Journal of Multivariate Analysis 155 (2017) 305–316

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing block-diagonal covariance structure for high-dimensional data under non-normality

Yuki Yamada [a], Masashi Hyodo [b,*], Takahiro Nishiyama [c]

[a] NEC Solution Innovators, Ltd., 1-18-7 Shinkiba, Koto-ku, Tokyo 136-8627, Japan

[b] Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai-shi, Osaka 599-8531, Japan

[c] Department of Business Administration, Senshu University, 2-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8580, Japan

ARTICLE INFO

Article history:
Received 30 March 2016
Available online 16 January 2017

AMS 2000 subject classifications:
primary 62H15
secondary 62F05

Keywords:
Tests of covariance structure
High dimension
Statistical hypothesis testing
Non-normality

ABSTRACT

In this article, we propose a test for making an inference about the block-diagonal covariance structure of a covariance matrix in non-normal high-dimensional data. We prove that the limiting null distribution of the proposed test is normal under mild conditions when its dimension is substantially larger than its sample size. We further study the local power of the proposed test. Finally, we study the finite-sample performance of the proposed test via Monte Carlo simulations. We demonstrate the relevance and benefits of the proposed approach for a number of alternative covariance structures.

1. Introduction

In recent years, a number of statistical methods have been proposed for high-dimensional data such as DNA microarray gene expressions, where the number of feature variables, p, exceeds the sample size, N. For analyzing high-dimensional data, the $p \times p$ covariance matrix Σ is an important measure for dependence among the components of a high-dimensional random vector $\boldsymbol{X}$, and thus plays a major role in many statistical procedures. Accordingly, statistical inference about Σ is an important problem.

A number of procedures for testing a high-dimensional covariance matrix, which include testing identity, sphericity and diagonal structure, have been proposed by Schott [9], Srivastava [10], Akita et al. [1], Srivastava and Reid [12], and so on. All these tests assume either, and sometimes even both, normality or moderate dimensionality such that $p/n \to c$ for a finite constant c. There has also been substantial research on inference concerning the covariance matrix under non-normal high-dimensional distributions; see, e.g., Chen et al. [3], Srivastava et al. [11], Zhong and Chen [16], Li and Chen [6], Qiu and Chen [8], Yata and Aoshima [14], Srivastava et al. [13], and Himeno and Yamada [4].

In this paper, we provide tests for covariance matrices without assuming normality while allowing the dimension p to be much larger than the sample size N. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ be $N \geq 4$ independent and identically distributed (i.i.d.) p-dimensional

* Corresponding author.
E-mail addresses: hyodo@ms.osakafu-u.ac.jp (M. Hyodo), nishiyama@isc.senshu-u.ac.jp (T. Nishiyama).

http://dx.doi.org/10.1016/j.jmva.2016.12.009
0047-259X/

random vectors with $\mathrm{E}(\boldsymbol{X}_i) = \boldsymbol{\mu}$ and $\mathrm{var}(\boldsymbol{X}_i) = \Sigma$. We partition $\boldsymbol{X}_i$, $\boldsymbol{\mu}$, and Σ into q components, viz.

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_i^{(1)} \\ \vdots \\ \boldsymbol{X}_i^{(q)} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \vdots \\ \boldsymbol{\mu}^{(q)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & \ddots & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix},$$

where $\boldsymbol{X}_i^{(g)}$ and $\boldsymbol{\mu}^{(g)}$ are $p_g \times 1$ vectors and Σ_{gh} is a $p_g \times p_h$ matrix, with $g, h \in \{1, \ldots, q\}$. Note that $p = p_1 + \cdots + p_q$. Our interest is to test covariance structure, viz.

$$\mathcal{H}_0 : \Sigma_{ij} = O \text{ for any } i \neq j,\ i, j \in \{1, \ldots, q\} \quad \text{vs.} \quad \mathcal{H}_A : \text{not } \mathcal{H}_0. \tag{1.1}$$

For $q = 2$ in hypothesis (1.1), i.e., to test the independence between two sub-vectors, Srivastava and Reid [12] proposed a procedure under the multivariate normal assumption, while Yata and Aoshima [15] proposed an alternative approach that dispenses with this assumption. Our testing procedure extends the results derived by Yata and Aoshima [15].

Let $\Sigma_d = \mathrm{diag}(\Sigma_{11}, \ldots, \Sigma_{qq})$. Then $\mathcal{H}_0$ is equivalent to $\Sigma = \Sigma_d$. Checking the block-diagonal structure is very important when $p \gg N$. For, if $\mathcal{H}_0$ holds, then the total number of unknown parameters to be estimated for Σ reduces from $p(p+1)/2$ to $\sum_{i=1}^{q} p_i(p_i + 1)/2$. Further, it is interesting to note that the block-diagonal covariance structure attracts much attention across the fields of convex optimization and machine learning, with the main focus being the Gaussian graph structure learning procedures. In particular, Pavlenko et al. [7] determined the block structure for colon cancer data by using gLasso and Cuthill–McKee algorithm. Testing hypothesis (1.1) may provide a procedure to check whether this structure is true.

For testing hypothesis (1.1), Hyodo et al. [5] proposed a test statistic based on the normalized Frobenius matrix norm under normality. They used an estimator of $\|\Sigma\|_F^2$ proposed by Bai and Saranadasa [2] and Srivastava [10]. Here, $\|A\|_F^2 = \mathrm{tr}(AA^\top)$ denotes the Frobenius matrix norm of the matrix A. This estimator is unbiased under the normality assumption but not in general. Thus, in this paper, we propose a new test statistic based on an unbiased estimator proposed by Himeno and Yamada [4] and Srivastava et al. [13] that does not rely on the normality assumption.

The rest of the paper is organized as follows. Section 2 presents test procedure after establishing the asymptotic normality of the test statistics, and derives asymptotic powers in Section 3. Further, in Section 4, the attained significance levels and powers of the suggested test are empirically analyzed, and applied to real dataset. Finally, Section 5 concludes this paper. Some technical details are relegated to the Appendix (see Appendix A).

2. Test procedure

2.1. Assumptions

In this section, we list the assumptions we need to construct a test for (1.1). Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ are i.i.d. p-dimensional random vectors such that

$$\boldsymbol{X}_i = \Gamma \boldsymbol{Z}_i + \boldsymbol{\mu} \quad \text{for } i \in \{1, \ldots, N\}, \tag{2.1}$$

where $\Gamma = (\Gamma^{(1)\top}, \ldots, \Gamma^{(q)\top})^\top$ is a $p \times m$ constant matrix such that $\Gamma\Gamma^\top = \Sigma$, and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N$ are i.i.d. m-dimensional random vectors such that $\mathrm{E}(\boldsymbol{Z}_i) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{Z}_i) = I_m$, an $m \times m$ identity matrix. Note that $\boldsymbol{X}_i^{(g)} = \Gamma^{(g)}\boldsymbol{Z}_i + \boldsymbol{\mu}^{(g)}$, and $\Gamma^{(g)}\Gamma^{(h)\top} = \Sigma_{gh}$ for all $g, h \in \{1, \ldots, q\}$.

Further, we use the following assumptions as necessary.

(A1) Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{im})^\top$, with each Z_{ij} having a uniformly bounded 4th moment, and there exist finite constants κ_3, κ_4 such that $\mathrm{E}(Z_{ij}^3) = \kappa_3$, $\mathrm{E}(Z_{ij}^4) = \kappa_4 + 3$ and for any positive integers r and α_ℓ such that $\alpha_\ell \leq 4$ and $\alpha_1 + \cdots + \alpha_r \leq 8$, $\mathrm{E}(\prod_{\ell=1}^{r} Z_{ij_\ell}^{\alpha_\ell}) = \prod_{\ell=1}^{r} \mathrm{E}(Z_{ij_\ell}^{\alpha_\ell})$ whenever $j_1, \ldots, j_r$ are distinct indices.

(A1′) Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{im})^\top$, with each Z_{ij} having a uniformly bounded 8th moment, and there exist finite constants κ_3, κ_4 such that $\mathrm{E}(Z_{ij}^3) = \kappa_3$, $\mathrm{E}(Z_{ij}^4) = \kappa_4 + 3$ and for any positive integers r and α_ℓ such that $\sum_{\ell=1}^{r} \alpha_\ell \leq 8$, $\mathrm{E}(\prod_{\ell=1}^{r} Z_{ij_\ell}^{\alpha_\ell}) = \prod_{\ell=1}^{r} \mathrm{E}(Z_{ij_\ell}^{\alpha_\ell})$, whenever $j_1, \ldots, j_r$ are distinct indices.

(A2) q is fixed. At least one of $p_1, \ldots, p_q$ is a function of N, and goes to infinity as $N \to \infty$.

(A3) $\dfrac{\max_{\substack{g \neq h \\ g,h \in \{1,\ldots,q\}}} \left\{\|\Sigma_{gg}^2\|_F^2 \|\Sigma_{hh}^2\|_F^2\right\}}{\left(\max_{\substack{g \neq h \\ g,h \in \{1,\ldots,q\}}} \left\{\|\Sigma_{gg}\|_F^2 \|\Sigma_{hh}\|_F^2\right\}\right)^2} \to 0$ as $p \to \infty$.

(A4) $\limsup_{m' \to \infty} \max_{\substack{g \neq h \\ g,h \in \{1,\ldots,q\}}} \left\{\dfrac{N^2 \|\Sigma_{gh}\|_F^4}{\|\Sigma_{gg}\|_F^2 \|\Sigma_{hh}\|_F^2}\right\} < \infty$.

(A5) $\limsup_{m' \to \infty} \min_{\substack{g \neq h \\ g,h \in \{1,\ldots,q\}}} \left\{\dfrac{N^2 \|\Sigma_{gh}\|_F^4}{\|\Sigma_{gg}\|_F^2 \|\Sigma_{hh}\|_F^2}\right\} \to \infty$.

Here $m' = \min(N, p)$. We state some remarks on the conditions.