Metrika
https://doi.org/10.1007/s00184-020-00796-w

Bayesian multi-way balanced nested MANOVA models with random effects and a large number of the main factor levels

Chun-Lung Su[1]

Received: 7 October 2019 / Accepted: 15 September 2020

Abstract
This article considers the balanced nested multi-way multivariate analysis of variance (MANOVA) models with random effects and a large number of main factor levels under certain prior assumptions. Two different parametrizations for the MANOVA models with random effects and the corresponding explicit asymptotics are established. The asymptotic approximations are then compared with those obtained from the classical large-sample approximation and Markov chain Monte Carlo method via a balanced nested two-way MANOVA model with random effects. Simulation results demonstrate that our approach is superior to the classical approximation method on estimating the posterior standard deviations of variance component parameters.

Keywords Asymptotic posterior · Balanced nested model · MANOVA · Sufficient reduction statistics

1 Introduction

In certain experimental designs, nested models are commonly used when constraints prevent us from crossing every level of one factor with every level of the other factor. In particular, the balanced nested case has the advantage of minimizing the sensitivity of statistical assumptions and maximizing the power of the test (Montgomery 2012). Classical large-sample inferences on testing treatment or sub-treatment effects are usually made under the assumption that the number of levels of main and nested factors is small and replication is large at each level of the nested factor (Arnold 1980). For instance, in the two-way balanced nested ANOVA or MANOVA models comparisons are made between the products from two suppliers with three shifts nested within each supplier and large replications made in each shift; in mammography or colonoscopy examination, the number of levels of clinics and laboratory scientists

✉ Chun-Lung Su
clsu@thu.edu.tw

1 Department of Statistics, Tunghai University, Taichung, Taiwan

Published online: 11 October 2020

nested within clinics is small while the number of physical examination for human subjects is large. However, in some experimental design, data may consist of a larger number of levels of the main factor and a small number of levels of nested factor and replications. For example, in some 24-h automatic factories, data consist of hundreds of machines (main factor) and a small number of shift operators (nested factor) and in-process destructive tests (replications). For this type of data, some authors, from the non-Bayesian perspective, have introduced asymptotic results for test statistics under univariate ANOVA models without normality assumption (see Boos and Brownie 1995; Akritas and Arnold 2000; Bathke 2002; Wang and Akritas 2006); similar asymptotics of three common test statistics for MANOVA models under non-normality have been shown in Gupta et al. (2006). For a Bayesian counterpart, Su (2017) has developed large sample results for balanced nested multi-way ANOVA models with random effects. This article aims to establish the explicit estimates of parameters and some functions of random effects, called the sufficient reduction statistics, simultaneously for the finite-sample case, and give the explicit asymptotic posterior results under the balanced nested multi-way MANOVA models with random effects and certain prior assumptions. For the finite-sample case, to the best of our knowledge, there is no analytical approach that can derive the posterior estimates for these multivariate models, though the approximate posterior mean estimates of parameters for the multivariate mixed models can be seen in Jelenkowska and Press (1997). For the univariate case, explicit estimates of parameters with Bayesian and non-Bayesian approaches can be found in Su (2017) for balanced nested ANOVA models with random effects and Szatrowski and Miller (1980) for more general ANOVA models, respectively (see also Sahai and Ojeda (2004)). A practical application of the sufficient reduction statistics can be seen in Ansia et al. (2014).

As to the large-sample results, unlike the maximum a posterior-based (MAP-based) approaches that require fully specifying the joint posterior (see Johnson 1967; Chen 1985; Press 2002 and the references therein), the technique used to derive asymptotic results is mainly based on the proposition in Su and Johnson (2006), where the joint posterior asymptotics are derived through a system of conditional moment functions and under the assumption that each conditional distribution is asymptotically normally distributed. For completeness, we will state this proposition in the following section. The benefit of using posterior mean and standard deviation instead of the MAP estimates and the corresponding estimated standard errors in Edgeworth approximations for the univariate posterior density can be seen in Kolassa and Kuffner (2020). Moreover, we propose two parametrizations and compare their locally linear convergence rates of iterative algorithms when there exists no analytical solution for the posterior estimates.

We organize the paper as follows. In Sect. 2, we introduce some notations and Su and Johnson's proposition used in the subsequent sections and appendix. In Sect. 3, we propose two different parametrizations for balanced nested two-way MANOVA models with random effects, and derive the corresponding asymptotics. We also make comparisons between these two parametrizations based on the asymptotic convergence rates. In Sect. 4, we extend the asymptotics from the two-way case to the multi-way case. In Sect. 5, we compare our approach with those from the Markov chain Monte

Carlo (MCMC) and the classical large-sample approach via a simulation study. Some comments are given in Sect. 6. The proofs are established in the appendix.

2 Notation and proposition

Notation We introduce some notations that can be found mainly in Magnus and Neudecker (1999) and Seber (2007). Other notations, such as matrix differential operator and duplication matrix, that have been used only for the proofs are presented in the appendix.

The Kronecker product of an $m \times n$ matrix A and an $p \times q$ matrix B, $A \otimes B$, is an $m \times n$ block-matrix whose (i, j)th block is $a_{ij}B$. Denote $A^{\otimes 2} = A \otimes A$. The abbreviation blkdiag$\{A_1, \ldots, A_k\}$ denotes a block diagonal matrix with block matrices $A_1, \ldots, A_k$. $A > 0$ ($A \geq 0$) represents A is symmetric positive definite (positive semi-definite). The vectorization of an $m \times n$ matrix A, vec(A), is an mn column vector and vec$(A) = (a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn})'$. The half-vectorization of an $n \times n$ matrix A, vech(A), is an $\frac{n(n+1)}{2}$ column vector and vech$(A) = (a_{11}, \ldots, a_{n1}, a_{22}, \ldots, a_{n2}, \ldots, a_{n-1,n-1}, a_{n,n-1}, a_{nn})'$. For brevity, we add an additional subscript v to represent the vectorization of a matrix, or its half-vectorization operator when the matrix is symmetric, e.g. $A_v =$ vec(A); $A_{1v} =$ vech(A_1) when A_1 is symmetric. We omit the subscript v to stand for the inverse of the (half-)vectorization. For any square matrix A, the spectral radius $\rho(A)$ denotes the largest absolute eigenvalue of A and $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A. I_n is an $n \times n$ identity matrix. $\mathbf{0}_n$ is an $n \times 1$ column vector of all zeros. $\mathbf{O}_{mn}$ is an $m \times n$ zero matrix. Denote $\mathbf{O}_{mn} = \mathbf{O}_n$ if $m = n$. For any $n \times n$ matrix A, define $A^{\otimes 2h} \equiv L_n N_n A^{\otimes 2} N_n L_n'$, where the elimination matrix L_n is the $\frac{n(n+1)}{2} \times n^2$ matrix such that $L_n \text{vec}(A) = \text{vech}(A)$, the $n^2 \times n^2$ matrix $N_n \equiv \frac{1}{2}(I_{n^2} + K_n)$, and the commutation matrix K_n is the $n^2 \times n^2$ matrix such that $K_n \text{vec}(A) = \text{vec}(A')$.

The abbreviation iid stands for "independently and identically distributed," and the abbreviation ind stands for "independently distributed." The statistical distributions $\mathrm{W}_p(\cdot,\cdot)$, $\mathrm{IW}_p(\cdot,\cdot)$, $\mathrm{t}_p(\cdot,\cdot,\cdot)$, and $N_p(\cdot,\cdot)$ denote the p-variate Wishart, inverse Wishart, multivariate t, and multivariate normal distributions, respectively. We suppress the subscript when there is no ambiguity.

Also, in the context, we use small letters without bold font to represent some scalars, and either small letters with bold font or capital letters without bold font to represent some vectors or matrices.

Su and Johnson's proposition Suppose there are m-block conditional distributions which may or may not be derived from the joint posterior of $\Theta_n = (\Theta_{1n}', \ldots, \Theta_{mn}')'$ given the observed data $\boldsymbol{y}_n$, and Θ_{in} is a q_i-dimensional random vector, $i = 1, \ldots, m$. Set $q = \sum_{i=1}^m q_i$. Define the conditional mean functions and their derivatives as follows.

$$\boldsymbol{g}_{in}(\boldsymbol{t}_{(-i)n}) \equiv E(\Theta_{in} | \Theta_{(-i)n} = \boldsymbol{t}_{(-i)n}, \boldsymbol{y}_n),\ i = 1, \ldots, m,$$
$$\boldsymbol{g}_n(\boldsymbol{t}_n) = (\boldsymbol{g}_{1n}'(\boldsymbol{t}_{(-1)n}), \ldots, \boldsymbol{g}_{mn}'(\boldsymbol{t}_{(-m)n}))',$$

where $\boldsymbol{t}_n = (\boldsymbol{t}'_{1n}, \ldots, \boldsymbol{t}'_{mn})'$ is any arbitrary fixed vector in the parameter space $\Omega = \Omega_1 \times \cdots \times \Omega_m \subseteq \mathbb{R}^q$, $\boldsymbol{t}_{(-i)n} = (\boldsymbol{t}'_{1n}, \ldots, \boldsymbol{t}'_{i-1,n}, \boldsymbol{t}'_{i+1,n}, \ldots, \boldsymbol{t}'_{mn})'$, $\boldsymbol{t}_{in} \in \Omega_i \subseteq \mathbb{R}^{q_i}$, $\Theta_{(-i)n} = (\Theta'_{1n}, \ldots, \Theta'_{i-1,n}, \Theta'_{i+1,n}, \ldots, \Theta'_{mn})'$, $\boldsymbol{g}_{in}$ is a q_i-dimensional vector-valued function from $\Omega_{(-i)} \subseteq \mathbb{R}^{q-q_i}$ into $\Omega_i \subseteq \mathbb{R}^{q_i}$, $\Omega_{(-i)} = \Omega_1 \times \cdots \times \Omega_{i-1} \times \Omega_{i+1} \times \cdots \times \Omega_m$, $i = 1, \cdots, m$, and $\boldsymbol{g}_n$ is a function from Ω into Ω. The derivative of $\boldsymbol{g}_n(\boldsymbol{t}_n)$, $\dot{\boldsymbol{g}}_n(\boldsymbol{t}_n)$, is an $m \times m$ block matrix whose (i, j)th block is a $q_i \times q_j$ matrix, and

$$\dot{\boldsymbol{g}}_{in}^{(j)}(\boldsymbol{t}_{(-i)n}) = \frac{\partial}{\partial \boldsymbol{t}_{jn}} \boldsymbol{g}_{in}(\boldsymbol{t}_{(-i)n}),\ 1 \le i, j \le m.$$

Also, define the conditional covariance function, Σ_{Dn}, as follows.

$$\Sigma_{Dn}(\boldsymbol{t}_n) = \text{blkdiag}\{\Sigma_{1n}(\boldsymbol{t}_{(-1)n}), \ldots, \Sigma_{mn}(\boldsymbol{t}_{(-m)n})\},$$
$$\Sigma_{in}(\boldsymbol{t}_{(-i)n}) \equiv n \cdot \text{Cov}(\Theta_{in} | \Theta_{(-i)n} = \boldsymbol{t}_{(-i)n}, \boldsymbol{y}_n),\ i = 1, \ldots, m,$$

where Σ_{in} is a square matrix-valued function of size q_i, $i = 1, \cdots, m$. Assume the observed data sequence $\boldsymbol{y}_n \overset{n\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\boldsymbol{y}}$, and, for any $\boldsymbol{t} \in \mathbb{R}^p$, $\boldsymbol{g}_n(\boldsymbol{t}) \overset{n\to\infty}{\longrightarrow} \boldsymbol{g}(\boldsymbol{t})$, and $\Sigma_{Dn}(\boldsymbol{t}) \overset{n\to\infty}{\longrightarrow} \Sigma_D(\boldsymbol{t})$.

Proposition 1 *(Su and Johnson 2006) Suppose the assumptions* ***(A1)*** *and* ***(A2)*** *hold, namely*
(A1) *$\boldsymbol{g}_n$ and $\boldsymbol{g}$ have fixed points, $\boldsymbol{\mu}_{\Theta_n}$ and $\boldsymbol{\mu}_\Theta$, respectively, $\boldsymbol{g}_n(\boldsymbol{\mu}_{\Theta_n}) = \boldsymbol{\mu}_{\Theta_n} \overset{n\to\infty}{\longrightarrow} \boldsymbol{\mu}_\Theta = \boldsymbol{g}(\boldsymbol{\mu}_\Theta)$, and $\Sigma_D(\boldsymbol{\mu}_\Theta) \equiv \Sigma_D = blkdiag\{\Sigma_1, \ldots, \Sigma_m\} > 0$.*
(A2) *$\Sigma_{in}^{-1/2}(\boldsymbol{t}_{(-i)n})\sqrt{n}\{\Theta_{in} - \boldsymbol{g}_{in}(\boldsymbol{t}_{(-i)n})\}|(\Theta_{(-i)n} = \boldsymbol{t}_{(-i)n}, \boldsymbol{y}_n) \overset{\mathcal{L}}{\longrightarrow} N_{q_i}(\mathbf{0}_{q_i}, I_{q_i})$, $i = 1, \ldots, m$.*
Then, under some regularity conditions (see the proof of Theorem 1 for further details), we have

$$\sqrt{n}(\Theta_n - \boldsymbol{\mu}_{\Theta_n}) \overset{\mathcal{L}}{\longrightarrow} N_p(\mathbf{0}_p, \Sigma) \Longleftrightarrow \Sigma > 0,$$

where $\Sigma = V^{-1}\Sigma_D$, and $V = I_q - \dot{\boldsymbol{g}}(\boldsymbol{\mu}_\Theta)$ is an $m \times m$ block matrix whose (i, j)th block $V_{ij} = I_{q_i}$, if $i = j$, $V_{ij} = -\dot{\boldsymbol{g}}_i^{(j)}(\boldsymbol{\mu}_{\Theta_{(-i)}})$, if $i \ne j$. For finite sample, the posterior distribution of Θ_n is approximated by $N_q(\boldsymbol{\mu}_{\Theta_n}, \frac{1}{n}(I_q - \dot{\boldsymbol{g}}_n(\boldsymbol{\mu}_{\Theta_n}))^{-1}\Sigma_{Dn}(\boldsymbol{\mu}_{\Theta_n}))$.

3 Two-way balanced nested MANOVA model with random effects

In Sect. 3.1, we consider two different parametrizations, called the standard and the centering of the balanced nested two-way MANOVA models with random effects. In Sects. 3.2 and 3.3, we derive the corresponding point estimates and asymptotics for parameters and the sufficient reduction statistics explicitly.

3.1 Two different parametrizations: standard and centering

Consider the following standard balanced nested two-way MANOVA model with random effects

$$\begin{aligned}\boldsymbol{y}_{ijk} =&\boldsymbol{\beta} + \boldsymbol{u}_i + \boldsymbol{u}_{ij} + \boldsymbol{\epsilon}_{ijk},\ \boldsymbol{\epsilon}_{ijk} \overset{\text{iid}}{\sim} N_p(\boldsymbol{0}, R_0),\ \boldsymbol{u}_{ij}|R_1 \overset{\text{iid}}{\sim} N_p(\boldsymbol{0}, R_1),\\ &\boldsymbol{u}_i|R_2 \overset{\text{iid}}{\sim} N_p(\boldsymbol{0}, R_2),\ i = 1, \ldots, a,\ j = 1, \ldots, b,\ k = 1, \ldots, c,\end{aligned} \tag{1}$$

where $\boldsymbol{y}_{ijk}$ denotes the p-dimensional random vector, and the error terms $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_{111}, \ldots, \boldsymbol{\epsilon}_{abc}\}$ and the random effects $\mathcal{U} = \{\boldsymbol{u}_{11}, \ldots, \boldsymbol{u}_{ab}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_a\}$ are assumed to be independent. The standard parametrization means the fixed effect $\boldsymbol{\beta}$ and the random effects $\mathcal{U}$ are treated as the first stage parameters and latent variables, respectively, whereas the variance component parameters $\{R_0, R_1, R_2\}$ to be the second stage ones. Owing to the symmetry of R_i's, set $R = (R'_{0v}, R'_{1v}, R'_{2v})'$ and $\boldsymbol{\phi} = (R', \boldsymbol{\beta}')'$. Let $p_1 = \frac{p(p+1)}{2}$ and $p_2 = 3p_1 + p$. Here R, R_{kv}, $k = 0, 1, 2$, and $\boldsymbol{\phi}$ are $3p_1$, p_1, and p_2 dimensional vectors, respectively. Although the standard model is commonly used in some statistical textbooks, its Gibbs sampling convergence rate is usually slower than that from the (hierarchical) centering case (Gelfand et al. 1995; Roberts and Sahu 1997). By setting $\boldsymbol{v}_{ij} = \boldsymbol{\beta} + \boldsymbol{u}_i + \boldsymbol{u}_{ij}$ and $\boldsymbol{v}_i = \boldsymbol{\beta} + \boldsymbol{u}_i$, the centering model is given as

$$\begin{aligned}\boldsymbol{y}_{ijk} &= \boldsymbol{v}_{ij} + \boldsymbol{\epsilon}_{ijk},\ \boldsymbol{\epsilon}_{ijk} \overset{\text{iid}}{\sim} N_p(\boldsymbol{0}, R_0), \boldsymbol{v}_{ij}|\boldsymbol{v}_i, R_1 \overset{\text{ind}}{\sim} N_p(\boldsymbol{v}_i, R_1),\\ &\boldsymbol{v}_i|\boldsymbol{\beta}, R_1 \overset{\text{iid}}{\sim} N_p(\boldsymbol{\beta}, R_2),\ i = 1, \ldots, a,\ j = 1, \ldots, b,\ i = 1, \ldots, c,\end{aligned} \tag{2}$$

where the error terms $\boldsymbol{\epsilon}$ and the random effects $\mathcal{V} = \{\boldsymbol{v}_{11}, \ldots, \boldsymbol{v}_{ab}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_a\}$ are assumed to be independent. In addition to the advantage of MCMC convergence, other benefits of the centering model are discussed in Remark 1 and Corollary 1. When we are only interested in the parameter $\boldsymbol{\phi}$, the posteriors of $\boldsymbol{\phi}$ are the same under both models with the same priors used after integrating out the random effects.

Assume the joint prior for R_0, R_1, R_2, and $\boldsymbol{\beta}$, $f_{pr}(R_0, R_1, R_2, \boldsymbol{\beta})$, is proportional to

$$\prod_{k=0}^{2} |R_k|^{-\frac{m_k+p+1}{2}} e^{-\frac{1}{2}\text{tr}(R_k^{-1}\Psi_k)} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)' Q_{\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}. \tag{3}$$

When the hyper-parameters $m_k > p - 1$, and Ψ_k, $Q_{\boldsymbol{\beta}} > 0$, the priors can be rewritten as $\prod_{k=0}^{2} \text{IW}(m_k, \Psi_k) N(\boldsymbol{\beta}_0, Q_{\boldsymbol{\beta}}^{-1})$ since the prior setting in (3) is the product of the three kernels of the inverse Wishart distributions (for R_k's) and the kernel of the normal distribution (for $\boldsymbol{\beta}$). When Ψ_k and $Q_{\boldsymbol{\beta}}$ are set to be equal to zero matrices, an improper prior is assumed for each R_k and a flat prior for $\boldsymbol{\beta}$. The reason why we use the inverse Wishart distribution, the conjugate prior for R_k's, is to derive the explicit conditional mean function and ensure the existence of the unique fixed point of $\boldsymbol{g}_a$ for any combination of (a, b, c) under certain hyper-parameter assumptions (see Lemma 1).

3.2 The centering case

From the Eq. (2) and the prior setting (3), the joint density of R_0, R_1, R_2, $\boldsymbol{\beta}$, $\{\boldsymbol{v}_{ij}\text{'s}\}$, and $\{\boldsymbol{v}_i\text{'s}\}$ given $\boldsymbol{y}$ is obtained as

$$\begin{aligned}
&f(R_0, R_1, R_2, \boldsymbol{\beta}, \{\boldsymbol{v}_{ij}\text{'s}\}, \{\boldsymbol{v}_i\text{'s}\}|\boldsymbol{y})\\
&\quad\propto |R_0|^{-\frac{abc}{2}} \exp\{-\frac{1}{2}\sum_{i,j,k}(\boldsymbol{y}_{ijk} - \boldsymbol{v}_{ij})' R_0^{-1}(\boldsymbol{y}_{ijk} - \boldsymbol{v}_{ij})\}\\
&\quad\times |R_1|^{-\frac{ab}{2}} \exp\{-\frac{1}{2}\sum_{i,j}(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)' R_1^{-1}(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)\}\\
&\quad\times |R_2|^{-\frac{a}{2}} \exp\{-\frac{1}{2}\sum_{i}(\boldsymbol{v}_i - \boldsymbol{\beta})' R_2^{-1}(\boldsymbol{v}_i - \boldsymbol{\beta})\} \times f_{pr}(R_0, R_1, R_2, \boldsymbol{\beta}),
\end{aligned} \tag{4}$$

where $\sum_{i,j,k}$, $\sum_{i,j}$, and $\sum_i$ represent $\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c$, $\sum_{i=1}^a \sum_{j=1}^b$, and $\sum_{i=1}^a$, respectively. Hence the conditional posteriors for R_0, R_1, R_2, and $\boldsymbol{\beta}$ are

$$\begin{aligned}
R_0|T, \boldsymbol{y} &\sim \mathrm{IW}_p\big(abc + m_0, abcT_0 + \Psi_0\big),\\
R_1|T, \boldsymbol{y} &\sim \mathrm{IW}_p\big(ab + m_1, abT_1 + \Psi_1\big),\\
R_2|T, \boldsymbol{\beta}, \boldsymbol{y} &\sim \mathrm{IW}_p\big(a + m_2, a[T_2 + (T_3 - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}} - \boldsymbol{\beta})' + (\bar{\boldsymbol{y}} - \boldsymbol{\beta})(T_2 - \bar{\boldsymbol{y}})'\\
&\qquad + (\bar{\boldsymbol{y}} - \boldsymbol{\beta})(\bar{\boldsymbol{y}} - \boldsymbol{\beta})'] + \Psi_2\big),\\
\boldsymbol{\beta}|T, R_2, \boldsymbol{y} &\sim N_p([aR_2^{-1} + Q_{\boldsymbol{\beta}}]^{-1}[aR_2^{-1}T_3 + Q_{\boldsymbol{\beta}}\boldsymbol{\beta}_0], [aR_2^{-1} + Q_{\boldsymbol{\beta}}]^{-1}),
\end{aligned} \tag{5}$$

where the sufficient reduction statistics T is a p_2-dimensional vector, $T = (T_{0v}', T_{1v}', T_{2v}', T_3')'$, $T_{kv} = \mathrm{vech}(T_k),\ k = 0, 1, 2$, and

$$\begin{aligned}
T_0 &= \frac{1}{abc}\sum_{i,j,k}(\boldsymbol{v}_{ij} - \boldsymbol{y}_{ijk})(\boldsymbol{v}_{ij} - \boldsymbol{y}_{ijk})',\ T_1 = \frac{1}{ab}\sum_{i,j}(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)',\\
T_2 &= \frac{1}{a}\sum_i(\boldsymbol{v}_i - \bar{\boldsymbol{y}})(\boldsymbol{v}_i - \bar{\boldsymbol{y}})',\ T_3 = \frac{1}{a}\sum_i \boldsymbol{v}_i,\ \bar{\boldsymbol{y}} = \frac{1}{abc}\sum_{i,j,k}\boldsymbol{y}_{ijk}.
\end{aligned} \tag{6}$$

Note that $T_0, \ldots, T_3$ are functions of random effects $\mathcal{V}$ (and data), and the conditional posteriors for parameters depend on the random effects only through T. Hence, as Proposition 1, define $\Theta_a = (\Theta_{1a}', \ldots, \Theta_{5a}')' = (T', R_{0v}', R_{1v}', R_{2v}', \boldsymbol{\beta}')'$, and define some sums of squares and the averages of the observed data used in the following lemma and theorems as follows.

$$S_0 = \frac{1}{abc}\sum_{i,j,k}(\boldsymbol{y}_{ijk} - \bar{\boldsymbol{y}}_{ij})(\boldsymbol{y}_{ijk} - \bar{\boldsymbol{y}}_{ij})',\ S_1 = \frac{1}{ab}\sum_{i,j}(\bar{\boldsymbol{y}}_{ij} - \bar{\boldsymbol{y}}_i)(\bar{\boldsymbol{y}}_{ij} - \bar{\boldsymbol{y}}_i)',$$
$$S_2 = \frac{1}{a}\sum_{i}(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})',\ \bar{\boldsymbol{y}}_{ij} = \frac{1}{c}\sum_k \boldsymbol{y}_{ijk},\ \bar{\boldsymbol{y}}_i = \frac{1}{bc}\sum_{j,k}\boldsymbol{y}_{ijk}.$$

Also, assume the hyper-parameters of priors and data satisfy the following conditions:

$$\begin{aligned}&\text{(hyper-parameters) } (a)\ \frac{m_k}{a} \stackrel{a\to\infty}{\longrightarrow} 0,\ \frac{\Psi_k}{a} \stackrel{a\to\infty}{\longrightarrow} \mathbf{O}_p,\ k=0,1,2,\ \frac{Q_{\boldsymbol{\beta}}}{a} \stackrel{a\to\infty}{\longrightarrow} \mathbf{O}_p,\\ &\text{(observed data) } (b)\ S_k \stackrel{a\to\infty}{\longrightarrow} \boldsymbol{\mu}_{S_k} > 0,\ k=0,1,2,\ \bar{\boldsymbol{y}} \stackrel{a\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\bar{\boldsymbol{y}}}.\end{aligned} \tag{7}$$

The hyper-parameter assumption is to ensure that the effects of the prior vanish as a goes to infinity (see also Remark 1), i.e. the asymptotic results do not depend on any hyper-parameters. The assumption on the observed data is mainly to ensure that the sequence $\boldsymbol{y}_a = \{S_k{}'s, \bar{\boldsymbol{y}}\}$ in Proportion 1 converges to $\boldsymbol{\mu}_{\boldsymbol{y}} = \{\boldsymbol{\mu}_{S_k}{}'s, \boldsymbol{\mu}_{\bar{\boldsymbol{y}}}\}$.

Throughout the paper, we suppress the notation $\boldsymbol{\mu}$ and the subscript of $\boldsymbol{y}_a$ for brevity on data part when there is no ambiguity. For example, we use S_0 and $\boldsymbol{y}$ instead of $\boldsymbol{\mu}_{S_0}$ and $\boldsymbol{y}_a$, respectively. Then the conditional mean function of Θ_a is defined as follows.

$$\begin{aligned}\boldsymbol{g}_a &= (\boldsymbol{g}'_{1a}, \ldots, \boldsymbol{g}'_{5a})'\\ &= (E[T'|\boldsymbol{\phi}, \boldsymbol{y}], E[R'_{0v}|T, \boldsymbol{y}], E[R'_{1v}|T, \boldsymbol{y}], E[R'_{2v}|T, \boldsymbol{\beta}, \boldsymbol{y}], E[\boldsymbol{\beta}'|T, R_{2v}, \boldsymbol{y}])'.\end{aligned}$$

Lemma 1 *Suppose $m_k = p+1$, $\Psi_k = \mathbf{O}_p$, $k = 0, 1, 2$, and $Q_{\boldsymbol{\beta}} = \mathbf{O}_p$. Then the conditional mean function, $\boldsymbol{g}_a$, has an explicit unique fixed point $\boldsymbol{\mu}_{\Theta_a} \stackrel{a\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\Theta}$ provided that $b, c > 1$ and $\boldsymbol{\mu}_{R_k} > 0$, $k = 0, 1, 2$, where*

$$\begin{aligned}&\boldsymbol{\mu}_{\Theta_a} = (\boldsymbol{\mu}'_T, \boldsymbol{\mu}'_{R_{0v}}, \boldsymbol{\mu}'_{R_{1v}}, \boldsymbol{\mu}'_{R_{2v}}, \boldsymbol{\mu}'_{\boldsymbol{\beta}})',\ \boldsymbol{\mu}_T = (\boldsymbol{\mu}'_{T_{0v}}, \boldsymbol{\mu}'_{T_{1v}}, \boldsymbol{\mu}'_{T_{2v}}, \boldsymbol{\mu}'_{T_3})',\\ &\boldsymbol{\mu}_{R_{kv}} = \mathrm{vech}(\boldsymbol{\mu}_{R_k}),\ \boldsymbol{\mu}_{T_{kv}} = \mathrm{vech}(\boldsymbol{\mu}_{T_k}),\ k = 0, 1, 2,\\ &\boldsymbol{\mu}_{R_0} = \boldsymbol{\mu}_{T_0} = \frac{c}{c_1}S_0,\ \boldsymbol{\mu}_{R_1} = \boldsymbol{\mu}_{T_1} = \frac{b}{b_1}S_1 - \frac{1}{c_1}S_0,\\ &\boldsymbol{\mu}_{R_2} = \boldsymbol{\mu}_{T_2} = S_2 - \frac{1}{b_1}S_1,\ \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{T_3} = \bar{\boldsymbol{y}},\ b_1 = b-1,\ c_1 = c-1.\end{aligned} \tag{8}$$

Note that $\boldsymbol{\mu}_{\Theta}$ and $\boldsymbol{\mu}_{\Theta_a}$ have the same functional forms except that we substitute data with their limiting values, and there maybe exist some hyper-parameters setting other than Lemma 1 to have the explicit unique fixed point of $\boldsymbol{g}_a$ for finite a, yet it needs to be justified.

Theorem 1 *Suppose the conditions in (7) hold and $\boldsymbol{g}_a$ has a fixed point $\boldsymbol{\mu}_{\Theta_a}$, which may not be the same as that in Lemma 1, converging to $\boldsymbol{\mu}_{\Theta}$ as $a \to \infty$, then*

$$\sqrt{a}[\Theta_a - \boldsymbol{\mu}_{\Theta_a}] \xrightarrow{\mathcal{L}} N_{2p_2}(\mathbf{0}, \Sigma),$$

where $\Sigma = \begin{pmatrix} \Sigma_T & \Sigma_{T\boldsymbol{\phi}} \\ \Sigma'_{T\boldsymbol{\phi}} & \Sigma_{\boldsymbol{\phi}} \end{pmatrix}$, $\Sigma_T = \Sigma_{T\boldsymbol{\phi}} = \Sigma'_{T\boldsymbol{\phi}} = \Sigma_{\boldsymbol{\phi}} - \Sigma_2^*$, *and*

$$\Sigma_{\boldsymbol{\phi}} = 2 \begin{pmatrix} \frac{c^2}{bc_1^3} S_0{}^{\otimes 2h} & -\frac{c}{bc_1^3} S_0{}^{\otimes 2h} & O_{p_1} & O_{p_1} \\ -\frac{c}{bc_1^3} S_0{}^{\otimes 2h} & (\frac{1}{bc_1^3} S_0{}^{\otimes 2h} + \frac{b^2}{b_1^3} S_1{}^{\otimes 2h}) & -\frac{b}{b_1^3} S_1{}^{\otimes 2h} & O_{p_1} \\ O_{p_1} & -\frac{b}{b_1^3} S_1{}^{\otimes 2h} & (\frac{1}{b_1^3} S_1{}^{\otimes 2h} + S_2{}^{\otimes 2h}) & O_{p_1} \\ O_{p_1} & O_{p_1} & O_{p_1} & \frac{1}{2} S_2 \end{pmatrix},$$

$$\Sigma_2^* = 2 \cdot \text{blkdiag}\{\frac{1}{bc} \boldsymbol{\mu}_{R_0}{}^{\otimes 2h}, \frac{1}{b} \boldsymbol{\mu}_{R_1}{}^{\otimes 2h}, \boldsymbol{\mu}_{R_2}{}^{\otimes 2h}, \frac{1}{2} \boldsymbol{\mu}_{R_2}\}. \tag{9}$$

3.3 The standard case

For the standard model in (1), the conditional posteriors of R_0, R_1, R_2, and $\boldsymbol{\beta}$ are

$$\begin{aligned} R_0 | T_s, \boldsymbol{\beta}, \boldsymbol{y} &\sim \text{IW}_p\big(abc + m_0, abc[T_0 - \boldsymbol{\beta} T_3' - T_3 \boldsymbol{\beta}' + \boldsymbol{\beta}\boldsymbol{\beta}'] + \Psi_0\big), \\ R_1 | T_s, \boldsymbol{y} &\sim \text{IW}_p\big(ab + m_1, abT_1 + \Psi_1\big), \\ R_2 | T_s, \boldsymbol{y} &\sim \text{IW}_p\big(a + m_2, aT_2 + \Psi_2\big), \\ \boldsymbol{\beta} | T_s, R_0, y &\sim N_p([abcR_0^{-1} + Q_{\boldsymbol{\beta}}]^{-1}[abcR_0^{-1} T_3 + Q_{\boldsymbol{\beta}} \boldsymbol{\beta}_0], [abcR_0^{-1} + Q_{\boldsymbol{\beta}}]^{-1}), \end{aligned} \tag{10}$$

where the sufficient reduction statistics $T_s = (T'_{0v}, T'_{1v}, T'_{2v}, T'_3)'$ and

$$\begin{aligned} T_0 &= \frac{1}{ab} \sum_{i,j} (\boldsymbol{u}_{ij} + \boldsymbol{u}_i - \bar{\boldsymbol{y}}_{ij})(\boldsymbol{u}_{ij} + \boldsymbol{u}_i - \bar{\boldsymbol{y}}_{ij})' + S_0, \ T_1 = \frac{1}{ab} \sum_{i,j} \boldsymbol{u}_{ij} \boldsymbol{u}'_{ij}, \\ T_2 &= \frac{1}{a} \sum_i \boldsymbol{u}_i \boldsymbol{u}'_i, \ T_3 = \frac{1}{ab} \sum_{i,j} (\bar{\boldsymbol{y}}_{ij} - \boldsymbol{u}_{ij} - \boldsymbol{u}_i). \end{aligned}$$

Define the conditional mean function of Θ_a as follows:

$$\boldsymbol{g}_a = (E[T'_s | \boldsymbol{\phi}, \boldsymbol{y}], E[R'_{0v} | T_s, \boldsymbol{\beta}, \boldsymbol{y}], E[R'_{1v} | T_s, \boldsymbol{y}], E[R'_{2v} | T_s, \boldsymbol{y}], E[\boldsymbol{\beta}' | T_s, R_{0v}, \boldsymbol{y}])'.$$

Then, by combining Lemma 1 and Theorem 1 as in the centering case, we have the following theorem.

Theorem 2 *Suppose the hyper-parameter assumptions in Lemma 1 and the condition of the observed data in (7) hold. Then*

(a) *The conditional mean function, $\boldsymbol{g}_a$, has an explicit unique fixed point $\boldsymbol{\mu}_{\Theta_a} \overset{a\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\Theta}$ provided that $b, c > 1$ and $\boldsymbol{\mu}_{R_k} > 0,\ k = 0, 1, 2$, where $\boldsymbol{\mu}_{\Theta_a} = (\boldsymbol{\mu}'_{T_s}, \boldsymbol{\mu}'_{\boldsymbol{\phi}})'$ is the same as that from the centering case except $\boldsymbol{\mu}_{T_{0v}} = \frac{c}{c_1} S_{0v} + \mathrm{vech}(\bar{\mathbf{y}}\bar{\mathbf{y}}')$;*

(b)

$$\sqrt{a}[\Theta_a - \boldsymbol{\mu}_{\Theta_a}] \overset{\mathcal{L}}{\longrightarrow} N_{2p_2}(\mathbf{0}, \Sigma_s),$$

where

$$\Sigma_s = \begin{pmatrix} \Sigma_{T_s} & \Sigma_{T_s\boldsymbol{\phi}} \\ \Sigma'_{T_s\boldsymbol{\phi}} & \Sigma_{\boldsymbol{\phi}} \end{pmatrix},\ \Sigma_{T_s} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma^*_{2s})B_s^{-1'},\ \Sigma_{T_s\boldsymbol{\phi}} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma^*_{2s}),$$

$$\Sigma^*_{2s} = 2 \cdot \mathrm{blkdiag}\{\frac{1}{bc}\boldsymbol{\mu}_{R_0}{}^{\otimes 2h}, \frac{1}{b}\boldsymbol{\mu}_{R_1}{}^{\otimes 2h}, \boldsymbol{\mu}_{R_2}{}^{\otimes 2h}, \frac{1}{2bc}\boldsymbol{\mu}_{R_0}\},$$

$$B_s = I_{3p_2} + \begin{pmatrix} \mathbf{O}_{p_1,3p_1} & -L_p[\bar{\mathbf{y}} \otimes I_p + I_p \otimes \bar{\mathbf{y}}] \\ \mathbf{O}_{2p_1+p,3p_1} & \mathbf{O}_{2p_1+p,p} \end{pmatrix},$$

and $\Sigma_{\boldsymbol{\phi}}$ is the same as that in Theorem 1. Note that we cannot have the asymptotic normality for $\mathcal{U}|\mathbf{y}$ since the number of random effects and main factor levels a grow to infinity at the same rate.

Remark 1 The centering model is "less correlated" than the standard case in the sense that the T_0 is asymptotically independent of T_3 and $\boldsymbol{\beta}$ for the centering model, while independence does not hold for the standard case. Also, R_1 and R_2 as well as T_1 and T_2 given data are (asymptotically) correlated.

When we use the dispersion parameter $Q_k = R_k^{-1}$ with the prior $Q_k \overset{\mathrm{ind}}{\sim} \mathrm{W}_p(m_k, \Psi_k)$ instead of R_k, the assumptions for the assertion of Lemma 1 are $m_k = 0$ and $\Psi_k = \mathbf{O}_p$, $k = 0, 1, 2$, and $Q_{\boldsymbol{\beta}} = \mathbf{O}_p$, i.e. we put the flat priors on Q_k and $\boldsymbol{\beta}$. The fixed point $\boldsymbol{\mu}_{Q_k}$ is just the inverse matrix of $\boldsymbol{\mu}_{R_k}$. For both parametrizations, when we use other priors, for instance, with their hyper-parameters all equal to constants or with the order of hyper-parameters equal to $o_a(1)$, the limiting fixed point as well as the asymptotic covariance matrix are identical to those using the prior setting in Eq. (3). For the finite-sample case, no explicit conditional mode function for T exists, i.e. no explicit MAP estimate exists for $\Theta_a = (T', \boldsymbol{\phi}')'$ though MAP estimate is asymptotically equivalent to the fixed point of $\boldsymbol{g}_a$. Furthermore, when we use the priors other than those specified in Lemmas 1, the fixed point for $\boldsymbol{g}_a$ can be calculated through the recursive algorithm $\boldsymbol{t}_a^{(k+1)} = \boldsymbol{g}_a(\boldsymbol{t}_a^{(k)}),\ k = 1, 2, \ldots$, and the asymptotic (or asymptotically locally linear) convergence rate of this algorithm is equal to $\rho(\dot{\boldsymbol{g}}(\boldsymbol{\mu}_{\boldsymbol{\Theta}}))$ (see Meng 1994 or Section 2.1 in Su and Johnson 2006).

Corollary 1 *Suppose the assumptions in Theorems 1 and 2 hold, then the asymptotic convergence rates for the standard and centering models are both less than one, and their values are $\rho_s = \sqrt{\rho(I_{p_2} - \Sigma^*_{2s}\Sigma_{\boldsymbol{\phi}}^{-1})}$ and $\rho_c = \sqrt{\rho(I_{p_2} - \Sigma^*_2\Sigma_{\boldsymbol{\phi}}^{-1})}$, respectively. Moreover, $\rho_c \le (\ge)\rho_s$ provided $\lambda_{min}(\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) \le (\ge)\ \lambda_{min}(\boldsymbol{\mu}_{R_2}S_2^{-1})$ or simply $\boldsymbol{\mu}_{R_2} - \frac{1}{bc}\boldsymbol{\mu}_{R_0} \ge (\le)0$, $\rho_c \to 0$ as b and $c \to \infty$, and $\rho_s \to 1$ as b or $c \to \infty$.*

As b or c is getting larger, we expect a lower asymptotic convergence rate for the centering case and this rate is, in general, less than that from the standard case.

Remark 2 For either the standard or centering case, more general settings by relaxing distribution assumption can be considered. For example, in Eq. (1), one can assume the distribution of $\boldsymbol{u}_i$ given R_2 follows from a multivariate t-distribution; i.e. $\boldsymbol{u}_i|R_2 \overset{\text{iid}}{\sim} t_p(\mathbf{0}, R_2, \nu)$, $i = 1, \dots, a$, where ν denotes a known degrees of freedom. This is equivalent to assuming, after introducing some latent variables τ_i's, $\boldsymbol{u}_i|R_2, \tau_i \overset{\text{ind}}{\sim} N_p(\mathbf{0}, R_2/\tau_i)$, and $\tau_i \overset{\text{iid}}{\sim} \Gamma(\nu/2, \nu/2)$, $i = 1, \dots, a$. The corresponding sufficient reduction statistic T_s and the conditional mean function $\boldsymbol{g}_a$ can be derived; however, there exists no explicit fixed point for $\boldsymbol{g}_a$.

4 Multi-way balanced nested MANOVA model with random effects

In this section, we generalize the finite-sample and asymptotic results from the two-way case to the three- and higher-way cases under two different parametrizations as shown in Sect. 3.

4.1 The centering case

For the centering r-way balanced nested case, $r \geq 3$, the model is given as

$$\begin{aligned} \boldsymbol{y}_{i_0\dots i_r} &= \boldsymbol{v}_{i_0\dots i_{r-1}} + \boldsymbol{\epsilon}_{i_0\dots i_r},\ \boldsymbol{\epsilon}_{i_0\dots i_r} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, R_0), \\ &\quad \boldsymbol{v}_{i_0\dots i_k}|\boldsymbol{v}_{i_0\dots i_{k-1}}, R_{r-k} \overset{\text{iid}}{\sim} N_p(\boldsymbol{v}_{i_0\dots i_{k-1}}, R_{r-k}),\ k = r-1, \dots, 1, \\ &\quad \boldsymbol{v}_{i_0}|\boldsymbol{\beta}, R_r \overset{\text{iid}}{\sim} N_p(\boldsymbol{\beta}, R_r),\ i_r = 1, \dots, n_r; \dots; i_0 = 1, \dots, n_0, \end{aligned} \tag{11}$$

where $\boldsymbol{y}_{i_0\dots i_r}$ denotes the p-dimensional random vector, and the error terms $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_{i_0\dots i_r}{}'s\}$ and the random effects $\mathcal{V} = \{\boldsymbol{v}_{i_0\dots i_{r-1}}{}'s, \dots, \boldsymbol{v}_{i_0}{}'s\}$ are assumed to be independent. In this case, $\boldsymbol{v}_{i_0\dots i_{k-1}}$ and R_{r-k} can be viewed as the $(r-k+1)$th-level latent variable and parameter, $k = r, \dots, 1$, respectively, and $\boldsymbol{\beta}$ and R_r as the $(r+1)$th-level parameters (Su 2017). For simplicity, assume the priors for $R_0, \dots, R_r$, and $\boldsymbol{\beta}$ are proportional to

$$\prod_{k=0}^{r} |R_k|^{-(p+1)}. \tag{12}$$

Then the conditional posteriors for $R_0, \dots, R_r$, and $\boldsymbol{\beta}$ are

$$\begin{aligned} &R_k|T, \boldsymbol{y} \sim \mathrm{IW}_p\big(n_0 \dots n_{r-k} + p + 1, n_0 \dots n_{r-k} T_k\big),\ k = 0, \dots, r-1, \\ &R_r|T, \boldsymbol{\beta}, \boldsymbol{y} \sim \mathrm{IW}_p\big(n_r + p + 1, a[T_r + (T_{r+1} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}} - \boldsymbol{\beta})' \\ &\quad +(\bar{\boldsymbol{y}} - \boldsymbol{\beta})(T_{r+1} - \bar{\boldsymbol{y}})' + (\bar{\boldsymbol{y}} - \boldsymbol{\beta})(\bar{\boldsymbol{y}} - \boldsymbol{\beta})']\big), \\ &\quad \boldsymbol{\beta}|T, R_r, \boldsymbol{y} \sim N_p(T_{r+1}, \frac{1}{n_0} R_r), \end{aligned} \tag{13}$$

where the sufficient reduction statistics T is an $(r+1)p_1+p$ dimensional vector, $T=(T_{0v}',\ldots,T_{rv}',T_{r+1}')'$,

$$\begin{aligned}
T_0 =&\frac{1}{n_0\ldots n_r}\sum_{i_0\ldots i_r}(\boldsymbol{y}_{i_0\ldots i_r}-\boldsymbol{v}_{i_0\ldots i_{r-1}})(\boldsymbol{y}_{i_0\ldots i_r}-\boldsymbol{v}_{i_0\ldots i_{r-1}})',\\
T_k =&\frac{1}{n_0\ldots n_{r-k}}\sum_{i_0\ldots i_{r-k}}(\boldsymbol{v}_{i_0\ldots i_{r-k}}-\boldsymbol{v}_{i_0\ldots i_{r-k-1}})\\
&(\boldsymbol{v}_{i_0\ldots i_{r-k}}-\boldsymbol{v}_{i_0\ldots i_{r-k-1}})',\ k=1,\ldots,r-1,\\
T_r =&\frac{1}{n_0}\sum_{i_0}(\boldsymbol{v}_{i_0}-\bar{\boldsymbol{y}})(\boldsymbol{v}_{i_0}-\bar{\boldsymbol{y}})',\ T_{r+1}=\frac{1}{n_0}\sum_{i_0}\boldsymbol{v}_{i_0},\ \bar{\boldsymbol{y}}=\frac{1}{n_0\ldots n_r}\sum_{i_0\ldots i_r}\boldsymbol{y}_{i_0\ldots i_r},
\end{aligned}\tag{14}$$

and $\sum_{i_0\ldots i_k}$ represents $\sum_{i_0=1}^{n_0}\ldots\sum_{i_k=1}^{n_k},\ k=0,\ldots,r$. As in Sect. 3 , define

$$\begin{aligned}
&\Theta_n=(\Theta_{1n}',\Theta_{2n}',\cdots,\Theta_{(r+2)n}',\Theta_{(r+3)n})'=(T',R_{0v}',\ldots,R_{rv}',\boldsymbol{\beta}')'=(T',\boldsymbol{\phi}')',\\
&\boldsymbol{\phi}=(R',\boldsymbol{\beta}')',\ R=(R_{0v}',\ldots,R_{rv}')'.
\end{aligned}$$

Also, define some sums of squares and the averages of the observed data.

$$\begin{aligned}
S_0 &= \frac{1}{n_0\ldots n_r}\sum_{i_0\ldots i_r}(\boldsymbol{y}_{i_0\ldots i_r}-\bar{\boldsymbol{y}}_{i_0\ldots i_{r-1}})(\boldsymbol{y}_{i_0\ldots i_r}-\bar{\boldsymbol{y}}_{i_0\ldots i_{r-1}})',\\
S_k &= \frac{1}{n_0\ldots n_{r-k}}\sum_{i_0\ldots i_{r-k}}(\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}}-\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k-1}})(\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}}-\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k-1}})',\\
& k=1,\ldots,r-1,\ S_r=\frac{1}{n_0}\sum_{i_0}(\bar{\boldsymbol{y}}_{i_0}-\bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_{i_0}-\bar{\boldsymbol{y}})',\\
\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}} &= \frac{1}{n_{r+1-k}\cdots n_r}\sum_{i_{r+1-k}\cdots i_r}\boldsymbol{y}_{i_0\ldots i_r},\quad k=1,\ldots,r.
\end{aligned}$$

Assume the observed data satisfy the following conditions:

$$S_k\overset{n_0\to\infty}{\longrightarrow}\boldsymbol{\mu}_{S_k}>0,\ k=0,\ldots,r,\ \bar{\boldsymbol{y}}\overset{n_0\to\infty}{\longrightarrow}\boldsymbol{\mu}_{\bar{\boldsymbol{y}}},\tag{15}$$

as in Eq. (7). Set $n_j\ldots n_k=1$ if $j>k$ and the notations n in Proposition 1 and n_0 in our model are used interchangeably hereinafter. Then, define the conditional mean function of Θ_n as follows.

$$\begin{aligned}
\boldsymbol{g}_n &= (\boldsymbol{g}_{1n}',\boldsymbol{g}_{2n}',\cdots,\boldsymbol{g}_{(r+2)n}',\boldsymbol{g}_{(r+3)n}')'\\
&=(E[T'|\boldsymbol{\phi},\boldsymbol{y}],E[R_{0v}'|T,\boldsymbol{y}],\ldots,E[R_{rv}'|T,\boldsymbol{\beta},\boldsymbol{y}],E[\boldsymbol{\beta}'|T,R_{rv},\boldsymbol{y}])'.
\end{aligned}$$

Theorem 3 *Suppose the prior setting in* (12) *and the conditions of the observed data in* (15) *hold. Then*

(a) *The conditional mean function, $\mathbf{g}_n$, has an explicit unique fixed point $\boldsymbol{\mu}_{\Theta_n} \overset{n\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\Theta}$ provided that $n_1, \ldots, n_r > 1$, $\boldsymbol{\mu}_{R_k} > 0$, $k = 0, \ldots, r$, where*

$$\boldsymbol{\mu}_{\Theta_n} = (\boldsymbol{\mu}_T', \boldsymbol{\mu}_{\boldsymbol{\phi}}')', \ \boldsymbol{\mu}_{\boldsymbol{\phi}}' = (\boldsymbol{\mu}_R', \boldsymbol{\mu}_{\boldsymbol{\beta}}')', \ \boldsymbol{\mu}_R' = (\boldsymbol{\mu}_{R_{0v}}', \ldots \boldsymbol{\mu}_{R_{rv}}')', \ \boldsymbol{\mu}_T = (\boldsymbol{\mu}_{T_{0v}}', \ldots, \boldsymbol{\mu}_{T_{rv}}', \boldsymbol{\mu}_{T_{r+1}}')',$$
$$\boldsymbol{\mu}_{R_0} = \boldsymbol{\mu}_{T_0} = \frac{n_r}{n_r - 1} S_0, \ \boldsymbol{\mu}_{R_k} = \boldsymbol{\mu}_{T_k} = \frac{n_{r-k}}{n_{r-k} - 1} S_k - \frac{1}{n_{r-k+1} - 1} S_{k-1}, \ k = 1, \ldots, r-1,$$
$$\boldsymbol{\mu}_{R_r} = \boldsymbol{\mu}_{T_r} = S_r - \frac{1}{n_1 - 1} S_{r-1}, \ \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{T_{r+1}} = \bar{\mathbf{y}}, \ \boldsymbol{\mu}_{R_{kv}} = \mathrm{vech}(\boldsymbol{\mu}_{R_k}), \ k = 0, \ldots, r. \quad (16)$$

(b) $\sqrt{n_0}[\Theta_n - \boldsymbol{\mu}_{\Theta_n}] \overset{\mathcal{L}}{\longrightarrow} N_{(2r+2)p_1+2p}(\mathbf{0}, \Sigma)$, *where*

$$\Sigma = \begin{pmatrix} \Sigma_T & \Sigma_{T\boldsymbol{\phi}} \\ \Sigma_{T\boldsymbol{\phi}}' & \Sigma_{\boldsymbol{\phi}} \end{pmatrix}, \ \Sigma_T = \Sigma_{T\boldsymbol{\phi}} = \Sigma_{T\boldsymbol{\phi}}' = \Sigma_{\boldsymbol{\phi}} - \Sigma_2^*,$$
$$\Sigma_2^* = 2 \cdot \mathrm{blkdiag}\{\frac{1}{n_1 \ldots n_r} \boldsymbol{\mu}_{R_0}{}^{\otimes 2h}, \ldots, \frac{1}{n_1} \boldsymbol{\mu}_{R_{r-1}}{}^{\otimes 2h}, \boldsymbol{\mu}_{R_r}{}^{\otimes 2h}, \frac{1}{2} \boldsymbol{\mu}_{R_r}\}.$$

Here $\Sigma_{\boldsymbol{\phi}}$ is a $(r+2) \times (r+2)$ symmetric block tri-diagonal matrix, and

$$\Sigma_{\boldsymbol{\phi}} = 2 \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & & & \\ \Sigma_{12} & \Sigma_{22} & \Sigma_{23} & & \\ & \Sigma_{23} & \ddots & \ddots & \\ & & \ddots & \Sigma_{r+1,r+1} & \Sigma_{r+1,r+2} \\ & & & \Sigma_{r+1,r+2} & \Sigma_{r+2,r+2} \end{pmatrix}, \quad (17)$$

where

$$\Sigma_{11} = \frac{n_r^2 S_0{}^{\otimes 2h}}{n_1 \ldots n_{r-1}(n_r - 1)^3},$$
$$\Sigma_{kk} = \frac{1}{n_1 \ldots n_{r-k}} \left[\frac{S_{k-2}{}^{\otimes 2h}}{n_{r-k+1}(n_{r-k+2} - 1)^3} + \frac{n_{r-k+1}^2 S_{k-1}{}^{\otimes 2h}}{(n_{r-k+1} - 1)^3} \right], \ k = 2, \ldots, r-1,$$
$$\Sigma_{rr} = \frac{S_{r-2}{}^{\otimes 2h}}{n_1(n_2 - 1)^3} + \frac{n_1^2 S_{r-1}{}^{\otimes 2h}}{(n_1 - 1)^3}, \ \Sigma_{r+1,r+1} = \frac{S_{r-1}{}^{\otimes 2h}}{(n_1 - 1)^3} + S_r{}^{\otimes 2h},$$
$$\Sigma_{r+2,r+2} = \frac{1}{2} S_r, \ \Sigma_{k,k+1} = -\frac{n_{r-k+1} S_{k-1}{}^{\otimes 2h}}{n_1 \ldots n_{r-k}(n_{r-k+1} - 1)^3}, \ k = 1, \ldots, r-1,$$
$$\Sigma_{r,r+1} = -\frac{n_1 S_{r-1}{}^{\otimes 2h}}{(n_1 - 1)^3}, \ \Sigma_{r+1,r+2} = \mathbf{O}_{p_1,p}.$$

4.2 The standard case

In the standard case, the fixed and all the random effects are postulated as the first-stage parameter and latent variables, respectively, whereas the variance component parameters as the second-stage parameters, i.e.

$$
\begin{aligned}
\boldsymbol{y}_{i_0\ldots i_r} &= \boldsymbol{\beta} + \boldsymbol{u}_{i_0} + \cdots + \boldsymbol{u}_{i_0\ldots i_{r-1}} + \boldsymbol{\epsilon}_{i_0\ldots i_r},\ \boldsymbol{\epsilon}_{i_0\ldots i_r} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, R_0),\\
&\boldsymbol{u}_{i_0\ldots i_k}|R_{r-k} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, R_{r-k}),\\
&k = r-1, \ldots, 0,\ i_r = 1, \ldots, n_r; \ldots; i_0 = 1, \ldots, n_0,
\end{aligned}
\tag{18}
$$

where the error terms $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_{i_0\ldots i_r}{}'s\}$ and the random effects $\mathcal{U} = \{\boldsymbol{u}_{i_0\ldots i_{r-1}}{}'s, \ldots, \boldsymbol{u}_{i_0}{}'s\}$ are assumed to be independent. For the standard model in (18), the conditional posteriors of $R_0, \ldots, R_r$, and $\boldsymbol{\beta}$ are

$$
\begin{aligned}
R_0|T_s, \boldsymbol{\beta}, \boldsymbol{y} &\sim \mathrm{IW}_p\big(n_0\ldots n_r + p + 1, n_0\ldots n_r[T_0 - \boldsymbol{\beta}T_{r+1}' - T_{r+1}\boldsymbol{\beta}' + \boldsymbol{\beta}\boldsymbol{\beta}'] + \Psi_0\big),\\
R_k|T_s, \boldsymbol{y} &\sim \mathrm{IW}_p\big(n_0\ldots n_{r-k} + p + 1, n_0\ldots n_{r-k}T_k + \Psi_k\big),\ k = 1, \ldots, r,\\
\boldsymbol{\beta}|T_s, R_0, \boldsymbol{y} &\sim N_p(T_{r+1}, \frac{1}{n_0\ldots n_r}R_0),
\end{aligned}
\tag{19}
$$

where the sufficient reduction statistics $T_s = (T_{0v}', \ldots, T_{rv}', T_{r+1}')'$, and

$$
\begin{aligned}
T_0 =& \frac{1}{n_0\ldots n_{r-1}} \sum_{i_0\ldots i_{r-1}} (\boldsymbol{u}_{i_0\ldots i_{r-1}} + \cdots + \boldsymbol{u}_{i_0} - \bar{\boldsymbol{y}}_{i_0\ldots i_{r-1}})\\
&(\boldsymbol{u}_{i_0\ldots i_{r-1}} + \cdots + \boldsymbol{u}_{i_0} - \bar{\boldsymbol{y}}_{i_0\ldots i_{r-1}})' + S_0,\\
T_k =& \frac{1}{n_0\ldots n_{r-k}} \sum_{i_0\ldots i_{r-k}} \boldsymbol{u}_{i_0\ldots i_{r-k}}\boldsymbol{u}_{i_0\ldots i_{r-k}}',\ k = 1, \ldots, r,\\
T_{r+1} =& \frac{1}{n_0\ldots n_{r-1}} \sum_{i_0} (\bar{\boldsymbol{y}}_{i_0\ldots i_{r-1}} - \boldsymbol{u}_{i_0} - \cdots - \boldsymbol{u}_{i_0\ldots i_{r-1}}).
\end{aligned}
$$

Then we have the following theorem and corollary.

Theorem 4 *Suppose the prior setting in (12) and the conditions of the observed data in (15) hold. Then*

(a) the conditional mean function, $\mathbf{g}_n$, has an explicit unique fixed point $\boldsymbol{\mu}_{\Theta_n} \xrightarrow{n\to\infty} \boldsymbol{\mu}_{\Theta}$ provided that $n_1, \ldots, n_r > 1$, $\boldsymbol{\mu}_{R_k} > 0,\ k = 0, \ldots, r$, where $\boldsymbol{\mu}_{\Theta_n} = (\boldsymbol{\mu}_T', \boldsymbol{\mu}_{\boldsymbol{\phi}}')'$ is the same as that from the centering case except $\boldsymbol{\mu}_{T_0} = \frac{n_r}{n_r-1}S_0 + \mathrm{vech}(\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}')$;

(b) $\sqrt{n_0}[\Theta_n - \boldsymbol{\mu}_{\Theta_n}] \xrightarrow{\mathcal{L}} N_{(2r+2)p_1+2p}(\mathbf{0}, \Sigma_s)$,

$$
\Sigma_s = \begin{pmatrix} \Sigma_{T_s} & \Sigma_{T_s\boldsymbol{\phi}} \\ \Sigma_{T_s\boldsymbol{\phi}}' & \Sigma_{\boldsymbol{\phi}} \end{pmatrix},\ \Sigma_{T_s} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma_{2s}^*)B_s^{-1'},\ \Sigma_{T_s\boldsymbol{\phi}} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma_{2s}^*),
$$

$$
\Sigma_{2s}^* = 2\cdot\mathrm{blkdiag}\{\frac{1}{n_1\ldots n_r}\boldsymbol{\mu}_{R_0}^{2\ \otimes 2h}, \ldots, \frac{1}{n_1}\boldsymbol{\mu}_{R_{r-1}}^{2\ \otimes 2h}, \boldsymbol{\mu}_{R_r}^{2\ \otimes 2h}, \frac{1}{2n_1\ldots n_r}\boldsymbol{\mu}_{R_0}\},
$$

where $B_s = I_{(r+1)p_1+p} + \begin{pmatrix} \mathbf{O}_{p_1,(r+1)p_1} & -L_p[\bar{\mathbf{y}} \otimes I_p + I_p \otimes \bar{\mathbf{y}}] \\ \mathbf{O}_{rp_1+p,(r+1)p_1} & \mathbf{O}_{rp_1+p,p} \end{pmatrix}$, *and* $\Sigma_{\boldsymbol{\phi}}$ *is the same as that in Theorem 3.*

Corollary 2 *Suppose the assumptions in Theorems 3 and 4 hold, then the asymptotically convergence rates for the standard and centering models are both less than one, and their values are* $\rho_s = \sqrt{\rho(I_{(r+1)p_1+p} - \Sigma_{2s}^{*}\Sigma_{\boldsymbol{\phi}}^{-1})}$ *and* $\rho_c = \sqrt{\rho(I_{(r+1)p_1+p} - \Sigma_{2}^{*}\Sigma_{\boldsymbol{\phi}}^{-1})}$, *respectively. Moreover,* $\rho_c \leq (\geq)\rho_s$ *provided* $\lambda_{min}(\frac{1}{n_1\cdots n_r}\boldsymbol{\mu}_{R_0}S_r^{-1}) \leq (\geq)\ \lambda_{min}(\boldsymbol{\mu}_{R_r}S_r^{-1})$ *or simply* $\boldsymbol{\mu}_{R_r} - \frac{1}{n_1\cdots n_r}\boldsymbol{\mu}_{R_0} \geq (\leq)0$, $\rho_c \to 0$ *as* $n_1, \ldots, n_r \to \infty$, *and* $\rho_s \to 1$ *as any* $n_k \to \infty$.

5 Simulation

In this section, we compare our asymptotics discussed in Sect. 3, called M_{SR}, with those from the Gibbs sampling approach, called M_{GS}, and from the classical large-sample numerical approach, called M_{EM}, by which we calculate the MAP estimates and their standard deviations through the Markov chain EM (MCEM) algorithm. To reduce the sampling error for a particular data set, we generate 100 data sets, each of which is generated from a balanced nested two-way MANOVA model with random effects with $\boldsymbol{\beta} = (10, -10)'$, $R_{0v} = (1, 0.7, 1)'$, $R_{1v} = (2, 1, 2)'$, $R_{2v} = (4, 2, 4)'$, $(b, c) = (4, 2)$, and different numbers of the main factor levels, i.e. a=15, 30, 50. Other combinations of (a, b, c) such as (15, 10, 2) and (15, 4, 4) are also calculated to investigate the trend of the posterior estimates as b or c increases larger (not reported here). We then calculate the averages of the estimated posterior means (medians or modes) and s.d.'s for $\boldsymbol{\phi}$ with the prior set as Lemma 1. The simulation results are given in Table 1.

In Table 1, the average of the estimated posterior means (medians or modes) for $\boldsymbol{\beta}$ are almost identical to the default setting values for all approaches, though the average of the estimated posterior s.d.'s using EM method is smaller than those from the other two approaches. For the variance component parameters, $R = (R_{0v}', R_{1v}', R_{2v}')'$, when the number of main effect levels a is small, the averages of the estimates from M_{GS} are closer to the default settings than those from M_{EM} and M_{SR}. Especially for the R_{2v}, the average estimates of s.d.'s from M_{EM}, which are much smaller than that from M_{SR} or M_{GS}'s, severely deviate from the default. This is mainly due to the highly skewed posterior distribution for R_{2v}. When a is getting larger, the results from M_{GS} and M_{SR} are, as expected, closer to the default settings and are still better than those from M_{EM}, and increasing b or c somewhat improves the results for all three approaches. Furthermore, for the simulated case $(a, b, c) = (30, 4, 2)$, the average asymptotic convergence rates of the standard and centering models are 0.99 and 0.44, respectively, and the rate of convergence in the standard case tends to one as b or c increases.

Table 1 Compare M_{SR} with M_{EM} and M_{GS} methods (simulated data)

Sample size	Parameters	R'_{0v}	R'_{1v}	R'_{2v}	β'
(a,b,c)	Default values	(1.00,0.70,1.00)	(2.00,1.00,2.00)	(4.00,2.00,4.00)	(10.00,−10.00)
	M_{EM}(mode)	(1.00,0.71,0.98)	(1.95,0.95,1.89)	(2.37,1.16,2.32)	(10.12,−10.08)
	M_{GS}(median)	(1.07,0.76,1.05)	(2.09,1.02,2.03)	(3.44,1.63,3.35)	(10.12,−10.08)
	M_{SR}(mean)	(0.98,0.67,0.97)	(2.22,1.15,2.15)	(3.84,1.89,3.74)	(10.12,−10.08)
(15,4,2)	M_{GS}(mean)	(1.10,0.78,1.08)	(2.19,1.08,2.13)	(3.83,1.87,3.73)	(10.12,−10.08)
	M_{EM}(s.d.)	(0.15,0.13,0.15)	(0.46,0.38,0.46)	(0.89,0.69,0.89)	(0.39,0.39)
	M_{SR}(s.d.)	(0.18,0.15,0.18)	(0.59,0.47,0.57)	(1.78,1.40,1.73)	(0.54,0.54)
	M_{GS}(s.d.)	(0.21,0.18,0.21)	(0.66,0.52,0.64)	(1.89,1.44,1.84)	(0.54,0.53)
	M_{EM}(mode)	(0.98,0.68,0.98)	(1.87,0.92,1.89)	(2.98,1.41,3.00)	(10.12,−10.05)
	M_{GS}(median)	(1.01,0.69,0.99)	(1.96,0.99,1.95)	(3.63,1.75,3.92)	(10.12,−10.15)
	M_{SR}(mean)	(1.01,0.71,1.01)	(2.02,1.01,2.03)	(3.80,1.81,3.82)	(10.12,−10.05)
(30,4,2)	M_{GS}(mean)	(1.03,0.72,1.03)	(2.00,0.99,2.02)	(3.82,1.81,3.84)	(10.12,−10.05)
	M_{EM}(s.d.)	(0.13,0.11,0.13)	(0.36,0.29,0.36)	(0.85,0.67,0.86)	(0.34,0.34)
	M_{SR}(s.d.)	(0.13,0.11,0.13)	(0.38,0.31,0.39)	(1.19,0.94,1.20)	(0.38,0.38)
	M_{GS}(s.d.)	(0.14,0.12,0.14)	(0.40,0.32,0.40)	(1.33,0.94,1.23)	(0.38,0.38)
	M_{EM}(mode)	(1.00,0.72,1.02)	(1.92,0.96,1.92)	(3.44,1.71,3.38)	(10.12,−10.13)
	M_{GS}(median)	(1.03,0.73,1.05)	(1.98,0.98,1.98)	(3.89,1.92,3.83)	(10.12,−10.13)
	M_{SR}(mean)	(1.02,0.73,1.04)	(2.02,1.01,2.01)	(3.99,1.98,3.93)	(10.12,−10.13)
(50,4,2)	M_{GS}(mean)	(1.03,0.74,1.05)	(2.00,1.00,2.00)	(4.01,1.99,3.95)	(10.12,−10.13)
	M_{EM}(s.d.)	(0.10,0.09,0.10)	(0.28,0.23,0.28)	(0.74,0.60,0.74)	(0.28,0.28)
	M_{SR}(s.d.)	(0.10,0.09,0.10)	(0.30,0.24,0.30)	(0.95,0.75,0.93)	(0.30,0.30)
	M_{GS}(s.d.)	(0.11,0.09,0.11)	(0.30,0.24,0.30)	(0.96,0.75,0.95)	(0.30,0.30)

6 Conclusion

For the multi-way balanced nested MANOVA models with random effects, we have derived the explicit asymptotic joint posterior for the parameters and the sufficient reduction statistics when the number of the main factors levels approaches infinity. Based on Proposition 1 and/or the sufficient reduction statistics, other types of asymptotic results under different assumptions might be established, e.g., relaxation of distributional assumption or balanced assumption (see Remark 2 or Su and Johnson 2006). We are working on this problem. Compared with the standard parametrization shown in the textbook, the proposed centering parametrization usually has a better (locally linear) convergence rate for calculating the posterior estimates of parameters and sufficient reduction statistics through iterative algorithm.

Simulation results demonstrate that our approach is superior to the classical approximation method on estimating the posterior s.d.'s of variance component parameters. Although numerical approaches based on MCMC are popular and somewhat easier to manipulate for finite sample cases, our approach provides the explicit asymptotic form for parameters, which allows us to look into the structure of the model and inference considered. For instance, one should be cautious in making inferences on the main and nested effects simultaneously for the two-way cases since the posteriors of R_1 and R_2 are correlated. Besides the inferences on variance components parameters, the asymptotics of the sufficient reduction statistics and other functional forms of random effects can be derived through similar procedure. For example, the use of averages of random effects as the prediction of non-sampled area can be seen in Anisa et al. (2014).

Acknowledgements We thank Prof. Pao-sheng Shen for linguistic improvements of this paper, and Editor in Chief and two anonymous reviewers whose remarks greatly contributed to this paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Proof of Lemma 1: Recall that the conditional mean function $\boldsymbol{g}_a = (E[T'|\boldsymbol{\phi}, \boldsymbol{y}], E[R'_{0v}|T, \boldsymbol{y}], E[R'_{1v}|T, \boldsymbol{y}], E[R'_{2v}|T, \boldsymbol{\beta}, \boldsymbol{y}], E[\boldsymbol{\beta}'|T, R_{2v}, \boldsymbol{y}])'$. From the Eq. (4), we obtain

$$\begin{aligned} \boldsymbol{v}_{ij}|\boldsymbol{v}_i, \boldsymbol{\phi}, \boldsymbol{y} &\overset{\text{ind}}{\sim} N_p(W_1^{*-1}[P_0\bar{\boldsymbol{y}}_{ij} + Q_1\boldsymbol{v}_i], W_1^{*-1}),\ j = 1, \ldots, b,\ i = 1, \ldots, a, \\ \boldsymbol{v}_i|\boldsymbol{\phi}, \boldsymbol{y} &\overset{\text{ind}}{\sim} N_p(W_2^{*-1}[P_1\bar{\boldsymbol{y}}_i + Q_2\boldsymbol{\beta}], W_2^{*-1}),\ i = 1, \ldots, a, \end{aligned}$$

where $P_0 = cQ_0$, $P_1 = bcQ_1W_1^{*-1}Q_0$, $Q_k = R_k^{-1}$, $k = 0, 1, 2$, and $W_k^* = P_{k-1} + Q_k$, $k = 1, 2$. Note that, for example, $T_0 = \frac{1}{abc}\sum_{i,j,k}(\boldsymbol{v}_{ij} - \boldsymbol{y}_{ijk})(\boldsymbol{v}_{ij} - \boldsymbol{y}_{ijk})' = \frac{1}{ab}\sum_{i,j}(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})' + S_0$, $E[T_0|\boldsymbol{\phi}, \boldsymbol{y}] = E[E[T_0|\{\boldsymbol{v}_i\text{'s}\}, \boldsymbol{\phi}, \boldsymbol{y}]|\boldsymbol{\phi}, \boldsymbol{y}]$, and

$$\begin{aligned} E[T_0|\{\boldsymbol{v}_i\text{'s}\}, \boldsymbol{\phi}, \boldsymbol{y}] =& \frac{1}{ab}\sum_{i,j}\big[\mathrm{Cov}[(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})|\boldsymbol{v}_i, \boldsymbol{\phi}, \boldsymbol{y}] \\ &+ E[(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})|\boldsymbol{v}_i, \boldsymbol{\phi}, \boldsymbol{y}]E[(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})'|\boldsymbol{v}_i, \boldsymbol{\phi}, \boldsymbol{y}]\big] + S_0 \\ =& W_1^{*-1} + W_1^{*-1}Q_1\Big[\frac{1}{ab}\sum_{i,j}(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_{ij})(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_{ij})'\Big]Q_1W_1^{*-1} + S_0 \\ =& W_1^{*-1} + W_1^{*-1}Q_1[S_1 + \frac{1}{a}\sum_{i}(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_i)(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_i)']Q_1W_1^{*-1} + S_0. \end{aligned}$$

Then, under the settings in Lemma 1, the components of $\boldsymbol{g}_a$ can be calculated through the following conditional mean functions.

$$\begin{aligned}
&E[T_0|\boldsymbol{\phi}, \boldsymbol{y}] = W_1^{*-1} + W_1^{*-1} Q_1[S_1 + E[T_{20}|\boldsymbol{\phi}, \boldsymbol{y}]]Q_1 W_1^{*-1} + S_0,\\
&E[T_1|\boldsymbol{\phi}, \boldsymbol{y}] = W_1^{*-1} + W_1^{*-1} P_0[S_1 + E[T_{20}|\boldsymbol{\phi}, \boldsymbol{y}]]P_0 W_1^{*-1},\\
&E[T_2|\boldsymbol{\phi}, \boldsymbol{y}] = W_2^{*-1} + W_2^{*-1}[P_1 S_2 P_1 + Q_2(\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})' Q_2] W_2^{*-1},\\
&E[T_3|\boldsymbol{\phi}, \boldsymbol{y}] = W_2^{*-1}[P_1 \bar{\boldsymbol{y}} + Q_2 \boldsymbol{\beta}],\\
&E[R_k|T, \boldsymbol{y}] = T_k,\ k = 0, 1,\ E[R_2|T, \boldsymbol{\beta}, \boldsymbol{y}] = T_2,\ E[\boldsymbol{\beta}|T, R_2, \boldsymbol{y}] = T_3,
\end{aligned} \tag{A.1}$$

where $T_{20} = \frac{1}{a}\sum_i(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_i)(\boldsymbol{v}_i - \bar{\boldsymbol{y}}_i)'$ and $E[T_{20}|\boldsymbol{\phi}, \boldsymbol{y}] = W_2^{*-1} + W_2^{*-1}Q_2[S_2 + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})']Q_2 W_2^{*-1}$. Some algebra implies (8) provided that $b, c > 1$ and $\boldsymbol{\mu}_{R_k} > 0,\ k = 0, 1, 2.$ □

Proof of Theorem 1: To prove the asymptotic normality of the posterior distribution for Θ_a, it suffices, by the Proposition 1, to verify **(i)** the first part of the assumption **(A1)**, the existence of fixed points of $\boldsymbol{g}_a$ and $\boldsymbol{g}$, and $\boldsymbol{\mu}_{\Theta_a} \overset{a\to\infty}{\longrightarrow} \boldsymbol{\mu}_{\Theta}$; **(ii)** the second part of the assumption **(A1)**, $\Sigma_D > 0$; **(iii)** the assumption **(A2)**, the asymptotic normality for each conditional; **(iv)** $\Sigma > 0$; **(v)** the regularity condition(s), $\boldsymbol{g}_a$ is continuously differentiable and $\dot{\boldsymbol{g}}_a(\boldsymbol{t}) \overset{a\to\infty}{\longrightarrow} \dot{\boldsymbol{g}}(\boldsymbol{t})$ (in the neighborhood of $\boldsymbol{\mu}_{\Theta_a}$). Note that we have shown that **(i)** holds in Lemma 1, and **(iv)** $\Sigma > 0$ implies **(ii)** holds.

As to the **(v)**, the regularity condition(s), by Eq. (A.1), it is sufficient to check if each component of $\boldsymbol{g}_a$, or simply $\boldsymbol{g}_{1a} = E[T|\boldsymbol{\phi}, \boldsymbol{y}]$, is continuously differentiable and all its partial derivatives converge since $\dot{\boldsymbol{g}}_{2a},\ldots,\dot{\boldsymbol{g}}_{5a}$ are just some constant vectors. From Eq. (A.1) or, more concisely, Eq. (A.4), the conditional mean functions of $T_k's$ can be expressed as a linear combinations of the products of some positive definite matrices, inverses of positive definite matrices, or vectors where each of which matrix or vector is a linear combinations of $\{R_k's, \boldsymbol{\beta}\}$. It is easy to see that if all the such matrices or vectors are continuously differentiable and their partial derivatives converge. Hence the conditions hold.

Recall that $\Theta_a = (\Theta_{1a}', \ldots, \Theta_{5a}')' = (T', R_{0v}', R_{1v}', R_{2v}', \boldsymbol{\beta}')'$, $R = (R_{0v}', R_{1v}', R_{2v}')'$, and $\boldsymbol{\phi} = (R', \boldsymbol{\beta}')'$. We then break the proof into two parts: **(a)** prove the asymptotic normality for each conditional, which is the **(iii)**; **(b)** simplify Σ and derive the relationship among Σ_T, $\Sigma_{T\boldsymbol{\phi}}$, and $\Sigma_{\boldsymbol{\phi}}$; **(c)** calculate $\Sigma_{\boldsymbol{\phi}}$ explicitly. We defer the proof of **(iv)** $\Sigma > 0$ to Theorem 3.

(a) The conditionally asymptotic normality of Θ_{1a} (or T), $T = (T_{0v}', T_{1v}', T_{2v}', T_3')'$, $T_{rv} = \mathrm{vech}(T_r),\ r = 0, 1, 2$:

Let $\boldsymbol{a} = (\boldsymbol{a}_0', \boldsymbol{a}_1', \boldsymbol{a}_2', \boldsymbol{a}_3')'$, where $\boldsymbol{a}_0, \boldsymbol{a}_1, \boldsymbol{a}_2$ are p_1-dimensional column vectors and $\boldsymbol{a}_3$ is a p-dimensional column vector. $p_1 = \frac{p(p+1)}{2}$. Then, by the Eq. (6),

$$\begin{aligned}
\boldsymbol{a}'T = \frac{1}{a}\sum_{i=1}^{a}\Bigg\{&\frac{1}{bc}\sum_{j,k}\boldsymbol{a}_0'\mathrm{vech}[(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})(\boldsymbol{v}_{ij} - \bar{\boldsymbol{y}}_{ij})']\\
&+\frac{1}{b}\sum_{j,k}\boldsymbol{a}_1'\mathrm{vech}[(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)(\boldsymbol{v}_{ij} - \boldsymbol{v}_i)']
\end{aligned}$$

$$
\left. + \boldsymbol{a}_2' \text{vech}[(\boldsymbol{v}_i - \bar{\boldsymbol{y}})(\boldsymbol{v}_i - \bar{\boldsymbol{y}})'] + \boldsymbol{a}_3' \boldsymbol{v}_i \right\} + \boldsymbol{a}_0' S_{0v}
$$
$$
= \frac{1}{a} \sum_{i=1}^{a} H_i + \frac{1}{ab} \sum_{i,j} \boldsymbol{a}_0' \text{vech}[\bar{\boldsymbol{y}}_{ij} \bar{\boldsymbol{y}}_{ij}'] + \boldsymbol{a}_2' \text{vech}[\bar{\boldsymbol{y}} \bar{\boldsymbol{y}}'] + \boldsymbol{a}_0' S_{0v},
$$

where H_i can be expressed as $(\mathbf{v}_i' B_T \mathbf{v}_i + \mathbf{v}_i' \mathbf{b}_i)$, a combination of a quadratic form and a linear combination of $\mathbf{v}_i = (\boldsymbol{v}_{i1}', \ldots, \boldsymbol{v}_{ib}', \boldsymbol{v}_i')'$, $i = 1, \ldots, a$, and all the coefficients of function of $\mathbf{v}_i$, H_i, are uniformly bounded. Moreover, from the Eq. (4), it is easily to see that $\mathbf{v}_i$'s given $\boldsymbol{\phi}$ and $\boldsymbol{y}$ are conditionally independent, normally distributed, i.e.

$$
\mathbf{v}_i | \boldsymbol{\phi}, \boldsymbol{y} \overset{\text{ind}}{\sim} N_{p(b+1)}(Q_{\mathbf{v}_i}^{-1} E_{\mathbf{v}_i}, Q_{\mathbf{v}_i}^{-1}),
$$

where $Q_{\mathbf{v}_i} = \begin{pmatrix} I_b \otimes (P_0 + Q_1) & -\mathbf{1}_b \otimes Q_1 \\ -\mathbf{1}_b' \otimes Q_1 & P_1 + Q_2 \end{pmatrix}$, $E_{\mathbf{v}_i} = (\bar{\boldsymbol{y}}_{i1}' P_0, \ldots, \bar{\boldsymbol{y}}_{ib}' P_0, \boldsymbol{\beta}' Q_2)'$, and $\mathbf{1}_b$ is a b-dimensional column vector of all ones. Hence, we have $E|H_i - E(H_i)|^{2+\delta}$ = O(1) for some $\delta > 0$. Thus, the asymptotic normality of T given $\boldsymbol{\phi}$ and $\boldsymbol{y}$ as a tends to infinity follows by the Liapounov criterion and the Cramér-Wold device.

The conditionally asymptotic normality of Θ_{ia}, $i = 2, \ldots, 5$ (R_{0v},R_{1v},R_{2v},$\boldsymbol{\beta}$):

Recall that each conditional distribution of R_k follows an inverse Wishart distribution and the conditional distribution of $\boldsymbol{\beta}$ is normally distributed. We then use the following lemma to derive the conditionally asymptotic normality of each R_{kv}.

Lemma A.1 *Suppose* $R \sim \text{IW}_p(m_n, \Psi_n)$, $\frac{m_n}{n} \overset{n\to\infty}{\longrightarrow} m > 0$, *and* $\frac{\Psi_n}{n} \overset{n\to\infty}{\longrightarrow} \Psi \geq 0$. *Then* $\sqrt{n}(R_v - E(R_v)) \overset{\mathcal{L}}{\longrightarrow} N_{p_1}(\mathbf{0}, \frac{2}{m^3} \Psi^{\otimes 2h})$.

Note that, from Jelenkowska (1995) or Press (2005), we have

$$
\begin{aligned}
\lim_{n\to\infty} n\text{Cov}(R_v) &= \lim_{n\to\infty} n\text{Cov}(L_p \text{vec}(R)) \\
&= \lim_{n\to\infty} n L_p \frac{(m_n - p + 1) I_{p^2} + (m_n - p - 1) K_p}{(m_n - p)(m_n - p - 1)^2 (m_n - p - 3)} \Psi_n{}^{\otimes 2} L_p' \\
&= L_p \frac{2}{m^3} N_p \Psi^{\otimes 2} L_p' = \frac{2}{m^3} L_p N_p \Psi^{\otimes 2} N_p L_p' = \frac{2}{m^3} \Psi^{\otimes 2h}.
\end{aligned}
$$

The proof of Lemma A.1 is similar to that shown in Anderson (2003) for the Wishart distribution case and hence is omitted.

Thus, the asymptotic normalities for conditional R_{0v}, R_{1v}, R_{2v}, and $\boldsymbol{\beta}$ are as follows.

$$
\begin{aligned}
&\sqrt{a}(R_{0v} - E(R_{0v}|T, \mathbf{y})|_{\Theta_a = \boldsymbol{\mu}_{\Theta_a}}) \overset{\mathcal{L}}{\longrightarrow} N_{p_1}(\mathbf{0}, \frac{2}{bc} \boldsymbol{\mu}_{T_0}{}^{\otimes 2h}), \\
&\sqrt{a}(R_{1v} - E(R_{1v}|T, \mathbf{y})|_{\Theta_a = \boldsymbol{\mu}_{\Theta_a}}) \overset{\mathcal{L}}{\longrightarrow} N_{p_1}(\mathbf{0}, \frac{2}{b} \boldsymbol{\mu}_{T_1}{}^{\otimes 2h}), \\
&\sqrt{a}(R_{2v} - E(R_{2v}|T, \boldsymbol{\beta}, \mathbf{y})|_{\Theta_a = \boldsymbol{\mu}_{\Theta_a}}) \overset{\mathcal{L}}{\longrightarrow} N_{p_1}(\mathbf{0}, 2\boldsymbol{\mu}_{T_2}{}^{\otimes 2h}), \\
&\sqrt{a}(\boldsymbol{\beta} - E(\boldsymbol{\beta}|T, R_2, \mathbf{y})|_{\Theta_a = \boldsymbol{\mu}_{\Theta_a}}) \overset{\mathcal{L}}{\longrightarrow} N_p(\mathbf{0}, \boldsymbol{\mu}_{R_2}).
\end{aligned}
\tag{A.2}
$$

(b) From the Proposition 1, we have $\varSigma = V^{-1}\varSigma_D$, where, from the Eqs. (8) and (A.2),

$$
\begin{aligned}
\varSigma_D &= \text{blkdiag}\{''_1, \ldots, ''_5\},\\
&= \text{blkdiag}\{\lim_{a\to\infty} a \cdot \text{Cov}(\text{T}|\boldsymbol{\phi}, \boldsymbol{y})\big|_{\boldsymbol{\phi}=\boldsymbol{\mu}_{\boldsymbol{\phi}}}, \frac{2}{\text{bc}}\boldsymbol{\mu}_{\text{R}_0}{}^{\otimes 2\text{h}}, \frac{2}{\text{b}}\boldsymbol{\mu}_{\text{R}_1}{}^{\otimes 2\text{h}}, 2\boldsymbol{\mu}_{\text{R}_2}{}^{\otimes 2\text{h}}, \boldsymbol{\mu}_{\text{R}_2}\},
\end{aligned}
$$

the matrix V is a 5×5 block matrix, $\{V_{ij}\}_{1\le i,j\le 5}$, and

$$
V = \lim_{a\to\infty}\begin{pmatrix}
I_{p_2} & -\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial R'_{0v}} & -\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial R'_{1v}} & -\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial R'_{2v}} & -\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial \boldsymbol{\beta}'}\\
-\frac{\partial E(R_{0v}|T,\boldsymbol{y})}{\partial T'} & I_{p_1} & -\frac{\partial E(R_{0v}|T,\boldsymbol{y})}{\partial R'_{1v}} & -\frac{\partial E(R_{0v}|T,\boldsymbol{y})}{\partial R'_{2v}} & -\frac{\partial E(R_{0v}|T,\boldsymbol{y})}{\partial \boldsymbol{\beta}'}\\
-\frac{\partial E(R_{1v}|T,\boldsymbol{y})}{\partial T'} & -\frac{\partial E(R_{1v}|T,\boldsymbol{y})}{\partial R'_{0v}} & I_{p_1} & -\frac{\partial E(R_{1v}|T,\boldsymbol{y})}{\partial R'_{2v}} & -\frac{\partial E(R_{1v}|T,\boldsymbol{y})}{\partial \boldsymbol{\beta}'}\\
-\frac{\partial E(R_{2v}|T,\boldsymbol{\beta},\boldsymbol{y})}{\partial T'} & -\frac{\partial E(R_{2v}|T,\boldsymbol{\beta},\boldsymbol{y})}{\partial R'_{0v}} & -\frac{\partial E(R_{2v}|T,\boldsymbol{\beta},\boldsymbol{y})}{\partial R'_{1v}} & I_{p_1} & -\frac{\partial E(R_{2v}|T,\boldsymbol{\beta},\boldsymbol{y})}{\partial \boldsymbol{\beta}'}\\
-\frac{\partial E(\boldsymbol{\beta}|T,R_{2v},\boldsymbol{y})}{\partial T'} & -\frac{\partial E(\boldsymbol{\beta}|T,R_{2v},\boldsymbol{y})}{\partial R'_{0v}} & -\frac{\partial E(\boldsymbol{\beta}|T,R_{2v},\boldsymbol{y})}{\partial R'_{1v}} & -\frac{\partial E(\boldsymbol{\beta}|T,R_{2v},\boldsymbol{y})}{\partial R'_{2v}} & I_p
\end{pmatrix}\Bigg|_{\varTheta_a=\boldsymbol{\mu}_{\varTheta_a}}.
$$

By the Eq. (A.1), the block sub-matrices of V, V_{ij}'s, are given as follows:

$$
\begin{aligned}
A \equiv -[V_{12}\ V_{13}\ V_{14}\ V_{15}] &= \lim_{a\to\infty}\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial \boldsymbol{\phi}'}\bigg|_{\boldsymbol{\phi}=\boldsymbol{\mu}_{\boldsymbol{\phi}}},\\
[V'_{21}\ V'_{31}\ V'_{41}\ V'_{51}]' = -I_{p_2},\ V_{23} &= V_{24} = V_{34} = V_{32} = V_{42} = V_{43} = \mathbf{O}_{p_1},\\
V_{25} = V_{35} &= V_{45} = V'_{52} = V'_{53} = V'_{54} = \mathbf{O}_{p_1,p}.
\end{aligned}
$$

Thus, the asymptotic covariance of $\varTheta_a = (T', R'_{0v}, R'_{1v}, R'_{2v}, \boldsymbol{\beta}')'$, $\varSigma$, can be simplified to a 2×2 block matrix (from a 5×5 block matrix), which is the asymptotic covariance of $(T', \boldsymbol{\phi}')'$, and owing to the symmetry of $\varSigma$, we have

$$
\begin{aligned}
\varSigma &= V^{-1}\varSigma_D = \begin{pmatrix} I_{p_2} & -A\\ -I_{p_2} & I_{p_2}\end{pmatrix}^{-1}\begin{pmatrix}\varSigma_1 & \mathbf{O}_{p_2}\\ \mathbf{O}_{p_2} & \varSigma_2^*\end{pmatrix}\\
&= \begin{pmatrix}(I_{p_2}-A)^{-1}\varSigma_1 & (I_{p_2}-A)^{-1}A\varSigma_2^*\\ (I_{p_2}-A)^{-1}\varSigma_1 & (I_{p_2}-A)^{-1}\varSigma_2^*\end{pmatrix} = \begin{pmatrix}\varSigma_T & \varSigma_{T\boldsymbol{\phi}}\\ \varSigma'_{T\boldsymbol{\phi}} & \varSigma_{\boldsymbol{\phi}}\end{pmatrix},\\
\varSigma_T &= \varSigma_{T\boldsymbol{\phi}} = \varSigma'_{T\boldsymbol{\phi}} = (I_{p_2}-A)^{-1}\varSigma_1 = \varSigma_{\boldsymbol{\phi}} - \varSigma_2^*,\\
\varSigma_{\boldsymbol{\phi}} &= (I_{p_2}-A)^{-1}\varSigma_2^*,\ \varSigma_2^* = \text{blkdiag}\{\varSigma_2, \ldots, \varSigma_5\}.
\end{aligned}
\tag{A.3}
$$

(c) To simplify the calculation of A and $\varSigma_{\boldsymbol{\phi}}$, define $\varLambda_0 = R_0$, $\varLambda_1 = cR_1 + R_0$, $\varLambda_2 = bcR_2 + \varLambda_1$, and $\varLambda = (\varLambda'_{0v}, \varLambda'_{1v}, \varLambda'_{2v}, \boldsymbol{\beta}')'$. Then we have

$$
\begin{aligned}
&R_0 = \varLambda_0,\ R_1 = \frac{\varLambda_1-\varLambda_0}{c},\ R_2 = \frac{\varLambda_2-\varLambda_1}{bc},\ Q_0 = \varLambda_0^{-1},\ Q_1 = [\frac{\varLambda_1-\varLambda_0}{c}]^{-1},\ Q_2 = [\frac{\varLambda_2-\varLambda_1}{bc}]^{-1},\\
&P_0 = c\varLambda_0^{-1},\ P_1 = bcQ_1W_1^{*-1}Q_0 = bc\varLambda_1^{-1},\\
W_1^{*-1} &= (cR_0^{-1}+R_1^{-1})^{-1} = R_0[cR_1+R_0]^{-1}R_1 (= R_1[cR_1+R_0]^{-1}R_0) = \varLambda_0\varLambda_1^{-1}[\frac{\varLambda_1-\varLambda_0}{c}],\\
W_2^{*-1} &= (P_1+R_2^{-1})^{-1} = P_1^{-1}[P_1^{-1}+R_2]^{-1}R_2 (= R_2[P_1^{-1}+R_2]^{-1}P_1^{-1}) = \varLambda_1\varLambda_2^{-1}\frac{\varLambda_2-\varLambda_1}{bc}.
\end{aligned}
$$

The conditional mean functions of T_i's given Λ and $\boldsymbol{y}$ are

$$
\begin{aligned}
E[T_0|\Lambda, \boldsymbol{y}] &= \Lambda_0\Lambda_1^{-1}\left[\frac{\Lambda_1-\Lambda_0}{c}\right] + \Lambda_0\Lambda_1^{-1}\left[S_1 + \Lambda_1\Lambda_2^{-1}\frac{\Lambda_2-\Lambda_1}{bc}\right]\Lambda_1^{-1}\Lambda_0 \\
&\quad + \Lambda_0\Lambda_2^{-1}[S_2 + (\boldsymbol{\beta}-\bar{\boldsymbol{y}})(\boldsymbol{\beta}-\bar{\boldsymbol{y}})']\Lambda_2^{-1}\Lambda_0 + S_0, \\
E[T_1|\Lambda, \boldsymbol{y}] &= \Lambda_0\Lambda_1^{-1}\left[\frac{\Lambda_1-\Lambda_0}{c}\right] + [I_{p_1} - \Lambda_0\Lambda_1^{-1}][S_1 + \Lambda_1\Lambda_2^{-1}\frac{\Lambda_2-\Lambda_1}{bc}][I_{p_1} - \Lambda_1^{-1}\Lambda_0] \\
&\quad + [I_{p_1} - \Lambda_0\Lambda_1^{-1}]\Lambda_1\Lambda_2^{-1}[S_2 + (\boldsymbol{\beta}-\bar{\boldsymbol{y}})(\boldsymbol{\beta}-\bar{\boldsymbol{y}})']\Lambda_2^{-1}\Lambda_1[I_{p_1} - \Lambda_1^{-1}\Lambda_0], \\
E[T_2|\Lambda, \boldsymbol{y}] &= \Lambda_1\Lambda_2^{-1}\frac{\Lambda_2-\Lambda_1}{bc} + [I_{p_1} - \Lambda_1\Lambda_2^{-1}]S_2[I_{p_1} - \Lambda_2^{-1}\Lambda_1] \\
&\quad + \Lambda_0\Lambda_2^{-1}(\boldsymbol{\beta}-\bar{\boldsymbol{y}})(\boldsymbol{\beta}-\bar{\boldsymbol{y}})'\Lambda_2^{-1}\Lambda_0, \\
E[T_3|\Lambda, \boldsymbol{y}] &= [I_{p_1} - \Lambda_1\Lambda_2^{-1}]\bar{\boldsymbol{y}} + \Lambda_1\Lambda_2^{-1}\boldsymbol{\beta},
\end{aligned}
\tag{A.4}
$$

and

$$
A = \lim_{a\to\infty}\frac{\partial E(T|\Lambda, \boldsymbol{y})}{\partial \Lambda'}\frac{\partial \Lambda}{\partial \boldsymbol{\phi}'}\Big|_{\Lambda=\boldsymbol{\mu}_\Lambda} = A^* \cdot \frac{\partial \Lambda}{\partial \boldsymbol{\phi}'},
$$

where A^* is a 4×4 block matrix, $A^* = \lim_{a\to\infty}\frac{\partial E(T|\Lambda,\boldsymbol{y})}{\partial \Lambda'}\big|_{\Lambda=\boldsymbol{\mu}_\Lambda}$,

$$
\boldsymbol{\mu}_\Lambda = (\boldsymbol{\mu}'_{\Lambda_{0v}}, \boldsymbol{\mu}'_{\Lambda_{1v}}, \boldsymbol{\mu}'_{\Lambda_{2v}}, \boldsymbol{\mu}'_{\boldsymbol{\beta}})',\ \boldsymbol{\mu}_{\Lambda_0} = \frac{c}{c_1}S_0,\ \boldsymbol{\mu}_{\Lambda_1} = \frac{bc}{b_1}S_1,\ \boldsymbol{\mu}_{\Lambda_2} = bcS_2,
\tag{A.5}
$$

$\frac{\partial \Lambda}{\partial \boldsymbol{\phi}'} = \text{blkdiag}\{L_1^*, I_p\}$, L_1^* is a lower block triangular matrix, and

$$
L_1^* = \begin{pmatrix} I_{p_1} & \mathbf{O}_{p_1} & \mathbf{O}_{p_1} \\ I_{p_1} & cI_{p_1} & \mathbf{O}_{p_1} \\ I_{p_1} & cI_{p_1} & bcI_{p_1} \end{pmatrix}.
$$

To calculate the matrices A^* and hence Σ, we introduce the first-order matrix differential operator and the following lemma.

Definition A.1 (Magnus and Neudecker 1980, 1999)

(a) (matrix differentials) Let $F(X)$ be a differentiable $m \times p$ matrix function of an $n \times q$ matrix of real variables X. Define the first differential $\mathrm{d}F(X)$ to be the matrix of differentials $\{\mathrm{d}F(X)\}_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, p$.

(b) The duplication matrix D_p is the $p^2 \times \frac{p(p+1)}{2}$ matrix such that $D_p\text{vech}(C) = \text{vec}(C)$ and $C^{\otimes 2l} \equiv L_p C^{\otimes 2} D_p$ for any $p \times p$ matrix C.

Lemma A.2 *(Magnus and Neudecker 1980, 1999)*

(a) $d(XY) = (dX)Y + X(dY)$; $dX^{-1} = -X^{-1}(dX)X^{-1}$ *for any nonsingular matrix* X;

(b) $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ *for any* A, B, C *matrices of appropriate sizes;*

(c) $N_p' = N_p$; $D_p L_p N_p = N_p$; $N_p D_p = D_p$; $N_p C^{\otimes 2} = C^{\otimes 2} N_p = N_p C^{\otimes 2} N_p$ *for any square matrix* C;
(d) $(L_p N_p C^{\otimes 2} N_p L_p')^{-1} = D_p' (C^{-1})^{\otimes 2} D_p$ *for any nonsingular matrix* C.

In the following, we calculate the (1, 1)th and (1, 2)th blocks of A^*. Calculations for other (i, j)th blocks of A^* just follow the same techniques.

By the Eq. (A.4) and Lemma A.2 (a), the first differential of $E[T_0|\Lambda, \boldsymbol{y}]$ with respect to Λ_0 evaluated at $\boldsymbol{\mu}_\Lambda$ is

$$\begin{aligned}
\mathrm{d}E[T_0|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda} =& \mathrm{d}\Bigg[\frac{1}{c}\Lambda_0 + \Lambda_0\left[-\frac{1}{c}\Lambda_1^{-1} + \Lambda_1^{-1}S_1\Lambda_1^{-1} + \frac{1}{bc}\Lambda_1^{-1}\right]\Lambda_0 \\
&+ \Lambda_0[-\frac{1}{bc}\Lambda_2^{-1} + \Lambda_2^{-1}[S_2 + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})']\Lambda_2^{-1}]\Lambda_0 + S_0\Bigg]\Bigg|_{\Lambda=\boldsymbol{\mu}_\Lambda} \\
=& \frac{1}{c}(\mathrm{d}\Lambda_0) + \mathrm{d}\left[\Lambda_0\left[-\frac{1}{c}\Lambda_1^{-1} + \Lambda_1^{-1}S_1\Lambda_1^{-1} + \frac{1}{bc}\Lambda_1^{-1}\right]\Bigg|_{\Lambda=\boldsymbol{\mu}_\Lambda}\Lambda_0\right] \\
&+ \mathrm{d}\left[\Lambda_0\left[-\frac{1}{bc}\Lambda_2^{-1} + \Lambda_2^{-1}[S_2 + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})'\right]\Lambda_2^{-1}\right]\Bigg|_{\Lambda=\boldsymbol{\mu}_\Lambda}\Lambda_0\Bigg] \\
=& \frac{1}{c}(\mathrm{d}\Lambda_0).
\end{aligned}$$

Thus,

$$\mathrm{d}E[T_{0v}|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda} = \mathrm{vech}\left[\mathrm{d}E[T_0|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda}\right] = \mathrm{vech}\left[\frac{1}{c}(\mathrm{d}\Lambda_0)\right] = \frac{1}{c}\mathrm{d}\Lambda_{0v},$$

and the (1, 1)th block of A^*, $\{A^*\}_{11} = \lim_{a\to\infty} \frac{\partial E(T|\Lambda, y)}{\partial \Lambda_{0v}'}\Big|_{\Lambda=\boldsymbol{\mu}_\Lambda} = \frac{1}{c} I_{p_1}$.

Next, the first differential of $E[T_0|\boldsymbol{\phi}, \boldsymbol{y}]$ with respect to Λ_1 evaluated at $\boldsymbol{\mu}_\Lambda$ is

$$\begin{aligned}
\mathrm{d}E[T_0|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda} =& \mathrm{d}\Bigg[\frac{1}{c}[\Lambda_0 - \Lambda_0\Lambda_1^{-1}\Lambda_0] + \Lambda_0\Lambda_1^{-1}S_1\Lambda_1^{-1}\Lambda_0 + \frac{1}{bc}\Lambda_0[\Lambda_1^{-1} - \Lambda_2^{-1}]\Lambda_0 \\
&+ \Lambda_0\Lambda_2^{-1}[S_2 + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})']\Lambda_2^{-1}\Lambda_0 + S_0\Bigg]\Bigg|_{\Lambda=\boldsymbol{\mu}_\Lambda} \\
=& \Bigg[\frac{1}{c}\Lambda_0\Lambda_1^{-1}(\mathrm{d}\Lambda_1)\Lambda_1^{-1}\Lambda_0 - [\Lambda_0\Lambda_1^{-1}(\mathrm{d}\Lambda_1)\Lambda_1^{-1}S_1\Lambda_1^{-1}\Lambda_0 \\
&+ \Lambda_0\Lambda_1^{-1}S_1\Lambda_1^{-1}(\mathrm{d}\Lambda_1)\Lambda_1^{-1}\Lambda_0] - \frac{1}{bc}\Lambda_0\Lambda_1^{-1}(\mathrm{d}\Lambda_1)\Lambda_1^{-1}\Lambda_0\Bigg]\Bigg|_{\Lambda=\boldsymbol{\mu}_\Lambda} \\
=& \left[\frac{1}{c}A_{01}(\mathrm{d}\Lambda_1)A_{01}' - \frac{2b_1}{bc}A_{01}(\mathrm{d}\Lambda_1)A_{01}' - \frac{1}{bc}A_{01}(\mathrm{d}\Lambda_1)A_{01}'\right] \\
=& \left[-\frac{b_1}{bc}A_{01}(\mathrm{d}\Lambda_1)A_{01}'\right],
\end{aligned}$$

where $A_{01} = \boldsymbol{\mu}_{\Lambda_0}\boldsymbol{\mu}_{\Lambda_1}^{-1} = \frac{b_1}{bc}S_0S_1^{-1}$. By Lemma A.2 (b),

$$\begin{aligned}
\mathrm{d}E[T_{0v}|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda} =& \mathrm{vech}\left[\mathrm{d}E[T_0|\Lambda, \boldsymbol{y}]|_{\Lambda=\boldsymbol{\mu}_\Lambda}\right] = -\frac{b_1}{bc}L_p\mathrm{vec}\left[A_{01}(\mathrm{d}\Lambda_1)A_{01}'\right] \\
=& -\frac{b_1}{bc}L_pA_{01}{}^{\otimes 2}\mathrm{vec}(\mathrm{d}\Lambda_1) = -\frac{b_1}{bc}L_pA_{01}{}^{\otimes 2}D_p\mathrm{d}\Lambda_{1v} = -\frac{b_1}{bc}A_{01}{}^{\otimes 2l}\mathrm{d}\Lambda_{1v}.
\end{aligned}$$

The (1,2) block of A^*, $\{A^*\}_{12} = \lim_{a\to\infty} \left.\frac{\partial E(T|\Lambda,\mathbf{y})}{\partial \Lambda'_{1v}}\right|_{\Lambda=\boldsymbol{\mu}_\Lambda} = -\frac{b_1}{bc} A_{01}{}^{\otimes 2l}$. Hence, we obtain $A^* = \mathrm{blkdiag}\{\mathrm{A}_1^*, \mathrm{I_p} - \boldsymbol{\mu}_{\mathrm{R}_2}\mathrm{S}_2^{-1}\}$ and $A_1^* =$

$$\begin{pmatrix} \frac{1}{c} I_{p_1} & -\frac{b_1}{bc} A_{01}{}^{\otimes 2l} & -\frac{1}{bc} A_{02}{}^{\otimes 2l} \\ -\frac{1}{c} I_{p_1} & \frac{1}{c}(I_{p_1} - \frac{b_1}{b}(I_{p_1} - A_{01})^{\otimes 2l}) & -\frac{1}{bc}(A_{02} - A_{12})^{\otimes 2l} \\ \mathbf{O}_{p_1} & -\frac{1}{bc} I_{p_1} & \frac{1}{bc}(I_{p_1} - (I_{p_1} - A_{12})^{\otimes 2l}) \end{pmatrix},$$

where $A_{12} = \boldsymbol{\mu}_{\Lambda_1}\boldsymbol{\mu}_{\Lambda_2}^{-1} = \frac{1}{b_1} S_1 S_2^{-1}$, and $A_{02} = \boldsymbol{\mu}_{\Lambda_0}\boldsymbol{\mu}_{\Lambda_2}^{-1} = \frac{1}{bc_1} S_0 S_2^{-1}$. Note that the calculation of A^* is based on the following formulas:

$$\boldsymbol{\mu}_{\Lambda_1}^{-1} = \frac{1}{c}\boldsymbol{\mu}_{R_1}^{-1}(I_{p_1} - A_{01}),\ \boldsymbol{\mu}_{\Lambda_2}^{-1} = \frac{1}{c}\boldsymbol{\mu}_{R_1}^{-1}(A_{12} - A_{02}),\ \boldsymbol{\mu}_{\Lambda_2}^{-1} = \frac{1}{bc}\boldsymbol{\mu}_{R_2}^{-1}(I_{p_1} - A_{12}),$$

and for $i < j$, by Lemma A.2 (c) and (d),

$$\begin{aligned} \boldsymbol{\mu}_{\Lambda_i}^{\otimes 2h}\boldsymbol{\mu}_{\Lambda_j}^{\otimes 2h^{-1}} &= L_p N_p \boldsymbol{\mu}_{\Lambda_i}^{\otimes 2} N_p L'_p (L_p N_p \boldsymbol{\mu}_{\Lambda_j}^{\otimes 2} N_p L'_p)^{-1} = L_p N_p \boldsymbol{\mu}_{\Lambda_i}^{\otimes 2} N_p L'_p D'_p (\boldsymbol{\mu}_{\Lambda_j}^{-1})^{\otimes 2} D_p \\ &= L_p N_p \boldsymbol{\mu}_{\Lambda_i}^{\otimes 2} N_p (\boldsymbol{\mu}_{\Lambda_j}^{-1})^{\otimes 2} D_p = L_p (\boldsymbol{\mu}_{\Lambda_i}\boldsymbol{\mu}_{\Lambda_j}^{-1})^{\otimes 2} D_p = A_{ij}{}^{\otimes 2l}. \end{aligned}$$

Thus, after some calculations,

$$\begin{aligned} \Sigma_{\boldsymbol{\phi}}^{-1} &= \Sigma_2^{*-1}(I_{p_2} - A) = \mathrm{blkdiag}\{\mathrm{Q_A}, \mathrm{S}_2^{-1}\},\ \mathrm{A} = \mathrm{blkdiag}\{\mathrm{A}_1, \mathrm{I_p} - \boldsymbol{\mu}_{\mathrm{R}_2}\mathrm{S}_2^{-1}\},\ \mathrm{A}_1 = \mathrm{A}_1^*\mathrm{L}^*, \\ Q_A &= (\mathrm{blkdiag}\{''_2, ''_3, ''_4\})^{-1}(\mathrm{I}_{3\mathrm{p}_1} - \mathrm{A}_1) \\ &= \frac{1}{2}\begin{pmatrix} bc_1\boldsymbol{\mu}_{\Lambda_0}^{\otimes 2h^{-1}} + b_1\boldsymbol{\mu}_{\Lambda_1}^{\otimes 2h^{-1}} + \boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & b_1 c\boldsymbol{\mu}_{\Lambda_1}^{\otimes 2h^{-1}} + c\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & bc\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} \\ b_1 c\boldsymbol{\mu}_{\Lambda_1}^{\otimes 2h^{-1}} + c\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & b_1 c^2\boldsymbol{\mu}_{\Lambda_1}^{\otimes 2h^{-1}} + c^2\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & bc^2\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} \\ bc\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & bc^2\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} & b^2c^2\boldsymbol{\mu}_{\Lambda_2}^{\otimes 2h^{-1}} \end{pmatrix}. \end{aligned} \tag{A.6}$$

Or simply

$$Q_A = \frac{1}{2} L_1^{*\prime}\mathrm{blkdiag}\{\mathrm{bc}_1\boldsymbol{\mu}_0^{\otimes 2\mathrm{h}^{-1}}, \mathrm{b}_1\boldsymbol{\mu}_1^{\otimes 2\mathrm{h}^{-1}}, \boldsymbol{\mu}_2^{\otimes 2\mathrm{h}^{-1}}\}\mathrm{L}_1^*.$$

Some algebra results in (9).

Proof of Theorem 2: The proofs of the existence and the convergence of the explicit fixed point (the first part of **(A1)** in Proposition 1), the asymptotic normality of the posterior distribution of Θ_a, the assumption **(A2)**, and the positive definiteness of Σ_s are similar to those in Lemma 1 and Theorem 1. We omit them to save space. Next, similar to the Eq. (A.3) in the centering case, Σ_s after some algebra can be expressed as

$$\begin{aligned} \Sigma_s &= \begin{pmatrix} I_{p_2} & -A_s \\ -B_s & I_{p_2} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{1s} & \mathbf{O}_{p_2} \\ \mathbf{O}_{p_2} & \Sigma_{2s}^* \end{pmatrix} \\ &= \begin{pmatrix} (I_{p_2} - A_s B_s)^{-1}\Sigma_{1s} & A_s(I_{p_2} - B_s A_s)^{-1}\Sigma_{2s}^* \\ B_s(I_{p_2} - A_s B_s)^{-1}\Sigma_{1s} & (I_{p_2} - B_s A_s)^{-1}\Sigma_{2s}^* \end{pmatrix} = \begin{pmatrix} \Sigma_{T_s} & \Sigma_{T_s\boldsymbol{\phi}} \\ \Sigma'_{T_s\boldsymbol{\phi}} & \Sigma_{\boldsymbol{\phi}} \end{pmatrix}, \end{aligned} \tag{A.7}$$

where $\Sigma^*_{2s} = \text{blkdiag}\{''2, ''3, ''4, ''5s\}$, $''5s = \frac{1}{bc}\boldsymbol{\mu}_{R_0}$, and

$$B_s = I_{p_2} + \begin{pmatrix} \mathbf{O}_{p_1,3p_1} & -L_p(\bar{\mathbf{y}} \otimes I_p + I_p \otimes \bar{\mathbf{y}}) \\ \mathbf{O}_{2p_1+p,3p_1} & \mathbf{O}_{2p_1+p,p} \end{pmatrix}.$$

Hence $(I_{p_2} - B_s A_s)^{-1}\Sigma^*_{2s} = \Sigma_{\boldsymbol{\phi}}$ implies $A_s = B_s^{-1}(I_{p_2} - \Sigma^*_{2s}\Sigma^{-1}_{\boldsymbol{\phi}})$. Thus, $\Sigma_{T_s\boldsymbol{\phi}} = A_s(I_{p_2} - B_s A_s)^{-1}\Sigma^*_{2s} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma^*_{2s})$ and $\Sigma_{T_s} = B_s^{-1}(\Sigma_{\boldsymbol{\phi}} - \Sigma^*_{2s})B_s^{-1'}$ as shown in Theorem 2.

Proof of Corollary 1 From the proofs of Theorems 1 and 2 (see the Eqs. (A.6) and (A.7)), the derivatives of mean functions for the standard and centering cases, $\dot{\boldsymbol{g}}_s$ and $\dot{\boldsymbol{g}}_c$, are

$$\dot{\boldsymbol{g}}_s = \begin{pmatrix} \mathbf{O}_{p_2} & A_s \\ B_s & \mathbf{O}_{p_2} \end{pmatrix}, \ \dot{\boldsymbol{g}}_c = \begin{pmatrix} \mathbf{O}_{p_2} & A \\ I_{p_2} & \mathbf{O}_{p_2} \end{pmatrix}.$$

Note that these two derivatives can be partitioned as $\begin{pmatrix} \mathbf{O}_{p_2} & E \\ F & \mathbf{O}_{p_2} \end{pmatrix}$. By Lemma A.3 in Su and Johnson (2006), $\rho_s = \rho(\dot{\boldsymbol{g}}_s)$ and $\rho_c = \rho(\dot{\boldsymbol{g}}_c)$ are less than one. Second, the characteristic polynomials of $\dot{\boldsymbol{g}}_c$ and $\dot{\boldsymbol{g}}_s$ are $p_{\dot{\boldsymbol{g}}_c}(\lambda)$=det$(\lambda^2 I_{p_2} - A)$ and $p_{\dot{\boldsymbol{g}}_s}(\lambda)$=det$(\lambda^2 I_{p_2} - B_s A_s)$, respectively. Here

$$A = \begin{pmatrix} A_1 & \mathbf{O}_{p_2} \\ \mathbf{O}_{p_2} & I_{p_2} - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1} \end{pmatrix}, \ B_s A_s = \begin{pmatrix} A_1 & \mathbf{O}_{p_2} \\ \mathbf{O}_{p_2} & I_{p_2} - \boldsymbol{\mu}_{R_2}S_2^{-1} \end{pmatrix},$$

and $A_1 = A_1^* L^* =$

$$-\begin{pmatrix} -\frac{1}{c}I + \frac{1}{bc}[b_1 A_{01}{}^{\otimes 2l} + A_{02}{}^{\otimes 2l}] & \frac{1}{b}[b_1 A_{01}{}^{\otimes 2l} + A_{02}{}^{\otimes 2l}] & A_{02}{}^{\otimes 2l} \\ \frac{1}{bc}[b_1(I - A_{01})^{\otimes 2l} + (A_{12} - A_{02})^{\otimes 2l}] & \frac{1}{b}[b_1(I - A_{01})^{\otimes 2l} + (A_{12} - A_{02})^{\otimes 2l}] - I & (I - A_{12})^{\otimes 2l} \\ \frac{1}{bc}(I - A_{12})^{\otimes 2l} & \frac{1}{b}(I - A_{12})^{\otimes 2l} & (I - A_{12})^{\otimes 2l} - I \end{pmatrix}.$$

Thus,

$$\rho_s = \max\left\{\sqrt{\rho(A_1)}, \sqrt{\rho(I_p - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1})}\right\},$$
$$\rho_c = \max\left\{\sqrt{\rho(A_1)}, \sqrt{\rho(I_p - \boldsymbol{\mu}_{R_2}S_2^{-1})}\right\}.$$

Next, any eigenvalue of $I_p - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}$, $\lambda(I_p - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) = \lambda(S_2^{-1/2}(\boldsymbol{\mu}_{R_2} + \frac{1}{b}\boldsymbol{\mu}_{R_1})S_2^{-1/2}) > 0$ and $0 < \rho(I_p - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) = 1 - \lambda_{\min}(\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) < 1$ provided $0 < \rho_s < 1$. Similarly, $\lambda(I_p - \boldsymbol{\mu}_{R_2}S_2^{-1})) = \lambda(S_2^{-1/2}(\frac{1}{b}\boldsymbol{\mu}_{R_1} + \frac{1}{bc}\boldsymbol{\mu}_{R_2})S_2^{-1/2}) > 0$ and $0 < \rho(I_p - \boldsymbol{\mu}_{R_2}S_2^{-1}) = 1 - \lambda_{\min}(\boldsymbol{\mu}_{R_0}S_2^{-1}) < 1$ provided $0 < \rho_c < 1$. Note that

$$S_2^{-1/2}\boldsymbol{\mu}_{R_2}S_2^{-1/2} = S_2^{-1/2}\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1/2} + S_2^{-1/2}(\boldsymbol{\mu}_{R_2} - \frac{1}{bc}\boldsymbol{\mu}_{R_0})S_2^{-1/2}.$$

By Weyl's Theorem for symmetric matrices (Horn and Johnson 2012), $\boldsymbol{\mu}_{R_2} - \frac{1}{bc}\boldsymbol{\mu}_{R_0} \geq (\leq) 0$ implies $\lambda_{\min}(S_2^{-1/2}\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1/2}) \leq (\geq) \lambda_{\min}(S_2^{-1/2}\boldsymbol{\mu}_{R_2}S_2^{-1/2})$. Thus, $\rho_c \leq (\geq)\rho_s$ provided $\lambda_{\min}(S_2^{-1/2}\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1/2}) \leq (\geq) \lambda_{\min}(S_2^{-1/2}\boldsymbol{\mu}_{R_2}S_2^{-1/2})$ or simply $\lambda_{\min}(\frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) \leq (\geq) \lambda_{\min}(\boldsymbol{\mu}_{R_2}S_2^{-1})$. Furthermore, $\rho(A_1) \to 0$ as both b and c tend to infinity, $\rho(I_p - \boldsymbol{\mu}_{R_2}S_2^{-1}) \to 0$ as b tends to infinity, and $\rho(I_p - \frac{1}{bc}\boldsymbol{\mu}_{R_0}S_2^{-1}) \to 1$ as b or $c \to \infty$. Thus, $\rho_c \to 0$ as both b and $c \to \infty$, and $\rho_s \to 1$ as b or $c \to \infty$. □

Proof of Theorem 3: Similar to the proof of Theorem 1, we first show the unique fixed point for $\boldsymbol{g}_n$ and then break the rest proof into three parts: **(a)** derive the asymptotic normality of conditional distribution of $R_{kv},\ k = 0, \ldots, r$, and $\boldsymbol{\beta}$; **(b)** compute Σ by simplifying Σ and derive the relationship between Σ_T, $\Sigma_{T\boldsymbol{\phi}}$, and $\Sigma_{\boldsymbol{\phi}}$; **(c)** calculate Σ explicitly; and **(d)** show that $\Sigma > 0$. We omit the proof of the asymptotic normality of the conditional T and the verification of the regularity condition(s) since it is similar to arguments used in the proof of Theorem 1.

First, for simplicity, we define some notations used in the following. Let

$$
\begin{aligned}
&Q_k = R_k^{-1},\ k = 0, \ldots, r,\\
&W_1 = W_1^* = n_r Q_0 + Q_1,\ P_0 = n_r Q_0,\\
&W_k = n_{r-k+1}Q_{k-1} + Q_k,\ W_k^* = W_k - n_{r-k+1}Q_{k-1}W_{k-1}^{*-1}Q_{k-1},\ P_{k-1} = W_k^* - Q_k,\\
&T_{k0} = \frac{1}{n_0 \ldots n_{r-k}} \sum_{i_0 \ldots i_{r-k}} (\boldsymbol{v}_{i_0\ldots i_{r-k}} - \bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}})(\boldsymbol{v}_{i_0\ldots i_{r-k}} - \bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}})',\ k = 2, \ldots, r.
\end{aligned}
$$

From the Eq. (11), we have

$$
\begin{aligned}
\boldsymbol{v}_{i_0\ldots i_{r-k}} | \boldsymbol{v}_{i_0\ldots i_{r-k-1}}, \boldsymbol{\phi}, \boldsymbol{y} &\sim N(W_k^{*-1}[P_{k-1}\bar{\boldsymbol{y}}_{i_0\ldots i_{r-k}} + Q_k \boldsymbol{v}_{i_0\ldots i_{r-k-1}}], W_k^{*-1}),\ k = 1, \ldots, r-1,\\
\boldsymbol{v}_{i_0} | \boldsymbol{\phi}, \boldsymbol{y} &\sim N(W_r^{*-1}[P_{r-1}\bar{\boldsymbol{y}}_{i_0} + Q_r \boldsymbol{\beta}], W_k^{*-1}).
\end{aligned} \tag{A.8}
$$

Then, the components of $\boldsymbol{g}_n$ can be derived through the following conditional mean functions.

$$
\begin{aligned}
E[T_0|\boldsymbol{\phi}, \boldsymbol{y}] &= W^{-1} + W^{-1}Q_1[S_1 + E[T_{20}|\boldsymbol{\phi}, \boldsymbol{y}]]Q_1 W^{-1} + S_0,\\
E[T_k|\boldsymbol{\phi}, \boldsymbol{y}] &= W_k^{*-1} + W_k^{*-1}P_{k-1}[S_k + E[T_{k+1,0}|\boldsymbol{\phi}, \boldsymbol{y}]]P_{k-1}W_k^{*-1},\ k = 1, \ldots, r-1,\\
E[T_r|\boldsymbol{\phi}, \boldsymbol{y}] &= W_r^{*-1} + W_r^{*-1}[P_{r-1}S_r P_{r-1} + Q_r(\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})'Q_r]W_r^{*-1},\\
E[T_{r+1}|\boldsymbol{\phi}, y] &= W_r^{*-1}[P_{r-1}\bar{\boldsymbol{y}} + Q_r\boldsymbol{\beta}],\\
E[R_k|T, \boldsymbol{y}] &= T_k, \quad k = 0, \ldots r-1,\\
E[R_r|T, \boldsymbol{\beta}, \boldsymbol{y}] &= T_r + (T_{r+1} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}} - \boldsymbol{\beta})' + (\bar{\boldsymbol{y}} - \boldsymbol{\beta})(T_{r+1} - \bar{\boldsymbol{y}})' + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})',\\
E[\boldsymbol{\beta}|T, R_r, \boldsymbol{y}] &= T_{r+1},
\end{aligned} \tag{A.9}
$$

where

$$
\begin{aligned}
E[T_{k0}|\boldsymbol{\phi}, \boldsymbol{y}] &= W_k^{*-1} + W_k^{*-1}Q_k[S_k + E[T_{k+1,0}|\boldsymbol{\phi}, \boldsymbol{y}]]Q_k W_k^{*-1},\ k = 2, \ldots, r-1,\\
E[T_{r0}|\boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{y}] &= W_r^{*-1} + W_r^{*-1}Q_r[S_r + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})']Q_r W_r^{*-1}.
\end{aligned}
$$

Next, we solve the unique fixed point of $\boldsymbol{g}_n$ iteratively. Let

$U_k(\boldsymbol{\phi}) = E[T_{k0}|\boldsymbol{\phi}, \boldsymbol{y}]$, a function of $\boldsymbol{\phi}$. First, solve

$$E[T_0|\boldsymbol{\phi}, \boldsymbol{y}] = W^{-1} + W^{-1}Q_1[S_1 + U_2(\boldsymbol{\phi})]Q_1W^{-1} + S_0,\ E[R_0|T, \boldsymbol{y}] = T_0,$$
$$E[T_1|\boldsymbol{\phi}, \boldsymbol{y}] = W^{-1} + W^{-1}P_0[S_1 + U_2(\boldsymbol{\phi})]P_0W^{-1},\ E[R_1|T, \boldsymbol{y}] = T_1.$$

The solution is

$$\boldsymbol{\mu}_{R_0} = \boldsymbol{\mu}_{T_0} = \frac{n_r}{n_r - 1}S_0,\ \boldsymbol{\mu}_{R_1} = \boldsymbol{\mu}_{T_1} = U_2(\boldsymbol{\phi}) + S_1 - \frac{1}{n_r - 1}S_0.$$

Then, by iteratively solving the following equations and plugging the solution into the next step,

$$\begin{aligned}
E[T_{k0}|\boldsymbol{\phi}, \boldsymbol{y}] &= W_k^{*-1} + W_k^{*-1}Q_k[S_k + U_{k+1}(\boldsymbol{\phi})]Q_kW_k^{*-1},\\
\boldsymbol{\mu}_{R_{k-1}} &= U_k(\boldsymbol{\phi}) + S_{k-1} - \frac{1}{n_{r-k+2} - 1}S_{k-2},\\
E[T_k|\boldsymbol{\phi}, \boldsymbol{y}] &= W_k^{*-1} + W_k^{*-1}P_{k-1}[S_k + U_{k+1}(\boldsymbol{\phi})]P_{k-1}W_k^{*-1},\\
E[R_k|T, \boldsymbol{y}] &= T_k,\ k = 2, \ldots, r-1,
\end{aligned}$$

we obtain

$$\begin{aligned}
\boldsymbol{\mu}_{R_{k-1}} = \boldsymbol{\mu}_{T_{k-1}} &= \frac{n_{r-k+1}}{n_{r-k+1} - 1}S_{k-1} - \frac{1}{n_{r-k+2} - 1}S_{k-2},\ \boldsymbol{\mu}_{U_k(\boldsymbol{\phi})} = \frac{1}{n_{r-k}}S_k,\\
\boldsymbol{\mu}_{R_k} = \boldsymbol{\mu}_{T_k} &= U_{k+1}(\boldsymbol{\phi}) + S_k - \frac{1}{n_{r-k+1} - 1}S_{k-1},\ k = 2, \ldots, r-1.
\end{aligned}$$

The last step is to solve

$$\begin{aligned}
E[T_{r0}|\boldsymbol{\phi}, \boldsymbol{y}] &= W_r^{*-1} + W_r^{*-1}Q_r[S_r + (\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})']Q_rW_r^{*-1},\\
\boldsymbol{\mu}_{R_{r-1}} &= U_r(\boldsymbol{\phi}) + S_{r-1} - \frac{1}{n_2 - 1}S_{r-2},\\
E[T_r|\boldsymbol{\phi}, \boldsymbol{y}] &= W_r^{*-1} + W_r^{*-1}[P_{r-1}S_rP_{r-1} + Q_r(\boldsymbol{\beta} - \bar{\boldsymbol{y}})(\boldsymbol{\beta} - \bar{\boldsymbol{y}})'Q_r]W_r^{*-1},\\
E[R_r|T, \boldsymbol{\beta}, \boldsymbol{y}] &= T_r,\ E[T_{r+1}|\boldsymbol{\phi}, \boldsymbol{y}] = W_r^{*-1}[P_{r-1}\bar{\boldsymbol{y}} + Q_r\boldsymbol{\beta}],\ E[\boldsymbol{\beta}|T, R_r, \boldsymbol{y}] = T_{r+1}.
\end{aligned}$$

This implies

$$\begin{aligned}
\boldsymbol{\mu}_{R_{r-1}} = \boldsymbol{\mu}_{T_{r-1}} &= \frac{n_1}{n_1 - 1}S_{r-1} - \frac{1}{n_2 - 1}S_{r-2},\ \boldsymbol{\mu}_{U_r} = \frac{1}{n_1 - 1}S_{r-1},\\
\boldsymbol{\mu}_{R_r} = \boldsymbol{\mu}_{T_r} &= S_r - \frac{1}{n_1 - 1}S_{r-1},\ \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{T_{r+1}} = \bar{\boldsymbol{y}}.
\end{aligned}$$

In each step, we require $n_k > 1$ and $\boldsymbol{\mu}_{R_k} > 0, i = 1, \ldots, r$, to have a unique solution.

(a) The asymptotic normalities for conditional $R_{0v}, \ldots, R_{rv}$, and $\boldsymbol{\beta}$ are as follows.

$$
\begin{aligned}
&\sqrt{n_0}[R_{kv} - E(R_{kv}|T, \boldsymbol{y})]|_{\Theta_n=\boldsymbol{\mu}_{\Theta_n}}) \xrightarrow{\mathcal{L}} N_{p_1}(\mathbf{0}, \frac{2\boldsymbol{\mu}_{T_k}{}^{\otimes 2h}}{n_1 \ldots n_{r-k}}),\ k = 0, \ldots, r-1,\\
&\sqrt{n_0}[R_{rv} - E(R_{rv}|T, \boldsymbol{\beta}, \boldsymbol{y})]|_{\Theta_n=\boldsymbol{\mu}_{\Theta_n}}) \xrightarrow{\mathcal{L}} N_{p_1}(\mathbf{0}, 2\boldsymbol{\mu}_{T_r}{}^{\otimes 2h}),\\
&\sqrt{n_0}[\boldsymbol{\beta} - E(\boldsymbol{\beta}|T, R_r, \boldsymbol{y})|_{\Theta_n=\boldsymbol{\mu}_{\Theta_n}}] \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \boldsymbol{\mu}_{R_r}).
\end{aligned}
\tag{A.10}
$$

(b) Note that conditioning on T and $\boldsymbol{\beta}$, R_k's are independent, and

$$
\begin{aligned}
&\lim_{n_0\to\infty} \frac{\partial E(\boldsymbol{\phi}|T, \boldsymbol{y})}{\partial T'}\Big|_{T=\boldsymbol{\mu}_T} = I_{(r+1)p_1+p},\ \lim_{n_0\to\infty} \frac{\partial E(R|T, \boldsymbol{\beta}, \boldsymbol{y})}{\partial \boldsymbol{\beta}'}\Big|_{T=\boldsymbol{\mu}_T,\ \boldsymbol{\beta}=\boldsymbol{\mu}_{\boldsymbol{\beta}}} = \mathbf{O}_{(r+1)p_1,p},\\
&\lim_{n_0\to\infty} \frac{\partial E(\boldsymbol{\beta}|T, R, \boldsymbol{y})}{\partial R'}\Big|_{T=\boldsymbol{\mu}_T,\ R=\boldsymbol{\mu}_R} = \mathbf{O}_{p,(r+1)p_1}.
\end{aligned}
$$

The asymptotic covariance can then be simplified to a 2×2 block matrix (from an (r+3)$\times$(r+3) block matrix), $(T', \boldsymbol{\phi}')' = (\Theta_{1n}', \Theta_{2n}')'$, and expressed as

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} (I_{(r+1)p_1+p} - A)^{-1}\Sigma_1 & (I_{(r+1)p_1+p} - A)^{-1} A \Sigma_2^* \\ (I_{(r+1)p_1+p} - A)^{-1}\Sigma_1 & (I_{(r+1)p_1+p} - A)^{-1} \Sigma_2^* \end{pmatrix} = \begin{pmatrix} \Sigma_T & \Sigma_{T\boldsymbol{\phi}} \\ \Sigma_{T\boldsymbol{\phi}}' & \Sigma_{\boldsymbol{\phi}} \end{pmatrix},\\
\Sigma_T &= \Sigma_{T\boldsymbol{\phi}} = \Sigma_{T\boldsymbol{\phi}}' = (I_{(r+1)p_1+p} - A)^{-1}\Sigma_1 = \Sigma_{\boldsymbol{\phi}} - \Sigma_2^*,\\
\Sigma_{\boldsymbol{\phi}} &= (I_{(r+1)p_1+p} - A)^{-1}\Sigma_2^*,
\end{aligned}
\tag{A.11}
$$

where, from the Eqs. (16) and (A.10),

$$
\begin{aligned}
&\Sigma_1 = \lim_{n_0\to\infty} n_0 \cdot \mathrm{Cov}(T|\boldsymbol{\phi}, \boldsymbol{y})\big|_{\boldsymbol{\phi}=\boldsymbol{\mu}_{\boldsymbol{\phi}}},\\
&\Sigma_2^* = \mathrm{blkdiag}\{″_2, \ldots, ″_{r+3}\} = \mathrm{blkdiag}\{\frac{2}{\mathrm{n}_1 \ldots \mathrm{n}_\mathrm{r}} \boldsymbol{\mu}_{\mathrm{R}_0}{}^{\otimes 2\mathrm{h}}, \ldots, \frac{2}{\mathrm{n}_1} \boldsymbol{\mu}_{\mathrm{R}_{\mathrm{r}-1}}{}^{\otimes 2\mathrm{h}}, 2\boldsymbol{\mu}_{\mathrm{R}_\mathrm{r}}{}^{\otimes 2\mathrm{h}}, \boldsymbol{\mu}_{\mathrm{R}_\mathrm{r}}\},\\
&A = \Sigma_1 \Sigma_2^{*-1} = \lim_{n_0\to\infty} \frac{\partial E(T|\boldsymbol{\phi}, \boldsymbol{y})}{\partial \boldsymbol{\phi}'}\Big|_{\boldsymbol{\phi}=\boldsymbol{\mu}_{\boldsymbol{\phi}}}.
\end{aligned}
$$

(c) To simplify the calculation of A and $\Sigma_{\boldsymbol{\phi}}$, define $\Lambda_0 = R_0,\ \Lambda_k = n_{r-k+1} \cdots n_r R_k + \Lambda_{k-1}$. We have $R_k = \frac{1}{n_{r-k+1} \ldots n_r}(\Lambda_k - \Lambda_{k-1}),\ k = 1, \ldots, r$. Let $\Lambda = (\Lambda_{0v}', \ldots, \Lambda_{rv}', \boldsymbol{\beta}')'$. From the Eqs. (A.9) and (16), we obtain

$$
\begin{aligned}
W_k^{*-1} &= \Lambda_{k-1}\Lambda_k^{-1}\left[\frac{\Lambda_k - \Lambda_{k-1}}{n_{r-k+1} \cdots n_r}\right] = \left[\frac{\Lambda_k - \Lambda_{k-1}}{n_{r-k+1} \cdots n_r}\right]\Lambda_k^{-1}\Lambda_{k-1},\\
W_k^{*-1} Q_k &= \Lambda_{k-1}\Lambda_k^{-1},\ W_k^{*-1} P_{k-1} = I_p - \Lambda_{k-1}\Lambda_k^{-1},\ k = 1, \ldots, r.
\end{aligned}
\tag{A.12}
$$

Note that $\frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial \boldsymbol{\phi}'} = \frac{\partial E(T|\boldsymbol{\phi},\boldsymbol{y})}{\partial \Lambda'}\frac{\partial \Lambda}{\partial \boldsymbol{\phi}'}$, where $\frac{\partial \Lambda}{\partial R'} = \text{blkdiag}\{L_1^*, I_p\}$, L_1^* is a lower block triangular matrix with its block components

$$\{L_1^*\}_{ij} = \begin{cases} I_{p_1}, & i \geq 1,\ j = 1, \\ n_{r+2-j}\cdots n_r I_{p_1}, & i \geq j,\ j \geq 2, \\ \mathbf{O}_{p_1}, & i < j. \end{cases}$$

Also, $E(T|\boldsymbol{\phi}, \boldsymbol{y}) = E(T|\Lambda, \boldsymbol{y})$ depends on Λ through W_k^{*-1}, $W_k^{*-1}Q_k$, and $\boldsymbol{\beta}$, which are simple functions of Λ_{k-1}, Λ_k, and $\boldsymbol{\beta}$ only. Some algebra results in $\Sigma_{\boldsymbol{\phi}}^{-1} = \text{blkdiag}\{Q_A, S_r^{-1}\}$ and

$$Q_A = \frac{1}{2}L_1^{*\prime}\text{blkdiag}\{n_1\cdots n_{r-1}(n_r-1)\boldsymbol{\mu}_{\Lambda_0}^{\otimes 2h^{-1}}, \ldots, (n_1-1)\boldsymbol{\mu}_{\Lambda_{r-1}}^{\otimes 2h^{\ -1}}, \boldsymbol{\mu}_{\Lambda_r}^{\otimes 2h^{-1}}\}L_1^*,$$

where $\boldsymbol{\mu}_{\Lambda_k} = \frac{n_r\cdots n_{r-k}}{n_{r-k}-1}S_k$, $k = 0, \ldots, r-1$, $\boldsymbol{\mu}_{\Lambda_r} = n_r \ldots n_1 S_r$. Note that Q_A satisfies the quasi-triangle property (Rizvi 1984). Hence, the inverse of Q_A is a block quasi-tridiagonal matrix. By Theorem 2 in Rizvi and some calculations, we obtain Eq. (17).

(d) From the Eq. (A.11), and the Schur complement condition of positive definiteness, $\Sigma > 0$ if and only if $\Sigma_T > 0$ and $\Sigma_{\boldsymbol{\phi}} - \Sigma_T\Sigma_T^{-1}\Sigma_T = \Sigma_2^* > 0$. Note that the block diagonal entries of Σ_2^* are all positive definite, thus $\Sigma_2^* > 0$.

Similarly, $\Sigma_T > 0$ if and only if

$$\Sigma_T^{(1)} > 0,\ \{\Sigma_T\}_{kk} - B^{(k)\prime}\Sigma_T^{(k-1)^{-1}}B^{(k)} > 0,\ k = 2, \ldots, r+1,\ \{\Sigma_T\}_{r+2,r+2} > 0, \tag{A.13}$$

where $\{\Sigma_T\}_{ij}$ denotes the (i, j)th block of Σ_T, and the first $(r+1)$ leading principal block sub-matrices of Σ_T are

$$\Sigma_T^{(1)} = \{\Sigma_T\}_{11},\ \Sigma_T^{(k)} = \begin{pmatrix} \Sigma_T^{(k-1)} & B^{(k)} \\ B^{(k)\prime} & \{\Sigma_T\}_{kk} \end{pmatrix},\ k = 2, \ldots, r+1,$$
$$B^{(2)} = \{\Sigma_T\}_{12},\ B^{(k)'} = (\mathbf{O}_{p_1,(k-2)p_1}, \{\Sigma_T\}_{k-1,k}),\ k = 3, \ldots, r+1. \tag{A.14}$$

Note that $\{\Sigma_T\}_{r+2,r+2} = S_r - \boldsymbol{\mu}_{R_r} = \frac{1}{n_r-1}S_{r-1} > 0$. Consider the following lemma.

Lemma A.3

(a). $\Sigma_T^{(1)} = -\{\Sigma_T\}_{12} = \dfrac{2}{n_1 \ldots n_r(n_r-1)}\boldsymbol{\mu}_{R_0}{}^{\otimes 2h} > 0;$

$\{\Sigma_T\}_{k,k+1} = \Sigma_{k,k+1} < 0,\ k = 1, \ldots, r;$

(b). $\{\Sigma_T\}_{kk} + \{\Sigma_T\}_{k-1,k} + \{\Sigma_T\}_{k,k+1} = \dfrac{2}{n_1 \ldots n_{r-k+1}(n_{r-k+2}-1)}$

$\times L_p N_p(\boldsymbol{\mu}_{R_{k-1}} \otimes S_{k-2} + S_{k-2} \otimes \boldsymbol{\mu}_{R_{k-1}})N_p L_p' > 0,\ k = 2, \ldots, r+1.$

When $k = r + 1$, we set $n_1 ... n_{r-k+1} = 1$.

Proof: We only show that (b) holds when $k = 2, \ldots, r$ since the proof of (b) when $k = r + 1$ is similar to the proof of (b) when $k < r + 1$ and the proof of (a) is straightforward. By the Eq. (16), we have $\boldsymbol{\mu}_{R_{k-1}} = \frac{n_{r-k+1}}{n_{r-k+1}-1} S_{k-1} - \frac{1}{n_{r-k+2}-1} S_{k-2}$ or

$$S_{k-1}{}^{\otimes 2h} = \frac{(n_{r-k+1}-1)^2}{n_{r-k+1}^2} \left[\boldsymbol{\mu}_{R_{k-1}} + \frac{1}{n_{r-k+2}-1} S_{k-2} \right]^{\otimes 2h}, \ k = 2, \ldots, r-1.$$

Thus,

$$\begin{aligned}
&\{\Sigma_T\}_{kk} + \{\Sigma_T\}_{k-1,k} + \{\Sigma_T\}_{k,k+1} = \Sigma_{kk} - \{\Sigma_2^*\}_{kk} + \Sigma_{k-1,k} + \Sigma_{k,k+1} \\
&= \frac{2}{n_1 \ldots n_{r-k}} \left[\frac{S_{k-2}{}^{\otimes 2h}}{n_{r-k+1}(n_{r-k+2}-1)^3} + \frac{n_{r-k+1}^2 S_{k-1}{}^{\otimes 2h}}{(n_{r-k+1}-1)^3} \right] - \frac{2}{n_1 \cdots n_{r-k+1}} \boldsymbol{\mu}_{R_{k-1}}{}^{\otimes 2h} \\
&\quad - \frac{2 \cdot n_{r-k+2} S_{k-2}{}^{\otimes 2h}}{n_1 \cdots n_{r-k+1}(n_{r-k+2}-1)^3} - \frac{2 \cdot n_{r-k+1} S_{k-1}{}^{\otimes 2h}}{n_1 \cdots n_{r-k}(n_{r-k+1}-1)^3} \\
&= \frac{2 \cdot n_{r-k+1} S_{k-1}{}^{\otimes 2h}}{n_1 \cdots n_{r-k}(n_{r-k+1}-1)^2} - \frac{2 \cdot S_{k-2}{}^{\otimes 2h}}{n_1 \cdots n_{r-k+1}(n_{r-k+2}-1)^2} - \frac{2}{n_1 \cdots n_{r-k+1}} \boldsymbol{\mu}_{R_{k-1}}{}^{\otimes 2h} \\
&= \frac{2}{n_1 \cdots n_{r-k+1}(n_{r-k+2}-1)} L_p N_p (\boldsymbol{\mu}_{R_{k-1}} \otimes S_{k-2} + S_{k-2} \otimes \boldsymbol{\mu}_{R_{k-1}}) N_p L_p' > 0, \ k = 2, \ldots, r.
\end{aligned}$$

We then prove that the Eq. (A.13) is true by Lemma A.3 and the mathematical induction. Also, for any two square symmetric matrices A and B, $A > B$ ($A \geq B$) denotes $A - B > 0$ ($A - B \geq 0$).

For $m = 1$, by Lemma A.3 (a)., we have $\Sigma_T^{(1)} > 0$. For $m = 2$, by the Eq. (A.14) and Lemma A.3 (a), $-B^{(2)\prime} {\Sigma_T^{(1)}}^{-1} B^{(2)} = -\{\Sigma_T\}_{12}' {\Sigma_T^{(1)}}^{-1} \{\Sigma_T\}_{12} = \{\Sigma_T\}_{12}$ and $\{\Sigma_T\}_{23} = \Sigma_{23} < 0$, and by Lemma A.3 (b), we have

$$\{\Sigma_T\}_{22} - B^{(2)\prime} {\Sigma_T^{(1)}}^{-1} B^{(2)} = \{\Sigma_T\}_{22} + \{\Sigma_T\}_{12} > \{\Sigma_T\}_{22} + \{\Sigma_T\}_{12} + \{\Sigma_T\}_{23} > 0.$$

Suppose for $m = k$, $2 \leq k \leq r - 1$,

$$-B^{(k)\prime} {\Sigma_T^{(k-1)}}^{-1} B^{(k)} \geq \{\Sigma_T\}_{k-1,k}, \ \{\Sigma_T\}_{kk} - B^{(k)\prime} {\Sigma_T^{(k-1)}}^{-1} B^{(k)} > 0.$$

For $m = k + 1$, by Lemma A.3, $\{\Sigma_T\}_{kk} + \{\Sigma_T\}_{k-1,k} + \{\Sigma_T\}_{k,k+1} > 0$ and $-\{\Sigma_T\}_{k,k+1} = -\Sigma_{k,k+1} > 0$. This implies

$$\{\Sigma_T\}_{kk} - B^{(k)\prime} {\Sigma_T^{(k-1)}}^{-1} B^{(k)} \geq \{\Sigma_T\}_{kk} + \{\Sigma_T\}_{k-1,k} > -\{\Sigma_T\}_{k,k+1} > 0,$$

and hence

$$\begin{aligned}
-B^{(k+1)\prime} {\Sigma_T^{(k)}}^{-1} B^{(k+1)} &= -(\mathbf{O}_{p_1,(k-1)p_1}, \{\Sigma_T\}_{k,k+1}) {\Sigma_T^{(k)}}^{-1} (\mathbf{O}_{p_1,(k-1)p_1}, \{\Sigma_T\}_{k,k+1})' \\
&= -\{\Sigma_T\}_{k,k+1} [\{\Sigma_T\}_{kk} - B^{(k)\prime} {\Sigma_T^{(k-1)}}^{-1} B^{(k)}]^{-1} \{\Sigma_T\}_{k,k+1} \geq \{\Sigma_T\}_{k,k+1}.
\end{aligned}$$

Similarly, by Lemma A.3, $\{\Sigma_T\}_{k+1,k+1} + \{\Sigma_T\}_{k,k+1} + \{\Sigma_T\}_{k+1,k+2} > 0$ and $-\{\Sigma_T\}_{k+1,k+2} = -\Sigma_{k+1,k+2} > 0$, we have

$$\begin{aligned}&\{\Sigma_T\}_{k+1,k+1} - B^{(k+1)'} \Sigma_T^{(k)^{-1}} B^{(k+1)} \\ &\geq \{\Sigma_T\}_{k+1,k+1} + \{\Sigma_T\}_{k,k+1} > -\{\Sigma_T\}_{k+1,k+2} > 0.\end{aligned}$$

This completes the proof. □

Proofs of Theorem 4 and Corollary 2: The proofs of Theorem 4 and Corollary 2 are similar to Theorems 2 and 3, and Corollary 1, respectively. We omit them to save space. □

References

Akritas M, Arnold S (2000) Asymptotics for ANOVA when the number of levels is large. J Am Stat Assoc 95(449):212–226

Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New York

Anisa R, Kurnia A, Indahwati (2014) Cluster information of non-sampled area in small area estimation. IOSR J Math 10(1):15–19

Arnold S (1980) Asymptotic validity of F tests for the ordinary linear model and the multiple correlation model. J Am Stat Assoc 75(372):890–894

Bathke A (2002) ANOVA for a large number of treatments. Math Methods Stat 11(1):118–132

Boos DD, Brownie C (1995) ANOVA and rank test when the number of treatments is large. Stat Probab Lett 23(2):183–191

Chen CF (1985) On asymptotic normality of limiting density function with Bayesian implications. J R Stat Soc B 47(3):540–546

Gelfand AE, Sahu SK, Carlin BP (1995) Efficient parametrisations for normal linear mixed models. Biometrika 82(3):479–488

Gupta AK, Harrar SW, Fujihoshi Y (2006) Asymptotics for testing hypothesis in some multivariate variance components model under non-normality. J Multivar Anal 97(1):148–178

Horn RA, Johnson CR (2012) Matrix analysis, 2nd edn. Cambridge University Press, New York

Jelenkowska TH (1995) A Bayesian analysis of the multivariate mixed linear model. Commun Stat Theory Methods 24(12):3183–3196

Jelenkowska TH, Press SJ (1997) Bayesian inference in the multivariate mixed model MANOVA. Am J Math Manag Sci 17(1–2):97–116

Johnson RA (1967) An asymptotic expansion for posterior distributions. Ann Math Stat 38(6):1899–1906

Kolassa JE, Kuffner TA (2020) On the validity of the formal Edgeworth expansion for posterior densities. Ann Stat 48(4):1940–1958

Magnus JR, Neudecker H (1980) The elimination matrix: some lemmas and applications. SIAM J Alg Disc Meth 1(4):422–449

Magnus JR, Neudecker H (1999) Matrix differential calculus with applications in statistics and econometrics, 2nd edn. Wiley, New York

Meng XL (1994) On the rate of convergence of the ECM algorithm. Ann Stat 22(1):326–339

Montgomery DC (2012) Design and analysis of experiments, 8th edn. Wiley, New York

Press SJ (2002) Subjective and objective Bayesian statistics: principles, models, and applications, 2nd edn. Wiley, New York

Press SJ (2005) Applied multivariate analysis: using Bayesian and frequentist methods of inference, 2nd edn. Dover, New York

Rizvi SAH (1984) Inverse of quasi-tridiagonal matrices. Linear Algebra Appl 56:177–184

Roberts GO, Sahu SK (1997) Updating scheme, correlation structure, blocking and parametrization for the Gibbs sampler. J R Stat Soc B 59(2):291–317

Sahai H, Ojeda MM (2004) Analysis of variance for random models volume I: balanced data. Birkhäuser, Boston

Seber GAF (2007) A matrix handbook for statisticians, 1st edn. Wiley, New York
Su CL, Johnson WO (2006) Large-sample joint posterior approximations when full conditionals are approximately normal: application to generalized linear mixed models. J Am Stat Assoc 101(474):795–811
Su CL (2017) Asymptotic posterior distributions for balanced nested multi-way models with a large number of main effect levels. Commun Stat Theory Methods 46(19):9425–9440
Szatrowski TH, Miller JJ (1980) Explicit maximum likelihod estimates form balanced data in the mixed model of the analysis of vairance. Ann Stat 8(4):811–819
Wang L, Akritas M (2006) Two-way heteroscedastic ANOVA when the number of levels is large. Stat Sin 16:1387–1408