The Canadian Journal of Statistics
Vol. 42, No. 1, 2014, Pages 61–77
La revue canadienne de statistique

An accurate test for the equality of covariance matrices from decomposable graphical Gaussian models

Yanyan WU,[1], Hélène MASSAM,[2]* and Augustine WONG[2]

[1] *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada*
[2] *Department of Mathematics and Statistics, York University, Toronto, Canada*

Key words and phrases: Adjusted likelihood ratio test statistic; Bartlett–Box M-statistic; cumulant generating function; decomposable graphical models; equality of covariance matrices; Lugannani and rice formula.

MSC 2010: Primary 62H15; secondary 62E20

Abstract: This paper derives a saddlepoint based approximation for the cumulative distribution function of the Bartlett–Box M-statistic that tests the equality of covariance matrices for several samples from graphical Gaussian models Markov with respect to a decomposable graph $\mathcal{G}$. The proposed saddlepoint-based method has third-order accuracy ($O(n^{-3/2})$). Simulation results show that the proposed method has extremely good coverage properties even when the sample size is small. We apply our method to the well-known Call Centre data set and show that the covariance matrix is not constant through time. *The Canadian Journal of Statistics* 42: 61–77; 2014 © 2014 Statistical Society of Canada

Résumé: Dans cet article, les auteurs obtiennent une approximation en point de selle de la fonction de répartition de la M-statistique de Bartlett-Box. Cette statistique sert à tester l'égalité des matrices de covariance de plusieurs échantillons issus de modèles graphiques gaussiens markoviens par rapport à un graphe décomposable $\mathcal{G}$. La méthode en point de selle proposée a une précision du troisième ordre ($O(n^{-3/2})$). Les résultats de simulations montrent que la méthode proposée offre de bons taux de couverture, même pour des échantillons de petite taille. Les auteurs appliquent leur méthode à un jeu de données bien connu à propos d'un centre d'appels et montrent que la matrice de covariance n'est pas constante dans le temps. *La revue canadienne de statistique* 42: 61–77; 2014 © 2014 Société statistique du Canada

1. INTRODUCTION

In this paper, we consider a classical multivariate analysis testing problem in the more general framework of graphical Gaussian models, that is, in the case where the Gaussian models considered are known to include certain given conditional independences represented by a graph. We will first outline the classical problem and its solutions and then describe the problem when dealing with graphical Gaussian models.

The classical problem is that of testing the equality of the covariance matrices Σ_i in q samples from multivariate normal $N_p(\mu_i, \Sigma_i)$ distributions. This test typically arises when performing a multivariate analysis of variance (abbreviated MANOVA) to determine whether the means μ_i, $i = 1, \ldots, q$ from the q multivariate normal samples are equal or not. If the covariance matrices Σ_i are assumed to be the same, then the standard MANOVA technique can be applied for testing for the equality of the μ_i. However if the Σ_i are not equal, the MANOVA problem becomes

* *Author to whom correspondence may be addressed.*
E-mail: massamh@yorku.ca

a Behrens–Fisher-type problem and solutions are approximate and complicated. It is therefore essential to test the equality of the $\Sigma_i, i = 1, \ldots, q$ before performing a MANOVA analysis.

Given q samples $(X_{i1}, \ldots, X_{in_i})$ from the normal $N(\mu_i, \Sigma_i), i = 1, \ldots, q$ distribution, it is well-known (see Muirhead, 1982, Theorem 8.2.1) that the likelihood ratio test statistic is

$$\Lambda^* = \frac{n^{\mathrm{d}n/2}}{\prod_{i=1}^q n_i^{\mathrm{d}n_i/2}} \cdot \frac{\prod_{i=1}^q |U_i|^{n_i/2}}{|U_{\cdot q}|^{n/2}},$$

where $U_i = \sum_{l=1}^n (X_{il} - \bar{X}_i)(X_{il} - \bar{X}_i)^t$, $\bar{X}_i = \sum_{l=1}^{n_i} X_{il}/n_i$, and $U_{\cdot q} = \sum_{i=1}^q U_i$. This quantity Λ^* is based on the maximum likelihood estimates obtained from the normal distributions of the X_{il}'s. It was shown by Das Gupta (1969) that the test based on Λ^* is biased while Perlman (1980) showed that the modified likelihood ratio statistic Λ is unbiased where

$$\Lambda = \frac{p^{\mathrm{d}p/2}}{\prod_{i=1}^q p_i^{\mathrm{d}p_i/2}} \cdot \frac{\prod_{i=1}^q |U_i|^{p_i/2}}{|U_{\cdot q}|^{p/2}} \tag{1}$$

with $p_i = n_i - 1$, $i = 1, \ldots, q$, $p = \sum_{i=1}^q p_i = p - q$. The statistic Λ is the modified likelihood ratio statistic suggested by Bartlett (1937) and $M = -2 \log \Lambda$ is what, following Booth et al. (1995), we call the Bartlett–Box M-statistic. The distribution of the likelihood ratio test statistic cannot be given explicitly and various asymptotic approximations to this distribution have been proposed. The traditional χ^2 approximation to the distribution of Λ is very poor. Booth et al. (1995) have given accurate approximations to the cumulative distribution function of the Bartlett–Box M-statistic using the Lugannani & Rice (1980) and Skovgaard (1987) approximations. The reader is referred to Chapter 11 of Butler (2007) for a review of MANOVA testing using the saddlepoint approximation.

Our aim is to give a Lugannani–Rice-type approximation to the distribution of M when the Gaussian models $N(0, \Sigma_i), i = 1, \ldots, q$ are known to include certain given fixed conditional independences. Such models are indeed very important: nowadays, in many fields such as finance, marketing or genomics, data are high-dimensional, and often, the number of data points is relatively low compared to the dimension of the data. The sample covariance matrix is no longer a good estimate of the population covariance matrix and it is essential to reduce the number of parameters in the model. For Gaussian models, one way to do so is to detect the conditional independences between variables: the conditional independence of X_i and X_j given the other variables in the random vector $X \in R^p$ is, of course, equivalent to a zero (i, j) entry in the inverse covariance matrix. Under these conditional independences, the dimension of the parameter space is then reduced significantly. We can use a graph to represent such conditional independences in the following way. Let $\mathcal{G} = (V, E)$ be an undirected graph where $V = \{1, \ldots, p\}$ is the set of vertices and E is the set of undirected edges. Then we will have that X_i is conditionally independent of X_j given all the other components of X if the edge (i, j) does not belong to E. Multivariate normal models with conditional independences represented by a graph are called graphical Gaussian models and the models are said to be Markov with respect to the graph $\mathcal{G}$. Graphical Gaussian models now belong to the main toolkit of any applied statisticians. They actually date back more than a century but in their more modern form graphical models were first introduced by Dempster (1972) under the name of covariance selection models. They are nowadays widely used in machine learning and statistics. For a classical treatment of graphical Gaussian models, the reader is referred to Lauritzen (1996). There are numerous other more recent books on the topic, both in statistics and in machine learning (see, e.g., Koller & Friedman, 2009).

Let $K = \Sigma^{-1}$ denote the inverse covariance or precision matrix. As mentioned above, in Gaussian models, the conditional independence of X_i and X_j given the other variables is

equivalent to $K_{ij} = 0$. The parameter space of a graphical Gaussian model, let us say with mean 0, is therefore equal to the cone of positive definite matrices with certain entries fixed and equal to 0. In Section 2 below, we will study this cone and call it $P_{\mathcal{G}}$.

In recent years, a considerable amount of work has been devoted to the detection, estimation and testing of such precision matrices (e.g., Dawid & Lauritzen, 1993; Letac & Massam, 2007; Bickel & Levina, 2008; Rajaratnam, Massam, & Carvalho, 2009, and many other references given within these papers). Both Bickel & Levina (2008) and Rajaratnam, Massam, & Carvalho (2009) study the example of the Call Centre data set, which consists of the number of calls received in a call centre, every day over the course of several months. This data set is subjected to a standard transformation to make it normal. Through different methods, both papers strive first to identify the conditional independences in the variables, that is, the correct graph underlying the model and to estimate it, assuming the the covariance matrix of the normalized data is constant throughout the months. It is important to make sure this assumption is correct since any subsequent inference depends on it. Using the new theoretical results in this paper, we will test the hypothesis that for this data set, the covariance matrix remains constant throughout the months. We will see that there is strong evidence that this is not so. The class of undirected graphs considered in both papers is the class of decomposable graphs. Decomposable graphs are undirected graphs such that any cycle of length greater than or equal to 4 has a chord: their properties will be recalled in Section 2 below. Decomposable graphs have many nice properties and can be used as an approximation to general undirected graphs in certain algorithms. We will therefore work here with this class of graphs.

In this paper, we therefore consider the problem of equality of the covariance matrices in q samples from graphical Gaussian models Markov with respect to a decomposable graph $\mathcal{G}$. We are given q independent samples from multivariate Gaussian $N(0, \Sigma_i)$ distributions that are known to be Markov with respect to a given decomposable graph $\mathcal{G}$. Without loss of generality, we can also assume that the Gaussian models are centered. We want to test whether the q covariance matrices $\Sigma_i,\ i = 1, \ldots, q$ of these distributions are equal, that is, we want to test

$$H_0 : \Sigma_1 = \cdots = \Sigma_q = \Sigma, \tag{2}$$

where the Σ_i are such that the corresponding distributions are Markov with respect to $\mathcal{G}$ or equivalently such that $K = \Sigma^{-1}$ belongs to the cone $P_{\mathcal{G}}$ of positive definite matrices with fixed zeros, defined below in Equation (3).

In the particular case where there are no conditional independencies between the variables the model is said to be saturated. We can still represent it by a graph $\mathcal{G}$ where all vertices are linked to all other vertices. The graph $\mathcal{G}$ is then said to be complete and testing problem (Eq. 2) becomes the classical problem of testing the equality of covariance matrices described above. This paper is therefore a generalization of Booth et al. (1995) to the case where the Gaussian distributions $N(0, \Sigma_i)$, $i = 1, \ldots, q$ are Markov with respect to the decomposable graph $\mathcal{G}$. Our aim is to derive first the expression of Λ in this case and then an accurate Lugannani–Rice-type approximation to the distribution of M.

In the saturated case considered by Booth et al. (1995), the likelihood ratio statistic is a function of the maximum likelihood estimates (abbreviated mle) of Σ under the null and alternative hypothesis. Under the null, this mle is proportional to the sample covariance matrix which follows the well-known Wishart distribution. Using the properties of the Wishart, it is then relatively easy to compute the moments of the likelihood ratio statistic and from there, an approximation to its distribution under the null hypothesis.

In the decomposable graphical Gaussian model case, as we shall see in the sequel, the covariance parameter of the normal distribution Markov with respect to $\mathcal{G}$ is no longer the full matrix Σ but its projection on the subspace of incomplete matrices with only entries the Σ_{ij} such

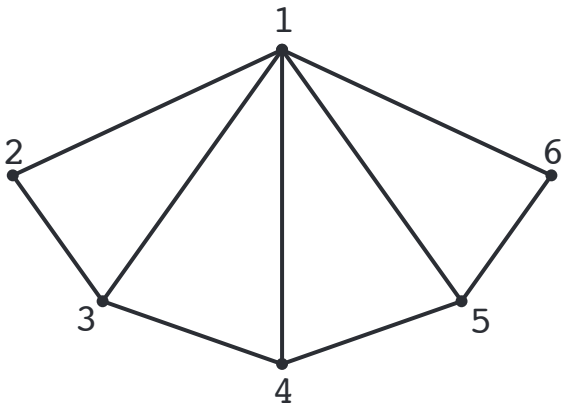

Figure 1: Decomposable graph with 4 cliques $C_1 = \{1, 2, 3\}$, $C_2 = \{1, 3, 4\}$, $C_3 = \{1, 4, 5\}$, and $C_4 = \{1, 5, 6\}$.

that $(i, j) \in E$. The distribution of the mle of the covariance parameter is no longer the Wishart distribution but rather the hyper Wishart distribution first identified by Dawid & Lauritzen (1993) and analysed further by Letac & Massam (2007). We will show that, as in the saturated case, we can compute the hth moment of the likelihood ratio statistic and subsequently derive an accurate asymptotic approximation to the distribution of the Bartlett–Box M-statistic using the Lugannani and Rice (henceforth abbreviated LR) formula.

We will see in Section 3 that in the decomposable case, the determinant $|U_i|$ is replaced by the product of determinants $|(U_i)_{C_1}| \prod_{j=2}^{k} |(U_i)_{R_j \cdot}|$ following a perfect ordering $C_1, \ldots, C_k$ of the cliques of $\mathcal{G}$ (see Eq. 4 below for the definition of $(U_i)_{R_j \cdot}$), and similarly for $|U_{\cdot q}|$. What will allow us to compute the hth moment of M and the LR approximation are the facts that the hyper Wishart distribution also belongs to a natural exponential family and that the components $(U_i)_{C_1}$, $(U_i)_{R_j \cdot}$, $j = 2, \ldots, k$ are statistically independent, following Wishart distributions of smaller dimensions. In the next section, we will recall some notions of graphical models needed in the sequel of the paper. In Section 3, we will derive the Bartlett–Box M-statistic $M = -2 \log \Lambda$ and the LR approximation to its distribution. In Section 4, we will illustrate through several examples the accuracy of this approximation and compare its performance with the standard asymptotic χ^2 approximation of order $O(n^{-1})$, the Box (1949) approximation (see Muirhead, 1982, Section 8.2.4) of order $O(n^{-2})$ as well as with Monte Carlo simulation. We will see that the LR approximation is very fast and very accurate even for small sample sizes . It always does better than the two standard methods for accuracy and is much faster than Monte Carlo simulation.

2. PRELIMINARIES

Let us first recall some basic notions on decomposable graphs and graphical Gaussian models. Let $\mathcal{G} = \{V, E\}$ be an undirected graph with vertex set $V = \{1, \ldots, d\}$ and edge set E. Vertices i and j are said to be neighbours in $\mathcal{G}$ if $(i, j) \in E$. A clique C is a subset of V such that in $\mathcal{G}$, every vertex of C is linked to any other vertex of C by an edge. A maximal clique is a clique which is maximal with respect to inclusion. In the sequel, we will use the term clique for a maximal clique. In this paper, we assume that $\mathcal{G}$ is decomposable. This implies in particular that we can always find a perfect ordering $C_1, \ldots, C_k$ of the cliques of $\mathcal{G}$ such that, for any $j = 2, \ldots, k$ there exists $j_0 < j$ such that $C_j \cap (\cup_{l=1}^{j-1} C_l) \subset C_{j_0}$. To illustrate the notion of a perfect ordering of the cliques, let us consider Figure 1. The cliques are C_i, $i = 1, 2, 3, 4$ as indicated in the caption.

 DOI: 10.1002/cjs

As one can readily verify, the order C_1, C_2, C_3, C_4 is perfect while the order C_1, C_2, C_4, C_3 is not: indeed $C_2 \cap C_1 = \{1, 3\} \subset C_1$, $C_4 \cap (C_1 \cup C_2) = \{1\} \subset C_2$ but $C_3 \cap (C_1 \cup C_2 \cup C_4) = \{1, 4, 5\}$, which cannot be included entirely in anyone of either C_1, C_2, or C_4.

Let X be a random vector in R^d and let us represent the components of X by the vertices of $\mathcal{G}$. The random vector X is said to follow a normal distribution Markov with respect to $\mathcal{G}$ if for any edge (i, j) not in E, the ith and the jth variables, X_i and X_j respectively, are conditionally independent given all the other variables $X_{V\setminus\{i,j\}}$. Let Σ be the covariance matrix of the distribution of X. It is a classical result (see Lauritzen, 1996, Proposition 5.2) that X_i is independent of X_j given $X_{V\setminus\{i,j\}}$ if and only if K_{ij}, the (i, j) entry of $K = \Sigma^{-1}$ is equal to 0. A Gaussian model is therefore Markov with respect to $\mathcal{G}$ if and only if $K = \Sigma^{-1}$ belongs to the cone $P_{\mathcal{G}}$, defined in Equation (3) below. Let M be the linear space of symmetric matrices of order d, M_d the space of symmetric matrices of order d and $M_d^+ \subset M$ the cone of positive definite (abbreviated > 0) matrices. We define the cone

$$P_{\mathcal{G}} = \{y \in M_d^+ | y_{ij} = 0, (i, j) \notin E\}. \tag{3}$$

Without loss of generality, we consider centered Gaussian distributions and we define the family of distributions

$$\mathcal{N}_{\mathcal{G}} = \{N(0, \Sigma), \quad \Sigma^{-1} \in P_{\mathcal{G}}\},$$

to be the Gaussian model Markov with respect to $\mathcal{G}$. Clearly the number of parameters in this model is equal to $d(d + 1)/2 - z$ where z is the number of missing edges in $\mathcal{G}$. Therefore, the number of free entries in Σ is also equal to $d(d + 1)/2 - z$. One can verify that the entries Σ_{ij}, $(i, j) \notin E$ are functions of $(\Sigma_{ij}, (i, j) \in E)$ and are therefore not free parameters of the Gaussian graphical model. Let $I_{\mathcal{G}}$ be the linear space of symmetric incomplete matrices x with missing entries x_{ij}, $(i, j) \notin E$, and $\pi : M_d \mapsto I_{\mathcal{G}}$ be the projection of M into $I_{\mathcal{G}}$ so that if $y \in M_d$, then $\pi(y)$ can be thought of as y with the y_{ij} entries discarded whenever $(i, j) \notin E$. The parameter space of the model $\mathcal{N}_{\mathcal{G}}$ can be described as

$$Q_{\mathcal{G}} = \{\pi(x) | \ x \in M_d^+ \text{ and } x^{-1} \in P_{\mathcal{G}}\}.$$

Equivalently, we can say that the parameter of a normal distribution in $\mathcal{N}_{\mathcal{G}}$ is

$$(\Sigma_{C_i}, \ i = 1, \ldots, k),$$

where $\Sigma_{C_i} \in M_{c_i}^+$ and we write $c_i = |C_i|$ for the cardinality of C_i, Σ_A for the $A \times A$ submatrix of Σ and $\Sigma_{A,B}$ for the rectangular $A \times B$ submatrix of Σ. Given a perfect ordering of the cliques of the decomposable graph $\mathcal{G}$, we define

$$H_j = \cup_{i \le j} C_i, \quad S_j = C_j \cap H_{j-1}, \quad R_j = C_j \setminus H_{j-1}, \ \ j = 2, \ldots, k \ .$$

For example in Figure 1, with the order C_1, C_2, C_3, C_4 of the cliques, $H_2 = \{1, 2, 3, 4\}$, while $S_2 = \{1, 3\}$, $R_2 = \{4\}$. Similarly, $H_3 = \{1, 2, 3, 4, 5\}$, $S_3 = \{1, 4\}$, $R_3 = \{5\}$ and $H_4 = \{1, 2, 3, 4, 5, 6\}$, $S_4 = \{1, 5\}$, $R_4 = \{6\}$.

The non-free entries of Σ are the $\Sigma_{R_j, H_{j-1}\setminus S_j}$ which are equal to

$$\Sigma_{R_j, H_{j-1}\setminus S_j} = \Sigma_{R_j, S_j} \Sigma_{S_j}^{-1} \Sigma_{S_j, H_{j-1}\setminus S_j} \ .$$

In the hypothesis testing problem (Eq. 2), it is therefore understood that even though the hypothesis is expressed in terms of the full Σ_i, the likelihood ratio statistic will be expressed in terms of the parameter $\pi(\Sigma)$ or equivalently $(\Sigma_{C_i}, \ i = 1, \ldots .k)$.

In the sequel we shall use the notation

$$\Sigma_{R_j\cdot} = \Sigma_{R_j} - \Sigma_{R_j,S_j}\Sigma_{S_j}^{-1}\Sigma_{S_j,R_j} \tag{4}$$

for the conditional covariance of the jth residual and the notation $n_i, i = 1, \ldots, q$ for the size of the sample from the decomposable graphical Gaussian model $N_d(0, \Sigma_i)$, $\Sigma_i^{-1} \in P_{\mathcal{G}}$. We also write

$$\Sigma_{\cdot q} = \sum_{i=1}^{q} \Sigma_i, \qquad n_{\cdot q} = \sum_{i=1}^{q} n_i, \qquad U_{\cdot q} = \sum_{i=1}^{q} U_i$$

and

$$\Sigma_{\cdot qC_j} = \sum_{i=1}^{q} (\Sigma_i)_{C_j}, \qquad \Sigma_{\cdot qR_j} = \sum_{i=1}^{q} (\Sigma_i)_{R_j}, \qquad \Sigma_{\cdot qS_j} = \sum_{i=1}^{q} (\Sigma_i)_{S_j},$$

$$\Sigma_{\cdot qR_j\cdot} = (\Sigma_{\cdot q})_{(R_j\cdot S_j)} = (\Sigma_{\cdot q})_{R_j} - (\Sigma_{\cdot q})_{(R_j,S_j)}(\Sigma_{\cdot q})_{S_j}^{-1}(\Sigma_{\cdot q})_{(S_j,R_j)},$$

and use similar notation for the U_i's. Moreover, we denote the inner product $tr(ab')$ of a and b in the space of $d \times d$ matrices by $\langle a, b\rangle$. This implies that, for a and b symmetric, $\langle a, b\rangle = tr(ab)$.

In the following section , we first derive the Bartlett–Box M-statistic and obtain its cumulant generating function. Then a saddlepoint-based approximation to its cumulative distribution is obtained.

3. THE BARTLETT–BOX M-STATISTIC

We will now derive the Bartlett–Box M-statistic for testing the hypothesis stated in Equation (2). As in the saturated case, it is distributed like the product of multivariate Beta distributions and therefore we cannot obtain its density directly. However, we can compute its moments and its cumulant generating function. This will allow us to compute the Box (1949) approximation and the LR approximation to its cumulative distribution function.

3.1. The Modified Likelihood Ratio Statistic Λ

We consider q samples $X_{i1} \ldots, X_{in_i}$, $i = 1, \ldots, q$ coming respectively from the d-dimensional Gaussian $N_d(0, \Sigma_i)$, $i = 1, \ldots, q$ distribution with $\Sigma_i^{-1} \in P_{\mathcal{G}}$. We want to test Equation (2) against the alternative H_a which is "not H_0." As mentioned above, in the saturated case, that is when $\mathcal{G}$ is complete, we know from Theorem 8.2.5 of Muirhead (1982) that the test using the modified likelihood ratio statistic (Eq. 1) is unbiased.

We have a parallel test statistic in the decomposable case. A sketch of its derivation is given here. Further details can be found in the Supplementary Material. In the decomposable case, as we recall from Section 2, the covariance parameter is $\pi(\Sigma)$ or equivalently $(\Sigma_{C_j}, j = 1, \ldots, k)$. From Proposition 5.9 of Lauritzen (1996), the mle of $\pi(\Sigma_i)$, $i = 1, \ldots, q$, obtained from the distribution of X_{il}, under H_a, has a completion $\hat{\Sigma}_i^a$ such that $\hat{\Sigma}_i^a$ and its inverse $\hat{K}_i{}^a = (\hat{\Sigma}_i^a)^{-1}$ satisfy

$$\hat{\Sigma}_{C_j}^a = (U_i)_{C_j}/n_i, \quad j = 1, \ldots, k, \tag{5}$$

$$\hat{K}_i^a = n_i \left[\sum_{j=1}^{k} (U_i)_{C_j}^{-1} - \sum_{j=2}^{k} (U_i)_{S_j}^{-1} \right], \tag{6}$$

 DOI: 10.1002/cjs

$$|\hat{K}_i^q|^{-1} = n_i^{-d}\,\frac{\prod_{j=1}^k |(U_i)_{C_j}|}{\prod_{j=2}^k |(U_i)_{S_j}|} = n_i^{-d}\,|(U_i)_{C_1}| \prod_{j=2}^k |(U_i)_{R_j\cdot}|\,. \tag{7}$$

Similarly, under H_0, $\hat{\Sigma}^0$, the completion of the mle of $\pi(\Sigma)$, and its inverse $\hat{K}^0$, satisfy

$$\hat{\Sigma}^0_{C_j} = \frac{(U_{\cdot q})_{C_j}}{n_{\cdot q}},\ j = 1, \ldots, k, \tag{8}$$

$$\hat{K}_i^0 = n_{\cdot q}\left[\sum_{j=1}^k (U_{\cdot q})^{-1}_{C_j} - \sum_{j=2}^k (U_{\cdot q})^{-1}_{S_j}\right], \tag{9}$$

$$|\hat{K}_i^0|^{-1} = n_{\cdot q}^{-d}\,\frac{\prod_{j=1}^k |(U_{\cdot q})_{C_j}|}{\prod_{j=2}^k |(U_{\cdot q})_{S_j}|} = n_{\cdot q}^{-d}\,|(U_{\cdot q})_{C_1}| \prod_{j=2}^k |(U_{\cdot q})_{R_j\cdot}|\,. \tag{10}$$

The mle of $\pi(\Sigma_i),\ i = 1, \ldots, q$ and of $\pi(\Sigma)$ follow hyper Wishart distributions on $Q_{\mathcal{G}}$ with densities of the general form

$$\begin{aligned} W_{Q_{\mathcal{G}}}(p, \Sigma; x)\,\mathrm{d}x &= \frac{\prod_{j=1}^k w_{c_j}(p, \Sigma_{C_j}; x_{C_j})}{\prod_{j=2}^k w_{s_j}(p, \Sigma_{S_j}; x_{S_j})} 1_{Q_{\mathcal{G}}}(x)\,\mathrm{d}x \\ &= \frac{\prod_{j=2}^k \Gamma_{s_j}(p/2)|\Sigma_{S_j}|^{p/2}}{\prod_{j=1}^k \Gamma_{c_j}(p/2)|\Sigma_{C_j}|^{p/2}} \frac{\prod_{j=1}^k |x_{C_j}|^{p/2-(c_j+1)/2}}{\prod_{j=2}^k |x_{S_j}|^{p/2-(c_j+1)/2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2}\left\{\sum_{j=1}^k \langle \Sigma^{-1}_{C_j}, x_{C_j}\rangle - \sum_{j=2}^k \langle \Sigma^{-1}_{S_j}, x_{S_j}\rangle\right\}\right\} 1_{Q_{\mathcal{G}}}(x)\,\mathrm{d}x \end{aligned} \tag{11}$$

where $p = p_i,\ i = 1, \ldots, q,\ p = p_{\cdot q}$ respectively, $\Gamma_{c_i}(p) = \prod_{j=1}^{c_i} \Gamma(p - (j-1)/2)$ and similarly for $\Gamma_{s_i}(p)$. From these distributions, we can derive the modified likelihood ratio statistic

$$\Lambda = \frac{p_{\cdot q}^{\mathrm{d}p_{\cdot q}/2}}{\prod_{i=1}^q p_i^{\mathrm{d}p_i/2}} \cdot \frac{\prod_{i=1}^q \left[|(U_i)_{C_1}| \prod_{j=2}^k |(U_i)_{R_j\cdot}|\right]^{p_i/2}}{\left[|(U_{\cdot q})_{C_1}| \prod_{j=2}^k |(U_{\cdot q})_{R_j\cdot}|\right]^{p_{\cdot q}/2}}, \tag{12}$$

which is going to be the basis of our inference. We are confident that, following an argument parallel to that given in the saturated case, one can prove that the test using Λ in Equation (12) is unbiased. It is not the purpose of this paper to prove unbiasedness. We next turn our attention to the various approximations to the distribution of $M = -2\log\Lambda$.

3.2. The Cumulant Generating Function of $M = -2\log\Lambda$

In order to compute the LR approximation to the distribution of M under H_0, we first need to compute the cumulant generating function of M, that is,

$$K_M(t) = \log \mathrm{E}(e^{Mt}),$$

under H_0. Let $c = p_{\cdot q}^{\mathrm{d}p_{\cdot q}/2} / \prod_{i=1}^q p_i^{\mathrm{d}p_i/2}$. We have the following result.

Theorem 1. *Let $X_{il},\ l = 1, \ldots, n_i$ for $i = 1, \ldots, q$ be q independent random samples from the $N_d(\mu_i, \Sigma_i)$, $\Sigma_i^{-1} \in P_{\mathcal{G}}$ distributions. With the notations given above and under the null hypothesis $H_0 : \Sigma_1^{-1} = \Sigma_2^{-1} = \cdots = \Sigma_q^{-1} \in P_{\mathcal{G}}$, the cumulant generating function of M is*

$$K(t) = \log \mathbf{E}(e^{Mt}) = t \log c + \log \mathbf{E} \left(\frac{\prod_{i=1}^{q} \left|U_{iC_1}\right|^{-p_i t} \prod_{j=2}^{k} \left|U_{iR_j.}\right|^{-p_i t}}{\left|U_{.qC_1}\right|^{-p_{.q} t} \prod_{j=2}^{k} \left|U_{.qR_j.}\right|^{-p_{.q} t}} \right) \tag{13}$$

$$= t \log c + \log \left(\frac{\prod_{i=1}^{q} \frac{\Gamma_{c_1}((p_i - 2p_i t)/2)}{\Gamma_{c_1}(p_i/2)} \prod_{j=2}^{k} \frac{\Gamma_{c_j - s_j}((p_i - s_j - 2p_i t)/2)}{\Gamma_{c_j - s_j}((p_i - s_j)/2)}}{\frac{\Gamma_{c_1}((p_{.q} - 2p_{.q} t)/2)}{\Gamma_{c_1}(p_{.q}/2)} \prod_{j=2}^{k} \frac{\Gamma_{c_j - s_j}((p_{.q} - s_j - 2p_{.q} t)/2)}{\Gamma_{c_j - s_j}((p_{.q} - s_j)/2)}} \right). \tag{14}$$

The proof is given in the Appendix.

3.3. Three Approximations to the Cumulative Distribution Function of M

In Section 4, we will compare the accuracy of the LR (1980) approximation $\hat{F}_M(m)$ to the cumulative distribution function $F(m) = P(M \leq m)$ of M with that of the two most common classical methods, namely, the standard χ^2 and Box (1949) approximations. The LR approximation is of order $O(n^{-3/2})$ and is now a classical result which can be found for example in Section 1.2.1 of Butler (2007) or Section 9.6 of Young & Smith (2005). For our particular problem, its expression can be readily derived from Equation (14) using the first and second derivatives of the Gamma function.

The standard χ^2 approximation to $F(m)$ is of order $O(n^{-1})$ and is equal to $F_M^*(m) = P(\chi_f^2 \leq m)$ where f is the drop in number of parameters between the models under H_0 and H_a, which in our case is

$$f = (q-1)\Big(\sum_{j=1}^{k} c_j(c_j+1)/2 - \sum_{j=2}^{k} s_j(s_j+1)/2 \Big). \tag{15}$$

As for the Box (1949) approximation, it applies to random variables that have moments that can be expressed as particular ratios of Gamma functions (see p.304 of Muirhead, 1982). This is the case for M and the Box (1949) approximation is $\tilde{F}_M(m) = P(\chi_f^2 \leq m\rho)$ where

$$f = q \left(\frac{c_1(c_1-1)}{2} + \sum_{j=2}^{q} \left[\frac{c_j(c_j-1)}{2} - \frac{s_j(s_j-1)}{2} \right] \right) + d(q-1)$$

$$\rho = 1 - \frac{2}{f} \Big(\sum_{j=1}^{q} \frac{1}{p_j} - \frac{1}{p_{\cdot q}} \Big) \cdot \left(\sum_{l=1}^{c_1} \left[\left(\frac{l-1}{2} \right)^2 + \frac{l-1}{2} + \frac{1}{6} \right] + \sum_{i=2}^{k} \sum_{l=1}^{r_i} \left[\left(\frac{s_i + l - 1}{2} \right)^2 + \frac{s_i + l - 1}{2} + \frac{1}{6} \right] \right).$$

Details of the derivation of the parameters for the standard chi-square and Box (1949) approximations are given in the Supplementary File (17).

4. NUMERICAL RESULTS FOR THE NULL DISTRIBUTION OF M

In this section, we will illustrate the accuracy of the LR approximation to the cumulative distribution of M through several examples. In each case, we will compare the three methods discussed in this paper: the standard χ^2 approximation, the Box (1949) approximation and the LR approximation. We will see that, in all cases, the LR approximation is extremely accurate even for small sample sizes while the other two approximations are completely unsatisfactory unless the sample sizes are very large.

As discussed in the introduction, our motivating example is the test of equality of covariance matrices in different months for the Call Centre data set. However, for illustrative purposes, we will start with a low-dimensional simulation example where the graph has a small number of cliques. The accuracy of the three approximations is evaluated by simulating data from Normal distributions with covariances $\Sigma_i, i = 1, \ldots, q$ under H_0 in Equation (2), that is when the Σ_i's are equal, and then computing, for each approximation, the proportion of P values that are less than a specific significance level α. The closer the proportion of P values to α, the more accurate the method is.

The second example concerns the Call Centre data, which have already been analysed in numerous papers such as Huang et al. (2006), Bickel & Levina (2008), and Rajaratnam, Massam, & Carvalho (2009). We will use the decomposable graphical model identified by Rajaratnam, Massam, & Carvalho (2009). The problem is high-dimensional and the corresponding graph has 99 cliques. We divide the data set into twelve months and we test for equality of covariance matrices in successive months. We will see that the results given by the three methods give very different P-value which can lead to contradictory conclusions. This, of course, only illustrates that the three methods yield different results. In order to show accuracy of the LR approximation, we do a simulation study as described below. The criteria for comparison are the same as in the first example. Again, the simulation results show that the LR approximation is the most accurate and the other two do not give satisfactory results.

The programs (available upon request from the authors) for the calculations below are written in R 2.10.1 and are run on a PC model Dell Optiplex 960 with processor Inter Quad CPU Q9650 3Ghz and a memory of 8GB. Examples of speed will be given below.

4.1. A Simulation Study for a Low-Dimensional Example

Consider the decomposable graph $\mathcal{G}$ as given in Figure 1. We consider the graphical Gaussian model Markov with respect to $\mathcal{G}$ and with precision matrix in $P_{\mathcal{G}}$ equal to

$$K = \Sigma^{-1} = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 & 2 & 1 \\ 1 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

We randomly selected $q \geq 2$ samples of size $n_i,\ i = 1, \ldots q$ from this distribution. We perform 10,000 iterations. We test $H_0 : \Sigma_1 = \cdots = \Sigma_q$. The proportions of P values less than the significance levels $\alpha = 0.025, 0.05$, and 0.1, obtained by the standard χ^2 approximation, $P(\chi^2_f \geq m) = 1 - F^*_M(m)$, (here $f = 15(q-1)$ in Eq. 15), the Box (1949) approximation, $P(\chi^2_f \geq m\rho) = 1 - \tilde{F}_M(m)$, and the LR approximation $P_{\mathrm{LR}} = 1 - \hat{F}(m)$, are recorded in Tables 1 and 2 , for equal sample sizes ($n_i = n,\ i = 1, \ldots, q$) and unequal sample sizes respectively. The closer the proportion of P values to α, the more accurate the method. It is obvious from the

Table 1: Simulation results: proportion of P values less than α for the test in Section 4.1 with equal sample size $n_i = n,\ i = 1, \ldots, q$ given by the three methods.

q	n	$P(\chi^2_f \geq m)$			$P(\chi^2_f \geq m\rho)$			P_{LR}		
α		0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
2	4	0.865	0.910	0.948	0.436	0.539	0.640	0.025	0.052	0.101
	5	0.541	0.642	0.744	0.177	0.260	0.375	0.022	0.049	0.098
	8	0.185	0.268	0.383	0.067	0.118	0.199	0.025	0.050	0.101
	10	0.117	0.192	0.302	0.049	0.088	0.169	0.024	0.047	0.095
	15	0.071	0.121	0.202	0.040	0.073	0.136	0.026	0.050	0.100
	20	0.053	0.096	0.170	0.032	0.064	0.123	0.023	0.048	0.098
	30	0.044	0.081	0.149	0.031	0.062	0.119	0.023	0.051	0.104
	50	0.033	0.064	0.125	0.027	0.053	0.111	0.024	0.047	0.100
5	4	0.995	0.998	0.999	0.386	0.492	0.617	0.025	0.049	0.097
	5	0.862	0.911	0.950	0.129	0.204	0.315	0.024	0.048	0.098
	8	0.345	0.459	0.588	0.048	0.087	0.161	0.023	0.049	0.098
	10	0.206	0.302	0.428	0.040	0.071	0.133	0.025	0.049	0.095
	15	0.106	0.170	0.277	0.033	0.062	0.115	0.026	0.050	0.098
	20	0.074	0.124	0.214	0.031	0.058	0.109	0.026	0.050	0.097
	30	0.054	0.100	0.174	0.028	0.057	0.114	0.025	0.051	0.107
	50	0.036	0.073	0.135	0.026	0.052	0.102	0.024	0.048	0.097
10	4	1.000	1.000	1.000	0.522	0.625	0.739	0.026	0.048	0.094
	5	0.982	0.991	0.996	0.147	0.224	0.339	0.024	0.051	0.103
	8	0.550	0.658	0.772	0.047	0.086	0.155	0.024	0.050	0.100
	10	0.340	0.448	0.584	0.037	0.069	0.133	0.025	0.049	0.097
	15	0.158	0.244	0.370	0.030	0.059	0.116	0.024	0.050	0.099
	20	0.110	0.176	0.277	0.031	0.060	0.117	0.027	0.054	0.108
	30	0.066	0.111	0.198	0.027	0.056	0.102	0.025	0.053	0.097
	50	0.040	0.081	0.151	0.023	0.048	0.101	0.022	0.045	0.098

two tables that the LR approximation gives results very close to α, whereas the P values obtained using the other two approximations are not satisfactory at all unless the sample sizes are large.

Calculations are fast. For $n = 50$ and $q = 10$ in Table 1, we have the following times in seconds

$P(\chi^2_f \geq m)$	$P(\chi^2_f \geq m\rho)$	P_{LR}
74.2	76.1	78.6

4.2. The Call Centre Data

The data set in this example records the number of incoming phone calls during 10-minute intervals from 7:00 am until midnight to a call Centre of a major financial institution in 2002. Weekends,

Table 2: Simulation results: proportion of P values less than α for the test in Section 4.1 with unequal sample sizes n_i $i = 1, \ldots, q$ given by the three methods.

q	n_i	$P(\chi^2_f \geq m)$			$P(\chi^2_f \geq m\rho)$			P_{LR}		
	α	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
2	4,5	0.746	0.814	0.879	0.314	0.408	0.531	0.026	0.049	0.100
	5,7	0.400	0.501	0.616	0.136	0.210	0.309	0.024	0.051	0.104
	8,11	0.131	0.202	0.306	0.057	0.097	0.172	0.026	0.050	0.098
	10,14	0.094	0.154	0.247	0.047	0.085	0.154	0.026	0.051	0.101
	15,20	0.062	0.106	0.182	0.038	0.071	0.130	0.027	0.052	0.104
	20,26	0.049	0.093	0.162	0.034	0.068	0.127	0.026	0.053	0.108
	30,37	0.036	0.069	0.131	0.028	0.055	0.110	0.024	0.047	0.096
	50,58	0.032	0.061	0.117	0.028	0.053	0.104	0.026	0.048	0.096
5	4,5,6,7,8	0.808	0.871	0.921	0.163	0.245	0.358	0.025	0.050	0.099
	5,7,9,11,13	0.410	0.519	0.645	0.070	0.117	0.208	0.026	0.052	0.100
	8,11,14,17,20	0.122	0.190	0.298	0.037	0.070	0.130	0.025	0.052	0.101
	10,14,18,22,26	0.083	0.143	0.238	0.033	0.062	0.119	0.026	0.048	0.100
	15,20,25,30,35	0.048	0.094	0.169	0.026	0.051	0.107	0.022	0.045	0.097
	20,26,32,38,44,	0.043	0.079	0.146	0.027	0.054	0.104	0.024	0.051	0.096
	30,37,44,51,58	0.037	0.069	0.133	0.027	0.051	0.103	0.025	0.049	0.098
	50,58,66,74,82	0.034	0.064	0.123	0.028	0.054	0.106	0.027	0.052	0.103

public holidays, and days with equipment malfunction are excluded, resulting in data for 239 days. The number of calls in each of these intervals is denoted as $N_{ij}, i = 1, 2, \ldots, 239$ and $j = 1, 2, \ldots, 102$. A standard transformation $x_{ij} = (N_{ij} + 1/4)^{1/2}$ is applied to the raw data to make it closer to the Normal distribution. We adopt the model chosen by cross-validation in Rajaratnam, Massam, & Carvalho (2009) which is the banded model with cliques of cardinality $k = 4$. This means that if the vertices are labelled $1, 2, 3, \ldots, 102$. Then the cliques are $C_1 = \{1, 2, 3, 4\}, C_2 = \{2, 3, 4, 5\}, \ldots, C_{99} = \{99, 100, 101, 102\}$. Thus this decomposable graph has 402 free parameters. We split the data set into 12 approximate months (20 days per month for the first 11 months and 19 days for the last month) and denote by Σ_i the covariance matrix for the ith month. The P values for testing the null hypotheses $\Sigma_i = \Sigma_{i+1},\ i = 1, \ldots, 11$ and $\Sigma_1 = \cdots = \Sigma_{12}$, obtained from the three methods are recorded in Table 3. Here in Equation (15), $f = 402$.

Clearly the three methods yield different P values and for some months can lead to contradictory conclusions. For example, when testing $H_0 : \Sigma_4 = \Sigma_5$, the P-value obtained by the χ^2 approximation gives strong evidence against the null hypothesis, and so is the P-value obtained by the Box (1949) approximation; however, the P-value obtained by the LR approximation gives no evidence against the null hypothesis. Similar conclusion can be reached when testing $H_0 : \Sigma_5 = \Sigma_6$, and also $H_0 : \Sigma_8 = \Sigma_9$. Furthermore, for testing $H_0 : \Sigma_9 = \Sigma_{10}$, the P-value obtained by the χ^2 approximation gives strong evidence against the null hypothesis, the P-value obtained by the Box approximation gives weak evidence against the null hypothesis, and the P-value obtained by the LR approximation gives no evidence against the null hypothesis. We will therefore next run a Monte Carlo simulation to compare the accuracy of the three methods for

Table 3: P values for the tests of Section 4.2 given by the three methods.

H_0 :	$P(\chi^2_f \geq m)$	$P(\chi^2_f \geq m\rho)$	P_{LR}
$\Sigma_1 = \Sigma_2$	0.24229	0.72793	0.94989
$\Sigma_2 = \Sigma_3$	0.11145	0.54079	0.87536
$\Sigma_3 = \Sigma_4$	0.00001	0.00144	0.03199
$\Sigma_4 = \Sigma_5$	0.00057	0.03071	0.22028
$\Sigma_5 = \Sigma_6$	0.00094	0.04193	0.26332
$\Sigma_6 = \Sigma_7$	0.00000	0.00000	0.00005
$\Sigma_7 = \Sigma_8$	0.00000	0.00000	0.00039
$\Sigma_8 = \Sigma_9$	0.00207	0.06742	0.34252
$\Sigma_9 = \Sigma_{10}$	0.00610	0.12603	0.47427
$\Sigma_{10} = \Sigma_{11}$	0.00005	0.00638	0.08451
$\Sigma_{11} = \Sigma_{12}$	0	0	0
$\Sigma_1 = \cdots = \Sigma_{12}$	0	0	0

this high-dimensional example. The computation times for all rows together in Table 3 are

$P(\chi^2_f \geq m)$	$P(\chi^2_f \geq m\rho)$	P_{LR}
0.46	0.49	0.52

4.3. A High-Dimensional Simulation Study

We consider two covariances matrices Σ_1 and Σ_2 which we take equal to Σ_0, the sample covariance matrix calculated from the Call Centre Data. We simulate a sample of size n from a multivariate normal distribution $N_{102}(0, \Sigma_1)$, and another sample of the same size from an independent multivariate normal distribution $N_{102}(0, \Sigma_2)$. We perform 10,000 iterations. We first take the sample size to be $n = 20$ in order to have a small sample case and then we take $n = 239$ in order to mimic the Call Centre Data. We test $H_0 : \Sigma_1 = \Sigma_2$. Table 4 records the proportion of P values that are less than $\alpha = 0.025$, 0.05, and 0.1, obtained using the three methods. The LR approximation gives results extremely close to α even when the sample size is small. Although the Box approximation gives results closer to α than the χ^2 approximation, the results are still far from α even for the large sample size $n = 239$. The proportions of P values less than the significance level α are plotted versus α in Figure 2. The graph shows that the P_{LR} (red) line almost overlaps the black diagonal line, α, which implies that the LR approximation is uniformly more accurate than the other two methods even when the sample size is small.

Here again, calculations are fast. For $n_1 = n_2 = 20$ in Table 4, the times in seconds are

	$P(\chi^2_f \geq m)$	$P(\chi^2_f \geq m\rho)$	P_{LR}
$n_1 = n_2 = 20$	348.6	349.1	371.8
$n_1 = n_2 = 239$	531.1	550.9	575.8

Table 4: Simulation results: proportion of P values less than α for the test in Section 4.1.

	α	$P(\chi^2_f \geq m)$	$P(\chi^2_f \geq m\rho)$	P_{LR}
	0.025	0.6602	0.2031	0.0292
$n = 20$	0.05	0.7565	0.2971	0.0575
	0.1	0.8459	0.4204	0.1105
	0.025	0.0444	0.0365	0.0305
$n = 239$	0.05	0.0786	0.0664	0.0584
	0.1	0.1418	0.1218	0.1068

4.4. Comparison With Monte Carlo Simulation

At this point, the speed and accuracy of our method to compute P values for our testing problem have been demonstrated. One might think, though, that a similar or better result might be achieved via Monte Carlo simulation, thus avoiding the analytical work needed to compute the saddlepoint approximation. We therefore computed the P values of Sections 4.1 and 4.3 through Monte Carlo simulations as follows. First, we generated one sample from q normal $N_p(0, \Sigma)$ distributions with Σ as given in Sections 4.1 and 4.3. We treated this sample as our data. From this we calculated

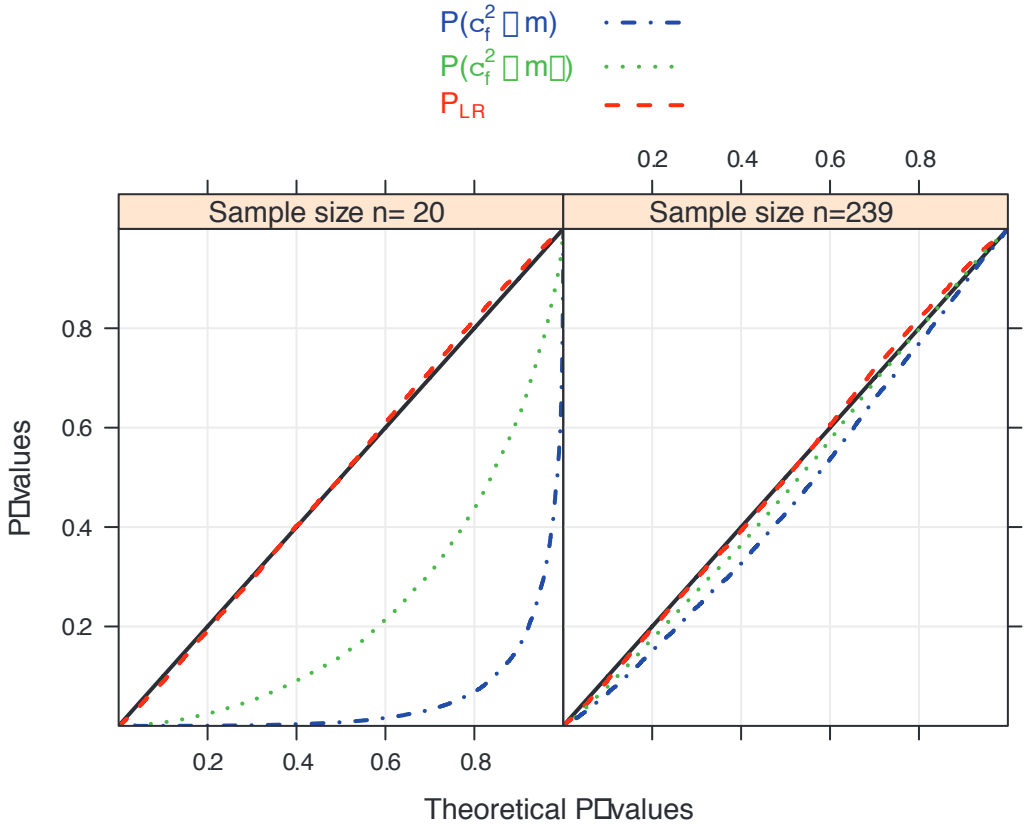

Figure 2: Plot of P values given by the three methods versus theoretical P values, Section 4.3. [Color figure can be seen in the online version of this article, available at http://wileyonli nelibrary.com/journal/cjs]

Table 5: Comparison of accuracy and computational times between our method and Monte Carlo simulation for the problem in Section 4.1: $q = 10, n_1 = \cdots = n_{10} = 50$.

Method	Time in seconds	P-value
$P(\chi_f^2 \geq m_0)$	0.008	0.007799
$P(\chi_f^2 \geq m_0\rho)$	0.009	0.0150384
P_{LR}	0.009	0.0157693
Empirical P-value for $N = 10^4$	79	0.0173
Empirical P-value for $N = 10^5$	782	0.0159
Empirical P-value for $N = 10^6$	7812	0.015995
Empirical P-value for $N = 5 \times 10^6$	78343	0.0157714

the observed M-statistic and called it m_0. We reported the values $P(\chi_f^2 \geq m_0)$, $P(\chi_f^2 \geq m_0\rho)$ and P_{LR} as well as the time required for each calculation. Second in each case, we generated N data sets with $N = 10^4, 10^5, 10^6, 5 \times 10^6$. Third, for each of the N data sets, we calculated the observed M-statistic m and computed the empirical P-value, that is, the proportion of the N observed M-statistics that are larger than or equal to m_0. We reported the P values and time required for each calculation. The results as summarized in the following table for Section 4.1. We see that to obtain the same 4th decimal place accuracy as in P_{LR}, the computation times are about 10^8 times longer. For the computations of Section 4.3, the results are even more striking since for $N = 5 \times 10^6$, we still do not have the same second decimal place accuracy as for P_{LR} with much longer computational times. The detailed results are not reported for the sake of brevity.

5. CONCLUSION

In this paper, we derived the Bartlett–Box M-statistic $M = -2 \log \Lambda$ for the equality of covariances of q samples from normal distributions Markov with respect to a decomposable graph $\mathcal{G}$. Using some recent powerful independence results derived from the hyper Markov properties of the hyper-Wishart distribution, we subsequently derived the cumulant generating function of M. This allowed us to give the Lugannani and Rice approximation to its cumulative distribution function and thus obtain extremely accurate approximation. Our proposed method can be viewed as an extension of that proposed by Booth et al. (1995) in the sense that when $\mathcal{G}$ is complete, the model is saturated and the LR approximation we obtain yields their approximation with the same numerical accuracy. Various numerical examples illustrate the extreme accuracy of our method compared to the two existing methods, in the general decomposable case. The computation times for the three methods are not significantly different from each other. However, simulation results show that our proposed method has extremely good coverage properties even when the sample size is small. Moreover, as demonstrated in Table 5, for the same level of accuracy, our method is immensely faster than Monte-Carlo simulation.

APPENDIX

Proof of Theorem 1. Since $e^{Mt} = \Lambda^{-2t}$, the equality in Equation (13) follows immediately from the definition of M and Equation (12). We now need to compute the $(-2t)$th moment of Λ. For

convenience, we will use the notation $A_i = \pi(U_i) = p_i\pi(\tilde{\Sigma}_i^a)$. We know, from Section 2, that under H_0, $A_i \in Q_{\mathcal{G}}$ follows the $W_{Q_{\mathcal{G}}}(a_i;\, p_i,\, \Sigma)$ distribution. Let

$$c_{p_i} = \frac{\prod_{j=2}^{k} \Gamma_{s_j}(p_i/2)|\Sigma_{S_j}|^{p_i/2}}{\prod_{j=1}^{k} \Gamma_{c_j}(p_i/2)|\Sigma_{C_j}|^{p_i/2}}.$$

From Equations (7), (10), and (11), it follows that, under H_0

$$\begin{aligned} E\left[\left(\frac{\Lambda}{c}\right)^{-2t}\right] &= \prod_{i=1}^{q} c_{p_i} \int_{Q_{\mathcal{G}}} \cdots \int_{Q_{\mathcal{G}}} \left|U_{.qC_1}\right|^{p_{\cdot q}t} \prod_{j=2}^{k} \left|U_{.qR_j.}\right|^{p_{\cdot q}t} \cdot \frac{\prod_{j=1}^{k} |(a_i)_{C_j}|^{p_i(1-2t)/2-(c_j+1)/2}}{\prod_{j=2}^{k} |(a_i)_{S_j}|^{p_i(1-2t)/2-(c_j+1)/2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2}\left\{\sum_{j=1}^{k} \langle \Sigma_{C_j}^{-1}, (a_i)_{C_j}\rangle - \sum_{j=2}^{k} \langle \Sigma_{S_j}^{-1}, (a_i)_{S_j}\rangle\right\}\right\} 1_{Q_{\mathcal{G}}}(a_i)\, \mathrm{d}a_i \\ &= \prod_{i=1}^{q} \frac{c_{p_i}}{c_{p_i(1-2t)}} \cdot \mathrm{E}\left(\left|U_{.qC_1}\right|^{p_{\cdot q}t} \prod_{j=2}^{k} \left|U_{.qR_j.}\right|^{p_{\cdot q}t}\right) \end{aligned}$$

where the expectation is with respect to $A_i \sim W_{Q_{\mathcal{G}}}(p_i(1-2t)),\ i = 1, \ldots, q$. Since the samples $X_{i1}, \ldots, X_{in_i},\ i = 1, \ldots, q$ are independent, the $A_i,\ i = 1, \ldots, q$ are also independent. Then, from Section 4.3 of Letac & Massam (2007), we know that $A_{\cdot q} = \sum_{i=1}^{q} A_i \sim W_{Q_{\mathcal{G}}}(p_{\cdot q}(1-2t), \Sigma)$. We also have that $U_{.qC_1} = (A_{\cdot q})_{C_1}$, $U_{.qR_j.} = (A_{\cdot q})_{R_j\cdot}$. It follows, from Theorem 4.5 of Letac & Massam (2007), that

$$\begin{aligned} &(U_{.qC_1},\ U_{.qR_j.},\ j = 2, \ldots, k) \text{ are mutually independent,} \\ &U_{.qC_1} \sim w_{c_1}(p_q(1-2t), \Sigma), \\ &U_{.qR_j.} \sim w_{c_j - s_j}(p_{\cdot q}(1-2t) - s_j, \Sigma_{R_j\cdot}). \end{aligned}$$

This property is crucial to our argument and follows from the so-called weak hyper Markov property of the hyper Wishart, first identified by Dawid & Lauritzen (1993) and further analysed in Theorem 4.5 of Letac & Massam (2007). So, since $p_{\cdot q}(1-2t)/2 + 2p_{\cdot q}t/2 = p_{\cdot q}/2$, we have

$$E\left[\left(\frac{\Lambda}{c}\right)^{-2t}\right] = \prod_{i=1}^{q} \frac{c_{p_i}}{c_{p_i(1-2t)}} \cdot E(\left|U_{.qC_1}\right|^{p_{\cdot q}t}) \cdot \prod_{j=2}^{k} E(\left|U_{.qR_j.}\right|^{p_{\cdot q}t})$$

where the expectation is taken with respect to $A_{\cdot q} \sim W_{Q_{\mathcal{G}}}(p_{\cdot q}(1-2t), \Sigma)$. It follows that

$$E\left[\left(\frac{\Lambda}{c}\right)^{-2t}\right] = \prod_{i=1}^{q} \frac{\prod_{j=2}^{k} \Gamma_{s_j}(p_i/2)|\Sigma_{S_j}|^{p_i/2}}{\prod_{j=1}^{k} \Gamma_{c_j}(p_i/2)|\Sigma_{C_j}|^{p_i/2}} \tag{1}$$

$$\cdot \prod_{i=1}^{q} \frac{\prod_{j=1}^{k} \Gamma_{c_j}(p_i(1-2t)/2)|\Sigma_{C_j}|^{p_i(1-2t)/2}}{\prod_{j=2}^{k} \Gamma_{s_j}(p_i(1-2t)/2)|\Sigma_{S_j}|^{p_i(1-2t)/2}} \tag{2}$$

$$\cdot \prod_{i=1}^{q} \frac{|\Sigma_{C_1}|^{p\cdot q/2}\Gamma_{c_1}(p_{\cdot q}/2)}{|\Sigma_{C_1}|^{p\cdot q(1-2t)/2}\Gamma_{c_1}(p_{\cdot q}(1-2t)/2)} \tag{3}$$

$$\cdot \prod_{j=2}^{k} \frac{|\Sigma_{R_{j\cdot}}|^{(p\cdot q-s_j)/2}\Gamma_{c_j-s_j}((p_{\cdot q}-s_j)/2)}{|\Sigma_{R_{j\cdot}}|^{(p\cdot q(1-2t)-s_j)/2}\Gamma_{c_j-s_j}((p_{\cdot q}(1-2t)-s_j)/2)} \,. \tag{4}$$

Now, for any integers s, c, p such that $s < c < p$, we have $\Gamma_c(p)/\,\Gamma_s(p) = \Gamma_{c-s}(p-s)$. Applying this to the various ratios in Equations (1)-(4) above and noticing that the terms in Σ cancel out, we obtain

$$E\left[\left(\frac{\Lambda}{c}\right)^{-2t}\right] = \frac{\prod_{i=1}^{q} \frac{\Gamma_{c_1}((p_i-2p_it)/2)}{\Gamma_{c_1}(p_i/2)} \prod_{j=2}^{k} \frac{\Gamma_{c_j-s_j}((p_i-s_j-2p_it)/2)}{\Gamma_{c_j-s_j}((p_i-s_j)/2)}}{\frac{\Gamma_{c_1}((p_{\cdot q}-2p_{\cdot q}t)/2)}{\Gamma_{c_1}(p_{\cdot q}/2)} \prod_{j=2}^{k} \frac{\Gamma_{c_j-s_j}((p_{\cdot q}-s_j-2p_{\cdot q}t)/2)}{\Gamma_{c_j-s_j}((p_{\cdot q}-s_j)/2)}},$$

which immediately yields Equation (14). j

ACKNOWLEDGEMENTS

H. Massam and A. Wong gratefully acknowledge the support of NSERC through their Discovery Grants. Y. Wu acknowledges the support of CIHR Training grant in Genetic Epidemiology and Statistical Genetics.

BIBLIOGRAPHY

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, 160, 268–282.

Bickel, P. J. & Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36, 199–227.

Booth, J. G., Butler, R. W., Huzurbazar, S. & Wood, A. T. A. (1995) Saddlepoint approximations for *P* values of some tests of covariance matrices. *Journal of Statistical Computation and Simulation*, 53, 165–180.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood ratio criteria. *Proceedings of the Royal Society A*, 160, 268–282.

Butler, R. W. (2007). *Saddlepoint Approximations With Applications*. Cambridge University Press, NY.

Das Gupta, S. (1969). Properties of power functions of some tests concerning dispersion matrices of multivariate normal distributions. *The Annals of Mathematical Statistics*, 40, 697–701.

Dawid, A. & Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21, 1272–1317.

Dempster, A. (1975). Covariance selection. *Biometrics*, 28, 157–175.

Huang, J. Z., Liu, N., Pourahmadi, M. & Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93, 85–98.

Koller, D. & Friedman, N. (2009). *Probabilistic Graphical Models, Principles and Techniques*. The MIT Press, Cambridge, MA.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford Science Publications, NY.

Letac, G. & Massam, H. (2007). Wishart distributions for decomposable graphs. *The Annals of Statistics*, 35, 1278–1323.

Lugannani, R. & Rice, S. (1980). Saddlepoint approximations for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12, 475–490.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, NY.

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *The Annals of Mathematical Statistics*, 8, 247–263.

Rajaratnam, B., Massam, H., & Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of Statistics*, 36, 2818–2849.

Skovgaard, I. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability*, 24, 275–287.

Young, G. A. & Smith, R. L. (2005). *Essentials of Statistical Inference*. Cambridge University Press, NY.

Received 18 October 2012
Accepted 13 July 2013