Improved Simplified T^2 Test Statistics for Mean Vector with Monotone Missing Data

Ayaka Yagi*, Takashi Seo* and Zofia Hanusz**

* *Department of Applied Mathematics*
Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

** *Department of Applied Mathematics and Computer Science*
University of Life Sciences
Głęboka 28, 20-612 Lublin, Poland

Abstract

In this study, the distribution of the simplified Hotelling's T^2 statistic for testing mean vector when the data matrix is of a monotone missing pattern is considered. The test statistic is decomposed and an asymptotic expansion of the null distribution of each term is derived. Using the result, an approximation to the upper percentiles of this statistic are derived. Further, the transformed test statistics based on the Bartlett adjustment are proposed. Finally, by a Monte Carlo simulation, it is shown that the transformed statistics usually perform better than the original statistic with respect to speed of convergence to the chi-squared limiting distribution.

Key Words and Phrases: Asymptotic expansion; Bartlett correction; Chi-squared approximation; Maximum likelihood estimator; Monte Carlo simulation.

1 Introduction

In this paper, we consider a test of a mean vector with k-step monotone missing data pattern. In the case of a two-step monotone missing data pattern, the usual Hotelling's T^2-type statistic for a test of the mean vector and its null distribution have been discussed, among others, by Chang and Richards (2009) and Seko et al. (2012). Krishnamoorthy and Pannala (1999) and Yagi and Seo (2014, 2017), applying different notations and definitions, gave the simplified T^2 statistic and approximated null distribution. Krishnamoorthy and Pannala (1999) considered approximation of upper percentiles using the F-distribution multiplied by a constant. Yagi and Seo (2014, 2017) applied another approximation using a linear interpolation based on the complete parts of the data set. Both procedures are heuristic but yield a good approximation. On the other hand, Kawasaki and Seo (2016) gave an asymptotic expansion for the expectation of usual Hotelling's T^2 statistic under the two-step monotone missing data case. However, an asymptotic expansion for the null distribution of the usual Hotelling's T^2 statistic for more than two-step monotone missing data matrix is complicated.

The main goal of this paper is to improve the chi-squared approximation for the simplified T^2 statistic when the data matrix has a k-step monotone missing pattern. To derive the asymptotic expansion of the null distribution for the simplified T^2 statistic, we used the perturbation method described in Fujikoshi et al. (2010). Decomposition of the test statistic proposed in the paper allowed to calculate the asymptotic expansion more easily. Furthermore, using the asymptotic expansion of the distribution, we present the transformed test statistics based on the Bartlett adjustment. The improved transformations for the general test statistic are discussed by Fujikoshi (2000).

The paper is organized in the following way. Section 2 presents the maximum likelihood estimators (MLEs) of mean vector and covariance matrix for k-step monotone missing data matrix. In Section 3, an asymptotic expansion of the null distribution of the simplified T^2 statistic is derived in two-step case. In k-step case, an approximation to an asymptotic expansion of null distribution of the statistic is derived. Transformed statistics for the simplified T^2 statistic with Bartlett corrections and Bartlett-type corrections are presented in Section 4. In Section 5 results of Monte Carlo simulations on the upper percentiles of test statistics and empirical Type I error are presented. Obtained results proved that the transformed statistics of Section 4 usually perform better than the original statistic (i.e. simplified T^2 statistic) with respect to speed of convergence to the chi-squared limiting distribution. Finally, the paper is ended by concluded remarks.

2 MLEs of mean vector and covariance matrix for monotone missing data

Let us consider k-step monotone missing data matrix of the form

$$
\left(\begin{array}{ccccccccccccc}
x_{11} & \cdots & x_{1,p_k} & x_{1,p_k+1} & \cdots & x_{1,p_{k-1}} & \cdots & x_{1,p_3+1} & \cdots & x_{1,p_2} & x_{1,p_2+1} & \cdots & x_{1,p_1} \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
x_{n_1,1} & \cdots & x_{n_1,p_k} & x_{n_1 1,p_k+1} & \cdots & x_{n_1,p_{k-1}} & \cdots & x_{n_1,p_3+1} & \cdots & x_{n_1,p_2} & x_{n_1,p_2+1} & \cdots & x_{n_1,p_1} \\
x_{n_1+1,1} & \cdots & x_{n_1+1,p_k} & x_{n_1+1,p_k+1} & \cdots & x_{n_1+1,p_{k-1}} & \cdots & x_{n_1+1,p_3+1} & \cdots & x_{n_1+1,p_2} & * & \cdots & * \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots & \cdots & \vdots \\
x_{n_1+n_2,1} & \cdots & x_{n_1+n_2,p_k} & x_{n_1+n_2,p_k+1} & \cdots & x_{n_1+n_2,p_{k-1}} & \cdots & x_{n_1+n_2,p_3+1} & \cdots & x_{n_1+n_2,p_2} & * & \cdots & * \\
 & \vdots & & & \vdots & & & \vdots & & & \vdots & & \\
x_{n_1+\cdots+n_{k-2}+1,1} & \cdots & x_{n_1+\cdots+n_{k-2}+1,p_k} & x_{n_1+\cdots+n_{k-2}+1,p_k+1} & \cdots & x_{n_1+\cdots+n_{k-2}+1,p_{k-1}} & \cdots & * & \cdots & * & * & \cdots & * \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \cdots & * & \cdots & * & * & \cdots & * \\
x_{n_1+\cdots+n_{k-1},1} & \cdots & x_{n_1+\cdots+n_{k-1},p_k} & x_{n_1+\cdots+n_{k-1},p_k+1} & \cdots & x_{n_1+\cdots+n_{k-1},p_{k-1}} & \cdots & * & \cdots & * & * & \cdots & * \\
x_{n_1+\cdots+n_{k-1}+1,1} & \cdots & x_{n_1+\cdots+n_{k-1}+1,p_k} & * & \cdots & * & \cdots & * & \cdots & * & * & \cdots & * \\
\vdots & \cdots & \vdots & * & \cdots & * & \cdots & * & \cdots & * & * & \cdots & * \\
x_{n_1+\cdots+n_k,1} & \cdots & x_{n_1+\cdots+n_k,p_k} & * & \cdots & * & \cdots & * & \cdots & * & * & \cdots & *
\end{array}\right),
$$

where "$*$" indicates a missing observation, and p_i and n_i $(i = 1, 2, \ldots, k)$ denote numbers of traits and sizes of complete submatrices, with the assumption that $n_1 > p_1 = p$.

Let us assume that the random vectors $\boldsymbol{x}_{ij}$ (transposed rows of the above matrix) are independently distributed as $N_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu})_i = (\mu_1, \mu_2, \ldots, \mu_{p_i})'$ and $\boldsymbol{\Sigma}_i$ is the $p_i \times p_i$ principal submatrix of $\boldsymbol{\Sigma}(= \boldsymbol{\Sigma}_1)$, $i = 1, 2, \ldots k$, $j = 1, 2, \ldots, n_i$. Thus, as for the partitions of $\boldsymbol{\Sigma}$, for $i = 2, 3, \ldots k$, we define

$$
\boldsymbol{\Sigma}_i = (\boldsymbol{\Sigma}_1)_i, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{\Sigma}_{i2} \\ \boldsymbol{\Sigma}_{i2}' & \boldsymbol{\Sigma}_{i3} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{i-1} = \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{\Sigma}_{(i-1,2)} \\ \boldsymbol{\Sigma}_{(i-1,2)}' & \boldsymbol{\Sigma}_{(i-1,3)} \end{pmatrix}.
$$

Let us define

$$
\overline{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}, \quad \boldsymbol{E}_i = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)', \quad i = 1, 2, \ldots, k,
$$

$$
\boldsymbol{d}_1 = \overline{\boldsymbol{x}}_1, \quad \boldsymbol{d}_i = \frac{n_i}{N_{i+1}}\left[\overline{\boldsymbol{x}}_i - \frac{1}{N_i}\sum_{j=1}^{i-1} n_j (\overline{\boldsymbol{x}}_j)_i\right], \quad i = 2, 3, \ldots, k,
$$

where

$$
N_1 = 0, \ N_{i+1} = N_i + n_i \ (= \sum_{j=1}^{i} n_j), \ i = 1, 2, \ldots, k.
$$

Further, for calculation of MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, i.e. $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, let us consider the following equalities:

$$
\boldsymbol{T}_1 = \boldsymbol{I}_{p_1}, \quad \boldsymbol{H}_1 = \boldsymbol{E}_1, \quad \boldsymbol{L}_1 = \boldsymbol{H}_1, \ \boldsymbol{F}_1 = \boldsymbol{G}_1, \quad \boldsymbol{G}_1 = \boldsymbol{I}_{p_1},
$$

$$\widehat{\boldsymbol{T}}_{i+1} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \widehat{\boldsymbol{\Sigma}}'_{(i,2)}\widehat{\boldsymbol{\Sigma}}^{-1}_{i+1}, \end{pmatrix}, \quad \boldsymbol{L}_{i1} = (\boldsymbol{L}_i)_{i+1}, \quad \boldsymbol{L}_i = \begin{pmatrix} \boldsymbol{L}_{i1} & \boldsymbol{L}_{i2} \\ \boldsymbol{L}'_{i2} & \boldsymbol{L}_{i3} \end{pmatrix},$$
$$\boldsymbol{G}_{i+1} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \boldsymbol{L}'_{i2}\boldsymbol{L}^{-1}_{i1} \end{pmatrix} \quad \text{for} \quad i = 1, 2, \ldots, k-1,$$

and

$$\boldsymbol{H}_i = \boldsymbol{E}_i + \frac{N_i N_{i+1}}{n_i}\boldsymbol{d}_i\boldsymbol{d}'_i, \quad \boldsymbol{L}_i = (\boldsymbol{L}_{i-1})_i + \boldsymbol{H}_i, \quad \boldsymbol{F}_i = \boldsymbol{F}_{i-1}\boldsymbol{G}_i \ \ \text{for} \ \ i = 2, 3, \ldots, k.$$

Then, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are given in the following theorem.

Theorem 1 (Jinadasa and Tracy, 1992)

The MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for the monotone sample are

$$\widehat{\boldsymbol{\mu}} = \sum_{i=1}^{k} \widehat{\boldsymbol{f}}_i$$

with

$$\widehat{\boldsymbol{f}}_1 = \boldsymbol{d}_1, \quad \widehat{\boldsymbol{f}}_i = \boldsymbol{T}_1\widehat{\boldsymbol{T}}_2 \cdots \widehat{\boldsymbol{T}}_i\boldsymbol{d}_i, \quad i = 2, 3, \ldots, k \quad \textit{and}$$
$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{N_2}\boldsymbol{E}_1 + \sum_{i=2}^{k} \frac{1}{N_{i+1}}\boldsymbol{F}_i\left[\boldsymbol{H}_i - \frac{n_i}{N_i}\boldsymbol{L}_{i-1,1}\right]\boldsymbol{F}'_i.$$

In the case where $\boldsymbol{\Sigma}$ is known, the MLE of $\boldsymbol{\mu}$ is given by

$$\widetilde{\boldsymbol{\mu}} = \sum_{i=1}^{k} \boldsymbol{f}_i,$$

where

$$\boldsymbol{f}_1 = \boldsymbol{d}_1, \quad \boldsymbol{f}_i = \boldsymbol{U}_i\boldsymbol{d}_i, \quad \boldsymbol{U}_1 = \boldsymbol{T}_1, \quad \boldsymbol{U}_i = \boldsymbol{U}_{i-1}\boldsymbol{T}_i \ \ \text{for} \ \ i = 2, 3, \ldots, k,$$
$$\boldsymbol{T}_1 = \boldsymbol{I}_{p_1}, \quad \boldsymbol{T}_{i+1} = \begin{pmatrix} \boldsymbol{I}_{p_{i+1}} \\ \boldsymbol{\Sigma}'_{(i,2)}\boldsymbol{\Sigma}^{-1}_{i+1}, \end{pmatrix} \quad \text{for} \quad i = 1, 2, \ldots, k-1.$$

3 Asymptotic expansion for the distribution of the simplified T^2 statistic for mean vector

Let us consider testing the hypothesis on mean vector

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \ \ \text{vs.} \ \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \qquad (1)$$

where $\boldsymbol{\mu}_0$ is known.

Without loss of generality, we can assume in (1) that $\boldsymbol{\mu}_0 = \mathbf{0}$. Then, the simplified T^2 statistic for null hypothesis in (1) is given by

$$Q = \widehat{\boldsymbol{\mu}}' \, \widetilde{\boldsymbol{\Gamma}}^{-1} \widehat{\boldsymbol{\mu}}, \tag{2}$$

where $\widehat{\boldsymbol{\mu}}$ is a MLE of $\boldsymbol{\mu}$, and $\widetilde{\boldsymbol{\Gamma}}$ is an estimator of $\boldsymbol{\Gamma} = \mathrm{Cov}(\widetilde{\boldsymbol{\mu}})$, and $\widetilde{\boldsymbol{\mu}}$ is the MLE of $\boldsymbol{\mu}$ when $\boldsymbol{\Sigma}$ is known. For detail, see Yagi and Seo (2017), where $Q = \widetilde{T}_1^2$.

3.1 Two-step case ($k = 2$)

For illustration, we consider the null distribution of the simplified T^2 statistic in the case of two-step monotone missing data. From Theorem 1 with $k = 2$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \overline{\boldsymbol{x}}_1 + \widehat{\boldsymbol{T}}_2 \boldsymbol{d}_2, \\
\widehat{\boldsymbol{\Sigma}} &= \frac{1}{n_1}\boldsymbol{E}_1 + \frac{1}{n_1+n_2}\boldsymbol{G}_2 \left[\boldsymbol{E}_2 + \frac{n_1(n_1+n_2)}{n_2}\boldsymbol{d}_2\boldsymbol{d}_2' - \frac{n_2}{n_1}\boldsymbol{L}_{11}\right]\boldsymbol{G}_2'.
\end{aligned}$$

Let us take the following notations $p_{[1]} = p_2$ and $p_{[2]} = p - p_2$, illustrated in Figure 1. Then $\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\mu}}_{[1]}', \ \widehat{\boldsymbol{\mu}}_{[2]}')'$, where $\widehat{\boldsymbol{\mu}}_{[1]} = (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_{p_{[1]}})'$, $\widehat{\boldsymbol{\mu}}_{[2]} = (\widehat{\mu}_{p_{[1]}+1}, \widehat{\mu}_{p_{[1]}+2}, \ldots, \widehat{\mu}_{p_{[1]}+p_{[2]}})'$. We note that

$$\underset{p\times p}{\boldsymbol{E}_1 = \boldsymbol{L}_1} = \left(\begin{array}{c|c} \overbrace{\boldsymbol{L}_{11}}^{p_{[1]}} & \overbrace{\boldsymbol{L}_{12}}^{p_{[2]}} \\ \hline \boldsymbol{L}_{12}' & \boldsymbol{L}_{13} \end{array}\right) \begin{array}{l} \}p_{[1]} \\ \}p_{[2]} \end{array}.$$

and $\boldsymbol{L}_1$ is the sum of squares and products matrix of $n_1 \times p$ data block (see Fig.1) and is distributed as $W_p(\boldsymbol{\Sigma}, n_1 - 1)$. Then, the statistic Q in (2) can be decomposed as follows,

$$Q = Q_1 + Q_2, \tag{3}$$

where

$$\begin{aligned}
&Q_1 = (n_1+n_2)\widehat{\boldsymbol{\mu}}_{[1]}'\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\mu}}_{[1]}, \quad Q_2 = n_1(\overline{\boldsymbol{x}}_{1[2]} - \boldsymbol{L}_{12}'\boldsymbol{L}_{11}^{-1}\overline{\boldsymbol{x}}_{1[1]})'\widehat{\boldsymbol{\Sigma}}_{(1,3)\cdot 2}^{-1}(\overline{\boldsymbol{x}}_{1[2]} - \boldsymbol{L}_{12}'\boldsymbol{L}_{11}^{-1}\overline{\boldsymbol{x}}_{1[1]}), \\
&\overline{\boldsymbol{x}}_1 = (\overline{\boldsymbol{x}}_{1[1]}', \ \overline{\boldsymbol{x}}_{1[2]}')', \quad \overline{\boldsymbol{x}}_{1[1]} = (\overline{\boldsymbol{x}}_1)_2 : p_{[1]} \times 1, \quad \overline{\boldsymbol{x}}_{1[2]} : p_{[2]} \times 1, \quad \widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{n_1+n_2}\boldsymbol{L}_2, \\
&\widehat{\boldsymbol{\Sigma}}_{(1,3)\cdot 2} = \frac{1}{n_1}\boldsymbol{L}_{13\cdot 1}, \quad \boldsymbol{L}_{13\cdot 1} = \boldsymbol{L}_{13} - \boldsymbol{L}_{12}'\boldsymbol{L}_{11}^{-1}\boldsymbol{L}_{12}.
\end{aligned}$$

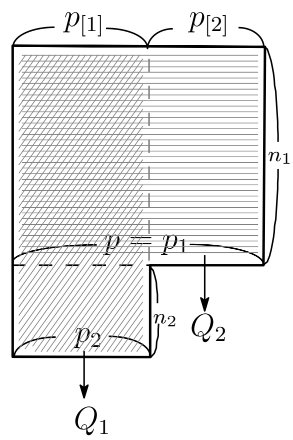

Figure 1. Illustration of 2-step monotone missing data

The decomposition in (3) is similar to the one of the usual T^2 statistic (see, e.g., Fujikoshi et al. (2010), p.49). Note that $\boldsymbol{L}_2 = \boldsymbol{L}_{11} + \boldsymbol{E}_2 + n_1(n_1+n_2)/n_2 \boldsymbol{d}_2\boldsymbol{d}_2'$ is the sum of squares and products matrix based on $(n_1+n_2) \times p_{[1]}$ data block (corresponding of Q_1 in Fig.1) and is distributed as $W_{p_{[1]}}(\boldsymbol{\Sigma}_2, n_1+n_2-1)$. The decomposition in (3) was previously used in Krsihnamoorthy and Pannala (1999). In the two-step case ($k=2$), we consider the null distribution of the simplified T^2 statistic Q when $n_1 \to \infty,\ n_2 \to \infty$ with $r_2 = n_2/n_1 \to \delta \in (0, \infty)$.

At the beginning we consider a stochastic expansion of Q_1.

Let

$$\widehat{\boldsymbol{\mu}}_{[1]} = \frac{1}{\sqrt{n_1+n_2}}\boldsymbol{z}, \quad \frac{1}{n_1+n_2-1}\boldsymbol{L}_2 = \boldsymbol{I} + \frac{1}{\sqrt{n_1+n_2-1}}\boldsymbol{V},$$

Then, we can expand Q_1 as

$$Q_1 = \boldsymbol{z}'\boldsymbol{z} - \frac{1}{\sqrt{n_1+n_2}}\boldsymbol{z}'\boldsymbol{V}\boldsymbol{z} + \frac{1}{n_1+n_2}(\boldsymbol{z}'\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{V}^2\boldsymbol{z}) + O_p((n_1+n_2)^{-\frac{3}{2}}), \tag{4}$$

and

$$\exp(itQ_1) = \exp(it\boldsymbol{z}'\boldsymbol{z})\left[1 + \frac{1}{\sqrt{(1+r_2)n_1}}A + \frac{1}{(1+r_2)n_1}B\right] + O_p(n_1^{-\frac{3}{2}}), \tag{5}$$

where

$$A = (-it)\boldsymbol{z}'\boldsymbol{V}\boldsymbol{z}, \quad B = it(\boldsymbol{z}'\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{V}^2\boldsymbol{z}) + \frac{1}{2}(it)^2(\boldsymbol{z}'\boldsymbol{V}\boldsymbol{z})^2.$$

On the other hand, let

$$\overline{\boldsymbol{x}}_{1[1]} = \frac{1}{\sqrt{n_1}}\boldsymbol{z}_{[1]}, \quad \overline{\boldsymbol{x}}_{1[2]} = \frac{1}{\sqrt{n_1}}\boldsymbol{z}_{[2]},$$

$$\frac{1}{n_1-1}\begin{pmatrix}\boldsymbol{L}_{11} & \boldsymbol{L}_{12}\\ \boldsymbol{L}_{12}' & \boldsymbol{L}_{13}\end{pmatrix} = \begin{pmatrix}\boldsymbol{I}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_{22}\end{pmatrix} + \frac{1}{\sqrt{n_1-1}}\begin{pmatrix}\boldsymbol{V}_{11} & \boldsymbol{V}_{12}\\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22}\end{pmatrix}.$$

In the similar way, we can express Q_2 as

$$\begin{aligned} Q_2 = {} & \boldsymbol{z}_{[2]}'\boldsymbol{z}_{[2]} + \frac{1}{\sqrt{n_1}}(-2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} - \boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{z}_{[2]}) \\ & + \frac{1}{n_1}\Big\{2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{V}_{11}\boldsymbol{z}_{[1]} + 2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} + \boldsymbol{z}_{[1]}'\boldsymbol{V}_{12}\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} + \boldsymbol{z}_{[2]}'\boldsymbol{z}_{[2]} \\ & \quad + \boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{V}_{12}\boldsymbol{z}_{[2]} + \boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}^2\boldsymbol{z}_{[2]}\Big\} + O_p(n_1^{-\frac{3}{2}}). \end{aligned} \tag{6}$$

Therefore, we can express

$$\exp(itQ_2) = \exp(it\boldsymbol{z}_{[2]}'\boldsymbol{z}_{[2]})\left[1 + \frac{1}{\sqrt{n_1}}C + \frac{1}{n_1}D\right] + O_p(n_1^{-\frac{3}{2}}), \tag{7}$$

where

$$\begin{aligned} C = {} & (-it)(2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} + \boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{z}_{[2]}), \\ D = {} & it(2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{V}_{11}\boldsymbol{z}_{[1]} + 2\boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} + \boldsymbol{z}_{[1]}'\boldsymbol{V}_{12}\boldsymbol{V}_{21}\boldsymbol{z}_{[1]} + \boldsymbol{z}_{[2]}'\boldsymbol{z}_{[2]} \\ & + \boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{V}_{12}\boldsymbol{z}_{[2]} + \boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}^2\boldsymbol{z}_{[2]}) \\ & + \frac{1}{2}(it)^2\left\{\, 4(\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{z}_{[1]})^2 + 4(\boldsymbol{z}_{[2]}'\boldsymbol{V}_{21}\boldsymbol{z}_{[1]})(\boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{z}_{[2]}) + (\boldsymbol{z}_{[2]}'\boldsymbol{V}_{22}\boldsymbol{z}_{[2]})^2 \,\right\}. \end{aligned}$$

Further, multiplying (5) by (7), we obtain the characteristic function of Q in (2) as

$$\begin{aligned} \phi(t) = {} & \mathrm{E}\Bigg[\exp(it\boldsymbol{z}'\boldsymbol{z})\exp(it\boldsymbol{z}_{[2]}'\boldsymbol{z}_{[2]}) \\ & \times\left\{1 + \frac{1}{\sqrt{n_1}}\left(C + \frac{1}{\sqrt{1+r_2}}A\right) + \frac{1}{n_1}\left(D + \frac{1}{\sqrt{1+r_2}}AC + \frac{1}{1+r_2}B\right) + O(n_1^{-\frac{3}{2}})\right\}\Bigg]. \end{aligned} \tag{8}$$

We note that Q_1 and Q_2 are not independent, and $\widehat{\boldsymbol{\mu}}_{[1]} = (n_1\overline{\boldsymbol{x}}_{1[1]} + n_2\overline{\boldsymbol{x}}_2)/(n_1+n_2)$. When we put $\boldsymbol{w} = \sqrt{n_2}\,\overline{\boldsymbol{x}}_2$, and substitute $\boldsymbol{z} = (\sqrt{n_1}\boldsymbol{z}_{[1]} + \sqrt{n_2}\boldsymbol{w})/\sqrt{n_1+n_2}$ in (8), we can calculate the characteristic function of Q. Hence, we can obtain

$$\phi(t) = (1-2it)^{-\frac{p}{2}} + \frac{1}{n_1}\sum_{j=0}^{2}\beta_j(1-2it)^{-\frac{p}{2}-j} + O(n_1^{-2}),$$

where

$$\beta_0 = -\frac{1}{4}\left\{s_1 p_{[1]}(p_{[1]}+2) + p_{[2]}(p_{[2]}+2) + 2(s_2+1)p_{[1]}p_{[2]}\right\}, \quad \beta_1 = s_2 p_{[1]}p_{[2]},$$
$$\beta_2 = \frac{1}{4}\left\{s_1 p_{[1]}(p_{[1]}+2) + p_{[2]}(p_{[2]}+2) - 2(s_2-1)p_{[1]}p_{[2]}\right\}, \quad s_i = \frac{n_i}{n_1+n_2}, \quad i = 1, 2.$$

Thus, we have

$$\Pr(Q \le x) = G_p(x) + \frac{1}{n_1}\sum_{j=0}^{2}\beta_j G_{p+2j}(x) + O(n_1^{-2}). \tag{9}$$

Therefore, an asymptotic expansion for the upper percentile of Q up to the order n_1^{-1} is given by

$$q(\alpha) = \chi_p^2(\alpha) - \frac{1}{n_1}\frac{2\chi_p^2(\alpha)}{p}\left\{\beta_0 - \frac{\beta_2}{p+2}\chi_p^2(\alpha)\right\} + O(n_1^{-2}),$$

where $\chi_p^2(\alpha)$ is the upper 100α percentile of χ^2 distribution with p degrees of freedom. Then, we can obtain the asymptotic expansion approximation as follows,

$$q_{\mathrm{AE}}(\alpha) = \chi_p^2(\alpha) - \frac{1}{n_1}\frac{2\chi_p^2(\alpha)}{p}\left\{\beta_0 - \frac{\beta_2}{p+2}\chi_p^2(\alpha)\right\}. \tag{10}$$

3.2 General case ($k > 2$)

In the general case with a k-step monotone missing data, we denote by $p_{[1]}$, $p_{[2]}$, $\cdots$, $p_{[k]}$ the dimensions of proper subblock matrices as presented in Figure 2. Let

$$\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\mu}}'_{[1]},\ \widehat{\boldsymbol{\mu}}'_{[2]}, \ldots,\ \widehat{\boldsymbol{\mu}}'_{[k]})',$$

where $\widehat{\boldsymbol{\mu}}_{[i]} = (\widehat{\mu}_{p_{k-i+2}+1}, \widehat{\mu}_{p_{k-i+2}+2}, \ldots, \widehat{\mu}_{p_{k-i+1}})'$ is a $p_{[i]}\times 1$ subvector of $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_p)'$, $p_{k+1} = 0$. Moreover, let $\boldsymbol{L}_{k-i+1}$, $i = 1, 2, \ldots, k$ be the sum of squares and products matrix of $N_{k-i+2} \times p_{k-i+1}$ data block (see Fig.2). Further, let $\boldsymbol{L}_{k-i+1}$, be partitioned as follows,

$$\underset{p_{k-i+1}\times p_{k-i+1}}{\boldsymbol{L}_{k-i+1}} = \left(\begin{array}{c|c} \overbrace{\boldsymbol{L}_{k-i+1,1}}^{p_{k-i+2}} & \overbrace{\boldsymbol{L}_{k-i+1,2}}^{p_{[i]}} \\ \hline \boldsymbol{L}'_{k-i+1,2} & \boldsymbol{L}_{k-i+1,3} \end{array}\right)\begin{array}{l}\}p_{k-i+2} \\ \}p_{[i]}\end{array} \quad \text{for} \quad i = 2, 3, \ldots, k.$$

Then, the simplified T^2 statistic can be written as

$$Q = \sum_{i=1}^{k} Q_i, \tag{11}$$

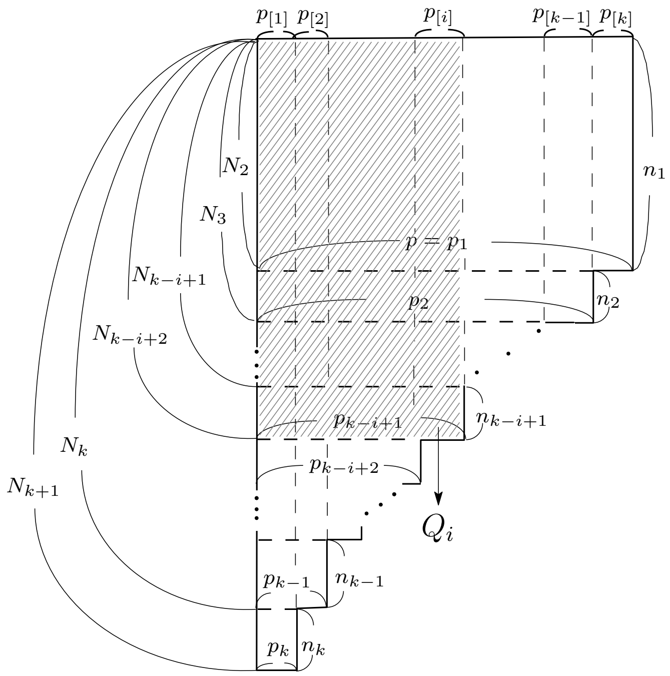

Figure 2. Illustration of Q_i in k-step monotone missing data

where

$$Q_1 = N_{k+1}^2 \widehat{\boldsymbol{\mu}}_{[1]}' \boldsymbol{L}_k^{-1} \widehat{\boldsymbol{\mu}}_{[1]},$$

$$Q_i = N_{k-i+2}^2 (\widehat{\boldsymbol{\mu}}_{[i]} - \boldsymbol{L}_{k-i+1,2}' \boldsymbol{L}_{k-i+1,1}^{-1} \widehat{\boldsymbol{\mu}}_{k-i+2})' \boldsymbol{L}_{k-i+1,3\cdot 1}^{-1} (\widehat{\boldsymbol{\mu}}_{[i]} - \boldsymbol{L}_{k-i+1,2}' \boldsymbol{L}_{k-i+1,1}^{-1} \widehat{\boldsymbol{\mu}}_{k-i+2}),$$

$$\boldsymbol{L}_{k-i+1,3\cdot 1} = \boldsymbol{L}_{k-i+1,3} - \boldsymbol{L}_{k-i+1,2}' \boldsymbol{L}_{k-i+1,1}^{-1} \boldsymbol{L}_{k-i+1,2}, \quad i = 2, 3, \ldots, k.$$

As a result, we have the following theorem.

Theorem 2

For large N_{k-i+2}, the distribution of Q_i, $i = 1, 2, \ldots, k$ can be expressed as

$$\Pr(Q_i \leq x) = G_{p_{[i]}}(x) + \frac{1}{N_{k-i+2}} \sum_{j=0}^{2} \beta_{j,i} G_{p_{[i]}+2j}(x) + O(N_{k-i+2}^{-2}),$$

and also its upper percentiles can be expanded as

$$q_i(\alpha) = \chi_{p_{[i]}}^2(\alpha) - \frac{1}{N_{k-i+2}} \frac{2\chi_{p_{[i]}}^2(\alpha)}{p_{[i]}} \left\{ \beta_{0,i} - \frac{\beta_{2,i}}{p_{[i]}+2} \chi_{p_{[i]}}^2(\alpha) \right\} + O(N_{k-i+2}^{-2}),$$

where

$$\beta_{0,i} = -\frac{1}{4} p_{[i]} \Big(p_{[i]} + 4 \sum_{j=1}^{i-1} p_{[j]} + 2 \Big), \quad \beta_{1,i} = p_{[i]} \sum_{j=1}^{i-1} p_{[j]}, \quad \beta_{2,i} = \frac{1}{4} p_{[i]} (p_{[i]} + 2).$$

For $k=2$, we obtain an asymptotic expansion for the distribution of Q_1 and Q_2 as

$$\Pr(Q_1 \le x) = G_{p_{[1]}}(x) - \frac{p_{[1]}(p_{[1]}+2)}{4N_3}\Big[G_{p_{[1]}}(x) - G_{p_{[1]}+4}(x)\Big] + O(N_3^{-2}),$$
$$\Pr(Q_2 \le x) = G_{p_{[2]}}(x) + \frac{1}{N_2}\sum_{j=0}^{2}\beta_{j,2}G_{p_{[2]}+2j}(x) + O(N_2^{-2}),$$

respectively, where

$$\beta_{0,2} = -\frac{1}{4}p_{[2]}(4p_{[1]}+p_{[2]}+2),\quad \beta_{1,2} = p_{[1]}p_{[2]},\quad \beta_{2,2} = \frac{1}{4}p_{[2]}(p_{[2]}+2).$$

In the general case of k-step monotone missing data, it is difficult to calculate the characteristic function of Q because of the dependence of Q_i's in (11). In this paper, using the property of asymptotically independence of Q_i's, we consider $\prod_{i=1}^{k}\mathrm{E}[\exp(itQ_i)]$ as an approximation of $\phi(t)$. Hence, for large n_1, an approximate expression for the distribution of Q can be written as

$$\Pr(Q \le x) \simeq G_p(x) + \frac{1}{n_1}\sum_{j=0}^{2}\beta_j G_{p+2j}(x), \tag{12}$$

where

$$\beta_0 = -\frac{1}{4}\sum_{i=1}^{k}\frac{1}{m_{k-i+1}}p_{[i]}\Big(p_{[i]} + 4\sum_{j=1}^{i-1}p_{[j]} + 2\Big),\quad \beta_1 = \sum_{i=1}^{k}\frac{1}{m_{k-i+1}}p_{[i]}\sum_{j=1}^{i-1}p_{[j]},$$
$$\beta_2 = \frac{1}{4}\sum_{i=1}^{k}\frac{1}{m_{k-i+1}}p_{[i]}(p_{[i]}+2),\quad m_{k-i+1} = \sum_{j=1}^{k-i+1}r_j.$$

Further, an asymptotic expansion approximation for the upper percentile of Q is given by

$$q_{\mathrm{AE}}(\alpha) = \chi_p^2(\alpha) + \frac{1}{n_1}\frac{\chi_p^2(\alpha)}{2p}\sum_{i=1}^{k}\frac{1}{m_{k-i+1}}p_{[i]}\left\{p_{[i]} + 4\sum_{j=1}^{i-1}p_{[j]} + 2 + \frac{1}{p+2}(p_{[i]}+2)\chi_p^2(\alpha)\right\}. \tag{13}$$

Indeed, we obtained the approximation of the distribution of Q with the assumption that n_i's tend to infinity with $r_i = n_i/n_1 \to \delta_i \in (0,\infty)$, $i = 1,2,\ldots,k$.

4 Transformed test statistics

In this section, we derive the transformations with Bartlett adjustment. A statistic with the Bartlett correction was originally proposed so that its mean was coincident with the one of χ^2 distribution up to the order n^{-1}, where sample size n tends to infinity. Whereas a statistic with the Bartlett-type correction is defined that its distribution is coincident with the one of χ^2 distribution up to the order n^{-1}. In this paper, we present the Bartlett corrections and Bartlett-type corrections. The transformations for the general case were derived by Fujikoshi (2000).

4.1 Two-step case ($k = 2$)

First we consider the transformations for Q_i, $i = 1, 2$. In this case, using the results of Th.2 for $k = 2$, the first moment of Q_i can be obtained as

$$\mathrm{E}[Q_1] = p_{[1]}\left(1 + \frac{p_{[1]} + 2}{n_1 + n_2}\right) + O((n_1+n_2)^{-2}), \quad \mathrm{E}[Q_2] = p_{[2]}\left(1 + \frac{2p_{[1]} + p_{[2]} + 2}{n_1}\right) + O(n_1^{-2}).$$

Therefore, the transformed test statistics with Bartlett corrections are given by

$$Q_1^* = \left\{1 - \frac{1}{n_1 + n_2}(p_{[1]} + 2)\right\} Q_1, \quad Q_2^* = \left\{1 - \frac{1}{n_1}(2p_{[1]} + p_{[2]} + 2)\right\} Q_2,$$

and as an improvement of Q, we propose to take

$$Q^* = Q_1^* + Q_2^*. \tag{14}$$

We note that $\mathrm{E}[Q_i^*] = p_{[i]} + O(N_{4-i}^{-2}), \ i = 1, 2,$ and $\mathrm{E}[Q^*] = p + O(n_1^{-2})$. Further, since the second moments of Q_i can be obtained as

$$\mathrm{E}[Q_1^2] = p_{[1]}(p_{[1]} + 2)\left\{1 + \frac{2(p_{[1]} + 3)}{n_1 + n_2}\right\} + O((n_1+n_2)^{-2}),$$
$$\mathrm{E}[Q_2^2] = p_{[2]}(p_{[2]} + 2)\left\{1 + \frac{2(2p_{[1]} + p_{[2]} + 3)}{n_1}\right\} + O(n_1^{-2}),$$

we have the transformed test statistics with the Bartlett-type corrections given by

$$Y_1 = \left\{n_1 + n_2 - \frac{1}{2}(p_{[1]} + 2)\right\} \log\left(1 + \frac{1}{n_1 + n_2}Q_1\right) \quad \text{for} \ \ n_1 + n_2 - (p_{[1]} + 2)/2 > 0,$$
$$Y_2 = \left\{n_1 - \frac{1}{2}(4p_{[1]} + p_{[2]} + 2)\right\} \log\left(1 + \frac{1}{n_1}Q_2\right) \quad \text{for} \ \ n_1 - (4p_{[1]} + p_{[2]} + 2)/2 > 0,$$
$$Z_1 = \left\{n_1 + n_2 - \frac{1}{2}(p_{[1]} + 2)\right\}\left\{1 - \exp\left(-\frac{1}{n_1 + n_2}Q_1\right)\right\},$$
$$Z_2 = \left\{n_1 - \frac{1}{2}(4p_{[1]} + p_{[2]} + 2)\right\}\left\{1 - \exp\left(-\frac{1}{n_1}Q_2\right)\right\}.$$

We note that $\Pr(Y_i \leq x) = G_{p_{[i]}}(x) + O(N_{4-i}^{-2})$ and $\Pr(Z_i \leq x) = G_{p_{[i]}}(x) + O(N_{4-i}^{-2})$, $i = 1, 2$. Thus, we can propose two transformed test statistics as

$$Y = Y_1 + Y_2 \quad \text{and} \quad Z = Z_1 + Z_2. \tag{15}$$

Moreover using (9), we can propose

$$Q^{\dagger} = \left(1 - \frac{c_1}{n_1}\right) Q, \tag{16}$$

$$Y^{\dagger} = (n_1 a + b) \log\left(1 + \frac{1}{n_1 a} Q\right) \quad \text{for } n_1 a + b > 0, \tag{17}$$

$$Z^{\dagger} = (n_1 a + b)\left\{1 - \exp\left(-\frac{1}{n_1 a} Q\right)\right\}, \tag{18}$$

where

$$\begin{aligned}
c_1 =& \frac{1}{p}\left\{s_1 p_{[1]}(p_{[1]} + 2) + p_{[2]}(p_{[2]} + 2) + 2p_{[1]}p_{[2]}\right\}, \\
a =& p(p+2)\left\{s_1 p_{[1]}(p_{[1]} + 2) + p_{[2]}(p_{[2]} + 2) + 2(1 - s_2)p_{[1]}p_{[2]}\right\}^{-1}, \\
b =& -\frac{p+2}{2}\left\{s_1 p_{[1]}(p_{[1]} + 2) + p_{[2]}(p_{[2]} + 2) + 2(1 + s_2)p_{[1]}p_{[2]}\right\} \\
& \times \left\{s_1 p_{[1]}(p_{[1]} + 2) + p_{[2]}(p_{[2]} + 2) + 2(1 - s_2)p_{[1]}p_{[2]}\right\}^{-1}.
\end{aligned}$$

We note that $Q^{\dagger}$ has the Bartlett correction, and $Y^{\dagger}$ and $Z^{\dagger}$ have the Bartlett-type corrections.

4.2 General case ($k > 2$)

We consider the transformations for Q_i, $i = 1, 2, \ldots, k$ in (11). In this case, since

$$\mathrm{E}[Q_i] = p_{[i]}\left(1 + \frac{c_{1,i}}{N_{k-i+2}}\right) + O(N_{k-i+2}^{-2}),$$

where

$$c_{1,i} = p_{[i]} + 2\sum_{j=1}^{i-1} p_{[j]} + 2,$$

the transformed test statistic with Bartlett correction is given by

$$Q_i^{*} = \left\{1 - \frac{1}{N_{k-i+2}}\left(p_{[i]} + 2\sum_{j=1}^{i-1} p_{[j]} + 2\right)\right\} Q_i.$$

As an improvement of Q in (11), we propose

$$Q^{*} = \sum_{i=1}^{k} Q_i^{*}. \tag{19}$$

We note that $\mathrm{E}[Q_i^*] = p_{[i]} + O(n_1^{-2})$ and $\mathrm{E}[Q^*] = p + O(n_1^{-2})$. Further, since

$$\mathrm{E}[Q_i^2] = p_{[i]}(p_{[i]}+2)\left\{1 + \frac{c_{2,i}}{N_{k-i+2}} + O(N_{k-i+2}^{-2})\right\},$$

where

$$c_{2,i} = 2\Big(p_{[i]} + 2\sum_{j=1}^{i-1} p_{[j]} + 3\Big),$$

we have the transformed test statistics with the Bartlett-type corrections given by

$$Y_i = \left\{N_{k-i+2} - \frac{1}{2}\Big(p_{[i]} + 4\sum_{j=1}^{i-1} p_{[j]} + 2\Big)\right\} \log\left(1 + \frac{1}{N_{k-i+2}} Q_i\right)$$

$$\text{for } N_{k-i+2} - \frac{1}{2}\Big(p_{[i]} + 4\sum_{j=1}^{i-1} p_{[j]} + 2\Big) > 0,$$

$$Z_i = \left\{N_{k-i+2} - \frac{1}{2}\Big(p_{[i]} + 4\sum_{j=1}^{i-1} p_{[j]} + 2\Big)\right\}\left\{1 - \exp\left(-\frac{1}{N_{k-i+2}} Q_i\right)\right\}.$$

We note that $\Pr(Y_i \le x) = G_{p_{[i]}}(x) + O(N_{k-i+2}^{-2})$ and $\Pr(Z_i \le x) = G_{p_{[i]}}(x) + O(N_{k-i+2}^{-2})$. Thus, we can propose two transformed test statistics as

$$Y = \sum_{i=1}^{k} Y_i \quad \text{and} \quad Z = \sum_{i=1}^{k} Z_i. \tag{20}$$

Further, using the result of (12), as with Q^*, Y and Z, we propose

$$Q^\dagger = \left(1 - \frac{c_1}{n_1}\right) Q, \tag{21}$$

$$Y^\dagger = (n_1 a + b)\log\left(1 + \frac{1}{n_1 a} Q\right) \quad \text{for } n_1 a + b > 0, \tag{22}$$

$$Z^\dagger = (n_1 a + b)\left\{1 - \exp\left(-\frac{1}{n_1 a} Q\right)\right\}, \tag{23}$$

where

$$c_1 = \frac{1}{p}\sum_{i=1}^{k} \frac{1}{m_{k-i+1}} p_{[i]}\Big(p_{[i]} + 2\sum_{j=1}^{i-1} p_{[j]} + 2\Big), \quad a = p(p+2)\left\{\sum_{i=1}^{k} \frac{1}{m_{k-i+1}} p_{[i]}(p_{[i]}+2)\right\}^{-1},$$

$$b = -\frac{p+2}{2}\left\{\sum_{i=1}^{k} \frac{1}{m_{k-i+1}} p_{[i]}\Big(p_{[i]} + 4\sum_{j=1}^{i-1} p_{[j]} + 2\Big)\right\}\left\{\sum_{i=1}^{k} \frac{1}{m_{k-i+1}} p_{[i]}(p_{[i]}+2)\right\}^{-1}.$$

We can note that the formulas (16), (17) and (18) and their counterparts in (21), (22) and (23) are the same for $k = 2$ and $k > 2$, but constants a and b are different. Therefore, it holds that $Q^\dagger$ in (16) coincides with $Q^\dagger$ in (21) for $k = 2$.

5 Simulation studies on test statistics and empirical Type I error

In this section, we investigate the numerical accuracy and the asymptotic behavior of the approximate upper percentiles of the test statistic Q. We compare the six transformed test statistics proposed in Section 4, with the original test statistic and asymptotic expansion approximation test. Moreover, we consider Krishnamoorthy and Pannala (1999) approximation test. Their approximation to the upper 100α percentiles of Q is given by

$$q_{\mathrm{KP}}(\alpha) = dF_{p,\nu}(\alpha), \tag{24}$$

where $\nu = \{4pM_2 - 2(p+2)M_1^2\}/\{pM_2 - (p+2)M_1^2\}$, $d = M_1(\nu-2)/\nu$, M_1 and M_2 are the exact first and the approximate second moments for Q, respectively.

We compare the nine procedures with respect to the upper percentiles of test statistics and empirical Type I errors.

Let us denote by the upper 100α percentiles and empirical Type I errors, respectively, as follows

- q and $\alpha_{\chi^2} = 100\,\Pr(Q > \chi_p^2(\alpha))$: for test Q in (2)
- q_{AE} and $\alpha_{\mathrm{AE}} = 100\,\Pr(Q > q_{\mathrm{AE}}(\alpha))$: for asymptotic expansion approximation (AE) test in (10) for $k = 2$ and in (13) for $k > 2$
- q_{KP} and $\alpha_{\mathrm{KP}} = 100\,\Pr(Q > q_{\mathrm{KP}}(\alpha))$: for Krishnamoorthy and Pannala approximation (KP) test in (24)
- q_{Q^*} and $\alpha_{Q^*} = 100\,\Pr(Q^* > \chi_p^2(\alpha))$: for test Q^* with Bartlett correction in (14) for $k = 2$ and in (19) for $k > 2$
- $q_{Q^\dagger}$ and $\alpha_{Q^\dagger} = 100\,\Pr(Q^\dagger > \chi_p^2(\alpha))$: for test $Q^\dagger$ with Bartlett correction in (16) for $k = 2$ and in (21) for $k > 2$
- q_Y and $\alpha_Y = 100\,\Pr(Y > \chi_p^2(\alpha))$: for test Y based on Bartlett-type correction in (15) for $k = 2$ and in (20) for $k > 2$
- q_Z and $\alpha_Z = 100\,\Pr(Z > \chi_p^2(\alpha))$: for test Z based on Bartlett-type correction in (15) for $k = 2$ and in (20) for $k > 2$

– $q_{Y^\dagger}$ and $\alpha_{Y^\dagger} = 100\,\Pr(Y^\dagger > \chi^2_p(\alpha))$: for test $Y^\dagger$ with Bartlett-type correction in (17) for $k = 2$ and based on Bartlett-type correction in (22) for $k > 2$

– $q_{Z^\dagger}$ and $\alpha_{Z^\dagger} = 100\,\Pr(Z^\dagger > \chi^2_p(\alpha))$: for test $Z^\dagger$ with Bartlett-type correction in (18) for $k = 2$ and based on Bartlett-type correction in (23) for $k > 2$

We compute the upper percentiles of these statistics in Monte Carlo simulations, taking into account significance level $\alpha = 0.05$. For different set (k, p_i, n_i), 10^6 of sample forming k-step monotone missing data matrices were generated from $N_{p_i}(\mathbf{0}, \boldsymbol{I}_{p_i})$, $i = 1, 2, \ldots, k$. Computations are carried out for the following two cases:

$$
\begin{aligned}
k = 2\ &: \text{(I)}\ \ (p_1, p_2) = (4, 2),\quad (n_1, n_2) = (m, m),\quad m = 10(10)50, 100, 200, 400.\\
&\ \ \ \text{(II)}\ (p_1, p_2) = (8, 4),\quad (n_1, n_2) = (m, m),\quad m = 20(10)50, 100, 200, 400,\\
k = 3\ &: (p_1, p_2, p_3) = (8, 4, 2),\quad (n_1, n_2, n_3) = (m, m, m),\ (m, \frac{m}{2}, \frac{m}{2}),\ (m, 2m, 2m),\\
&\ \ \ (m, \frac{m}{2}, 2m),\ (m, 2m, \frac{m}{2}),\quad m = 20(10)50, 100, 200, 400, 800.
\end{aligned}
$$

The upper percentiles of tests under consideration were calculated as upper 0.05 quantiles of calculated test statistics, while empirical Type I error was calculated as the ratio of number of rejections of null hypothesis in (1) to 10^6.

Simulation results for $k = 2$ are given in Tables 1–2 and for $k = 3$ in Tables 3–4, respectively. To discuss empirical Type I errors in Tables 1–4 we regard criterion given in Dale (1986). Dale (1986) defined the closeness to normal percentiles considering inequality $|\mathrm{logit}(1-\widehat{\alpha}) - \mathrm{logit}(1-\alpha)| \leq 0.35$ for "close" tests, where $\mathrm{logit}(\alpha) = \log(\alpha/(1-\alpha))$. Note that for significance level $\alpha = 0.05$ the inequality is equivalent to $\widehat{\alpha} \in [0.0357, 0.0695]$. The value of empirical Type I error satisfying this equality are presented in tables in bold.

It may be noted from Tables 1–4 that the upper percentiles of Q, AE and KP tests are decreasing with n_1 and n_2 increasing, and converging to the upper percentile of χ^2 distribution. Moreover, the upper percentiles of KP test is a very good approximation of Q test. The rest of six test under consideration, Q^*, $Q^\dagger$, Y, Z, $Y^\dagger$ and $Z^\dagger$, behave very similar for all cases considered in simulations with respect to upper percentiles and empirical Type I error.

Table 1 : The upper percentiles of test statistics and empirical Type I errors for $(p_1, p_2) = (4, 2)$ and $\alpha = 0.05$.

n_1	n_2	q (α_{χ^2})	q_{AE} (α_{AE})	q_{KP} (α_{KP})	q_{Q^*} (α_{Q^*})	$q_{Q^\dagger}$ ($\alpha_{Q^\dagger}$)	q_Y (α_Y)	q_Z (α_Z)	$q_{Y^\dagger}$ ($\alpha_{Y^\dagger}$)	$q_{Z^\dagger}$ ($\alpha_{Z^\dagger}$)
10	10	26.98	15.33	27.65	9.22	13.49	8.80	7.81	10.81	8.76
		(30.07)	(14.65)	**(4.75)**	**(4.61)**	(10.00)	**(3.73)**	(2.13)	(7.55)	(2.85)
20	20	13.92	12.41	13.92	9.95	10.44	9.45	9.24	9.72	9.47
		(14.10)	(7.06)	**(5.00)**	**(5.83)**	**(6.68)**	**(4.92)**	**(4.49)**	**(5.48)**	**(4.96)**
30	30	12.04	11.44	11.96	9.86	10.04	9.54	9.46	9.60	9.51
		(10.40)	**(5.96)**	**(5.12)**	**(5.72)**	**(6.04)**	**(5.10)**	**(4.95)**	**(5.24)**	**(5.05)**
40	40	11.26	10.95	11.20	9.76	9.85	9.52	9.49	9.54	9.50
		(8.76)	**(5.52)**	**(5.09)**	**(5.51)**	**(5.71)**	**(5.07)**	**(4.99)**	**(5.12)**	**(5.02)**
50	50	10.83	10.66	10.80	9.69	9.75	9.51	9.49	9.51	9.48
		(7.86)	**(5.30)**	**(5.06)**	**(5.41)**	**(5.51)**	**(5.05)**	**(5.01)**	**(5.05)**	**(4.99)**
100	100	10.12	10.07	10.09	9.61	9.62	9.52	9.52	9.50	9.50
		(6.35)	**(5.10)**	**(5.07)**	**(5.24)**	**(5.26)**	**(5.07)**	**(5.06)**	**(5.03)**	**(5.02)**
200	200	9.79	9.78	9.78	9.54	9.55	9.50	9.50	9.49	9.49
		(5.64)	**(5.02)**	**(5.03)**	**(5.12)**	**(5.12)**	**(5.03)**	**(5.03)**	**(5.01)**	**(5.00)**
400	400	9.65	9.63	9.63	9.53	9.53	9.51	9.51	9.50	9.50
		(5.32)	**(5.03)**	**(5.03)**	**(5.08)**	**(5.08)**	**(5.04)**	**(5.04)**	**(5.02)**	**(5.02)**
10	5	27.74	15.83	28.26	9.43	12.95	8.87	7.76	10.35	7.96
		(31.03)	(14.61)	**(4.81)**	**(4.93)**	(9.37)	**(3.88)**	(1.99)	**(6.74)**	(0.00)
20	10	14.24	12.66	14.15	10.02	10.45	9.48	9.25	9.63	9.30
		(14.73)	(7.11)	**(5.11)**	**(5.95)**	**(6.68)**	**(4.98)**	**(4.52)**	**(5.31)**	**(4.59)**
30	15	12.21	11.60	12.11	9.90	10.04	9.55	9.46	9.55	9.42
		(10.76)	**(5.93)**	**(5.15)**	**(5.77)**	**(6.04)**	**(5.12)**	**(4.95)**	**(5.13)**	**(4.86)**
40	20	11.42	11.07	11.31	9.82	9.89	9.56	9.52	9.54	9.47
		(9.08)	**(5.57)**	**(5.17)**	**(5.64)**	**(5.78)**	**(5.15)**	**(5.06)**	**(5.10)**	**(4.97)**
50	25	10.97	10.76	10.88	9.75	9.80	9.55	9.52	9.52	9.48
		(8.17)	**(5.35)**	**(5.13)**	**(5.52)**	**(5.60)**	**(5.13)**	**(5.07)**	**(5.06)**	**(4.98)**
100	50	10.17	10.12	10.13	9.62	9.63	9.52	9.52	9.50	9.49
		(6.43)	**(5.09)**	**(5.08)**	**(5.27)**	**(5.29)**	**(5.07)**	**(5.06)**	**(5.01)**	**(5.00)**
200	100	9.79	9.81	9.80	9.53	9.53	9.48	9.48	9.46	9.46
		(5.63)	**(4.97)**	**(4.98)**	**(5.08)**	**(5.08)**	**(4.98)**	**(4.98)**	**(4.95)**	**(4.94)**
400	200	9.64	9.65	9.64	9.51	9.51	9.49	9.49	9.48	9.48
		(5.33)	**(4.99)**	**(5.00)**	**(5.06)**	**(5.06)**	**(5.01)**	**(5.01)**	**(4.98)**	**(4.98)**
10	20	26.37	14.83	27.13	9.01	14.06	8.72	7.80	11.37	9.69
		(29.16)	(14.84)	**(4.72)**	**(4.32)**	(10.76)	**(3.58)**	(2.18)	(8.46)	**(5.50)**
20	40	13.67	12.16	13.71	9.87	10.48	9.43	9.23	9.85	9.67
		(13.61)	(7.10)	**(4.95)**	**(5.70)**	**(6.75)**	**(4.88)**	**(4.46)**	**(5.72)**	**(5.37)**
30	60	11.84	11.27	11.82	9.77	10.00	9.48	9.41	9.64	9.57
		(9.96)	**(5.91)**	**(5.03)**	**(5.54)**	**(5.97)**	**(4.99)**	**(4.84)**	**(5.30)**	**(5.17)**
40	80	11.14	10.82	11.10	9.72	9.84	9.51	9.48	9.58	9.55
		(8.51)	**(5.54)**	**(5.08)**	**(5.46)**	**(5.69)**	**(5.05)**	**(4.98)**	**(5.19)**	**(5.13)**
50	100	10.72	10.56	10.71	9.65	9.72	9.50	9.47	9.52	9.50
		(7.62)	**(5.29)**	**(5.01)**	**(5.32)**	**(5.45)**	**(5.01)**	**(4.98)**	**(5.06)**	**(5.02)**
100	200	10.05	10.02	10.05	9.57	9.58	9.49	9.48	9.48	9.48
		(6.18)	**(5.05)**	**(5.00)**	**(5.16)**	**(5.18)**	**(5.00)**	**(4.99)**	**(4.99)**	**(4.98)**
200	400	9.75	9.76	9.76	9.52	9.52	9.48	9.48	9.48	9.47
		(5.54)	**(4.99)**	**(4.98)**	**(5.06)**	**(5.07)**	**(4.98)**	**(4.98)**	**(4.97)**	**(4.97)**
400	800	9.62	9.62	9.62	9.51	9.51	9.49	9.49	9.49	9.48
		(5.28)	**(5.00)**	**(5.00)**	**(5.04)**	**(5.05)**	**(5.01)**	**(5.00)**	**(4.99)**	**(4.99)**

Note. $\chi^2_4(0.05) = 9.49$.

Table 2 : The upper percentiles of test statistics and empirical Type I errors for $(p_1, p_2) = (8, 4)$ and $\alpha = 0.05$.

n_1	n_2	q (α_{χ^2})	q_{AE} (α_{AE})	q_{KP} (α_{KP})	q_{Q^*} (α_{Q^*})	$q_{Q^\dagger}$ ($\alpha_{Q^\dagger}$)	q_Y (α_Y)	q_Z (α_Z)	$q_{Y^\dagger}$ ($\alpha_{Y^\dagger}$)	$q_{Z^\dagger}$ ($\alpha_{Z^\dagger}$)
20	20	34.75	23.49	34.73	15.44	19.98	14.62	13.64	17.15	15.36
		(35.61)	(15.12)	(**5.01**)	(**4.91**)	(10.56)	(**3.69**)	(2.38)	(7.68)	(**4.72**)
30	30	24.36	20.83	24.12	16.12	17.46	15.34	14.98	16.14	15.62
		(21.42)	(8.91)	(**5.20**)	(**5.85**)	(7.79)	(**4.74**)	(**4.17**)	(**6.04**)	(**5.18**)
40	40	21.20	19.50	21.01	16.07	16.70	15.50	15.32	15.83	15.58
		(15.79)	(7.09)	(**5.21**)	(**5.87**)	(**6.81**)	(**4.98**)	(**4.70**)	(**5.54**)	(**5.13**)
50	50	19.72	18.70	19.54	15.98	16.36	15.52	15.43	15.71	15.57
		(12.92)	(**6.31**)	(**5.21**)	(**5.74**)	(**6.34**)	(**5.03**)	(**4.86**)	(**5.34**)	(**5.10**)
100	100	17.31	17.10	17.24	15.76	15.84	15.56	15.54	15.54	15.51
		(8.29)	(**5.31**)	(**5.11**)	(**5.42**)	(**5.56**)	(**5.08**)	(**5.05**)	(**5.06**)	(**5.01**)
200	200	16.37	16.31	16.31	15.66	15.67	15.56	15.55	15.53	15.52
		(**6.50**)	(**5.10**)	(**5.09**)	(**5.25**)	(**5.28**)	(**5.08**)	(**5.07**)	(**5.04**)	(**5.02**)
400	400	15.94	15.91	15.90	15.60	15.60	15.55	15.55	15.53	15.53
		(**5.75**)	(**5.06**)	(**5.07**)	(**5.16**)	(**5.16**)	(**5.08**)	(**5.08**)	(**5.04**)	(**5.04**)
20	10	35.70	24.12	35.29	15.77	19.63	14.78	13.68	16.68	14.42
		(36.76)	(15.05)	(**5.19**)	(**5.31**)	(10.02)	(**3.94**)	(2.44)	(**6.94**)	(2.87)
30	15	24.91	21.25	24.49	16.31	17.44	15.43	15.04	15.96	15.26
		(22.32)	(8.93)	(**5.35**)	(**6.12**)	(7.74)	(**4.88**)	(**4.26**)	(**5.75**)	(**4.56**)
40	20	21.63	19.81	21.28	16.23	16.76	15.59	15.40	15.76	15.42
		(16.52)	(7.13)	(**5.35**)	(**6.07**)	(**6.86**)	(**5.13**)	(**4.83**)	(**5.41**)	(**4.85**)
50	25	19.98	18.95	19.76	16.08	16.38	15.58	15.47	15.63	15.43
		(13.40)	(**6.28**)	(**5.25**)	(**5.87**)	(**6.34**)	(**5.11**)	(**4.94**)	(**5.19**)	(**4.87**)
100	50	17.45	17.23	17.34	15.82	15.88	15.59	15.56	15.54	15.49
		(8.49)	(**5.32**)	(**5.16**)	(**5.50**)	(**5.60**)	(**5.13**)	(**5.09**)	(**5.05**)	(**4.98**)
200	100	16.43	16.37	16.36	15.68	15.69	15.57	15.56	15.52	15.51
		(**6.63**)	(**5.09**)	(**5.10**)	(**5.29**)	(**5.31**)	(**5.10**)	(**5.09**)	(**5.02**)	(**5.01**)
400	200	15.96	15.94	15.92	15.60	15.61	15.55	15.55	15.52	15.52
		(**5.77**)	(**5.04**)	(**5.07**)	(**5.15**)	(**5.16**)	(**5.07**)	(**5.06**)	(**5.03**)	(**5.02**)
20	40	34.07	22.85	34.24	15.19	20.44	14.51	13.58	17.77	16.41
		(34.55)	(15.39)	(**4.92**)	(**4.61**)	(11.22)	(**3.54**)	(2.35)	(8.64)	(**6.61**)
30	60	23.87	20.40	23.80	15.94	17.50	15.26	14.92	16.36	15.99
		(20.65)	(9.01)	(**5.05**)	(**5.62**)	(7.94)	(**4.59**)	(**4.06**)	(**6.40**)	(**5.82**)
40	80	20.88	19.18	20.76	15.95	16.71	15.43	15.27	15.97	15.80
		(15.19)	(7.13)	(**5.13**)	(**5.68**)	(**6.83**)	(**4.86**)	(**4.61**)	(**5.76**)	(**5.49**)
50	100	19.44	18.44	19.34	15.89	16.33	15.49	15.40	15.78	15.69
		(12.40)	(**6.31**)	(**5.12**)	(**5.60**)	(**6.29**)	(**4.97**)	(**4.82**)	(**5.46**)	(**5.29**)
100	200	17.17	16.98	17.14	15.71	15.80	15.52	15.50	15.55	15.53
		(7.99)	(**5.28**)	(**5.05**)	(**5.32**)	(**5.47**)	(**5.02**)	(**4.99**)	(**5.08**)	(**5.04**)
200	400	16.28	16.24	16.26	15.61	15.63	15.52	15.51	15.51	15.51
		(**6.34**)	(**5.07**)	(**5.03**)	(**5.17**)	(**5.21**)	(**5.02**)	(**5.01**)	(**5.01**)	(**5.00**)
400	800	15.89	15.87	15.87	15.57	15.57	15.52	15.52	15.51	15.51
		(**5.64**)	(**5.02**)	(**5.02**)	(**5.10**)	(**5.10**)	(**5.02**)	(**5.02**)	(**5.01**)	(**5.00**)

Note. $\chi_8^2(0.05) = 15.51$.

Table 3 : The upper percentiles of test statistics and empirical Type I errors for $(p_1, p_2, p_3) = (8, 4, 2)$ and $\alpha = 0.05$.

n_1	n_2	n_3	q	q_{AE}	q_{KP}	q_{Q^*}	$q_{Q^\dagger}$	q_Y	q_Z	$q_{Y^\dagger}$	$q_{Z^\dagger}$
			(α_{χ^2})	(α_{AE})	(α_{KP})	(α_{Q^*})	$(\alpha_{Q^\dagger})$	(α_Y)	(α_Z)	$(\alpha_{Y^\dagger})$	$(\alpha_{Z^\dagger})$
20	20	20	34.56	22.79	34.54	15.28	20.16	14.63	13.71	18.01	17.16
			(35.38)	(16.03)	**(5.01)**	**(4.72)**	(10.80)	**(3.72)**	(2.51)	(8.79)	(7.67)
30	30	30	24.22	20.36	23.97	16.03	17.49	15.35	15.01	16.59	16.37
			(21.19)	(9.46)	**(5.21)**	**(5.75)**	(7.87)	**(4.76)**	**(4.21)**	**(6.71)**	**(6.40)**
40	40	40	21.06	19.15	20.89	15.97	16.67	15.47	15.31	16.11	16.01
			(15.52)	(7.43)	**(5.19)**	**(5.70)**	**(6.78)**	**(4.94)**	**(4.68)**	**(5.98)**	**(5.83)**
50	50	50	19.61	18.42	19.44	15.94	16.34	15.55	15.46	15.92	15.87
			(12.75)	**(6.59)**	**(5.20)**	**(5.68)**	**(6.29)**	**(5.07)**	**(4.92)**	**(5.68)**	**(5.60)**
100	100	100	17.28	16.96	17.18	15.75	15.84	15.57	15.55	15.66	15.65
			(8.20)	**(5.47)**	**(5.14)**	**(5.40)**	**(5.54)**	**(5.11)**	**(5.08)**	**(5.25)**	**(5.23)**
200	200	200	16.37	16.24	16.29	15.67	15.68	15.58	15.57	15.60	15.60
			(6.47)	**(5.21)**	**(5.12)**	**(5.26)**	**(5.29)**	**(5.12)**	**(5.11)**	**(5.15)**	**(5.14)**
400	400	400	15.89	15.87	15.88	15.56	15.56	15.52	15.51	15.52	15.52
			(5.66)	**(5.03)**	**(5.01)**	**(5.08)**	**(5.09)**	**(5.01)**	**(5.01)**	**(5.02)**	**(5.02)**
800	800	800	15.69	15.69	15.69	15.52	15.52	15.50	15.50	15.50	15.50
			(5.30)	**(5.00)**	**(4.99)**	**(5.02)**	**(5.03)**	**(4.99)**	**(4.99)**	**(4.99)**	**(4.99)**
20	10	10	35.41	23.25	35.06	15.55	19.77	14.78	13.81	17.57	16.59
			(36.51)	(16.17)	**(5.16)**	**(5.06)**	(10.24)	**(3.96)**	(2.68)	(8.11)	**(6.76)**
30	15	15	24.71	20.67	24.30	16.19	17.44	15.45	15.09	16.47	16.21
			(21.99)	(9.58)	**(5.35)**	**(5.97)**	(7.75)	**(4.91)**	**(4.35)**	**(6.50)**	**(6.13)**
40	20	20	21.43	19.38	21.13	16.12	16.70	15.57	15.41	16.08	15.97
			(16.20)	(7.56)	**(5.31)**	**(5.92)**	**(6.81)**	**(5.10)**	**(4.83)**	**(5.93)**	**(5.75)**
50	25	25	19.89	18.60	19.64	16.05	16.38	15.62	15.52	15.92	15.85
			(13.27)	**(6.69)**	**(5.29)**	**(5.83)**	**(6.36)**	**(5.18)**	**(5.02)**	**(5.69)**	**(5.58)**
100	50	50	17.42	17.06	17.28	15.81	15.88	15.61	15.59	15.68	15.67
			(8.45)	**(5.52)**	**(5.20)**	**(5.50)**	**(5.59)**	**(5.18)**	**(5.15)**	**(5.29)**	**(5.27)**
200	100	100	16.39	16.28	16.33	15.65	15.67	15.55	15.55	15.57	15.57
			(6.55)	**(5.17)**	**(5.09)**	**(5.23)**	**(5.26)**	**(5.08)**	**(5.07)**	**(5.11)**	**(5.10)**
400	200	200	15.94	15.89	15.91	15.58	15.58	15.53	15.53	15.54	15.54
			(5.74)	**(5.07)**	**(5.05)**	**(5.12)**	**(5.13)**	**(5.04)**	**(5.04)**	**(5.05)**	**(5.05)**
800	400	400	15.75	15.70	15.70	15.57	15.57	15.55	15.55	15.55	15.55
			(5.40)	**(5.07)**	**(5.07)**	**(5.10)**	**(5.11)**	**(5.07)**	**(5.07)**	**(5.07)**	**(5.07)**
20	40	40	33.81	22.36	34.11	15.06	20.51	14.51	13.61	18.42	17.69
			(34.33)	(16.03)	**(4.86)**	**(4.45)**	(11.38)	**(3.54)**	(2.39)	(9.44)	(8.51)
30	60	60	23.74	20.08	23.69	15.86	17.51	15.24	14.91	16.67	16.50
			(20.33)	(9.34)	**(5.03)**	**(5.51)**	(7.95)	**(4.58)**	**(4.04)**	**(6.85)**	**(6.58)**
40	80	80	20.78	18.93	20.68	15.87	16.69	15.41	15.25	16.17	16.09
			(14.94)	(7.35)	**(5.11)**	**(5.56)**	**(6.81)**	**(4.84)**	**(4.58)**	**(6.06)**	**(5.94)**
50	100	100	19.34	18.25	19.27	15.83	16.30	15.47	15.39	15.92	15.87
			(12.26)	**(6.48)**	**(5.08)**	**(5.52)**	**(6.25)**	**(4.95)**	**(4.81)**	**(5.67)**	**(5.60)**
100	200	200	17.17	16.88	17.10	15.72	15.81	15.55	15.53	15.65	15.64
			(7.99)	**(5.44)**	**(5.10)**	**(5.34)**	**(5.51)**	**(5.07)**	**(5.04)**	**(5.24)**	**(5.22)**
200	400	400	16.26	16.19	16.24	15.60	15.62	15.52	15.52	15.54	15.54
			(6.32)	**(5.11)**	**(5.03)**	**(5.15)**	**(5.19)**	**(5.02)**	**(5.02)**	**(5.06)**	**(5.06)**
400	800	800	15.87	15.85	15.86	15.56	15.56	15.52	15.52	15.52	15.52
			(5.61)	**(5.04)**	**(5.01)**	**(5.08)**	**(5.09)**	**(5.02)**	**(5.02)**	**(5.02)**	**(5.02)**
800	1600	1600	15.68	15.68	15.68	15.53	15.53	15.51	15.51	15.51	15.51
			(5.30)	**(5.01)**	**(5.00)**	**(5.04)**	**(5.04)**	**(5.01)**	**(5.01)**	**(5.01)**	**(5.01)**

Note. $\chi^2_8(0.05) = 15.51$.

Table 4 : The upper percentiles of test statistics and empirical Type I errors for $(p_1, p_2, p_3) = (8, 4, 2)$, $\alpha = 0.05$, and $n_2 \neq n_3$.

n_1	n_2	n_3	q (α_{χ^2})	q_{AE} (α_{AE})	q_{KP} (α_{KP})	q_{Q^*} (α_{Q^*})	$q_{Q^\dagger}$ ($\alpha_{Q^\dagger}$)	q_Y (α_Y)	q_Z (α_Z)	$q_{Y^\dagger}$ ($\alpha_{Y^\dagger}$)	$q_{Z^\dagger}$ ($\alpha_{Z^\dagger}$)
20	10	40	35.08	23.03	34.88	15.40	19.96	14.68	13.74	17.80	16.90
			(36.08)	(16.19)	(**5.09**)	(**4.87**)	(10.51)	(**3.82**)	(2.57)	(8.47)	(7.23)
30	15	60	24.45	20.53	24.18	16.05	17.42	15.35	15.00	16.50	16.27
			(21.68)	(9.52)	(**5.22**)	(**5.77**)	(7.77)	(**4.75**)	(**4.20**)	(**6.57**)	(**6.23**)
40	20	80	21.27	19.27	21.04	16.03	16.69	15.51	15.36	16.10	16.00
			(15.95)	(7.49)	(**5.24**)	(**5.80**)	(**6.77**)	(**5.01**)	(**4.75**)	(**5.94**)	(**5.77**)
50	25	100	19.75	18.52	19.56	15.95	16.35	15.56	15.47	15.92	15.86
			(13.05)	(**6.63**)	(**5.23**)	(**5.71**)	(**6.30**)	(**5.08**)	(**4.94**)	(**5.66**)	(**5.57**)
100	50	200	17.32	17.01	17.24	15.73	15.82	15.55	15.53	15.64	15.63
			(8.25)	(**5.43**)	(**5.11**)	(**5.38**)	(**5.50**)	(**5.07**)	(**5.04**)	(**5.21**)	(**5.19**)
200	100	400	16.32	16.26	16.31	15.60	15.62	15.51	15.50	15.53	15.53
			(**6.44**)	(**5.10**)	(**5.02**)	(**5.15**)	(**5.19**)	(**5.00**)	(**5.00**)	(**5.04**)	(**5.04**)
400	200	800	15.91	15.88	15.90	15.56	15.56	15.52	15.52	15.52	15.52
			(**5.68**)	(**5.04**)	(**5.01**)	(**5.09**)	(**5.09**)	(**5.02**)	(**5.01**)	(**5.02**)	(**5.02**)
800	400	1600	15.70	15.70	15.70	15.53	15.53	15.51	15.51	15.51	15.51
			(**5.33**)	(**5.01**)	(**5.00**)	(**5.04**)	(**5.04**)	(**5.00**)	(**5.00**)	(**5.00**)	(**5.00**)
20	40	10	34.01	22.45	34.18	15.15	20.49	14.55	13.65	18.36	17.60
			(34.41)	(16.01)	(**4.93**)	(**4.55**)	(11.27)	(**3.62**)	(2.48)	(9.33)	(8.35)
30	60	15	23.84	20.13	23.74	15.90	17.52	15.27	14.94	16.66	16.48
			(20.52)	(9.35)	(**5.09**)	(**5.57**)	(7.92)	(**4.62**)	(**4.09**)	(**6.83**)	(**6.56**)
40	80	20	20.84	18.98	20.71	15.92	16.69	15.44	15.29	16.16	16.07
			(15.07)	(7.36)	(**5.14**)	(**5.62**)	(**6.79**)	(**4.89**)	(**4.65**)	(**6.04**)	(**5.92**)
50	100	25	19.41	18.28	19.30	15.88	16.32	15.51	15.42	15.93	15.88
			(12.36)	(**6.51**)	(**5.13**)	(**5.58**)	(**6.28**)	(**5.00**)	(**4.86**)	(**5.70**)	(**5.62**)
100	200	50	17.21	16.90	17.11	15.75	15.84	15.57	15.55	15.67	15.66
			(8.07)	(**5.47**)	(**5.14**)	(**5.38**)	(**5.54**)	(**5.10**)	(**5.08**)	(**5.27**)	(**5.25**)
200	400	100	16.31	16.20	16.25	15.64	15.66	15.56	15.56	15.58	15.58
			(**6.41**)	(**5.17**)	(**5.10**)	(**5.23**)	(**5.26**)	(**5.09**)	(**5.08**)	(**5.13**)	(**5.12**)
400	800	200	15.89	15.85	15.87	15.57	15.57	15.53	15.53	15.53	15.53
			(**5.65**)	(**5.05**)	(**5.03**)	(**5.10**)	(**5.10**)	(**5.04**)	(**5.04**)	(**5.04**)	(**5.04**)
800	1600	400	15.72	15.68	15.68	15.56	15.56	15.54	15.54	15.54	15.54
			(**5.36**)	(**5.06**)	(**5.06**)	(**5.08**)	(**5.09**)	(**5.05**)	(**5.05**)	(**5.06**)	(**5.06**)

Note. $\chi^2_8(0.05) = 15.51$.

6 Concluding remarks

We have considered the null distribution of the simplified T^2 statistic for testing mean vector when the data matrix is of a monotone missing data pattern. Using the decomposition of the test statistic and their asymptotic expansion results, an asymptotic expansion of null distribution of the simplified T^2 statistic was derived in two-step case. In general k-step case, an approximation to asymptotic expansion of the null distribution of the simplified T^2 statistic was presented. Further, to improve the chi-squared approximation, the transformed test statistics based on the Bartlett adjustment were proposed.

Simulation results show us that all tests for mean vector considered in the paper, with monotone missing data matrix, behave very similar with respect to upper percentiles and empirical Type I error for bigger number of replications. Krishnamoorthy and Pannala approximation test turned out to be a very good approximation of simplified T^2 test, although from theoretical point of view, for missing data set, test Q may not be F distributed. Proposed test based on an asymptotic expansion of the distribution of the simplified T^2 statistic using the Bartlett corrections and Bartlett-type corrections improved the simple chi-squared test. From theoretical point of view, Bartlett-type corrections should give the better test but based on simulation results, we can not confirm this fact. We can also conclude that simulation results gave no grounds to indicate the best test for mean vector with monotone missing data matrix. The transformed test statistics with Bartlett-type correction proposed in this paper are expected to be useful in most cases if the missing data is of monotone type, although it must be checked.

Acknowledgment

The first and second authors' research was partly supported by a Grant-in-Aid for JSPS Fellows (15J00414) and a Grant-in-Aid for Scientific Research (C) (26330050, 17K00058), respectively.

References

[1] Chang, W. -Y. and Richards, D. St. P. (2009). Finite-sample inference with monotone incomplete multivariate normal data, I. *Journal of Multivariate Analysis*, **100**, 1883–1899.

[2] Dale, J. R. (1986). Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *Journal of the Royal Statistical Society. Series B (Methodological)* **48**, 48–59.

[3] Fujikoshi, Y. (2000). Transformations with improved chi-squared approximations. *Journal of Multivariate Analysis*, **72**, 249–263.

[4] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Hoboken, NJ: Wiley.

[5] Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics – Theory and Methods*, **21**, 41–50.

[6] Kawasaki, T. and Seo, T. (2016). Bias correction for T^2 type statistic with two-step monotone missing data. *Statistics*, **50**, 76–88.

[7] Krishnamoorthy, K. and Pannala, M. K. (1999). Confidence estimation of a normal mean vector with incomplete data. *The Canadian Journal of Statistics*, **27**, 395–407.

[8] Seko, N., Yamazaki, A. and Seo, T. (2012). Tests for mean vector with two-step monotone missing data. *SUT Journal of Mathematics*, **48**, 13–36.

[9] Yagi, A. and Seo, T. (2014). A test for mean vector and simultaneous confidence intervals with three-step monotone missing data. *American Journal of Mathematical and Management Sciences*, **33**, 161–175.

[10] Yagi, A. and Seo, T. (2017). Tests for normal mean vectors with monotone incomplete data. *American Journal of Mathematical and Management Sciences*, **36**, 1–20.