Maximum Pairwise Bayes Factors for Covariance Structure Testing

Kyoungjae Lee¹, Lizhen Lin¹, and David Dunson²

¹Department of Applied and Computational Mathematics and Statistics, The University of Notre Dame
²Department of Statistical Science, Duke University

September 12, 2018

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. The performance of our graph selection method is demonstrated through applications to a sonar data set.

Key words: Bayesian hypothesis test; Modularization; Covariance matrix; Graph selection; High-dimensional.

1 Introduction

Consider a sample of observations from a high-dimensional normal model in which the number of variables $p$ can grow with the sample size $n$,

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

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 \( X_n = (X_1, \ldots, X_n)^T \in \mathbb{R}^{n \times p} \) be the data matrix. A one-sample covariance test can be reduced to the following simple form:

\[
H_0 : \Sigma_n = I_p \quad \text{versus} \quad H_1 : \Sigma_n \neq I_p,
\]

by noting that \( H_0 : \Sigma_n = I_p \) is equivalent to a null hypothesis \( H_0 : \Sigma_n = \Sigma_0 \) for any given positive definite matrix \( \Sigma_0 \) by applying the linear transformation \( X_i \mapsto \Sigma_0^{-1/2} X_i \).

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 tests 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. Throughout the paper, 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 et al. (2010) and Cai and Ma (2013) suggested consistent one-sample covariance tests based on unbiased estimators of \( \| \Sigma_n - I_p \|^2_F \), 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, \( \| \Sigma_n - I_p \|^2_F n/p \to \infty \) as \( n \to \infty \). This condition implies that they essentially adopted \( H_1 = \{ \Sigma_n : \| \Sigma_n - I_p \|^2_F \geq b_n p/n \} \) for some \( b_n \to \infty \) as \( n \to \infty \) as the alternative class. Cai and Ma (2013) proved that if we consider an alternative class \( H_1 = \{ \Sigma_n : \| \Sigma_n - I_p \|^2_F \geq \epsilon_n \} \), 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 \( \sqrt{p/n} \), although \( \sqrt{p/n} \) 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 (e.g., a consistent test) even when the condition \(\|\Sigma_n - I_p\|_F^2 n/p \to \infty\) fails to hold. We achieve this by proposing a novel Bayesian testing framework based on the maximum pairwise Bayes factor (mxPBF) 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 enables us to consider a different alternative class,

\[ H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_{\text{max}}^2 \geq C \log p/n\} \]

for some constant \(C > 0\), where \(\|A\|_{\text{max}} = \max_{i,j} |a_{ij}|\) 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 based on mxPBF is much more effective than the frequentist methods mentioned above.

The main contributions of this paper are as follows. The mxPBF is a general theoretically-supported method with a simple implementation that can be readily used by practitioners. mxPBF is also the first Bayesian test that has been shown to be consistent in the high-dimensional setting for the one-sample or diagonal covariance testing problems. In the one-sample case, the proposed test is rate-optimal in the sense that it can distinguish the elementwise maximum norm-based alternative class \(H_1 = \{\Sigma_n : \|\Sigma_n - I_p\|_{\text{max}}^2 \geq \epsilon_n\} \) 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.

The rest of the paper is organized as follows. Section 2 introduces the notations and definition of the mxPBF. Section 3 presents the main results of the paper: the one-sample covariance test, diagonality test and support recovery of the covariance matrix. In Section 4 the practical performance of the proposed methods is evaluated based on simulation study and real data analysis. R codes for implements of our empirical results are available at https://github.com/leekjstat/mxPBF. Concluding remarks are given in Section 5. Proofs of our main results are included in an Appendix, with additional results in Supplemental Materials.

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 \(a_n/b_n \to 0\) as \(n \to \infty\). For any vector \(x = (x_1, \ldots, x_p)^T \in \mathbb{R}^p\), we define the vector \(l_1\)- and \(l_2\)-norm as \(\|x\|_1 = \sum_{j=1}^p |x_i|\) 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 \( \chi_\lambda^2 = \chi_k^2(\lambda = 0) \). 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 (mxPBF)

In this subsection, we introduce the mxPBF 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 \( \tilde{X}_j \in \mathbb{R}^a \) be the \( j \)th column vector of \( X_n \). For any indices \( i \) and \( j \), based on the joint distribution (1), the conditional distribution of \( \tilde{X}_i \) given \( \tilde{X}_j \) is

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

where \( a_{ij} \in \mathbb{R} \) and \( \tau_{ij} > 0 \). We can view model (2) as a linear regression model given a design matrix \( \tilde{X}_j \). For each paired conditional model (2), we consider a testing problem

\[
H_{0,ij} : a_{ij} = 0, \quad \tau_{ij}^2 = 1 \quad \text{versus} \quad H_{1,ij} : \text{not } H_{0,ij}.
\]

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 (3),

\[
a_{ij} \mid \tau_{ij}^2 \sim N\left(0, \frac{\tau_{ij}^2}{\gamma} \|\tilde{X}_j\|^{-2}_1\bigg)\right), \quad \tau_{ij}^2 \sim IG(a_0, b_0),
\]

where \( \gamma = c_\gamma(n \vee p)^{-\alpha} \) and \( a_0, b_0, c_\gamma \) and \( \alpha \) are positive constants. The induced Bayes factor is

\[
B_{10}(\tilde{X}_i, \tilde{X}_j) := \frac{p(\tilde{X}_i \mid \tilde{X}_j, H_{1,ij})}{p(\tilde{X}_i \mid \tilde{X}_j, H_{0,ij})} = \frac{b_{10}^{\alpha_0}}{\Gamma(a_0)} \left(\frac{\gamma}{1 + \gamma}\right)^{1/2} \Gamma\left(\frac{n}{2} + a_0\right) e^{n\tau_{ij}^2/2} \left(\frac{\tau_{ij}^2}{2} \gamma + b_0\right)^{-n/2 - a_0},
\]

where \( n\tau_i^2 = \|\tilde{X}_i\|_2^2, n\tau_{ij,\gamma}^2 = \tilde{X}_i^T(I_n - (1 + \gamma)^{-1}H_j)\tilde{X}_i \) and \( H_j = \tilde{X}_j(\tilde{X}_j^T \tilde{X}_j)^{-1} \tilde{X}_j^T \).

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 (mxPBF),

\[
B_{\text{max},10}(X_n) := \max_{i \neq j} B_{10}(\tilde{X}_i, \tilde{X}_j).
\]

A large value for \( B_{\text{max},10}(X_n) \) provides evidence supporting the alternative hypothesis. By taking a maximum, the mxPBF supports the alternative hypothesis if at least one of the pairwise Bayes factors support 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 the mxPBF defined in (5) for the one-sample covariance test

\[ H_0 : \Sigma_n = I_p \quad \text{versus} \quad H_1 : \Sigma_n \neq I_p. \] (6)

We first introduce assumptions for consistency under \( H_1 : \Sigma_n \neq I_p \). Let \( \Sigma_0 = (\sigma_{0,ij}) \in C_p \) be the true covariance matrix, implying the conditional distribution of \( \tilde{X}_i \) given \( \tilde{X}_j \) is

\[ \tilde{X}_i \mid \tilde{X}_j \sim N_n(a_{0,ij}\tilde{X}_j, \tau^2_{0,ij}I_n) \] (7)

under \( P_0 \), where \( a_{0,ij} = \sigma_{0,ij}/\sigma_{0,jj}, \tau^2_{0,ij} = \sigma_{0,ii}(1 - \sigma^2_{0,ij}/(\sigma_{0,ii}\sigma_{0,jj})) \) and \( P_0 \) is the probability measure corresponding to model (1) with \( \Sigma_0 \). Note that \( \tau^2_{0,ij} = \sigma_{0,ii} \) if and only if \( a_{0,ij} = 0 \).

Under the true alternative \( H_1 : \Sigma_n \neq I_p \), we assume that \( \Sigma_0 \) satisfies at least one of the following conditions:

(A1) There exists an index \( i \) satisfying

\[ |\sigma_{0,ii} - 1| \geq \left(4\sigma_{0,ii}\sqrt{C_1} + C_2 + \frac{2b_0}{\sqrt{n\log(n \vee p)}}\right)\sqrt{\frac{\log(n \vee p)}{n}} \] (8)

for some constants \( C_1 > 0 \) and \( C_2 > 2\sqrt{\alpha + 2} \).

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

\[ |\tau^2_{0,ij} - 1| \geq \left(4\tau^2_{0,ij}\sqrt{C_1} + C_2 + \frac{2b_0 + \tau^2_{0,ij}}{\sqrt{n\log(n \vee p)}}\right)\sqrt{\frac{\log(n \vee p)}{n}}. \] (9)

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

\[ \sigma^2_{0,ij} \geq \frac{\sigma_{0,jj}}{1 - 2\sqrt{C_1}\epsilon_0} \left\{ 9C_1\tau^2_{0,ij} \sqrt{\frac{C_4(\alpha + 2)}{C_3}} \frac{\log(n \vee p)}{n} \right\} \] (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 \( C_1 = \epsilon, C_2 = 2\sqrt{\alpha + 2} + \epsilon, C_3 = 1 - \epsilon^{1/4} \) and \( C_4 = 1 + \epsilon \).

Condition (A1) is required to detect a non-unit variance \( \sigma_{0,ii} \). The condition (8) 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 et al.; 2015; Martin et al.; 2017). Interestingly, the rate of lower bound in (A1) is given by \( \sqrt{\log(n \vee p)/n} \), 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 also the beta-min condition for off-diagonal elements of the covariance matrix. The right hand side of (10) implies that the larger \(\tau_{0,ij}\) and \(\sigma_{0,jj}\) are, the larger value of \(\sigma_{0,ij}^2\) is required to consistently detect the dependency between \(\tilde{X}_i\) and \(\tilde{X}_j\). Based on (7), the former means that larger value of residual variance \(\tau_{0,ij}\) makes the inference harder. The conditional expectation in (7) can be considered as 

\[
\sigma_{0,ij}/\sigma_{0,jj}^{1/2} \cdot (\tilde{X}_j/\sigma_{0,jj}^{1/2}) \equiv \sigma_{0,ij}/\sigma_{0,jj}^{1/2} \cdot \tilde{Z}_j, \text{ where } \tilde{Z}_j \sim N_n(0, I_n).
\]

Thus, a large value of \(\sigma_{0,jj}\) makes it hard to detect nonzero \(\sigma_{0,ij}\). In summary, conditions (A1)–(A3) imply that

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

for some constant \(C > 0\), which corresponds to the meaningful difference we mentioned earlier. In fact, the rate \(\log p/n\) is optimal to guarantee the consistency under both hypotheses (Theorem 3.2).

Theorem 3.1 shows that the mxPBF is consistent 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 \(H_0 : \Sigma_n = I_p\) versus \(H_1 : \Sigma_n \neq I_p\). Consider prior (4) under \(H_{1,ij}\) in (3) with \(a_0 > 1\) and \(\alpha > 8(1+\sqrt{2}\epsilon_0)/\{1 - 2\sqrt{2}\epsilon_0(1 + \sqrt{2}\epsilon_0)\}\) for some small constant \(0 < \epsilon_0 < 3(4C_0)^{-1}\). Assume that \(\log p \leq \epsilon_0^2 n\) for all large \(n\), and \(\Sigma_0\) satisfies at least one of conditions (A1)–(A3) or \(\Sigma_0 = I_p\). Then, the mxPBF defined in (5) is consistent under \(\mathbb{P}_0\). Moreover, under \(H_0 : \Sigma_n = I_p\), we have

\[
-\log B_{\text{max},10}(X_n) = O_p(\log(n \lor p)),
\]

and under \(H_1 : \Sigma_n \neq I_p\),

\[
\log B_{\text{max},10}(X_n) = O_p(\log(n \lor p)).
\]

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 \(\mathbb{P}_0(E_{ij}^c) \to 0\) as \(n \to \infty\). To show the consistency of the mxPBF under \(H_0\), we need to prove that \(\sum_{i \neq j} \mathbb{P}_0(E_{ij}^c) \to 0\) as \(n \to \infty\), which means that the false discovery rate converges to zero. The condition for \(\alpha\) in Theorem 3.1 is closely related to this requirement. To show the consistency of the mxPBF under \(H_1\), it suffices to show \(\mathbb{P}_0(E_{ij}^c) \to 0\) as \(n \to \infty\) 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).

**Remark** Compared to existing frequentist one-sample covariance tests in Chen et al. (2010) and Cai and Ma (2013), our proposed test can detect \(H_1 : \Sigma_n \neq I_p\) if there is at least one meaningful nonzero entry in \(\Sigma_0 - I_p\), i.e. \(\|\Sigma_0 - I_p\|_{\text{max}}^2 \geq C \log p/n\) for some constant \(C > 0\). For example,
suppose $p \geq n$ and $\Sigma_0 = I_p + \tau I (i = j = 1)$ for some constant $\tau > 0$, which implies $\|\Sigma_0 - I_p\|_F^2 = \tau^2$. Then, the mxPBF consistently detects the true $H_1$. However, the powers of the frequentist tests do not tend to 1 in this case, because the condition $\|\Sigma_0 - I_p\|_F^2 \gg p/n$ is not met.

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 $\|\Sigma_0 - I_p\|_{\text{max}}^2 \geq C \log p/n$ 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 $E_\Sigma$ be the expectation corresponding to model (1). For a given constant $C_* > 0$, define

$$H_1(C_*) = \left\{ \Sigma \in C_p : \|\Sigma - I_p\|_{\text{max}}^2 \geq C_* \log p / n \right\}.$$  

If $C_*^2 \leq 2$, then for any consistent test $\phi$ such that $E_{I_p,\phi} \to 0$ as $n \to \infty$,

$$\limsup_{n \to \infty} \inf_{\Sigma \in H_1(C_*)} E_{\Sigma,\phi} \leq \frac{1}{2}.$$

### 3.2 Testing Diagonality

We now extend mxPBF to test diagonality of the covariance matrix:

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

where $\Sigma_n = (\sigma_{i j})$. The above hypothesis testing problem can be modularized into many pairwise independence tests

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

for all $1 \leq i < j \leq p$. We can adopt the mxPBF 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 (2), the null hypothesis $H_{0,ij}$ in (12) is equivalent to $H'_{0,ij} : a_{ij} = 0$. We suggest the prior $\tau_{ij}^2 \sim IG(a_0, b_0)$ under $H_{0,ij}$ and prior (4) under $H_{1,ij}$, which leads to the pairwise Bayes factor

$$\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = \left( \frac{\gamma}{1 + \gamma} \right)^{1/2} \left( \frac{n \tau_{ij,0}^2 + 2 b_0}{n \tau_{ij}^2 + 2 b_0} \right)^{-n/2-a_0}.$$  

We suggest using the mxPBF,

$$\tilde{B}_{\text{max},10}(X_n) := \max_{i \neq j} \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j),$$  

(13)
for the hypothesis testing problem (11). Theorem 3.3 states the consistency of the mxPBF 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

\[
\sigma_{0,ij}^2 \geq \frac{C_4 \sigma_{0,ij}}{1 - 2\epsilon_0 \sqrt{C_1}} \left\{ \frac{9C_1 \tau_{0,ij}^2}{(1 - C_3)^2} \right\} \frac{\alpha(1 + \gamma)(1 + 4\epsilon_0 \sqrt{C_1}) \sigma_{0,ii} + 2b_0 n^{-1}}{C_3} \log(n \lor p) \]  

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 \(\tau_{0,ij}^2 \sim \text{IG}(a_0, b_0)\) under \(H_{0,ij}\) and prior (4) under \(H_{1,ij}\) in (12) with \(\alpha > 4/(1 - 3\sqrt{2}\epsilon_0)\) for some small constant \(0 < \epsilon_0 < 1/(3\sqrt{2})\). Assume that \(\log p \leq \epsilon_2 n\) for all large \(n\), and \(\Sigma_0\) satisfies condition (A4) or \(\Sigma_0 = I_p\). Then the mxPBF defined in (13) is consistent under \(P_0\). Moreover, under \(H_0 : \sigma_{ij} = 0\) for any \(i \neq j\),

\[-\log \tilde{B}_{\text{max,10}}(\tilde{X}_i, \tilde{X}_j) = O_p(\log(n \lor p)),\]

and under \(H_1 : \text{not } H_0\),

\[\log \tilde{B}_{\text{max,10}}(\tilde{X}_i, \tilde{X}_j) = O_p(\log(n \lor p)).\]

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), the mxPBF consistently detects the true alternative hypothesis. Similar to Theorem 3.1, the condition for \(\alpha\) is required to control the false discovery rate, and the mxPBF has similar rates of convergence under both hypotheses.

As a by-product, when pairwise independence testing (12) itself is of interest, we suggest a pairwise Bayes factor

\[
\tilde{B}_{\text{pair,10}}(\tilde{X}_i, \tilde{X}_j) := \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) \lor \tilde{B}_{10}(\tilde{X}_j, \tilde{X}_i),
\]

which can be shown to be consistent. Note that decisions from tests based on \(\tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j)\) and \(\tilde{B}_{10}(\tilde{X}_j, \tilde{X}_i)\) can differ in finite samples, which is undesirable. To obtain an order-invariant test, we suggest using the pairwise Bayes factor (14). For consistency under the alternative hypothesis, we assume

\[
\sigma_{0,ij}^2 \geq \frac{C_4 \sigma_{0,ij}}{1 - 2\epsilon_0 \sqrt{C_1}} \left\{ \frac{9C_1 \tau_{0,ij}^2}{(1 - C_3)^2} \right\} \frac{\alpha(1 + \gamma)(1 + 4\epsilon_0 \sqrt{C_1}) \sigma_{0,ii} + 2b_0 n^{-1}}{C_3} \log n \frac{n}{n} \]  

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 \(\log(n \lor p)\), it coincides with condition (A4). The proof of Corollary 3.4 is obvious from that of Theorem 3.3, thus it is omitted.
Corollary 3.4 Consider model (1) and a hypothesis testing problem (12) for a given pair \((i, j)\) such that \(i \neq j\). Suppose we use prior \(\tau_{ij}^2 \sim IG(a_0, b_0)\) under \(H_{0,ij}\) and prior (4) under \(H_{1,ij}\) with \\
\(\gamma = c_\gamma n^{-\alpha}\) for some positive constants \(a_0, b_0, c_\gamma\) and \(\alpha\). Assume that either \(\sigma_{0,ij} = 0\) or at least one of \(\sigma_{0,ij}\) and \(\sigma_{0,ji}\) satisfy (15). Then, the pairwise Bayes factor \(\tilde{B}_{\text{pair},10}(\tilde{X}_i, \tilde{X}_j)\) is consistent under \(\mathbb{P}_0\). Moreover, under \(H_{0,ij}: \sigma_{ij} = 0\), we have \\
\[-\log \tilde{B}_{\text{pair},10}(\tilde{X}_i, \tilde{X}_j) = O_p(\log n),\]
and under \(H_{1,ij}: \sigma_{ij} \neq 0\),
\[\log \tilde{B}_{\text{pair},10}(\tilde{X}_i, \tilde{X}_j) = O_p(\log n).\]

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. (2013) proposed the regularized inverse Wishart (RIW) 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 et al. (2018) adopted the spike-and-slab Lasso prior (Ročková and George; 2016; Ročková et al.; 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.

To tackle this gap, we propose a scalable graph selection scheme for high-dimensional covariance matrices based on pairwise Bayes factors. By looking closer at the proof of Theorem 3.3, it reveals that each pairwise Bayes factor \(\tilde{B}_{\text{pair},10}(\tilde{X}_i, \tilde{X}_j)\) can consistently determine whether the corresponding covariance element \(\sigma_{0,ij}\) is zero or not. Thus, we suggest using the estimated index set \\
\[\tilde{S}_{\text{pair}} = \{(i, j): \, 2\log \tilde{B}_{\text{pair},10}(\tilde{X}_i, \tilde{X}_j) > C_{\text{sel}}, \quad 1 \leq i < j \leq p\},\]
for some constant \(C_{\text{sel}} > 0\). Asymptotically, any constant \(C_{\text{sel}}\) can be used for consistent selection. However, in practice, we suggest using the threshold \(C_{\text{sel}} = 10\), which corresponds to ‘very strong evidence’ under the Bayes factor thresholds of Kass and Raftery (1995). In the frequentist literature, Drton and Perlman (2004) and Drton and Perlman (2007) proposed selection procedures using a related idea to (16), 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 $\hat{S}_{\text{pair}}$, we introduce the following condition for some constants $0 < C_3 < 1, C_4 > 1$ and $C_5 > 2$:

(A.ij) For a given pair $(i, j)$ such that $i \neq j$,

$$
\sigma^2_{0,ij} \geq \frac{C_4 \sigma_{0,jj}^2}{1 - 2\epsilon_0 \sqrt{C_5}} \left\{ \frac{9C_5 \sigma_{0,ij}^2}{(1 - C_3)^2} \right\} \left[ \frac{\alpha(1 + \gamma)\{(1 + 4\epsilon_0 \sqrt{C_5})\sigma_{0,ii} + 2b_0 n^{-1}\}}{C_3} \right] \log(n \vee p) \frac{\sigma_{0,ii} + 2b_0 n^{-1}}{n}.
$$

The beta-min condition (A.ij) is almost the same with (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 3.5 states that (16) achieves model selection consistency if at least conditions (A.ij) or (A.ji) hold for all $(i,j) \in S(\Sigma_0)$.

**Theorem 3.5** Consider model (1) and prior (4) with $\alpha > 4/(1 - 3\epsilon_0 \sqrt{2})$ for some small constant $0 < \epsilon_0 < (3\sqrt{2})^{-1}$ and each pair $(i, j)$ such that $i \neq j$. Assume that $\log p \leq \epsilon_0^2 n$ for all large $n$ and at least one of conditions (A.ij) and (A.ji) hold for all $(i, j) \in S(\Sigma_0)$. Then, we have

$$
\lim_{n \to \infty} P_0 \left( \hat{S}_{\text{pair}} = S(\Sigma_0) \right) = 1.
$$

We note that $\hat{S}_{\text{pair}}$ consistently recovers the support of the true covariance matrix $\Sigma_0$ regardless of the true sparsity as long as $\log p \leq \epsilon_0^2 n$ and nonzero entries satisfy the beta-min condition (A.ij).

Rothman et al. (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 the consistent support recovery result of the covariance matrix using adaptive thresholding. In terms of the condition on $n$ and $p$, our condition, $\log p \leq \epsilon_0^2 n$, is much weaker than the conditions used in the literature. The beta-min condition (A.ij) is similar to that in Cai and Liu (2011) and also has the same rate to that in Rothman et al. (2009) if we assume $\max_i \sigma_{0,ii} \leq M$ for some $M > 0$. Thus, the required condition in Theorem 3.5 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. The four hyperparameters were chosen as $a_0 = 2, b_0 = 2, c_\gamma = 10$ and $\alpha = 8.01(1 - 1/\log n)$, which satisfy the sufficient conditions in Theorem 3.1. If we assume a small $\epsilon_0 > 0$, the above choice of $\alpha$ asymptotically satisfies $\alpha > 8(1 + \sqrt{2}\epsilon_0)^2/(1 - 2\sqrt{2}\epsilon_0(1 + \sqrt{2}\epsilon_0))$.

We compare our one-sample covariance test with existing frequentist tests, proposed by Cai and
Ma (2013) and Srivastava et al. (2014). The test suggested by Srivastava et al. (2014) is based on estimating the squared Frobenius norm, and has a similar perspective to the test proposed by Cai and Ma (2013). The test statistic is asymptotically normal if \( n \leq p^\delta \) for some \( 1/2 < \delta < 1 \), but does not have any theoretical guarantee on power.

We generated 100 data sets from each hypothesis \( H_0 : \Sigma_n = I_p \) and \( H_1 : \Sigma_n \neq I_p \), 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 a tridiagonal matrix \( \Sigma_0 = (\sigma_{0,ij}) \) with

\[
\sigma_{0,ij} = I(i = j) + \rho \cdot I(|i - j| = 1)
\]

(17)

for some signal strength constant \( \rho > 0 \) ranging from 0.1 to 0.5 by 0.025. For each value of \( \rho \), we generated 100 data sets from the multivariate normal distribution \( N_p(0, \Sigma_0) \). As a second case for \( \Sigma_0 \), we let

\[
\sigma_{0,ij} = I(i = j) + \rho \cdot I(i = 1, j = 2) + \rho \cdot I(i = 2, j = 1),
\]

(18)

for some constant \( \rho > 0 \). Because (18) has signals at only two locations, the difference between \( \Sigma_0 \) and \( I_p \) is sparse. mxPBF and the frequentist tests require \( \|\Sigma_0 - I_p\|_{\text{max}}^2 \geq C \log p/n \) and \( \|\Sigma_0 - I_p\|_F^2 \gg p/n \), respectively for consistency, while we have \( \|\Sigma_0 - I_p\|_F^2 = 2\|\Sigma_0 - I_p\|_{\text{max}}^2 \) for (18). Hence, mxPBF is potentially much more powerful than the tests in Cai and Ma (2013) and Srivastava et al. (2014) in this setting. To check this conjecture, we increased \( \rho \) from 0.3 to 0.8 by 0.025 and generated 100 simulated data from \( N_p(0, \Sigma_0) \).

We calculated receiver operating characteristic (ROC) curves to illustrate and compare the performance of the tests. For each setting, points of the ROC curves were obtained based on various thresholds and significance levels for mxPBF and frequentist tests, respectively. We tried \( n = 100, 200, 300 \) and \( p = 200, 500 \) for each setting. Figure 1 shows ROC curves based on 100 simulated data from \( N_p(0, I_p) \) and 100 simulated data from \( N_p(0, \Sigma_0) \) with a tridiagonal \( \Sigma_0 \) given in (17), when \((n, p) = (100, 200)\) and \((n, p) = (200, 500)\). In this setting, the tests in Cai and Ma (2013) and Srivastava et al. (2014) seem to work better than the test proposed in this paper. This is as expected since these earlier tests are based on the squared Frobenius norm \( \|\Sigma_0 - I_p\|_F^2 = 2(p - 1)\rho^2 \), while mxPBF focuses on the maximum difference \( \|\Sigma_0 - I_p\|_{\text{max}} = \rho \). Thus, when the difference between the true covariance and the point null is dense, the test based on the Frobenius norm difference would be more powerful. Figure 2 shows ROC curves based on 100 simulated data from \( N_p(0, I_p) \) and 100 simulated data from \( N_p(0, \Sigma_0) \) with a tridiagonal \( \Sigma_0 \) given in (18), when \((n, p) = (100, 200)\) and \((n, p) = (200, 500)\). As expected, the mxPBF is much more powerful than the frequentist tests when \( \Sigma_0 - I_p \) is sparse. The mxPBF test works reasonably well when the signal \( \rho \) is larger than 0.5, but the frequentist tests do not work well even when \( \rho \) is 0.8. Furthermore, unlike the previous result described in Figure 1, the performance of the frequentist tests is not getting better as \( n \) and \( p \) increase, while mxPBF has better performance when \((n, p) = (100, 200)\) than \((n, p) = (200, 500)\).
Figure 1: ROC curves are represented for the three tests based on 100 simulated data sets for each hypothesis $H_0$ and $H_1$, where (17) was used for $H_1$. mxPBF, CM and SYK represent the test proposed in this paper, Cai and Ma (2013) and Srivastava et al. (2014), respectively.

It confirms the theoretical property that the tests in Cai and Ma (2013) and Srivastava et al. (2014) require $\|\Sigma_0 - I_p\|_F^2 \gg p/n$ to have power tending to one. Note that $\|\Sigma_0 - I_p\|_F^2 = 2\rho^2$ remains the same while $p/n$ increases as the setting is changed from $(n,p) = (100,200)$ to $(n,p) = (200,500)$.

4.2 Simulation Study: Testing Diagonality

We conducted a simulation study to illustrate the performance of the diagonality test based on the mxPBF. The hyperparameters in the mxPBF were chosen as in section 4.1 except $\alpha = 4.01(1 - 1/\log n)$. For each hypothesis $H_0 : \sigma_{ij}$ for any $i \neq j$ and $H_1 : \not \sim H_0$, 100 data sets were generated. The two structures of $\Sigma_0$ under $H_1$ used in the previous section, (17) and (18), were considered.

We compare our test with frequentist tests suggested by Cai and Jiang (2011) and Lan et al. (2015). Cai and Jiang (2011) proposed a diagonality test and showed its asymptotic null distribution in the high-dimensional setting. Their test is based on the maximum of sample correlations between variables. Note that $\hat{\tau}_{ij}^2$ in the pairwise Bayes factor $\hat{B}_{10}(\hat{X}_i, \hat{X}_j)$ is a decreasing function of the sample correlation between $\hat{X}_i$ and $\hat{X}_j$. Thus, their test has a similar aspect to our test. Lan et al. (2015) developed a test in the regression setting and showed asymptotic normality. Their test statistic is based on the squared Frobenius norm of a sample covariance matrix, so it is expected...
Figure 2: ROC curves are represented for the three tests based on 100 simulated data sets for each hypothesis $H_0$ and $H_1$, where (18) was used for $H_1$. mxPBF, CM and SYK represent the test proposed in this paper, Cai and Ma (2013) and Srivastava et al. (2014), respectively.

that it will have a high power when off-diagonal entries of the true covariance matrix are dense.

The results for the case $(n, p) = (200, 500)$ are reported in Figure 3. ROC curves of tests based on the mxPBF and proposed by Cai and Jiang (2011) are almost identical in every setting. As expected, the test suggested by Lan et al. (2015) is more powerful when $\Sigma_0 - I_p$ is dense, which corresponds to “Type 1” in Figure 3. The tests proposed in this paper and Cai and Jiang (2011) have less power, but also work reasonably well when the signal $\rho$ is larger than 0.15. On the other hand, they tend to have a high power when $\Sigma_0 - I_p$ is sparse, which corresponds to “Type 2” in Figure 3. The ROC curve of the test of Lan et al. (2015) is close to the straight line, so it does not work well in this setting; similar behavior occurs even when $\rho$ is 0.8.

4.3 Support Recovery using Sonar Data

To describe the practical performance of the support recovery procedure (16), $\hat{S}_{pair}$, we analyzed the sonar data set used by Gorman and Sejnowski (1988). The data record sonar signals collected from a rock and a metal cylinder, which were both approximately 5 feet long and embedded in the seabed. The impinging pulse sent to the targets was a wide-band linear frequency-modulated (FM) chirp $(ka = 55.6)$. Data were collected within 10 meters of each object at various angles, spanning
Figure 3: ROC curves are represented for the three tests based on 100 simulated data sets for each hypothesis $H_0$ and $H_1$. Type 1 and Type 2 mean that the true covariance matrix $\Sigma_0$ under $H_1$ were generated from (17) and (18), respectively. mxPBF, Lan and CJ represent the test proposed in this paper, Lan et al. (2015) and Cai and Jiang (2011), respectively.

$180^\circ$ degrees for the rock and $90^\circ$ degrees for the metal cylinder. The rock data contains $n_r = 97$ returned patterns, while the metal data contains $n_m = 111$ returns. Each pattern consists of $p = 60$ values supported in the range $[0, 1]$ representing the normalized frequency of energy measurements integrated over a certain period of time. We transformed the data by adding a small value and taking the logarithm to convert from $[0, 1]$ to $\mathbb{R}$ support. After transformation, the data sets were centered.

For pairwise Bayes factors, the hyperparameters were set at $a_0 = 2$, $b_0 = 2$, $c_\gamma = 10$ and $\alpha = 4.01(1 - 1/\log n)$. The threshold $C_{sel}$ in (16) was set at 10 as a default. Furthermore, as a pragmatic approach, we also incorporated cross-validation (CV) 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 $I_1 \cup I_2 = \{1, \ldots, n\}$. 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)

\[ MSE(C_{sel}) = \frac{1}{p} \sum_{j=1}^{p} \sum_{i \in I_1} \{ \sum_{i \in I_1} (X_{ij} - X_{il} \hat{\beta}_{jl})^2 / (n_1 - 1) \} / |S_j(C_{sel})|, \]

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 0 to 20 with increment 0.2, and we selected \( \hat{C}_{sel} \) which minimizes \( 50^{-1} \sum_{\nu=1}^{50} MSE_{\nu}(C_{sel}) \), where \( MSE_{\nu}(C_{sel}) \) is the MSE based on the \( \nu \)th split. The selected \( \hat{C}_{sel} \) were 15.2 and 19 for the rock and metal data, respectively.

We compared our method with generalized thresholding estimators of Rothman et al. (2009) and Cai and Liu (2011). Both estimators are based on the sample covariance with the elementwise thresholding rule \( s_\lambda(\cdot) \), satisfying certain conditions, where \( \lambda \) is the threshold. The difference is that Rothman et al. (2009) used a universal threshold \( \lambda = \delta \sqrt{\log p / n} \), while Cai and Liu (2011) used an individual threshold \( \hat{\lambda}_{ij} = \delta \sqrt{\hat{\theta}_{ij} \log p / n} \) with a data-dependent \( \hat{\theta}_{ij} \). We denote thresholding estimators proposed by Rothman et al. (2009) and Cai and Liu (2011) by \( \hat{\Sigma}_\delta \) and \( \hat{\Sigma}_\delta^* \), respectively. For the thresholding rule, we used the “adaptive lasso” rule, \( s_\lambda(\sigma) = \sigma \cdot \max(1 - |\lambda/\sigma|^{\eta},0) \) with \( \eta = 4 \), because it gave good support recovery results in simulation studies in Rothman et al. (2009) and Cai and Liu (2011). We adopted the CV method, described in section 4 of Cai and Liu (2011), to select the tuning parameter \( \delta \). We randomly divided the data 50 times into two subsamples with sizes \( n_1 = \lceil n/3 \rceil \) and \( n_2 = n - n_1 \), and found \( \hat{\delta} \) minimizing the objective function \( 50^{-1} \sum_{\nu=1}^{50} \| S_{\nu}^{1,\nu} - \hat{\Sigma}_{\delta}^{2,\nu} \|_F \), where \( S_{\nu}^{1,\nu} \) is the sample covariance based on subsamples with size \( n_1 \) from the \( \nu \)th split, and \( \hat{\Sigma}_{\delta}^{2,\nu} \) is the generalized thresholding estimator based on the remaining samples from the \( \nu \)th split. The tuning parameter \( \delta \) was varied from 0 to 4 with increment 0.1.

Figure 4 shows the support recovery results for the three methods described above and absolute sample correlation matrices. The first row represents the estimated supports based on \( \hat{S}_{pair,10} \), \( \hat{S}_{pair,\hat{C}_{sel}} \), \( \hat{\Sigma}_\delta \) and \( \hat{\Sigma}_\delta^* \) for the rock data set, where \( \hat{S}_{pair,C_{sel}} \) denotes the estimated support via pairwise Bayes factors with the threshold \( C_{sel} \). One can see that \( \hat{S}_{pair,10} \), \( \hat{S}_{pair,\hat{C}_{sel}} \) and \( \hat{\Sigma}_\delta^* \) give sparse support estimates, while \( \hat{\Sigma}_\delta \) provides a relatively dense support estimate. Note that \( \hat{\Sigma}_\delta^* \) does not capture the small block structure in the upper right side, although it seems to be significant. It is due to relatively large values of \( \hat{\theta}_{ij} \)'s, which correspond to empirical variances of sample covariances between \( \hat{X}_i \) and \( \hat{X}_j \)'s. Similarly, the second row represents the estimated supports for the metal data set. Our support recovery procedure provides a sparse support estimate. On the other hand, \( \hat{\Sigma}_\delta \) and \( \hat{\Sigma}_\delta^* \) give dense support recovery results. Our approach has the advantage of producing a sparser, and hence potentially more interpretable estimate of support. \( \hat{S}_{pair,10} \) should give a sparse support estimate because we used the threshold \( C_{sel} = 10 \), removing all but ‘very strong’ covariances. If one wants to have more dense support estimates, a smaller threshold \( C_{sel} \) can be used.
Figure 4: The absolute sample correlation matrix and estimated supports of the rock data (top row) and metal data (bottom row), with 1’s representing the estimated support from each method. The first, second, third and fourth columns represent the estimated supports based on $\hat{S}_{\text{pair},10}$, $\hat{S}_{\text{pair},\hat{C}_{\text{set}}}$, $\hat{\Sigma}_{\delta}$ and $\hat{\Sigma}^\star_{\delta}$, respectively.

5 Discussion

We have focused on covariance matrix structure testing in this paper, but the mxPBF 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 mxPBF 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 scheme achieving desirable theoretical properties similar to the mxPBF. Especially when $p$ is huge, it will effectively reduce the computational complexity. The mxPBF 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 (or signals), 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

Proof of Theorem 3.1

Proof For a given pair \((i,j)\),

\[
\log B_{10}(\tilde{X}_i, \tilde{X}_j) = a_0 \log b_0 + \frac{1}{2} \log \left(\frac{\gamma}{1 + \gamma}\right) + \left\{ \log \Gamma\left(\frac{n}{2} + a_0\right) - \log \Gamma(a_0) - \left(\frac{n}{2} + a_0\right) \log \frac{n}{2} + \frac{n}{2} \right\} + \left\{ \frac{n \hat{\tau}^2_i - \left(\frac{n}{2} + a_0\right) \log \left(\hat{\tau}^2_{ij,\gamma} + \frac{2b_0}{n}\right) - \frac{n}{2} \right\}.
\]

Note that by Theorem 1 of Kečkić and Vasić (1971),

\[
(b - 1) \log b - (a - 1) \log a + a - b \leq \log \Gamma(b) - \log \Gamma(a) \leq (b - 1/2) \log b - (a - 1/2) \log a + a - b
\]

for any \(b > a > 1\). Thus, it is easy to see that (20) is equal to

\[
-C \log \left(\frac{n}{2} + a_0\right) + \left( a_0 + \frac{2a_0^2}{n} \right) \log \left(1 + \frac{2a_0}{n}\right) \frac{n}{2b_0} + C'
\]

for some constants \(1/2 < C < 1\) and \(-(a_0 - 1/2) \log a_0 < C' < -(a_0 - 1) \log a_0\), which is of order \(O(-\log n)\). Since \(\gamma = c_\gamma(n \lor p)^{-\alpha}\), (19) is equal to

\[
a_0 \log b_0 + \frac{1}{2} \log c_\gamma - \frac{\alpha}{2} \log (n \lor p) - \frac{1}{2} \log \left(1 + c_\gamma(n \lor p)^{-\alpha}\right)
\]

\[
= -\frac{\alpha}{2} \log (n \lor p) + C
\]

for some constant \(C\). Thus, we only need to focus on the behavior of (21).

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 \(i\),

Case (iii). \(\tau^2_{0,ij}\) satisfies condition (A2) for some pair \((i,j)\), and

Case (iv). \(\sigma_{0,ij}\) satisfies condition (A3) for some pair \((i,j)\).

Note that (21) can be expressed as

\[
\frac{n}{2} \left\{ \hat{\tau}^2_{ij,\gamma} + \frac{2b_0}{n} - \log \left(\hat{\tau}^2_{ij,\gamma} + \frac{2b_0}{n}\right) - 1 \right\}
\]

\[
+ \frac{n}{2} \left(\hat{\tau}^2_i - \hat{\tau}^2_{ij,\gamma}\right)
\]

\[
- \left\{ a_0 \log \left(\hat{\tau}^2_{ij,\gamma} + \frac{2b_0}{n}\right) + b_0 \right\}.
\]

We will calculate the rate of the above three terms (24)-(26) for every possible case. More precisely, we will show that for all sufficiently large \(n\), \(\log B_{10}(\tilde{X}_i, \tilde{X}_j) \leq -C \log (n \lor p)\) with probability at least
1 - (n \lor p)^{-c}$ for some constants $C > 0$ and $c > 2$ under Case (i), and \( \log B_{10}(\hat{X}_i, \hat{X}_j) \geq C' \log(n \lor p) \) with probability at least $1 - (n \lor p)^{-c'}$ for some constants $C' > 0$ and $c' > 0$ under Cases (ii)–(iv). Then, we have

$$
P_0\left( \log B_{\text{max}, 10}(X_n) \leq -C \log(n \lor p) \right) = 1 - P_0\left( \max_{(i,j):i \neq j} \log B_{10}(\hat{X}_i, \hat{X}_j) > -C \log(n \lor p) \right)
\geq 1 - \sum_{(i,j):i \neq j} P_0\left( \log B_{10}(\hat{X}_i, \hat{X}_j) > -C \log(n \lor p) \right)
\geq 1 - (n \lor p)^{-c + 2}
$$

if all pairs $\langle i, j \rangle$ satisfy Case (i) (i.e. $H_0 : \Sigma_n = I_p$ holds), and

$$
P_0\left( \log B_{\text{max}, 10}(X_n) \geq C' \log(n \lor p) \right) \geq \max_{(i,j):i \neq j} P_0\left( \log B_{10}(\hat{X}_i, \hat{X}_j) \geq C' \log(n \lor p) \right)
\geq 1 - (n \lor p)^{-c'}
$$

if there exists at least one pair $\langle i, j \rangle$ satisfying one of Cases (ii)–(iv) (i.e. $H_1 : \Sigma_n \neq I_p$ holds).

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

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

with probability at least $1 - 2(n \lor p)^{-C}$ for some constant $C > 2$. Since $\gamma = c_r(n \lor p)^{-\alpha}$ and the conditions on $\alpha$, it means

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

(27)

with probability at least $1 - 2(n \lor p)^{-C}$ 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 $n(\hat{\tau}_{ij}^2 - \hat{\tau}_{ij, \gamma}^2) \sim \chi_1^2$ under Case (i). Thus, we have

$$
P_0\left( \left| n(\hat{\tau}_{ij}^2 - \hat{\tau}_{ij, \gamma}^2) - 1 \right| \leq 2C' c \log(n \lor p) \right) \geq 1 - 2(n \lor p)^{-c},
$$

(28)

$$
P_0\left( \left| n - n^{-1}\hat{\tau}_{ij}^2 - 1 \right| \leq 2C' \sqrt{\frac{c \log(n \lor p)}{n - 1}} \right) \geq 1 - 2(n \lor p)^{-c}
$$

(29)

for any constants $c > 2$, $C' > 1 + \sqrt{2}c_0$ and all large $n$. The inequalities (27) and (29) imply

$$
P_0\left( \left| \hat{\tau}_{ij, \gamma}^2 + \frac{2b_0}{n} - 1 \right| < C \sqrt{\frac{\log(n \lor p)}{n}} \right) \geq 1 - 4(n \lor p)^{-c}
$$

(30)
for any constant $C > 2(\sqrt{2} + 2\epsilon_0)$ and all large $n$. Note that by the Taylor expansion of $\log(1 + x)$,

$$
x - \log(1 + x) = x - \left(x - \frac{1}{2} x^2 + \frac{1}{3} x^3 - \frac{1}{4} x^4 + \cdots\right)
$$

$$
= \frac{1}{2} x^2 - \frac{1}{3} x^3 + \frac{1}{4} x^4 + \cdots \leq \frac{1}{2} x^2 \cdot (1 - |x|)^{-1}
$$

for small $|x|$. Thus, on the event in (30), (24) is bounded above by $4^{-1}C^2(1 - C\epsilon_0)^{-1} \log(n \lor p)$ for any constant $2(\sqrt{2} + 2\epsilon_0) < C < \epsilon_0^{-1}$. Since (26) is of order $O(1)$ on the event in (30), we have

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

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

**Case (ii).** Now assume that $\sigma_{0,ij}$ satisfies condition (A1) for some $i$. Note that (21) can be expressed as

$$
\frac{n}{2} \left\{ \tilde{\tau}_{1i}^2 + \frac{2b_0}{n} - \log \left( \tilde{\tau}_{1i}^2 + \frac{2b_0}{n} \right) - 1 \right\}
$$

$$
+ \left( \frac{n}{2} + a_0 \right) \log \left( \frac{\tilde{\tau}_{1i}^2 + \frac{2b_0}{n}}{\tilde{\tau}_{ij, \gamma}^2 + \frac{2b_0}{n}} \right)
$$

$$
- a_0 \log \left( \tilde{\tau}_{1i}^2 + \frac{2b_0}{n} \right) - b_0.
$$

(31)

(32)

(33)

We will show that for given constants $C_1 > 0$ and $C_2 > 2\sqrt{\alpha + 2}$,

$$
P_0 \left( \left| \tilde{\tau}_{1i}^2 + \frac{2b_0}{n} - 1 \right| \geq C_2 \sqrt{\frac{\log(n \lor p)}{n}} \right) \geq 1 - 2(n \lor p)^{-C_1}.
$$

(34)

On this event, we can show that (31) is larger than $4^{-1}C_2^2(1 - 2C_2\epsilon_0/3) \log(n \lor p)$ 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 (32) is positive and (33) is negligible compared to (31). Then, by (22) and (23),

$$
\log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq -\frac{\alpha}{2} \log(n \lor p) - \log n + 4^{-1}C_2^2(1 - 2C_2\epsilon_0/3) \log(n \lor p) + C
$$

$$
\geq 8^{-1} \left( C_2^2 - 4\alpha - 8 \right) \log(n \lor p) + C
$$

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

19
Now, we only need to show (34). By Lemma 1 in Laurent and Massart (2000), one can show that
\[
\mathbb{P}_0 \left( \sigma_{0,ii} - 1 + \frac{2b_0}{n} - 2\sigma_{0,ii} \sqrt{\frac{C_1 \log(n \vee p)}{n}} \leq \frac{\tau_i^2 + 2b_0}{n} - 1 \right) \leq \sigma_{0,ii} - 1 + \frac{2b_0}{n} + 4\sigma_{0,ii} \sqrt{\frac{C_1 \log(n \vee p)}{n}} \geq 1 - (n \vee p)^{-2C_1},
\]
because \(n\tau_i^2 / \sigma_{0,ii} \sim \chi_n^2\). Thus, it suffices to prove
\[
\sigma_{0,ii} - 1 + \frac{2b_0}{n} - 2\sigma_{0,ii} \sqrt{\frac{C_1 \log(n \vee p)}{n}} \geq C_2 \sqrt{\frac{\log(n \vee p)}{n}}
\]
or
\[
\sigma_{0,ii} - 1 + \frac{2b_0}{n} + 4\sigma_{0,ii} \sqrt{\frac{C_1 \log(n \vee p)}{n}} \leq -C_2 \sqrt{\frac{\log(n \vee p)}{n}},
\]
which is satisfied by (A1).

**Case (iii).** If \(\sigma_{0,ii}\) satisfies (A1) for some \(i\), 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 \(C_2 > 2\sqrt{\alpha + 2}\),
\[
\mathbb{P}_0 \left( \left| \tau_{ij,\gamma}^2 + \frac{2b_0}{n} - 1 \right| \geq C_2 \sqrt{\frac{\log(n \vee p)}{n}} \right) \geq 1 - 4(n \vee p)^{-C_1},
\]
which gives the desired result by (24)–(26). Note that we have \(n\tau_{ij,\gamma}^2 / \tau_{0,ij}^2 \sim \chi_n^2 - 1\). Then, similar to (29),
\[
\mathbb{P}_0 \left( \left| \tau_{ij,\gamma}^2 - \frac{n - 1}{n} \tau_{0,ij}^2 \right| \leq 4\tau_{0,ij}^2 \sqrt{\frac{C_1 \log(n \vee p)}{n}} \right) \geq 1 - 4(n \vee p)^{-C_1}
\]
by (27). To prove (35), we only need to show that
\[
\left(1 - \frac{1}{n}\right) \tau_{0,ij}^2 + \frac{2b_0}{n} - 1 - 4\tau_{0,ij}^2 \sqrt{\frac{C_1 \log(n \vee p)}{n}} \geq C_2 \sqrt{\frac{\log(n \vee p)}{n}},
\]
when \(\tau_{0,ij}^2 > 1\), and
\[
\left(1 - \frac{1}{n}\right) \tau_{0,ij}^2 + \frac{2b_0}{n} - 1 + 4\tau_{0,ij}^2 \sqrt{\frac{C_1 \log(n \vee p)}{n}} \leq -C_2 \sqrt{\frac{\log(n \vee p)}{n}},
\]
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),
\[
\log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq 8^{-1} \left(C_2^2 - 4\alpha - 8\right) \log(n \vee p) + C
\]
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 \( n(\tilde{\tau}_{ij}^2 - \tilde{\tau}_{ij}^2)/\tilde{\tau}_{ij}^2 \sim \chi_1^2(\lambda_{ij}) \) given \( \tilde{X}_j \), where \( \lambda_{ij} = \| \tilde{X}_j \|^2 a_{0,ij}^2 / \tilde{\tau}_{ij}^2 \). Note that for a random variable \( X \sim \chi_1^2(\lambda) \),

\[
P(X \geq k + \lambda - 2\sqrt{(k + 2\lambda)x}) \geq 1 - e^{-x}
\]

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

\[
1 - (n \vee p)^{-C_1} \leq P_0 \left( \frac{n}{2} (\tilde{\tau}_{ij}^2 - \tilde{\tau}_{ij}^2) \geq \frac{\tau_{0,ij}^2}{2} \left( 1 + \lambda_{ij} - 2\sqrt{(1 + 2\lambda_{ij})C_1 \log(n \vee p)} \right) \mid \tilde{X}_j \right) \leq P_0 \left( \frac{n}{2} (\tilde{\tau}_{ij}^2 - \tilde{\tau}_{ij}^2) \geq \frac{\tau_{0,ij}^2}{2} \lambda_{ij} \left( 1 - 2\sqrt{(1 + 2\lambda_{ij})C_1 \log(n \vee p)} \right) \mid \tilde{X}_j \right).
\]

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

\[
P_0 \left( \frac{\tau_{0,ij}^2}{a_{0,ij}^2 \sigma_{0,ij}} \lambda_{ij} \geq n \left( 1 - 2\sqrt{\frac{C_1 \log(n \vee p)}{n}} \right) \right) \geq 1 - (n \vee p)^{-C_1}
\]  

(36)

by Lemma 1 in Laurent and Massart (2000). On the event in (36), we have

\[
\lambda_{ij} \geq \frac{a_{0,ij}^2 \sigma_{0,ij}}{\tau_{0,ij}^2} n \left( 1 - 2\sqrt{\frac{C_1 \log(n \vee p)}{n}} \right) \geq \frac{\sigma_{0,ij}^2}{\sigma_{0,ij}^2 \tau_{0,ij}^2} \left( 1 - 2\sqrt{\frac{C_1 \log(n \vee p)}{n}} \right) \geq \frac{9C_1}{(1 - C_3)^2 \log(n \vee p)}
\]

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

\[
1 - 2\sqrt{(1 + 2\lambda_{ij})C_1 \log(n \vee p)} \geq C_3
\]  

(37)

for a given constant \( 0 < C_3 < 1 \). Again by (36) and condition (A3),

\[
\frac{\tau_{0,ij}^2}{2} \lambda_{ij} \left( 1 - 2\sqrt{(1 + 2\lambda_{ij})C_1 \log(n \vee p)} \right) \geq \frac{\tau_{0,ij}^2}{2} \lambda_{ij} C_3 \geq \frac{1}{2} \sigma_{0,ij}^2 C_3 n \left( 1 - 2\sqrt{\frac{C_1 \log(n \vee p)}{n}} \right) \geq \frac{1}{2} C_4 (\alpha + 2) \log(n \vee p)
\]

with probability at least \( 1 - (n \vee p)^{-C_1} \) for a given constant \( C_4 > 1 \) and all large \( n \). Note that (24) is positive and (26) is negligible compared to (24). Thus, by similar arguments used in Case (ii),

\[
\log B_{10}(\tilde{X}_i, \tilde{X}_j) \geq 2^{-1}(C_4 - 1)(\alpha - 2) \log(n \vee p) + C
\]

with probability at least \( 1 - 2(n \vee p)^{-C''} \) for some constants \( C, C'' > 0 \) and all large \( n \).
Proof of Theorem 3.2

Proof For a given constant $C_*$, define a parameter class

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

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

$$\sup_\nu \{ E_{I_p} \phi + E_{\Sigma_\nu} (1 - \phi) \} \geq \inf_\phi \frac{1}{p} \sum_\nu \{ E_{I_p} \phi + E_{\Sigma_\nu} (1 - \phi) \} \geq \inf_\phi \{ E_{I_p} \phi + E_{mix} (1 - \phi) \} \geq \int (f_{I_p} \land f_{mix}) = 1 - \frac{1}{2} \int |f_{mix} - f_{I_p}|,$$

where $f_{mix}$ and $f_\Sigma$ are density functions of $P_{mix}$ and $P_\Sigma$, respectively. Also note that

$$\left( \int |f_{mix} - f_{I_p}| \right)^2 \leq \int |f_{mix} - f_{I_p}|^2 \leq \int \left| \frac{f_{mix}}{f_{I_p}} - 1 \right|^2 \leq \int \frac{f_{mix}^2}{f_{I_p}} - 1. \quad (38)$$

Thus, for any test $\phi$,

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

by the above arguments. Now we only need to deal with the upper bound of (38). An upper bound of $\int (f_{mix}^2 / f_{I_p})$ can be derived as follows:

$$\int \frac{f_{mix}^2}{f_{I_p}} = \int \frac{1}{f_{I_p}} \left( \frac{1}{p} \sum_\nu f_{\Sigma_\nu} \right)^2$$

$$= \frac{1}{p^2} \sum_{\nu_1, \nu_2} \int \frac{f_{\Sigma_{\nu_1}} f_{\Sigma_{\nu_2}}}{f_{I_p}}$$

$$= \frac{1}{p^2} \sum_{\nu_1, \nu_2} \left\{ \det \left( I_p - (\Sigma_{\nu_1} - I_p)(\Sigma_{\nu_2} - I_p) \right) \right\}^{-\frac{n}{2}}$$

$$= \frac{1}{p^2} \left\{ p^2 - p + p \left( 1 - C^2_* \frac{\log p}{n} \right) \right\}^{-\frac{n}{2}}$$

$$\leq 1 - \frac{1}{p} + \frac{1}{p} \frac{C^2_*}{2} \leq 2 - \frac{1}{p}$$

22
because $C^2_s \leq 2$. The third equality follows from Lemma B.3 of Lee and Lee (2018). It gives the upper bound

$$\inf_{\Sigma \in H_1(C_s)} \mathbb{E}_{\Sigma \phi} \leq \mathbb{E}_{\Sigma_0} \phi + \frac{1}{2} \{2 - p^{-1}\}^{\frac{3}{2}},$$

which completes the proof.

**Proof of Theorem 3.3**

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

$$\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) - \left( \frac{n}{2} + a_0 \right) \log \left( \frac{n\hat{\tau}_{ij,\gamma}^2 + 2b_0}{n\hat{\tau}_i^2 + 2b_0} \right)$$

$$\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \left( \frac{n}{2} + a_0 \right) \cdot \frac{n(\hat{\tau}_{ij,\gamma}^2 - \hat{\tau}_{ij}^2)}{n\hat{\tau}_{ij,\gamma}^2 + 2b_0}$$

$$\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \left( \frac{n}{2} + a_0 \right) \cdot \frac{n(\hat{\tau}_{ij,\gamma}^2 - \hat{\tau}_{ij}^2)}{n\hat{\tau}_i^2 + 2b_0}$$

where the first inequality holds because $\log(1 + x) \leq x$ for all $x$. Note that $n\hat{\tau}_{ij}^2/\sigma_{0,ij} \sim \chi^2_{n-1}$ and $n(\hat{\tau}_{ij,\gamma}^2 - \hat{\tau}_{ij}^2)/\sigma_{0,ij} \sim \chi^2_1$. By Lemma 1 in Laurent and Massart (2000),

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

and

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

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

$$\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) \leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \left( \frac{n}{2} + a_0 \right) \frac{1 + 2C_4 \log(n \vee p)}{n - 1 - 2\sqrt{C(n-1) \log(n \vee p)} + 2b_0 \sigma_{0,ij}^{-1}}$$

$$\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \left( \frac{n}{2} + a_0 \right) \frac{n + 2C_4 n \log(n \vee p)}{n - 1 - 2\sqrt{C(n-1) \log(n \vee p)}}$$

$$\leq \frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} \right) + \frac{1}{2} \cdot C'' \log(n \vee p) + C''$$

for any constants $C'' > 2C_4/(1 - 3\varepsilon_0\sqrt{C})$, $C'' > 0$ with probability at least $1 - 6(n \vee p)^{-C}$ for some constant $C > 2$ and large $n$. Since $\gamma = c_\gamma (n \vee p)^{-\alpha}$ and $\alpha > 4/(1 - 3\varepsilon_0\sqrt{2})$, 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 $n \to \infty$ on the above event. By the similar arguments used in the proof of Theorem 3.1, it implies the consistency of the mxPBF under $H_0$. 

23
Now, suppose the alternative hypothesis is true. Without loss of generality, assume that \( \sigma_{0,ij} \) satisfies condition (15). Then,

\[
-\log \tilde{B}_{10}(\tilde{X}_i, \tilde{X}_j) = -\frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} - \left( \frac{n}{2} + a_0 \right) \log \left( 1 - \frac{n(\tilde{\tau}_{ij,\gamma}^2 - \tilde{\tau}_i^2)}{n\tilde{\tau}_{ij,\gamma}^2 + 2b_0} \right) \right)
\leq -\frac{1}{2} \log \left( \frac{\gamma}{1 + \gamma} + \left( \frac{n}{2} + a_0 \right) \frac{n(\tilde{\tau}_{ij,\gamma}^2 - \tilde{\tau}_i^2)}{n\tilde{\tau}_{ij,\gamma}^2 + 2b_0} \right)
\]

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

\[
\left( \frac{n}{2} + a_0 \right) \frac{1}{n\tilde{\tau}_i^2 + 2b_0} \geq \frac{1}{2} \left( \frac{\tilde{\tau}_i^2 + 2b_0}{n} \right)^{-1}
\geq \frac{1}{2} \left[ \sigma_{0,ii} \left( 1 + 4\sqrt{\frac{C_1 \log(n \lor p)}{n}} \right) + \frac{2b_0}{n} \right]^{-1}
\geq \frac{1}{2} \left[ \sigma_{0,ii}(1 + 4\epsilon_0\sqrt{C_1}) + \frac{2b_0}{n} \right]^{-1}
\]

and

\[
n(\tilde{\tau}_{ij,\gamma}^2 - \tilde{\tau}_i^2) = -\frac{1}{1 + \gamma} n(\tilde{\tau}_{ij}^2 - \tilde{\tau}_{ij}^2)
\leq -\frac{C_3}{1 + \gamma} \tilde{\tau}_{0,ij}^2 \lambda_{ij}
\leq -\frac{C_3}{1 + \gamma} \sigma_{0,ij}^2 \|\tilde{X}_j\|^2 / \sigma_{0,ij}^2
\leq -\frac{C_3}{1 + \gamma} \sigma_{0,ij}^2 (1 - 2\epsilon_0\sqrt{C_1})
\]

with probability at least \(1 - C(n \lor p)^{-c}\) for some constants \(c > 0\) and \(C > 0\) and all large \(n\), by condition (15). The fourth inequality holds by condition (A4) and the similar arguments used in (37). Thus, on this event,

\[
\left( \frac{n}{2} + a_0 \right) \frac{n(\tilde{\tau}_{ij,\gamma}^2 - \tilde{\tau}_i^2)}{n\tilde{\tau}_i^2 + 2b_0} \leq -\frac{C_3}{1 + \gamma} \sigma_{0,ij}^2 n(1 - 2\epsilon_0\sqrt{C_1}) \times \frac{1}{2} \left[ \sigma_{0,ii}(1 + 4\epsilon_0\sqrt{C_1}) + \frac{2b_0}{n} \right]^{-1}
\leq -\frac{\alpha}{2} C_4 \log(n \lor p)
\]

which implies

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

for all sufficiently large \(n\). Since \(C_4 > 1\), it completes the proof.
Proof of Theorem 3.5

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

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

and $S_{ij}(\Sigma_0) = I(\sigma_{0, ij} \neq 0)$. By the proof of Theorem 3.3,

$$
\mathbb{P}_0\left( \hat{S}_{ij} = 1, S_{ij}(\Sigma_0) = 0 \right) = \mathbb{P}_0\left( \log \hat{B}_{pair, 10}(\hat{X}_i, \hat{X}_j) > C_{sel}, \sigma_{0, ij} = 0 \right) \\
\leq (n \lor p)^c
$$

for some constant $c > 2$ and all sufficiently large $n$. Now, assume that $\sigma_{0, ij}$ satisfies condition (A.ij). Note that condition (A.ij) is the same with 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

$$
\mathbb{P}_0\left( \hat{S}_{ij} = 0, S_{ij}(\Sigma_0) = 1 \right) = \mathbb{P}_0\left( \log \hat{B}_{pair, 10}(\hat{X}_i, \hat{X}_j) \leq C_{sel}, \sigma_{0, ij} \neq 0 \right) \\
\leq (n \lor p)^c
$$

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

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

as $n \rightarrow \infty$.

References

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation, *Journal of the American Statistical Association* **106**(494): 672–684.

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, *The Annals of Statistics* **39**(3): 1496–1525.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices, *Bernoulli* **19**(5B): 2359–2388.

Castillo, I., Schmidt-Hieber, J. and Van der Vaart, A. (2015). Bayesian linear regression with sparse priors, *The Annals of Statistics* **43**(5): 1986–2018.

Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices, *Journal of the American Statistical Association* **105**(490): 810–819.

25
Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs, *Statistical science* pp. 204–218.

Drton, M. and Perlman, M. D. (2004). Model selection for gaussian concentration graphs, *Biometrika* 91(3): 591–602.

Drton, M. and Perlman, M. D. (2007). Multiple testing and error control in gaussian graphical model selection, *Statistical Science* pp. 430–449.

Gan, L., Narisetty, N. N. and Liang, F. (2018). Bayesian Regularization for Graphical Models with Unequal Shrinkage, *ArXiv e-prints*.

Gorman, R. P. and Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets, *Neural networks* 1(1): 75–89.

Johnson, V. E. and Rossell, D. (2010). On the use of non-local prior densities in bayesian hypothesis tests, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(2): 143–170.

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings, *Journal of the American Statistical Association* 107(498): 649–660.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions, *J. Amer. Statist. Assoc.* 104(486): 682–693.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the american statistical association* 90(430): 773–795.

Keckić, J. D. and Vasić, P. M. (1971). Some inequalities for the gamma function, *Publications de l'Institut Mathématique* 11(31): 107–114.

Kolar, M. and Liu, H. (2012). Marginal regression for multitask learning., *AISTATS*, pp. 647–655.

Kundu, S., Baladandayuthapani, V. and Mallick, B. K. (2013). Bayes regularized graphical model estimation in high dimensions, *arXiv preprint arXiv:1308.3915*.

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, *Journal of Business & Economic Statistics* 33(1): 76–86.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection, *Annals of Statistics* 28(5): 1302–1338.

Lee, K. and Lee, J. (2018). Optimal Bayesian Minimax Rates for Unconstrained Large Covariance Matrices, *Bayesian Analysis*. Accepted.
Martin, R., Mess, R. and Walker, S. G. (2017). Empirical bayes posterior concentration in sparse high-dimensional linear models, *Bernoulli* 23(3): 1822–1847.

Ročková, V. and George, E. I. (2016). Fast bayesian factor analysis via automatic rotations to sparsity, *Journal of the American Statistical Association* 111(516): 1608–1622.

Ročková, V. et al. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior, *The Annals of Statistics* 46(1): 401–437.

Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices, *Journal of the American Statistical Association* 104(485): 177–186.

Srivastava, M. S., Yanagihara, H. and Kubokawa, T. (2014). Tests for covariance matrices in high dimension with less sample size, *Journal of Multivariate Analysis* 130: 289–309.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(1): 49–67.