Testing the Mean Matrix in High-Dimensional Transposable Data

Anestis Touloumis[1,*], Simon Tavaré[1,], and John C. Marioni[2,***]**

[1] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, U.K.

[2] EMBL-European Bioinformatics Institute, Hinxton CB10 1SD, U.K.

email: Anestis.Touloumis@cruk.cam.ac.uk

**email:* Simon.Tavare@cruk.cam.ac.uk

***email:* marioni@ebi.ac.uk

Summary: The structural information in high-dimensional transposable data allows us to write the data recorded for each subject in a matrix such that both the rows and the columns correspond to variables of interest. One important problem is to test the null hypothesis that the mean matrix has a particular structure without ignoring the potential dependence structure among and/or between the row and column variables. To address this, we develop a simple and computationally efficient nonparametric testing procedure to assess the hypothesis that, in each predefined subset of columns (rows), the column (row) mean vector remains constant. In simulation studies, the proposed testing procedure seems to have good performance and unlike traditional approaches, it is powerful without leading to inflated nominal sizes. Finally, we illustrate the use of the proposed methodology via two empirical examples from gene expression microarrays.

Key words: High-dimensional transposable data; Hypothesis testing; Mean matrix; Nonparametric test.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

In some applications, the measurements related to each subject are naturally organized in a matrix, especially when the rows and columns correspond to two different sets of variables and dependencies are expected to occur between and/or among them. Allen and Tibshirani (2010) introduced the term 'transposable data' to acknowledge the structural information and the presence of two-way dependencies in matrix-valued random variables. Examples of transposable data can be found in spatiotemporal studies (Genton, 2007; Mardia and Goodall, 1993), in cross-classified multivariate data (Galecki, 1994; Naik and Rao, 2001), in genetics (Allen and Tibshirani, 2010, 2012; Efron, 2009; Teng and Huang, 2009), in functional MRI (Allen and Tibshirani, 2010) and in time-series (Carvalho and West, 2007; Lee, Daniels, and Joo, 2013).

Our work is motivated by biological studies that use microarrays to study gene expression patterns in multiple tissue samples taken from the same subject (Sottoriva et al., 2013; Zahn et al., 2007). For each subject, the row variables correspond to genes, the column variables to tissue samples and the measurements are mRNA gene expression levels. A biologically important objective is to check whether different subsets of the samples share a common mean vector of gene expression levels. Application of standard statistical methods, such as the analysis of variance (ANOVA), is hindered by two factors: (i) the 'large p, small N' setting (the number of genes will typically exceed the number of subjects); and (ii) the high-dimensional dependence structure (since the genes and the tissue samples are not expected to be independent of each other).

To introduce these concepts in mathematical terms, suppose that an experimentalist collects N independent and identically distributed (i.i.d.) transposable $r \times c$ random matrices $\mathbf{X}_1, \ldots, \mathbf{X}_N$. For each subject, there are r row variables and c column variables and the high-dimensional setting is indicated by letting the sample size (N) be much smaller than the

number of observations (rc) for a single subject. The goal is to test pre-specified hypotheses about $\mathbf{M} = \mathrm{E}[\mathbf{X}_i]$, the $r \times c$ mean matrix of the transposable data.

To illustrate some difficulties of this task, consider the simple hypothesis

$$\mathrm{H}_0 : \mathbf{M} = \boldsymbol{\mu}\mathbf{1}_c^T \text{ vs } \mathrm{H}_1 : \mathbf{M} \neq \boldsymbol{\mu}\mathbf{1}_c^T, \tag{1}$$

where $\boldsymbol{\mu}$ is an unknown r-variate parameter vector and $\mathbf{1}_s$ denotes an s-variate vector of ones. The null hypothesis suggests that the mean-relationship between the row and column variables is completely determined by the row variables. In the motivating examples, H_0 in (1) is consistent with no genes showing differential expression across the multiple tissue samples. To the best of our knowledge, no statistical procedure exists to test hypothesis (1) directly in high-dimensional transposable data unless there are only two dependent column variables $(c = 2)$. In this case, one approach is to employ the test proposed by Chen and Qin (2010) for comparing the mean vector of paired high-dimensional random vectors. To accomplish this, one needs to form the random vector of the difference of the two columns for each subject and then test the hypothesis of a zero mean vector. Unfortunately, there is no straightforward way to apply or extend this test when $c > 2$. In particular, attempts to do this essentially infer rather than test the mean relationship between the row and column variables. For example, suppose that $\mathbf{M} = [\boldsymbol{\mu}, -\boldsymbol{\mu}, \boldsymbol{\mu}, -\boldsymbol{\mu}]$ and consider the following naive algorithm to test hypothesis (1). First, create two groups of column variables, one based on the first two columns and the other based on the last two. Second, for each group create N r-variate random vectors by averaging the appropriate columns in each matrix, and then for each subject create the r-variate vectors of the difference of the two groups. Thirdly, test hypothesis (1) using the test statistic of Chen and Qin (2010) as above. It can be shown that this vector-based test statistic will be powerless since the transformed random vectors will indeed have a zero mean vector.

By contrast, we propose a simple approach to test hypothesis (1) that overcomes theoretical

problems. In this direction, let $\mathbf{P}_c = \mathbf{I}_c - \mathbf{J}_c/c$ where $\mathbf{I}_s$ is the identity matrix of size s and $\mathbf{J}_s$ is the $s \times s$ matrix of ones, and let $\mathrm{tr}(\mathbf{A})$ denote the trace operator of the matrix $\mathbf{A}$. Note that $\mathbf{P}_c$ is a symmetric and idempotent $(\mathbf{P}_c^2 = \mathbf{P}_c)$ matrix such that $\mathrm{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P}_c) = 0$ if and only if H_0 in (1) holds. Since the Frobenious norm, $\mathrm{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P}_c)$, measures deviations from H_0, it seems meaningful to develop a test statistic based on the unbiased estimator of this norm,

$$\frac{1}{N(N-1)} \sum_{i \neq j} \mathrm{tr}(\mathbf{X}_i^T\mathbf{X}_j\mathbf{P}_c).$$

Under rather weak conditions about the dependence structure of the row and column variables, outlined in Section 2.2, this estimator asymptotically follows a normal distribution, and hence, the critical region of the test statistic can be defined under H_0.

The main contribution of this paper is that we extend this idea to test more complicated hypotheses than hypothesis (1) for the mean matrix. In particular, we consider the hypothesis

$$\mathrm{H}_0 : \mathbf{M} = [\boldsymbol{\mu}_1 \mathbf{1}_{c_1}^T, \boldsymbol{\mu}_2 \mathbf{1}_{c_2}^T, \ldots, \boldsymbol{\mu}_g \mathbf{1}_{c_g}^T] \text{ vs } \mathrm{H}_1 : \mathbf{M} \neq [\boldsymbol{\mu}_1 \mathbf{1}_{c_1}^T, \boldsymbol{\mu}_2 \mathbf{1}_{c_2}^T, \ldots, \boldsymbol{\mu}_g \mathbf{1}_{c_g}^T], \qquad (2)$$

where $c_1, \ldots, c_g$ are positive integers such that $\sum_{q=1}^{g} c_q = c$ with at least one $c_q \geqslant 2$ and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g$ are g unknown r-variate parameter vectors. H_0 in (2) states that in each one of the given g column groups there is no column effect upon the mean of the row variables. Since g is known but arbitrary, the proposed testing procedure is not limited by the number of column groups or the group size under consideration. For example, hypothesis (1) is a special case of hypothesis (2) with $g = 1$ and $c_1 = c$ while the hypothesis that two column variables, say the first two, have a common mean vector is obtained by setting $g = c - 1$, $c_1 = 2$ and $c_2, \ldots, c_g = 1$. Similarly to testing hypothesis (1), the proposed test statistic will be based on an asymptotic argument via a pivotal quantity that is the unbiased estimator of the distance of the mean matrix from H_0 in (2).

This article is structured as follows. In Section 2, we introduce the high-dimensional working framework and we construct the test statistic for testing hypothesis (2). Also, we

discuss the asymptotic power of the proposed test, we argue that the required assumptions that make the high-dimensional setting manageable are weak, and we provide guidelines to ease the computational task whenever straightforward calculation of the test statistic is time-consuming. In Section 3, we examine the performance of the proposed testing methodology in finite samples using simulations. In Section 4, we apply the proposed testing methodology to two microarrays studies where gene expression levels are measured in different tissue samples (Sottoriva et al., 2013; Zahn et al., 2007). In Section 5, we summarize the main findings of our research and discuss how this method can be extended to test additional hypotheses for the mean matrix. The Appendix contains technical details.

2. Test Statistics for the Mean Matrix

As the generative process for transposable data, consider a matrix-valued extension of the nonparametric model for vectors considered in Bai and Saranadasa (1996) and Chen and Qin (2010)

$$\mathbf{X}_i = \mathbf{W}_i + \mathbf{M} \tag{3}$$

for $i = 1, \ldots, N$, where

(1) $\mathbf{M} = \mathrm{E}[\mathbf{X}_i]$ is the $r \times c$ mean matrix,

(2) $\mathbf{W}_i$ is an $r \times c$ matrix of random variables such that $\mathrm{vec}(\mathbf{W}_i) = \boldsymbol{\Sigma}^{1/2}\mathrm{vec}(\mathbf{Z}_i)$, and where $\mathrm{vec}(\mathbf{A})$ denotes vectorization of the matrix $\mathbf{A}$,

(3) $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \mathrm{cov}[\mathrm{vec}(\mathbf{X}_i)]$ is an $(rc) \times (rc)$ positive-definite covariance matrix,

(4) $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ are i.i.d. $r \times c$ random matrices and Z_{iab} is the (a, b)-th element of $\mathbf{Z}_i$,

(5) $\mathrm{E}[Z_{iab}] = 0$, $\mathrm{E}[Z_{iab}^2] = 1$, $\mathrm{E}[Z_{iab}^4] = 3 + B$ for a finite constant $B > -2$, $\mathrm{E}[Z_{iab}^8] < \infty$ and for any positive integers $l_1, \ldots, l_q$ with $\sum_{\nu=1}^{q} l_\nu \leqslant 8$

$$\mathrm{E}[Z_{ia_1b_1}^{l_1} Z_{ia_2b_2}^{l_2} \ldots Z_{ia_qb_q}^{l_q}] = \mathrm{E}[Z_{ia_1b_1}^{l_1}]\mathrm{E}[Z_{ia_2b_2}^{l_2}] \ldots \mathrm{E}[Z_{ia_qb_q}^{l_q}]$$

for $(a_1, b_1) \neq (a_2, b_2) \neq \cdots \neq (a_q, b_q)$.

The matrix-variate normal distribution (Dawid, 1981; Gupta and Nagar, 2000), a common and sensible choice for modeling transposable data, is a special case of model (3). To see this, let Z_{iab} be i.i.d. random variables from a standard normal distribution $N(0,1)$ and let $\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$, where $\mathbf{\Sigma}_1$ is the covariance matrix of the row variables, $\mathbf{\Sigma}_2$ is the covariance matrix of the column variables and $\otimes$ denotes the Kronecker product operator applied to matrices. However, we underline that model (3) is more general. It can handle departures from the matrix-variate normal model by relaxing the normality and/or the covariance structure assumption. First, the distribution of the "white-noise" random variables in $\mathbf{Z}_i$ remains unspecified. In fact, the white noise random variables do not need to be independent or identically distributed. The latter holds because the results of this paper are not affected even if the assumption of common fourth moments for the elements of $\mathbf{Z}_i$ is dropped. Finally, the dependence structure between and among the row and column variables is not limited to a Kronecker product form.

To construct the test statistic for testing hypothesis (2), we need additional notation. Let $\mathbf{P}_{\{c_1,c_2,\ldots,c_g\}} = \text{diag}(\mathbf{P}_{c_1}, \mathbf{P}_{c_2}, \ldots, \mathbf{P}_{c_g})$ be the $c \times c$ block diagonal matrix where the positive integers are defined by H_0 in (2). For notational ease, suppress the index set in $\mathbf{P}_{\{c_1,c_2,\ldots,c_g\}}$ and write instead $\mathbf{P}$. Further, note that $\mathbf{P}$ is a projection matrix as it is both idempotent and symmetric. The key to our proposal is to observe that $\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P}) = 0$ if and only if H_0 in (2) holds. To see this, note that $\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P}) = \text{tr}(\mathbf{P}\mathbf{M}^T\mathbf{M}\mathbf{P})$ is the sum of squares of the elements of $\mathbf{M}\mathbf{P}$, whose (a,b)-th element equals the difference between μ_{ab}, the (a,b)-th element of $\mathbf{M}$, and $\bar{\mu}_a^{(k)}$, the average of the a-th row in the mean matrix when this is restricted to the column group, say k, to which column b belongs under H_0 in (2). Therefore, it is sensible to consider the unbiased estimator of $\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})$

$$G_N = \frac{1}{N(N-1)} \sum_{i \neq j} \text{tr}(\mathbf{X}_i^T \mathbf{X}_j \mathbf{P}),$$

whose variance is

$$\text{Var}[G_N] = \frac{2}{N(N-1)}\text{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right) + \frac{4}{N}\text{vec}(\mathbf{MP})^T\boldsymbol{\Sigma}\text{vec}(\mathbf{MP}).$$

Next, we define the asymptotic framework needed to derive the limiting distribution of G_N. We handle the high-dimensional setting without specifying the limiting rate of the pairwise ratios of the triplet (N, r, c). This is a pragmatic restriction in many applications, including our motivating examples, where the number of row (genes) and/or column (multiple samples) variables are not expected to increase proportionally to the sample size. Instead, we assume that as $N \to \infty$ and $rc = r(N)c(N) \to \infty$, the following conditions hold:

$$\text{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^4\right) = o\left\{\text{tr}^2\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right)\right\} \tag{4}$$

and

$$\text{vec}(\mathbf{MP})^T\boldsymbol{\Sigma}\text{vec}(\mathbf{MP}) = o\left\{\frac{1}{N}\text{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right)\right\} \tag{5}$$

or

$$\frac{1}{N}\text{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right) = \text{o}\left\{\text{vec}(\mathbf{MP})^T\boldsymbol{\Sigma}\text{vec}(\mathbf{MP})\right\}. \tag{6}$$

The assumption $rc \to \infty$ is quite flexible and it does not require $r \to \infty$ and $c \to \infty$ simultaneously. On the contrary, the number of row or column variables is allowed to be fixed provided that the other dimension of the transposable data tends to ∞. Condition (4) specifies the class of covariance matrices for $\boldsymbol{\Sigma}$ under consideration. In Section 2.2, we argue that this class is quite large and thus, the proposed testing procedure is not seriously restricted. At least one of the conditions (5) and (6) is needed to control the asymptotic variance of G_N and to derive the asymptotic distribution of G_N, given in Theorem 1.

THEOREM 1:   *Under the nonparametric model (3), condition (4) and either condition (5) or condition (6)*

$$\frac{G_N - \text{tr}(\mathbf{M}^T\mathbf{MP})}{\sqrt{\text{Var}[G_N]}} \rightsquigarrow \text{N}(0, 1)$$

where $\rightsquigarrow$ denotes convergence in distribution as $N \to \infty$ and $rc = r(N)c(N) \to \infty$.

Consequently, under H_0 *in (2),*

$$\frac{G_N}{\sqrt{2\operatorname{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right)/[N(N-1)]}} \rightsquigarrow \mathrm{N}(0,1).$$

The final step to construct the test statistic is to estimate $\operatorname{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right)$ consistently. To do this without estimating the high-dimensional covariance matrix $\boldsymbol{\Sigma}$, note that $\operatorname{tr}\left([\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)]^2\right) = \operatorname{tr}(\boldsymbol{\Omega}^2)$ where $\boldsymbol{\Omega} = (\mathbf{P}\otimes\mathbf{I}_r)\boldsymbol{\Sigma}(\mathbf{P}\otimes\mathbf{I}_r)$ is the positive semi-definite covariance matrix of $\mathbf{Y}_i = \operatorname{vec}(\mathbf{X}_i\mathbf{P})$, and let

$$T_N = \frac{1}{D_2^N}\sum\nolimits_{i\neq j}(\mathbf{Y}_i^T\mathbf{Y}_j)^2 - 2\frac{1}{D_3^N}\sum\nolimits_{i\neq j\neq k}^{*}\mathbf{Y}_i^T\mathbf{Y}_j\mathbf{Y}_i^T\mathbf{Y}_k + \frac{1}{D_4^N}\sum\nolimits_{i\neq j\neq k\neq l}^{*}\mathbf{Y}_i^T\mathbf{Y}_j\mathbf{Y}_k^T\mathbf{Y}_l$$

where $D_t^s = (s-t)!/s!$ and $\sum^{*}$ denotes summation over mutually exclusive indices. Since $\operatorname{tr}(\boldsymbol{\Omega}^4) = o\left\{\operatorname{tr}^2(\boldsymbol{\Omega}^2)\right\}$, $\mathrm{E}[T_N] = \operatorname{tr}(\boldsymbol{\Omega}^2)$ and $\operatorname{Var}[T_N] = o\{\operatorname{tr}^2(\boldsymbol{\Omega}^2)\}$, it follows that T_N is a ratio-consistent estimator of $\operatorname{tr}(\boldsymbol{\Omega}^2)$. Therefore, the proposed test rejects H_0 in (2) with an α significance level if and only if

$$G_N^{*} = \frac{G_N}{\sqrt{2T_N/[N(N-1)]}} \geqslant z_a,$$

where z_a is the upper α-quantile of $\mathrm{N}(0,1)$.

To study some properties of G_N^{*}, consider the transformation $\mathbf{X}_i \longmapsto a\mathbf{A}\mathbf{X}_i + \mathbf{C}$ where $a \neq 0 \in \Re$, $\mathbf{A}$ is an $r \times r$ orthogonal matrix and $\mathbf{C}$ is an $r \times c$ matrix of constants such that $\mathbf{CP} = \mathbf{0}_{r\times c}$, and where $\mathbf{0}_{s\times t}$ denotes the zero matrix of size $s \times t$. As desired, the test statistic is invariant to (i) orthogonal rotations of the row variables (ii) to scalar multiplication and (iii) to a location shift of the mean matrix under H_0 in (2). In addition, (iii) also implies that the nominal size of the test statistic is not affected by the magnitude of the true mean matrix $\mathbf{M}$ given that this satisfies H_0 in (2). Although the test statistic was derived for testing the mean structure of row variables across groups of column variables, we emphasize that the same procedure can be used to test the mean structure of column variables across groups of row variables. For doing this, one needs to apply the transformation $\mathbf{X}_i \longmapsto \mathbf{X}_i^T$ prior to calculating the test statistic.

2.1 *Asymptotic power*

Investigation of the asymptotic power of the proposed test relies on Theorem 1. Under condition (5), the leading order power is

$$\beta_N = \Phi\left(-z_a + N\frac{\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})}{\sqrt{2\text{tr}(\mathbf{\Omega}^2)}}\right),$$

where Φ is the cumulative distribution function of $N(0,1)$. The power of the proposed test is bounded since

$$\Phi\left(-z_a + N\frac{\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})}{\sqrt{2\text{tr}(\mathbf{\Sigma}^2)}}\right) \leqslant \beta_N \leqslant \Phi\left(-z_a + N\frac{\text{tr}(\mathbf{M}^T\mathbf{M})}{\sqrt{2\text{tr}(\mathbf{\Omega}^2)}}\right),$$

and thus a sufficient condition for the proposed test to have non-trivial power is

$$\lim_{N,(rc)\to\infty} N\frac{\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})}{\sqrt{2\text{tr}(\mathbf{\Sigma}^2)}} > 0.$$

To gain insight into the asymptotic power, assume independent row and column variables and no row-effect in the mean structure, that is $\mu_{ab} = \mu_b$ where μ_{ab} is the (a,b)-th element of $\mathbf{M}$. Then

$$N\frac{\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})}{\sqrt{2\text{tr}(\mathbf{\Omega}^2)}} = N\sqrt{\frac{r}{2(c-g)}}\sum_{k=1}^{g}\sum_{b=c_{k-1}+1}^{c_k}(\mu_b - \bar{\mu}^{(k)})^2,$$

where $c_0 = 0$ and $\bar{\mu}^{(k)}$ is the average of the mean of the row variable a in group k. Contrary to univariate tests that require multiple testing correction, the asymptotic power of the proposed test is an increasing function of the number of row variables r regardless of the differences $\mu_b - \bar{\mu}^{(k)}$. Thus, in high-dimensional settings it should be more beneficial to utilize the global test proposed herein than to apply univariate tests, a speculation that was empirically verified in simulations.

Under condition (6), similar conclusions can be drawn about the power of the test statistic since the leading order power term becomes

$$\beta_N = \Phi\left(N\frac{\text{tr}(\mathbf{M}^T\mathbf{M}\mathbf{P})}{\sqrt{2\text{tr}(\mathbf{\Omega}^2)}}\right).$$

2.2 *Class of covariance matrices under consideration*

Condition (4) defines the class of covariance matrices to which the proposed test is applicable. To provide a few examples of this class, note that if $\mathrm{tr}(\mathbf{\Sigma}^4) = o\left\{\mathrm{tr}^2(\mathbf{\Sigma}^2)\right\}$ then condition (4) holds. This includes the cases where $\mathbf{\Sigma}$ has bounded eigenvalues or it has a few eigenvalues that diverge slowly to infinity (Chen and Qin, 2010) or when $\mathbf{\Sigma}$ implies a (banded) first order autoregressive correlation pattern such that the rc variances are bounded away from 0 or ∞ (Chen et al., 2010). In addition, condition (4) is also satisfied when $\mathbf{\Sigma}$ is a compound symmetry correlation matrix, i.e. $\mathbf{\Sigma} = \rho\mathbf{I}_{rc} + (1-\rho)\mathbf{J}_{rc}$ for $-1/(rc-1) < \rho \leqslant 1$. As before, this result extends to the case where $\mathbf{\Sigma}$ satisfies a compound symmetry correlation pattern and the variances are bounded away from 0 and ∞.

Because of the popularity of the matrix-variate normal distribution in modelling transposable data, we study the implications of condition (4) when $\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$. In this case, condition (4) becomes

$$\mathrm{tr}\left[(\mathbf{P}\mathbf{\Sigma}_2)^4\right]\mathrm{tr}(\mathbf{\Sigma}_1^4) = o\left\{\mathrm{tr}^2\left[(\mathbf{P}\mathbf{\Sigma}_2)^2\right]\mathrm{tr}^2(\mathbf{\Sigma}_1^2)\right\},$$

which is met if, for example, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ satisfy any of the covariance structures mentioned above. Therefore, a Kronecker product dependence pattern for $\mathbf{\Sigma}$ does not seem to impose restrictions on applying the proposed testing procedure.

When the number of row or column variables is fixed, the class of covariance matrices for $\mathbf{\Sigma}$ might be slightly restricted depending on which set of variables (row or column) is fixed. To see this, suppose that $\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$. If the number of column variables is fixed, then condition (4) becomes $\mathrm{tr}(\mathbf{\Sigma}_1^4) = o\left\{\mathrm{tr}^2(\mathbf{\Sigma}_1^2)\right\}$, and it follows that $\mathbf{\Sigma}_1$ cannot satisfy a compound symmetry correlation structure. However, if the number of row variables is fixed, then condition (4) becomes $\mathrm{tr}\left[(\mathbf{P}\mathbf{\Sigma}_2)^4\right] = o\left\{\mathrm{tr}^2\left[(\mathbf{P}\mathbf{\Sigma}_2)^2\right]\right\}$, and therefore the compound symmetry correlation structure is an acceptable dependence structure for $\mathbf{\Sigma}_2$.

2.3 *Computational details*

The proposed testing procedure does not require explicit estimation of the high dimensional matrix parameters $\mathbf{M}$ or $\boldsymbol{\Sigma}$ but only of $\operatorname{tr}(\boldsymbol{\Omega}^2)$. This is computed efficiently through T_N by applying the transformation $\mathbf{X}_i \longmapsto \mathbf{X}_i \mathbf{P}$. Although the provided form of T_N is computationally intensive for large N, the task of calculating T_N can be greatly simplified as shown in Himeno and Yamada (2012).

One way to calculate G_N is to apply first the transformation $\mathbf{X}_i \longmapsto \mathbf{X}_i \mathbf{P}$ and then use the corresponding expression given in the Appendix. However, if $c >> r$, then it seems more sensible to ignore the transformation step and calculate G_N by applying the circular property of the trace operator, i.e. writing $\operatorname{tr}(\mathbf{X}_i^T \mathbf{X}_j \mathbf{P})$ as $\operatorname{tr}(\mathbf{X}_j \mathbf{P} \mathbf{X}_i^T)$.

To this end, note that column groups of size one do not contribute to the test statistic, meaning that the value of G_N^* does not change if column groups of size one ($c_k = 1$) are ignored. This is not surprising since no mean comparisons are performed therein. Hence, these column variables should be removed prior to calculating the test statistic.

3. Simulation Studies

We investigated the nominal size and the power of the proposed testing procedure using simulations. The simulated random matrices $\mathbf{X}_1, \ldots, \mathbf{X}_N$ satisfied model (3). To study the nonparametric nature of the proposed methodology, three distributional scenarios were considered for the elements of $\mathbf{Z}_i$:

Scenario 1: A normality scenario, in which $Z_{iab} \overset{i.i.d}{\sim} N(0, 1)$.

Scenario 2: A centralized gamma distributed scenario, in which $Z_{iab} = (Z_{iab}^* - 8)/4$ and $Z_{iab}^* \overset{i.i.d}{\sim} Gamma(4, 0.5)$.

Scenario 3: A mixture of Scenarios 1 and 2, in which the random variables in the upper

half of $\mathbf{Z}_i$ are distributed as in Scenario 1, while the remaining random variables are distributed as in Scenario 2.

Conditional on N, $\mathbf{M}$, $\boldsymbol{\Sigma}$ and the distributional scenario, we draw 1000 replicates while keeping the significance level fixed at 5%. For each competing testing procedure, we calculated the empirical size as the proportion of rejections when $\mathbf{M} = \mathbf{0}_{r \times c}$ and the empirical power as the proportion of rejections when $\mathbf{M} \neq \mathbf{0}_{r \times c}$ as defined in Sections 3.1 and 3.3. To distinguish the test statistics of the proposed methodology used in the simulations, we denoted by $H_{\{c_1,c_2,\ldots,c_g\}}$ the test statistic G_N^* of the proposed methodology based on $\mathbf{P}_{\{c_1,c_2,\ldots,c_g\}}$. Further, we let $[t]$ denote the integer part of $t \in \Re$.

3.1 *Comparison with univariate tests that require multiple testing correction*

For testing the hypothesis of no column effect in the mean matrix, i.e. testing hypothesis (1), we compared the proposed testing procedure, evaluated using $H_{\{c\}}$, to univariate tests that require multiple testing correction.

First, we considered the ANOVA test of no group effect and the Kruskal-Wallis test. These tests were applied sequentially to each of the r row variables and the resulting p-values were adjusted using the false discovery rate (FDR) correction and the Bonferroni (BON) correction. To ensure a fair comparison, no dependence structure ($\boldsymbol{\Sigma} = \mathbf{I}_{rc}$) was assumed between and among the row and column variables. To evaluate the empirical power of the three competing testing procedures, we used $\mathbf{M} = [\mathbf{0}_{r \times 7}, t\mathbf{J}_{r \times 3}]$ where $\mathbf{J}_{k \times l}$ denotes the $k \times l$ matrix of ones. The constant t was selected such that $\mathrm{tr}(\mathbf{M}^T\mathbf{M})/\sqrt{r(c-1)} = 0.1$, i.e. by fixing the quantity that determines the upper bound of the asymptotic power of any test statistic of the proposed methodology equal to 0.1. Table 1 displays the results under Scenario 3 - similar patterns were observed under the other two scenarios. Unlike the Kruskal-Wallis test which seemed to be conservative unless $N = 100$, the empirical sizes for $H_{\{c\}}$ and the ANOVA test appeared to be a good approximation of the nominal size even for

$N = 10$. Despite the conservativeness of the proposed test for $N = 10$, it was always more powerful than the ANOVA and the Kruskal-Wallis test. In fact, the power of the proposed test increased as r increased while that of the competing testing procedures decreased. This trend validates our speculations about the power of the proposed test over univariate tests in high-dimensional settings.

[Table 1 about here.]

Next, note that if the column variables are independent, then the columns of the N random matrices can be classified in c groups, one group for each column variable, such that each group contains N independent r-variate random vectors. In this case, an alternative approach to test hypothesis (1) is to apply the two-sample mean test proposed by Chen and Qin (2010) to all possible pairs of groups, and then adjust the resulting p-values for multiple testing. To satisfy the required assumptions of the Chen-Qin test, $\mathbf{\Sigma}$ was set equal to a block diagonal matrix with c blocks. Each block of $\mathbf{\Sigma}$ satisfied a first-order autoregressive form ($\{\rho^{|a-b|}\}_{1 \leqslant a,b \leqslant r}$) where $\rho = 0.5$ in the first $c/2$ blocks and $\rho = 0.4$ elsewhere. Table 2 shows the empirical sizes of the two competing testing procedures across the three distributional scenarios. The Chen-Qin test appeared to have a highly inflated empirical size even when $r = 500$ while the proposed test seemed to preserve the nominal size. For this reason, we did not proceed to power investigations.

[Table 2 about here.]

3.2 *Nominal size*

Using $H_{\{c\}}$, $H_{\{[0.7c],[0.3c]\}}$ and $H_{\{[0.5c],[0.2c],[0.3c]\}}$, we examined in greater detail the size of the proposed methodology. For the dependence structure, we assumed that $\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$ where $\mathbf{\Sigma}_1 = \{0.85^{|a-b|}\}_{1 \leqslant a,b \leqslant r}$ and $\mathbf{\Sigma}_2 = 0.5(\mathbf{I}_c + \mathbf{J}_c)$. We also employed an exchangeable form for $\mathbf{\Sigma}$ but the simulation results were similar and so we do not present them here. To reflect

practical situations where the dimension of the mean vector is at least equal to the sample size (N) and the number of row variables (r) is greater or equal to the number of column variables (c), we set $N = 10, 30, 50, 100$, $r = 100, 500$ and $c = 10, 100$. Also, we covered the case where the number of row variables is much smaller than the number of column variables by using $r = 10$ and $c = 100, 500$. Table 3 contains the empirical sizes under Scenario 3. Again, similar results were observed for the other two distributional scenarios, a fact that validates empirically the non-parametric nature of the methodology. The discrepancy between the empirical and nominal size was small for all three test statistics. As desired, this confirms the robustness of the testing procedure to the number of groups and to the group sizes.

[Table 3 about here.]

3.3 *Power considerations*

Using $H_{\{c\}}$, $H_{\{[0.6c],[0.4c]\}}$ and $H_{\{[0.4c],[0.2c],[0.4c]\}}$, we evaluated the empirical power of the proposed methodology under two different configurations of the mean matrix. First, we defined $\mathbf{M} = [\mathbf{0}^T_{r \times [0.7c]}, \boldsymbol{\mu} \mathbf{1}^T_{[0.3c]}]$ and similarly to Chen and Qin (2010), we let $\boldsymbol{\mu}$ contain a varying proportion (0%, 25%, 50%, 75%, 95% and 99%) of non-zero elements. At each proportion level, we employed two types of allocations for the non-zero elements: (i) equal allocation and (ii) linearly increasing allocation where two nonzero-elements of $\boldsymbol{\mu}$ satisfy $\mu_{l_1} < \mu_{l_2}$ if and only if $l_1 < l_2$. We set $r = 100$, $c = 10$ and we used a Kronecker product form for $\boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma}_1 = \{0.8^{|a-b|}\}_{1 \leqslant a,b \leqslant r}$ and $\boldsymbol{\Sigma}_2 = 0.5(\mathbf{I}_c + \mathbf{J}_c)$. To make the results comparable across the different proportion levels, the non-zero elements of $\boldsymbol{\mu}$ were defined in such a way that

$$\frac{\text{tr}(\mathbf{M}^T \mathbf{M})}{\sqrt{\text{tr}(\boldsymbol{\Sigma}_1^2)\text{tr}(\boldsymbol{\Sigma}_2^2)}} = 0.1.$$

Table 4 displays the simulation results for $H_{\{6,4\}}$. Similar trends occurred for $H_{\{4,2,4\}}$ but not for $H_{\{10\}}$, which was extremely powerful in these settings. This indicates that as we move away from H_0, the power of the proposed methodology increases. Conditional on the sample size, the empirical power was similar across the three distributional scenario and it did not

depend on the type of allocation or the proportion level. The proposed testing procedure was powerful to the sparsity scenarios considered and their empirical power approached 1.00 as N increased.

[Table 4 about here.]

The second configuration of $\mathbf{M}$ we employed involved a multiplicative mean vectors scenario in which $\mathbf{M} = [\boldsymbol{\mu}\mathbf{1}_{[0.9c]}^{T}, t\boldsymbol{\mu}\mathbf{1}_{[0.1c]}^{T}]$ with $t \neq 0 \in \Re$. In this simulation scheme, the following settings were used: $\boldsymbol{\mu} = \mathbf{1}_r$, $t = 1.15$, $\boldsymbol{\Sigma}_1 = \{0.85^{|a-b|}\}_{1\leqslant a,b\leqslant r}$ and $\boldsymbol{\Sigma}_2 = 0.5(\mathbf{I}_c + \mathbf{J}_c)$ for $r = 100, 500$ and $c = 10, 100$. Table 5 displays the simulation results based on $H_{\{c\}}$ across the three distributional scenarios. The tests based on $H_{\{[0.6c],[0.4c]\}}$ and $H_{\{[0.4c],[0.2c],[0.4c]\}}$ were more powerful and hence we do not show these results. In addition to the desired trends mentioned above, the test appeared to be consistent as the number of row and/or column variables increased.

[Table 5 about here.]

4. Two Examples

We applied our proposed testing methodology to two datasets.

4.1 *The glioblastoma dataset*

The glioblastoma (GB) dataset describes an experimental study designed to investigate the heterogeneity of GB (Sottoriva et al., 2013). For each of the patients ($N = 8$) included in the study, mRNA samples were extracted from the tumor mass, the tumor margin and the subventricular zone. From the tumor mass, 5 RNA samples were collected from different fragments such that earlier fragments were closer to the tumor margin and later fragments closer to the subventricular zone. Gene expression levels were then measured from the mRNA samples using microarrays. The data for each subject were organized in a matrix where the

row variables ($r = 16810$) correspond to genes and the column variables ($c = 7$) to the margin, the subventricular zone and the five tumor fragments ordered in the spatial order described above.

A biological objective in this study was to assess whether for each gene the mean expression level was constant across the tumor fragments, i.e., testing the hypothesis $H_0 : \mathbf{M} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3 \mathbf{1}_5^T]$, where $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ are unknown r-variate vectors. The test statistic equals -0.282 (p-value$= 0.611$), meaning that we fail to reject the null hypothesis. Since we rejected any hypothesis that implies a simpler mean structure than the one tested (results not shown), it seems sensible to assume that the mean vectors of the margin, subventricular zone and tumor fragments are not equal.

We investigated further the mean gene expression levels using Gene Ontology (GO) terms. The GO terms classify genes into groups such that the genes within a group are involved in the same biological process and therefore, similar behavior in the gene expression levels might be expected. From the 1316 gene groups in the GB dataset, we selected 166 groups that had more than 10 genes in order to be closer to the high-dimensional assumptions. To determine the mean structure in each of these groups, we employed the following stepwise strategy. First we tested the assumption of no differentially expressed genes in each group, that is the hypothesis of a constant mean vector across the margin, the subventricular zone and the tumor fragments. For the groups that we rejected the null hypothesis, we tested the hypothesis of a common mean vector between the subventricular zone and the tumor fragments. The gene groups where the null hypothesis was rejected were then tested for the hypothesis of a common mean vector between the margin and the subventricular zone. Finally, the groups that rejected all the null hypotheses were tested for the hypothesis of a common mean vector between the margin and the tumor fragments. Note that at each step of this procedure, we applied an FDR correction. In this way, we identified 5 gene groups

in which each gene had a constant mean expression level across the different tissue samples, 102 groups in which each gene had the same mean expression level in the subventricular zone and in the tumor fragments, 25 groups in which each gene had the same mean expression level in the margin and the subventricular zone, and 34 groups that satisfied the global mean relationship described in the previous paragraph.

4.2 *The mouse aging dataset*

The atlas of gene expression in the mouse aging data (Zahn et al., 2007) contains mouse mRNA gene expression levels measured in different tissues. For each mouse ($N = 40$), mRNA expression levels were extracted for $r = 8932$ genes from up to 16 tissues. Here, we considered $c = 9$ tissues (adrenal glands, cerebrum, hippocampus, kidney, lung, muscle, spinal cord, spleen and thymus) for which mRNA gene expression levels were available for all the mice.

Our goal was to determine if there exist groups of tissues that have similar gene expression levels. To explore this, first we tested the hypothesis of no tissue effect which was rejected since $G_N^* = 481.28$ (p-value< 0.001). This is not surprising since we do not expect conservation of the gene expression across the tissues. However, a small subset of genes called 'housekeeping' genes are presumably expressed at a relatively constant level across many or all known experimental conditions. Hence, an interesting question is whether they have constant gene expression levels across tissues. To do this, we created a list of 22 housekeeping genes by considering 8 genes that are commonly classified as housekeeping genes in the literature (de Jonge et al., 2007) and 14 genes that were suggested as housekeeping genes in de Jonge et al. (2007). We rejected the hypothesis of no tissue effect upon the mean expression level of the housekeeping genes ($G_N^* = 382.93$ and p-value< 0.001). Hence, the expression levels of the housekeeping genes in mice should not be considered to be constant across different tissues in mice.

5. Discussion

We proposed a novel non-parametric procedure to test the mean matrix in high-dimensional transposable data. In particular, our methodology can determine whether in each of the given groups of column variables the mean of every row variable remains constant. Of course, the role of the row and column variables is interchangeable in transposable data and hence the proposed tests can be applied to check the effect of the row variables upon the mean vector of the column variables. The simulation studies verified the robustness of the proposed testing procedure to the number of row or column groups, to the size of each group, to the number of column and row variables relative to the sample size, and to the underlying dependence structure between and among the row and column variables. Compared to univariate testing procedures that preserve the nominal size, the proposed test was substantially more powerful. In a sense, we developed a theoretically sound non-parametric testing procedure that extends the application of ANOVA based tests to high-dimensional matrix-valued random variables by making mild assumptions about the dependence structure. A practical advantage of the proposed test is the computationally simplicity because the cumbersome task of estimating high-dimensional matrix parameters is avoided.

We emphasize that a critical point in our proposal is the choice of the projection matrix $\mathbf{P}$. Although Theorem 1 holds for any projection matrix, say $\mathbf{P}^*$, to avoid trivial power under certain alternatives it is essential to require that $\mathbf{M}\mathbf{P}^* = \mathbf{0}_{r \times c}$ if and only if the corresponding null hypothesis is true. To illustrate this, note that an alternative way to test hypothesis (1) is to consider the projection matrix $\mathbf{P}^* = \mathbf{J}_c/c$ (instead of $\mathbf{P}_c = \mathbf{I}_c - \mathbf{J}_c/c$). It can be shown that the asymptotic power of the resulting test statistic is trivial if, for example, c is even and the mean vector is $\boldsymbol{\mu}$ for the odd columns of $\mathbf{M}$ and $-\boldsymbol{\mu}$ for the even columns. Thus attention is required when projection matrices other than the suggested ones are used.

It is important to note that it might be possible to modify the proposed methodology in

such a way that the same type of test statistic can be used to test hypotheses other than

hypothesis (2). For example, consider testing the hypothesis of a known $r \times c$ matrix of

constants $\mathbf{M}_0$, i.e. $\mathrm{H}_0 : \mathbf{M} = \mathbf{M}_0$. To do this, we can center the data by subtracting $\mathbf{M}_0$ and

then employ the test statistic G_N^* calculated using $\mathbf{P} = \mathbf{J}_c/c$. Another example is testing the

hypothesis $\mathrm{H}_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the unknown r-variate mean vectors

of the first and second column variable respectively, and $\boldsymbol{\mu}_0$ is an r-variate vector of known

constants. To test this hypothesis, one needs to subtract $\boldsymbol{\mu}_0$ from the first column of each

data matrix and then test hypothesis (2) with $g = 2$, $c_1 = 2$ and $c_2 = \ldots = c_g = 1$ using

the transformed data. In a similar way, the proposed method can be extended to test known

differences in the mean vectors of two or more column groups. In future work, we aim to

develop test statistics for hypotheses that cannot be directly handled by the proposed testing

methodology, e.g. the hypothesis of a mean-restricted matrix (Allen and Tibshirani, 2010),

that is $\mathbf{M} = \boldsymbol{\mu}\mathbf{1}_c^T + \boldsymbol{\nu}^T\mathbf{1}_r$ where $\boldsymbol{\mu}$ is an r-variate vector of constants and $\boldsymbol{\nu}$ is a c-variate

vector of constants.

From a practical point of view, we expect that the experimental design will dictate the null

hypothesis of interest about the mean-relationship between the row and column variables,

as was the case with the glioblastoma dataset. In applications like the mouse aging dataset,

where it is not clear which column groups should be formed under the null hypothesis, we

suggest the following simple strategy. First, test whether there is no column (or row) effect

upon the mean of the row (column) variables. If we fail to reject this hypothesis, assume

that the mean of the row (column) variables is independent of the column (row) variables.

Otherwise, perform the test that two column (row) variables have the same mean vector for

all pairs of column (row) variables, and then apply a multiple testing correction. If all the

adjusted p-values are very small, then assume an unstructured mean matrix $\mathbf{M}$. Otherwise,

record the column (row) pairs for which the adjusted *p*-values< 0.05, form *g* column (row) groups and test hypothesis (2) as defined by the *g* groups.

We plan to include our testing methodology in an R package but currently R functions are available upon request.

References

Allen, G. and Tibshirani, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *The Annals of Applied Statistics* **4,** 764–790.

Allen, G. and Tibshirani, R. (2012). Inference with transposable data: Modelling the effects of row and column correlations. *Journal of the Royal Statistical Society B* **74,** 721–743.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica* **6,** 311–330.

Carvalho, C. and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis* **2,** 69–97.

Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics* **38,** 808–835.

Chen, S., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association* **105,** 810–819.

Dawid, P. A. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68,** 265–274.

de Jonge, H., Fehrmann, R., de Bont, E., Hofstra, R., Gerbens, F., Kamps, W. A., de Vries, E., van der Zee, A., te Meerman, G., and ter Elst, A. (2007). Evidence based selection of housekeeping genes. *PLoS One* **2,** e898.

Efron, B. (2009). Are a set of microarrays independent of each other? *The Annals of Applied Statistics* **3,** 922–942.

Galecki, A. (1994). General class of covariance structures for two or more repeated factors

in longitudinal data analysis. *Communications in Statistics-Theory and Methods* **23,** 3105–3119.

Genton, M. G. (2007). Spatial-temporal analysis of multivariate environmental monitoring data. *Environmetrics* **18,** 681–695.

Gupta, A. and Nagar, D. (2000). *Matrix Variate Distributions.* Chapman & Hall/CRC.

Himeno, T. and Yamada, T. (2012). Estimations for some functions of covariance matrix in high dimension under non-normality. Technical report, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan.

Lee, K., Daniels, M., and Joo, Y. (2013). Flexible marginalized models for bivariate longitudinal ordinal data. *Biostatistics* **14,** 462–476.

Mardia, K. V. and Goodall, C. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. *Environmetrics* **6,** 347–385.

Naik, D. N. and Rao, S. S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *Journal of Applied Statistics* **28,** 91–105.

Sottoriva, A., Spiteri, I., Piccirillo, S., Touloumis, A., Collins, P., Marioni, J., Curtis, C., Watts, C., and Tavaré, S. (2013). Intra-tumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 4009–4014.

Teng, S. L. and Huang, H. (2009). A statistical framework to infer functional gene relationships from biologically interrelated microarray experiments. *Journal of the American Statistical Association* **104,** 465–473.

Zahn, J. M., Poosala, S., Owen, A. B., Ingram, D. K., Lustig, A., Carter, A., Weeraratna, A. T., Taub, D. D., Gorospe, M., Mazan-Mamczarz, K., Lakatta, E., Boheler, K., Xu, X., Mattson, M., Falco, G., Ko, S., Schlessinger, D., Firman, J., Kummerfeld, S., Wood III,

W., Zonderman, A., Kim, S., and Becker, K. (2007). AGEMAP: A gene expression database for aging in mice. *PLoS Genetics* **3,** 2326–2337.

<div align="center">APPENDIX</div>

<div align="center">*Sketch of the proof of Theorem 1*</div>

Without loss of generality, let $\mathbf{P}$ be an idempotent and symmetric matrix and define $\mathbf{Y}_i = \mathrm{vec}(\mathbf{X}_i\mathbf{P})$ for all i, where $\mathrm{E}[\mathbf{Y}_i] = \mathrm{vec}(\mathbf{MP})$ and $\mathrm{cov}[\mathbf{Y}_i] = \mathbf{\Omega} = (\mathbf{P} \otimes \mathbf{I}_r)\mathbf{\Sigma}(\mathbf{P} \otimes \mathbf{I}_r)$. Rewrite relations (4), (5) and (6) as $\mathrm{tr}(\mathbf{\Omega}^4) = o\left\{\mathrm{tr}^2(\mathbf{\Omega}^2)\right\}$, $\mathrm{vec}(\mathbf{M})^T\mathbf{\Omega}\mathrm{vec}(\mathbf{M}) = o\left\{\mathrm{tr}(\mathbf{\Omega}^2)/N\right\}$ and $\mathrm{tr}(\mathbf{\Omega}^2)/N = o\left\{\mathrm{vec}(\mathbf{M})^T\mathbf{\Omega}\mathrm{vec}(\mathbf{M})\right\}$ respectively, and note that

$$G_N = \frac{1}{N(N-1)}\sum_{i \neq j}\mathrm{tr}(\mathbf{X}_i^T\mathbf{X}_j\mathbf{P}) = \frac{1}{N(N-1)}\sum_{i \neq j}\mathbf{Y}_i^T\mathbf{Y}_j.$$

With this parameterization, the asymptotic distribution of $(G_N - \mathrm{E}[G_N])/\sqrt{\mathrm{Var}[G_N]}$ can be derived in a similar fashion as in the proof of Theorem 1 in Chen and Qin (2010).

Table 1

Empirical size and power of $H_{\{10\}}$, ANOVA and Kruskal-Wallis test at 5% significance.

| | | $H_{\{10\}}$ | | ANOVA | | | | Kruskal-Wallis | | | |
| | | | | FDR | | BON | | FDR | | BON | |
r	N	Power	Size	Power	Size	Power	Size	Power	Size	Power	Size
100	10	0.138	0.063	0.051	0.047	0.051	0.046	0.013	0.014	0.013	0.014
	30	0.412	0.057	0.091	0.045	0.088	0.045	0.062	0.040	0.060	0.039
	50	0.756	0.053	0.136	0.045	0.125	0.044	0.115	0.043	0.112	0.043
	100	0.997	0.044	0.319	0.047	0.294	0.045	0.317	0.048	0.285	0.047
500	10	0.186	0.063	0.075	0.066	0.075	0.066	0.011	0.008	0.011	0.008
	30	0.703	0.039	0.096	0.060	0.094	0.059	0.051	0.033	0.047	0.033
	50	0.974	0.040	0.102	0.042	0.093	0.040	0.082	0.026	0.077	0.026
	100	1.000	0.051	0.261	0.054	0.244	0.053	0.253	0.048	0.233	0.047

Table 2

Empirical size of $H_{\{10\}}$ and the Chen-Qin test at 5% significance.

		Scenario 1		Scenario 2		Scenario 3	
r	N	$H_{\{10\}}$	Chen-Qin	$H_{\{10\}}$	Chen-Qin	$H_{\{10\}}$	Chen-Qin
100	10	0.066	0.179	0.048	0.179	0.065	0.173
	20	0.058	0.144	0.050	0.150	0.059	0.144
	30	0.069	0.147	0.059	0.157	0.056	0.158
	50	0.046	0.142	0.057	0.126	0.063	0.169
500	10	0.059	0.114	0.045	0.104	0.057	0.097
	20	0.046	0.115	0.051	0.090	0.054	0.091
	30	0.046	0.084	0.054	0.081	0.040	0.078
	50	0.050	0.091	0.054	0.090	0.050	0.077

Table 3
Empirical size of the proposed methodology under the mixture distributional scenario at 5% significance.

N	r	$H_{\{c\}}$		$H_{\{[0.7c],[0.3c]\}}$		$H_{\{[0.5c],[0.2c],[0.3c]\}}$	
	c	10	100	10	100	10	100
10	100	0.064	0.056	0.059	0.056	0.057	0.058
	500	0.068	0.067	0.068	0.067	0.060	0.067
30	100	0.063	0.053	0.061	0.050	0.060	0.049
	500	0.049	0.054	0.053	0.048	0.049	0.049
50	100	0.058	0.046	0.059	0.048	0.064	0.048
	500	0.060	0.058	0.066	0.062	0.054	0.059
100	100	0.047	0.055	0.050	0.053	0.057	0.058
	500	0.047	0.048	0.049	0.040	0.048	0.044
	c	100	500	100	500	100	500
10	10	0.055	0.065	0.052	0.067	0.052	0.068
30	10	0.061	0.059	0.057	0.058	0.057	0.055
50	10	0.054	0.053	0.057	0.053	0.056	0.054
100	10	0.062	0.045	0.065	0.045	0.058	0.045

Table 4
Empirical power of $H_{\{6,4\}}$ for the sparsity scenario with $r = 100$ at 5% significance.

N	$\#\{\mu_{1l} = \mu_{2l}\}$	Equal Allocation			Increasing Allocation		
		Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
10	99%	0.194	0.213	0.173	0.194	0.213	0.173
	95%	0.175	0.207	0.164	0.172	0.213	0.163
	75%	0.166	0.204	0.171	0.168	0.205	0.173
	50%	0.174	0.211	0.169	0.172	0.207	0.169
	25%	0.173	0.203	0.170	0.169	0.203	0.168
	0%	0.167	0.201	0.165	0.164	0.199	0.166
30	99%	0.605	0.609	0.606	0.605	0.609	0.606
	95%	0.626	0.582	0.605	0.623	0.589	0.605
	75%	0.632	0.634	0.635	0.637	0.637	0.642
	50%	0.643	0.646	0.649	0.651	0.648	0.650
	25%	0.645	0.647	0.654	0.647	0.645	0.653
	0%	0.658	0.644	0.663	0.662	0.643	0.666
50	99%	0.903	0.868	0.882	0.903	0.868	0.882
	75%	0.896	0.897	0.899	0.904	0.898	0.896
	50%	0.938	0.936	0.934	0.947	0.941	0.936
	25%	0.962	0.955	0.949	0.965	0.958	0.954
	5%	0.964	0.959	0.954	0.967	0.964	0.958
	0%	0.965	0.966	0.961	0.969	0.967	0.963

Table 5
Empirical power of $H_{\{c\}}$ for the multiplicity scenario at 5% significance.

	c	10	100	10	100	10	100
N	r	Scenario 1		Scenario 2		Scenario 3	
10	100	0.097	0.317	0.128	0.282	0.103	0.303
	500	0.210	0.778	0.207	0.813	0.206	0.781
30	100	0.291	0.975	0.313	0.964	0.294	0.966
	500	0.809	1.000	0.782	1.000	0.790	1.000
50	100	0.590	1.000	0.551	1.000	0.576	1.000
	500	0.997	1.000	0.992	1.000	0.998	1.000