Asymptotic second-order consistency for two-stage estimation methodologies and its applications

著者別名	青嶋 誠, 矢田 和善
journal or publication title	Annals of the Institute of Statistical Mathematics
volume	62
number	3
page range	571-600
year	2010-06
権利	(C) The Institute of Statistical Mathematics, Tokyo 2008. The original publication is available at www.springerlink.com
URL	http://hdl.handle.net/2241/114080

doi: 10.1007/s10463-008-0188-y

Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Asymptotic second-order consistency for two-stage estimation methodologies and its applications

Makoto Aoshima · Kazuyoshi Yata

Received: date / Revised: date

Abstract We consider fixed–size estimation for a linear function of means from independent and normally distributed populations having unknown and respective variances. We construct a fixed–width confidence interval with required accuracy about the magnitude of the length and the confidence coefficient. We propose a two–stage estimation methodology having the asymptotic second–order consistency with the required accuracy. The key is the asymptotic second–order analysis about the risk function. We give a variety of asymptotic characteristics about the estimation methodology, such as asymptotic sample size and asymptotic Fisher–information. With the help of the asymptotic second–order analysis, we also explore a number of generalizations and extensions of the two–stage methodology to such as bounded risk point estimation, multiple comparisons among components between the populations, and power analysis in equivalence tests to plan the appropriate sample size for a study.

Keywords Bounded risk · Confidence interval · Efficiency · Equivalence tests · Fisher information · Multiple comparisons · Sample size determination · Second–order consistency · Two–stage estimation

Research of the first author was partially supported by Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 18300092.

M. Aoshima
Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan
E-mail: aoshima@math.tsukuba.ac.jp

K. Yata
Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki 305-8571, Japan
E-mail: yata@math.tsukuba.ac.jp

1 Introduction

Suppose that there exist k independent and normally distributed populations π_i : $N(\mu_i, \sigma_i^2)$, $i = 1, ..., k$, where μ_i's and σ_i^2's are both unknown. Let $X_{i1}, X_{i2}, ...$ be a sequence of independent and identically distributed random variables from each π_i. Having recorded $X_{i1}, ..., X_{in_i}$ for each π_i, let us write $\overline{X}_{in_i} = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $\boldsymbol{n} = (n_1, ..., n_k)$. We are interested in estimating the linear function $\mu = \sum_{i=1}^{k} b_i\mu_i$, where b_i's are known and nonzero scalars. Let $T_{\mathbf{n}} = \sum_{i=1}^{k} b_i\overline{X}_{in_i}$. We want to construct a fixed–width confidence interval such that

$$P_{\boldsymbol{\theta}}(|T_{\mathbf{n}} - \mu| < d) \geq 1 - \alpha \tag{1}$$

for all $\boldsymbol{\theta} = (\mu_1, ..., \mu_k, \sigma_1^2, ..., \sigma_k^2)$, where d (> 0) and $\alpha \in (0, 1)$ are both prespecified. Since

$$P_{\boldsymbol{\theta}}(|T_{\mathbf{n}} - \mu| < d) = G\left(d^2\left(\sum_{i=1}^{k}\frac{b_i^2\sigma_i^2}{n_i}\right)^{-1}\right) \tag{2}$$

with $G(\cdot)$ the cumulative distribution function (c.d.f.) of a chi–square random variable having one degree of freedom (d.f.), requirement (1) is satisfied if

$$d^2\left(\sum_{i=1}^{k}\frac{b_i^2\sigma_i^2}{n_i}\right)^{-1} \geq a, \tag{3}$$

where a is the constant such that $G(a) = 1 - \alpha$. It is easy to see that the sample sizes $\boldsymbol{n}$ which minimize the sum $\sum_{i=1}^{k} n_i$ subject to (3) are given as the smallest integer such that

$$n_i \geq \frac{a}{d^2}|b_i|\sigma_i\sum_{j=1}^{k}|b_j|\sigma_j \quad (= C_i, \text{ say}) \tag{4}$$

for each π_i. However, since σ_i's are unknown, the optimal fixed–sample–sizes C_i's should be estimated by using pilot samples from every π_i. It should be noted from Dantzig (1940) that any fixed–sample–size design cannot claim requirement (1).

Takada and Aoshima (1997) gave a two–stage estimation methodology in the spirit of Stein (1945) to satisfy requirement (1) for all the parameters. For the two–sample problem, see Banerjee (1967), Schwabe (1995) and Takada and Aoshima (1996). However, it tends to be oversampling especially when the pilot sample is fixed small compared to the size of C_i. Later, Takada (2004) gave a modification of the Takada–Aoshima procedure so as to make it *asymptotically second–order efficient*, i.e., $\limsup_{d\to 0} E_{\boldsymbol{\theta}}(N_i - C_i) < \infty$. Such a modification had been created and explored for the one–sample problem and the other problems by Mukhopadhyay and Duggan (1997, 1999), Aoshima and Takada (2000), and Aoshima and Mukhopadhyay (2002) among others. One may refer to Aoshima (2005) for a review of two–stage estimation methodologies.

Here, we summarize a modified two–stage procedure due to Takada (2004): Along the lines of Mukhopadhyay and Duggan (1997, 1999), we assume that there exists a known and positive lower bound $\sigma_{i\star}$ for σ_i such that

$$\sigma_i > \sigma_{i\star}, \quad i = 1, ..., k. \tag{5}$$

(T1) Having m_0 (≥ 4) fixed, define

$$m = \max\left\{ m_0, \left[\frac{a}{d^2} \min_{1\leq i\leq k} |b_i|\sigma_{i\star} \sum_{j=1}^{k} |b_j|\sigma_{j\star} \right] + 1 \right\}, \tag{6}$$

where $[x]$ denotes the largest integer less than x. Take a pilot sample $X_{i1}, ...,$ X_{im} of size m and calculate $S_i^2 = \sum_{j=1}^{m}(X_{ij} - \overline{X}_{im})^2/\nu$ for each π_i, where $\overline{X}_{im} = \sum_{j=1}^{m} X_{ij}/m$ and $\nu = m-1$. Define the total sample size of each π_i by

$$N_i = \max\left\{ m, \left[\frac{u}{d^2} |b_i| S_i \sum_{j=1}^{k} |b_j| S_j \right] + 1 \right\}, \tag{7}$$

where the design constant u is chosen as

$$u = a\left(1 + \frac{a + 2k - 1}{2\nu}\right). \tag{8}$$

Let $\boldsymbol{N} = (N_1, ..., N_k)$.

(T2) Take an additional sample $X_{im+1}, ..., X_{iN_i}$ of size $N_i - m$ from each π_i. By combining the initial sample and the additional sample, calculate $\overline{X}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} X_{ij}$ for each π_i. Finally, construct the fixed–width confidence interval with $T_{\mathbf{N}} = \sum_{i=1}^{k} b_i \overline{X}_{iN_i}$.

Then, it holds as $d \to 0$ that

$$P_{\boldsymbol{\theta}}(|T_{\mathbf{N}} - \mu| < d) \geq 1 - \alpha + o(d^2) \quad \text{for all } \boldsymbol{\theta}.$$

However, the modification in those literatures has as yet been unable to prevent oversampling in two–stage estimation methodologies.

In this paper, we make an improvement on the two–stage procedure so as to make it *asymptotically second–order consistent* with the required accuracy as $d \to 0$, i.e.,

$$P_{\boldsymbol{\theta}}(|T_{\mathbf{N}} - \mu| < d) = 1 - \alpha + o(d^2) \quad \text{for all } \boldsymbol{\theta}. \tag{9}$$

With such an improvement, the required sample size is drastically reduced especially when k is large. The key is the asymptotic second–order analysis about the risk function. In Section 2, we show the asymptotic second–order consistency for such the modified two–stage procedure along with its asymptotic second–order characteristics. Also, we discuss asymptotic Fisher–information in the modified two–stage estimation methodology. In Section 3, with the help of the asymptotic second–order analysis, we explore a number of generalizations and extensions of the modified two–stage methodology to such as bounded risk point estimation, and multiple comparisons among components between the populations. In Section 4, we apply the modified two–stage methodology to power analysis in equivalence tests to plan the appropriate sample size for a study. In Section 5, we report the findings of simulation studies and compare performance of our methodology with those of earlier literatures.

2 Asymptotic second–order consistency

Throughout this section, we write that

$$\tau_\star = \min_{1\le i\le k} |b_i|\sigma_{i\star}\sum_{j=1}^k |b_j|\sigma_{j\star}, \quad f_i = |b_i|\sigma_i\left(\sum_{j=1}^k |b_j|\sigma_j\right)^{-1} \quad (i=1,...,k).$$

Theorem 1 *Choose u in (7) as $u = a(1+\nu^{-1}\hat{s})$ instead of (8), where*

$$\hat{s} = 1 + \frac{(a-1)\sum_{i=1}^k b_i^2 S_i^2 - k\tau_\star}{2(\sum_{i=1}^k |b_i|S_i)^2} \tag{10}$$

with S_i^2's calculated in (T1). Then, the two–stage procedure (6)–(7) is asymptotically second–order consistent as $d\to 0$ as stated in (9).

Proof We have from (2) that

$$\begin{aligned} P_{\boldsymbol\theta}\left(|T_{\mathbf N}-\mu|<d\right) =& E_{\boldsymbol\theta}\left\{G\left(d^2\left(\sum_{i=1}^k \frac{b_i^2\sigma_i^2}{N_i}\right)^{-1}\right)\right\} \\ =& E_{\boldsymbol\theta}\left\{G\left(a\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right)^{-1}\right)\right\}. \end{aligned} \tag{11}$$

Now, let us define a new function as follows. We write

$$g(u_1,...,u_k) = G(av^{-1}), \quad v = f_1u_1^{-1}+\cdots+f_ku_k^{-1} \quad \text{for } u_i>0,\ i=1,...,k.$$

Denoting $G'(w)$, $G''(w)$ for the first and second derivatives of $G(w)$ respectively, one can verify the following expressions of the partial derivatives of $g(u_1,...,u_k)$. For all $1\le i\ne j\le k$, we have that

$$\begin{aligned} \frac{\partial g}{\partial u_i} &= aG'(a/v)f_iv^{-2}u_i^{-2}, \\ \frac{\partial^2 g}{\partial u_i^2} &= a\{aG''(a/v)f_i^2v^{-4}u_i^{-4} + 2G'(a/v)f_i^2v^{-3}u_i^{-4} - 2G'(a/v)f_iv^{-2}u_i^{-3}\}, \\ \frac{\partial^2 g}{\partial u_i\partial u_j} &= a\{aG''(a/v)f_if_jv^{-4}u_i^{-2}u_j^{-2} + 2G'(a/v)f_if_jv^{-3}u_i^{-2}u_j^{-2}\}. \end{aligned}$$

From (11), we use the Taylor expansion to claim that

$$
\begin{aligned}
P_{\boldsymbol{\theta}}(|T_{\mathbf{N}}-\mu|<d) &= E_{\boldsymbol{\theta}}\left\{ g\left(\frac{N_1}{C_1},...,\frac{N_k}{C_k}\right)\right\} \\
&= 1-\alpha + aG'(a)\sum_{i=1}^{k} f_i E_{\boldsymbol{\theta}}\left(\frac{N_i-C_i}{C_i}\right) \\
&\quad +\frac{a}{2}\sum_{i=1}^{k}(aG''(a)f_i^2+2G'(a)f_i^2-2G'(a)f_i)E_{\boldsymbol{\theta}}\left\{\left(\frac{N_i-C_i}{C_i}\right)^2\right\} \\
&\quad +\frac{a}{2}\sum_{i\neq j}(aG''(a)f_if_j+2G'(a)f_if_j)E_{\boldsymbol{\theta}}\left\{\left(\frac{N_i-C_i}{C_i}\right)\left(\frac{N_j-C_j}{C_j}\right)\right\} \\
&\quad +E_{\boldsymbol{\theta}}(\Re), \qquad (12)
\end{aligned}
$$

where

$$
E_{\boldsymbol{\theta}}(\Re) = \frac{1}{6}\sum_{i,j,\ell} E_{\boldsymbol{\theta}}\left\{\frac{\partial^3 g}{\partial u_i \partial u_j \partial u_\ell}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}\left(\frac{N_i-C_i}{C_i}\right)\left(\frac{N_j-C_j}{C_j}\right)\left(\frac{N_\ell-C_\ell}{C_\ell}\right)\right\} \qquad (13)
$$

with suitable random variables ξ_i's between 1 and $N_i/C_i,\ i=1,...,k,\ \boldsymbol{u}=(u_1,...,u_k)$ and $\boldsymbol{\xi}=(\xi_1,...,\xi_k)$. With the help of Lemmas 5 and 6 in Appendix, we obtain the following expansion from (12):

$$
\begin{aligned}
&P_{\boldsymbol{\theta}}(|T_{\mathbf{N}}-\mu|<d) = 1-\alpha \\
&+\frac{aG'(a)}{\nu}\left(s-1+\frac{1}{2}\sum_{i=1}^{k} f_iB_i+\sum_{i=1}^{k} f_i^2 + a\frac{G''(a)}{G'(a)}\sum_{i=1}^{k} f_i^2\right)+o(\nu^{-1}), \qquad (14)
\end{aligned}
$$

where $B_i = C_i^{-1}\nu$ and s is a constant such that $E_{\boldsymbol{\theta}}(\hat{s}) = s+o(1)$. Combining the results that $\sum_{i=1}^{k} f_iB_i = k\tau_\star(\sum_{i=1}^{k}|b_i|\sigma_i)^{-2}+O(d^2)$ and $aG''(a)/G'(a) = (-a-1)/2$ with (14), we claim assertion (9) as $d\to 0$. □

Remark 1 Liu and Wang (2007) gave a three–stage estimation methodology satisfying requirement (9) when $k=2$. In fact, their results are verified under the assumption (3.1), in the literature, that requires known lower bounds such as (5) tacitly.

Remark 2 From Lemma 2 in Takada (2004), the constant u given by (8) is coincident with the one originally given by Takada and Aoshima (1997) upto the order $O(\nu^{-1})$. For the two–stage procedure (6)–(7) with (8), by putting $s=(a+2k-1)/2$ in (14), one has as $d\to 0$ that

$$
\begin{aligned}
&P_{\boldsymbol{\theta}}(|T_{\mathbf{N}}-\mu|<d) = 1-\alpha \\
&+\frac{aG'(a)}{2\nu}\left(a+2k-3+\frac{k\tau_\star+(1-a)\sum_{i=1}^{k} b_i^2\sigma_i^2}{(\sum_{i=1}^{k}|b_i|\sigma_i)^2}\right)+o(d^2) \quad \text{for all } \boldsymbol{\theta}.
\end{aligned}
$$

Note that $\hat{s}<(a+2k-1)/2$ w.p.1. The use of (10) saves more samples when k is large.

Theorem 2 *The two–stage procedure (6)–(7) with (10) has as* $d \to 0$*:*

(i) $E_{\boldsymbol\theta}(N_i - C_i) = (2\tau_\star)^{-1}\{\,|b_i|\sigma_i \sum_{j=1}^k |b_j|\sigma_j + (a-1)f_i \sum_{j=1}^k b_j^2\sigma_j^2 + b_i^2\sigma_i^2\,\} + \frac{1}{2}(1-kf_i) + o(1)$ *for* $i = 1, ..., k,$

(ii) $E_{\boldsymbol\theta}(\sum_{i=1}^k N_i - \sum_{i=1}^k C_i) = (2\tau_\star)^{-1}\{\,(\sum_{i=1}^k |b_i|\sigma_i)^2 + a\sum_{i=1}^k b_i^2\sigma_i^2\,\} + o(1).$

Proof The results are obtained by Lemma 5 in Appendix straightforwardly.

□

Remark 3 Let us consider two cases that the lower bounds $\sigma_{i\star}$'s are misidentified: (i) $\sigma_{i\star}$ is much smaller than the true value of σ_i; (ii) several $\sigma_{i\star}$'s are larger than the true values of σ_i's so that it causes $m > \min_{1\le i\le k} C_i$. For case (i), as observed in Theorem 2, it causes oversampling although requirement (9) is satisfied. For case (ii), the two–stage procedure (6)–(7) with (10) has as $d \to 0$ that

$$P_{\boldsymbol\theta}(|T_{\mathbf N} - \mu| < d) > 1 - \alpha + O(d^2) \quad \text{for all } \boldsymbol\theta.$$

Now, we evaluate the Fisher information in the statistic $T_{\mathbf N}$ that is calculated in (T2) with the constant u given by (10). We write the Fisher information in $T_{\mathbf N}$ about μ as $\mathcal{F}_{T_{\mathbf N}}(\mu)$.

Theorem 3 *The two–stage procedure (6)–(7) with (10) has the Fisher information in* $T_{\mathbf N}$ *as* $d \to 0$*:*

$$\frac{\mathcal{F}_{T_{\mathbf N}}(\mu)}{\mathcal{F}_{T_{\boldsymbol C}}(\mu)} = 1 + \frac{d^2(a+1)\sum_{i=1}^k b_i^2\sigma_i^2}{2a\tau_\star(\sum_{i=1}^k |b_i|\sigma_i)^2} + o(d^2), \tag{15}$$

where $\boldsymbol C = (C_1, ..., C_k)$ *is defined by (4).*

Proof In a way similar to Theorem 2.1 in Mukhopadhyay (2005), we have that

$$\begin{aligned}\mathcal{F}_{T_{\mathbf N}}(\mu) &= E_{\boldsymbol\theta}\left\{\left(\sum_{i=1}^k \frac{b_i^2\sigma_i^2}{N_i}\right)^{-1}\right\}\\ &= E_{\boldsymbol\theta}\left\{\frac{a}{d^2}\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right)^{-1}\right\}.\end{aligned}$$

Then, one has that $\mathcal{F}_{T_{\boldsymbol C}}(\mu) = (\sum_{i=1}^k b_i^2\sigma_i^2/C_i)^{-1} = ad^{-2}$. So, we may write that

$$\frac{\mathcal{F}_{T_{\mathbf N}}(\mu)}{\mathcal{F}_{T_{\boldsymbol C}}(\mu)} = E_{\boldsymbol\theta}\left\{\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right)^{-1}\right\}. \tag{16}$$

From (16), we use the Taylor expansion to claim that

$$\frac{\mathcal{F}_{T_{\mathbf{N}}}(\mu)}{\mathcal{F}_{T_{\boldsymbol{C}}}(\mu)} = 1 + \sum_{i=1}^{k} f_i E_{\boldsymbol{\theta}} \left(\frac{N_i - C_i}{C_i} \right) + \sum_{i=1}^{k} (f_i^2 - f_i) E_{\boldsymbol{\theta}} \left\{ \left(\frac{N_i - C_i}{C_i} \right)^2 \right\}$$
$$+ \sum_{i \neq j} f_i f_j E_{\boldsymbol{\theta}} \left\{ \left(\frac{N_i - C_i}{C_i} \right) \left(\frac{N_j - C_j}{C_j} \right) \right\} + E_{\boldsymbol{\theta}}(\Re), \qquad (17)$$

where

$$E_{\boldsymbol{\theta}}(\Re) = \frac{1}{6} \sum_{i,j,\ell} E_{\boldsymbol{\theta}} \left\{ \frac{\partial^3 v^{-1}}{\partial u_i \partial u_j \partial u_\ell} \Big|_{\boldsymbol{u}=\boldsymbol{\xi}} \left(\frac{N_i - C_i}{C_i} \right) \left(\frac{N_j - C_j}{C_j} \right) \left(\frac{N_\ell - C_\ell}{C_\ell} \right) \right\}$$

with $v = \sum_{i=1}^{k} f_i u_i^{-1}$ for $u_i > 0,\ i = 1, ..., k$, suitable random variables ξ_i's between 1 and $N_i/C_i,\ i = 1, ..., k$, $\boldsymbol{u} = (u_1, ..., u_k)$ and $\boldsymbol{\xi} = (\xi_1, ..., \xi_k)$. With the help of Lemmas 5 and 6 in Appendix, we obtain the following expansion from (17):

$$\frac{\mathcal{F}_{T_{\mathbf{N}}}(\mu)}{\mathcal{F}_{T_{\boldsymbol{C}}}(\mu)} = 1 + \nu^{-1} \left(s - 1 + \sum_{i=1}^{k} f_i^2 + \frac{1}{2} \sum_{i=1}^{k} f_i B_i \right) + o(\nu^{-1}), \qquad (18)$$

where $B_i = C_i^{-1} \nu$ and s is a constant such that $E_{\boldsymbol{\theta}}(\hat{s}) = s + o(1)$. Combining the result that $\sum_{i=1}^{k} f_i B_i = k\tau_\star (\sum_{i=1}^{k} |b_i|\sigma_i)^{-2} + O(d^2)$ with (18), we claim assertion (15) as $d \to 0$. □

Remark 4 For simplicity, we let $k = 1$ ($b = 1$). Then, $C = a\sigma^2/d^2$. Under the assumption that $\mathcal{F}_{\overline{X}_N}(\mu)$ exceeds $\mathcal{F}_{\overline{X}_C}(\mu)$ for every fixed (μ, σ^2), Mukhopadhyay (2005) proposed to determine the pilot sample size m for Stein's (1945) two–stage estimation methodology as

$$m = \text{smallest positive integer such that } \ \mathcal{F}_{\overline{X}_N}(\mu)/\mathcal{F}_{\overline{X}_C}(\mu) \le 1 + \varepsilon$$

for a prespecified quantity ε (> 0) which is free from (μ, σ^2). Mukhopadhyay showed that $\mathcal{F}_{\overline{X}_N}(\mu) = \sigma^{-2} E_{\sigma^2}(N)$ and suggested that one may determine the pilot sample size m as

$$m = \text{smallest positive integer such that } \ E_{\sigma^2}(N)/C \le 1 + \varepsilon + o(m^{-1}).$$

Let us write that $E_{\sigma^2}(N)/C = 1 + x/m + o(m^{-1})$ with the design constant $u = a(1 + s/m) + O(m^{-2})$ where x is a constant free from m and $s = (a+1)/2$ for Stein's methodology. If m is completely free from σ^2, we should choose m in order $O(d^c)$ with $c \in (-1, 0)$ in order to specify quantity ε free from σ^2. Then, we have that $x = s$, so that $m = s/\varepsilon$ which is exactly the one given by (3.7) in Mukhopadhyay (2005). Now, let us say $c = -0.5$ and choose m in order $O(d^{-1/2})$. Let us simply write $m = sd^{-1/2}$. Then, we have that $\varepsilon = s/m = d^{1/2}$. When ε is specified as $\varepsilon = 0.1$ (0.01), we have that $d = 10^{-2}$ (10^{-4}), so that C should be very large. It would cause oversampling in the two–stage estimation methodology.

Remark 5 From (15), we have as $d \to 0$ that

$$\mathcal{F}_{T_{\mathbf{N}}}(\mu)/\mathcal{F}_{T_C}(\mu) \le 1 + \varepsilon + o(m^{-1}),$$

with $\varepsilon = (2a\tau_\star)^{-1}(a+1)d^2$. On the other hand, from (18) with $s = (a+2k-1)/2$, which is coincide with the one for Stein's (1945) methodology for $k = 1$, the two–stage procedure (6)–(7) with (8) has the Fisher information in $T_{\mathbf{N}}$ as $d \to 0$:

$$\frac{\mathcal{F}_{T_{\mathbf{N}}}(\mu)}{\mathcal{F}_{T_C}(\mu)} = 1 + \frac{d^2}{2a\tau_\star}\left(a + 2k - 3 + \frac{2\sum_{i=1}^k b_i^2\sigma_i^2 + k\tau_\star}{(\sum_{i=1}^k |b_i|\sigma_i)^2}\right) + o(d^2). \qquad (19)$$

From (19), we have $\varepsilon = (2a\tau_\star)^{-1}(a+3k-1)d^2$. It should be noted that the ε part (redundancy) becomes small when we utilize (10) instead of (8).

Remark 6 If we choose u in (7) as $u = a(1+\nu^{-1}\hat{s})$ with

$$\hat{s} = 1 - \frac{2\sum_{i=1}^k b_i^2 S_i^2 + k\tau_\star}{2(\sum_{i=1}^k |b_i|S_i)^2} \qquad (20)$$

instead of (10), the two–stage procedure (6)–(7) has the Fisher information in $T_{\mathbf{N}}$ as $d \to 0$:

$$\mathcal{F}_{T_{\mathbf{N}}}(\mu)/\mathcal{F}_{T_C}(\mu) = 1 + o(m^{-1}).$$

Then, it holds as $d \to 0$:

(i) $E_{\boldsymbol{\theta}}(N_i - C_i) = (2\tau_\star)^{-1}\{\,|b_i|\sigma_i\sum_{j=1}^k |b_j|\sigma_j - (2\sum_{j=1}^k b_j^2\sigma_j^2 + k\tau_\star)f_i + b_i^2\sigma_i^2\,\}$
$+\frac{1}{2} + o(1)$ for $i = 1, ..., k$,

(ii) $E_{\boldsymbol{\theta}}(\sum_{i=1}^k N_i - \sum_{i=1}^k C_i) = (2\tau_\star)^{-1}\{\,(\sum_{i=1}^k |b_i|\sigma_i)^2 - \sum_{i=1}^k b_i^2\sigma_i^2\} + o(1)$.

3 Applications

3.1 Bounded risk estimation

Suppose that there exist k independent and normally distributed populations $\pi_i :\ N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, ..., k$, where $\boldsymbol{\mu}_i$'s $\in R^p$ and $\boldsymbol{\Sigma}_i$'s are both unknown, but $\boldsymbol{\Sigma}_i$'s are $p \times p$ p.d. matrices. Let $\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, ...$ be a sequence of independent and identically distributed random vectors from each π_i. Having recorded $\boldsymbol{X}_{i1}, ..., \boldsymbol{X}_{in_i}$ for each π_i, let us write $\overline{\boldsymbol{X}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{X}_{ij}/n_i$ and $\boldsymbol{n} = (n_1, ..., n_k)$. We are interested in estimating the linear function $\boldsymbol{\mu} = \sum_{i=1}^k b_i\boldsymbol{\mu}_i$, where b_i's are known and nonzero scalars. Let $\boldsymbol{T}_{\mathbf{n}} = \sum_{i=1}^k b_i\overline{\boldsymbol{X}}_{in_i}$. For a prespecified constant $W\ (>0)$, we want to construct $\boldsymbol{T}_{\mathbf{n}}$ such that

$$E_{\boldsymbol{\theta}}(||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2) \le W \qquad (21)$$

for all $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k)$, where $||\cdot||$ is the Euclidean norm. Since

$$E_{\boldsymbol{\theta}}(||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2) = \sum_{i=1}^{k} b_i^2 \mathrm{tr}(\boldsymbol{\Sigma}_i)/n_i, \tag{22}$$

it is easy to see that the sample sizes $\boldsymbol{n}$ which minimize the sum $\sum_{i=1}^{k} n_i$ subject to (21) are given as the smallest integer such that

$$n_i \geq \frac{1}{W}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^{k}|b_j|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)} \quad (= C_i, \text{ say}) \tag{23}$$

for each π_i.

When $p = 1$, Ghosh et al. (1997, Chap. 6) considered a two–stage estimation methodology to satisfy requirement (21). Later, Aoshima and Takada (2002) considered the present problem and gave a different two–stage estimation methodology. Aoshima and Takada showed that their procedure satisfies requirement (21) with fewer samples than those in Ghosh et al. When applying the asymptotic second–order analysis to the present problem, we make an improvement on the two–stage estimation methodology to hold the asymptotic second–order consistency as $W \to 0$ as stated in (28): We assume that there exists a known and positive lower bound $\sigma_{i\star}$ for $(\mathrm{tr}(\boldsymbol{\Sigma}_i))^{1/2}$ such that

$$\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)} > \sigma_{i\star}, \quad i = 1, ..., k. \tag{24}$$

(T1) Having m_0 (≥ 4) fixed, define

$$m = \max\left\{m_0, \left[\frac{1}{W}\min_{1\leq i\leq k}|b_i|\sigma_{i\star}\sum_{j=1}^{k}|b_j|\sigma_{j\star}\right] + 1\right\}. \tag{25}$$

Take a pilot sample $\boldsymbol{X}_{i1}, ..., \boldsymbol{X}_{im}$ of size m and calculate $\boldsymbol{S}_i = \sum_{j=1}^{m}(\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_{im})(\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_{im})'/\nu$ for each π_i, where $\overline{\boldsymbol{X}}_{im} = \sum_{j=1}^{m}\boldsymbol{X}_{ij}/m$ and $\nu = m - 1$. Define the total sample size of each π_i by

$$N_i = \max\left\{m, \left[\frac{u}{W}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{S}_i)}\sum_{j=1}^{k}|b_j|\sqrt{\mathrm{tr}(\boldsymbol{S}_j)}\right] + 1\right\}, \tag{26}$$

where u is chosen as $u = 1 + \nu^{-1}\hat{s}$ with $\hat{s}$ given by (27). Let $\boldsymbol{N} = (N_1, ..., N_k)$.

(T2) Take an additional sample $\boldsymbol{X}_{im+1}, ..., \boldsymbol{X}_{iN_i}$ of size $N_i - m$ from each π_i. By combining the initial sample and the additional sample, calculate $\overline{\boldsymbol{X}}_{iN_i} = N_i^{-1}\sum_{j=1}^{N_i}\boldsymbol{X}_{ij}$ for each π_i. Finally, estimate $\boldsymbol{\mu}$ by $\boldsymbol{T}_{\mathbf{N}} = \sum_{i=1}^{k} b_i\overline{\boldsymbol{X}}_{iN_i}$.

Theorem 4 *Let* $\tau_\star = \min_{1\le i\le k}|b_i|\sigma_{i\star}\sum_{j=1}^k|b_j|\sigma_{j\star}$*, where* $\sigma_{i\star}$ *is given by (24). Choose* u *in (26) as* $u=1+\nu^{-1}\hat{s}$*, where*

$$\hat{s} = \frac{\sum_{i=1}^k (tr(\boldsymbol{S}_i^2)/(tr(\boldsymbol{S}_i))^2)\left(b_i^2 tr(\boldsymbol{S}_i)+|b_i|\sqrt{tr(\boldsymbol{S}_i)}\sum_{j=1}^k|b_j|\sqrt{tr(\boldsymbol{S}_j)}\right)}{\left(\sum_{i=1}^k|b_i|\sqrt{tr(\boldsymbol{S}_i)}\right)^2} - \frac{k\tau_\star}{2\left(\sum_{i=1}^k|b_i|\sqrt{tr(\boldsymbol{S}_i)}\right)^2} \tag{27}$$

with $\boldsymbol{S}_i$*'s calculated in (T1). Then, the two–stage procedure (25)–(26) is asymptotically second–order consistent as* $W\to 0$*, i.e.,*

$$E_{\boldsymbol{\theta}}(||\boldsymbol{T}_{\mathbf{N}}-\boldsymbol{\mu}||^2) = W + o(W^2) \quad \textit{for all } \boldsymbol{\theta}. \tag{28}$$

Proof We have from (22) that

$$\begin{aligned} E_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{N}}-\boldsymbol{\mu}\|)^2 =& E_{\boldsymbol{\theta}}\left(\sum_{i=1}^k b_i^2 \mathrm{tr}(\boldsymbol{\Sigma}_i)/N_i\right) \\ =& W E_{\boldsymbol{\theta}}\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right), \end{aligned}$$

where $f_i = |b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}/\sum_{j=1}^k|b_j|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)}$. Use the Taylor expansion to claim that

$$\begin{aligned} E_{\boldsymbol{\theta}}\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right) =& 1 - \sum_{i=1}^k f_i E_{\boldsymbol{\theta}}\left(\frac{N_i-C_i}{C_i}\right) + \sum_{i=1}^k f_i E_{\boldsymbol{\theta}}\left\{\left(\frac{N_i-C_i}{C_i}\right)^2\right\} \\ &+ E_{\boldsymbol{\theta}}(\Re), \end{aligned} \tag{29}$$

where $E_{\boldsymbol{\theta}}(\Re) = -\sum_{i=1}^k f_i E_{\boldsymbol{\theta}}\left\{\xi_i^{-4}C_i^{-3}(N_i-C_i)^3\right\}$ with suitable random variables ξ_i's between 1 and N_i/C_i, $i=1,...,k$. One may apply Lemma 6 in Appendix to claim that $E_{\boldsymbol{\theta}}(\Re)=o(\nu^{-1})$ as $W\to 0$. With the help of Remark 18 in Appendix, we obtain the following expansion from (29):

$$\begin{aligned} &E_{\boldsymbol{\theta}}\left(\sum_{i=1}^k f_i\frac{C_i}{N_i}\right) \\ =&1+\frac{1}{2\nu}\sum_{i=1}^k f_i\left(-2s-B_i+A_i\left(f_i+\frac{3}{2}\right)+\sum_{j=1}^k f_jA_j\left(f_j+\frac{1}{2}\right)\right)+o(\nu^{-1}), \end{aligned} \tag{30}$$

where $A_i = \mathrm{tr}(\boldsymbol{\Sigma}_i^2)/(\mathrm{tr}(\boldsymbol{\Sigma}_i))^2$, $B_i=\nu C_i^{-1}$, and s is a constant such that $E_{\boldsymbol{\theta}}(\hat{s}) = s+o(1)$. From (30), we obtain (28) straightforwardly. □

Remark 7 The two–stage procedure (25)–(26) with (27) has as $W\to 0$:

(i) $E_{\boldsymbol{\theta}}(N_i - C_i) = (2\tau_\star)^{-1}\{\frac{3}{2}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^k |b_j|A_j\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)} + b_i^2 A_i \mathrm{tr}(\boldsymbol{\Sigma}_i)$
$+(2\sum_{j=1}^k b_j^2 A_j \mathrm{tr}(\boldsymbol{\Sigma}_j))f_i - \frac{1}{2}|b_i|A_i\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^k |b_j|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)}\}$
$+\frac{1}{2}(1-kf_i) + o(1)$ for $i = 1, ..., k$,

(ii) $E_{\boldsymbol{\theta}}(\sum_{i=1}^k N_i - \sum_{i=1}^k C_i) = (2\tau_\star)^{-1}\{\sum_{i,j}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}|b_j|A_j\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)}$
$+3\sum_{j=1}^k |b_j|A_j\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)}\} + o(1)$.

Remark 8 Aoshima and Takada (2002) gave a two–stage estimation methodology to satisfy requirement (21) without assumption (24). In their methodology, the constant u in (26) is given by $u = \nu/(\nu-2) = 1 + 2/\nu + O(\nu^{-2})$. Then, for the two–stage procedure (25)–(26) with $u = 1 + 2/\nu$, one has from (30) with $s = 2$ that

$$\begin{aligned} &E_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{N}} - \boldsymbol{\mu}\|)^2 \\ &= W + \frac{\tau_\star}{2\nu^2}\Big(2\sum_{i=1}^k f_i(f_i+1)A_i - 4 - k\tau_\star\Big(\sum_{i=1}^k |b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}\Big)^{-2}\Big) \\ &+ o(W^2) \quad \text{for all } \boldsymbol{\theta}. \end{aligned}$$

Note that $\hat{s} < 2$ w.p.1. The use of (27) saves more samples when k is large.

3.2 Multiple comparisons among components

Suppose that there exist k independent and normally distributed populations $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, ..., k$, where $p \geq 2$, and $\boldsymbol{\mu}_i$'s $\in R^p$ and $\boldsymbol{\Sigma}_i$'s are both unknown, but $\boldsymbol{\Sigma}_i = (\sigma_{(i)rs})$ ($> \mathbf{0}$) has a spherical structure such that

$$\sigma_{(i)rr} + \sigma_{(i)ss} - 2\sigma_{(i)rs} = 2\delta_i^2 \quad (1 \leq r < s \leq p) \tag{31}$$

with δ_i (> 0) unknown parameter for each π_i. A special case of such the model is the intraclass correlation model, that is, $\boldsymbol{\Sigma}_i = \sigma_i^2\{(1-\rho_i)\boldsymbol{I}_p + \rho_i \boldsymbol{J}\}$ for some ρ_i, where $\boldsymbol{J}$ denotes a $p \times p$ matrix of all 1's. We consider multiple comparisons experiments for correlated components of $\boldsymbol{\mu} = \sum_{i=1}^k b_i \boldsymbol{\mu}_i$. Let us write $\boldsymbol{\mu} = (\xi_1, ..., \xi_p)$. Similarly to Section 3.1, we use $\boldsymbol{T}_{\mathbf{n}} = \sum_{i=1}^k b_i \overline{\boldsymbol{X}}_{in_i}$ as an estimate of $\boldsymbol{\mu}$. Let us write $\boldsymbol{T}_{\mathbf{n}} = (T_{1\mathbf{n}}, ..., T_{p\mathbf{n}})$. For a prespecified constant d (> 0), we define three types of simultaneous confidence intervals for $(\xi_1, ..., \xi_p)$:

(MCA) $R_{\mathbf{n}} = \{\boldsymbol{\mu}|\ \xi_r - \xi_s \in [T_{r\mathbf{n}} - T_{s\mathbf{n}} - d,\ T_{r\mathbf{n}} - T_{s\mathbf{n}} + d],\ 1 \leq r < s \leq p\}$;

(MCB)
$R_{\mathbf{n}} = \{\boldsymbol{\mu}|\ \xi_r - \max_{s \neq r}\xi_s \in [-(T_{r\mathbf{n}} - \max_{s\neq r} T_{s\mathbf{n}} - d)^-,\ +(T_{r\mathbf{n}} - \max_{s\neq r} T_{s\mathbf{n}} + d)^+],$
$r = 1, ..., p\}$,
where $+x^+ = \max\{0, x\}$ and $-x^- = \min\{0, x\}$;

(MCC) $R_{\mathbf{n}} = \{\boldsymbol{\mu}|\ \xi_r - \xi_p \in [T_{r\mathbf{n}} - T_{p\mathbf{n}} - d,\ T_{r\mathbf{n}} - T_{p\mathbf{n}} + d],\ r = 1, ..., p-1\}$.

For the details of these multiple comparisons methods, see Aoshima and Kushida (2005) and its references. For each of them, for d (>0) and $\alpha \in (0,1)$ both specified, we want to construct $R_{\mathbf{n}}$ such that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{n}}) \geq 1-\alpha \quad \text{for all } \boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k) \tag{32}$$

with $\boldsymbol{\Sigma}_i$'s defined by (31).

It is shown for MCA and MCC that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{n}}) = G_p\left(d^2\left(\sum_{i=1}^k \frac{b_i^2\delta_i^2}{n_i}\right)^{-1}\right),$$

where $G_p(y)$ for $y>0$ is defined by

$$G_p(y) = p\int_{-\infty}^{\infty}\{\Phi(x)-\Phi(x-\sqrt{y})\}^{p-1}d\Phi(x) \quad \text{(for MCA)}, \tag{33}$$

$$G_p(y) = \int_{-\infty}^{\infty}\{\Phi(x+\sqrt{y})-\Phi(x-\sqrt{y})\}^{p-1}d\Phi(x) \quad \text{(for MCC)} \tag{34}$$

with $\Phi(\cdot)$ the c.d.f. of a $N(0,1)$ random variable. It is shown for MCB that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{n}}) \geq G_p\left(d^2\left(\sum_{i=1}^k \frac{b_i^2\delta_i^2}{n_i}\right)^{-1}\right),$$

where

$$G_p(y) = \int_{-\infty}^{\infty}\{\Phi(x+\sqrt{y})\}^{p-1}d\Phi(x). \tag{35}$$

So, the sample sizes $\boldsymbol{n}$ that minimize the sum $\sum_{i=1}^k n_i$ while satisfying requirement (32) are given as the smallest integer such that

$$n_i \geq \frac{a}{d^2}|b_i|\delta_i\sum_{j=1}^k |b_j|\delta_j \quad (= C_i, \text{ say})$$

for each π_i, where a (>0) is a constant such that $G_p(a) = 1-\alpha$ with $G_p(\cdot)$ defined for each method by (33), (34) or (35), respectively.

When applying the asymptotic second–order analysis to this problem, we make an improvement on the two–stage estimation methodology to hold the asymptotic second–order consistency as $d \to 0$ as stated in (40)–(41): We assume that there exists a known and positive lower bound $\sigma_{i\star}$ for δ_i such that

$$\delta_i > \sigma_{i\star}, \quad i = 1, ..., k. \tag{36}$$

(T1) Having m_0 (≥ 4) fixed, define

$$m = \max\left\{m_0, \left[\frac{a}{d^2}\min_{1\leq i\leq k}|b_i|\sigma_{i\star}\sum_{j=1}^k |b_j|\sigma_{j\star}\right]+1\right\}. \tag{37}$$

Take a pilot sample $\boldsymbol{X}_{ij} = (X_{ij1}, ..., X_{ijp}),\ j = 1, ..., m$, and calculate $S_{ip}^2 = \nu_p^{-1} \sum_{r=1}^{p} \sum_{j=1}^{m} (X_{ijr} - \overline{X}_{ij.} - \overline{X}_{i.r} + \overline{X}_{i..})^2$ with $\nu_p = (p-1)(m-1)$ for each π_i. Here, $\overline{X}_{ij.} = p^{-1} \sum_{r=1}^{p} X_{ijr}$, $\overline{X}_{i.r} = m^{-1} \sum_{j=1}^{m} X_{ijr}$ and $\overline{X}_{i..} = (pm)^{-1} \sum_{r=1}^{p} \sum_{j=1}^{m} X_{ijr}$. Note that $\nu_p S_{ip}^2/\delta_i^2$ is distributed as a chi–square distribution with ν_p d.f. Define the total sample size of each π_i by

$$N_i = \max\left\{ m, \left[\frac{u}{d^2} |b_i| S_{ip} \sum_{j=1}^{k} |b_j| S_{jp} \right] + 1 \right\}, \tag{38}$$

where u is chosen as $u = a(1+\nu_p^{-1}\hat{s})$ with a given for each method and $\hat{s}$ given by (39). Let $\boldsymbol{N} = (N_1, ..., N_k)$.

(T2) Take an additional sample $\boldsymbol{X}_{im+1}, ..., \boldsymbol{X}_{iN_i}$ of size $N_i - m$ from each π_i. By combining the initial sample and the additional sample, calculate $\overline{\boldsymbol{X}}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{X}_{ij}$ for each π_i. Finally, for each method, construct $R_{\mathbf{N}}$ with the components $(T_{1\mathbf{N}}, ..., T_{p\mathbf{N}})$ of $\boldsymbol{T}_{\mathbf{N}} = \sum_{i=1}^{k} b_i \overline{\boldsymbol{X}}_{iN_i}$.

The following theorem can be obtained similarly to Theorem 1.

Theorem 5 *Let* $\tau_\star = \min_{1\le i\le k} |b_i|\sigma_{i\star} \sum_{j=1}^{k} |b_j|\sigma_{j\star}$, *where* $\sigma_{i\star}$ *is given by (36). Choose* u *in (38) as* $u = a(1+\nu_p^{-1}\hat{s})$ *with* a *given for each method, where*

$$\hat{s} = 1 - \frac{2(a\frac{G_p''(a)}{G_p'(a)} + 1)\sum_{i=1}^{k} b_i^2 S_{ip}^2 + k(p-1)\tau_\star}{2(\sum_{i=1}^{k} |b_i| S_{ip})^2} \tag{39}$$

with S_{ip}^2*'s calculated in (T1). Then, the two–stage procedure (37)–(38) is asymptotically second–order consistent as* $d \to 0$*, i.e.,*

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{N}}) = 1 - \alpha + o(d^2) \quad \textit{for all } \boldsymbol{\theta} \quad \textit{(MCA, MCC)}; \tag{40}$$

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{N}}) \ge 1 - \alpha + o(d^2) \quad \textit{for all } \boldsymbol{\theta} \quad \textit{(MCB)}. \tag{41}$$

Remark 9 The two–stage procedure (37)–(38) with (39) has as $d \to 0$:

(i) $E_{\boldsymbol{\theta}}(N_i - C_i)$
$= (2(p-1)\tau_\star)^{-1}\{\, |b_i|\delta_i \sum_{j=1}^{k} |b_j|\delta_j - 2\left(a\frac{G_p''(a)}{G_p'(a)} + 1\right) f_i \sum_{j=1}^{k} b_j^2\delta_j^2 + b_i^2\delta_i^2 \,\}$
$+\frac{1}{2}(1 - kf_i) + o(1)$ for $i = 1, ..., k$,

(ii) $E_{\boldsymbol{\theta}}(\sum_{i=1}^{k} N_i - \sum_{i=1}^{k} C_i)$
$= (2(p-1)\tau_\star)^{-1}\{\, (\sum_{i=1}^{k} |b_i|\delta_i)^2 - 2\left(a\frac{G_p''(a)}{G_p'(a)} + \frac{1}{2}\right) \sum_{i=1}^{k} b_i^2\delta_i^2 \,\} + o(1)$,

where $f_i = |b_i|\delta_i(\sum_{j=1}^{k} |b_j|\delta_j)^{-1}$.

Remark 10 The two–stage estimation methodology (37)–(38) was given by Ao-shima and Kushida (2005), but they chose the constant u in (38) as $u =$

$a(1+\nu_p^{-1}s)$ with $s = k-1-aG_p''(a)/G_p'(a)$. For their two–stage procedure, we have as $d \to 0$ that

$$
\begin{aligned}
&P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in R_{\mathbf{N}}) \geq 1-\alpha \\
&+\frac{aG_p'(a)}{\nu_p}\left(k-2+\frac{1}{2}\frac{k(p-1)\tau_\star}{(\sum_{i=1}^k |b_i|\delta_i)^2}+\sum_{i=1}^k f_i^2 + a\frac{G_p''(a)}{G_p'(a)}(\sum_{i=1}^k f_i^2 - 1)\right) \\
&+o(d^2) \quad \text{for all} \ \ \boldsymbol{\theta},
\end{aligned}
$$

where the equality holds for MCA and MCC. For a nominal value of α, note that $aG_p'(a)/G_p(a) \leq -1$. Then, from (39), we have that $\hat{s} < s$ w.p.1. The use of (39) saves more samples when k is large.

4 Testing for equivalence

We consider the problem to test the equivalence of two independent normal populations π_i : $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, with μ_i's and σ_i^2's both unknown. We want to design a test of

$$
H_0 : |\mu| = |\mu_1 - \mu_2| \geq d \quad \text{against} \quad H_a : |\mu| < d \tag{42}
$$

which has size α and power no less than $1-\beta$ at $|\mu| \leq \gamma d$ for all $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, where α, $\beta \in (0,1)$, $\gamma \in [0,1)$, and $d > 0$ (the limit of equivalence) are four prescribed constants. Let us write $\overline{X}_{in_i} = \sum_{j=1}^{n_i} X_{ij}/n_i$, $i = 1, 2$, similarly to Section 1. If σ_i^2's had been known, we would take a sample from each π_i of size

$$
n_i \geq \frac{\delta^2}{d^2}\sigma_i \sum_{j=1}^2 \sigma_j \quad (= C_i, \ \text{say})
$$

and test the hypothesis by

$$
\begin{aligned}
\text{rejecting} \ \ H_0 \iff |\overline{X}_{1n_1} - \overline{X}_{2n_2}| < \left(\sum_{i=1}^2 \frac{\sigma_i^2}{C_i}\right)^{1/2} R\left(d\left(\sum_{i=1}^2 \frac{\sigma_i^2}{C_i}\right)^{-1/2}\right) \\
= \frac{dR(\delta)}{\delta}.
\end{aligned}
$$

Here, the function $R(\cdot)$ is determined uniquely by the equation

$$
P(|N(0,1)+x| < R(x)) = \alpha
$$

with $N(0,1)$ a standard normal random variable, and $\delta = \delta(\alpha, \beta, \gamma)$ is the unique solution of the equation

$$
P(|N(0,1)+\gamma\delta| < R(\delta)) = 1-\beta.
$$

When σ_i^2's are unknown but common ($\sigma_1^2 = \sigma_2^2$), Liu (2003) proposed k (≥ 3)–stage procedure having the size $\alpha+o(n^{-1})$ and the minimum power $1-\beta+o(n^{-1})$. When applying the asymptotic second–order analysis to the present problem, we give a two–stage estimation methodology to hold the asymptotic second–order consistency, which has the accuracy of the same degree as in Liu, as stated in (49): We assume that there exists a known and positive lower bound $\sigma_{i\star}$ for σ_i such that

$$\sigma_i > \sigma_{i\star}, \quad i = 1, 2. \tag{43}$$

(T1) Having m_0 (≥ 4) fixed, define

$$m = \max\left\{ m_0, \left[\frac{\delta^2}{d^2} \min_{1\leq i\leq 2} \sigma_{i\star} \sum_{j=1}^{2} \sigma_{j\star} \right] + 1 \right\}. \tag{44}$$

Take a pilot sample $X_{i1}, ..., X_{im}$ of size m and calculate $S_i^2 = \sum_{j=1}^{m}(X_{ij} - \overline{X}_{im})^2/\nu$ with $\nu = m - 1$ for each π_i. Define the total sample size of each π_i by

$$N_i = \max\left\{ m, \left[\frac{u}{d^2} S_i \sum_{j=1}^{2} S_j \right] + 1 \right\}, \tag{45}$$

where u is chosen as $u = \delta^2(1 + \nu^{-1}\hat{s})$ with $\hat{s}$ given by (47).

(T2) Take an additional sample $X_{im+1}, ..., X_{iN_i}$ of size $N_i - m$ from each π_i. By combining the initial sample and the additional sample, calculate $\overline{X}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} X_{ij}$ for each π_i. Then, test the hypothesis by

$$\text{rejecting } \; H_0 \iff |\overline{X}_{1N_1} - \overline{X}_{2N_2}| < \sqrt{\lambda}\frac{dR(\delta)}{\delta}, \tag{46}$$

where λ is chosen as $\lambda = 1 + \nu^{-1}\hat{t}$ with $\hat{t}$ given by (48).

Theorem 6 *Let $\tau_\star = \min_{1\leq i\leq 2} \sigma_{i\star} \sum_{j=1}^{2} \sigma_{j\star}$, where $\sigma_{i\star}$ is given by (43). Choose u and λ in (45)–(46) as $u = \delta^2(1 + \nu^{-1}\hat{s})$ and $\lambda = 1 + \nu^{-1}\hat{t}$, respectively, with*

$$\hat{s} = 1 + \left(\frac{\varepsilon_1\eta_3 - \eta_1\varepsilon_3}{\varepsilon_1\eta_2 - \eta_1\varepsilon_2} - 1 \right) \frac{\sum_{i=1}^{2} S_i^2}{2(\sum_{i=1}^{2} S_i)^2} - \frac{\tau_\star}{(\sum_{i=1}^{2} S_i)^2}, \tag{47}$$

$$\hat{t} = \frac{\varepsilon_3\eta_2 - \varepsilon_2\eta_3}{2R(\delta)\,(\varepsilon_1\eta_2 - \varepsilon_2\eta_1)} \frac{\sum_{i=1}^{2} S_i^2}{(\sum_{i=1}^{2} S_i)^2}, \tag{48}$$

where S_i^2's are calculated in (T1), $\phi(\cdot)$ is the p.d.f. of $N(0,1)$, and

$$\begin{aligned}
\varepsilon_1 &= \phi(R(\delta)-\delta)+\phi(R(\delta)+\delta),\\
\varepsilon_2 &= (R(\delta)-\delta)\phi(R(\delta)-\delta)+(R(\delta)+\delta)\phi(R(\delta)+\delta),\\
\varepsilon_3 &= (R(\delta)-\delta)^3\phi(R(\delta)-\delta)+(R(\delta)+\delta)^3\phi(R(\delta)+\delta),\\
\eta_1 &= \phi(R(\delta)-\gamma\delta)+\phi(R(\delta)+\gamma\delta),\\
\eta_2 &= (R(\delta)-\gamma\delta)\phi(R(\delta)-\gamma\delta)+(R(\delta)+\gamma\delta)\phi(R(\delta)+\gamma\delta),\\
\eta_3 &= (R(\delta)-\gamma\delta)^3\phi(R(\delta)-\gamma\delta)+(R(\delta)+\gamma\delta)^3\phi(R(\delta)+\gamma\delta).
\end{aligned}$$

Then, the test (46) of (42), with (44)–(45), is asymptotically second–order consistent as $d \to 0$, i.e.,

$$\textit{size} = \alpha + o(d^2) \quad \textit{and} \quad \textit{minimum power} = 1-\beta+o(d^2) \quad \textit{for all } \boldsymbol{\theta}. \tag{49}$$

Proof From (46), we have the size at $|\mu_1-\mu_2| = d$ that

$$\begin{aligned}
&E_{\boldsymbol{\theta}}\left\{\Phi\left((\sqrt{\lambda}R(\delta)-\delta)\left(\sum_{i=1}^2 f_i\frac{C_i}{N_i}\right)^{-1/2}\right)\right\}\\
&\quad - E_{\boldsymbol{\theta}}\left\{\Phi\left(-(\sqrt{\lambda}R(\delta)+\delta)\left(\sum_{i=1}^2 f_i\frac{C_i}{N_i}\right)^{-1/2}\right)\right\}\\
&=\Phi(R(\delta)-\delta)-\Phi(-R(\delta)-\delta)+\frac{R(\delta)t\varepsilon_1}{2\nu}\\
&+\frac{\varepsilon_2}{4\nu}\left(2s-2+\sum_{i=1}^2 f_iB_i+\sum_{i=1}^2 f_i^2\right)-\frac{\varepsilon_3}{4\nu}\sum_{i=1}^2 f_i^2+E_{\boldsymbol{\theta}}(\Re_\alpha)+o(\nu^{-1}),
\end{aligned}\tag{50}$$

where $f_i = \sigma_i(\sum_{j=1}^2\sigma_j)^{-1},\ i=1,2$, and

$$\begin{aligned}
&E_{\boldsymbol{\theta}}(\Re_\alpha)\\
&=\sum_{i=1}^2 E_{\boldsymbol{\theta}}\left\{\frac{\partial^2 g_\alpha}{\partial\lambda\partial u_i}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}(\lambda-1)\left(\frac{N_i-C_i}{C_i}\right)\right\}+\frac{1}{2}E_{\boldsymbol{\theta}}\left\{\frac{\partial^2 g_\alpha}{\partial\lambda^2}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}(\lambda-1)^2\right\}\\
&+\frac{1}{6}\sum_{i,j,\ell}E_{\boldsymbol{\theta}}\left\{\frac{\partial^3 g_\alpha}{\partial u_i\partial u_j\partial u_\ell}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}\left(\frac{N_i-C_i}{C_i}\right)\left(\frac{N_j-C_j}{C_j}\right)\left(\frac{N_\ell-C_\ell}{C_\ell}\right)\right\},
\end{aligned}$$

with

$$\begin{aligned}
g_\alpha(\lambda,u_1,u_2) &= \Phi\left((\lambda^{1/2}R(\delta)-\delta)v^{-1/2}\right)-\Phi\left(-(\lambda^{1/2}R(\delta)+\delta)v^{-1/2}\right),\\
v &= f_1u_1^{-1}+f_2u_2^{-1}
\end{aligned}$$

for $u_i > 0$, $i = 1, 2$. With suitable random variables ξ_λ between 1 and λ and ξ_i's between 1 and N_i/C_i, $i = 1, 2$, $\boldsymbol{u} = (\lambda, u_1, u_2)$ and $\boldsymbol{\xi} = (\xi_\lambda, \xi_1, \xi_2)$. Similarly, we have the minimum power at $|\mu_1 - \mu_2| = \gamma d$ that

$$\begin{aligned}
&E_{\boldsymbol{\theta}}\left\{\Phi\left((\sqrt{\lambda}R(\delta)-\gamma\delta)\left(\sum_{i=1}^{2} f_i\frac{C_i}{N_i}\right)^{-1/2}\right)\right\}\\
&-E_{\boldsymbol{\theta}}\left\{\Phi\left(-(\sqrt{\lambda}R(\delta)+\gamma\delta)\left(\sum_{i=1}^{2} f_i\frac{C_i}{N_i}\right)^{-1/2}\right)\right\}\\
=&\Phi\left(R(\delta)-\gamma\delta\right)-\Phi\left(-R(\delta)-\gamma\delta\right)+\frac{R(\delta)t\eta_1}{2\nu}\\
&+\frac{\eta_2}{4\nu}\left(2s-2+\sum_{i=1}^{2} f_iB_i+\sum_{i=1}^{2} f_i^2\right)-\frac{\eta_3}{4\nu}\sum_{i=1}^{2} f_i^2+E_{\boldsymbol{\theta}}(\Re_\beta)+o(\nu^{-1}), \quad (51)
\end{aligned}$$

where $E_{\boldsymbol{\theta}}(\Re_\beta)$ is defined by replacing $g_\alpha(\lambda, u_1, u_2)$ with

$$g_\beta(\lambda,u_1,u_2)=\Phi\left((\lambda^{1/2}R(\delta)-\gamma\delta)v^{-1/2}\right)-\Phi\left(-(\lambda^{1/2}R(\delta)+\gamma\delta)v^{-1/2}\right)$$

in $E_{\boldsymbol{\theta}}(\Re_\alpha)$. Here, in both (50)–(51), s and t are constants such that $E_{\boldsymbol{\theta}}(\hat{s}) = s + o(1)$ and $E_{\boldsymbol{\theta}}(\hat{t}) = t + o(1)$. One may apply Lemma 6 and Remark 19 in Appendix to claim that $E_{\boldsymbol{\theta}}(\Re_\alpha) = o(\nu^{-1})$ and $E_{\boldsymbol{\theta}}(\Re_\beta) = o(\nu^{-1})$ as $d \to 0$ in (50)–(51). Note that $\Phi(R(\delta) - \delta) - \Phi(-R(\delta) - \delta) = \alpha$ and $\Phi(R(\delta) - \gamma\delta) - \Phi(-R(\delta) - \gamma\delta) = 1 - \beta$. The assertion (49) can be shown straightforwardly. □

Remark 11 When σ_i^2's are unknown but common ($\sigma_1^2 = \sigma_2^2$), define the total sample size as $N_1 = N_2 = \max\{m, [(u/d^2)\sum_{j=1}^{2} S_j^2] + 1\}$. Choose

$$\hat{s}=\frac{1}{4}\left(\frac{\varepsilon_1\eta_3-\eta_1\varepsilon_3}{\varepsilon_1\eta_2-\eta_1\varepsilon_2}+1\right)-\frac{\tau_\star}{2\sum_{i=1}^{2} S_i^2},\quad \hat{t}=\frac{\varepsilon_3\eta_2-\varepsilon_2\eta_3}{4R(\delta)\left(\varepsilon_1\eta_2-\varepsilon_2\eta_1\right)}.$$

Then, the test (46) of (42) is asymptotically second–order consistent as $d \to 0$ as stated in (49).

Remark 12 The two–stage procedure (44)–(45) with (47) has as $d \to 0$:

(i) $E_{\boldsymbol{\theta}}(N_i - C_i) = (2\tau_\star)^{-1}\{\sigma_i\sum_{j=1}^{2}\sigma_j + (s_t - 1)f_i\sum_{j=1}^{2}\sigma_j^2 + \sigma_i^2\} + \frac{1}{2}(1 - 2f_i) + o(1)$ for $i = 1, 2$,

(ii) $E_{\boldsymbol{\theta}}(\sum_{i=1}^{2} N_i - \sum_{i=1}^{2} C_i) = (2\tau_\star)^{-1}\{(\sum_{i=1}^{2}\sigma_i)^2 + s_t\sum_{i=1}^{2}\sigma_i^2\} + o(1)$,

where $s_t = (\varepsilon_1\eta_3 - \eta_1\varepsilon_3)/(\varepsilon_1\eta_2 - \eta_1\varepsilon_2)$. It has the Fisher information in $T_{\mathbf{N}}$ as $d \to 0$:

$$\frac{\mathcal{F}_{T_{\mathbf{N}}}(\mu)}{\mathcal{F}_{T_{\boldsymbol{C}}}(\mu)}=1+\frac{d^2(s_t+1)\sum_{i=1}^{2}\sigma_i^2}{2\delta^2\tau_\star(\sum_{i=1}^{2}\sigma_i)^2}+o(d^2).$$

Remark 13 Let us consider the case that our goal is to design a one–sided equivalence test of

$$H_0 : \mu = \mu_1 - \mu_2 \le -d \quad \text{against} \quad H_a : \mu > -d \tag{52}$$

which has size α and power no less than $1-\beta$ at $\mu \ge -\gamma d$ for all $\boldsymbol{\theta}$. So, one wants to demonstrate that a treatment is no worse than a standard or one treatment is no worse than another treatment in paired comparison by amount d. If σ_i^2's had been known, we would take a sample from each π_i of size

$$n_i \ge \left(\frac{z_\alpha - z_{1-\beta}}{(1-\gamma)d}\right)^2 \sigma_i \sum_{j=1}^{2} \sigma_j,$$

and test the hypothesis by

$$\text{rejecting } H_0 \iff \overline{X}_{1n_1} - \overline{X}_{2n_2} > -d\left(\frac{\gamma z_\alpha - z_{1-\beta}}{z_\alpha - z_{1-\beta}}\right).$$

One may utilize the two–stage procedure for this goal as well. Replace δ^2 with $(z_\alpha - z_{1-\beta})^2/(1-\gamma)^2$ in (44) and in the choice of u of (45). Choose

$$\hat{s} = 1 + (z_\alpha^2 + z_{1-\beta}^2 + z_\alpha z_{1-\beta} - 1)\frac{\sum_{i=1}^2 S_i^2}{2(\sum_{i=1}^2 S_i)^2} - \frac{\tau_\star}{(\sum_{i=1}^2 S_i)^2}, \tag{53}$$
$$\hat{t} = z_\alpha z_{1-\beta}(z_\alpha + z_{1-\beta})\frac{1-\gamma}{\gamma z_\alpha - z_{1-\beta}}\frac{\sum_{i=1}^2 S_i^2}{2(\sum_{i=1}^2 S_i)^2}.$$

Then, the test of (52), given by

$$\text{rejecting} \quad H_0 \iff \overline{X}_{1N_1} - \overline{X}_{2N_2} > -\sqrt{\lambda} d\left(\frac{\gamma z_\alpha - z_{1-\beta}}{z_\alpha - z_{1-\beta}}\right)$$

with $\lambda = 1 + \nu^{-1}\hat{t}$, is asymptotically second–order consistent as $d \to 0$ as stated in (49). Then, it holds as $d \to 0$:

(i) $E_{\boldsymbol{\theta}}(N_i - C_i) = (2\tau_\star)^{-1}\{\sigma_i \sum_{j=1}^2 \sigma_j + (s_o - 1)f_i \sum_{j=1}^2 \sigma_j^2 + \sigma_i^2\} + \frac{1}{2}(1-2f_i) + o(1) \quad \text{for } i = 1, 2,$

(ii) $E_{\boldsymbol{\theta}}(\sum_{i=1}^2 N_i - \sum_{i=1}^2 C_i) = (2\tau_\star)^{-1}\{(\sum_{i=1}^2 \sigma_i)^2 + s_o \sum_{i=1}^2 \sigma_i^2\} + o(1),$

where $s_0 = z_\alpha^2 + z_{1-\beta}^2 + z_\alpha z_{1-\beta}$. It has the Fisher information in $T_{\mathbf{N}}$ as $d \to 0$:

$$\frac{\mathcal{F}_{T_{\mathbf{N}}}(\mu)}{\mathcal{F}_{T_{\mathbf{C}}}(\mu)} = 1 + \frac{(1-\gamma)^2 d^2 (s_o + 1)\sum_{i=1}^2 \sigma_i^2}{2\tau_\star (z_\alpha - z_{1-\beta})^2 (\sum_{i=1}^2 \sigma_i)^2} + o(d^2).$$

Remark 14 Let us consider the case that our goal is to design a two–sided test of

$$H_0 : \mu = \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_a : \mu \neq 0 \tag{54}$$

which has size α and power $1-\beta$ at $|\mu| = d$ for all $\boldsymbol{\theta}$, where $\alpha,\ \beta \in (0,1)$ and $d > 0$ are three prescribed constants. If σ_i^2's had been known, we would take a sample from each π_i of size

$$n_i \geq \frac{c^2(\alpha,\beta)}{d^2}\sigma_i \sum_{j=1}^{2} \sigma_j$$

and test the hypothesis by

$$\text{rejecting } \ H_0 \iff \left|\overline{X}_{1n_1} - \overline{X}_{2n_2}\right| > \frac{dz_{\alpha/2}}{c(\alpha,\beta)},$$

where z_x is the upper x point of $N(0,1)$, and $c(\alpha,\beta)$ (> 0) is the unique solution of the equation

$$P(|N(0,1) + c(\alpha,\beta)| > z_{\alpha/2}) = 1 - \beta.$$

One may utilize the two–stage procedure described above for this goal as well after replacing $(\delta, R(\delta), \gamma)$ with $(c(\alpha,\beta), z_{\alpha/2}, 0)$, respectively, in (44)–(45) and (47)–(48). Then, the test of (54), given by

$$\text{rejecting } \ H_0 \iff \left|\overline{X}_{1n_1} - \overline{X}_{2n_2}\right| > \sqrt{\lambda}\frac{dz_{\alpha/2}}{c(\alpha,\beta)},$$

is asymptotically second–order consistent as $d \to 0$ as stated in (49).

For a one–sided equivalence test of

$$H_0 : \mu = \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_a : \mu < 0 \tag{55}$$

which has size α and power $1-\beta$ at $\mu = -d$ for all $\boldsymbol{\theta}$, we would take a sample from each π_i of size

$$n_i \geq \left(\frac{z_\alpha - z_{1-\beta}}{d}\right)^2 \sigma_i \sum_{j=1}^{2} \sigma_j,$$

and test the hypothesis by

$$\text{rejecting } \ H_0 \iff \overline{X}_{1n_1} - \overline{X}_{2n_2} < -d\left(\frac{z_\alpha}{z_\alpha - z_{1-\beta}}\right).$$

So, replace δ^2 with $(z_\alpha - z_{1-\beta})^2$ in (44) and in the choice of u of (45). Choose $\hat{s}$ as in (53) and choose

$$\hat{t} = -z_{1-\beta}(z_\alpha + z_{1-\beta})\frac{\sum_{i=1}^{2} S_i^2}{2(\sum_{i=1}^{2} S_i)^2}.$$

Then, the test of (55), given by

$$\text{rejecting} \quad H_0 \iff \overline{X}_{1N_1} - \overline{X}_{2N_2} < -\sqrt{\lambda} d \left(\frac{z_\alpha}{z_\alpha - z_{1-\beta}} \right)$$

with $\lambda = 1 + \nu^{-1}\hat{t}$, is asymptotically second–order consistent as $d \to 0$ as stated in (49).

5 Computer simulations

In order to study the performance of our methodology, we take resort to computer simulations. We shall compare our procedure given in Section 2 with the earlier two–stage procedure or the three-stage procedure. We fix $k = 2$ and $(b_1, b_2) = (-1, 1)$. Our goal is to construct 95% fixed–width confidence intervals for $\mu = \mu_1 - \mu_2$. In other words, we have $\alpha = 0.05$ (that is, $a = 3.841$) and we set $d = 0.5$. Let $C = \sum_{i=1}^{2} C_i$. We set $(C_1, C_2) = (40, 60)$, whereas with $C = 100$ one easily obtains from (4) that $(\sigma_1, \sigma_2) = (1.02, 1.53)$. We consider three cases that $m = 10, 20, 30$ ($m_0 = 4$ which is kept fixed throughout) and for each case $(\sigma_{1\star}, \sigma_{2\star})$ are chosen as $\sigma_{1\star}/\sigma_1 = \sigma_{2\star}/\sigma_2 = \sqrt{m/C_1}$. Table 1 examines the performance of the two–stage procedure (6)–(7) with (10) in the first block, the earlier two–stage procedure (6)–(7) with (8) due to Takada (2004) in the second block, and the three–stage procedure due to Liu and Wang (2007, Section 3) with $c = 0.5, 0.7, 0.9$ according to each set of fixed $(\sigma_{1\star}, \sigma_{2\star})$ in the third block.

The findings obtained by averaging the outcomes from 10,000 (= R, say) replications are summarized in each situation. Under a fixed scenario, suppose that the rth replication ends with $N_i = n_{ir}$ $(i = 1, 2)$ observations and the corresponding fixed–width confidence interval $R_{\mathbf{n}_r} = \{\mu \in \boldsymbol{R} : |T_{\mathbf{n}_r} - \mu| < d\}$ based on $\boldsymbol{n}_r = (n_{1r}, n_{2r})$ for $r = 1, ..., R$. Now, $\overline{n}_i = R^{-1}\sum_{r=1}^{R} n_{ir}$ which estimates C_i with its estimated standard error $s(\overline{n}_i)$, where $s^2(\overline{n}_i) = (R^2 - R)^{-1}\sum_{r=1}^{R}(n_{ir} - \overline{n}_i)^2,\ i = 1, 2$. Then, $\overline{n}$ $(= \overline{n}_1 + \overline{n}_2)$ estimates the total fixed sample size C with its estimated standard error $s(\overline{n})$, computed analogously. In the end of the rth replication, we also check whether μ belongs to the constructed confidence interval $R_{\mathbf{n}_r}$ and define $p_r = 1$ (or 0) accordingly as μ does (or does not) belong to $R_{\mathbf{n}_r},\ r = 1, ..., R$. Let $\overline{p} = R^{-1}\sum_{r=1}^{R} p_r$, which estimates the target coverage probability, having its estimated standard error $s(\overline{p})$ where $s^2(\overline{p}) = R^{-1}\overline{p}(1 - \overline{p})$. For the two–stage procedure (6)–(7) with (10), the value of u is given as the average number of the outcomes from 10,000 replications. At the last column, we gave the approximate value of $E_{\boldsymbol{\theta}}(N_i - C_i)$, which was obtained from Theorem 2 in Section 2, from Theorem 3 in Takada (2004), and from Theorem with (3.2) in Liu and Wang (2007), respectively for each procedure.

Let us explain, for example, the entries from the first block for the case when $m = 20$ in Table 1, and hence $(\sigma_{1\star}, \sigma_{2\star}) = (0.72, 1.08)$. From 10,000 independent simulations, we observed $u = 4.152,\ \overline{n}_1 = 43.11,\ s(\overline{n}_1) = 0.106,\ \overline{n}_2 =$

Table 1 Simulated results

		u	$\overline{n}$	$s(\overline{n})$	$\overline{p}$	$s(\overline{p})$	$\overline{n}-C$	$E(N-C)$
$m=10,\ (\sigma_{1\star},\sigma_{2\star})=(0.51,0.77)$								
Two–stage procedure (6)–(7) with (10)								
C	100	4.541	116.02	0.403	0.9482	0.00222	16.02	14.98
C_1	40		46.20	0.167			6.20	5.86
C_2	60		69.82	0.279			9.82	9.13
Two–stage procedure of Takada (2004)								
C	100	5.302	135.16	0.464	0.9584	0.00200	35.16	31.81
C_1	40		53.83	0.193			13.83	12.58
C_2	60		81.33	0.321			21.33	19.22
Three–stage procedure of Liu and Wang (2007) with $c=0.5$								
C	100		106.66	0.228	0.9508	0.00216	6.66	5.84
C_1	40		43.43	0.101			3.43	3.32
C_2	60		63.23	0.148			3.23	2.52
$m=20,\ (\sigma_{1\star},\sigma_{2\star})=(0.72,1.08)$								
Two–stage procedure (6)–(7) with (10)								
C	100	4.152	108.14	0.253	0.9515	0.00215	8.14	7.49
C_1	40		43.11	0.106			3.11	2.98
C_2	60		65.03	0.175			5.03	4.52
Two–stage procedure of Takada (2004)								
C	100	4.533	117.08	0.274	0.9556	0.00206	17.08	15.90
C_1	40		46.79	0.115			6.79	6.34
C_2	60		70.29	0.189			10.29	9.56
Three–stage procedure of Liu and wang (2007) with $c=0.7$								
C	100		104.36	0.179	0.9461	0.00226	4.36	4.17
C_1	40		42.66	0.080			2.66	2.37
C_2	60		61.70	0.116			1.70	1.80
$m=30,\ (\sigma_{1\star},\sigma_{2\star})=(0.88,1.33)$								
Two–stage procedure (6)–(7) with (10)								
C	100	4.031	105.43	0.196	0.9485	0.00221	5.43	5.00
C_1	40		42.32	0.081			2.32	2.02
C_2	60		63.11	0.137			3.11	2.98
Two–stage procedure of Takada (2004)								
C	100	4.295	111.87	0.210	0.9573	0.00202	11.87	10.60
C_1	40		44.83	0.088			4.83	4.26
C_2	60		67.04	0.145			7.04	6.34
Three–stage procedure of Liu and wang (2007) with $c=0.9$								
C	100		105.02	0.161	0.9463	0.00225	5.02	3.25
C_1	40		42.46	0.068			2.46	1.84
C_2	60		62.56	0.107			2.56	1.40

65.03, $s(\overline{n}_2)=0.175$, and $\overline{n}=108.14$, $s(\overline{n})=0.253$. Also, we had $\overline{p}=0.9515$, $s(\overline{p})=0.00215$, and $\overline{n}_1-C_1=3.11$, $\overline{n}_2-C_2=5.03$, $\overline{n}-C=8.14$. At the last column, we had $E(N_1-C_1)=2.98$, $E(N_2-C_2)=4.52$, $E(N-C)=7.49$ where $N=\sum_{i=1}^2 N_i$. Theorem 2 indicates that one may expect $\overline{n}_i-C_i$ to fall in the vicinity of the value of $E(N_i-C_i)$, $i=1,2$. One will observe that the values of $E_{\boldsymbol{\theta}}(N_i-C_i)$ are approximated fairly well by these asymptotic values for small d.

Throughout, the two–stage procedure (6)–(7) with (10) reduces the sample size required in the two–stage procedure due to Takada (2004). When $\sigma_{i\star}$ is specified well, the performance of the two–stage procedure (6)–(7) with (10) can even compare with the performance of the three–stage procedure due to Liu and Wang (2007). If the experimenter considers the cost of each sampling seriously, the two–stage procedure (6)–(7) with (10) might be the most likely candidate in such a real world.

Appendix

Throughout, we write that

$$\tau_i = |b_i|\sigma_i \sum_{j=1}^{k} |b_j|\sigma_j, \quad Y_i = |b_i|S_i \sum_{j=1}^{k} |b_j|S_j$$

for $i = 1, ..., k$. From (4), we write that $C_i = a\tau_i/d^2$. Let d (> 0) go to zero thorough a sequence such that $a\tau_\star/d^2$ always remains an integer. Then, from (6), we may write that $m = a\tau_\star/d^2$. We note that $\nu S_i^2/\sigma_i^2,\ i = 1, ..., k$, are independently distributed as a chi-square distribution with ν d.f. Let $W_i,\ i = 1, ..., k$, denote random variables such that $\nu W_i,\ i = 1, ..., k$, are independently distributed as the chi–square distribution with ν d.f. Let $w_i = W_i - 1$. Then, we have that $S_i^2 = \sigma_i^2(1+w_i)$, and $E(w_i) = 0,\ E(w_i^2) = 2\nu^{-1},\ E(w_i^{2t-1}) = O(\nu^{-t})$ and $E(w_i^{2t}) = O(\nu^{-t}),\ t = 1, 2, ...$

Lemma 1 *For each i, we have as $\nu \to \infty$ that*

$$E_{\boldsymbol{\theta}}(|Y_i - \tau_i|^t) = O(\nu^{-t/2}) \quad (t \geq 2).$$

Proof We write that

$$\begin{aligned} &S_iS_j - \sigma_i\sigma_j \\ &= \sigma_i\sigma_j\{(\sqrt{1 + w_i} - 1)(\sqrt{1 + w_j} - 1) + (\sqrt{1 + w_i} - 1) + (\sqrt{1 + w_j} - 1)\}. \end{aligned}$$

By noting that $E_{\boldsymbol{\theta}}(|(1 + w_i)^{1/2} - 1|^t) = O(\nu^{-t/2})$ ($t \geq 2$), we have that $E_{\boldsymbol{\theta}}(|S_iS_j - \sigma_i\sigma_j|^t) = O(\nu^{-t/2})$ ($t \geq 2$). Hence, it holds that

$$E_{\boldsymbol{\theta}}(|Y_i - \tau_i|^t) = E_{\boldsymbol{\theta}}\left(\left|\sum_{j=1}^{k} |b_i||b_j|(S_iS_j - \sigma_i\sigma_j)\right|^t\right) = O(\nu^{-t/2}) \quad (t \geq 2).$$

The proof is completed. □

Remark 15 As for (26), let $\tau_i = |b_i|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)} \sum_{j=1}^{k} |b_j|\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_j)}$ and $Y_i = |b_i|\sqrt{\mathrm{tr}(\boldsymbol{S}_i)}\ \sum_{j=1}^{k} |b_j|\sqrt{\mathrm{tr}(\boldsymbol{S}_j)}$. Let $W_{ij},\ i = 1, ..., k;\ j = 1, ..., p$, denote random variables such that $\nu W_{ij},\ i = 1, ..., k;\ j = 1, ..., p$, are independently distributed as a chi–square distribution with ν d.f. One may write that $\mathrm{tr}(\boldsymbol{S}_i) = \mathrm{tr}(\boldsymbol{\Sigma}_i) + \sum_{j=1}^{p} \lambda_{ij}(W_{ij} - 1)$, where λ_{ij}'s are latent roots of $\boldsymbol{\Sigma}_i$. Then, we can obtain the same result as in Lemma 1 for (26) as well.

Lemma 2 *For the two–stage procedure (6)–(7) with (10), we have as $d \to 0$ that*

$$E_{\boldsymbol{\theta}}\left(N_i - \left[\frac{u}{d^2}Y_i\right] - 1\right) = O(d).$$

Proof Let $I_{\{N_i=m\}}$ be the indicator function. Then, we have that

$$E_{\boldsymbol{\theta}}\left(N_i-\left[\frac{u}{d^2}Y_i\right]-1\right)=E_{\boldsymbol{\theta}}\left\{I_{\{N_i=m\}}\left(m-\left[\frac{u}{d^2}Y_i\right]-1\right)\right\}$$
$$\leq\sqrt{P_{\boldsymbol{\theta}}(N_i=m)E_{\boldsymbol{\theta}}\left\{\left(m-\left[\frac{u}{d^2}Y_i\right]-1\right)^2\right\}}. \quad (56)$$

Then, it follows that

$$\begin{aligned}P_{\boldsymbol{\theta}}(N_i=m)&=P_{\boldsymbol{\theta}}\left(\frac{uY_i}{d^2}\leq m\right)\\&=P_{\boldsymbol{\theta}}\left(\frac{uY_i}{d^2C_i}-\frac{C_i+1}{C_i}\leq\frac{m-(C_i+1)}{C_i}\right)\\&\leq P_{\boldsymbol{\theta}}\left(\frac{uY_i}{a\tau_i}-1-\frac{1}{C_i}\leq\frac{\tau_\star-\tau_i}{\tau_i}\right)\\&\leq P_{\boldsymbol{\theta}}\left(\left|\frac{uY_i}{a\tau_i}-1\right|+C_i^{-1}\geq\frac{\tau_i-\tau_\star}{\tau_i}\right)\\&\leq\left(\frac{\tau_i-\tau_\star}{\tau_i}\right)^{-6}E_{\boldsymbol{\theta}}\left\{\left(\left|\frac{uY_i}{a\tau_i}-1\right|+C_i^{-1}\right)^6\right\}. \quad (57)\end{aligned}$$

Now, one can yield that

$$E_{\boldsymbol{\theta}}\left\{\left|\frac{uY_i}{a\tau_i}-1\right|^t\right\}\leq E_{\boldsymbol{\theta}}\left\{\left(\frac{1}{\tau_i}\left(|Y_i-\tau_i|+\left|\frac{\hat{s}Y_i}{\nu}\right|\right)\right)^t\right\}=O(\nu^{-t/2})\quad(t\geq2). \quad (58)$$

Here, (58) follows from the result that for any x (≥ 0) and y (≥ 0) such that $x+y=t$ (≥ 2), we have from Lemma 1 that

$$\begin{aligned}E_{\boldsymbol{\theta}}(|Y_i-\tau_i|^x|\nu^{-1}\hat{s}Y_i|^y)\leq&\sqrt{E_{\boldsymbol{\theta}}(|Y_i-\tau_i|^{2x})E_{\boldsymbol{\theta}}(|\nu^{-1}\hat{s}Y_i|^{2y})}\\=&O\left(\nu^{-(x/2+y)}\right)=O\left(\nu^{-(t/2+y/2)}\right).\end{aligned}$$

By combining (58) with (57), we have that

$$P_{\boldsymbol{\theta}}(N_i=m)=O(d^6). \quad (59)$$

The result can be obtained in view of (56) and (59). □

Lemma 3 *Let q (>0) and h (≥ 0) be constants. For a fixed b (≥ 1), let $X_{b\nu}$ denote a chi–square random variable with $b\nu$ d.f. Then, we have as $\nu\to\infty$ that*

$$E(qX_{b\nu}-h-[qX_{b\nu}-h])=\frac{1}{2}+O(\nu^{-1/2}).$$

Proof Let $U = qX_{b\nu} - h - [qX_{b\nu} - h]$. Then, we have for $x \in (0,1)$ and $x_i \in (0,x)$ that

$$P(U \le x) = \sum_{i=0}^{\infty} P(U \le x,\ i \le qX_{b\nu} - h < i+1)$$
$$= \sum_{i=0}^{\infty} P(i \le qX_{b\nu} - h < i+x)$$
$$= \sum_{i=0}^{\infty} \left(F_{b\nu}\left(\frac{i+h+x}{q}\right) - F_{b\nu}\left(\frac{i+h}{q}\right)\right)$$
$$= \frac{x}{q} \sum_{i=0}^{\infty} F'_{b\nu}\left(\frac{i+h+x_i}{q}\right), \tag{60}$$

where $F_{b\nu}(\cdot)$ is the c.d.f. of a chi–square random variable with $b\nu$ d.f., and $F'_{b\nu}(\cdot)$ denotes the first derivative of $F_{b\nu}(\cdot)$. Since $m \ge 4$ and $b \ge 1$, we have that $b\nu \ge 3$. Here, there is at most one constant c $(= b\nu - 2)$ satisfying $\sup_z F'_{b\nu}(z) = F'_{b\nu}(c),\ z > 0$. If $(h+x_i)/q \le b\nu - 2$, there exists integer $i_\star$ such that $(i_\star + h + x_i)/q \le b\nu - 2 < (i_\star + 1 + h + x_i)/q$. Then, we have that

$$\int_i^{i+1} F'_{b\nu}\left(\frac{z+h+x_i}{q}\right) dz \ge \begin{cases} F'_{b\nu}\left(\frac{i+h+x_i}{q}\right) & (i < i_\star), \\ F'_{b\nu}\left(\frac{i+1+h+x_i}{q}\right) & (i \ge i_\star + 1). \end{cases}$$

Hence, it follows that

$$\sum_{i=0}^{\infty} F'_{b\nu}\left(\frac{i+h+x_i}{q}\right) \le \int_{h+x_i}^{\infty} F'_{b\nu}\left(\frac{z}{q}\right) dz + F'_{b\nu}\left(\frac{i_\star+h+x_i}{q}\right)$$
$$\le \int_0^{\infty} F'_{b\nu}\left(\frac{z}{q}\right) dz + \sup_z F'_{b\nu}(z). \tag{61}$$

Similarly, we have that

$$\int_i^{i+1} F'_{b\nu}\left(\frac{z+h+x_i}{q}\right) dz \le \begin{cases} F'_{b\nu}\left(\frac{i+1+h+x_i}{q}\right) & (i < i_\star), \\ F'_{b\nu}\left(\frac{i+h+x_i}{q}\right) & (i \ge i_\star + 1). \end{cases}$$

Hence, it follows that

$$\int_{h+x_i}^{\infty} F'_{b\nu}\left(\frac{z}{q}\right) dz - \sup_z F'_{b\nu}(z) \le \sum_{i=0}^{\infty} F'_{b\nu}\left(\frac{i+h+x_i}{q}\right). \tag{62}$$

If $(h+x_i)/q > b\nu - 2$, we can claim both (61) and (62). Combining (61) and (62) with (60), we have that

$$x - xF_{b\nu}\left(\frac{h+x_i}{q}\right) - \frac{x}{q}\sup_z F'_{b\nu}(z) \le P(U \le x) \le x + \frac{x}{q}\sup_z F'_{b\nu}(z). \tag{63}$$

Here, we note that

$$F_{b\nu}\left(\frac{h+x_i}{q}\right)=\frac{h+x_i}{q}F'_{b\nu}\left(\frac{h'_i}{q}\right)\le\frac{h+x_i}{q}\sup_z F'_{b\nu}(z) \tag{64}$$

with $h'_i\in(0,h+x_i)$, and by Stirling's formula that

$$\sup_z F'_{b\nu}(z)=F'_{b\nu}(b\nu-2)=O(\nu^{-1/2})\quad\text{as}\quad\nu\to\infty. \tag{65}$$

By combining (64) and (65) with (63), we conclude that

$$P(U\le x)=x+O(\nu^{-1/2})\quad\text{as}\quad\nu\to\infty.$$

It completes the proof. □

Lemma 4 *For the two–stage procedure (6)–(7) with (10), we have as* $d\to0$ *that*

$$E_{\boldsymbol\theta}\left\{\frac{u}{d^2}|b_i|S_i\sum_{j=1}^k|b_j|S_j-\left[\frac{u}{d^2}|b_i|S_i\sum_{j=1}^k|b_j|S_j\right]\right\}=\frac12+O(d).$$

Proof Let $X_{k\nu}=\nu\sum_{i=1}^k W_i$ and $V_i=\nu W_i/X_{k\nu}$, $i=1,...,k$. Then, $X_{k\nu}$ is distributed as the chi–square distribution with $k\nu$ d.f., V_i is distributed as the beta distribution with parameters $\nu/2$ and $(k-1)\nu/2$, and $X_{k\nu}$ and $\widetilde V=(V_1,...,V_k)$ are independent. We write $\hat s$ as

$$\hat s=1+\frac{(a-1)b_i^2\sigma_i^2V_i\sum_{j=1}^k b_j^2\sigma_j^2V_j}{2Z_i^2}-\nu\frac{b_i^2\sigma_i^2V_ik\tau_\star}{2X_{k\nu}Z_i^2},$$

where $Z_i=|b_i|\sigma_i\sqrt{V_i}\sum_{j=1}^k|b_j|\sigma_j\sqrt{V_j}$. Then, we have that

$$\frac{u}{d^2}|b_i|S_i\sum_{j=1}^k|b_j|S_j=\frac{u}{d^2\nu}X_{k\nu}Z_i=QX_{k\nu}-H,$$

where

$$Q=\frac{aZ_i}{d^2\nu}\left\{1+\frac1\nu\left(1+\frac{(a-1)b_i^2\sigma_i^2V_i\sum_{j=1}^k b_j^2\sigma_j^2V_j}{2Z_i^2}\right)\right\},\quad H=\frac{ab_i^2\sigma_i^2V_ik\tau_\star}{2d^2Z_i\nu}.$$

Let us define that $U=QX_{k\nu}-H-[QX_{k\nu}-H]$. From Lemma 3, the conditional distribution of U, given $\widetilde V=\widetilde v$ $(H=h,\ Q=q)$, is given for $x\in(0,1)$ that

$$x-\frac{x(1+h+x_i)}{q}\sup_z F'_{k\nu}(z)\le P_{\boldsymbol\theta}(U\le x|\widetilde V=\widetilde v)\le x+\frac{x}{q}\sup_z F'_{k\nu}(z),$$

where $x_i\in(0,x)$. We evaluate that $1/Q\le\tau_\star/Z_i\le\tau_\star/(b_i^2\sigma_i^2V_i)$, and $H/Q\le k\tau_\star/(2\sum_{j=1}^k b_j^2\sigma_j^2V_j)\ \le\ k\tau_\star/(2\min_{1\le i\le k}b_i^2\sigma_i^2)\ (=\ \gamma)$. Then, we have that

$E_{\boldsymbol{\theta}}(1/Q) \le (\tau_\star/b_i^2\sigma_i^2)(k\nu-2)/(\nu-2)$. Here, H/Q is uniformly integrable since $|H/Q| \le \gamma$, and $1/Q$ is uniformly integrable since $|1/Q| \le \tau_\star/(b_i^2\sigma_i^2 V_i)$ with $\tau_\star/(b_i^2\sigma_i^2 V_i)$ being uniformly integrable. From (65), one can yield that

$$E_{\boldsymbol{\theta}}\left\{\frac{x}{Q}\right\}\sup_z F'_{k\nu}(z) \le E_{\boldsymbol{\theta}}\left\{\frac{x(1+H+x_i)}{Q}\right\}\sup_z F'_{k\nu}(z) = O(d).$$

From the fact that $E_{\boldsymbol{\theta}}\{P_{\boldsymbol{\theta}}(U \le x|\widetilde{V} = \widetilde{v})\} = P_{\boldsymbol{\theta}}(U \le x)$, we obtain that

$$P_{\boldsymbol{\theta}}(U \le x) = x + O(d) \quad \text{as} \quad d \to 0. \tag{66}$$

Hence, U is asymptotically uniform on $(0,1)$ as $d \to 0$. The proof is completed. □

Remark 16 For $\hat{s}$ given by (20), (27), (39), (47) and (53), one can write Q and H similar to Lemma 4. Note that, for nominal values of α and β, it holds that $(\varepsilon_1\eta_3 - \eta_1\varepsilon_3)/(\varepsilon_1\eta_2 - \eta_1\varepsilon_2) \ge -1$ in (47) and $G''_p(a)/G'_p(a) < 0$ in (39). Then, we can evaluate that $E_{\boldsymbol{\theta}}(1/Q) = O(1)$ and $E_{\boldsymbol{\theta}}(H/Q) = O(1)$ for Q and H given by each $\hat{s}$. Hence, the result similar to Lemma 4 is obtained for those cases as well.

Remark 17 When the design constant is defined as a constant, the asymptotic uniformity of $P(U \le x)$ was studied by several authors. See Hall (1981) for $k=1$ and Takada (2004) for $k \ge 2$.

Lemma 5 *The two–stage procedure (6)–(7) with (10) has as* $d \to 0$*:*

(i) $E_{\boldsymbol{\theta}}\{C_i^{-1}(N_i - C_i)\} = (2\nu)^{-1}(2s - 1 + f_i + B_i) + O(d^3)$,

(ii) $E_{\boldsymbol{\theta}}\{C_i^{-2}(N_i - C_i)^2\} = (2\nu)^{-1}(1 + 2f_i + \sum_{i'=1}^k f_{i'}^2) + O(d^3)$,

(iii) $E_{\boldsymbol{\theta}}\{C_i^{-1}(N_i - C_i)C_j^{-1}(N_j - C_j)\} = (2\nu)^{-1}(f_i + f_j + \sum_{i'=1}^k f_{i'}^2)$ $+O(d^3)$ *(*$i \ne j$*);*

where $B_i = \nu/C_i$ *and* s *is a constant such that* $E_{\boldsymbol{\theta}}(\hat{s}) = s + o(1)$*.*

Proof Let us write that

$$N_i = rC_iT_i + (1 + [rC_iT_i] - rC_iT_i) + (N_i - [rC_iT_i] - 1),$$

where $r = u/a = 1 + \nu^{-1}\hat{s}$ and $T_i = \tau_i^{-1}Y_i$. Here, from Lemma 4, $U_i = 1 + [rC_iT_i] - rC_iT_i$ is asymptotically distributed as $U(0,1)$. Let $D_i = N_i - [rC_iT_i] - 1$. From Lemma 2, it follows that $E\{(D_i/\nu)^c\} = O(\nu^{-3/2})$ as $d \to 0$, where $c\ (\ge 1)$ is fixed. Then, we have that

$$C_i^{-1}(N_i - C_i) = (rT_i - 1) + \nu^{-1}B_iU_i + C_i^{-1}D_i. \tag{67}$$

By noting that $E_{\boldsymbol{\theta}}(\hat{s}) = s + o(1)$, we obtain the following results:

$$\begin{aligned} &E_{\boldsymbol{\theta}}(rT_i - 1) = (2\nu)^{-1}(2s - 1 + f_i) + O(d^3), \\ &E_{\boldsymbol{\theta}}\{(rT_i - 1)^2\} = (2\nu)^{-1}(1 + 2f_i + \sum_{i'=1}^k f_{i'}^2) + O(d^3), \\ &E_{\boldsymbol{\theta}}\{(rT_i - 1)(rT_j - 1)\} = (2\nu)^{-1}(f_i + f_j + \sum_{i'=1}^k f_{i'}^2) + O(d^3) \quad (i \ne j). \end{aligned} \tag{68}$$

Let us combine these results with the expectations of (67). The results are obtained straightforwardly. □

Remark 18 For the two–stage procedure (25)–(26) with (27), we have as $W \to 0$ that

(i) $E_{\boldsymbol{\theta}}\{C_i^{-1}(N_i - C_i)\} = (2\nu)^{-1}\{2s + B_i + A_i(f_i - 0.5) - 0.5\sum_{j=1}^{k} f_j A_j\} + O(W^{3/2})$,

(ii) $E_{\boldsymbol{\theta}}\{C_i^{-2}(N_i - C_i)^2\} = (2\nu)^{-1}\{A_i(1 + 2f_i) + \sum_{j=1}^{k} f_j^2 A_j\} + O(W^{3/2})$,

where $A_i = \mathrm{tr}(\boldsymbol{\Sigma}_i^2)/(\mathrm{tr}(\boldsymbol{\Sigma}_i))^2,\ B_i = \nu/C_i,\ C_i$ is defined by (23), and s is a constant such that $E_{\boldsymbol{\theta}}(\hat{s}) = s + o(1)$.

Lemma 6 *For the two–stage procedure (6)–(7) with (10), one has as $d \to 0$ that $E_{\boldsymbol{\theta}}(\Re) = o(\nu^{-1})$ in (13).*

Proof In order to verify this lemma, we have to deal with the terms such as $E_{\boldsymbol{\theta}}(I_i),\ E_{\boldsymbol{\theta}}(I_{ij})$ and $E_{\boldsymbol{\theta}}(I_{ij\ell})$, where

$$I_i = \frac{\partial^3 g}{\partial u_i^3}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}} \left(\frac{N_i - C_i}{C_i}\right)^3, \quad I_{ij} = \frac{\partial^3 g}{\partial u_i^2 \partial u_j}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}} \left(\frac{N_i - C_i}{C_i}\right)^2 \left(\frac{N_j - C_j}{C_j}\right),$$
$$I_{ij\ell} = \frac{\partial^3 g}{\partial u_i \partial u_j \partial u_\ell}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}} \left(\frac{N_i - C_i}{C_i}\right)\left(\frac{N_j - C_j}{C_j}\right)\left(\frac{N_\ell - C_\ell}{C_\ell}\right)$$

for all $1 \le i < j < \ell \le k$. Note that each third–order partial derivative's magnitude can be bounded from above by a finite sum of terms of the type

$$A\xi_1^{-p_1}\xi_2^{-p_2}\cdots\xi_k^{-p_k} \tag{69}$$

with $A \ge 0,\ p_r \ge 0,\ r = 1, ..., k$, which are independent of d. Let A also denote a generic positive constant, independent of d. Let us write $N_i^\star = C_i^{-1}(N_i - C_i)$ for $i = 1, ..., k$. Then, we obtain that

$$|E_{\boldsymbol{\theta}}(I_i)| \le A E_{\boldsymbol{\theta}}(\xi_1^{-p_1}\xi_2^{-p_2}\cdots\xi_k^{-p_k}|N_i^\star|^3). \tag{70}$$

We observe that $\xi_i > m/C_i = \tau_\star/\tau_i$ w.p.1 for all $i = 1, ..., k$. Also, we observe that $E_{\boldsymbol{\theta}}(|N_i^\star|^3) = O(\nu^{-3/2})$ since $E_{\boldsymbol{\theta}}(|N_i^\star|^4) = O(\nu^{-2})$ from the facts that $E_{\boldsymbol{\theta}}\{(rT_i - 1)^3\} = O(\nu^{-2}),\ E_{\boldsymbol{\theta}}\{(rT_i - 1)^4\} = O(\nu^{-2})$ and so on together with (68). Hence, from (70), it follows that $|E_{\boldsymbol{\theta}}(I_i)| = O(\nu^{-3/2})$. Similarly, one may use the facts that $E_{\boldsymbol{\theta}}(|N_i^\star|^2|N_j^\star|) = O(\nu^{-3/2})$ and $E_{\boldsymbol{\theta}}(|N_i^\star||N_j^\star||N_\ell^\star|) = O(\nu^{-3/2})$ to show that $|E_{\boldsymbol{\theta}}(I_{ij})| = O(\nu^{-3/2})$ and $|E_{\boldsymbol{\theta}}(I_{ij\ell})| = O(\nu^{-3/2})$ for $1 \le i < j < \ell \le k$. Therefore, we conclude that $E_{\boldsymbol{\theta}}(\Re) = O(\nu^{-3/2}) = o(\nu^{-1})$. □

Remark 19 Second–order partial derivative's magnitude can be bounded from above by a finite sum of terms of the type similar to (69). We observe that $E_{\boldsymbol{\theta}}\{(\lambda - 1)N_i^\star\} < (E_{\boldsymbol{\theta}}(\nu^{-2}\hat{t}^2)E_{\boldsymbol{\theta}}(|N_i^\star|^2))^{1/2} = O(\nu^{-3/2}),\ E_{\boldsymbol{\theta}}\{(\lambda - 1)^2\} = O(\nu^{-2})$ and $\xi_\lambda > \min\{1, 1 + \nu^{-1}\hat{t}\}$ in (50)–(51). Note that, for nominal values of α and β, it holds that $\hat{t} > -1$ in (48). Hence, we have that

$$E_{\boldsymbol{\theta}}\left\{\frac{\partial^2 g_\alpha}{\partial\lambda\partial u_i}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}(\lambda - 1)\left(\frac{N_i - C_i}{C_i}\right)\right\} = o(\nu^{-1}),$$
$$E_{\boldsymbol{\theta}}\left\{\frac{\partial^2 g_\alpha}{\partial\lambda^2}\Big|_{\boldsymbol{u}=\boldsymbol{\xi}}(\lambda - 1)^2\right\} = o(\nu^{-1}).$$

Acknowledgements The authors would like to thank an anonymous referee and the Associate Editor for their valuable comments.

References

Aoshima, M. (2005). Statistical inference in two–stage sampling. *American Mathematical Society Translations Series 2, 215*, 125–145.

Aoshima, M. and Kushida, T. (2005). Asymptotic second–order efficiency for two–stage multiple comparisons with components of a linear function of mean vectors. In N. Balakrishnan, N. Kannan and H.N. Nagaraja (Ed.) *Advances in Ranking and Selection, Multiple Comparisons, and Reliability: Methodology and Applications* (pp. 191–211). Boston: Birkhäuser.

Aoshima, M. and Mukhopadhyay, N. (2002). Two–stage estimation of a linear function of normal means with second–order approximations. *Sequential Analysis, 21*, 109–144.

Aoshima, M. and Takada, Y. (2000). Second–order properties of a two–stage procedure for comparing several treatments with a control. *Journal of The Japan Statistical Society, 30*, 27–41.

Aoshima, M. and Takada, Y. (2002). Bounded risk point estimation of a linear function of k multinormal mean vectors when covariance matrices are unknown. In N. Balakrishnan (Ed.) *Advances on Theoretical and Methodological Aspects of Probability and Statistics* (pp. 279–287). New York: Taylor & Francis.

Banerjee, S. (1967). Confidence interval of preassigned length for the Behrens–Fisher problem. *Annals of Mathematical Statistics, 38*, 1175–1179.

Dantzig, G.B. (1940). On the non–existence of tests of "Student's" hypothesis having power functions independent of σ. *Annals of Mathematical Statistics, 11*, 186–192.

Ghosh, M., Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation.* New York: Wiley.

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Annals of Statistics, 9*, 1229–1238.

Liu, W. (2003). Some sequential equivalence tests that control both the size and power. *South African Statistical Journal, 37*, 105–125.

Liu, W. and Wang, N. (2007). A note on Hall's triple sampling procedure: A multiple sample second order sequential analogue of the Behrens-Fisher problem. *Journal of Statistical Planning and Inference, 137*, 331–343.

Mukhopadhyay, N. (2005). A new approach to determine the pilot sample size in two–stage sampling. *Communications in Statistics–Theory and Methods, 34*, 1275–1295.

Mukhopadhyay, N. and Duggan, W.T. (1997). Can a two–stage procedure enjoy second–order properties ?. *Sankhyā, Series A, 59*, 435–448.

Mukhopadhyay, N. and Duggan, W.T. (1999). On a two–stage procedure having second–order properties with applications. *Annals of the Institute of Statistical Mathematics, 51*, 621–636.

Schwabe, R. (1995). Some two–stage procedures for treating the Behrens–Fisher problem. In W.G. Müller, H.P. Wynn and A.A. Zhigljavsky (Ed.) *Model-Oriented Data Analysis* (pp. 81–89). Heidelberg: Physica.

Stein, C. (1945). A two–sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics, 16*, 243–258.

Takada, Y. (2004). Asymptotic second–order efficiency of a two–stage procedure for estimating a linear function of normal means. *Sequential Analysis, 23*, 103–120.

Takada, Y. and Aoshima, M. (1996). Two–stage procedures for the difference of two multinormal means with covariance matrices different only by unknown scalar multipliers. *Communications in Statistics–Theory and Methods, 25*, 2371–2379.

Takada, Y. and Aoshima, M. (1997). Two–stage procedures for estimating a linear function of multinormal mean vectors. *Sequential Analysis, 16*, 353–362.