

# Maximum pairwise Bayes factors for covariance structure testing\*

Kyoungjae Lee

*Department of Statistics, Sungkyunkwan university*  
*e-mail: [leekjstat@gmail.com](mailto:leekjstat@gmail.com)*

Lizhen Lin

*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame*  
*e-mail: [lizhen.lin@nd.edu](mailto:lizhen.lin@nd.edu)*

and

David Dunson

*Department of Statistical Science, Duke University*  
*e-mail: [dunson@duke.edu](mailto:dunson@duke.edu)*

**Abstract:** Hypothesis testing of structure in covariance matrices is of significant importance, but faces great challenges in high-dimensional settings. Although consistent frequentist one-sample covariance tests have been proposed, there is a lack of simple, computationally scalable, and theoretically sound Bayesian testing methods for large covariance matrices. Motivated by this gap and by the need for tests that are powerful against sparse alternatives, we propose a novel testing framework based on the maximum pairwise Bayes factor. Our initial focus is on one-sample covariance testing; the proposed test can optimally distinguish null and alternative hypotheses in a frequentist asymptotic sense. We then propose diagonal tests and a scalable covariance graph selection procedure that are shown to be consistent. A simulation study evaluates the proposed approach relative to competitors. We illustrate advantages of our graph selection method on a gene expression data set.

**MSC 2010 subject classifications:** Primary 62C20, 62F15; secondary 62C12.

**Keywords and phrases:** Bayesian hypothesis test, covariance structure testing, modularization.

Received November 2020.

## 1. Introduction

Consider a sample of observations from a high-dimensional normal model

$$X_1, \dots, X_n \mid \Sigma_n \stackrel{i.i.d.}{\sim} N_p(0, \Sigma_n), \quad (1)$$

---

\*We would like to acknowledge the generous support of NSF grants DMS CAREER 1654579 and DMS 2113642. This research was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2020R1A4A1018207).

where  $\Sigma_n \in \mathbb{R}^{p \times p}$  is a covariance matrix. There is often interest in inferring the structure in  $\Sigma_n$  and in comparing different alternative covariance structures. This article focuses on this problem from a hypothesis testing perspective. Let  $X = (X_1, \dots, X_n)^T \in \mathbb{R}^{n \times p}$  be the data matrix. A *one-sample covariance test* can be reduced to the following simple form:

$$H_0 : \Sigma_n = I_p \quad \text{versus} \quad H_1 : \Sigma_n \neq I_p, \quad (2)$$

by noting that  $H_0 : \Sigma_n = I_p$  is equivalent to a null hypothesis  $H_0 : \Sigma_n = \Sigma_\star$  for a known positive definite matrix  $\Sigma_\star$  by applying the linear transformation  $X_i \mapsto \Sigma_\star^{-1/2} X_i$ .

Another important problem is testing diagonality

$$H_0 : \sigma_{ij} = 0 \text{ for any } i \neq j \quad \text{versus} \quad H_1 : \text{not } H_0,$$

where  $\Sigma_n = (\sigma_{ij})$ . Finally, we consider the problem of *support recovery*, corresponding to estimating the nonzero elements of covariance matrices.

We are interested in constructing novel Bayesian procedures that are practically applicable with theoretical guarantees for the (i) one-sample covariance test, (ii) diagonality test, and (iii) support recovery of the covariance matrix. We consider the high-dimensional setting in which the number of variables  $p$  can grow to infinity as the sample size  $n$  gets larger and possibly be much larger than  $n$ . Although it is well known that assuming a restricted covariance class is necessary for consistent *estimation* of large covariance matrices (Johnstone and Lu, 2009; Lee and Lee, 2018), in a *testing* context we focus on alternative hypotheses  $H_1$  that are unconstrained. One natural possibility is to assume a conjugate inverse-Wishart prior  $IW_p(\nu_n, A_n)$  for  $\Sigma_n$  under  $H_1$ . However, in order for the resulting posterior to be proper, it is necessary to choose the degrees of freedom  $\nu_n > p - 1$ , suggesting an extremely informative prior in high-dimensional settings. The resulting test will certainly be highly sensitive to the choice of  $A_n$ , and hence is not very useful outside of narrow applications having substantial prior information. One could instead choose a non-conjugate prior for  $\Sigma_n$  under  $H_1$ , but then substantial computational issues arise in attempting to estimate the Bayes factor.

From a frequentist perspective, Chen, Zhang and Zhong (2010) and Cai and Ma (2013) suggested consistent one-sample covariance tests based on unbiased estimators of  $\|\Sigma_n - I_p\|_F^2$ , where  $\|A\|_F = (\sum_{ij} a_{ij}^2)^{1/2}$  is the Frobenius norm of a matrix  $A = (a_{ij})$ . Under the null hypothesis, they showed that their test statistic is asymptotically normal. The test also has power tending to one as  $n$  goes to infinity, but it requires the condition,  $\|\Sigma_n - I_p\|_F^2 n/p \rightarrow \infty$  as  $n \rightarrow \infty$ . This condition implies that they essentially adopted  $H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_F^2 \geq b_n p/n\}$  for some  $b_n \rightarrow \infty$  as  $n \rightarrow \infty$  as the alternative class. Cai and Ma (2013) proved that if we consider an alternative class  $H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_F^2 \geq \epsilon_n\}$ , say a dense alternative, the condition  $\epsilon_n \geq b_n p/n$  is inevitable for any level  $\alpha$  test to have power tending to one. This excludes cases in which a finite number of the components of  $\Sigma_n - I_p$  have a magnitude  $(p/n)^{1/2}$ , although  $(p/n)^{1/2}$  can be a significant signal when  $p \geq n$ .

The above discussion motivates us to develop hypothesis tests that are easy to implement in practice while possessing theory guarantees. In particular, we wish to construct tests that can perform well even when the condition  $\|\Sigma_n - I_p\|_F^2 n/p \rightarrow \infty$  fails to hold. We achieve this by proposing a novel Bayesian testing framework based on the maximum pairwise Bayes factor which will be introduced in Section 2.2. The basic strategy is to focus on the pairwise difference between  $\Sigma_n$  and  $I_p$  rather than the Frobenius norm or other matrix norms. More precisely, instead of considering a usual Bayes factor based on a prior on the whole covariance matrix, we first consider the pairwise Bayes factors for each element of the matrix and combine them by taking a maximum over all possible pairs. This approach is analogous to frequentist tests based on maximum-type statistics (Jeng, Cai and Li, 2013; Enikeeva and Harchaoui, 2019). Our construction enables us to consider a different alternative class,  $H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_{\max}^2 \geq C \log p/n\}$  for some constant  $C > 0$ , say a sparse alternative, where  $\|A\|_{\max} = \max_{i,j} |a_{ij}|$  for a matrix  $A = (a_{ij})$ . When the primary interest is not on a collection of very weak signals, but on detecting at least one *meaningful signal*, our test is much more effective than the frequentist methods mentioned above.

We would like to clarify that although we use the terminologies “dense alternative” and “sparse alternative”, there is no sparsity assumption anywhere. They just indicate whether the alternative hypothesis is based on Frobenius or maximum norm. For example, we use the terminology “sparse alternative” because the maximum-type alternative considered in this paper is less stringent compared to the Frobenius-type alternative when the difference  $\Sigma_n - I_p$  is sparse.

The proposed testing method is general, easily implementable and theoretically supported, being the first Bayesian test shown to be consistent in the high-dimensional setting for the one-sample or diagonal covariance testing problems. Our procedure yields proven false discovery rate control and power improvement compared to existing methods. The proposed one-sample test is rate-optimal in the sense that it can distinguish the sparse alternative class  $H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_{\max}^2 \geq \epsilon_n\}$  from the null with the fastest rate of  $\epsilon_n$ , while guaranteeing consistency under the null. We also propose a scalable graph selection method for high-dimensional covariance graph models using pairwise Bayes factors. The proposed method consistently recovers the true covariance graph structure under a weaker or comparable condition to those in the existing frequentist literature.

Recently, Leday and Richardson (2018) suggested a fundamentally different pairwise approach to test marginal or conditional independence between two variables. Their focus is on the joint distribution of the  $i$ th and  $j$ th variables and an inverse-Wishart prior for  $\Sigma_n$  was imposed. For each  $i \neq j$ , the hypothesis testing problem  $H_{0,ij}^M : \sigma_{ij} = 0$  versus  $H_{1,ij}^M : \sigma_{ij} \neq 0$  was considered. Since the resulting Bayes factors for the pairwise tests are not scale-invariant, they proposed *scaled* versions. P-values under the *conditional null distribution* were obtained by shuffling or permuting labels of observations (Jiang, Ye and Liu, 2017). For support recovery, they suggest using standard multiplicity correction

procedures to control the false discovery rate, obtaining a frequentist procedure. Selection consistency results were not provided.

R code for implementation of our empirical results are available at <https://github.com/leekjstat/mxPBF>. Simulation studies for support recovery and proofs of our main results are included in the Appendix.

## 2. Preliminaries

### 2.1. Notations

For any real values  $a$  and  $b$ , we denote  $a \vee b$  as the maximum between  $a$  and  $b$ . For any positive sequences  $a_n$  and  $b_n$ , we denote  $a_n \ll b_n$  or  $a_n = o(b_n)$  if  $a_n/b_n \rightarrow 0$  as  $n \rightarrow \infty$ . For any vector  $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ , we define the vector  $\ell_1$ - and  $\ell_2$ -norm as  $\|x\|_1 = \sum_{j=1}^p |x_j|$  and  $\|x\|_2 = (\sum_{j=1}^p x_j^2)^{1/2}$ , respectively. Let  $\mathcal{C}_p$  be the set of all  $p \times p$  positive definite matrices. We denote  $\chi_k^2(\lambda)$  as the non-central chi-square distribution with degrees of freedom  $k$  and non-centrality  $\lambda \geq 0$ , and let  $\chi_k^2 = \chi_k^2(\lambda = 0)$ . For positive real values  $a$  and  $b$ ,  $IG(a, b)$  denotes the inverse gamma distribution with shape  $a$  and scale  $b$ .

### 2.2. Maximum pairwise Bayes factor

In this subsection, we introduce our approach focusing on the one-sample covariance test. As described before, the basic strategy is to concentrate on the pairwise difference between  $\Sigma_n$  and  $I_p$ . Let  $\tilde{X}_j \in \mathbb{R}^n$  be the  $j$ th column vector of  $X$ . For any indices  $i$  and  $j$  such that  $i \neq j$ , based on the joint distribution (1), the conditional distribution of  $\tilde{X}_i$  given  $\tilde{X}_j$  is

$$\tilde{X}_i \mid \tilde{X}_j \sim N_n\left(a_{ij}\tilde{X}_j, \tau_{ij}^2 I_n\right), \tag{3}$$

where  $a_{ij} \in \mathbb{R}$  and  $\tau_{ij} > 0$ . We can view (3) as a linear regression model given a design matrix  $\tilde{X}_j$ . For each paired conditional model (3), we consider a testing problem

$$H_{0,ij} : a_{ij} = 0, \tau_{ij}^2 = 1 \quad \text{versus} \quad H_{1,ij} : \text{not } H_{0,ij}. \tag{4}$$

If  $H_{0,ij}$  is true,  $\sigma_{ij} = 0$  and  $\sigma_{ii} = 1$  because  $a_{ij} = \sigma_{ij}/\sigma_{jj}$  and  $\tau_{ij}^2 = \sigma_{ii}(1 - \rho_{ij}^2)$ , where  $\Sigma_n = (\sigma_{ij})$  and  $R_n = (\rho_{ij})$  are covariance and correlation matrices, respectively. We suggest the following prior distribution under the alternative hypothesis  $H_{1,ij}$  in (4),

$$a_{ij} \mid \tau_{ij}^2 \sim N\left(0, \frac{\tau_{ij}^2}{\gamma} \|\tilde{X}_j\|_2^{-1}\right), \quad \tau_{ij}^2 \sim IG(\kappa_0, b_{0,ij}), \tag{5}$$

where  $\gamma = (n \vee p)^{-\alpha}$  and  $\kappa_0, b_{0,ij}$  and  $\alpha$  are positive constants. The induced Bayes factor is

$$B_{10}(\tilde{X}_i, \tilde{X}_j) = \frac{p(\tilde{X}_i \mid \tilde{X}_j, H_{1,ij})}{p(\tilde{X}_i \mid \tilde{X}_j, H_{0,ij})}$$

$$= \frac{b_{0,ij}^{\kappa_0}}{\Gamma(\kappa_0)} \left( \frac{\gamma}{1+\gamma} \right)^{1/2} \Gamma\left(\frac{n}{2} + \kappa_0\right) e^{n\hat{\tau}_i^2/2} \left( \frac{n}{2} \hat{\tau}_{ij,\gamma}^2 + b_{0,ij} \right)^{-n/2 - \kappa_0},$$

where one has  $n\hat{\tau}_i^2 = \|\tilde{X}_i\|_2^2$ ,  $n\hat{\tau}_{ij,\gamma}^2 = \tilde{X}_i^T \{I_n - (1+\gamma)^{-1}H_j\} \tilde{X}_i$  and  $H_j = \tilde{X}_j(\tilde{X}_j^T \tilde{X}_j)^{-1} \tilde{X}_j^T$ . More details for the derivation of  $B_{10}(\tilde{X}_i, \tilde{X}_j)$  can be found in Appendix D. The choice of hyperparameters  $\kappa_0$  and  $b_{0,ij}$  is discussed in Section 4.1.

The null hypothesis in the one-sample covariance test,  $H_0 : \Sigma_n = I_p$ , is true if  $H_{0,ij}$  is true for all pairs  $(i, j)$  such that  $i \neq j$ . We aggregate the information from each pairwise Bayes factor  $B_{10}(\tilde{X}_i, \tilde{X}_j)$  via the *maximum pairwise Bayes factor*,

$$B_{\max,10}(X) = \max_{i \neq j} B_{10}(\tilde{X}_i, \tilde{X}_j). \quad (6)$$

A large value for  $B_{\max,10}(X)$  provides evidence supporting the alternative hypothesis. By taking a maximum,  $B_{\max,10}(X)$  supports the alternative hypothesis if at least one of the pairwise Bayes factors supports the alternative. A natural question is whether false positives increase as we take a maximum over more and more pairs. Indeed, we find that this is not the case, either asymptotically based on our consistency results (Theorems 3.1 and 3.3) or in finite samples based on simulations.

### 3. Main results

#### 3.1. One-sample covariance test

In this subsection, we show consistency of  $B_{\max,10}(X)$  defined in (6) for the one-sample covariance test (2). We first introduce assumptions for consistency under  $H_1 : \Sigma_n \neq I_p$ . Let  $\Sigma_0 = (\sigma_{0,ij}) \in \mathcal{C}_p$  be the true covariance matrix, implying the conditional distribution of  $\tilde{X}_i$  given  $\tilde{X}_j$  is

$$\tilde{X}_i | \tilde{X}_j \sim N_n(a_{0,ij}\tilde{X}_j, \tau_{0,ij}^2 I_n) \quad (7)$$

under  $\mathbb{P}_0$ , where  $a_{0,ij} = \sigma_{0,ij}/\sigma_{0,jj}$ ,  $\tau_{0,ij}^2 = \sigma_{0,ii}\{1 - \sigma_{0,ij}^2/(\sigma_{0,ii}\sigma_{0,jj})\}$ ,  $\mathbb{P}_0$  is the probability measure corresponding to model (1) with  $\Sigma_n = \Sigma_0$ , and  $\tau_{0,ij}^2 = \sigma_{0,ii}$  if and only if  $a_{0,ij} = 0$ . Under the alternative  $H_1 : \Sigma_n \neq I_p$ , we assume that  $\Sigma_0$  satisfies *at least one* of the following conditions:

(A1) There exists a pair  $(i, j)$  satisfying

$$|\sigma_{0,ii} - 1| \geq \left[ 4\sigma_{0,ii}C_1^{1/2} + C_2 + \frac{2b_{0,ij}}{\{n \log(n \vee p)\}^{1/2}} \right] \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2} \quad (8)$$

for some constants  $C_1 > 0$  and  $C_2 > 2(\alpha + 2)^{1/2}$ .

(A2) There exists a pair  $(i, j)$  satisfying

$$|\tau_{0,ij}^2 - 1| \geq \left[ 4\tau_{0,ij}^2C_1^{1/2} + C_2 + \frac{2b_{0,ij} + \tau_{0,ij}^2}{\{n \log(n \vee p)\}^{1/2}} \right] \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2} \quad (9)$$

(A3) There exists a pair  $(i, j)$  satisfying

$$\sigma_{0,ij}^2 \geq \frac{\sigma_{0,jj}}{1 - 2C_1^{1/2}\epsilon_0} \left\{ \frac{9C_1\tau_{0,ij}^2}{(1 - C_3)^2} \vee \frac{C_4(\alpha + 2)}{C_3} \right\} \frac{\log(n \vee p)}{n} \quad (10)$$

for some constants  $0 < C_3 < 1$  and  $C_4 > 1$ .

Throughout the paper,  $C_1, C_2, C_3$  and  $C_4$  are fixed global constants. For a given small constant  $\epsilon > 0$ , they can be considered as  $C_1 = \epsilon, C_2 = 2(\alpha + 2)^{1/2} + \epsilon, C_3 = 1 - \epsilon^{1/4}$  and  $C_4 = 1 + \epsilon$ .

Condition (A1) is required to detect a non-unit variance  $\sigma_{0,ii}$ , and can be interpreted as a *beta-min condition* for  $|\sigma_{0,ii} - 1|$ . The beta-min condition gives a lower bound for nonzero parameters and is essential for model selection consistency (Castillo, Schmidt-Hieber and Van der Vaart, 2015; Martin, Mess and Walker, 2017). Interestingly, the rate of lower bound in (A1) is given by  $\{\log(n \vee p)/n\}^{1/2}$ , which has been commonly used in the variable selection literature. Condition (A2) is similar to condition (A1), which can be interpreted as a beta-min condition for  $|\tau_{0,ij}^2 - 1|$ . Condition (A3) is a beta-min condition for off-diagonal elements of the covariance matrix. In summary, conditions (A1)–(A3) imply the sparse alternative

$$\Sigma_0 \in H_1 = \left\{ \Sigma_n : \|\Sigma_n - I_p\|_{\max}^2 \geq C \frac{\log p}{n} \right\}$$

for some constant  $C > 0$ , which corresponds to the *meaningful* difference we mentioned earlier. When  $\log n \asymp \log p$ , the above alternative  $H_1$  is equivalent to conditions (A1)–(A3) in terms of the rate because  $\|\Sigma_0 - I_p\|_{\max}^2 \geq C \log p/n$  holds for some  $C > 0$  if and only if when at least one of conditions (A1)–(A3) holds. In fact, the rate  $\log p/n$  is *optimal* for guaranteeing the consistency under both hypotheses (Theorem 3.2). Our method is not designed to detect dense alternatives in which all differences are very small, but requires at least one difference to be sufficiently large.

Theorem 3.1 shows consistency for the one-sample covariance test even in the high-dimensional setting as long as  $\log p \leq \epsilon_0^2 n$  for some small constant  $\epsilon_0 > 0$ .

**Theorem 3.1.** *Consider model (1) and the one-sample covariance testing problem (2). Consider prior (5) under  $H_{1,ij}$  in (4) with  $\alpha > 8(1 + 2^{1/2}\epsilon_0)^2 / \{1 - 2^{3/2}\epsilon_0(1 + 2^{1/2}\epsilon_0)\}$  for some small constant  $0 < \epsilon_0 < 3(4C_2)^{-1}$ . Assume that  $\log p \leq \epsilon_0^2 n$  for all large  $n$ . Then under  $H_0 : \Sigma_n = I_p$ , for some constant  $C_{H_0} > 0$ ,*

$$B_{\max,10}(X) = O_p\{(n \vee p)^{-C_{H_0}}\}.$$

*If, under  $H_1 : \Sigma_n \neq I_p$ ,  $\Sigma_0$  satisfies at least one of conditions (A1)–(A3), for some constant  $C_{H_1} > 0$ ,*

$$B_{\max,10}(X)^{-1} = O_p\{(n \vee p)^{-C_{H_1}}\}.$$

The condition for  $\alpha$  in Theorem 3.1 is required to prove consistency of the maximum pairwise Bayes factor  $B_{\max,10}(X)$  under the null  $H_0 : \Sigma_n = I_p$ . We first prove that the pairwise Bayes factor  $B_{10}(\tilde{X}_i, \tilde{X}_j)$  is consistent on a large event  $E_{ij}$  such that  $\mathbb{P}_0(E_{ij}^c) \rightarrow 0$  as  $n \rightarrow \infty$ . To show consistency under  $H_0$ , it suffices to prove that  $\sum_{i \neq j} \mathbb{P}_0(E_{ij}^c) \rightarrow 0$  as  $n \rightarrow \infty$ , which means that the false discovery rate converges to zero. The condition for  $\alpha$  is closely related to this requirement. It also has connections with the variable selection literature in regression (Fernandez, Ley and Steel, 2001; Narisetty and He, 2014; Yang, Wainwright and Jordan, 2016) where the prior dispersion needs to depend on  $(n \vee p^2)$  or  $p$  for consistency. Note that as  $\alpha$  increases, so does the variance of the prior (5). Thus, the condition for  $\alpha$  implies that the variance of  $a_{ij}$  under the alternative hypothesis should rapidly increase at certain rate. More specifically, our theory requires a larger dispersion of order roughly  $(n \vee p)^8$  mainly due to the larger number of parameters compared to the regression setting.

To show consistency under  $H_1$ , it suffices to show  $\mathbb{P}_0(E_{ij}^c) \rightarrow 0$  as  $n \rightarrow \infty$  for some index  $(i, j)$  satisfying at least one of conditions (A1)–(A3). Interestingly, the rate of convergence is similar under both hypotheses, unlike most Bayesian testing procedures with the notable exception of non-local prior based methods (Johnson and Rossell, 2010, 2012).

The next theorem shows the optimality of the alternative class which is considered in Theorem 3.1 (Conditions (A1)–(A3)). It says, when the alternative class is defined based on the element-wise maximum norm, the condition  $\|\Sigma_0 - I_p\|_{\max}^2 \geq C \log p/n$  for some constant  $C > 0$  is necessary for any consistent test to have power tending to one. Thus, conditions (A1)–(A3) are rate-optimal to guarantee the consistency under  $H_0$  as well as  $H_1$ .

**Theorem 3.2.** *Let  $\mathbb{E}_\Sigma$  be the expectation corresponding to model (1). For a given constant  $C_\star > 0$ , define  $H_1(C_\star) = \{\Sigma \in \mathcal{C}_p : \|\Sigma - I_p\|_{\max}^2 \geq C_\star^2 \log p/n\}$ . If  $C_\star^2 \leq 2$ , then for any consistent test  $\phi$  such that  $\mathbb{E}_{I_p} \phi \rightarrow 0$  as  $n \rightarrow \infty$ ,*

$$\limsup_{n \rightarrow \infty} \inf_{\Sigma \in H_1(C_\star)} \mathbb{E}_\Sigma(\phi) \leq \frac{1}{2}.$$

To show that  $C_\star^2 = 2$  is “the” boundary constant above which type I and type II errors both converge to zero, it suffices to show that there exists a consistent test having power tending to 1 for  $\Sigma \in H_1(C_\star)$  when  $C_\star^2 > 2$ . For now, the maximum pairwise Bayes factor  $B_{\max,10}(X)$  is the only candidate that we have. However, for example, condition (A1) can be roughly considered as  $|\sigma_{0,ii} - 1|^2 \geq 4(\alpha + 2) \log(n \vee p)/n$ , which corresponds to  $C_\star^2 \geq 4(\alpha + 2) > 40$  due to the condition on  $\alpha$ . Based on current techniques used in the proof, it is not clear whether we can improve the constants in conditions (A1)–(A3) so that they match  $C_\star^2 = 2$ . It would be an interesting direction for the future research to investigate the optimal choice of  $C_\star > 0$ .

### 3.2. Testing diagonality

We now consider testing of diagonality of the covariance matrix:

$$H_0 : \sigma_{ij} = 0 \text{ for any } i \neq j \quad \text{versus} \quad H_1 : \text{not } H_0, \tag{11}$$

where  $\Sigma_n = (\sigma_{ij})$ . Note that the null hypothesis in (11) is composite as  $\sigma_{ii}$ 's are unknown. The above hypothesis testing problem can be modularized into many pairwise independence tests

$$H_{0,ij} : \sigma_{ij} = 0 \quad \text{versus} \quad H_{1,ij} : \sigma_{ij} \neq 0 \tag{12}$$

for all  $1 \leq i < j \leq p$ . We can adopt the maximum pairwise Bayes factor idea to aggregate the pairwise testing information from (12) for all possible pairs  $(i, j)$  such that  $i \neq j$  to test (11). Based on the conditional distribution (3), the null hypothesis  $H_{0,ij}$  in (12) is equivalent to  $H'_{0,ij} : a_{ij} = 0$ . We suggest the prior  $\pi(\tau_{ij}^2) \propto \tau_{ij}^{-2}$  under both  $H_{0,ij}$  and  $H_{1,ij}$ , and the prior  $\pi(a_{ij} | \tau_{ij}^2)$  defined in (5) under  $H_{1,ij}$ , which leads to the pairwise Bayes factor

$$\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = \left(\frac{\gamma}{1 + \gamma}\right)^{1/2} \left(\frac{\hat{\tau}_{ij,\gamma}^2}{\hat{\tau}_i^2}\right)^{-n/2}.$$

The improper prior  $\pi(\tau_{ij}^2) \propto \tau_{ij}^{-2}$  does not cause any problem because we use the same priors under  $H_{0,ij}$  and  $H_{1,ij}$ . In fact, theoretical results in this section still holds for the inverse gamma prior on  $\tau_{ij}^2$  with a slightly modified condition (A4). However, we use an improper prior  $\pi(\tau_{ij}) \propto \tau_{ij}^{-2}$  for the test of testing diagonality to reduce the number of hyperparameters. Note that the inverse gamma prior has two hyperparameters, while the improper prior does not have any hyperparameter.

We suggest using

$$\tilde{B}_{\max,10}(X) = \max_{i < j} \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) \tag{13}$$

for the hypothesis testing problem (11). Theorem 3.3 states the consistency of  $\tilde{B}_{\max,10}(X)$  for testing (11) under regularity conditions. For consistency under the alternative hypothesis, we assume the following condition:

(A4) There exists a pair  $(i, j)$  satisfying

$$\sigma_{0,ij}^2 \geq \frac{C_4 \sigma_{0,jj}}{1 - 2\epsilon_0 C_1^{1/2}} \left\{ \frac{9C_1 \tau_{0,ij}^2}{(1 - C_3)^2} \vee \frac{\alpha(1 + \gamma)(1 + 4\epsilon_0 C_1^{1/2}) \sigma_{0,ii}}{C_3} \right\} \frac{\log(n \vee p)}{n}$$

for constants  $C_1 > 0, 0 < C_3 < 1$  and  $C_4 > 1$  defined in Section 3.1.

**Theorem 3.3.** Consider model (1) and the diagonality testing problem (11). For a given pair  $(i, j)$  such that  $i \neq j$ , consider the prior  $\pi(\tau_{ij}^2) \propto \tau_{ij}^{-2}$  under both  $H_{0,ij}$  and  $H_{1,ij}$ , and the prior  $\pi(a_{ij} | \tau_{ij}^2)$  defined in (5) under  $H_{1,ij}$  in (12)



with  $\alpha > 4/(1 - 2^{1/2}3\epsilon_0)$  for some small constant  $0 < \epsilon_0 < 1/(2^{1/2}3)$ . Assume that  $\log p \leq \epsilon_0^2 n$  for all large  $n$ . Then under  $H_0 : \sigma_{ij} = 0$  for any  $i \neq j$ , for some constant  $c > 0$ ,

$$\tilde{B}_{\max,10}(X) = O_p\{(n \vee p)^{-c}\}.$$

If, under  $H_1 : \text{not } H_0$ ,  $\Sigma_0$  satisfies condition (A4), for some constant  $c' > 0$ ,

$$\tilde{B}_{\max,10}(X)^{-1} = O_p\{(n \vee p)^{-c'}\}.$$

Condition (A4) is the beta-min condition for off-diagonal elements of the true covariance matrix. It indicates that if one of the off-diagonal elements satisfies the beta-min condition (A4),  $\tilde{B}_{\max,10}(X)$  consistently detects the true alternative hypothesis. Similar to Theorem 3.1, the condition for  $\alpha$  is required to control the false discovery rate, and  $\tilde{B}_{\max,10}(X)$  has similar rates of convergence under both hypotheses.

Although the maximum pairwise Bayes factor idea is not limited to the test of diagonality, we introduce a few procedures that have been proposed for testing diagonality in the literature. Yao, Zhang and Shao (2018) and Leung and Drton (2018) proposed  $L_2$ -type tests for dependence in model-free settings. These tests are powerful against dense alternatives, while our focus is on the sparse setting. Han, Chen and Liu (2017) proposed two families of maximum-type rank tests of diagonality, which include Kendall's tau and Spearman's rho as special cases, respectively.

Although our procedure has a Bayesian motivation, one can use it as a frequentist test statistic. In the following proposition, we derive the *limiting null distribution* of the maximum pairwise Bayes factor in (13). This enables us to construct a test having size  $\alpha$  asymptotically.

**Proposition 1.** *Under the conditions of Theorem 3.3, further assume that  $p = p_n \rightarrow \infty$  as  $n \rightarrow \infty$  and  $\log p = o(n^{1/3})$ . If  $H_0 : \sigma_{ij} = 0$  for any  $i \neq j$  is true,  $2 \log \tilde{B}_{\max,10}(X) - C_{n,p}$  converges in distribution to a type I extreme value distribution with distribution function*

$$F(z) = \exp\{- (8\pi)^{-1/2} e^{-z/2}\}, \quad z \in \mathbb{R},$$

as  $n \rightarrow \infty$ , where  $C_{n,p} = 0.5 \log\{\gamma/(1 + \gamma)\} + 4 \log p - \log(\log p)$ .

### 3.3. Support recovery of covariance matrices

The primary interest of this section is on the recovery of  $S(\Sigma_0)$ , where  $S(\Sigma_0) \subseteq \{(i, j) : 1 \leq i < j \leq p\}$  is the nonzero index set of the true covariance matrix  $\Sigma_0$ . We call  $S(\Sigma_0)$  the *support* of  $\Sigma_0$ . Estimating  $S(\Sigma_0)$  corresponds to graph selection in covariance graph models (Cox and Wermuth, 1993). Despite its importance, few Bayesian articles have investigated this problem. Kundu et al. (2019) proposed the regularized inverse Wishart prior, which can be

viewed as a group Lasso penalty (Yuan and Lin, 2006) on the Cholesky factor. They showed the consistency of their selection procedure for the support of precision matrices when the dimension  $p$  is fixed. Recently, Gan, Narisetty and Liang (2018) adopted the spike-and-slab Lasso prior (Rockova and George, 2016; Rockova, 2018) for off-diagonal entries of the precision matrix. Their proposed graph selection procedure for the precision matrix also yields selection consistency. To the best of our knowledge, in the Bayesian literature, a consistent support recovery result for covariance matrices has not been established. Although Leday and Richardson (2018) proposed a graph selection procedure based on Bayesian modeling, their procedure relies on  $p$ -values and they do not show consistency.

To tackle this gap, we propose a scalable graph selection scheme for high-dimensional covariance matrices based on pairwise Bayes factors. Looking closely at the proof of Theorem 3.3, each pairwise Bayes factor  $\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)$  can consistently determine whether the corresponding covariance element  $\sigma_{0,ij}$  is zero or not. Thus, we suggest using the estimated index set

$$\hat{S}_{pair, C_{sel}} = \left\{ (i, j) : 2 \log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) > C_{sel}, \quad 1 \leq i < j \leq p \right\} \quad (14)$$

for some constant  $C_{sel} > 0$ . Although any threshold  $C_{sel}$  can be used for consistent selection asymptotically, the choice is crucial in practice. As a default method, we suggest using cross-validation to select  $C_{sel}$ , as described in detail in Section 4.3. A simulation study investigating the quality of support recovery for various threshold values can be found in Appendix C.

In the frequentist literature, Drton and Perlman (2004, 2007) proposed selection procedures using a related idea to (14), which select a graph by multiple hypothesis testing on each edge. However, they considered only the low-dimensional setting,  $n \geq p + 1$ .

For the consistency of  $\hat{S}_{pair, C_{sel}}$ , we introduce the following condition for some constants  $0 < C_3 < 1, C_4 > 1$  and  $C_5 > 2$ :

(A5) For a given pair  $(i, j)$  such that  $i \neq j$ ,

$$\sigma_{0,ij}^2 \geq \frac{C_4 \sigma_{0,jj}}{1 - 2\epsilon_0 C_5^{1/2}} \left[ \frac{9C_5 \tau_{0,ij}^2}{(1 - C_3)^2} \vee \frac{\alpha(1 + \gamma)(1 + 4\epsilon_0 C_5^{1/2})\sigma_{0,ii}}{C_3} \right] \frac{\log(n \vee p)}{n}.$$

The beta-min condition (A5) is almost the same as (A4) except using  $C_5 > 2$  instead of  $C_1 > 0$  to control the probabilities of small events on which the pairwise Bayes factor might not be consistent. Theorem 3.4 states that (14) achieves model selection consistency if condition (A5) holds with  $(i, j)$  or  $(j, i)$  for any  $(i, j) \in S(\Sigma_0)$ .

**Theorem 3.4.** Consider model (1) and prior (5) with  $\alpha > 4/(1 - 2^{1/2}3\epsilon_0)$  for some small constant  $0 < \epsilon_0 < (2^{1/2}3)^{-1}$  and each pair  $(i, j)$  such that  $i \neq j$ . Assume that  $\log p \leq \epsilon_0^2 n$  for all large  $n$  and condition (A5) holds with  $(i, j)$  or

$(j, i)$  for any  $(i, j) \in S(\Sigma_0)$ . Then, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_0(\widehat{S}_{pair, C_{sel}} = S(\Sigma_0)) = 1.$$

We note that  $\widehat{S}_{pair, C_{sel}}$  consistently recovers the support of the true covariance matrix  $\Sigma_0$  as long as  $\log p \leq \epsilon_0^2 n$  and nonzero entries satisfy the beta-min condition (A5). Rothman, Levina and Zhu (2009) proved a similar support recovery result for generalized thresholding of the sample covariance matrix while assuming  $\log p = o(n)$ ,  $\max_i \sigma_{0,ii} \leq M$  for some  $M > 0$  and  $\min_{(i,j) \in S(\Sigma_0)} \sigma_{0,ij}^2 \geq M' \log p/n$  for some sufficiently large  $M' > 0$ . Cai and Liu (2011) assumed  $\log p = o(n^{1/3})$  and  $\min_{(i,j) \in S(\Sigma_0)} \sigma_{0,ij}^2 \geq C \sigma_{0,ii} \sigma_{0,jj} \log p/n$  for some  $C > 0$  and obtained consistent support recovery using adaptive thresholding. Our condition,  $\log p \leq \epsilon_0^2 n$ , is much weaker than the conditions used in the literature. The beta-min condition (A5) is similar to that in Cai and Liu (2011) and also has the same rate to that in Rothman, Levina and Zhu (2009) if we assume  $\max_i \sigma_{0,ii} \leq M$  for some  $M > 0$ . Thus, the required condition in Theorem 3.4 is weaker or comparable to the conditions used in the literature.

## 4. Numerical results

### 4.1. Simulation study: One-sample covariance test

In this section, we demonstrate the performance of our one-sample covariance test in various simulation cases. For the hyperparameters, we suggest using  $\kappa_0 = 2 + K^{-2}$  and  $b_{0,ij} = \widehat{\tau}_{ij, \gamma=0}^2 (\kappa_0 - 1)$  for some large constant  $K > 0$ , which leads to  $\mathbb{E}^\pi(\tau_{ij}^2) = \widehat{\tau}_{ij, \gamma=0}^2$  and a prior coefficient of variation  $\{\text{Var}^\pi(\tau_{ij}^2)\}^{1/2} / \mathbb{E}^\pi(\tau_{ij}^2) = K$ . In the simulation studies,  $K = 100$  was used and the results are not sensitive to the choice of  $K$ . The hyperparameter  $\alpha$  was chosen as  $\alpha = 8.01(1 - 1/\log n)$ . If we assume a small  $\epsilon_0 > 0$ , the above choice of  $\alpha$  asymptotically satisfies  $\alpha > 8(1 + 2^{1/2}\epsilon_0)^2 / \{1 - 2^{3/2}\epsilon_0(1 + 2^{1/2}\epsilon_0)\}$ . We compare our one-sample covariance test with frequentist tests, proposed by Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014). The test suggested by Srivastava, Yanagihara and Kubokawa (2014) is based on estimating the squared Frobenius norm, and has a similar perspective to the test proposed by Cai and Ma (2013). Gupta and Bodnar (2014) proposed an exact one-sample covariance test based on fixed columns of the sample covariance matrix.

We first generated 100 data sets from the null hypothesis  $H_0 : \Sigma_n = I_p$  for various choices of  $n$  and  $p$ . We considered two structures for the alternative hypothesis  $H_1 : \Sigma_n \neq I_p$ . First, we chose  $\Sigma_0 = (\sigma_{0,ij})$  to have a compound symmetry structure

$$\sigma_{0,ij} = I(i = j) + \rho I(i \neq j) \quad (15)$$

for some signal strength constant  $\rho$  ranging from 0.025 to 0.15 by 0.025. In this

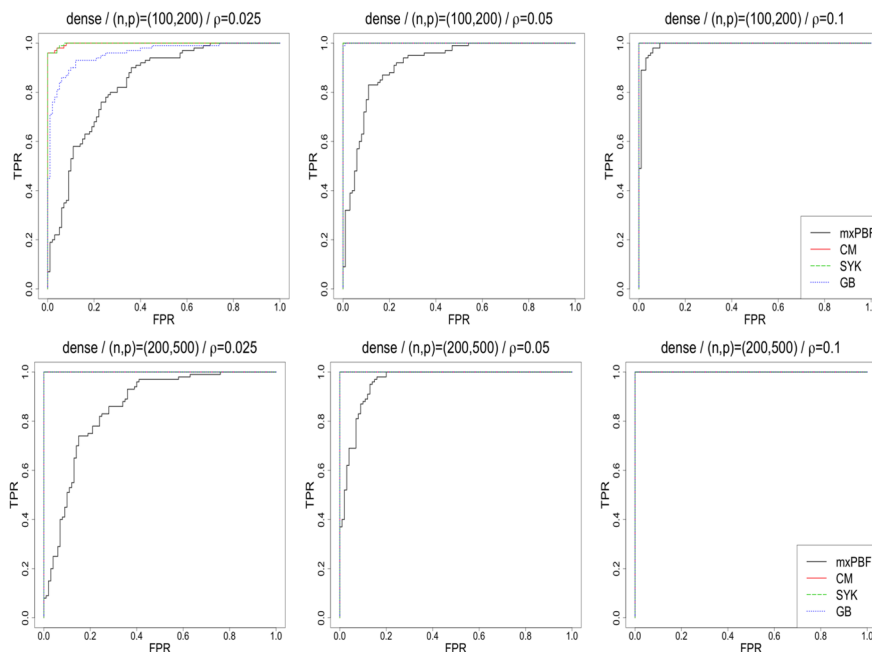


FIG 1. Receiver operating characteristic curves are represented for the three tests based on 100 simulated data sets for each hypothesis  $H_0 : \Sigma_n = I_p$  and  $H_1 : \Sigma_n \neq I_p$ , where (15) was used for  $H_1$ . TPR and FPR represent the true positive rate and the false positive rate, respectively. mxPBF, CM, SYK and GB represent the test proposed in this paper, Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014), respectively.

case, the difference between  $\Sigma_0$  and  $I_p$  is *dense*. As a second case for  $\Sigma_0$ , we let

$$\sigma_{0,ij} = I(i = j) + \rho I(i = 1, j = 2) + \rho I(i = 2, j = 1), \quad (16)$$

for some constant  $\rho$  ranging from 0.3 to 0.8 by 0.025. Because (16) has signals at only two locations, the difference between  $\Sigma_0$  and  $I_p$  is *sparse*. We generated 100 simulated data from  $N_p(0, \Sigma_0)$  for each setting. Note that (15) and (16) correspond to extremely sparse and dense settings, respectively. A more exhaustive simulation study can be found in Appendix B.

We calculated receiver operating characteristic (ROC) curves to illustrate and compare the performance of the tests. A ROC curve is drawn by plotting the true positive rate (TPR) against the false positive rate (FPR) while varying the thresholds. For each setting, points of the curves were obtained based on various thresholds and significance levels for  $B_{\max,10}(X)$  and the frequentist tests, respectively. We tried  $n = 100, 200, 300$  and  $p = 200, 500$  for each setting. Figure 1 shows results based on 100 simulated data from  $N_p(0, I_p)$  ( $H_0$ ) and 100 simulated data from  $N_p(0, \Sigma_0)$  with a compound symmetry structured  $\Sigma_0$  ( $H_1$ ) given in (15), for  $(n, p) = (100, 200)$  and  $(n, p) = (200, 500)$ . The false positive rate corresponds to the rate of  $H_0$ 's falsely detected as  $H_1$ 's. Similarly, the true

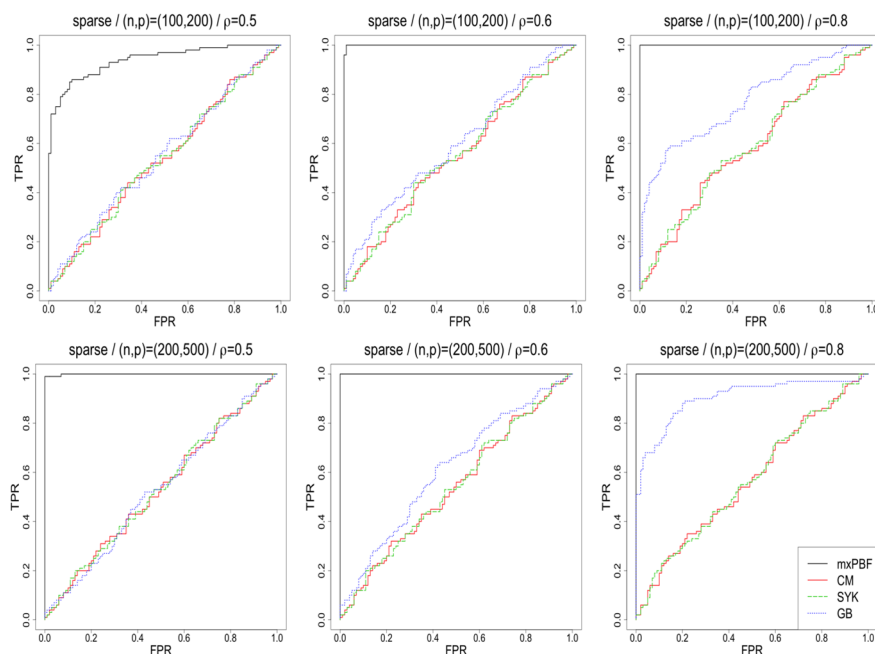


FIG 2. Receiver operating characteristic curves are represented for the three tests based on 100 simulated data sets for each hypothesis  $H_0 : \Sigma_n = I_p$  and  $H_1 : \Sigma_n \neq I_p$ , where (16) was used for  $H_1$ . TPR and FPR represent the true positive rate and the false positive rate, respectively. *mxPBF*, *CM*, *SYK* and *GB* represent the tests proposed in this paper, Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014), respectively.

positive rate is the rate of  $H_1$ 's correctly detected as  $H_1$ 's. In this setting, as expected, the tests in Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014) work better than the test proposed in this paper. However, as we can see from the second and third columns in Figure 1,  $B_{\max,10}(X)$  also performs well as the signal size  $\rho$  increases. The only case when our method is not as powerful is when weak signals are spread through the alternative covariance matrix, in which case one may question the meaningfulness of the signals.

Figure 2 shows results based on 100 simulated data from  $N_p(0, I_p)$  and 100 simulated data from  $N_p(0, \Sigma_0)$  with  $\Sigma_0$  given in (16), when  $(n, p) = (100, 200)$  and  $(n, p) = (200, 500)$ . As expected,  $B_{\max,10}(X)$  is much more powerful than the tests based on global norms when  $\Sigma_0 - I_p$  is sparse. Furthermore, the performances of the frequentist tests based on the Frobenius norm are almost the same for every setting, while  $B_{\max,10}(X)$  has better performance when  $(n, p) = (200, 500)$  than  $(n, p) = (100, 200)$ . Interestingly, the performance of the test in Gupta and Bodnar (2014) improves as the signal strength  $\rho$  increases. Thus, the test in Gupta and Bodnar (2014) is more sensitive to sparse changes than other tests based on the Frobenius norm difference. This makes

sense because it focuses on the changes in a column of the covariance matrix rather than in the whole covariance matrix.

#### 4.2. Simulation study: Testing diagonality

We conducted a simulation study to illustrate the performance of our proposed diagonality test. The hyperparameter  $\alpha$  was chosen as  $\alpha = 4.01(1 - 1/\log n)$ . We generated 100 data sets from the null  $H_0 : \sigma_{ij} = 0$  for any  $i \neq j$  using  $\Sigma_0 = I_p$ . The two structures of  $\Sigma_0$  under  $H_1$  used in the previous section, (15) and (16), were considered. For each setting, 100 data sets were generated.

We compare our test with some existing frequentist tests. Cai and Jiang (2011) proposed a diagonality test based on the maximum of sample correlations. Here  $\hat{\tau}_{ij,\gamma}^2$  in the pairwise Bayes factor  $\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)$  is a decreasing function of the sample correlation between  $\tilde{X}_i$  and  $\tilde{X}_j$ . Lan et al. (2015) developed a test in the regression setting based on the squared Frobenius norm of a sample covariance matrix. Their test should be powerful against dense alternatives. We also conducted maximum-type tests based on Kendall's tau and Spearman's rho (Han, Chen and Liu, 2017). Chen and Liu (2018) assumed  $p$ -dimensional observations from a common multivariate normal distribution and investigated the dependence *among samples*. Since their method can be applied to the diagonality test by considering  $X^T$  instead of  $X$ , we included it as a contender. Their test requires  $p = O(n)$  and the uniformly bounded condition for the eigenvalues of  $\Sigma_0$  for theoretical properties, excluding the high-dimensional setting  $p \gg n$  and some interesting covariance classes like compound symmetry. Finally, we also considered frequentist union-intersection tests based on the p-values associated with the marginal independence tests. A  $t$ -test for Pearson's correlation was conducted for testing  $H_{0,ij} : \sigma_{ij} = 0$  for each pair  $i > j$ , and the null hypothesis  $H_0 : \sigma_{ij} = 0$  for any  $i \neq j$  was rejected if at least one  $H_{0,ij}$  was rejected. To calculate the p-values, we used the `cor0.test` function in the `GeneNet` package.

Figure 3 shows the area under the ROC curve for varying signal strength  $\rho$  for each fixed  $(n, p)$ . We omit the results of Cai and Jiang (2011), which were almost identical to our test in every setting. As expected, the test of Lan et al. (2015) is more powerful against dense alternatives. The other tests, except the test of Chen and Liu (2018), have less power, but work reasonably well as the signal  $\rho$  grows. The test of Chen and Liu (2018) does not work well, likely because (15) violates their assumptions. When  $\Sigma_0 - I_p$  is sparse, the test of Lan et al. (2015) does not work well even when  $\rho$  is large. The other tests show good results against sparse alternatives, but our test has better performance.

#### 4.3. Support recovery using gene expression data

To describe the practical performance of the support recovery procedure (14),  $\hat{S}_{pair}$ , we analyzed a data set from a small round blue-cell tumor microarray experiment (Khan et al., 2001). The data set originally had 6,567 gene expression values, and 2,308 gene expressions were selected by an initial filtering (Khan

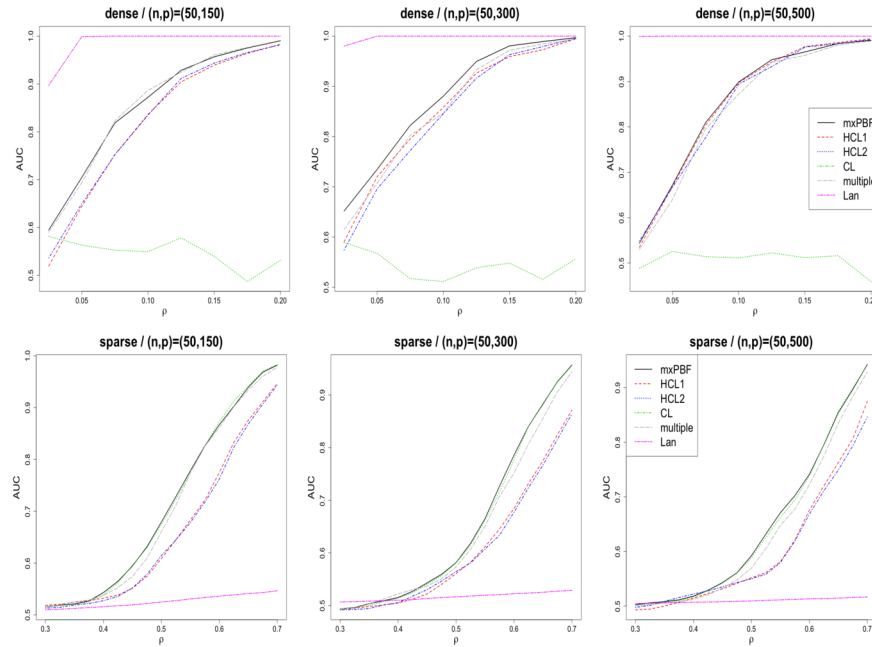


FIG 3. Area under the curves (AUC) are represented for the tests based on 100 simulated data sets for each hypothesis  $H_0 : \sigma_{ij} = 0$  for any  $i \neq j$  and  $H_1 : \text{not } H_0$ . “dense” and “sparse” mean that the true covariance matrix  $\Sigma_0$  under  $H_1$  were generated from (15) and (16), respectively. mxPBF, CL and Lan represent the tests proposed in this paper, Chen and Liu (2018) and Lan et al. (2015), respectively. HCL1 and HCL2 represent the test based on Kendall’s tau and Spearman’s rho, respectively. “multiple” means the frequentist union-intersection test.

et al., 2001). For comparison purposes, we focus on the preprocessed data used in Rothman, Levina and Zhu (2009) and Cai and Liu (2011), consisting of  $p = 200$  gene expression values for each of  $n = 64$  training tissue samples. There are four types of tumors represented in these tissue samples. Data were centered prior to analysis.

For pairwise Bayes factors, the hyperparameter was set at  $\alpha = 4.01(1 - 1/\log n)$ . We used cross-validation to select  $C_{sel}$ . Let  $n$  be the number of observations for a given data set. We randomly divided the data 50 times into two subsamples with size  $n_1 = \lceil n/3 \rceil$  and  $n_2 = n - n_1$  as a test set and training set, respectively. Denote  $I_1$  and  $I_2$  as indices for the test set and training set, respectively, thus,  $|I_1| = n_1$ ,  $|I_2| = n_2$  and  $I_1 \cup I_2 = \{1, \dots, n\}$ . Let  $\hat{S}_j(C_{sel})$  be the estimated support for the  $j$ th column of the covariance matrix via pairwise Bayes factors, based on  $\{X_i\}_{i \in I_2}$  and a given threshold  $C_{sel}$ . We calculated the averaged mean squared error

$$MSE(C_{sel}) = \sum_{j=1}^p \sum_{l \in \hat{S}_j(C_{sel})} \left\{ \sum_{i \in I_1} (X_{ij} - X_{il} \hat{\beta}_{jl})^2 / (n_1 - 1) \right\} / |\hat{S}_j(C_{sel})|,$$

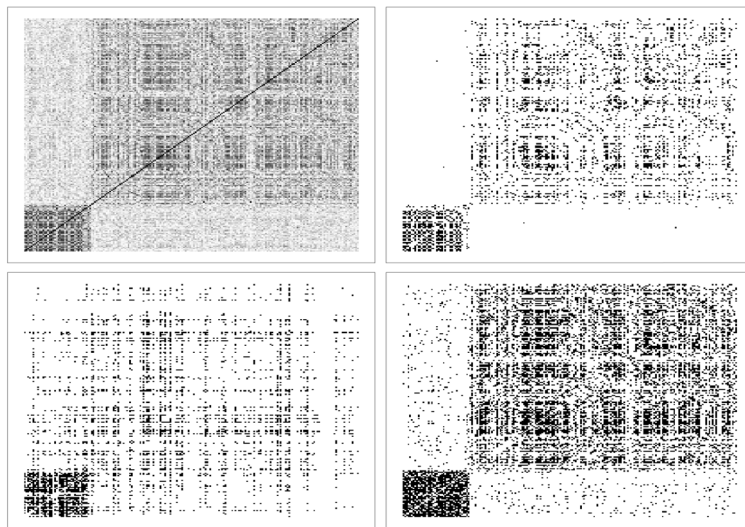


FIG 4. The absolute sample correlation matrix (top left) and estimated supports from various methods. Clockwise from the top right are plots for the estimated supports based on  $\hat{S}_{pair, \hat{C}_{sel}}$ ,  $\hat{\Sigma}_{\delta}^*$  and  $\hat{\Sigma}_{\delta}$ , respectively.

where  $\hat{\beta}_{jl}$  is a least square estimate with respect to the dependent variable  $\{X_{ij}\}_{i \in I_1}$  and covariate  $\{X_{il}\}_{i \in I_1}$ . The threshold  $C_{sel}$  was varied from  $-7$  to  $10$  with increment  $0.2$ , and  $\hat{C}_{sel}$  was set to minimize  $\sum_{\nu=1}^{50} MSE_{\nu}(C_{sel})/50$ , where  $MSE_{\nu}(C_{sel})$  is the averaged mean squared error based on the  $\nu$ th split.

We compared our method with generalized thresholding estimators of Rothman, Levina and Zhu (2009) and Cai and Liu (2011). Rothman, Levina and Zhu (2009) used a universal threshold  $\lambda = \delta(\log p/n)^{1/2}$ , while Cai and Liu (2011) used an individual threshold  $\hat{\lambda}_{ij} = \delta(\hat{\theta}_{ij} \log p/n)^{1/2}$  with a data-dependent  $\hat{\theta}_{ij}$ . We denote thresholding estimators proposed by Rothman, Levina and Zhu (2009) and Cai and Liu (2011) by  $\hat{\Sigma}_{\delta}$  and  $\hat{\Sigma}_{\delta}^*$ , respectively. We used the adaptive lasso thresholding rule,  $s_{\lambda}(\sigma) = \sigma \max(1 - |\lambda/\sigma|^{\eta}, 0)$  with  $\eta = 4$ , because it gave good support recovery results in simulation studies in Rothman, Levina and Zhu (2009) and Cai and Liu (2011). We adopted the cross-validation method described in Section 4 of Cai and Liu (2011) to select  $\delta$  and denote the selected tuning parameter by  $\hat{\delta}$ .

Figure 4 shows the support recovery results and the absolute sample correlation matrix. The estimated supports based on  $\hat{S}_{pair, \hat{C}_{sel}}$ ,  $\hat{\Sigma}_{\delta}^*$  and  $\hat{\Sigma}_{\delta}$  are represented. One can see that  $\hat{S}_{pair, \hat{C}_{sel}}$  and  $\hat{\Sigma}_{\delta}^*$  show the clustering structure between informative (top 40) and non-informative (bottom 160) genes, while the structure is somewhat blurred in  $\hat{\Sigma}_{\delta}$ . To compare  $\hat{S}_{pair, \hat{C}_{sel}}$  and  $\hat{\Sigma}_{\delta}^*$  in more detail, we further focused on the top 40 genes. We applied hierarchical clustering to the genes based on the complete linkage method using R function `hclust`, and the



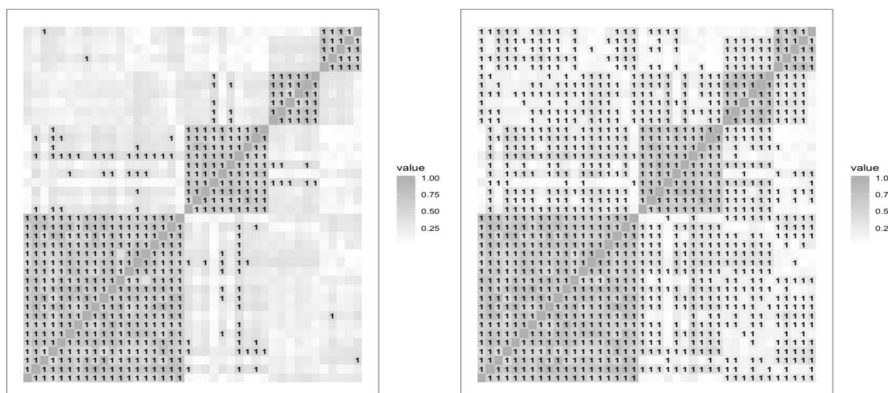


FIG 5. The ordered absolute sample correlation matrix and estimated supports for top 40 genes, with 1's representing the estimated supports from  $\hat{S}_{pair, \hat{C}_{sel}}$  (left) and  $\hat{\Sigma}_{\delta}^*$  (right).

genes were ordered according to the clustering result. Figure 5 shows the ordered absolute sample correlation matrix and estimated supports for the top 40 genes. The clustering result suggests that there are four clusters, consistent with the four tumor types. Both support recovery procedures detect significant blocks in the sample correlation matrix. However, our support recovery procedure shows the clustering structure much clearer, while  $\hat{\Sigma}_{\delta}^*$  gives a blurred structure due to a dense support estimate. The estimated support based on pairwise Bayes factors has the advantage of producing a sparser, and hence potentially more interpretable, estimate of support.

## 5. Discussion

We have focused on covariance matrix structure testing in this paper, but the maximum pairwise Bayes factor idea can be easily applied to other related settings. For example, testing differences across groups in high-dimensional mean vectors is an interesting possibility. When the two mean vectors are almost the same but differ only at a few locations, a maximum pairwise Bayes factor approach should have relatively high power. Similarly, it can be applied to the high-dimensional two-sample covariance test. Two covariances from two populations may differ only in a small number of entries.

There are some possible generalizations of the pairwise Bayes factor idea. To accelerate the speed of computation, a random subsampling method can be used instead of calculating the pairwise Bayes factor for every single pair  $(i, j)$ . It should be interesting to develop a suitable random subsampling or random projection scheme achieving desirable theoretical properties. Especially when  $p$  is huge, it will effectively reduce the computational complexity. The maximum pairwise Bayes factor approach is also trivially parallelizable. Another possibility is considering alternative combining approaches to the max in merging the

information from every pairwise Bayes factor. If there are many weak non-zero covariances, then the average or summation may be preferable to the maximum. A suitable modification to learn parameters in the combining operator can potentially make the test powerful to a broad class of alternative hypotheses.

**Appendix A: Testing pairwise independence**

As a by-product of Theorem 3.3, when pairwise independence testing (12) itself is of interest, we suggest a pairwise Bayes factor  $\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)$ , which can be shown to be consistent. For consistency under the alternative hypothesis, we assume

$$\sigma_{0,ij}^2 \geq \frac{C_4 \sigma_{0,jj}}{1 - 2\epsilon_0 C_1^{1/2}} \left\{ \frac{9C_1 \tau_{0,ij}^2}{(1 - C_3)^2} \vee \frac{\alpha(1 + \gamma)(1 + 4\epsilon_0 C_1^{1/2})\sigma_{0,ii}}{C_3} \right\} \frac{\log n}{n} \tag{17}$$

for constants  $C_1 > 0, 0 < C_3 < 1$  and  $C_4 > 1$  defined in Section 3.1. If we substitute  $\log n$  in the above condition with  $\log(n \vee p)$ , it coincides with condition (A4). The proof of Corollary 1 follows from that of Theorem 3.3, and thus it is omitted.

**Corollary 1.** *Consider model (1) and a hypothesis testing problem (12) for a given pair  $(i, j)$  such that  $i \neq j$ . Suppose we use the prior  $\pi(\tau_{ij}^2) \propto \tau_{ij}^{-2}$  under both  $H_{0,ij}$  and  $H_{1,ij}$ , and the prior  $\pi(a_{ij} | \tau_{ij}^2)$  defined in (5) under  $H_{1,ij}$  with  $\gamma = n^{-\alpha}$  for some positive constant  $\alpha$ . Then under  $H_{0,ij} : \sigma_{ij} = 0$ , for some constant  $c > 0$ ,*

$$\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = O_p(n^{-c}).$$

*If, under  $H_{1,ij} : \sigma_{ij} \neq 0$ , at least one of  $\sigma_{0,ij}$  and  $\sigma_{0,ji}$  satisfies (17), for some constant  $c' > 0$ ,*

$$\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)^{-1} = O_p(n^{-c'}).$$

**Appendix B: Additional simulation study for one-sample covariance test**

In this section, we conduct an additional simulation study for one-sample covariance test. To see a phase-transition from a regime where the maximum-based test (i.e., maximum pairwise Bayes factor) is preferable to a regime where global norm-based tests are preferable. As global norm-based tests, we choose the tests of Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014).

We fix  $(n, p) = (100, 200)$  and generate 100 data sets from the null  $H_0 : \Sigma_n = I_p$ . To generate the true covariance matrix  $\Sigma_0 = (\sigma_{0,ij})$  for the alternative  $H_1 : \Sigma_n \neq I_p$ , (1) we randomly select  $s_0$  locations, say  $S_0$ , among the lower triangular parts of  $\Sigma_0$  and (2) let  $\sigma_{0,ij} = \rho I((i, j) \in S_0)$ ,  $\sigma_{0,ji} = \sigma_{0,ij}$  for all  $1 \leq i < j \leq p$

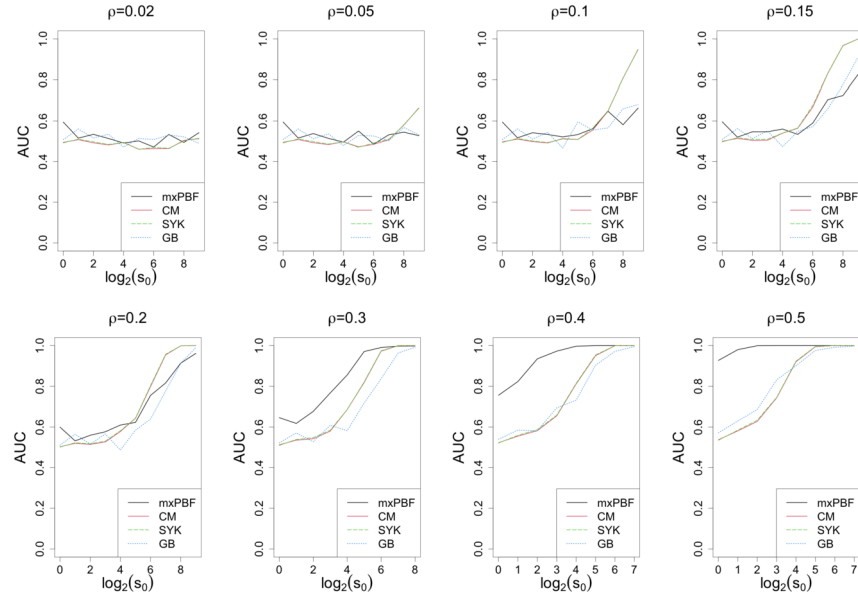


FIG 6. Area under the curves (AUC) are represented for the tests based on 100 simulated data sets for each hypothesis  $H_0 : \Sigma_n = I_p$  and  $H_1 : \Sigma_n \neq I_p$ . mxPBF, CM, SYK and GB represent the tests proposed in this paper, Cai and Ma (2013), Srivastava, Yanagihara and Kubokawa (2014) and Gupta and Bodnar (2014), respectively.

and  $\sigma_{0,ii} = 1$  for all  $1 \leq i \leq p$ . Here,  $\rho$  and  $s_0$  correspond to the magnitude and the number of signals, respectively. We generate 100 data sets for each  $\rho \in \{0.02, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5\}$  and  $s_0 \in \{1, 2, 2^2, \dots, 2^9\}$ . For  $\rho = 0.3$  and  $\rho \in \{0.4, 0.5\}$ , we use  $s_0 \in \{1, 2, 2^2, \dots, 2^8\}$  and  $s_0 \in \{1, 2, 2^2, \dots, 2^7\}$ , respectively, to make the resulting  $\Sigma_0$  positive definite.

Figure 6 shows the (averaged) area under the curves (AUC) for varying the number of signals  $s_0$  for each fixed  $\rho$ . For  $\rho \in \{0.02, 0.05\}$ , all the tests have AUC values around 0.5, which means that the signals are too small so that the performances of the tests are similar to that of random guess. For  $\rho \in \{0.1, 0.15\}$ , the Frobenius norm-based tests (Cai and Ma (2013) and Srivastava, Yanagihara and Kubokawa (2014)) outperform the maximum pairwise Bayes factor when the number of signals  $s_0$  larger than  $2^6$ . However, the performance of the maximum pairwise Bayes factor tends to better than the Frobenius norm-based tests when  $s_0$  is smaller than  $2^4$ . Furthermore, the performance of the maximum pairwise Bayes factor is also improved as the strength of signals grows. For  $\rho = 0.2$ , the maximum pairwise Bayes factor works slightly better than the Frobenius norm-based tests when  $s_0 \leq 2^4$ , while the reverse is true when  $s_0 \geq 2^6$ . For  $\rho \in \{0.3, 0.4, 0.5\}$ , the maximum pairwise Bayes factor outperforms the global norm-based tests regardless of the number of signals, which implies that  $\rho \geq 0.3$  is large enough to be detected as a *meaningful signal* by the maximum-based test.

Based on the simulation results, we see that the maximum pairwise Bayes factor is preferable for large  $\rho$  and small  $s_0$ . On the other hand, the Frobenius norm-based tests are preferable for small  $\rho$  and large  $s_0$ . Thus, there is a phase-transition from a regime where the maximum pairwise Bayes factor is preferable to a regime where global norm-based tests are preferable. A rigorous theoretical investigation for this phase-transition would be an interesting topic for future research.

**Appendix C: Simulation study for support recovery**

The proposed support recovery procedure,  $\widehat{S}_{pair, C_{sel}}$ , consistently recovers the true support  $S(\Sigma_0)$  for any constant threshold  $C_{sel}$  (Theorem 3.4). However, in practice, the choice of the threshold is crucial. In this section, we investigate the quality of  $\widehat{S}_{pair, C_{sel}}$  as a function of the threshold  $C_{sel}$  to assess the importance of the choice of  $C_{sel}$ . Furthermore, the performance of the cross-validation-based threshold  $\widehat{C}_{sel}$  proposed in Section 4.3 also will be demonstrated.

Two structures of covariance matrices were investigated. In the first setting, we consider

$$\sigma_{0,ij}^* = 2 \max \left( 1 - \frac{|i-j|}{10}, 0 \right) I \left\{ |i-j| \leq 5, (i \vee j) \leq \frac{p}{2} \right\}$$

for  $i \neq j$ . In the second setting, we consider

$$\begin{aligned} \sigma_{0,ij}^* &= 2 \max \left( 1 - \frac{|i-j|}{10}, 0 \right) I \left\{ |i-j| \leq 5, i \leq \frac{p}{2} \right\} \\ &\quad + 2 \max \left( 1 - \frac{|i-j|}{20}, 0 \right) I \left\{ |i-j| \leq 10, i > \frac{p}{2} \right\} \end{aligned}$$

for  $i < j$ . In both settings, we set  $\sigma_{0,ii}^* = 1$  and  $\sigma_{0,ij}^* = \sigma_{0,ji}^*$ , where  $\Sigma_0^* = (\sigma_{0,ij}^*)$ . Let  $\Sigma_0 = (\sigma_{0,ij})$  be the true covariance matrix. If  $\Sigma_0^*$  is positive definite,  $\Sigma_0 = \Sigma_0^*$  is used, and if  $\Sigma_0^*$  is not positive definite,  $\Sigma_0 = (\sigma_{0,ij})$  is used where  $\sigma_{0,ij} = \sigma_{0,ij}^*$  and  $\Sigma_{0,ii} = \Sigma_{0,ii}^* - \lambda_{\min}(\Sigma_0) + 0.01$  for all  $1 \leq i \neq j \leq p$ . The data were generated from  $N_p(0, \Sigma_0)$  with  $n = 50, 100$  and  $p = 100, 300$ . To illustrate the performance of the estimated support  $\widehat{S}_{pair, C_{sel}}$ , we consider (i) the Matthews correlation coefficient,

$$MCC = \frac{TP \times TN - FP \times FN}{\{(TP + FP)(TP + FN)(TN + FP)(TN + FN)\}^{1/2}},$$

and (ii) the number of errors,  $FP + FN$ , where  $TP, TN, FP$  and  $FN$  are true positive, true negative, false positive and false negative, respectively.

Figures 7 and 8 show the performance of  $\widehat{S}_{pair, C_{sel}}$  for the first setting and the second setting, respectively. Each plot represents the performance of  $\widehat{S}_{pair, C_{sel}}$  as a function of the threshold  $C_{sel}$ , where the red vertical line is the cross-validation-based threshold  $\widehat{C}_{sel}$ . We found that the quality of a fixed threshold changes with the sample size  $n$ , the dimension  $p$  and the structure of  $\Sigma_0$ . The

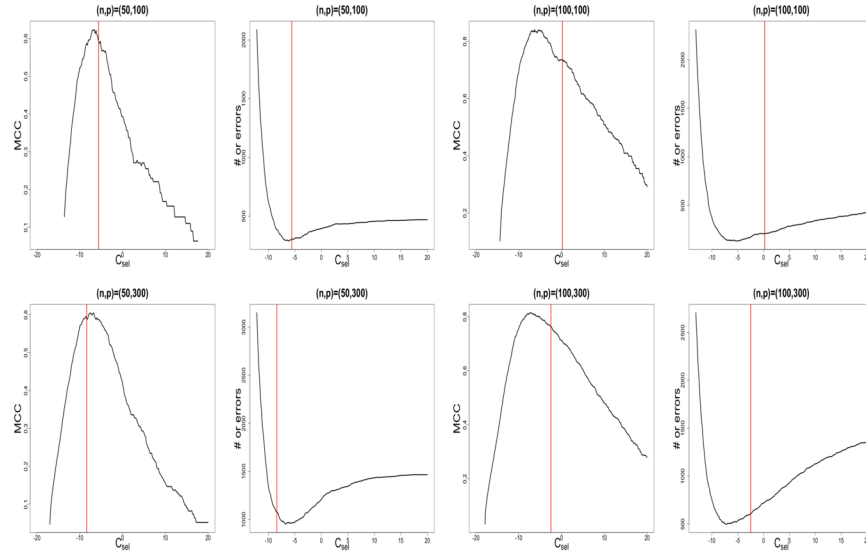


FIG 7. MCC (Mathews correlation coefficient) and the number of errors for the first setting. The red vertical line is the cross-validation-based threshold  $C_{sel}$ .

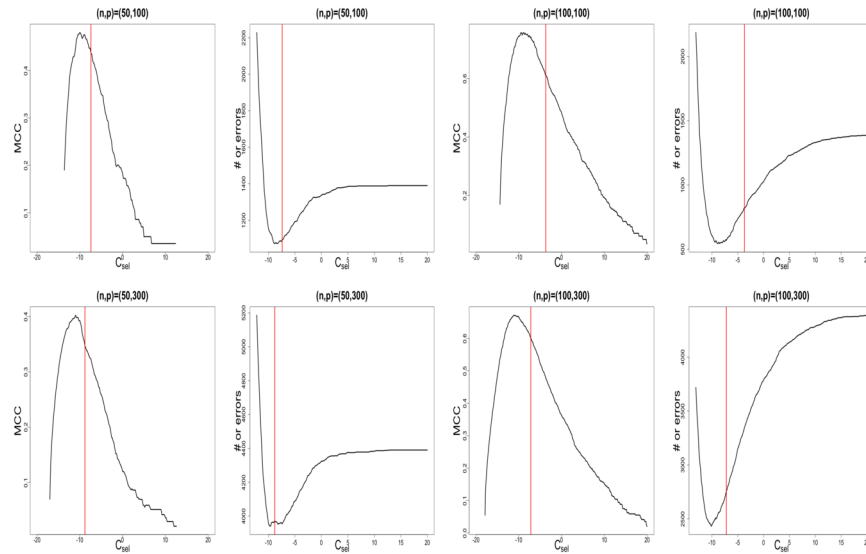


FIG 8. MCC (Mathews correlation coefficient) and the number of errors for the second setting. The red vertical line is the cross-validation-based threshold  $C_{sel}$ .

estimated  $\hat{C}_{sel}$  has reasonable performance in terms of the two criteria. Therefore, based on the simulation results, the cross-validation-based threshold  $\hat{C}_{sel}$  seems adequate for adaptive selection of the threshold. Figure 9 represents the

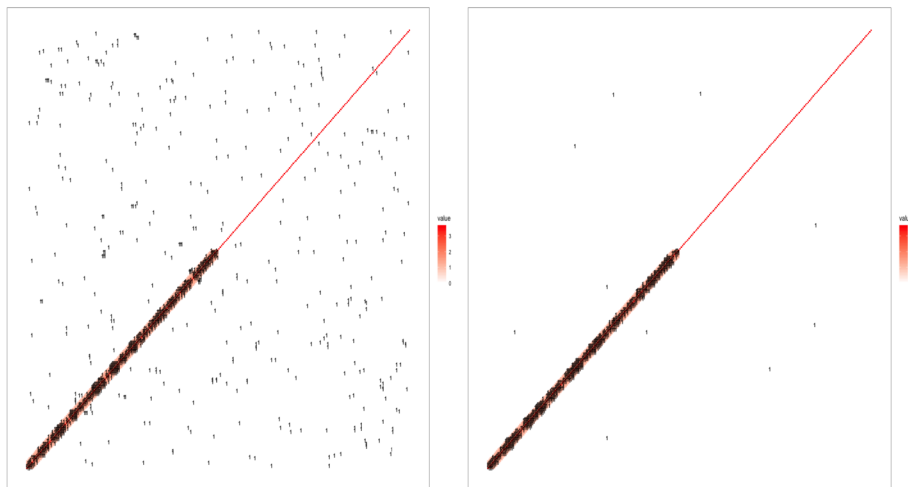


FIG 9. The character 1 indicates the estimated supports based on  $\hat{C}_{sel}$  for the first setting with  $p = 300$  when  $n = 50$  (left) and  $n = 100$  (right). The values in red represent the entries of the true covariance matrix.

estimated support  $\hat{S}_{pair, \hat{C}_{sel}}$  for the first setting with  $p = 300$ . It shows that the cross-validation-based threshold  $\hat{C}_{sel}$  has reasonable performance and the quality of the support recovery increases as the sample size gets larger.

### Appendix D: Proof of Theorem 3.1

We first give more details about

$$\begin{aligned} B_{10}(\tilde{X}_i, \tilde{X}_j) &= \frac{p(\tilde{X}_i | \tilde{X}_j, H_{1,ij})}{p(\tilde{X}_i | \tilde{X}_j, H_{0,ij})} \\ &= \frac{b_{0,ij}^{\kappa_0}}{\Gamma(\kappa_0)} \left(\frac{\gamma}{1+\gamma}\right)^{1/2} \Gamma\left(\frac{n}{2} + \kappa_0\right) e^{n\hat{\tau}_i^2/2} \left(\frac{n}{2}\hat{\tau}_{ij,\gamma}^2 + b_{0,ij}\right)^{-n/2-\kappa_0}. \end{aligned}$$

Note that

$$\begin{aligned} p(\tilde{X}_i | \tilde{X}_j, H_{0,ij}) &= f(\tilde{X}_i | \tilde{X}_j, a_{ij} = 0, \tau_{ij}^2 = 1) \\ &= (2\pi)^{-n/2} e^{-\|\tilde{X}_i\|_2^2/2} = (2\pi)^{-n/2} e^{-n\hat{\tau}_i^2/2} \end{aligned}$$

and

$$\begin{aligned} &p(\tilde{X}_i | \tilde{X}_j, H_{1,ij}) \\ &= \int \int f(\tilde{X}_i | \tilde{X}_j, a_{ij}, \tau_{ij}^2) \pi(a_{ij} | \tau_{ij}^2) \pi(\tau_{ij}^2) da_{ij} d\tau_{ij}^2 \\ &= \int \int (2\pi\tau_{ij}^2)^{-n/2} \exp\left(-\frac{\|\tilde{X}_i - a_{ij}\tilde{X}_j\|_2^2}{2\tau_{ij}^2}\right) \left(\frac{2\pi\tau_{ij}^2}{\gamma\|\tilde{X}_j\|_2^2}\right)^{-1/2} \end{aligned}$$

$$\begin{aligned} & \times \exp\left(-\frac{\gamma\|\tilde{X}_j\|_2^2}{2\tau_{ij}^2}a_{ij}\right)\frac{b_{0,ij}^{\kappa_0}}{\Gamma(\kappa_0)}(\tau_{ij}^2)^{-\kappa_0-1}e^{-b_{0,ij}/\tau_{ij}^2}da_{ij}d\tau_{ij}^2 \\ & = (2\pi)^{-n/2}\frac{b_{0,ij}^{\kappa_0}}{\Gamma(\kappa_0)}\Gamma\left(\frac{n}{2}+\kappa_0\right)\left(\frac{n}{2}\hat{\tau}_{ij,\gamma}^2+b_{0,ij}\right)^{-n/2-\kappa_0}\left(\frac{\gamma}{1+\gamma}\right)^{1/2}. \end{aligned}$$

*Proof of Theorem 3.1.* For a given pair  $(i, j)$ ,

$$\begin{aligned} & \log B_{10}(\tilde{X}_i, \tilde{X}_j) \\ & = \kappa_0 \log b_{0,ij} + \frac{1}{2} \log\left(\frac{\gamma}{1+\gamma}\right) \end{aligned} \tag{18}$$

$$+ \left\{ \log \Gamma\left(\frac{n}{2} + \kappa_0\right) - \log \Gamma(\kappa_0) - \left(\frac{n}{2} + \kappa_0\right) \log \frac{n}{2} + \frac{n}{2} \right\} \tag{19}$$

$$+ \left\{ \frac{n}{2}\hat{\tau}_i^2 - \left(\frac{n}{2} + \kappa_0\right) \log\left(\hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n}\right) - \frac{n}{2} \right\}. \tag{20}$$

Equation (19) can be written as

$$\log \Gamma\left(\frac{n}{2} + \kappa_0\right) - \log \Gamma(1 + \kappa_0) + \kappa_0 - \left(\frac{n}{2} + \kappa_0\right) \log \frac{n}{2} + \frac{n}{2},$$

where by Theorem 1 of Keckic and Vasic (1971),

$$\begin{aligned} & \left(\frac{n}{2} + \kappa_0 - 1\right) \log\left(\frac{n}{2} + \kappa_0\right) - \kappa_0 \log(1 + \kappa_0) - \frac{n}{2} + 1 \\ & \leq \log \Gamma\left(\frac{n}{2} + \kappa_0\right) - \log \Gamma(1 + \kappa_0) \\ & \leq \left(\frac{n}{2} + \kappa_0 - \frac{1}{2}\right) \log\left(\frac{n}{2} + \kappa_0\right) - \left(\frac{1}{2} + \kappa_0\right) \log(1 + \kappa_0) - \frac{n}{2} + 1. \end{aligned}$$

Thus, it is easy to see that (19) is equal to

$$-C \log\left(\frac{n}{2} + \kappa_0\right) + \left(\kappa_0 + \frac{2\kappa_0^2}{n}\right) \log\left(1 + \frac{2\kappa_0}{n}\right)^{n/2\kappa_0} + C' \tag{21}$$

for some constants  $1/2 < C < 1$  and  $\kappa_0\{1 - \log(1 + \kappa_0)\} + 1 - 0.5 \log(1 + \kappa_0) < C' < \kappa_0\{1 - \log(1 + \kappa_0)\} + 1$ , which is of order  $O(-\log n)$ . Since  $\gamma = (n \vee p)^{-\alpha}$ , (18) is equal to

$$\begin{aligned} & \kappa_0 \log b_{0,ij} - \frac{\alpha}{2} \log(n \vee p) - \frac{1}{2} \log\{1 + (n \vee p)^{-\alpha}\} \\ & = -\frac{\alpha}{2} \log(n \vee p) + C \end{aligned} \tag{22}$$

for some constant  $C$ . Thus, we only need to focus on the behavior of (20).

For the true covariance matrix  $\Sigma_0$ , one of the following cases holds:

- Case (i).**  $\Sigma_0 = I_p$ ,
- Case (ii).**  $\sigma_{0,ii}$  satisfies condition (A1) for some pair  $(i, j)$ ,
- Case (iii).**  $\tau_{0,ij}^2$  satisfies condition (A2) for some pair  $(i, j)$ ,
- Case (iv).**  $\sigma_{0,ij}$  satisfies condition (A3) for some pair  $(i, j)$ .

Note that (20) can be expressed as

$$\frac{n}{2} \left\{ \hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n} - \log \left( \hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n} \right) - 1 \right\} \tag{23}$$

$$+ \frac{n}{2} (\hat{\tau}_i^2 - \hat{\tau}_{ij,\gamma}^2) \tag{24}$$

$$- \left\{ \kappa_0 \log \left( \hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n} \right) + b_{0,ij} \right\}. \tag{25}$$

We will calculate the rate of the above three terms (23)-(25) for every possible case. More precisely, we will show that for all sufficiently large  $n$ ,  $\log B_{10}(\tilde{X}_i, \tilde{X}_j) \leq -C_{H_0} \log(n \vee p)$  with probability at least  $1 - (n \vee p)^{-c}$  for some constants  $C_{H_0} > 0$  and  $c > 2$  under Case (i), and  $\log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq C_{H_1} \log(n \vee p)$  with probability at least  $1 - (n \vee p)^{-c'}$  for some constants  $C_{H_1} > 0$  and  $c' > 0$  under Cases (ii)-(iv). Then, we have

$$\begin{aligned} & \mathbb{P}_0 \left\{ \log B_{\max,10}(X) \leq -C \log(n \vee p) \right\} \\ &= 1 - \mathbb{P}_0 \left\{ \max_{(i,j):i \neq j} \log B_{10}(\tilde{X}_i, \tilde{X}_j) > -C \log(n \vee p) \right\} \\ &\geq 1 - \sum_{(i,j):i \neq j} \mathbb{P}_0 \left\{ \log B_{10}(\tilde{X}_i, \tilde{X}_j) > -C \log(n \vee p) \right\} \\ &\geq 1 - (n \vee p)^{-c+2} \end{aligned}$$

if Case (i) holds, and

$$\begin{aligned} & \mathbb{P}_0 \left\{ \log B_{\max,10}(X) \geq C' \log(n \vee p) \right\} \\ &\geq \max_{(i,j):i \neq j} \mathbb{P}_0 \left\{ \log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq C' \log(n \vee p) \right\} \\ &\geq 1 - (n \vee p)^{-c'} \end{aligned}$$

if there exists at least one pair  $(i, j)$  satisfying one of Cases (ii)-(iv).

**Case (i).** Define  $n\hat{\tau}_{ij}^2 = n\hat{\tau}_{ij,\gamma=0}^2 = \tilde{X}_i^T(I_n - H_j)\tilde{X}_i$ , then we have  $n\hat{\tau}_{ij,\gamma}^2 = n\hat{\tau}_{ij}^2 + \gamma(1 + \gamma)^{-1}(\tilde{X}_j^T \tilde{X}_i)^2 \|\tilde{X}_j\|_2^{-2}$ . Note that  $(\tilde{X}_j^T \tilde{X}_i)^2 \leq \|\tilde{X}_j\|_2^2 \|\tilde{X}_i\|_2^2$  and  $\|\tilde{X}_i\|_2^2 \sim \chi_n^2$  under Case (i). By Lemma 1 in Laurent and Massart (2000), we have  $P\{k^{-1}\chi_k^2 - 1 \geq 2(k^{-1}x)^{1/2} + 2k^{-1}x\} \leq 2 \exp(-x)$  and  $P\{1 - k^{-1}\chi_k^2 \geq 2(k^{-1}x)^{1/2}\} \leq 2 \exp(-x)$  for all  $x > 0$ , which implies

$$\hat{\tau}_{ij}^2 \leq \hat{\tau}_{ij,\gamma}^2 \leq \hat{\tau}_{ij}^2 + \frac{\gamma}{1 + \gamma} \left[ 1 + 2 \left\{ \frac{C \log(n \vee p)}{n} \right\}^{1/2} + \frac{2C \log(n \vee p)}{n} \right]$$

with probability at least  $1 - 2(n \vee p)^{-C}$  for some constant  $C > 2$ . Since  $\gamma = (n \vee p)^{-\alpha}$  under the conditions on  $\alpha$ , we have

$$|\hat{\tau}_{ij,\gamma}^2 - \hat{\tau}_{ij}^2| \leq (n \vee p)^{-2} \tag{26}$$



with probability at least  $1 - 2(n \vee p)^{-C}$  for some constant  $C > 2$  and all large  $n$ . It is easy to check that  $n\hat{\tau}_{ij}^2 \sim \chi_{n-1}^2$  and  $n(\hat{\tau}_i^2 - \hat{\tau}_{ij}^2) \sim \chi_1^2$  under Case (i). Thus, we have

$$\mathbb{P}_0 \left\{ |n(\hat{\tau}_i^2 - \hat{\tau}_{ij}^2) - 1| \leq 2C'c \log(n \vee p) \right\} \geq 1 - 2(n \vee p)^{-c}, \tag{27}$$

$$\mathbb{P}_0 \left[ \left| \frac{n}{n-1} \hat{\tau}_{ij}^2 - 1 \right| \leq 2C' \left\{ \frac{c \log(n \vee p)}{n-1} \right\}^{1/2} \right] \geq 1 - 2(n \vee p)^{-c} \tag{28}$$

for any constants  $c > 2$ ,  $C' > 1 + 2^{1/2}\epsilon_0$  and all large  $n$ . The inequalities (26) and (28) imply

$$\mathbb{P}_0 \left[ \left| \hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n} - 1 \right| < C \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2} \right] \geq 1 - 4(n \vee p)^{-c} \tag{29}$$

for any constant  $C > 2(2^{1/2} + 2\epsilon_0)$  and all large  $n$ . By the Taylor expansion of  $\log(1+x)$ ,

$$\begin{aligned} x - \log(1+x) &= x - \left( x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots \right) \\ &= \frac{1}{2}x^2 - \frac{1}{3}x^3 + \frac{1}{4}x^4 - \dots \leq \frac{1}{2}x^2(1 - |x|)^{-1} \end{aligned}$$

for small  $|x|$ . Thus, on the event in (29), (23) is bounded above by  $4^{-1}C^2(1 - C\epsilon_0)^{-1} \log(n \vee p)$  for any constant  $2(2^{1/2} + 2\epsilon_0) < C < \epsilon_0^{-1}$ . Since (25) is of order  $O(1)$  on the event in (29), we have

$$\begin{aligned} &\log B_{10}(\tilde{X}_i, \tilde{X}_j) \\ &\leq -2^{-1}\alpha \log(n \vee p) - 2^{-1} \log n + 2^{-1}C^2(1 - C\epsilon_0)^{-1} \log(n \vee p) \\ &\leq -2^{-1}\{\alpha - C^2(1 - C\epsilon_0)^{-1}\} \log(n \vee p) \end{aligned}$$

with probability at least  $1 - 4(n \vee p)^{-c}$  for any constants  $2(\sqrt{2} + 2\epsilon_0) < C < \epsilon_0^{-1}$  and all large  $n$ . Thus, if  $\alpha > 8(1 + \sqrt{2}\epsilon_0)^2 / \{1 - 2\sqrt{2}\epsilon_0(1 + \sqrt{2}\epsilon_0)\}$ ,

$$\log B_{10}(\tilde{X}_i, \tilde{X}_j) \leq -2^{-1}C \log(n \vee p)$$

with probability at least  $1 - 4(n \vee p)^{-c}$  for any constants  $0 < C < \alpha - 8(1 + \sqrt{2}\epsilon_0)^2 / \{1 - 2^{3/2}\epsilon_0(1 + \sqrt{2}\epsilon_0)\}$ ,  $c > 2$  and all large  $n$ . This completes the proof for Case (i) with the constant  $C_{H_0} = [\alpha - 8(1 + \sqrt{2}\epsilon_0)^2 / \{1 - 2^{3/2}\epsilon_0(1 + \sqrt{2}\epsilon_0)\}] / 2$ .

**Case (ii).** Now assume that  $\sigma_{0,ii}$  satisfies condition (A1) for some pair  $(i, j)$ . Note that (20) can be expressed as

$$\frac{n}{2} \left\{ \hat{\tau}_i^2 + \frac{2b_{0,ij}}{n} - \log \left( \hat{\tau}_i^2 + \frac{2b_{0,ij}}{n} \right) - 1 \right\} \tag{30}$$

$$+ \left( \frac{n}{2} + \kappa_0 \right) \log \left( \frac{\hat{\tau}_i^2 + \frac{2b_{0,ij}}{n}}{\hat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n}} \right) \tag{31}$$

$$- \kappa_0 \log \left( \hat{\tau}_i^2 + \frac{2b_{0,ij}}{n} \right) - b_{0,ij}. \tag{32}$$

We will show that for given constants  $C_1 > 0$  and  $C_2 > 2(\alpha + 2)^{1/2}$ ,

$$\mathbb{P}_0 \left\{ \left| \widehat{\tau}_i^2 + \frac{2b_{0,ij}}{n} - 1 \right| \geq C_2 \left( \frac{\log(n \vee p)}{n} \right)^{1/2} \right\} \geq 1 - 2(n \vee p)^{-C_1}. \quad (33)$$

On this event, we can show that (30) is larger than  $4^{-1}C_2^2(1 - 2C_2\epsilon_0/3) \log(n \vee p)$  for all large  $n$  by the Taylor expansion of  $\log(1+x)$  and the fact that  $x - \log x - 1$  is increasing in  $|x - 1|$ . Note that (31) is positive and (32) is negligible compared to (30). Then, by (21) and (22),

$$\begin{aligned} \log B_{10}(\tilde{X}_i, \tilde{X}_j) &> -\frac{\alpha}{2} \log(n \vee p) - \log n + 4^{-1}C_2^2(1 - 2C_2\epsilon_0/3) \log(n \vee p) + C \\ &\geq 8^{-1}(C_2^2 - 4\alpha - 8) \log(n \vee p) + C \end{aligned}$$

with probability at least  $1 - 4(n \vee p)^{-C_1}$  for some constant  $C$  and all large  $n$ , by the condition on  $\epsilon_0$ .

Now, we only need to show (33). By Lemma 1 in Laurent and Massart (2000), one can show that

$$\begin{aligned} \mathbb{P}_0 \left[ \sigma_{0,ii} - 1 + \frac{2b_{0,ij}}{n} - 2\sigma_{0,ii} \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \leq \widehat{\tau}_i^2 + \frac{2b_{0,ij}}{n} - 1 \right. \\ \left. \leq \sigma_{0,ii} - 1 + \frac{2b_{0,ij}}{n} + 4\sigma_{0,ii} \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \\ \geq 1 - (n \vee p)^{-2C_1}, \end{aligned}$$

because  $n\widehat{\tau}_i^2/\sigma_{0,ii} \sim \chi_n^2$ . Thus, it suffices to prove

$$\sigma_{0,ii} - 1 + \frac{2b_{0,ij}}{n} - 2\sigma_{0,ii} \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \geq C_2 \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2}$$

or

$$\sigma_{0,ii} - 1 + \frac{2b_{0,ij}}{n} + 4\sigma_{0,ii} \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \leq -C_2 \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2},$$

which is satisfied by (A1).

**Case (iii).** If  $\sigma_{0,ii}$  satisfies (A1) for some pair  $(i, j)$ , the previous case gives the desired result. Here we assume that  $\sigma_{0,ii}$  does not satisfy (A1) for all  $i$ , and  $\tau_{0,ij}^2$  satisfies condition (A2) for some pair  $(i, j)$ . Similar to Case (ii), we will show that for given constants  $C_1 > 0$  and  $C_2 > 2(\alpha + 2)^{1/2}$ ,

$$\mathbb{P}_0 \left[ \left| \widehat{\tau}_{ij,\gamma}^2 + \frac{2b_{0,ij}}{n} - 1 \right| \geq C_2 \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2} \right] \geq 1 - 4(n \vee p)^{-C_1}, \quad (34)$$

which gives the desired result by (23)–(25). Note that we have  $n\widehat{\tau}_{ij,\gamma}^2/\tau_{0,ij}^2 \sim \chi_{n-1}^2$ . Then, similar to (28),

$$\mathbb{P}_0 \left[ \left| \widehat{\tau}_{ij,\gamma}^2 - \frac{n-1}{n} \tau_{0,ij}^2 \right| \leq 4\tau_{0,ij}^2 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \geq 1 - 4(n \vee p)^{-C_1}$$

by (26). To prove (34), we only need to show that

$$\left(1 - \frac{1}{n}\right)\tau_{0,ij}^2 + \frac{2b_{0,ij}}{n} - 1 - 4\tau_{0,ij}^2 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \geq C_2 \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2},$$

when  $\tau_{0,ij}^2 > 1$ , and

$$\left(1 - \frac{1}{n}\right)\tau_{0,ij}^2 + \frac{2b_{0,ij}}{n} - 1 + 4\tau_{0,ij}^2 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \leq -C_2 \left\{ \frac{\log(n \vee p)}{n} \right\}^{1/2}$$

when  $\tau_{0,ij}^2 < 1$ . It is satisfied because we have condition (A1). Thus, we have proved that if a pair  $(i, j)$  satisfies (A1),

$$\log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq 8^{-1}(C_2^2 - 4\alpha - 8) \log(n \vee p) + C$$

with probability at least  $1 - 4(n \vee p)^{-C_1}$  for some constant  $C$  and all sufficiently large  $n$ .

**Case (iv).** Suppose  $\sigma_{0,ij}$  satisfies condition (A3). In this case, we have  $n(\hat{\tau}_i^2 - \hat{\tau}_{ij}^2)/\tau_{0,ij}^2 \sim \chi_1^2(\lambda_{ij})$  given  $\tilde{X}_j$ , where  $\lambda_{ij} = \|\tilde{X}_j\|_2^2 a_{0,ij}^2 / \tau_{0,ij}^2$ . For a random variable  $X \sim \chi_k^2(\lambda)$ ,

$$\mathbb{P} \left[ X \geq k + \lambda - 2\{(k + 2\lambda)x\}^{1/2} \right] \geq 1 - e^{-x}$$

for all  $x > 0$ , by Lemma 8 in Kolar and Liu (2012). Then,

$$\begin{aligned} & 1 - (n \vee p)^{-C_1} \\ & \leq \mathbb{P}_0 \left( \frac{n}{2} (\hat{\tau}_i^2 - \hat{\tau}_{ij}^2) \geq \frac{\tau_{0,ij}^2}{2} \left[ 1 + \lambda_{ij} - 2\{(1 + 2\lambda_{ij})C_1 \log(n \vee p)\}^{1/2} \right] \mid \tilde{X}_j \right) \\ & \leq \mathbb{P}_0 \left( \frac{n}{2} (\hat{\tau}_i^2 - \hat{\tau}_{ij}^2) \geq \frac{\tau_{0,ij}^2}{2} \lambda_{ij} \left[ 1 - 2\left\{ \frac{(1 + 2\lambda_{ij})C_1 \log(n \vee p)}{\lambda_{ij}^2} \right\}^{1/2} \right] \mid \tilde{X}_j \right). \end{aligned}$$

Also  $\tau_{0,ij}^2 \lambda_{ij} / (a_{0,ij}^2 \sigma_{0,jj}) = \|\tilde{X}_j\|_2^2 / \sigma_{0,jj} \sim \chi_n^2$ , so we have

$$\mathbb{P}_0 \left( \frac{\tau_{0,ij}^2}{a_{0,ij}^2 \sigma_{0,jj}} \lambda_{ij} \geq n \left[ 1 - 2\left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \right) \geq 1 - (n \vee p)^{-C_1} \tag{35}$$

by Lemma 1 in Laurent and Massart (2000). On the event in (35), we have

$$\begin{aligned} \lambda_{ij} & \geq \frac{a_{0,ij}^2 \sigma_{0,jj}}{\tau_{0,ij}^2} n \left[ 1 - 2\left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \\ & \geq \frac{\sigma_{0,ij}^2}{\sigma_{0,jj} \tau_{0,ij}^2} n \left[ 1 - 2\left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \\ & \geq \frac{9C_1}{(1 - C_3)^2} \log(n \vee p) \end{aligned}$$

with probability at least  $1 - (n \vee p)^{-C_1}$  by condition (A3), which implies

$$1 - 2 \left\{ \frac{(1 + 2\lambda_{ij})C_1 \log(n \vee p)}{\lambda_{ij}^2} \right\}^{1/2} \geq C_3 \tag{36}$$

for a given constant  $0 < C_3 < 1$ . Again by (35) and condition (A3),

$$\begin{aligned} & \frac{\tau_{0,ij}^2}{2} \lambda_{ij} \left[ 1 - 2 \left\{ \frac{(1 + 2\lambda_{ij})C_1 \log(n \vee p)}{\lambda_{ij}^2} \right\}^{1/2} \right] \\ & \geq \frac{\tau_{0,ij}^2}{2} \lambda_{ij} C_3 \\ & \geq \frac{1}{2} \frac{\sigma_{0,ij}^2}{\sigma_{0,jj}} C_3 n \left[ 1 - 2 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \\ & \geq \frac{1}{2} C_4 (\alpha + 2) \log(n \vee p) \end{aligned}$$

with probability at least  $1 - (n \vee p)^{-C_1}$  for a given constant  $C_4 > 1$  and all large  $n$ . Note that (23) is positive and (25) is negligible compared to (23). Thus, by similar arguments used in Case (ii),

$$\begin{aligned} & \log B_{10}(\tilde{X}_i, \tilde{X}_j) \\ & \geq \frac{1}{2} \frac{\sigma_{0,ij}^2}{\sigma_{0,jj}} C_3 n \left[ 1 - 2 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] - \frac{1}{2} (\alpha - 2) \log(n \vee p) \\ & \geq \frac{1}{2} (C_4 - 1) (\alpha - 2) \log(n \vee p) + C \end{aligned}$$

with probability at least  $1 - 2(n \vee p)^{-C'}$  for some constants  $C, C' > 0$  and all large  $n$ . This completes the proof for cases (ii)–(iv) with the constant  $C_{H_1} = \{8^{-1}(C_2^2 - 4\alpha - 8)\} \vee \{2^{-1}(C_4 - 1)(\alpha - 2) \log(n \vee p)\}$ , where  $C_2 > 2(\alpha + 2)^{1/2}$  and  $C_4 > 1$ .  $\square$

### Appendix E: Proof of Theorem 3.2

*Proof.* For a given constant  $C_\star > 0$ , define a parameter class

$$H_1^*(C_\star) = \left\{ \Sigma_\nu : \Sigma_\nu = I_p + \left\{ C_\star \left( \frac{\log p}{n} \right)^{1/2} I(i = j = \nu) \right\}_{1 \leq i, j \leq p}, 1 \leq \nu \leq p \right\},$$

which trivially satisfies  $H_1^*(C_\star) \subset H_1(C_\star)$ . Let  $\mathbb{P}_{mix} = p^{-1} \sum_\nu \mathbb{P}_{\Sigma_\nu}$  and  $\mathbb{E}_{mix}$  be the corresponding expectation under  $\mathbb{P}_{mix}$ . For any  $\Sigma_\nu \in H_1^*(C_\star)$  and test  $\phi$ ,

$$\begin{aligned} \sup_\nu \{ \mathbb{E}_{I_p}(\phi) + \mathbb{E}_{\Sigma_\nu}(1 - \phi) \} & \geq \inf_\phi \sup_\nu \{ \mathbb{E}_{I_p}(\phi) + \mathbb{E}_{\Sigma_\nu}(1 - \phi) \} \\ & \geq \inf_\phi \frac{1}{p} \sum_\nu \{ \mathbb{E}_{I_p}(\phi) + \mathbb{E}_{\Sigma_\nu}(1 - \phi) \} \end{aligned}$$

$$\begin{aligned}
 &= \inf_{\phi} \{ \mathbb{E}_{I_p}(\phi) + \mathbb{E}_{mix}(1 - \phi) \} \\
 &= \int (f_{I_p} \wedge f_{mix}) \\
 &= 1 - \frac{1}{2} \int |f_{mix} - f_{I_p}|,
 \end{aligned}$$

where  $f_{mix}$  and  $f_{\Sigma}$  are density functions of  $\mathbb{P}_{mix}$  and  $\mathbb{P}_{\Sigma}$ , respectively. Also

$$\begin{aligned}
 \left( \int |f_{mix} - f_{I_p}| \right)^2 &\leq \int |f_{mix} - f_{I_p}|^2 \\
 &\leq \int \left| \frac{f_{mix}}{f_{I_p}} - 1 \right|^2 f_{I_p} \\
 &\leq \int \frac{f_{mix}^2}{f_{I_p}} - 1. \tag{37}
 \end{aligned}$$

Thus, for any test  $\phi$ ,

$$\inf_{\Sigma \in H_1(C_*)} \mathbb{E}_{\Sigma}(\phi) \leq \inf_{\Sigma \in H_1^*(C_*)} \mathbb{E}_{\Sigma}(\phi) \leq \mathbb{E}_{I_p}(\phi) + \frac{1}{2} \left( \int \frac{f_{mix}^2}{f_{I_p}} - 1 \right)^{1/2}$$

by the above arguments. Now we only need to deal with the upper bound of (37). An upper bound of  $\int (f_{mix}^2/f_{I_p})$  can be derived as follows:

$$\begin{aligned}
 \int \frac{f_{mix}^2}{f_{I_p}} &= \int \frac{1}{f_{I_p}} \left( \frac{1}{p} \sum_{\nu} f_{\Sigma_{\nu}} \right)^2 \\
 &= \frac{1}{p^2} \sum_{\nu_1, \nu_2} \int \frac{f_{\Sigma_{\nu_1}} f_{\Sigma_{\nu_2}}}{f_{I_p}} \\
 &= \frac{1}{p^2} \sum_{\nu_1, \nu_2} \left[ \det \left\{ I_p - (\Sigma_{\nu_1} - I_p)(\Sigma_{\nu_2} - I_p) \right\} \right]^{-n/2} \\
 &= \frac{1}{p^2} \left\{ p^2 - p + p \left( 1 - C_*^2 \frac{\log p}{n} \right)^{-n/2} \right\} \\
 &\leq 1 - \frac{1}{p} + \frac{1}{p} C_*^2/2 \leq 2 - \frac{1}{p}
 \end{aligned}$$

because  $C_*^2 \leq 2$ . The third equality follows from Lemma B.3 of Lee and Lee (2018). It gives the upper bound

$$\inf_{\Sigma \in H_1(C_*)} \mathbb{E}_{\Sigma}(\phi) \leq \mathbb{E}_{I_p}(\phi) + \frac{1}{2} (2 - p^{-1})^{1/2},$$

which completes the proof. □

**Appendix F: Proof of Theorem 3.3**

*Proof.* For a given pair  $(i, j)$  such that  $i \neq j$ , suppose the null hypothesis is true, so that  $\sigma_{0,ij} = 0$ . We have

$$\begin{aligned} \log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) &= \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) - \frac{n}{2} \log \left( \frac{\hat{\tau}_{ij,\gamma}^2}{\hat{\tau}_i^2} \right) \\ &\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{n}{2} \frac{\hat{\tau}_i^2 - \hat{\tau}_{ij,\gamma}^2}{\hat{\tau}_{ij,\gamma}^2} \\ &\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{n}{2} \frac{\hat{\tau}_i^2 - \hat{\tau}_{ij}^2}{\hat{\tau}_{ij}^2}, \end{aligned}$$

where the first inequality holds because  $\log(1 + x) \leq x$  for all  $x$ . Note that  $n\hat{\tau}_{ij}^2/\sigma_{0,ii} \sim \chi_{n-1}^2$  and  $n(\hat{\tau}_i^2 - \hat{\tau}_{ij}^2)/\sigma_{0,ii} \sim \chi_1^2$ . By Lemma 1 in Laurent and Massart (2000),

$$\frac{n(\hat{\tau}_i^2 - \hat{\tau}_{ij}^2)}{\sigma_{0,ii}} \leq 1 + 2C C_4 \log(n \vee p)$$

and

$$\frac{n\hat{\tau}_{ij}^2}{\sigma_{0,ii}} - (n - 1) \geq 2\{C(n - 1) \log(n \vee p)\}^{1/2}$$

with probability at least  $1 - 6(n \vee p)^{-C}$  for any constant  $C > 2$  and  $C_4 > 1$ . This implies

$$\begin{aligned} \log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) &\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{n}{2} \frac{1 + 2C C_4 \log(n \vee p)}{n - 1 - 2\{C(n - 1) \log(n \vee p)\}^{1/2}} \\ &\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{1}{2} \frac{n + 2C C_4 n \log(n \vee p)}{n - 1 - 2\{C(n - 1) \log(n \vee p)\}^{1/2}} \\ &\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{1}{2} C' \log(n \vee p) + C'' \end{aligned}$$

for any constants  $C' > 2C C_4/(1 - 3\epsilon_0 C^{1/2})$ ,  $C'' > 0$  with probability at least  $1 - 6(n \vee p)^{-C}$  for some constant  $C > 2$  and large  $n$ . Since  $\gamma = (n \vee p)^{-\alpha}$  and  $\alpha > 4/(1 - 2^{1/2} 3\epsilon_0)$ , by choosing  $C > 2$  and  $C_4 > 1$  such that  $\alpha > C'$ , the log Bayes factor  $\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)$  tends to minus infinity as  $n \rightarrow \infty$  on the above event. By similar arguments used in the proof of Theorem 3.1, this implies consistency of the maximum pairwise Bayes factor under  $H_0$ .

Now suppose the alternative hypothesis is true. Without loss of generality, assume that  $\sigma_{0,ij}$  satisfies condition (A4). Then,

$$-\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = -\frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) - \frac{n}{2} \log \left( 1 - \frac{\hat{\tau}_{ij,\gamma}^2 - \hat{\tau}_i^2}{\hat{\tau}_{ij,\gamma}^2} \right)$$

$$\leq -\frac{1}{2} \log\left(\frac{\gamma}{1+\gamma}\right) + \frac{n}{2} \frac{\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_i^2}{\widehat{\tau}_i^2}$$

because  $-\log(1-x) \leq x/(1-x)$  for any  $x < 1$ . Since  $n\widehat{\tau}_i^2/\sigma_{0,ii} \sim \chi_n^2$  and  $n(\widehat{\tau}_i^2 - \widehat{\tau}_{ij}^2)/\tau_{0,ij}^2 \sim \chi_1^2(\lambda_{ij})$  given  $\tilde{X}_j$ , where  $\lambda_{ij} = \|\tilde{X}_j\|_2^2 a_{0,ij}^2/\tau_{0,ij}^2$ , we have

$$\begin{aligned} \frac{1}{2\widehat{\tau}_i^2} &\geq \frac{1}{2} \left( \sigma_{0,ii} \left[ 1 + 4 \left\{ \frac{C_1 \log(n \vee p)}{n} \right\}^{1/2} \right] \right)^{-1} \\ &\geq \frac{1}{2} \left\{ \sigma_{0,ii} (1 + 4\epsilon_0 C_1^{1/2}) \right\}^{-1} \end{aligned}$$

and

$$\begin{aligned} n(\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_i^2) &= -\frac{1}{1+\gamma} n(\widehat{\tau}_i^2 - \widehat{\tau}_{ij}^2) \\ &\leq -\frac{C_3}{1+\gamma} \tau_{0,ij}^2 \lambda_{ij} \\ &= -\frac{C_3}{1+\gamma} \sigma_{0,jj} a_{0,ij}^2 \frac{\|\tilde{X}_j\|_2^2}{\sigma_{0,jj}} \\ &\leq -\frac{C_3}{1+\gamma} \frac{\sigma_{0,ij}^2}{\sigma_{0,jj}} n(1 - 2C_1^{1/2}\epsilon_0) \end{aligned}$$

with probability at least  $1 - C(n \vee p)^{-c}$  for some constants  $c > 0$  and  $C > 0$  and all large  $n$ , by condition (15). The fourth inequality holds by condition (A4) and similar arguments used in (36). Thus, on this event,

$$\begin{aligned} \frac{n}{2} \frac{\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_i^2}{\widehat{\tau}_i^2} &\leq -\frac{C_3}{1+\gamma} \frac{\sigma_{0,ij}^2}{\sigma_{0,jj}} n(1 - 2\epsilon_0 C_1^{1/2}) \times \frac{1}{2} \left\{ \sigma_{0,ii} (1 + 4\epsilon_0 C_1^{1/2}) \right\}^{-1} \\ &\leq -\frac{\alpha}{2} C_4 \log(n \vee p) \end{aligned}$$

which implies

$$-\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) \leq \frac{\alpha}{2} \log(n \vee p) + \frac{1}{2} \log(1+\gamma) - \frac{\alpha}{2} C_4 \log(n \vee p)$$

for all sufficiently large  $n$ . Since  $C_4 > 1$ , this completes the proof.  $\square$

### Appendix G: Proof of Proposition 1

*Proof.* Let  $\widehat{\sigma}_{ij}^2 = n^{-1}(\tilde{X}_i^T \tilde{X}_j)^2$  and  $\widehat{\rho}_{ij}^2 = \widehat{\sigma}_{ij}^2/(\widehat{\tau}_i^2 \widehat{\tau}_j^2)$ . Then

$$\begin{aligned} \frac{\widehat{\tau}_{ij,\gamma}^2}{\widehat{\tau}_i^2} &= \frac{\widehat{\tau}_{ij}^2}{\widehat{\tau}_i^2} + \frac{\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_{ij}^2}{\widehat{\tau}_i^2} \\ &= 1 - \frac{\widehat{\sigma}_{ij}^2}{\widehat{\tau}_i^2 \widehat{\tau}_j^2} + \frac{\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_{ij}^2}{\widehat{\tau}_i^2} \end{aligned}$$

$$\equiv 1 - \hat{\rho}_{ij}^2 + \hat{\delta}_{ij}.$$

Then we have

$$\log \tilde{B}_{\max,10}(X) = \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) - \frac{n}{2} \log \left\{ 1 - \max_{i \neq j} (\hat{\rho}_{ij}^2 - \hat{\delta}_{ij}) \right\}.$$

We will show that upper and lower bounds of  $2 \log \tilde{B}_{\max,10}(X) - C_{n,p}$  converges in distribution to an extreme distribution of type I with distribution function

$$F(z) = \exp \{ - (8\pi)^{-1/2} e^{-z/2} \}, \quad z \in \mathbb{R}.$$

By the definitions of  $\hat{\tau}_{ij}^2$  and  $\hat{\tau}_{ij,\gamma}^2$ , we have  $\hat{\delta}_{ij} \geq 0$ , thus

$$\begin{aligned} & 2 \log \tilde{B}_{\max,10}(X) - C_{n,p} \\ & \leq \log \left( \frac{\gamma}{1 + \gamma} \right) - n \log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right) - C_{n,p} \\ & = \frac{\log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right)}{- \max_{i \neq j} \hat{\rho}_{ij}^2} n \max_{i \neq j} \hat{\rho}_{ij}^2 - 4 \log p + \log(\log p) \\ & = \frac{\log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right)}{- \max_{i \neq j} \hat{\rho}_{ij}^2} \{ n \max_{i \neq j} \hat{\rho}_{ij}^2 - 4 \log p + \log(\log p) \} \end{aligned} \tag{38}$$

$$+ \left\{ \frac{\log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right)}{- \max_{i \neq j} \hat{\rho}_{ij}^2} - 1 \right\} \{ 4 \log p - \log(\log p) \}. \tag{39}$$

By Theorem 4 in Cai and Jiang (2011),  $\log p \max_{i \neq j} \hat{\rho}_{ij}^2 \xrightarrow{p} 0$  as  $n \rightarrow \infty$  under  $\mathbb{P}_0$ . Taylor expansion of  $\log(1 - x)$  gives

$$\begin{aligned} 0 \leq \frac{\log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right)}{- \max_{i \neq j} \hat{\rho}_{ij}^2} - 1 &= \frac{1}{2} \max_{i \neq j} \hat{\rho}_{ij}^2 + \frac{1}{3} \max_{i \neq j} \hat{\rho}_{ij}^4 + \dots \\ &\leq \frac{\max_{i \neq j} \hat{\rho}_{ij}^2}{1 - \max_{i \neq j} \hat{\rho}_{ij}^2}, \end{aligned}$$

which implies that (39) converges to 0 in probability and

$$\frac{\log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 \right)}{- \max_{i \neq j} \hat{\rho}_{ij}^2} \xrightarrow{p} 1$$

as  $n \rightarrow \infty$ . Since  $n \max_{i \neq j} \hat{\rho}_{ij}^2 - 4 \log p + \log(\log p)$  converges in distribution to an extreme distribution of type I by Theorem 4 in Cai and Jiang (2011), Slutsky's theorem says that (38) converges in distribution to an extreme distribution of type I.

On the other hand, we have

$$\begin{aligned} & 2 \log \tilde{B}_{\max,10}(X) - C_{n,p} \\ & \geq \log \left( \frac{\gamma}{1 + \gamma} \right) - n \log \left( 1 - \max_{i \neq j} \hat{\rho}_{ij}^2 + \max_{i \neq j} \hat{\delta}_{ij} \right) - C_{n,p} \end{aligned}$$



$$= \frac{\log(1 - \max_{i \neq j} \widehat{\rho}_{ij}^2 + \max_{i \neq j} \widehat{\delta}_{ij})}{-\max_{i \neq j} \widehat{\rho}_{ij}^2 + \max_{i \neq j} \widehat{\delta}_{ij}} n(\max_{i \neq j} \widehat{\rho}_{ij}^2 - \max_{i \neq j} \widehat{\delta}_{ij}) - 4 \log p + \log(\log p)$$

Thus, if we can show that  $n \max_{i \neq j} \widehat{\delta}_{ij} \xrightarrow{p} 0$  as  $n \rightarrow \infty$  under  $\mathbb{P}_0$ , with the same arguments used in the previous paragraph it implies that the lower bound of  $2 \log \tilde{B}_{\max,10}(X) - C_{n,p}$  converges in distribution to an extreme distribution of type I, which completes the proof. By the proof of Theorem 3.1, for a given pair  $(i, j)$ ,

$$\begin{aligned} \mathbb{P}_0 \left[ 1 - 2 \left\{ \frac{C \log(n \vee p)}{n} \right\}^{1/2} \leq \frac{\widehat{\tau}_i^2}{\sigma_{0,ii}} \leq 1 + 4 \left\{ \frac{C \log(n \vee p)}{n} \right\}^{1/2} \right] \\ \geq 1 - (n \vee p)^{-2C} \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}_0 \left( 0 < \frac{\widehat{\tau}_{ij,\gamma}^2 - \widehat{\tau}_{ij}^2}{\sigma_{0,ii}} \leq \frac{\gamma}{1 + \gamma} \left[ 1 + 2 \left\{ \frac{C \log(n \vee p)}{n} \right\}^{1/2} + \frac{2C \log(n \vee p)}{n} \right] \right) \\ \geq 1 - 2(n \vee p)^{-C} \end{aligned}$$

for some constant  $C > 2$ . Since  $\gamma = (n \vee p)^{-\alpha}$  and  $\alpha > 1$ , it implies that  $n \max_{i \neq j} \widehat{\delta}_{ij} \xrightarrow{p} 0$  as  $n \rightarrow \infty$  under  $\mathbb{P}_0$ . □

**Appendix H: Proof of Theorem 3.4**

*Proof.* For a given constant  $C_{sel} > 0$  and a pair  $(i, j)$ , define

$$\widehat{S}_{ij} = I \left\{ \log \tilde{B}_{pair,10}(\tilde{X}_i, \tilde{X}_j) > C_{sel} \right\}$$

and  $S_{ij}(\Sigma_0) = I(\sigma_{0,ij} \neq 0)$ . By the proof of Theorem 3,

$$\begin{aligned} \mathbb{P}_0 \left\{ \widehat{S}_{ij} = 1, S_{ij}(\Sigma_0) = 0 \right\} &= \mathbb{P}_0 \left\{ \log \tilde{B}_{pair,10}(\tilde{X}_i, \tilde{X}_j) > C_{sel}, \sigma_{0,ij} = 0 \right\} \\ &\leq (n \vee p)^{-c} \end{aligned}$$

for some constant  $c > 2$  and all sufficiently large  $n$ . Now, assume that  $\sigma_{0,ij}$  satisfies condition (A5). Condition (A5) is the same as condition (A4) except using  $C_5$  instead of  $C_1$ . Thus, by similar arguments used in the proof of Theorem 3.3, it is easy to check that

$$\begin{aligned} \mathbb{P}_0 \left\{ \widehat{S}_{ij} = 0, S_{ij}(\Sigma_0) = 1 \right\} &= \mathbb{P}_0 \left\{ \log \tilde{B}_{pair,10}(\tilde{X}_i, \tilde{X}_j) \leq C_{sel}, \sigma_{0,ij} \neq 0 \right\} \\ &\leq (n \vee p)^{-c} \end{aligned}$$

for some constant  $c > 2$  and all sufficiently large  $n$ . Therefore, we have

$$\mathbb{P}_0 \left\{ \widehat{S} \neq S(\Sigma_0) \right\} \leq \sum_{i \neq j, i < j} \mathbb{P}_0 \left\{ \widehat{S}_{ij} \neq S_{ij}(\Sigma_0) \right\} \rightarrow 0$$

as  $n \rightarrow \infty$ . □

## Acknowledgments

We are very grateful to the Associate Editor and the reviewer for their valuable comments which have led to improvement in our paper.

## References

- CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. [MR2850210](#)
- CAI, T. T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Statist. Assoc.* **106** 672–684. [MR2847949](#)
- CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. [MR3160557](#)
- CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. [MR3375874](#)
- CHEN, X. and LIU, W. (2018). Testing independence with high-dimensional correlated samples. *The Annals of Statistics* **46** 866–894. [MR3782387](#)
- CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Am. Statist. Assoc.* **105** 810–819. [MR2724863](#)
- COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* 204–218. [MR1243593](#)
- DRTON, M. and PERLMAN, M. D. (2004). Model selection for Gaussian concentration graphs. *Biometrika* **91** 591–602. [MR2090624](#)
- DRTON, M. and PERLMAN, M. D. (2007). Multiple testing and error control in Gaussian graphical model selection. *Statist. Sci.* 430–449. [MR2416818](#)
- ENIKEEVA, F. and HARCHAOU, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics* **47** 2051–2079. [MR3953444](#)
- FERNANDEZ, C., LEY, E. and STEEL, M. F. (2001). Benchmark priors for Bayesian model averaging. *J. Econometrics* **100** 381–427. [MR1820410](#)
- GAN, L., NARISSETTY, N. N. and LIANG, F. (2018). Bayesian regularization for graphical models with unequal shrinkage. *J. Am. Statist. Assoc.* 1–14. [MR4011774](#)
- GUPTA, A. K. and BODNAR, T. (2014). An exact test about the covariance matrix. *Journal of Multivariate Analysis* **125** 176–189. [MR3163837](#)
- HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. [MR3737306](#)
- JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. [MR3034330](#)
- JIANG, B., YE, C. and LIU, J. S. (2017). Bayesian nonparametric tests via sliced inverse modeling. *Bayesian Anal.* **12** 89–112. [MR3597568](#)
- JOHNSON, V. E. and ROSSELL, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *J. R. Stat. Soc. Ser. B Stat. Methodol* **72** 143–170. [MR2830762](#)

- JOHNSON, V. E. and ROSSELL, D. (2012). Bayesian model selection in high-dimensional settings. *J. Am. Statist. Assoc.* **107** 649–660. [MR2980074](#)
- JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. [MR2751448](#)
- KECKIC, J. D. and VASIC, P. M. (1971). Some inequalities for the gamma function. *Publications de l'Institut Mathématique* **11** 107–114. [MR0308446](#)
- KHAN, J., WEI, J. S., RINGNER, M., SAAL, L. H., LADANYI, M., WESTERMANN, F., BERTHOLD, F., SCHWAB, M., ANTONESCU, C. R. and PETERSON, C. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **7** 673.
- KOLAR, M. and LIU, H. (2012). Marginal regression for multitask learning. In *Artificial Intelligence and Statistics* 647–655.
- KUNDU, S., MALLICK, B. K., BALADANDAYUTHAPANI, V. et al. (2019). Efficient Bayesian regularization for graphical model selection. *Bayesian Analysis* **14** 449–476. [MR3934093](#)
- LAN, W., LUO, R., TSAI, C.-L., WANG, H. and YANG, Y. (2015). Testing the diagonality of a large covariance matrix in a regression setting. *J. Bus. Econom. Statist.* **33** 76–86. [MR3303743](#)
- LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. [MR1805785](#)
- LEDAY, G. G. and RICHARDSON, S. (2018). Fast Bayesian inference in large Gaussian graphical models. *arXiv preprint arXiv:1803.08155*. [MR4041830](#)
- LEE, K. and LEE, J. (2018). Optimal Bayesian minimax rates for unconstrained large covariance matrices. *Bayesian Anal.* **13** 1211–1229. [MR3855369](#)
- LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics* **46** 280–307. [MR3766953](#)
- MARTIN, R., MESS, R. and WALKER, S. G. (2017). Empirical Bayes posterior concentration in sparse high-dimensional linear models. *Bernoulli* **23** 1822–1847. [MR3624879](#)
- NARISSETTY, N. N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. [MR3210987](#)
- ROCKOVA, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Statist.* **46** 401–437. [MR3766957](#)
- ROCKOVA, V. and GEORGE, E. I. (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *J. Am. Statist. Assoc.* **111** 1608–1622. [MR3601721](#)
- ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Am. Statist. Assoc.* **104** 177–186. [MR2504372](#)
- SRIVASTAVA, M. S., YANAGIHARA, H. and KUBOKAWA, T. (2014). Tests for covariance matrices in high dimension with less sample size. *J. Multivariate Anal.* **130** 289–309. [MR3229539](#)
- YANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. [MR3576552](#)

- YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 455–480. [MR3798874](#)
- YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol* **68** 49–67. [MR2212574](#)