Anti-ridge regression for communality estimation in factor analysis

Preprint · April 2024

DOI: 10.13140/RG.2.2.22659.92966

1 author:

Haruhiko Ogasawara
Otaru University of Commerce

Anti-ridge regression for communality estimation in factor analysis

April 13, 2024

Haruhiko Ogasawara*

*Professor Emeritus, Otaru University of Commerce, Otaru 047-8501 Japan; Email: emt-hogasa@emt.otaru-uc.ac.jp; Web: https://www.otaru-uc.ac.jp/~emt-hogasa/

Abstract: For communality estimation in factor analysis, a method called anti-ridge (AR) regression using a negative ridge parameter is presented to improve the squared multiple correlation coefficient (SMC) for a lower bound of the communality. It is shown that the optimal improved SMC called the ARSMC is given when the anti-ridge parameter (the absolute value of the negative ridge parameter) is the smallest uniqueness in a correlation matrix. When the one-factor model holds, a method to have the smallest uniqueness is provided, which gives the exact communalities. For the multi-factor model, a communality estimator using the Moore-Penrose generalized inverse in the ARSMC is provided.

Keywords: ridge parameter, uniqueness, correlation matrix, positive semidefinite, Moore-Penrose generalized inverse.

1. Introduction

Let $\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}^{\mathrm{T}} + \mathbf{\Psi}$ be a covariance matrix for a *p*-dimensional random vector **X** following the exploratory factor analysis (EFA) model with $k(< p)$ common factors, where $\mathbf{\Lambda} = (\boldsymbol{\lambda}_1, ...\boldsymbol{\lambda}_p)^{\mathrm{T}}$ with $\boldsymbol{\lambda}_i = (\lambda_{i1}, ..., \lambda_{ik})^{\mathrm{T}} (i = 1, ..., p)$; and $\mathbf{\Psi}$ is the positive definite (p.d.) $p \times p$ diagonal matrix with the diagonal elements $\psi_i (i = 1, ..., p)$, which is denoted by $\mathbf{\Psi} = \mathrm{diag}(\psi_1, ..., \psi_p) > \mathbf{O}$ in Löwner's sense. Suppose that $\boldsymbol{\lambda}_i (i = 1, ..., p)$ are not zero vectors. Then, we have $\mathbf{\Sigma} \geq \mathbf{\Psi}$ or $\mathbf{\Sigma} - \mathbf{\Psi} \geq \mathbf{O}$, which indicates that $\mathbf{\Sigma} - \mathbf{\Psi}$ is positive semidefinite (p.s.d). Let $h_i^2 = \boldsymbol{\lambda}_i^{\mathrm{T}} \boldsymbol{\lambda}_i (i = 1, ..., p)$ be communalities typically defined when $\sigma_{ii} = 1 \ (i = 1, ..., p)$ with $\mathbf{\Sigma} = \{\sigma_{ij}\}(i, j = 1, ..., p)$. In the following $\sigma_{ii} = 1$ is assumed unless otherwise stated. Define $\mathbf{\Lambda}_{(-i)}$ as the $(p-1) \times k$ matrix removing the *i*-th row in $\mathbf{\Lambda}$. Similarly, the $(p-1) \times (p-1)$ diagonal matrix $\mathbf{\Psi}_{(-i)}$ is defined with $\mathbf{\Sigma}_{(-i)} \equiv \mathbf{\Lambda}_{(-i)} \mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)} > \mathbf{O} \, (i = 1, ..., p)$.

It has been known that the squared multiple correlation coefficient (SMC) of the *i*-th variable of **X** in multiple regression on the remaining $p-1$ variables denoted by SMC_i is a lower bound of h_i^2 (Roff, 1936, Corollary) i.e., $h_i^2 \geq \mathrm{SMC}_i = \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}} \mathbf{\Sigma}_{(-i)}^{-1} \boldsymbol{\sigma}_{(-i)}$, where $\boldsymbol{\sigma}_{(-i)} = (\sigma_{i1}, ...\sigma_{i,i-1}, \sigma_{i,i+1}, ..., \sigma_{ip})^{\mathrm{T}} \ (i = 1, ..., p)$. When *k* common factors consist of major k_1 and relatively minor k_2 ones with $k = k_1 + k_2$, we have the minimum trace factor analysis (MTFA; Bentler, 1972; ten Berge, Snijders & Zegers, 1981) and minimum-rank FA (MRFA; ten Berge & Kiers,1991; Shapiro & ten Berge, 2002) giving largest communalities in some similar senses with h_i^2 defined for *k* common factors. These results are seen as exact expressions of h_i^2. However, generally they require iterative computations as in low rank FA or semidefinite optimization (Bertsimas, Copenhaver & Mazumder, 2017; Bertsimas, Cory-Wright & Pauphilet, 2023; Tunçel, Vavasis & Xu, 2023).

Non-iterative methods by Gibson (1963), Ihara and Kano (1986) and Kano (1989, 1990) gave exact h_i^2 when there are two off-diagonal submatrices of rank *k* satisfying Anderson and Rubin's condition (1956, Theorem 5.1) under correct model specification. Yanai and

Ichikawa (1990) gave an improved lower bound for h_i^2 based on the eigenvalues of $\mathbf{\Sigma}$. However, they admitted the difficulty of using their result when there are more-than-one common factors. Ramsey and Gibson (2006) provided modifications of the Ihara-Kano method.

The purpose of this paper is to propose a method of estimating h_i^2 in EFA. The remainder of this paper is organized as follows. Section 2 gives theoretical considerations for lower bounds based on a modified SMC using a negative ridge parameter, which is called anti-ridge regression. In Section 3, a method of deriving the exact communalities in the one-factor model is provided while in Section 4, a method of communality estimation in the multi-factor model is shown. Section 5 gives numerical illustrations of the one- and multi-factor models. In Section 6, some discussions are given.

2. Lower bounds for the communality

As mentioned earlier, SMC_i gives a lower bound of h_i^2, whose derivation is shown to have an extension later. Note that $\boldsymbol{\sigma}_{(-i)} = \mathbf{\Lambda}_{(-i)}\boldsymbol{\lambda}_i$. Then, using the Guttman (1946, Equation (13)) formula (often called the Woodbury formula due to the rediscovery by Max A. Woodbury in 1950), we have

$$\begin{aligned} h_i^2 - \mathrm{SMC}_i &= \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\lambda}_i - \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\mathbf{\Sigma}_{(-i)}^{-1}\boldsymbol{\sigma}_{(-i)} = \boldsymbol{\lambda}_i^{\mathrm{T}}(\mathbf{I}_k - \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Sigma}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)})\boldsymbol{\lambda}_i \\ &= \boldsymbol{\lambda}_i^{\mathrm{T}}\left[\mathbf{I}_k - \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\left\{\mathbf{\Psi}_{(-i)}^{-1} - \mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)}(\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)} + \mathbf{I}_k)^{-1}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\right\}\mathbf{\Lambda}_{(-i)}\right]\boldsymbol{\lambda}_i \\ &= \boldsymbol{\lambda}_i^{\mathrm{T}}\left\{\mathbf{I}_k - \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)} + \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)}(\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)} + \mathbf{I}_k)^{-1}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)}\right\}\boldsymbol{\lambda}_i, \end{aligned}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix. In the above expression, take the spectral decomposition

$$\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(-i)} = \mathbf{\Gamma}_i\mathrm{diag}(\gamma_{i1},\ldots,\gamma_{ik})\mathbf{\Gamma}_i^{\mathrm{T}} \text{ with } \mathbf{\Gamma}_i\mathbf{\Gamma}_i^{\mathrm{T}} = \mathbf{\Gamma}_i^{\mathrm{T}}\mathbf{\Gamma}_i = \mathbf{I}_k \text{ and } \gamma_{i1} \ge \cdots \ge \gamma_{ik} > 0.$$

Then, we obtain

$$\begin{aligned} h_i^2 - \mathrm{SMC}_i &= \boldsymbol{\lambda}_i^{\mathrm{T}}\mathbf{\Gamma}_i\left\{\mathbf{I}_k - \mathrm{diag}(\gamma_{i1},\ldots,\gamma_{ik}) + \mathrm{diag}\left(\frac{\gamma_{i1}^2}{1+\gamma_{i1}},\ldots,\frac{\gamma_{ik}^2}{1+\gamma_{ik}}\right)\right\}\mathbf{\Gamma}_i^{\mathrm{T}}\boldsymbol{\lambda}_i \\ &= \boldsymbol{\lambda}_i^{\mathrm{T}}\mathbf{\Gamma}_i\mathrm{diag}\left(\frac{1}{1+\gamma_{i1}},\ldots,\frac{1}{1+\gamma_{ik}}\right)\mathbf{\Gamma}_i^{\mathrm{T}}\boldsymbol{\lambda}_i > 0 \ (i = 1,\ldots,p), \end{aligned}$$

which are the required inequalities.

Consider the re-expression of the communalities in the following way:

$$h_i^2 = \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-}\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\lambda}_i\,(i=1,\ldots,p),$$

where $\mathbf{A}^{-}$ is the generic expression for a generalized (*g*-) inverse of $\mathbf{A}$ satisfying $\mathbf{A}\,\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$. The above equation is derived by the definition $\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-}\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} = \boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}$ and pre- and post-multiplying the both sides of the equation by $\boldsymbol{\Lambda}_{(-i)}^{+} \equiv (\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)})^{-1}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}$ and $\boldsymbol{\Lambda}_{(-i)}^{+\mathrm{T}} \equiv (\boldsymbol{\Lambda}_{(-i)}^{+})^{\mathrm{T}}$, respectively, yielding $\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-}\boldsymbol{\Lambda}_{(-i)} = \mathbf{I}_k$. Note that $(\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)})^{-1}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}$ is known as the left-inverse of $\boldsymbol{\Lambda}_{(-i)}$ of full column rank (Browne, 1974/1977, Equation (7)), which is also the Moore-Penrose *g*-inverse of $\boldsymbol{\Lambda}_{(-i)}$ denoted by $\boldsymbol{\Lambda}_{(-i)}^{+}$ (see e.g., Yanai, 1990) with $(\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{+} = (\boldsymbol{\Lambda}_{(-i)}^{+})^{\mathrm{T}}$ in this case. Then, we have

$$\begin{aligned} h_i^2 - \mathrm{SMC}_i &= \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\lambda}_i - \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(-i)}^{-1}\boldsymbol{\sigma}_{(-i)} = \boldsymbol{\lambda}_i^{\mathrm{T}}(\mathbf{I}_k - \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(-i)}^{-1}\boldsymbol{\Lambda}_{(-i)})\boldsymbol{\lambda}_i \\ &= \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\{(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-} - (\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} + \boldsymbol{\Psi}_{(-i)})^{-1}\}\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\lambda}_i. \end{aligned}$$

An improved lower bound for h_i^2 may be obtained when we can find a matrix $\mathbf{A}_i$ from $\boldsymbol{\Sigma}_{(-i)}$ satisfying

$$\mathbf{I}_k = \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-}\boldsymbol{\Lambda}_{(-i)} > \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\Lambda}_{(-i)} > \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} + \boldsymbol{\Psi}_{(-i)})^{-1}\boldsymbol{\Lambda}_{(-i)} > \mathbf{O},$$

which gives $h_i^2 > \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\lambda}_i \equiv h_{\mathrm{A}i}^2 > \mathrm{SMC}_i$. Note that $(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{-}$ of rank *k* is positive semidefinite (p.s.d.), whose convenient special case is $(\boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}})^{+} = \boldsymbol{\Lambda}_{(-i)}(\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Lambda}_{(-i)})^{-2}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}$.

In order to have $\mathbf{A}_i$ with $\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\Lambda}_{(-i)}$ hopefully close to $\mathbf{I}_k$, consider the decomposition

$$\boldsymbol{\Sigma}_{(-i)} = \boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} + \boldsymbol{\Psi}_{(-i)} = \boldsymbol{\Sigma}_{\mathrm{S}(-i)} + \mathbf{D}_{\mathrm{P}(-i)},$$

where $\boldsymbol{\Sigma}_{\mathrm{S}(-i)}$ is a p.s.d. matrix of rank $p-2$ or smaller; and $\mathbf{D}_{\mathrm{P}(-i)} = \lambda_{(-i)p-1}\mathbf{I}_{p-1}$ is the positive definite (p.d.) diagonal matrix whose common diagonal element $\lambda_{(-i)p-1}$ is the smallest eigenvalue of $\mathbf{\Sigma}_{(-i)}$ with multiplicity $m_{(-i)}$. The positive eigenvalues of $\mathbf{\Sigma}_{\mathrm{S}(-i)}$ of rank $p - m_{(-i)} - 1$ are $\lambda_{(-i)1} - \lambda_{(-i)p-1} \geq \cdots \geq \lambda_{(-i)p-m_{(-i)}-1} - \lambda_{(-i)p-1} > 0$ with similar definitions of $\lambda_{(-i)k} (k = 1,..., p-2)$. When $\mathbf{D}_{\mathrm{P}(-i)}$ is relaxed to be a p.s.d. diagonal matrix with a single positive diagonal element given by the corresponding $1-\mathrm{SMC}$ from regression on the $p-2$ remaining variables, $p-1$ cases of $\mathbf{\Sigma}_{\mathrm{S}(-i)}$ are similarly obtained, as was discussed by Guttman (1958, p. 300).

An intriguing candidate is given from $\mathbf{\Sigma}_{\mathrm{S}(-i)}$ as $\mathbf{A}_i = \mathbf{\Sigma}^{+}_{\mathrm{S}(-i)}$ when $\mathbf{D}_{\mathrm{P}(-i)} = \lambda_{(-i)p-1}\mathbf{I}_{p-1}$ i.e., $\mathbf{\Sigma}^{+}_{\mathrm{S}(-i)} = (\mathbf{\Sigma}_{(-i)} - \mathbf{D}_{\mathrm{P}(-i)})^{+} = (\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^{+}$. In the special case of $\mathbf{\Sigma}_{(-i)} = a\mathbf{1}_{p-1}\mathbf{1}^{\mathrm{T}}_{p-1} + (1-a)\mathbf{I}_{p-1}$ ($0 < a < 1$) with compound symmetry or mutually exchangeable $p-1$ variables with respect to the covariance matrix, and $\mathbf{1}_{p-1} = (1,...,1)^{\mathrm{T}}$ ($p-1$ times of 1), there are two distinct eigenvalues i.e., the single largest $\lambda_{(-i)1} = (p-1)a + 1 - a$ and $\lambda_{(-i)2} = \cdots = \lambda_{(-i)p-1} = 1-a$ with $m_{(-i)} = p-2$. In this case, using $\mathbf{D}_{\mathrm{P}(-i)} = (1-a)\mathbf{I}_{p-1}$, we exactly obtain $\mathbf{\Sigma}_{\mathrm{S}(-i)} = a\mathbf{1}_{p-1}\mathbf{1}^{\mathrm{T}}_{p-1} = \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}^{\mathrm{T}}_{(-i)}$. On the other hand, the method using the diagonal matrix with a single positive diagonal element of the associated $1-\mathrm{SMC}_i$ does not restore $\mathbf{\Sigma}_{\mathrm{S}(-i)} = \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}^{\mathrm{T}}_{(-i)}$ in this case.

However, generally the diagonal elements of $\mathbf{\Psi}_{(-i)}$ are not the same, which does not satisfy a necessary condition for an improved lower bound

$$\mathbf{\Lambda}^{\mathrm{T}}_{(-i)}\mathbf{A}_i\mathbf{\Lambda}_{(-i)} > \mathbf{\Lambda}^{\mathrm{T}}_{(-i)}(\mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}^{\mathrm{T}}_{(-i)} + \mathbf{\Psi}_{(-i)})^{-1}\mathbf{\Lambda}_{(-i)}$$

when $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^{+}$. This is found from the following theorem.

Theorem 1 (Milliken & Akdeniz, 1977, Theorem 3.1). *Let* $\mathbf{A}, \mathbf{B}$ *and* $\mathbf{A} - \mathbf{B}$ *be* $p \times p$ *p.s.d. matrices. Then, a necessary and sufficient condition for* $\mathbf{B}^{+} - \mathbf{A}^{+}$ *to be p.s.d. or equivalently* $\mathbf{B}^{+} \geq \mathbf{A}^{+} \geq \mathbf{O}$ *is that* $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$.

Note that the rank of $\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1}$ in $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^{+}$ is less than or equal to $p-2$ as addressed earlier while that of $\mathbf{\Sigma}_{(-i)} = \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}^{\mathrm{T}}_{(-i)} + \mathbf{\Psi}_{(-i)}$ is $p-1$ by assumption, which does not satisfy the condition of the Milliken-Akdeniz (M-K) theorem.

Consequently, $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^+$ does not satisfy the condition for an improved lower bound over SMC_i though it can give a reasonable estimator of h_i^2. The M-K theorem for the Löwner ordering $\mathbf{B}^+ \geq \mathbf{A}^+ \geq \mathbf{O}$ has been well known in linear algebra (see Neudecker, 1989; Baksalary, Nordström & Styan, 1990; Wang & Dai, 2010). In psychometrics, however, to the author's knowledge the M-K theorem has not been used in the literature.

An alternative candidate of $\mathbf{A}_i$ with the same rank as that of $\mathbf{\Sigma}_{(-i)}^{-1}$ or $\mathbf{\Sigma}_{(-i)}$ satisfying the condition $\mathbf{A}_i > \mathbf{\Sigma}_{(-i)}^{-1} = \mathbf{\Sigma}_{(-i)}^{+}$ is easily found as $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}$, when $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$. Recall the other condition for an improved lower bound

$$\mathbf{I}_k = \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}(\mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}})^{-}\mathbf{\Lambda}_{(-i)} = \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}(\mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}})^{+}\mathbf{\Lambda}_{(-i)} > \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\mathbf{\Lambda}_{(-i)}.$$

Whether $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}$ satisfies the above condition depends on the choice of $\delta_{(-i)}$ among $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$. We have the following result:

Theorem 2 (Anti-ridge regression for lower bounds of the communalities). *Define* $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}$, *when* $\delta_{(-i)}$ *with* $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$ *is sufficiently close to* 0, *we have an improved lower bound* $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\sigma}_{(-i)}$ *of the communality over* SMC_i *as*

$$h_i^2 = \boldsymbol{\lambda}_i^{\mathrm{T}}\boldsymbol{\lambda}_i > \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\sigma}_{(-i)} > \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\mathbf{\Sigma}_{(-i)}^{-1}\boldsymbol{\sigma}_{(-i)} = \mathrm{SMC}_i \ \ (i = 1, \ldots, p).$$

Proof. When $\delta_{(-i)}$ is sufficiently close to 0, by continuity $\mathbf{A}_i = (\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}$ satisfies the condition $\mathbf{I}_k > \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\mathbf{\Lambda}_{(-i)}$. The remaining condition $\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\mathbf{\Lambda}_{(-i)} > \mathbf{\Lambda}_{(-i)}^{\mathrm{T}}(\mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)})^{-1}\mathbf{\Lambda}_{(-i)}$ can be shown to be satisfied by the M-K theorem. Q.E.D.

Remark 1. Note that $\mathrm{SMC}_i = \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}\mathbf{\Sigma}_{(-i)}^{-1}\boldsymbol{\sigma}_{(-i)}$ is given by usual multiple regression. When using ridge regression, the SMC_i counterpart (ridge SMC_i) becomes e.g., $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} + \varepsilon\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_{(-i)}$ or $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} + \varepsilon\mathbf{I}_{p-1})^{-1}\mathbf{\Sigma}_{(-i)}(\mathbf{\Sigma}_{(-i)} + \varepsilon\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_{(-i)}$, where the ridge parameter ε is a positive constant. The ridge SMC_i is majorized by SMC_i and is a poorer lower bound than SMC_i. From the form of $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Sigma}_{(-i)} - \varepsilon\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_{(-i)}$ for the improved lower bound, the term "anti-ridge (AR) regression" is used with a positive "anti-ridge parameter" ε and the notation $\mathrm{ARSMC}_i(\varepsilon) = \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}(\boldsymbol{\Sigma}_{(-i)} - \varepsilon\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_{(-i)}$.

Theorem 2 is given for theoretical purposes since when $\delta_{(-i)}$ is sufficiently close to 0, the amount of improvement as a lower bound of the communality is small. On the other hand, when $\delta_{(-i)}$ in $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$ is close to $\lambda_{(-i)p-1}$, while the condition $\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\Lambda}_{(-i)} > \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(-i)}^{-1}\boldsymbol{\Lambda}_{(-i)}$ is satisfied, $\mathbf{I}_k > \boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\Lambda}_{(-i)}$ may not always be satisfied. For this property, the following theorem is provided.

Theorem 3 (Beckenbach & Bellman, 1965, Chapter 2, Section 26, Theorem 19, Equation (3)). *Let* $\lambda_i(\mathbf{A})$ *be the i-th largest eigenvalue of the* $p \times p$ *matrix* $\mathbf{A}\,(i = 1,\ldots,p)$. *Suppose that* $\mathbf{A}$ *is symmetric and the* $p \times p$ *matrix* $\mathbf{B}$ *is p.s.d. Then,* $\lambda_i(\mathbf{A}+\mathbf{B}) \geq \lambda_i(\mathbf{A})$ $(i = 1,\ldots,p)$.

When $\mathbf{A} = \boldsymbol{\Psi}_{(-i)}$ and $\mathbf{B} = \boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}$ in the Beckenbach-Bellman (B-B) inequality, the smallest eigenvalue $\lambda_{p-1}(\mathbf{A}+\mathbf{B}) = \lambda_{p-1}(\boldsymbol{\Sigma}_{(-i)}) \geq \lambda_{p-1}(\boldsymbol{\Psi}_{(-i)}) = \min\{\psi_j \mid j = 1,\ldots,p, j \neq i\}$. That is, the smallest eigenvalue $\lambda_{(-i)p-1}(= \lambda_{p-1}(\boldsymbol{\Sigma}_{(-i)}))$ of $\boldsymbol{\Sigma}_{(-i)}$ is larger than or equal to the smallest value of $\psi_j\,(j = 1,\ldots,p, j \neq i)$. Consequently, when the strict inequality $\lambda_{p-1}(\boldsymbol{\Sigma}_{(-i)}) > \lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$ holds and $\delta_{(-i)}$ in $\mathbf{A}_i = (\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}$ with $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$ is close to $\lambda_{(-i)p-1}$, at least a diagonal element of $\boldsymbol{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I}_p$ in

$$\mathbf{A}_{(-i)}^{-1} = \boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1} = \boldsymbol{\Lambda}_{(-i)}\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} + \boldsymbol{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I}_p$$

becomes negative. Note that the B-B inequality was used in the communality problem for $\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}_{(-i)}$ by Yanai and Ichikawa (1990, Theorem 1).

Let $\boldsymbol{\Psi}_{(-i)}^{(\delta)} \equiv \boldsymbol{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I}_p$ for simplicity of notation. When $\delta_{(-i)}$ is close to $\lambda_{(-i)p-1}$, and when no diagonal element of $\boldsymbol{\Psi}_{(-i)}^{(\delta)}$ is zero, consider the spectral decomposition

$$\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}}\boldsymbol{\Psi}_{(-i)}^{(\delta)-1}\boldsymbol{\Lambda}_{(i))} = \boldsymbol{\Gamma}_i^{(\delta)}\mathrm{diag}(\gamma_{i1}^{(\delta)},\ldots,\gamma_{ik}^{(\delta)})\boldsymbol{\Gamma}_i^{(\delta)\mathrm{T}},\ \boldsymbol{\Gamma}_i^{(\delta)}\boldsymbol{\Gamma}_i^{(\delta)\mathrm{T}} = \boldsymbol{\Gamma}_i^{(\delta)\mathrm{T}}\boldsymbol{\Gamma}_i^{(\delta)} = \mathbf{I}_k$$
$$(\gamma_{i1}^{(\delta)} \geq \cdots \geq \gamma_{ik}^{(\delta)};\ i = 1,\ldots,p)$$

corresponding to that of $\mathbf{\Lambda}_{(-i)}^{\mathrm{T}}\mathbf{\Psi}_{(-i)}^{-1}\mathbf{\Lambda}_{(i)}$ given earlier. Since at least a diagonal element of $\mathbf{\Psi}_{(-i)}^{(\delta)}$ is negative, the smallest eigenvalue $\gamma_{ik}^{(\delta)}$ among $\gamma_{i1}^{(\delta)},...,\gamma_{ik}^{(\delta)}$ can be negative. Then, as before

$$
\begin{aligned}
& h_i^2 - \boldsymbol{\sigma}_i^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_i \\
&= \boldsymbol{\lambda}_i^{\mathrm{T}}\mathbf{\Gamma}_i^{(\delta)}\left\{\mathbf{I}_k - \mathrm{diag}(\gamma_{i1}^{(\delta)},...,\gamma_{ik}^{(\delta)}) + \mathrm{diag}\left(\frac{\gamma_{i1}^{(\delta)2}}{1+\gamma_{i1}^{(\delta)}},...,\frac{\gamma_{ik}^{(\delta)2}}{1+\gamma_{i1}^{(\delta)}}\right)\right\}\mathbf{\Gamma}_i^{(\delta)\mathrm{T}}\boldsymbol{\lambda}_i \\
&= \boldsymbol{\lambda}_i^{\mathrm{T}}\mathbf{\Gamma}_i^{(\delta)}\mathrm{diag}\left(\frac{1}{1+\gamma_{i1}^{(\delta)}},...,\frac{1}{1+\gamma_{ik}^{(\delta)}}\right)\mathbf{\Gamma}_i^{(\delta)\mathrm{T}}\boldsymbol{\lambda}_i \ (i=1,...,p).
\end{aligned}
$$

In the above result, when $\gamma_{ik}^{(\delta)} < -1$, which can happen, $h_i^2 - \boldsymbol{\sigma}_i^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_i > 0$ is not always satisfied, which shows that when $\delta_{(-i)}$ is close to $\lambda_{(-i)p-1}$, $\boldsymbol{\sigma}_i^{\mathrm{T}}\mathbf{A}_i\boldsymbol{\sigma}_i$ $= \boldsymbol{\sigma}_i^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_i$ cannot be a lower bound of the communality.

Theorem 4 (the largest anti-ridge parameter for a lower bound). *When $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) < \lambda_{(-i)p-1}$, the largest $\delta_{(-i)}$ among $0 < \delta_{(-i)} < \lambda_{(-i)p-1}$ to be a lower bound is $\delta_{(-i)} = \lambda_{p-1}(\mathbf{\Psi}_{(-i)})$. When $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) = \lambda_{(-i)p-1}$, this value is the supremum of the lower bound.*

Proof. When $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) < \lambda_{(-i)p-1}$, $\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1} = \mathbf{\Sigma}_{(-i)} - \lambda_{p-1}(\mathbf{\Psi}_{(-i)})$ is p.d. since $\lambda_{(-i)p-1} = \lambda_{p-1}(\mathbf{\Sigma}_{(-i)}) > \lambda_{p-1}(\mathbf{\Psi}_{(-i)})$ though $\mathbf{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1} = \mathbf{\Psi}_{(-i)} - \lambda_{p-1}(\mathbf{\Psi}_{(-i)})\mathbf{I}_{p-1}$ is p.s.d. or singular. When $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) = \lambda_{(-i)p-1} = \delta_{(-i)}$, this value cannot be used as a lower bound since $\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1}$ becomes singular. However, $\mathbf{\Sigma}_{(-i)} - (\delta_{(-i)} - \varepsilon_0)\mathbf{I}_{p-1}$ with ε_0 being arbitrarily small positive constant is p.d. and can be used as a lower bound. Q.E.D.

Of course, in practice, it is not easy to obtain $\lambda_{p-1}(\mathbf{\Psi}_{(-i)})$, since $\mathbf{\Psi}_{(-i)}$ is unknown.

3. A method of deriving the communalities in the one-factor model

To find $\lambda_{p-1}(\mathbf{\Psi}_{(-i)})$, when the one-factor model with $\mathbf{\Lambda}_{(-i)}$ being a column vector holds, we can use a property of

$$\mathbf{A}_i^{-1} = \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)} - \lambda_{p-1}(\mathbf{\Psi}_{(-i)})\mathbf{I}_{p-1} \equiv \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)}^{(\mathrm{S})} \equiv \mathbf{\Sigma}_{(-i)}^{(\mathrm{S})}$$

with $\mathbf{\Psi}_{(-i)}^{(\mathrm{S})} = \mathbf{\Psi}_{(-i)} - \lambda_{p-1}(\mathbf{\Psi}_{(-i)})\mathbf{I}_{p-1}$ being singular. Suppose that when $\lambda_{p-1}(\mathbf{\Psi}_{(-i)})$ is given by the *j*-th original variable(s) $(j \neq i)$ among the $p-1$ variables yielding $\mathbf{\Sigma}_{(-i)}^{(\mathrm{S})}$. Then, when the *j*-th original variable is partialed out in the covariance matrix $\mathbf{\Sigma}_{(-i)}^{(\mathrm{S})}$, the residual covariance matrix of $\mathbf{\Sigma}_{(-i)}^{(\mathrm{S})}$ for the remaining $p-2$ variables become diagonal, since the *j*-th original variable has no uniqueness in $\mathbf{\Sigma}_{(-i)}^{(\mathrm{S})}$ and becomes a scaled single common factor. When the eigenvalue $\lambda_{p-1}(\mathbf{\Psi}_{(-i)})$ has more-than-one multiplicity i.e., $m_{(-i)} > 1$, the residual covariance matrix for the remaining $p-1-m_{(i)}$ variables becomes diagonal.

When the one-factor model holds, the above finding gives the following results. Define the submatrices of $\mathbf{\Sigma}_{(-i)} = \begin{pmatrix} \sigma_{(-i)11} & \boldsymbol{\sigma}_{(-i)12}^{\mathrm{T}} \\ \boldsymbol{\sigma}_{(-i)21} & \mathbf{\Sigma}_{(-i)22} \end{pmatrix}$ with $\sigma_{(-i)11} = 1$, where the first variable is partialed out from the covariance matrix for the remaining $p-2$ variables. When the *j*-th $(j = 2,..., p-1)$ variable is partialed out, the $p-1$ variables in $\mathbf{\Sigma}_{(-i)}$ are reordered with the *j*-th variable moved to the first one, followed by the redefinition of $\mathbf{\Sigma}_{(-i)}$ as

$$\mathbf{\Sigma}_{(-i)}^{(j)} = \begin{pmatrix} \sigma_{(-i)11}^{(j)} & \boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \\ \boldsymbol{\sigma}_{(-i)21}^{(j)} & \mathbf{\Sigma}_{(-i)22}^{(j)} \end{pmatrix} \text{ with } \sigma_{(-i)11}^{(j)} = 1 \quad (j = 1,..., p-1).$$

Theorem 5 (the anti-ridge parameter for the communality in the one-factor model). *When the one-factor model holds for the* $p \times p$ *correlation matrix* $\mathbf{\Sigma}$ *and* $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) < \lambda_{(-i)p-1}$ *in* $\mathbf{\Sigma}_{(-i)}$, *the communality* h_i^2 *of the i-th variable is given by* $h_i^2 = \boldsymbol{\sigma}_i^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_i$, *where*

$$\delta_{(-i)} = 1 - \frac{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}\text{Off-diag}(\boldsymbol{\sigma}_{(-i)21}^{(j)}\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}})\boldsymbol{\sigma}_{(-i)21}^{(j)}}{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}\text{Off-diag}(\mathbf{\Sigma}_{(-i)22}^{(j)})\boldsymbol{\sigma}_{(-i)21}^{(j)}} \quad (i = 1,..., p; i \neq j);$$

$\text{Off-diag}(\mathbf{A}) \equiv \mathbf{A} - \text{Diag}(\mathbf{A})$; $\text{Diag}(\cdot)$ *is the diagonal matrix with the same diagonal elements of a matrix in parentheses; the variable* j *is chosen such that the right-hand side of the above equation becomes the smallest among the* $p-1$ *variables in* $\mathbf{\Sigma}_{(-i)}$; *all the off-diagonal elements of* $\mathbf{\Sigma}_{(-i)22}^{(j)} - \boldsymbol{\sigma}_{(-i)12}^{(j)}(1-\delta_{(-i)})^{-1}\boldsymbol{\sigma}_{(-i)21}^{(j)\mathrm{T}}$ $(i, j = 1,..., p;\ i \neq j)$ *including the* *matrices not giving the smallest right-hand side vanish. When* $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) = \lambda_{(-i)p-1}$, *the value less than* $\lambda_{p-1}(\mathbf{\Psi}_{(-i)}) = \lambda_{(-i)p-1}$ *by an arbitrarily small amount can be used for* $\delta_{(-i)}$ *yielding a lower bound with the difference* $h_i^2 - \boldsymbol{\sigma}_i^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^{-1}\boldsymbol{\sigma}_i (> 0)$ *being made as small as desired.*

Proof. Let

$$\mathbf{\Sigma}_{(-i)}^{(j)} - \delta_{(-i)}\mathbf{I}_{p-1} = \mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I}_p \equiv \begin{pmatrix} c_{11} & \mathbf{c}_{12}^{\mathrm{T}} \\ \mathbf{c}_{12} & \mathbf{C}_{22} \end{pmatrix}$$

for simplicity of notation as long as confusion does not occur including possible reordering of the rows and columns of $\mathbf{\Lambda}_{(-i)}$ and $\mathbf{\Psi}_{(-i)}$. When the *j*-th variable in $\mathbf{\Sigma}_{(-i)}$ is partialed out, the residual covariance matrix corresponding to $\mathbf{C}_{22}$ becomes $\mathbf{C}_{22}^{*} = \mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}}$.

Define the function

$$\begin{aligned} g_j(\delta_{(-i)}) &= (1/2)\mathrm{tr}\left[\left\{\text{Off-diag}\left(\mathbf{\Sigma}_{(-i)22}^{(j)} - \delta_{(-i)}\mathbf{I}_{p-2} - \boldsymbol{\sigma}_{(-i)21}^{(j)}(1-\delta_{(-i)})^{-1}\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}\right)\right\}^2\right] \\ &= (1/2)\mathrm{tr}\left[\left\{\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\right\}^2\right]. \end{aligned}$$

Due to the one-factor model, when a variable in the $p-1$ variables is partialed out, the residual covariance matrix $\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}}$ for the remaining $p-2$ variables becomes diagonal though $\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}}$ may not be p.d. Since $g_j(\delta_{(-i)}) \geq 0$, by continuity we have the necessary condition of $\delta_{(-i)}$ for the vanishing off-diagonal elements of $\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}}$ as

$$\begin{aligned} 0 = \frac{\partial g_j(\delta_{(-i)})}{\partial \delta_{(-i)}} &= \mathrm{tr}\left\{\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\frac{\partial}{\partial \delta_{(-i)}}\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\right\} \\ &= \mathrm{tr}\left\{\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\text{Off-diag}(-\mathbf{I}_{p-2} - \mathbf{c}_{12}c_{11}^{-2}\mathbf{c}_{12}^{\mathrm{T}})\right\} \\ &= -\mathrm{tr}\left\{\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\text{Off-diag}(\mathbf{c}_{12}c_{11}^{-2}\mathbf{c}_{12}^{\mathrm{T}})\right\} \\ &= -\mathbf{c}_{12}^{\mathrm{T}}\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\mathbf{c}_{12}c_{11}^{-2}. \end{aligned}$$

where $\mathrm{tr}\{\text{Off-diag}(\mathbf{A})\,\text{Off-diag}(\mathbf{a}\,\mathbf{a}^{\mathrm{T}})\} = \mathrm{tr}\{\text{Off-diag}(\mathbf{A})\,\mathbf{a}\,\mathbf{a}^{\mathrm{T}}\} = \mathbf{a}^{\mathrm{T}}\text{Off-diag}(\mathbf{A})\,\mathbf{a}$ is used. Noting that $c_{11} = 1 - \delta_{(-i)}$, $\mathbf{c}_{21} = \boldsymbol{\sigma}_{(-i)21}^{(j)}$, $\mathbf{c}_{12} = \mathbf{c}_{21}^{\mathrm{T}}$ and $\mathbf{C}_{22} = \mathbf{\Sigma}_{(-i)22}^{(j)} - \delta_{(-i)}\mathbf{I}_{p-2}$, we obtain

$$
\begin{aligned}
0 &= \partial g_j(\delta_{(-i)}) / \partial \delta_{(-i)} \\
&= -\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}\left\{ \boldsymbol{\Sigma}_{(-i)22}^{(j)} - \delta_{(-i)} \mathbf{I}_{p-2} - \boldsymbol{\sigma}_{(-i)21}^{(j)} (1-\delta_{(-i)})^{-1} \boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \right\} \boldsymbol{\sigma}_{(-i)21} (1-\delta_{(-i)})^{-2} \\
&= -\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}\left\{ \boldsymbol{\Sigma}_{(-i)22}^{(j)} - \boldsymbol{\sigma}_{(-i)21}^{(j)} (\sigma_{(-i)11}^{(j)} - \delta_{(-i)})^{-1} \boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \right\} \boldsymbol{\sigma}_{(-i)21} (1-\delta_{(-i)})^{-2},
\end{aligned}
$$

The value of $\delta_{(-i)}$ satisfying the above equation is

$$
\delta_{(-i)} = 1 - \frac{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\sigma}_{(-i)21}^{(j)} \boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}) \boldsymbol{\sigma}_{(-i)21}^{(j)}}{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\Sigma}_{(-i)22}^{(j)}) \boldsymbol{\sigma}_{(-i)21}^{(j)}}.
$$

Among the $p-1$ values of $\delta_{(-i)}$ when $j \neq i$, the smallest value can be chosen as the smallest uniqueness.

The remaining property to be derived is that the smallest uniqueness i.e., $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$ for the anti-ridge parameter $\delta_{(-i)}$ gives $h_i^2 = \boldsymbol{\sigma}_{(-i)}^{\mathrm{T}} \{\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1}\}^{-1} \boldsymbol{\sigma}_{(-i)}$. For the one-factor model, consider the case when $\delta_{(-i)}$ goes to $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$ from below but not reach $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$. Since the spectral decomposition $\boldsymbol{\Lambda}_{(-i)}^{\mathrm{T}} \boldsymbol{\Psi}_{(-i)}^{(\delta)-1} \boldsymbol{\Lambda}_{(i))} = \boldsymbol{\lambda}_{(-i)}^{\mathrm{T}} \boldsymbol{\Psi}_{(-i)}^{(\delta)-1} \boldsymbol{\lambda}_{(-i)} = \gamma_{i1}^{(\delta)}$ with $\boldsymbol{\Psi}_{(-i)}^{(\delta)} = \boldsymbol{\Psi}_{(-i)} - \delta_{(i)} \mathbf{I}_{p-1}$ becomes a scalar, the Guttman formula gives

$$
h_i^2 - \boldsymbol{\sigma}_i^{\mathrm{T}} (\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1})^{-1} \boldsymbol{\sigma}_i = \boldsymbol{\lambda}_i^{\mathrm{T}} \boldsymbol{\Gamma}_i^{(\delta)} \text{diag}\left(\frac{1}{1+\gamma_{i1}^{(\delta)}} \right) \boldsymbol{\Gamma}_i^{(\delta)\mathrm{T}} \boldsymbol{\lambda}_i = \boldsymbol{\lambda}_i^{\mathrm{T}} \frac{1}{1+\gamma_{i1}^{(\delta)}} \boldsymbol{\lambda}_i
$$

with $\boldsymbol{\Gamma}_i^{(\delta)} = 1$. When $\delta_{(-i)}$ goes to $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$, $\gamma_{i1}^{(\delta)}$ goes to ∞. The above equation shows that in this limiting case, $\boldsymbol{\sigma}_i^{\mathrm{T}} (\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1})^{-1} \boldsymbol{\sigma}_i$ goes to h_i^2. Though $\boldsymbol{\Psi}_{(-i)}^{(\delta)-1}$ with $\delta_{(-i)} = \lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$ cannot be taken in the Guttman formula because $\boldsymbol{\Psi}_{(-i)}^{(\delta)} = \boldsymbol{\Psi}_{(-i)} - \lambda_{p-1}(\boldsymbol{\Psi}_{(-i)}) \mathbf{I}_{p-1}$ is singular, $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}} \{\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1}\}^{-1} \boldsymbol{\sigma}_{(-i)}$ with $\delta_{(-i)} = \lambda_{p-1}(\boldsymbol{\Psi}_{(-i)})$ can be taken since $\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1}$ is p.d. due to the assumption $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)}) < \lambda_{(-i)p-1}$. That is, the limiting value $h_i^2 = \boldsymbol{\sigma}_i^{\mathrm{T}} (\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1})^{-1} \boldsymbol{\sigma}_i$ is attained.

When $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)}) = \lambda_{(-i)p-1}$, the value $\delta_{(-i)}$ less than $\lambda_{p-1}(\boldsymbol{\Psi}_{(-i)}) = \lambda_{(-i)p-1}$ by an arbitrarily small amount can be used, yielding a lower bound of h_i^2 with the difference $h_i^2 - \boldsymbol{\sigma}_i^{\mathrm{T}} (\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)} \mathbf{I}_{p-1})^{-1} \boldsymbol{\sigma}_i (> 0)$ made as small as desired. Q.E.D.

Remark 2. In Theorem 5, there are $p-1$ candidates of the optimal $\delta_{(-i)}$ for the *i*-th variable. Each of them becomes equal to a uniqueness in $\mathbf{\Sigma}_{(-i)}$ since even when $\delta_{(-i)} = \lambda_{p^*}(\mathbf{\Psi}_{(-i)})$ $(p^* = 1,...,p-2)$ and $\delta_{(-i)}$ goes to $\lambda_{p^*}(\mathbf{\Psi}_{(-i)})$, $\gamma_{i1}^{(\delta)}$ goes to ∞ as in $\delta_{(-i)} = \lambda_{p-1}(\mathbf{\Psi}_{(-i)})$. As addressed earlier, the optimal $\delta_{(-i)}$ can be obtained from the smallest $\delta_{(-i)}$ in the $p-1$ candidates as will be shown in a numerical illustration. The value of $\delta_{(-i)}$ for variable j $(j=1,...,p; j \neq i)$ in Theorem 5 is shown to be equal to ψ_j as follows. Note that $\boldsymbol{\sigma}_{(-i)21}^{(j)} = \lambda_{(j)}\boldsymbol{\lambda}_{(-i,-j)}$ with $h_j^2 = \lambda_{(j)}^2$, where $\lambda_{(j)}$ is the *j*-th element of $\mathbf{\Lambda} = \boldsymbol{\lambda}$ $(j \neq i)$ to avoid confusion with $\lambda_i(\cdot)$ defined earlier; $\boldsymbol{\lambda}_{(-i,-j)}$ is $\boldsymbol{\lambda}$ whose *i*- and *j*-th elements are deleted. Then, from Theorem 5,

$$\begin{aligned}\delta_{(-i)} &= 1 - \frac{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}\text{Off-diag}(\boldsymbol{\sigma}_{(-i)21}^{(j)}\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}})\boldsymbol{\sigma}_{(-i)21}^{(j)}}{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}\text{Off-diag}(\mathbf{\Sigma}_{(-i)22}^{(j)})\boldsymbol{\sigma}_{(-i)21}^{(j)}} \\ &= 1 - \frac{\lambda_{(j)}^4\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}}\left\{\boldsymbol{\lambda}_{(-i,-j)}\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}} - \text{Diag}(\boldsymbol{\lambda}_{(-i,-j)}\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}})\right\}\boldsymbol{\lambda}_{(-i,-j)}}{\lambda_{(j)}^2\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}}\{\boldsymbol{\lambda}_{(-i,-j)}\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}} - \text{Diag}(\boldsymbol{\lambda}_{(-i,-j)}\boldsymbol{\lambda}_{(-i,-j)}^{\mathrm{T}})\}\boldsymbol{\lambda}_{(-i,-j)}} \\ &= 1 - \lambda_{(j)}^2 = 1 - h_j^2 = \psi_j \quad (j=1,...,p;\ j \neq i).\end{aligned}$$

4. A method of communality estimation in the multi-factor model

In Section 2, it was argued that an extension of ARSMC_i i.e., $\boldsymbol{\sigma}_{(-i)}^{\mathrm{T}}(\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^{+}\boldsymbol{\sigma}_{(-i)}$ with $\lambda_{(-i)p-1}$ being the smallest eigenvalue of $\mathbf{\Sigma}_{(-i)}$ cannot be a lower bound of h_i^2 since the extension can be greater than h_i^2. Suppose that $\mathbf{\Sigma}$ is p.d. though the EFA model may or may not hold. Note that $\mathbf{\Sigma} - \lambda_p(\mathbf{\Sigma})\mathbf{I}_p \equiv \mathbf{\Sigma} - \lambda_p\mathbf{I}_p$ with λ_p being the smallest eigenvalue of $\mathbf{\Sigma}$ is p.s.d. of rank $p-1$ or smaller. Then, $\mathbf{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1}$ similar to $\mathbf{\Sigma} - \lambda_p\mathbf{I}_p$ in the extended ARSMC_i is p.s.d. of rank $p-2$ or smaller depending on the multiplicity $m_{(-i)}$ of $\lambda_{(-i)p-1}$. The following property is easily found.

Property 1 (the conjugate, maximum-rank or spherical FA model). Suppose that $\mathbf{\Sigma}$ is p.d., then the following FA model always holds even when the EFA model does not hold:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\mathrm{T}} + \lambda_p \mathbf{I}_p \equiv \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\mathrm{T}} + \boldsymbol{\Psi}^{(\lambda_p)},$$

where $\boldsymbol{\Lambda}^*$ is the $p \times (p - m_p)$ loading matrix of rank $p - m_p$ with m_p being the multiplicity of $\lambda_p = \lambda_p(\boldsymbol{\Sigma})$ and $\boldsymbol{\Psi}^{(\lambda_p)}$ of full rank has the common uniqueness $\lambda_p (0 < \lambda_p \le 1)$. In this case, the communalities are also the same as $1 - \lambda_p$. The case $\lambda_p = 1$ yielding $\boldsymbol{\Sigma} = \mathbf{I}_p$ with $m_p = p$ is included as a FA model with no common factors. Suppose that the EFA model $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}$ holds with k $(1 \le k < p)$ common factors. Then, it is found that we generally have a pair of the FA models satisfying

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\mathrm{T}} + \boldsymbol{\Psi}^{(\lambda_p)} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}.$$

The case of $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^*$ and consequently $\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(\lambda_p)}$ when $k = 1$ with $m_p = p - 1$ was mentioned earlier as the compound symmetric model. The model $\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\mathrm{T}} + \boldsymbol{\Psi}^{(\lambda_p)}$ may be called the conjugate, maximum-rank or spherical FA model.

Using the conjugate FA model defined in Property 1 for $\boldsymbol{\Sigma}_{(-i)} = \boldsymbol{\Lambda}^*_{(-i)}\boldsymbol{\Lambda}^{*\mathrm{T}}_{(-i)} + \boldsymbol{\Psi}^{(\lambda_{p-1})}$, the extended ARSMC_i given by $\boldsymbol{\sigma}^{\mathrm{T}}_{(-i)}(\boldsymbol{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^+ \boldsymbol{\sigma}_{(-i)} = \boldsymbol{\sigma}^{\mathrm{T}}_{(-i)}(\boldsymbol{\Lambda}^*_{(-i)}\boldsymbol{\Lambda}^{*\mathrm{T}}_{(-i)})^+ \boldsymbol{\sigma}_{(-i)}$ is called the Moore-Penrose ARSMC_i (MPSMC_i). Though $\boldsymbol{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1} < \boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1}$ $(0 \le \delta_{(-i)} < \lambda_{(-i)p-1})$ since

$$\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1} - (\boldsymbol{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1}) = (\lambda_{(-i)p-1} - \delta_{(-i)})\mathbf{I}_{p-1} > \mathbf{O}$$

by construction, the inequality

$$(\boldsymbol{\Sigma}_{(-i)} - \delta_{(-i)}\mathbf{I}_{p-1})^+ > (\boldsymbol{\Sigma}_{(-i)} - \lambda_{(-i)p-1}\mathbf{I}_{p-1})^{-1}$$

does not always hold since the necessary condition of the rank equality of the two matrices in the M-K theorem (see Theorem 1) is not satisfied. That is, MPSMC_i can be smaller than the ARSMC_i with $\delta_{(-i)}(0 \le \delta_{(-i)} < \lambda_{(-i)p-1})$ as will be illustrated in a numerical example. However, this property may be seen as an advantage since ARSMC_i when $\delta_{(-i)}$ is slightly smaller than $\lambda_{(-i)p-1}$ tends to become too large as will be illustrated later. The following gives another reassuring property of MPSMC_i.

Property 2 (A lower bound of MPSMC_i). Noting that the conjugate FA model associated with MPSMC_i is a FA model for p.d. $\mathbf{\Sigma}_{(-i)}$, we can use the Guttman formula as in the usual EFA, yielding

$$\mathrm{MPSMC}_i > \mathrm{SMC}_i \ \ (i = 1,..., p)\,,$$

which will be numerically illustrated in the next section. Note that the common communality of the conjugate FA model behind MPSMC_i is $1-\lambda_{p-1}(\mathbf{\Sigma}_{(-i)})$, where $\lambda_{p-1}(\mathbf{\Sigma}_{(-i)})$ is the smallest eigenvalue of $\mathbf{\Sigma}_{(i)}$.

5. Numerical illustrations

An artificial example of the one-factor model with $p = 6$ is given for illustration. Table 1 shows the correlation matrix and related values. The communalities have a relatively wide range i.e., from 0.04 to 0.49. It is found that $\lambda_p(\mathbf{\Psi}) = 0.51 < 0.57 = \lambda_p(\mathbf{\Sigma})$ as theory indicates. Table 2 gives $p(p-1)$ values of $\delta_{(-i)}$ minimizing $(1/2)\mathrm{tr}\left[\left\{\text{Off-diag}(\mathbf{C}_{22} - \mathbf{c}_{12}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}})\right\}^2\right]$ with $p-1$ values for each $\mathbf{\Sigma}_{(-i)}(i = 1,..., p)$. It is found a value of $\delta_{(-i)}$ is equal to the corresponding $\psi_j\,(j \neq i)$ when variable *j* is a dependent one in regression. Mostly, only the smallest value $\delta_{(-i)}$ among $p-2$ values of $\delta_{(-i)}$ is smaller than the smallest eigenvalue of $\mathbf{C}_{22}$ with an exceptional case when *i* = 4 and *j* = 2 giving $\lambda_{p-2}(\mathbf{C}_{22}) = 0.641 > 0.64 = \delta_{(-i)}$. As expected, $\mathbf{C}_{22}^*$ is p.d. only when $\delta_{(-i)}$ is the smallest among $p-2$ values. It is to be noted the smallest $\delta_{(-i)}$ can be the smallest or the second smallest i.e., 0.51 or 0.64 of the uniquenesses depending on *i* = 1 or *i* = 2,...,6, respectively since the variables are ordered with the ascending uniqueness.

Table 3 shows the computational values of $\mathrm{ARSMC}_i(\delta_{(-i)})$ using the values of $\delta_{(-i)}$ which are the smallest i.e., 0.51 or 0.64 in each $\mathbf{\Sigma}_{(-i)}$ of Table 2. It is found that they correctly give the corresponding communalities. Note also that the $\delta_{(-i)}$'s are smaller than the corresponding smallest eigenvalues of $\mathbf{\Sigma}_{(-i)}(i = 1,..., p)$ due to Theorem 3.

Table 4 gives an artificial correlation matrix used by Yanai and Ichikawa (1990, p. 409) when the 3-factor model holds with p = 6. The values of $\mathrm{ARSMC}_{(-i)}$ are shown as maximum lower bounds of h_i^2 when $\delta_{(-i)}$ is the smallest uniqueness in $\mathbf{\Sigma}_{(-i)}$ though in practice they are unknown. It is of interest to find that when i = 1, 2 and 6, the $\mathrm{ARSMC}_{(-i)}$'s are close to the corresponding h_i^2 's.

In Table 4, the lower bounds by Yanai and Ichikawa (Y-I for short; 1990, Table 1) are shown. The lower bounds were obtained by their Theorem 1 i.e., $h_i^2 \geq 1-\lambda_{p-i+1}(\mathbf{\Sigma})\ (i=1,\ldots,p)$ based on the B-B inequality (see our Theorem 3). Note that only the four lower bounds are presented by Y-I since the remaining two become lower than the SMC_i 's. For instance, when i = 5, $1-\lambda_{p-i+1}(\mathbf{\Sigma}) = 1-0.648 = 0.352 < 0.45 = \mathrm{SMC}_i$ and similarly when i = 6. Y-I also gave Theorem 2 i.e., $1-\lambda_p(\mathbf{\Sigma}) \geq \mathrm{SMC}_i\ (i=1,\ldots,p)$ and stated that "Theorem 2 implies that out of the p variables, at least one highest communality has a lower bound at least as good as using the SMC, although we do not know in general to which bound the proposed bound shous be assigned" (p. 406). In Table 4, it is found that the maximum lower bounds $\mathrm{ARSMC}_{(-i)}$ when i = 1,...,4 are greater than the corresponding Y-I lower bound, indicating that there is some room to improve their lower bounds. Note also that their Theorem 3 gave improved lower bounds of h_i^2 's when the one-factor model holds, while our Theorem 5 gives the exact values of h_i^2 in this case.

In Table 4, the values of MPSMC_i are also shown, where except for

$$\mathrm{MPSMC}_1 = 0.8030 > h_1^2 = 0.77 > \mathrm{ARSMC}_1 = 0.7675 \ ,$$

ARSMC_i and MPSMC_i are comparable. It is to be noted that ARSMC_i is not available in practice since ARSMC_i uses the unknown smallest uniqueness in Table 4 while MPSMC_i can always be obtained from $\mathbf{\Sigma}_{(-i)}$. As Property 2 shows, we find that $\mathrm{MPSMC}_i > \mathrm{SMC}_i (i=1,\ldots,p)$. Table 4 also gives $\overline{\mathrm{ARSMC}_i}$, which is the $\mathrm{ARSMC}_i(\delta_{(i)})$ when $\delta_{(i)} = 0.99\lambda_4(\mathbf{\Sigma}_{(-i)})$, a slightly smaller value than the smallest eigenvalue of $\mathbf{\Sigma}_{(-i)}$ retaining the p.d. property of $\mathbf{\Sigma}_{(-i)} - \delta_{(i)}\mathbf{I}_{(5)}$. All the values of

$\overline{\text{ARSMC}_i}$ are larger than the corresponding MPSMC_i's. Note that $\overline{\text{ARSMC}_2}$ is as large as 2.83. The values MPSMC_i $(i = 1, 2, 5, 6)$ are closer to the corresponding h_i^2's than $\overline{\text{ARSMC}_i}$ suggesting an advantage of MPSMC_i.

6. Discussions

(a) The method of minimizing the sum of the squared off-diagonal elements of the residual covariance matrix: In Remark 2, when the one-factor model holds, the following result was shown:

$$h_j^2 = \frac{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\sigma}_{(-i)21}^{(j)} \boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}}) \boldsymbol{\sigma}_{(-i)21}^{(j)}}{\boldsymbol{\sigma}_{(-i)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\Sigma}_{(-i)22}^{(j)}) \boldsymbol{\sigma}_{(-i)21}^{(j)}} (i, j = 1, \ldots, p;\ j \neq i),$$

which was used in the context of deriving the optimal anti-ridge parameter in each $\boldsymbol{\Sigma}_{(-i)} (i = 1, \ldots, p)$. We find that the result is seen as a special case of

$$h_j^2 = \frac{\boldsymbol{\sigma}_{(p^*)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\sigma}_{(p^*)21}^{(j)} \boldsymbol{\sigma}_{(p^*)12}^{(j)\mathrm{T}}) \boldsymbol{\sigma}_{(p^*)21}^{(j)}}{\boldsymbol{\sigma}_{(p^*)12}^{(j)\mathrm{T}} \text{Off-diag}(\boldsymbol{\Sigma}_{(p^*)22}^{(j)}) \boldsymbol{\sigma}_{(p^*)21}^{(j)}} (p^* \subseteq \{1, 2, \ldots, p\};\ 3 \leq \text{card}(p^*) \leq p; j \in p^*),$$

where p^* is a subset of $\{1, 2, \ldots, p\}$ whose cardinality or the number of the members denoted by $\text{card}(p^*)$ is larger than 2. The number of subsets p^*'s becomes $2^p - \{p(p-1)/2\} - p - 1$ due to the binomial expansion. In each subset, member *j* is chosen from p^*. The simplest case is given when *p* = 3, $\text{card}(p^*) = 3$ with three subsets $\{1, 2\}, \{1, 3\}$ and $\{2, 3\}$. Then, the above formula gives $\sigma_{12}\sigma_{13} / \sigma_{23} = h_1^2$, which is seen as a special case of Gibson (1963) and Ihara and Kano (1986). The "largest" case is given when $\text{card}(p^*) = p$ or when $\boldsymbol{\Sigma}$ is employed in place of $\boldsymbol{\Sigma}_{(-i)} (i = 1, \ldots, p)$ used earlier.

(b) The case of the B-B inequality when the FA model holds: For generality, using $\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}_{(-i)}$, Theorem 3 gives $\lambda_p(\boldsymbol{\Sigma}) \geq \lambda_p(\boldsymbol{\Psi})$, which was used by Y-I (1990). As addressed earlier, when $\boldsymbol{\Psi} = \psi \mathbf{I}_p$ with $\boldsymbol{\Sigma}$ showing compound symmetry, the equality $\lambda_p(\boldsymbol{\Sigma}) = \lambda_p(\boldsymbol{\Psi})$ ("the eigenvalue-uniqueness equality" or "the E-U equality" for short) holds with the multiplicity $m_p = p - 1$ for the smallest eigenvalue i.e., ψ of $\boldsymbol{\Sigma}$. Note that this is a case when the EFA model and the corresponding conjugate or spherical FA model defined in Property 1 coincide. Other cases with the E-U equality when the one-factor model holds are those with block diagonal correlation matrices, where a single block shows compound symmetry while the remaining block(s) with no common factor are simple diagonal element(s) with possible more-then-one multiplicity of the smallest eigen value. Note that the smallest eigenvalue is equal to the that of the single block showing compound symmetry or that of the remaining diagonal elements. The author conjectures that the E-U equality in the one-factor model with $p \geq 3$ holds only in these cases. Note that when the conjugate FA model satisfying the one-factor model is considered, the model is restricted to the compound symmetric one after possible reflection of observable variable(s), since the sums of the squared loadings are the same when the uniquenesses are the same for correlation matrices.

Consider the cases with k $(0 \leq k < p)$ common factors in a block diagonal correlation matrix. Suppose that the *i*-th $p_i \times p_i$ block $(i = 1,\ldots,B; p_1 + \ldots + p_B = p)$ has the conjugate FA model with k_i common factors $0 \leq k_i < p_i$. It is found that when $k_1 + \ldots + k_B = k$, the correlation matrix has the E-U equality. Note that the smallest possibly multiple eigenvalue $\lambda_p(\mathbf{\Sigma})$ is the smallest among the *B* smallest ones in the diagonal blocks. The cases of the one-factor model given earlier are special cases with $k_1 = 1$ and $k_i = 0\,(i = 2,\ldots,B)$. When $k = 0$, we have $\mathbf{\Sigma} = \mathbf{I}_p$ without common factor yielding $\lambda_i(\mathbf{\Sigma}) = \lambda_i(\mathbf{\Psi}) = 1$ $(i = 1,\ldots, p)$ as well as the E-U equality. It is unknown whether other cases in the *k*-factor model have the E-U equality.

(c) The problem to obtain the maximum anti-ridge parameter or the smallest uniqueness in the multi-factor cases: While the optimal anti-ridge parameters to yield exact h_i^2's when the one-factor model holds were obtained, the corresponding counterpart in the general multi-factor model was not derived. The optimal value is given when $\mathbf{\Psi}_{(-i)} - \delta_{(-i)}\mathbf{I} = \mathbf{\Psi}_{(-i)}^{(\mathrm{S})}$ becomes p.s.d. or singular with $\delta_{(-i)} < \lambda_{p-1}(\mathbf{\Sigma}_{(-i)})$. A property of the model $\mathbf{\Lambda}_{(-i)}\mathbf{\Lambda}_{(-i)}^{\mathrm{T}} + \mathbf{\Psi}_{(-i)}^{(\mathrm{S})}$, which may be called the p.s.d. factor model, is that at least one variable has no unique factor. For the derivation of the optimal $\delta_{(-i)}$, some semidefinite mathematical programming may be required as mentioned in the introductory section.

(d) An elementary proof of Y-I's (1990) Theorem 2 using the conjugate FA model: Recall that the common communality of the conjugate FA model for $\mathbf{\Sigma} = \mathbf{\Lambda}^{*}\mathbf{\Lambda}^{*\mathrm{T}} + \lambda_p \mathbf{I}_p$ defined in Property 1 is $1-\lambda_p$ with $\lambda_p = \lambda_p(\mathbf{\Sigma})$. As addressed in Section 5, Y-I's (1990) Theorem 2 gave $1-\lambda_p \geq \mathrm{SMC}_i \ (i=1,\ldots,p)$, which was obtained using a corollary of the Poincaré separation theorem (Magnus & Neudecker, 2007, Chapter 11, Theorems 10 and 12; for the separation theorem and its applications, see Rao, 1979, Takane & Shibayama, 1991, Section 2.2 and Appendix A, Rolle, 2000, and Yanai, Takeuchi & Takane, 2011, Theorem 5.8). An alternative elementary short proof of this result is given as follows. Noting that the conjugate FA model with the common communality $1-\lambda_p$ satisfies $1-\lambda_p \geq \mathrm{SMC}_i (i=1,\ldots,p)$ due to the property of SMC_i as a lower bound of the common communality for each variable, the required result follows.

References

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman, (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V (pp. 111-150). Berkeley: University of California Press.

Baksalary, J. K., Nordström, K., & Styan, G. P. (1990). Löwner-ordering antitonicity of generalized inverse of Hermitian matrices. *Linear Algebra and its Applications*, *127*, 171-182.

Beckenbach, E., & Bellman, R. (1965). *Inequalities* (2nd revised printing). New York: Springer.

Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*, 343-357.

Bertsimas, D., Copenhaver, M. S., & Mazumder, R. (2017). Certifiably optimal low rank factor analysis. *The Journal of Machine Learning Research*, *18* (1), 907-959.

Bertsimas, D., Cory-Wright, R., & Pauphilet, J. (2023). A new perspective on low-rank optimization. *Mathematical Programming, 202*, 47-92.

Browne, M. W. (1974). Generalized least-squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*, 1-24. Reprinted in D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models* (pp. 205-226). Amsterdam: North Holland, 1977.

Gibson, W. A. (1963). On the symmetric treatment of an asymmetric approach to factor analysis. *Psychometrika*, *28* (4), 423-426.

Guttman, L. (1946). Enlargement methods for computing the inverse matrix. *The Annals of Mathematical Statistics*, *17* (3), 336–343.

Guttman, L. (1958). To what extent can communalities reduce rank? *Psychometrika*, *23* (4), 297-308.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, *51*, 563-566.

Kano, Y. (1989). A new estimation procedure using *g*-inverse matrix in factor analysis. *Mathematica Japonica*, *34*, 43-52.

Kano, Y. (1990). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. *Psychometrika*, *55*, 277-291.

Magnus, J. R., & Neudecker, H. (2007). *Matrix differential calculus with applications in statistics and econometrics* (3rd ed.). New York: Wiley.

Milliken, G. A., & Akdeniz, F. (1977). A theorem on the difference of the generalized inverses of two nonnegative matrices. *Communications in Statistics-Theory and Methods*, *6* (1), 73-79.

Neudecker, H. (1989). A new proof of the Milliken-Akdeniz theorem. *Qüestiió*, *13* (1, 2, 3), 81-82.

Ramsey, P. H., & Gibson, W. A. (2006). Improved communality estimation in factor analysis. *Journal of Statistical Computation and Simulation*, *76* (2), 93-101.

Rao, C. R. (1979). Separation theorems for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis, 9* (3), 362-377.

Roff, M. (1936). Some properties of the communality in multiple factor theory. *Psychometrika*, *1*, 1-6.

Rolle, J. D. (2000). *Some applications of the Poincaré separation theorem in data analysis*. Unpublished report, Université de Genève. Available at: http://archiveouverte.unige.ch/unige:5852.

Shapiro, A., & ten Berge, J. M. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika*, *67*, 79-94.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, *56* (1), 97-120.

ten Berge, J. M., & Kiers, H. A. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309-315.

ten Berge, J. M. F., Snijders, T. A. B., & Zegers, F. E. (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 201-213.

Tunçel, L., Vavasis, S. A., & Xu, J. (2023). Computational complexity of decomposing a symmetric matrix as a sum of positive semidefinite and diagonal matrices. *Preprint arXiv:2209.05678v3*.

Wang, S., & Dai, M. (2010, October). The generalized inverse inequalities for symmetric nonnegative definite matrices. In *2010 International Conference on Artificial Intelligence and Computational Intelligence* (Vol. 3, pp. 547-550). Los Alamitos, CA: IEEE Computer Society.

Yanai, H. (1990). Some generalized forms of least squares *g*-inverses, minimum norm *g*-inverses, and Moore-Penrose inverse matrices. *Computational Statistics and Data Analysis, 10*, 251–260.

Yanai, H., & Ichikawa, M. (1990). New lower and upper bounds for communality in factor analysis. *Psychometrika*, *55*, 405-409.

Yanai, H., & Takane, Y. (2007). Matrix methods and their applications to factor analysis. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 345-366). Amsterdam: North-Holland.

Yanai, H., Takeuchi, K., & Takane, Y. (2011). *Projection matrices, generalized inverse matrices, and singular value decomposition*. New York: Springer.

Table 1. An artificial correlation matrix when the one-factor model holds

$\boldsymbol{\Sigma}$						$\boldsymbol{\lambda}$	h_i^2	$\boldsymbol{\psi}$	Eigen-values
1.00						.7	.49	.51	2.07
.42	1.00		symmetric			.6	.36	.64	.95
.35	.30	1.00				.5	.25	.75	.89
.28	.24	.20	1.00			.4	.16	.84	.81
.21	.18	.15	.12	1.00		.3	.09	.91	.70
.14	.12	.10	.08	.06	1.00	.2	.04	.96	.57

Note. $\boldsymbol{\Psi} = \mathrm{diag}(\boldsymbol{\psi})$.

Table 2. The values of $\delta_{(-i)}$ giving the zero off-diagonal elements of the $(p-2)\times(p-2)$ residual covariance matrix $\mathbf{C}_{22}^{*}=\mathbf{C}_{22}-\mathbf{c}_{21}c_{11}^{-1}\mathbf{c}_{12}^{\mathrm{T}}$ in the one-factor model

Variable i	1					2					3				
Variable j in $\mathbf{\Sigma}_{(-i)}$	2	3	4	5	6	1	3	4	5	6	1	2	4	5	6
Eigenvalues of $\mathbf{C}_{22}$	1.37	1.42	1.48	1.55	1.61	1.37	1.47	1.54	1.62	1.68	1.42	1.47	1.62	1.69	1.75
	.95	.95	.95	.95	.89	.95	.95	.95	.95	.89	.95	.95	.95	.95	.89
	.89	.88	.87	.81	.81	.89	.88	.87	.81	.80	.88	.88	.86	.79	.78
	.79	.745	.70	.69	.69	.79	.69	.64	.63	.63	.75	.69	.58	.58	.58
$\delta_{(-i)}$	.64	.75	.84	.91	.96	.51	.75	.84	.91	.96	.51	.64	.84	.91	.96
Eigenvalues of $\mathbf{C}_{22}^{*}$	.11	-.11	-.20	-.27	-.32	.24	-.24	-.33	-.40	-.45	.13	-.13	-.33	-.40	-.45
	.20	.09	-.09	-.16	-.21	.33	.09	-.09	-.16	-.21	.33	.20	-.20	-.27	-.32
	.27	.16	.07	-.07	-.12	.40	.16	.07	-.07	-.12	.40	.27	.07	-.07	-.12
	.32	.21	.12	.05	-.05	.45	.21	.12	.05	-.05	.45	.32	.12	.05	-.05
Variable i	4					5					6				
Variable j in $\mathbf{\Sigma}_{(-i)}$	1	2	3	5	6	1	2	3	4	6	1	2	3	4	5
Eigenvalues of $\mathbf{C}_{22}$	1.48	1.54	1.62	1.77	1.83	1.55	1.62	1.69	1.77	1.91	1.61	1.68	1.75	1.83	1.91
	.95	.95	.95	.94	.89	.95	.95	.95	.94	.82	.89	.89	.89	.89	.82
	.87	.87	.86	.71	.71	.81	.81	.79	.71	.71	.81	.80	.78	.71	.71
	.70	.641	.58	.57	.57	.69	.63	.58	.57	.57	69	.63	.58	.57	.57
$\delta_{(-i)}$	.51	.64	.75	.91	.96	.51	.64	.75	.84	.96	.51	.64	.75	.84	.91
Eigenvalues of $\mathbf{C}_{22}^{*}$	.13	-.13	-.24	-.40	-.45	.13	-.13	-.24	-.33	-.45	.13	-.13	-.24	-.33	-.40
	.24	.11	-.11	-.27	-.32	.24	.11	-.11	-.20	-.32	.24	.11	-.11	-.20	-.27
	.40	.27	.16	-.16	-.21	.33	.20	.09	-.09	-.21	.33	.20	.09	-.09	-.16
	.45	.32	.21	.05	-.05	.45	.32	.21	.12	-.12	.40	.27	.16	.07	-.07

Table 3. The values of $\mathrm{ARSMC}_i(\delta_{(-i)})$ when $\delta_{(-i)} = \min(\boldsymbol{\psi}_{(-i)})$ and associated results in the one-factor model

Variable i	1	2	3	4	5	6
Eigenvalues of $\boldsymbol{\Sigma}_{(-i)}$	1.66	1.73	1.81	1.89	1.96	2.02
	.9519	.9519	.9518	.9514	.9500	.90
	.89	.89	.89	.88	.81	.81
	.81	.80	.78	.71	.71	.70
	.69	.63	.58	.5712	.5697	.5690
$\delta_{(-i)} = \min(\boldsymbol{\psi}_{(-i)})$	.64	.51	*	*	*	*
$\mathrm{ARSMC}_i(\delta_{(-i)}) = h_i^2$	.49	.36	.25	.16	.09	.04
SMC_i	.27	.22	.16	.11	.06	.03

Note. $\boldsymbol{\Psi}_{(-i)} = \mathrm{diag}(\boldsymbol{\psi}_{(-i)})$. The asterisk indicates the same value as $\delta_{(-2)} = .51$.

Table 4. Yanai and Ichikawa's (1990) artificial correlation matrix and associated results

$\boldsymbol{\Sigma}$						$\boldsymbol{\Lambda}$			Eigen-values	$\boldsymbol{\psi}$	h_i^2	ARSMC_i	Y-I LB	SMC_i	MPSMC_i	$\overline{\text{ARSMC}_i}$
1.00						.6	.5	.4	3.810	.23	.77	.7675	.735	.654	.803	.818
.72	1.00				symmetric	.6	.4	.4	.648	.32	.68	.6758	.645	.573	.632	2.83
.62	.56	1.00				.2	.6	.5	.490	.35	.65	.583	.568	.518	.593	.652
.61	.56	.64	1.00			.2	.5	.6	.432	.35	.65	.568	.510	.498	.581	.626
.62	.56	.54	.50	1.00		.4	.6	.2	.355	.44	.56	.510		.453	.497	.815
.58	.54	.44	.42	.48	1.00	.5	.4	.2	.265	.55	.45	.444		.384	.434	.713

Note. $\text{ARSMC}_i = \text{ARSMC}_i(\delta_{(-i)})$ when $\delta_{(-i)}$ is the smallest uniqueness in $\boldsymbol{\Sigma}_{(-i)}$. Y-I LB = Yanai and Ichikawa's (1990) lower bound. $\overline{\text{ARSMC}_i} = \text{ARSMC}_i(\delta_{(-i)})$ when $\delta_{(-i)} = .99\lambda_5(\boldsymbol{\Sigma}_{(-i)})$ with $\lambda_5(\boldsymbol{\Sigma}_{(-i)})$ being the smallest eigenvalue of $\boldsymbol{\Sigma}_{(-i)}$.