XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 - September 4, 2015, Prague, Czech Republic

A SYNTHESIZED ESTIMATOR IN UNBALANCED NESTED DESIGNS WITH TWO STAGES

Seiichi Yasui[1] , *Yoshikazu Ojima*[2]

[1]Tokyo University of Science, Chiba, Japan, yasui@rs.tus.ac.jp.
[2]Tokyo University of Science, Chiba, Japan, ojima@rs.tus.ac.jp.

Abstract – Nested designs are one of experimental designs to evaluate and/or determine precision of measurement results, statistically. Unbalanced designs have interesting statistical problems, one of which is that an unbiased estimator of variance components is not unique. In our study, a new estimator of a variance component for the top stage of two-stages unbalanced nested designs is proposed. The performance of our estimator outperforms conventional estimators from the aspect of estimation accuracy.

Keywords: estimation of variance components, unbalanced nested designs, precision of measurement results

1. INTRODUCTION

Nested designs are one of experimental designs to evaluate and/or determine precision of measurement results, statistically. A nested design enables us to evaluate the several types of precision which is defined in ISO 5725-1 (1994).

For instance, if we obtain some measurement results from several laboratories, we can evaluate the two types of precision which are between-laboratories variation and within-laboratories variation. The statistical model to evaluate the precision is a random effect model with nested (hierarchical) structure corresponding to each stage of a nested design, where the stage is a laboratory or measurement. Thus, the above nested design has two stages. If we obtain some measurement results from each measurement technician and each laboratory, the nested deign has three stages. If the number of measurements for all stages is equal, the nested design is called balanced, otherwise, unbalanced. In this paper, we focus on unbalanced nested designs with two stages.

While two-stages nested design is the simplest one, the unbalanced design has interesting statistical matters that an unbiased estimator of variance components corresponding to between-laboratories variation is not unique. Yamazaki et al. (2008) proposed alternative unbiased estimator to an ANOVA type and described the differences between these estimators through numerical evaluation. The performances of these estimators are complementary to each other over the ratio of variance components.

In our study, we obtain more outperformed unbiased estimator than those proposed by Yamazaki et al. (2008) from the viewpoint of a variance of an estimator. Our estimator is the weighted average of two estimators that are an ANOVA type and an alternative estimator proposed by Yamazaki et al. (2008). Our estimator needs the optimal weight so that a variance of the estimator is minimized. The weight is a function of variance components. Thus, we cannot obtain the estimate in practice. We describe an iterative estimation, and evaluate it through Monte Carlo method.

Section 2 reviews two kinds of estimators of a variance component for a top stage of a two-stages unbalanced nested design. In Section 3, We propose weighted average of these estimators as a synthesized estimator and demonstrate that new estimator dominates existing estimators. In Section 4, an iterative procedure of the synthesized estimator is described. The estimation of the variance component is completed by this procedure. In Section 5, the proposal is evaluated through Monte Carlo method. Finally, we conclude in Section 6.

2. ESTIMATORS OF VARIANCE COMPONENTS

The statistical model of an unbalanced nested design with two stages is

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{1}$$
$$\alpha_i \sim i.i.d.\ N(0, \sigma_\alpha^2)\ ,\ \varepsilon_{ij} \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$$
$$i = 1, \cdots, a\ ,\ j = 1, \cdots n_i,$$

where μ is a general mean. σ_α^2 and σ_ε^2 are called variance components. In precise experiments of measurements, σ_α^2 is the variation between laboratories, and σ_ε^2 is the variation between the instantly repeated measurements. The purpose of our study is to estimate the variance components.

In general, the σ_α^2 is estimated by

$$\widehat{\sigma_\alpha^2} = \frac{n(a-1)}{(n^2 - m_2)^2}(MSA - MSE), \tag{2}$$

where MSA is a mean square error of the random effects α_i's, MSE is a mean square error of the random effects ε_{ij}'s, $n = \sum_{i=1}^{a} n_i$ and $m_2 = \sum_{i=1}^{a} n_i^2$, which are obtained through ANalysis Of Variance (ANOVA). The

MSA is

$$MSA = \frac{\sum_{i=1}^{a} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{a-1},$$

where

$$\bar{y}_{i\cdot} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}\ ,\ \bar{y}_{\cdot\cdot} = \frac{\sum_{i=1}^{n_a}\sum_{j=1}^{n_i} y_{ij}}{n}.$$

The MSA is an unbiased estimator of σ_α^2.

Yamazaki et al. (2008) introduced the other unbiased estimator of σ_α^2 ;

$$\widetilde{\sigma_\alpha^2} = MSA' - \left(\frac{m_{-1}}{a}\right)^2 MSE, \tag{3}$$

where $m_{-1} = \sum_{i=1}^{a} 1/n_i$ and

$$MSA' = \frac{\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{\bar{y}}_{\cdot\cdot})^2}{a-1}\ ,\ \bar{\bar{y}}_{\cdot\cdot} = \frac{\sum_{i=1}^{a}\bar{y}_{i\cdot}}{a}.$$

There is the difference between these estimators regarding to the estimation of the overall mean μ. In the MSA, the μ is estimated by an average of all the observations. On the other hand, the the estimator of the μ in the MSA' is an average of averages of observations in each laboratory.

3. WEIGHTED AVERAGE OF THE ESTIMATORS

The weighted average of (2) and (3) is $\lambda\widehat{\sigma_\alpha^2} + (1-\lambda)\widetilde{\sigma_\alpha^2}$, where the λ is a constant weight. This type of an estimator is unbiased as well. The variance of the weighted average is

$$C_1\lambda^2 - 2C_2\lambda + C_3 \tag{4}$$

$$C_1 = V(\widehat{\sigma_\alpha^2}) + V(\widetilde{\sigma_\alpha^2}) - 2Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})$$

$$C_2 = V(\widetilde{\sigma_\alpha^2}) - Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})\ ,\ C_3 = V(\widetilde{\sigma_\alpha^2}).$$

If $C_1 \neq 0$, then this variance is the minimum at $\lambda = C_2/C_1$, and the variance at $\lambda = C_2/C_1$ is $(C_1C_3 - C_2^2)/C_1$.

Due to $C_1 = V(\widehat{\sigma_\alpha^2} - \widetilde{\sigma_\alpha^2})$, $C_1 = 0$ is equivalent to $\widehat{\sigma_\alpha^2} = \widetilde{\sigma_\alpha^2}$. If a design is balanced and/or the number of levels is two ($a = 2$), $\widehat{\sigma_\alpha^2}$ is same as $\widetilde{\sigma_\alpha^2}$. It is trivial for a balanced design. In case of $a = 2$, due to $MSA = 2n_1n_2/(n_1+n_2)MSA'$, $\widehat{\sigma_\alpha^2} = (n_1+n_2)/(2n_1n_2)(MSA - MSE) = \widetilde{\sigma_\alpha^2}|_{a=2}$ holds.

Hence, for unbalanced designs with $a \geq 3$, the weighted average which is the estimator of σ_α^2

$$\overline{\sigma_\alpha^2} = \frac{C_2}{C_1}\widehat{\sigma_\alpha^2} + \left(1 - \frac{C_2}{C_1}\right)\widetilde{\sigma_\alpha^2}. \tag{5}$$

is unbiased, and it has the smaller or equal variance than the those of $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$, because the weighted average includes $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$. In fact, we can easily see it by calculating those variances. The numerator of the variance of this estimator (5) is

$$C_1C_3 - C_2^2 = V(\widehat{\sigma_\alpha^2})V(\widetilde{\sigma_\alpha^2}) - Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})^2.$$

Thus,

$$V(\overline{\sigma_\alpha^2}) - V(\widehat{\sigma_\alpha^2}) = -\frac{\left(V(\widehat{\sigma_\alpha^2}) - Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})\right)^2}{C_1} \leq 0$$

, due to $C_1 = V(\widehat{\sigma_\alpha^2} - \widehat{\sigma_\alpha^2}) \geq 0$. In the same manner,

$$V(\overline{\sigma_\alpha^2}) - V(\widetilde{\sigma_\alpha^2}) \leq 0$$

holds. Consequently, the new estimator $\overline{\sigma_\alpha^2}$ dominates $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$ in any design and over the space of variance components.

4. AN ITERATIVE METHOD FOR ESTIMATION

The estimator (5) is a function of variance components σ_α^2 and σ_ε^2. The σ_ε^2 is plugged in by MSE. We propose an iterative method to estimate σ_α^2 by the sythesized estimator (5). The iterative method consists of the four steps as follows :

Step 0	$\sigma_\alpha^2(0) \leftarrow$ an initial value
Step 1	σ_α^2 of $C_1, C_2 \leftarrow \sigma_\alpha^2(k)$
Step 2	$\sigma_\alpha^2(k+1) \leftarrow \overline{\sigma_\alpha^2}$
Step 3	repeat steps 1 and 2
	until $\lvert\sigma_\alpha^2(k+1) - \sigma_\alpha^2(k)\rvert < \delta$ is met
	, where δ is a small value such as 10^{-6}.

Yamasaki et al. (2008) derived $V(\widehat{\sigma_\alpha^2})$ and $V(\widetilde{\sigma_\alpha^2})$, analytically. We derive the covariance between $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$.

The covariance between $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$ is

$$Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2}) = \frac{n(a-1)}{n^2 - m_2}\left(Cov(MSA, MSA') + \frac{2m_{-1}}{a(n-a)}\sigma_\varepsilon^4\right). \tag{6}$$

The covariance between MSA and MSA' is

$$Cov(MSA, MSA') = \frac{1}{\phi_\alpha^2}\left(\sum_{i=1}^{a}\left(1 - \frac{n_i}{n}\right)n_ic_{ii} - \frac{1}{n}\sum_{i=1}^{a}\sum_{j\neq i}^{a} n_in_jc_{ij}\right) \tag{7}$$

, where $\phi_\alpha = a-1$ and

$$c_{ii} = \left(1-\frac{1}{a}\right)\left(E[\bar{y}_{i\cdot}^2]\sum_{k\neq i}^{a} E[\bar{y}_{k\cdot}^2] - (a-2)E[\bar{y}_{i\cdot}]^2 E[\bar{y}_{i\cdot}^2] - 2E[\bar{y}_{i\cdot}]E[\bar{y}_{i\cdot}^3] + E[\bar{y}_{i\cdot}^4]\right)$$

$$\begin{aligned} c_{ij} &= \left(1-\frac{1}{a}\right)E[\bar{y}_{i\cdot}]^2 \sum_{k\neq i,j}^{a} E[\bar{y}_{k\cdot}^2] \\ &\quad -\left(\frac{2(a-2)}{a}\right)E[\bar{y}_{i\cdot}]^2\left(E[\bar{y}_{i\cdot}^2]+E[\bar{y}_{j\cdot}^2]\right) \\ &\quad +\left(1-\frac{1}{a}\right)E[\bar{y}_{i\cdot}]\left(E[\bar{y}_{i\cdot}^3]+E[\bar{y}_{j\cdot}^3]\right) \\ &\quad -\frac{2}{a}E[\bar{y}_{i\cdot}^2]E[\bar{y}_{j\cdot}^2] - \frac{(a-2)(a-3)}{a}E[\bar{y}_{i\cdot}]^4 \end{aligned}$$

The random variates $\bar{y}_{i\cdot}, i=1,\cdots,a$ are identically and independently distributed to normal distribution $N(\mu, \sigma_\alpha^2 + \sigma_\varepsilon^2/n_i)$. About the derivation, see an Appendix.

We can obtain the simple representations of c_{ii} and c_{ij} due to substituting those expectations of the covariance (7) to the moments of a normal distribution, which are

$$\begin{aligned} c_{ii} &= \frac{\phi_\alpha}{a}\left(\mu^2\sigma_\cdot^2 + \sigma_\cdot^2\sigma_i^2 + 2\sigma_i^2\right) \\ c_{ij} &= \frac{\phi_\alpha}{a}\mu^2\sigma_\cdot^2 - \frac{2}{a}\sigma_i^2\sigma_j^2\ ,\ i\neq j. \end{aligned}$$

where $\sigma_i^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2/n_i$ and $\sigma_\cdot^2 = \sum_{i=1}^a \sigma_i^2$. By substituting the simplified c_{ii} and c_{ij} to the covariance (7), we can obtain

$$\begin{aligned} &Cov(MSA, MSA') \\ &\quad = \frac{1}{a\phi_a}\left[\sigma_\cdot^2\sum_{i=1}^{a}\left(1-\frac{n_i}{n}\right)n_i\sigma_i^2 + 2\sum_{i=1}^{a} n_i\sigma_i^4\right] \\ &\qquad - \frac{2}{n\phi_\alpha^2}\sum_{i=1}^{a}(n_i\sigma_i^2)^2 + \frac{2}{an\phi_\alpha^2}\left(\sum_{i=1}^{a} n_i\sigma_i^2\right)^2. \quad (8) \end{aligned}$$

Furthermore, by substituting σ_i^2 to the formula (8), the covariance between the mean squares is

$$\begin{aligned} Cov(MSA, MSA') &= \frac{(1+a)(n^2-m_2)}{n(a-1)^2}\sigma_\alpha^4 \\ &\quad + \left[a+3+\left(\frac{n^2-m_2}{an}\right)m_{-1}\right]\frac{\sigma_\alpha^2\sigma_\varepsilon^2}{a-1} \\ &\quad + \frac{(1+a)m_{-1}}{a(a-1)}\sigma_\varepsilon^2. \quad (9) \end{aligned}$$

Hence,

$$\begin{aligned} &Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2}) \\ &\quad = \frac{a+1}{a-1}\sigma_\alpha^4 + \frac{an(a+3)+(n^2-m_2)m_{-1}}{a(n^2-m_2)}\sigma_\alpha^2\sigma_\varepsilon^2 \\ &\qquad + \frac{(a-1)nm_{-1}}{a(n^2-m_2)}\left(\frac{a+1}{a-1}+\frac{2}{n-a}\right)\sigma_\varepsilon^4, \quad (10) \end{aligned}$$

can be obtained due to combining the equations (6) and (9).

The synthesized estimator is,

$$\overline{\sigma_\alpha^2} = \frac{C_2}{C_1}\left(\widehat{\sigma_\alpha^2} - \widetilde{\sigma_\alpha^2}\right) + \widetilde{\sigma_\alpha^2},$$

the function of σ_α^2 because the $\lambda = C_2/C_1$ is the function of σ_α^2 and the form is

$$\lambda(\sigma_\alpha^2) \equiv \lambda = \frac{V(\widehat{\sigma_\alpha^2}) - Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})}{V(\widetilde{\sigma_\alpha^2}) - Cov(\widehat{\sigma_\alpha^2}, \widetilde{\sigma_\alpha^2})} + 1,$$

which is a rational function of σ_α^2 whose numerator and denominator are quadratic functions because both variances are

$$\begin{aligned} V(\widehat{\sigma_\alpha^2}) &= 2n^2\left(m_2 + m_2^2/n^2 - 2m_3/n\right)/(n^2-m_2)^2\sigma_\alpha^4 \\ &\quad + 4n^2\left(n - m_2/n\right)/(n^2-m_2)^2\sigma_\alpha^2\sigma_\varepsilon^2 \\ &\quad + 2(a-1)^2n^2\left(\frac{1}{a-1}+\frac{1}{n-a}\right)/(n^2-m_2)^2\sigma_\varepsilon^4 \quad (11) \end{aligned}$$

and

$$\begin{aligned} V(\widetilde{\sigma_\alpha^2}) &= \frac{2}{a-1}\sigma_\alpha^4 + \frac{2(2am_{-1}+2(a-2)am_{-1})}{(a-1)^2a^2}\sigma_\alpha^2\sigma_\varepsilon^2 \\ &\quad + \left(\frac{2m_{-1}^2}{a^2(n-a)} + \frac{2(a-2)am_{-2}+m_{-1}^2}{(a-1)^2a^2}\right)\sigma_\varepsilon^4, \quad (12) \end{aligned}$$

respectively, which can be obtained from Yamazaki et al. (2008) by a few calculations. Let the synthesized estimator be $\overline{\sigma_\alpha^2}(\sigma_\alpha^2)$. If the $\overline{\sigma_\alpha^2}(\sigma_\alpha^2)$ is convergence, the limit value is the solution of the equation $\overline{\sigma_\alpha^2}(\sigma_\alpha^2) = \sigma_\alpha^2$.

5. PREFERMANCE STUDY

We calculate variances of estimators and compare performances of three estimators $\widehat{\sigma_\alpha^2}$, $\widetilde{\sigma_\alpha^2}$ and $\overline{\sigma_\alpha^2}$. It is difficult to evaluate the variance of $\overline{\sigma_\alpha^2}$ analytically because the $\overline{\sigma_\alpha^2}$ is calculated by iteration. Hence, we evaluate the variance of the $\overline{\sigma_\alpha^2}$ through Monte Carlo method.

We assume six cases of two-stages unbalanced nested designs which are denoted in Table 1. The first three cases (from cases 1 to 3) have three levels, and the remains have ten levels. In an experiment to evaluate the precision of the measurement, the number of laboratories is the number of levels. For instance, Case 1 shows that we obtained nine measurement results from the first two laboratories and twelve ones from the last laboratory. Any case has thirty observations. The contribution of the total number of measurement results to precision of estimation is same in any case.

Without loss of generality, we can assume $\sigma_\varepsilon^2 = 1$, because the covariance between $\widehat{\sigma_\alpha^2}$ and $\widetilde{\sigma_\alpha^2}$ depends on the moments of $\bar{y}_{i\cdot}$ which is a random variable according to a normal distribution. We consider three conditions $\gamma^2 = \sigma_\alpha^2/\sigma_\varepsilon^2 = 0.5^2, 1.0^2, 2.0^2$. Variances and expectations of the proposed estimator are calculated based on empirical distributions generated by Monte Carlo method with $10,000$ replications that is implemented by the software R. In this performance study, we set the convergence criterion δ on 10^{-5}.

Table 1. The number of observations in each level for each case

Case 1	(9 , 9 , 12)
Case 2	(8, 10, 12)
Case 3	(5, 5, 20)
Case 4	(2, 3, 3, 3, 3, 3, 3, 3, 3, 4)
Case 5	(2, 2, 2, 2, 2, 2, 2, 2, 7, 7)
Case 6	(2, 2, 2, 2, 3, 3, 4, 4, 4, 4)

Tables 2, 3 and 4 show variances of the proposed and two conventional estimators for each case. Variances of the estimators $\widehat{\sigma_\alpha^2}$ (ANOVA type) and $\widetilde{\sigma_\alpha^2}$ (introduced by Yamasaki et al. (2008)) are exact values. Values in parentheses denote estimated Monte Carlo error which is two times standard deviation calculated from 5,000 bootstrap samples.

Yamasaki et al. (2008) showed that the ANOVA type estimator $\widehat{\sigma_\alpha^2}$ has the better performance than the estimator $\widetilde{\sigma_\alpha^2}$. In this study, the difference between them is not too large in any of three levels case and the case 4 which has the smallest degree of imbalance among ten levels cases. In cases 5 and 6, for $\gamma^2 = 1.0^2$ and 2.0^2, the proposed estimator has the best performance if comparing them without Monte Carlo error. Even if considering Monte Carlo error, the proposed estimator has approximately same or better performance than the other estimators. Our estimator is useful if the degree of imbalance is large and the ratio γ^2 is lager than or equal to one. In most of actual experiments, between-laboratories variation is larger than within-laboratories variation, i.e. $\gamma^2 > 1$. The proposed estimator is reasonable in practice.

Tables 5 and 6 show biases of the proposed estimator. If considering Monte Carlo error, there is no bias in all the situations except for $\gamma^2 = 1$. In cases 2, 3, 5, and 6, there are biases beyond Monte Carlo error.

Table 2. Variances of the proposed and existing estimators

	Case 1			Case 2		
γ^2	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$
0.5^2	0.124	0.125	0.124 (0.007)	0.125	0.125	0.125 (0.008)
1.0^2	1.22	1.21	1.19 (0.06)	1.23	1.22	1.24 (0.06)
2.0^2	17.0	16.8	17.2 (0.47)	17.0	16.8	17.4 (0.98)

Table 3. Variances of the proposed and existing estimators (cont'd.)

	Case 3			Case 4		
γ^2	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$
0.5^2	0.155	0.164	0.168 (0.009)	0.087	0.090	0.106 (0.004)
1.0^2	1.40	1.33	1.37 (0.07)	0.411	0.412	0.431 (0.016)
2.0^2	18.9	17.2	16.5 (0.88)	4.26	4.20	4.27 (0.15)

6. CONCLUSION

In unbalanced nested designs, the estimators of variance components for upper stages are not unique. We focused on two-stages unbalanced nested designs. We proposed the synthesized estimator by weighted average of two kinds of conventional estimators and its iterative estimation. We demonstrate that the estimator is relatively more effective than existing estimators if the ratio of variance components is equal to or lager than one. The estimator obtained from the iterative procedure preserves unbiasedness. The evaluation is carried out through Monte Carlo method, thus, we estimate Monte Carlo errors by bootstrap method. Hence, our estimator is useful in many of practical situations to evaluate and determine precision of measurements.

REFERENCES

[1] ISO 5725-1, "Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions", *International Organization for Standardization*, Geneva, Switzerland.

[2] M. Yamasaki, Y. Ojima, and S. Yasui, Estimation of Variance Components In Unbalanced Nested Designs", *Proceedings of Asian Network for Quality Congress 2008* , 10 pages in CD-ROM, 2008.

Table 4. Variances of the proposed and existing estimators (cont'd.)

	Case 5			Case 6		
γ^2	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$	$\widehat{\sigma_\alpha^2}$	$\widetilde{\sigma_\alpha^2}$	$\overline{\sigma_\alpha^2}$
0.5^2	0.097	0.125	0.099 (0.004)	0.089	0.100	0.107 (0.004)
1.0^2	0.495	0.476	0.444 (0.016)	0.426	0.431	0.415 (0.014)
2.0^2	5.46	4.38	4.33 (0.16)	4.48	4.25	4.16 (0.16)

Table 5. Biases of the proposed estimator

γ^2	C1	C2	C3
0.5^2	0.002 (0.008)	-0.004 (0.008)	0.018 (0.008)
1.0^2	0.022 (0.022)	0.030 (0.024)	0.048 (0.024)
2.0^2	0.056 (0.084)	0.038 (0.084)	-0.004 (0.082)

APPENDIX

Let a vector $\boldsymbol{y}$ be $(\bar{y}_{1\cdot}, \cdots, \bar{y}_{a\cdot})^T$. Then, MSA and MSA' are expressed as

$$\begin{aligned} MSA &= \boldsymbol{y}^T \mathrm{diag}\{n_i\} \left(I - J\mathrm{diag}\{n_i/n\}\right) \boldsymbol{y}, \\ MSA' &= \boldsymbol{y}^T \left(I - J/a\right) \boldsymbol{y} \end{aligned}$$

, respectively, where I is an identity matrix and J is a square matrix in which all the elements are one.

$E[MSA'MSA]$ is

$$\begin{aligned} & E[\boldsymbol{y}^T \left(I - J/a\right) \boldsymbol{y}\boldsymbol{y}^T \mathrm{diag}\{n_i\} \left(I - J\mathrm{diag}\{n_i/n\}\right) \boldsymbol{y}] \\ & = \mathrm{tr} \left\{ \mathrm{diag}\{n_i\} \left(I - J\mathrm{diag}\{n_i/n\}\right) E[\boldsymbol{y}\boldsymbol{y}^T \left(I - J/a\right) \boldsymbol{y}\boldsymbol{y}^T] \right\}, \end{aligned}$$

where $\mathrm{diag}\{n_i\}$ is a diagonal matrix with element n_i's. Let a vector $\boldsymbol{z}_i$ be the i-th column vector of the matrix $\boldsymbol{y}\boldsymbol{y}^T$. Then, $\boldsymbol{z}_i = (\bar{y}_{1\cdot}\bar{y}_{i\cdot}, \cdots, \bar{y}_{a\cdot}\bar{y}_{i\cdot})^T$ and

$$\boldsymbol{y}\boldsymbol{y}^T \left(I - J/a\right) \boldsymbol{y}\boldsymbol{y}^T = \begin{pmatrix} \boldsymbol{z}_1^T A \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_1^T A \boldsymbol{z}_a \\ \vdots & \ddots & \vdots \\ \boldsymbol{z}_a^T A \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_a^T A \boldsymbol{z}_a \end{pmatrix}$$

, where $A = (I - J/a)$. Hence, it holds that

$$\begin{aligned} & E[MSA'MSA] \\ & \quad = \mathrm{tr}B \begin{pmatrix} \mathrm{tr}AE[\boldsymbol{z}_1\boldsymbol{z}_1^T] & \cdots & \mathrm{tr}AE[\boldsymbol{z}_a\boldsymbol{z}_1^T] \\ \vdots & \ddots & \vdots \\ \mathrm{tr}AE[\boldsymbol{z}_a\boldsymbol{z}_1^T] & \cdots & \mathrm{tr}AE[\boldsymbol{z}_a\boldsymbol{z}_a^T] \end{pmatrix} \end{aligned}$$

Table 6. Biases of the proposed estimator (cont'd)

γ^2	C4	C5	C6
0.5^2	0.005 (0.006)	0.032 (0.006)	0.018 (0.006)
1.0^2	0.005 (0.014)	0.020 (0.014)	0.027 (0.014)
2.0^2	-0.007 (0.042)	0.002 (0.042)	-0.020 (0.040)

, where $B = \mathrm{diag}\{n_i\} \left(I - J\mathrm{diag}\{n_i/n\}\right)$. Consequently, c_{ii} is $\mathrm{tr}AE[\boldsymbol{z}_i\boldsymbol{z}_i^T]$ and $c_{ij}, i \neq j$ is $\mathrm{tr}AE[\boldsymbol{z}_i\boldsymbol{z}_j^T], i \neq j$.