5)	0.8(1.2)	0.7(1.2)	1.1(1.3)	1.1(1.2)	0.8(0.5)	0.8(0.5)	0.8(1.1)	0.8(1.1)	1.0(1.2)	0.9(1.2)
	0.2	0.8(0.6)	0.5(0.5)	1.1(1.3)	0.7(1.1)	1.1(1.1)	0.8(0.9)	0.7(0.4)	0.5(0.4)	1.0(1.5)	0.7(1.4)	1.0(1.0)	0.7(0.8)
	0.4	0.9(0.6)	0.1(0.2)	1.0(1.2)	0.1(0.5)	1.2(1.1)	0.3(0.6)	0.7(0.6)	0.1(0.2)	1.0(1.1)	0.2(0.6)	1.1(1.1)	0.3(0.7)
	0.6	0.9(0.6)	0.0(0.1)	1.1(1.2)	0.1(0.5)	1.3(0.8)	0.1(0.3)	0.9(0.6)	0.0(0.1)	1.1(1.3)	0.0(0.3)	1.3(1.0)	0.1(0.3)
(200, 200)	0.0	0.8(0.5)	0.8(0.5)	1.1(1.2)	1.1(1.2)	0.7(0.4)	0.7(0.4)	0.9(0.5)	0.9(0.5)	0.9(0.9)	0.8(0.9)	0.9(0.5)	0.9(0.4)
	0.2	0.9(0.6)	0.5(0.5)	1.1(1.3)	0.8(1.0)	0.9(0.4)	0.6(0.4)	0.8(0.5)	0.5(0.4)	1.2(1.2)	0.7(0.9)	0.9(0.4)	0.5(0.3)
	0.4	1.1(0.7)	0.2(0.2)	1.0(1.0)	0.2(0.4)	0.8(0.4)	0.2(0.2)	0.9(0.5)	0.2(0.2)	1.1(1.0)	0.3(0.5)	0.8(0.5)	0.2(0.2)
	0.6	1.1(0.6)	0.0(0.1)	1.1(1.2)	0.0(0.2)	1.0(0.4)	0.0(0.1)	1.0(0.6)	0.0(0.1)	1.5(1.2)	0.0(0.2)	0.9(0.5)	0.0(0.1)
(800, 200)	0.0	0.7(0.2)	0.7(0.2)	0.9(0.7)	0.9(0.6)	0.7(0.2)	0.7(0.2)	0.7(0.2)	0.7(0.2)	0.9(0.6)	0.9(0.6)	0.8(0.2)	0.8(0.2)
	0.2	0.8(0.3)	0.5(0.2)	0.9(0.6)	0.5(0.5)	0.8(0.2)	0.5(0.2)	0.7(0.3)	0.4(0.2)	0.9(0.7)	0.5(0.5)	0.8(0.2)	0.5(0.1)
	0.4	0.9(0.3)	0.1(0.1)	1.1(0.7)	0.3(0.4)	0.8(0.2)	0.1(0.1)	0.8(0.3)	0.1(0.1)	0.8(0.6)	0.3(0.4)	0.8(0.2)	0.1(0.1)
	0.6	0.9(0.3)	0.0(0.1)	1.1(0.8)	0.2(0.3)	0.9(0.2)	0.0(0.0)	0.9(0.3)	0.0(0.0)	1.1(0.6)	0.3(0.4)	0.9(0.2)	0.0(0.0)
		Empirical power (in %)											
(200, 50)	0.0	90.7	90.8	54.8	54.9	19.4	19.4	90.2	90.2	54.9	54.9	19.3	19.4
	0.2	93.0	91.2	58.6	55.4	22.8	18.9	92.8	91.0	55.2	51.1	23.8	19.7
	0.4	96.8	91.2	69.8	53.8	33.2	17.0	97.2	91.9	68.7	53.1	31.6	16.0
	0.6	99.1	92.4	85.3	43.0	45.6	14.8	99.3	93.4	86.4	42.4	46.5	15.5
(200, 200)	0.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	99.9	99.9
	0.2	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9
	0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
(800, 200)	0.0	100.0	100.0	62.0	62.2	99.6	99.6	100.0	100.0	61.1	61.4	99.5	99.5
	0.2	100.0	100.0	60.3	60.6	99.8	99.7	100.0	100.0	60.1	60.5	99.8	99.7
	0.4	100.0	100.0	57.7	50.1	100.0	99.8	100.0	100.0	57.7	50.2	99.9	99.8
	0.6	100.0	100.0	52.2	47.1	100.0	99.8	100.0	100.0	54.7	47.2	100.0	99.8

Table 2. The empirical FDR (with standard error in the parenthesis) and the empirical power, both in percentage, for Setting II

Temporal structure		Moving average						Autoregressive					
Spatial structure		Banded		Hub		Small		Banded		Hub		Small	
Variance correction		✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
(p, q)	γ	Empirical FDR (SE) (in %)											
(200, 50)	0.0	0.7(0.6)	0.7(0.6)	0.5(1.2)	0.4(0.9)	1.0(1.2)	1.1(1.2)	0.7(0.5)	0.7(0.5)	0.8(1.3)	0.7(1.3)	1.1(1.2)	1.1(1.3)
	0.2	0.5(0.4)	0.6(0.4)	1.0(1.3)	0.9(1.2)	0.9(1.0)	1.0(1.1)	0.8(0.6)	0.9(0.6)	0.8(1.1)	0.8(1.1)	1.1(0.8)	1.3(1.1)
	0.4	0.7(0.5)	1.7(0.8)	1.1(1.3)	2.5(2.3)	0.9(0.7)	2.7(1.5)	0.7(0.5)	1.9(0.8)	1.0(1.2)	1.9(1.9)	1.1(1.0)	2.7(2.0)
	0.6	0.7(0.5)	1.8(0.9)	0.9(1.1)	2.8(2.6)	1.1(0.8)	2.7(1.9)	0.6(0.4)	1.9(0.8)	0.8(1.3)	3.0(2.6)	1.1(0.9)	2.8(1.7)
(200, 200)	0.0	0.7(0.5)	0.7(0.5)	0.8(0.9)	0.8(0.9)	0.9(0.4)	0.9(0.4)	0.8(0.6)	0.8(0.5)	0.9(1.0)	0.8(0.9)	0.8(0.4)	0.8(0.4)
	0.2	0.9(0.5)	1.0(0.5)	1.1(1.1)	1.1(1.1)	0.7(0.4)	0.8(0.4)	0.8(0.6)	0.8(0.6)	1.1(1.2)	1.0(1.0)	0.7(0.4)	0.8(0.4)
	0.4	1.0(0.7)	2.2(0.9)	1.0(1.0)	2.6(1.9)	0.9(0.5)	1.8(0.7)	0.9(0.5)	2.0(0.9)	1.1(1.2)	2.2(1.8)	0.8(0.4)	1.8(0.6)
	0.6	0.7(0.5)	2.3(0.8)	1.1(0.9)	2.9(1.6)	0.8(0.4)	1.8(0.6)	0.8(0.5)	2.1(0.7)	1.0(1.0)	3.2(1.6)	0.8(0.3)	1.9(0.6)
(800, 200)	0.0	0.8(0.3)	0.8(0.3)	0.9(0.7)	0.9(0.7)	0.7(0.2)	0.7(0.2)	0.7(0.2)	0.7(0.2)	0.9(0.6)	0.9(0.6)	0.8(0.2)	0.8(0.2)
	0.2	0.8(0.3)	0.9(0.3)	0.9(0.6)	0.9(0.7)	0.7(0.2)	0.8(0.2)	0.7(0.2)	0.9(0.3)	1.0(0.7)	1.0(0.7)	0.7(0.2)	0.8(0.2)
	0.4	0.8(0.2)	2.5(0.5)	0.9(0.6)	2.1(0.9)	0.7(0.2)	2.2(0.3)	0.7(0.2)	2.5(0.5)	1.0(0.7)	2.1(1.0)	0.8(0.2)	2.3(0.4)
	0.6	0.8(0.2)	2.4(0.4)	0.9(0.6)	2.2(0.9)	0.8(0.2)	2.3(0.4)	0.8(0.2)	2.4(0.4)	0.9(0.5)	2.2(0.9)	0.8(0.2)	2.3(0.4)
		Empirical power (in %)											
(200, 50)	0.0	91.0	90.9	52.8	53.4	18.3	18.2	91.3	91.3	54.8	54.4	19.1	19.2
	0.2	88.1	89.3	53.3	52.9	20.5	19.7	90.2	90.9	51.6	50.7	19.6	18.8
	0.4	84.9	89.9	58.9	56.4	25.3	20.7	86.2	90.8	57.1	53.0	23.3	18.3
	0.6	85.6	90.9	60.3	54.3	24.8	19.9	85.3	90.4	59.7	54.8	25.4	20.3
(200, 200)	0.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	99.9	99.9
	0.2	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	99.9	99.9
	0.4	100.0	100.0	99.9	100.0	99.6	99.9	100.0	100.0	100.0	100.0	99.4	99.8
	0.6	100.0	100.0	99.9	100.0	99.5	99.9	100.0	100.0	100.0	100.0	99.7	99.9
(800, 200)	0.0	100.0	100.0	61.1	60.6	99.6	99.6	100.0	100.0	61.5	60.9	99.6	99.6
	0.2	100.0	100.0	61.8	61.8	99.5	99.6	100.0	100.0	62.4	62.6	99.6	99.6
	0.4	100.0	100.0	64.5	63.3	98.5	99.5	100.0	100.0	64.1	63.0	98.6	99.5
	0.6	100.0	100.0	65.1	62.9	98.6	99.5	100.0	100.0	64.4	62.8	98.6	99.5

For Setting II, for different combinations of (p, q) and spatial structures, the test with variance correction again controls the FDR close to the significance level, while the test without correction fails to control FDR as γ increases. When $(p, q) = (200, 50)$, the test with correction is slightly inferior to that without correction for the banded graph in terms of power. This is not surprising though, as it is attributed to the inflated FDR. For other spatial structures, the test with correction clearly outperforms the one without correction. When $(p, q) = (200, 200)$, we observe that the power of both tests increases to 100% or close. For FDR, the inflation issue still remains for the test without correction. When $(p, q) = (800, 200)$, we observe a similar qualitative pattern.

In summary, our proposed test with variance correction can control the false discovery and attain a good power for a range of strength of correlation before and after the stimulus. In contrast, the test without correction has inferior power performance for Setting I, and fails to control the FDR and yields an inflated power for Setting II as this correlation increases. We also report the mean squared error of $\hat{\Theta}^{(d)}$ in Section S5 of the Supplementary material available at *Biostatistics* online.

4.2. *Sensitivity analysis*

We next carry out sensitivity analysis to evaluate the performance of our test when the data deviates from the matrix normal distribution. We first replace the normal distribution by a t distribution. Specifically, we follow the data generation mechanism as before, while we set $\mathbf{\Sigma}_{S_{1,2}}$ as a diagonal matrix with $\mathbf{\Sigma}_{S_{1,2},i,i} = \gamma \cdot \mathbf{\Sigma}_{S_1,i,i}(1-2\cdot\mathbf{1}[i \equiv k \pmod 7, k \in \{1,3,5\}])$ and $\gamma = 0.6$. Since a normal random vector $\boldsymbol{X} \sim N(\mathbf{0}, \mathbf{\Sigma})$ can be represented as $\boldsymbol{X} = \mathbf{\Sigma}^{1/2}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N(\mathbf{0}, I_p)$, we replace the Gaussian entries in $\boldsymbol{Z}$ with t-distributed random variables with degree of freedom $df \in \{4, 6, 8\}$. We report the empirical FDR and power out of 100 data replications in Table 3, part I. It is seen that, our test manages to control the FDR reasonably well, and attains a good power under different dependence structures.

We then examine the performance of our method with regard to the off-diagonal Kronecker product structure, i.e., $\mathrm{cov}(\mathrm{vec}\{\mathbf{X}^{(1)}\}, \mathrm{vec}\{\mathbf{X}^{(2)}\}) = \mathbf{\Sigma}_{S_{1,2}} \otimes \mathbf{\Sigma}_{T_{1,2}}$. We again follow the data generation mechanism as before, but set $\mathbf{\Sigma}_{S_{1,2}}$ as a diagonal matrix with $\mathbf{\Sigma}_{S_{1,2},i,i} = \gamma \cdot \mathbf{\Sigma}_{S_1,i,i}(1-2\cdot\mathbf{1}[i \equiv k \pmod 7, k \in \{1,3,5\}])$, $\gamma = 0.6$, and $(p, q) = (200, 50)$. We further perturb $\mathbf{\Sigma}_{1,2} = \mathbf{\Sigma}_{S_{1,2}} \otimes \mathbf{\Sigma}_{T_{1,2}}$ in two steps: we randomly sample $p^\star\%$ entries of $\mathbf{\Sigma}_{1,2}$, where $p^\star \in \{0, 1, 5, 10\}$, then replace those entries with i.i.d. Gaussian random variables of mean zero and standard deviation ν, where $\nu = l^\star \times$ the magnitude for the entries of $\mathbf{\Sigma}_{1,2}$, and $l^\star \in \{0.1, 1\}$. We report the empirical FDR and power out of 100 data replications in Table 3, part II. It is seen that, our test maintains a reasonably good performance in this setup too.

These results show that our method is relatively robust with regard to the joint matrix normal assumption (2.1). We also comment that, it is possible to extend our test to semiparametric normal copula setting. Liu *and others* (2012) and Xue and Zou (2012) studied the vector-valued case. Following a similar idea of marginal monotonic transformation, it is possible to develop a paired test in the matrix-valued setting. We leave the full investigation as future research.

5. AD DATA ANALYSIS

AD is an irreversible neurodegenerative disorder and is characterized by progressive impairment of cognitive and memory functions. It is the leading form of dementia in the elderly subjects. With the aging of the worldwide population, the number of affected people is rapidly increasing and is projected to be 13.8 million in the United States, and 1 in 85 worldwide by year 2050 (Brookmeyer *and others*, 2007, 2011). It thus has become an international imperative to understand, diagnose, and treat this disorder. Accumulated evidences have suggested that alterations in brain connectivity networks are predictive of cognitive function and decline, and hold crucial insights about the disease pathology of AD (Fox and Greicius, 2010).

Table 3. Sensitivity (part I): The empirical FDR (with standard error in the parenthesis) and the empirical power, both in percentage, for the sensitivity analysis on the normality.

Sensitivity I													
Temporal structure		Moving average						Autoregressive					
Spatial structure		Banded		Hub		Small		Banded		Hub		Small	
Correction		✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
(p, q)	df	Empirical FDR (SE) (in %)											
	4	0.7(0.5)	1.9(0.8)	1.1(1.6)	4.1(2.6)	1.1(0.8)	2.7(1.4)	0.8(0.6)	1.8(0.9)	1.2(1.5)	4.3(3.2)	1.2(0.8)	2.8(1.5)
	6	0.8(0.4)	1.8(0.8)	1.0(1.3)	3.7(2.6)	1.4(0.9)	2.8(1.4)	0.6(0.4)	1.7(0.8)	0.9(1.2)	3.9(2.3)	1.3(1.0)	3.2(1.7)
(200, 50)	8	0.7(0.5)	1.6(0.8)	1.4(1.8)	3.6(2.6)	1.3(1.1)	2.6(1.3)	0.6(0.5)	1.8(0.8)	0.6(1.1)	3.2(2.7)	1.0(0.9)	2.9(1.5)
	4	0.8(0.6)	1.9(0.8)	1.1(1.0)	3.3(1.8)	0.9(0.4)	2.1(0.6)	0.7(0.5)	2.0(0.8)	0.9(1.2)	3.2(2.0)	0.9(0.4)	1.9(0.6)
	6	0.8(0.5)	2.0(0.9)	0.8(0.8)	3.1(1.9)	0.9(0.4)	1.8(0.6)	0.9(0.6)	2.1(0.9)	0.9(1.0)	3.1(2.0)	0.8(0.5)	1.7(0.6)
(200, 200)	8	1.0(0.5)	2.1(0.9)	0.7(0.9)	3.2(1.9)	0.8(0.4)	1.9(0.6)	0.8(0.6)	2.1(0.7)	1.0(1.1)	3.2(1.6)	0.8(0.4)	1.9(0.6)
	4	0.8(0.3)	2.6(0.4)	0.9(0.7)	2.6(1.1)	0.8(0.2)	2.2(0.3)	0.8(0.3)	2.7(0.5)	1.0(0.5)	2.4(1.1)	0.8(0.2)	2.2(0.3)
	6	0.8(0.3)	2.5(0.4)	0.9(0.6)	2.4(0.8)	0.8(0.2)	2.1(0.3)	0.8(0.3)	2.5(0.5)	1.0(0.6)	2.3(0.9)	0.8(0.2)	2.1(0.3)
(800, 200)	8	0.8(0.3)	2.4(0.4)	0.8(0.6)	2.2(1.1)	0.8(0.2)	2.1(0.3)	0.8(0.2)	2.5(0.4)	0.8(0.5)	2.0(0.7)	0.8(0.2)	2.2(0.3)
		Empirical power (SE) (in %)											
	4	90.3	94.4	59.4	52.5	29.5	23.6	90.4	94.5	58.7	53.4	28.9	23.3
	6	90.3	94.5	59.2	53.4	28.5	23.2	90.9	94.9	58.0	51.3	27.8	21.9
(200, 50)	8	90.7	94.8	59.1	52.7	29.3	22.7	90.6	95.1	59.7	53.6	28.7	23.0
	4	100.0	100.0	99.9	100.0	99.7	99.9	100.0	100.0	99.9	100.0	99.6	99.9
	6	100.0	100.0	99.9	100.0	99.7	99.9	100.0	100.0	99.9	100.0	99.7	99.9
(200, 200)	8	100.0	100.0	99.9	100.0	99.8	99.9	100.0	100.0	99.9	100.0	99.7	99.9
	4	100.0	100.0	64.6	66.4	98.9	99.6	100.0	100.0	65.3	67.2	98.9	99.6
	6	100.0	100.0	65.1	66.6	99.0	99.6	100.0	100.0	65.3	67.4	98.9	99.6
(800, 200)	6	100.0	100.0	64.6	66.4	99.0	99.7	100.0	100.0	65.5	66.5	99.0	99.6

(Continued)

Table 3. (Continued) Sensitivity (part II): The empirical FDR (with standard error in the parenthesis) and the empirical power, both in percentage, for the sensitivity analysis on the Kronecker structure of the off-diagonal block, i.e., $\text{cov}(\text{vec}\{\mathbf{X}^{(1)}\}, \text{vec}\{\mathbf{X}^{(2)}\}) = \mathbf{\Sigma}_{S_{1,2}} \otimes \mathbf{\Sigma}_{T_{1,2}}$.

Sensitivity II													
Temporal structure		Moving average						Autoregressive					
Spatial structure		Banded		Hub		Small		Banded		Hub		Small	
Correction		✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
$p^\star$	$l^\star$	Empirical FDR (SE) (in %)											
0	0.1	0.7(0.4)	0.7(0.5)	0.7(1.2)	1.1(1.6)	1.1(0.9)	1.2(1.0)	0.7(0.5)	0.7(0.5)	0.8(1.1)	1.3(1.4)	1.0(1.1)	1.1(1.0)
	1	0.6(0.5)	0.7(0.6)	1.2(1.9)	1.5(1.9)	1.3(1.1)	1.4(1.0)	0.6(0.6)	0.7(0.5)	0.9(1.3)	1.2(1.5)	1.3(0.9)	1.4(1.2)
1	0.1	0.7(0.5)	0.7(0.5)	1.0(1.6)	1.1(1.7)	0.9(0.9)	1.0(0.9)	0.7(0.4)	0.7(0.4)	0.7(1.1)	1.1(1.3)	1.3(0.9)	1.3(1.0)
	1	0.8(0.5)	0.8(0.5)	0.7(1.2)	0.8(1.2)	1.2(1.0)	1.2(1.1)	0.7(0.5)	0.7(0.4)	0.9(1.2)	0.9(1.1)	1.1(0.9)	1.1(0.9)
5	0.1	0.6(0.5)	0.7(0.5)	0.9(1.3)	1.4(1.5)	1.0(0.9)	1.2(1.0)	0.7(0.5)	0.7(0.5)	0.7(1.2)	0.7(1.4)	1.2(1.0)	1.3(1.1)
	1	0.8(0.6)	0.8(0.6)	1.3(1.4)	1.4(1.4)	1.0(0.8)	1.1(0.9)	0.7(0.4)	0.7(0.5)	0.5(1.0)	0.6(1.3)	0.9(0.9)	1.1(1.0)
10	0.1	0.9(0.6)	0.9(0.5)	0.9(1.2)	0.8(1.1)	1.2(1.0)	1.3(1.1)	0.9(0.7)	0.9(0.6)	1.0(1.3)	1.1(1.6)	1.1(0.8)	1.1(0.8)
	1	0.5(0.5)	0.6(0.5)	1.1(1.7)	1.1(1.7)	1.2(1.1)	1.2(1.2)	0.6(0.5)	0.6(0.5)	1.2(1.6)	1.2(1.7)	1.1(0.9)	1.2(1.1)
		Empirical power (SE) (in %)											
0	0.1	94.6	95.0	50.9	49.3	23.8	23.5	94.7	95.2	53.4	52.2	23.2	22.7
	1	95.1	95.2	54.0	52.0	22.9	22.6	95.1	95.4	51.9	51.0	23.2	22.9
1	0.1	94.8	95.0	53.8	52.8	23.2	22.9	94.8	95.1	52.9	51.8	23.3	23.0
	1	94.6	94.9	53.6	53.1	23.2	22.6	95.2	95.4	51.8	50.9	23.0	22.5
5	0.1	94.8	95.2	52.1	52.1	23.3	22.9	94.7	95.0	53.3	51.9	23.5	23.1
	1	94.9	95.2	52.2	52.0	22.9	22.6	94.6	95.0	52.1	52.0	23.1	23.1
10	0.1	94.6	95.1	53.3	51.6	23.3	22.8	94.7	94.8	52.7	51.5	22.9	22.3
	1	95.1	95.4	52.8	52.1	22.8	22.7	94.7	94.9	52.6	51.6	23.8	23.5

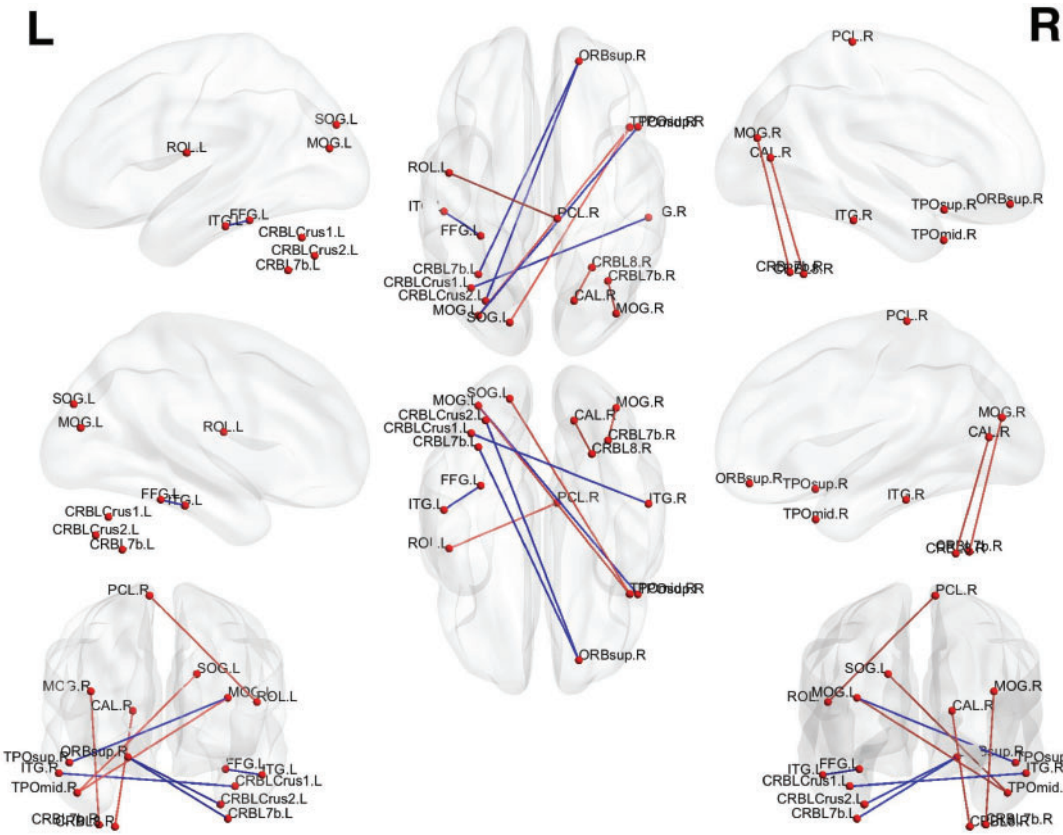

Fig. 1. Top 10 differentiating links of the brain connectivity networks of the 23 subjects of the ADNI database before and after the conversion from MCI to AD. All the associated p-values are smaller than 1e−13. Table 4 displays the status of these links (being enhanced or weakened after the conversion).

We analyzed a dataset from the Alzheimer's Disease Neuroimaing Initiative (ADNI). ADNI is an ongoing, longitudinal, multi-center study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. We focused on 23 subjects from ADNI who experienced conversion from mild cognitive impairment (MCI), a prodromal stage of AD, to AD during the 24-month follow-up. The primary scientific question of interest is to investigate the change of brain connectivity patterns before and after the conversion. All fMRI scans were resting-state and preprocessed, including slice timing correction, motion correction, spatial smoothing, denoising by regressing out motion parameters, white matter and cerebrospinal fluid time courses, and band-pass filtering. The data were then aligned and parcellated using the Anatomical Automatic Labeling atlas (Tzourio-Mazoyer *and others*, 2002). The resulting data are a region by time matrix for each subject, with the spatial dimension $p = 116$ and the temporal dimension $q = 130$.

We first examined the quantile–quantile plot, which shows no clear deviation from the normal distribution. We next applied the testing procedure of Aston *and others* (2017) to test if the data conforms with the Kronecker product structure. The p-values of the test before and after the conversion are 0.17 and 0.11, respectively, which suggests that the product structure seems to reasonably hold for this dataset. We then applied our proposed variance-corrected testing procedure to this data. In our analysis, we did not

Table 4. *Top 30 differentiating links of the brain connectivity networks of the 23 subjects of the ADNI database before and after the conversion from MCI to AD. The last column shows the direction of the link change. "+" represents the link gets enhanced after the conversion, and "−" represents the link gets weakened after the conversion*

Rank	Differentiating links	*p*-value	+/−
1	Cerebellum_Crus1_L↔Temporal_Inf_R	0	−
2	Temporal_Pole_Sup_R↔Occipital_Mid_L	1.11e−16	−
3	Temporal_Pole_Mid_R↔Occipital_Sup_L	2.22−16	+
4	Temporal_Pole_Mid_R↔Occipital_Mid_L	3.33−16	+
5	Paracentral_Lobule_R↔Rolandic_Oper_L	7.77e−16	+
6	Cerebellum_Crus2_L↔Frontal_Sup_Orb_R	9.99e−15	−
7	Cerebellum_7b_L↔Frontal_Sup_Orb_R	1.07e−14	−
8	Cerebellum_7b_R↔Occipital_Mid_R	1.74e−14	+
9	Cerebellum_8_R↔Calcarine_R	2.72e−14	+
10	Temporal_Inf_L↔Fusiform_L	3.30e−14	−
11	Fusiform_R↔Cuneus_R	2.02e−13	−
12	Cerebellum_Crus2_L↔Temporal_Inf_R	3.85e−13	−
13	Occipital_Inf_R↔Rectus_L	7.37e−13	+
14	Cerebellum_7b_L↔Fusiform_R	9.01e−13	+
15	ParaHippocampal_R↔Frontal_Inf_Orb_L	1.62e−12	+
16	Temporal_Pole_Mid_R↔Temporal_Pole_Sup_L	2.77e−12	+
17	Heschl_L↔Lingual_R	3.22e−12	−
18	Cerebellum_10_R↔Olfactory_R	3.46e−12	+
19	Cerebellum_9_L↔Frontal_Mid_Orb_L	4.89e−12	−
20	Cerebellum_Crus1_R↔Cerebellum_Crus1_L	9.00e−12	−
21	Cerebellum_10_R↔Frontal_Mid_Orb_R	1.42e−11	+
22	Cerebellum_10_L↔Frontal_Mid_Orb_R	1.75e−11	−
23	Cerebellum_Crus1_L↔Frontal_Mid_Orb_L	1.96e−11	+
24	SupraMarginal_L↔Cuneus_R	2.45e−11	−
25	Cerebellum_6_L↔Cerebellum_Crus2_L	3.70e−11	+
26	Cerebellum_7b_L↔Rectus_L	4.96e−11	−
27	Cerebellum_3_R↔Frontal_Med_Orb_R	5.69e−11	+
28	Insula_R↔Frontal_Inf_Oper_R	5.88e−11	−
29	Angular_R↔Angular_L	7.00e−11	−
30	Cerebellum_7b_L↔Temporal_Inf_R	7.02e−11	+

correct for potential confounder effects, but our test can be equally applied to the corrected data. Figure 1 plots those top differentiating links whose corresponding *p*-values are smaller than 1e−13, and the associated brain regions visualized with the BrainNet Viewer (Xia *and others*, 2013). Table 4 further reports the top 30 links that were found different before and after the conversion, with their associated *p*-values and the directions of the change. It is seen that the differentiating links concentrate on the cerebellum. The cerebellum is critical in the distributed neural circuits subserving not only motor function but also autonomic, limbic, and cognitive behaviors. There is recently increased interest in exploring the role of the cerebellum in neurodegenerative disorders, in particular AD (Jacobs *and others*, 2018). Our findings provide a useful support to the existing literature.

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

The computer code in R for the simulation and data analysis can be found at https://github.com/Elric2718/PairedTestPrecisionMatrix with the commit number 631c4e4. The Alzheimer's disease dataset can be found at https://doi.org/10.6084/m9.figshare.9643010.v3.

ACKNOWLEDGMENTS

The authors thank the Associate Editor and two anonymous reviewers for constructive suggestions and questions that helped improve the manuscript.

Conflict of Interest: None declared.

Funding

NSFC (11771094, 11690013) and The Recruitment Program of Global Experts Youth Project to Y.X. (in part); NSF (DMS-1613137) and NIH (R01AG034570 and R01AG061303) to L.L. (in part).

REFERENCES

AHN, M., SHEN, H., LIN, W. AND ZHU, H. (2015). A sparse reduced rank framework for group analysis of functional neuroimaging data. *Statistica Sinica* **25**, 295.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, Inc.

ASTON, J. A. D, PIGOLI, D. AND TAVAKOLI, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**, 1431–1461.

BICKEL, P. J. AND LEVINA, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

BROOKMEYER, R., EVANS, D. A., HEBERT, L., LANGA, K. M., HEERINGA, S. G., PLASSMAN, B. L. AND KUKULL, W. A. (2011). National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimer's & Dementia* **7**, 61–73.

BROOKMEYER, R., JOHNSON, E., ZIEGLER-GRAHAM, K. AND ARRIGHI, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimers Dementia* **3**, 186–191.

CAI, F., GAO, L., GONG, H., JIANG, F., PEI, C., ZHANG, X., ZENG, X. AND HUANG, R. (2015). Network centrality of resting-state fMRI in primary angle-closure glaucoma before and after surgery. *PLoS One* **10**, e0141389.

CAI, T. AND LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

CAI, T., LIU, W. AND LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106**, 594–607.

CHEN, S., KANG, J., XING, Y. AND WANG, G. (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. *Human Brain Mapping* **36**, 5196–5206.

CHEN, T., RYALI, S., QIN, S. AND MENON, V. (2013). Estimation of resting-state functional connectivity using random subspace based partial correlation: a novel method for reducing global artifacts. *Neuroimage* **82**, 87–100.

CHEN, X. AND LIU, W. (2018). Testing independence with high-dimensional correlated samples. *The Annals of Statistics* **46**, 866–894.

CHEN, X. AND LIU, W. (2019). Graph estimation for matrix-variate Gaussian data. *Statistica Sinica* **29**, 479–504..

FICEK, B. N., WANG, Z., ZHAO, Y., WEBSTER, K. T., DESMOND, J. E., HILLIS, A. E., FRANGAKIS, C., VASCONCELLOS FARIA, A., CAFFO, B. AND TSAPKINI, K. (2018). The effect of tDCS on functional connectivity in primary progressive aphasia. *Neuroimage Clinical* **19**, 703–715.

Fox, M. D. and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience* **4**, 19.

Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E. and Penny, W. D. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2011.

Gianaros, P. J., Sheu, L. K., Matthews, K. A., Jennings, J. R., Manuck, S. B. and Hariri, A. R. (2008). Individual differences in stressor-evoked blood pressure reactivity vary with activation, volume, and functional connectivity of the amygdala. *Journal of Neuroscience* **28**, 990–999.

Hedden, T., Van Dijk, K. R. A., Becker, J. A., Mehta, A., Sperling, R. A., Johnson, K. A. and Buckner, R. L. (2009). Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *Journal of Neuroscience* **29**, 12686–12694.

Jacobs, H. I. L., Hopkins, D. A., Mayrhofer, H. C., Bruner, E., van Leeuwen, F. W., Raaijmakers, W. and Schmahmann, J. D. (2018). The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. *Brain* **141**, 37–47.

Kang, J., Bowman, F. D., Mayberg, H. and Liu, H. (2016a). A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs. *Neuroimage* **141**, 431–441.

Kang, S.-G., Yoon, H.-K, Cho, C.-H., Kwon, S., Kang, J., Park, Y.-M., Lee, E., Kim, L. and Lee, H.-J. (2016b). Decrease in fMRI brain activation during working memory performed after sleeping under 10 lux light. *Scientific Reports* **6**, 36731.

Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of American Statistical Association* **107**, 1187–1200.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40**, 2293–2326.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics* **41**, 2948–2978.

Narayan, M., Allen, G. I. and Tomson, S. (2015). Two sample inference for populations of graphical models with applications to functional connectivity. arXiv preprint arXiv:1502.03853.

Peck, K. K., Moore, A. B., Crosson, B. A., Gaiefsky, M., Gopinath, K. S., White, K. and Briggs, R. W. (2004). Functional magnetic resonance imaging before and after aphasia therapy. *Stroke* **35**, 554–559.

Qiu, H., Han, F., Liu, H. and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**, 487–504.

Quaedflieg, C. W. E. M., van de Ven, V., Meyer, T., Siep, N., Merckelbach, H. and Smeets, T. (2015). Temporal dynamics of stress-induced alternations of intrinsic amygdala connectivity and neuroendocrine levels. *PLoS One* **10**, 1–16.

Rudie, J. D., Brown, J. A., Beck-Pancer, D., Hernandez, L. M., Dennis, E. L., Thompson, P. M., Bookheimer, S. Y. and Dapretto, M. (2013). Altered functional and structural brain network organization in autism. *Neuroimage: Clinical* **2**, 79–94.

Ryali, S., Chen, T., Supekar, K. and Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* **59**, 3852–3861.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E. *and others*. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289.

VAN MARLE, H. J. F., HERMANS, E. J., QIN, S. AND FERNÁNDEZ, G. (2010). Enhanced resting-state connectivity of amygdala in the immediate aftermath of acute psychological stress. *Neuroimage* **53**, 348–354.

VAN WIERINGEN, W. N. AND PEETERS, C. F. W. (2016). Ridge estimation of inverse covariance matrices from high-dimensional data. *Computational Statistics & Data Analysis* **103**, 284–303.

VAROQUAUX, G. AND CRADDOCK, R. C. (2013). Learning and comparing functional connectomes across subjects. *Neuroimage* **80**, 405–415.

WANG, Y., KANG, J., KEMMER, P. B. AND GUO, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Frontiers in Neuroscience* **10**, 123.

XIA, M., WANG, J. AND HE, Y. (2013). Brainnet viewer: a network visualization tool for human brain connectomics. *PLoS One* **8**, 1–15.

XIA, Y., CAI, T. AND CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102**, 247–266.

XIA, Y. AND LI, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics* **73**, 780–791.

XIA, Y. AND LI, L. (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statistica Sinica* **29**, 303–328.

XUE, L. AND ZOU, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics* **40**, 2541–2571.

YIN, J. AND LI, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis* **107**, 119–140.

ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. AND WASSERMAN, L. (2012). The huge package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research* **13**, 1059–1062.

[*Received November 4, 2018; revised August 20, 2019; accepted for publication September 17, 2019*]