J. Japan Statist. Soc.
Vol. 37 No. 2 2007 239–251

DISTRIBUTIONS AND THE BOOTSTRAP METHOD OF SOME STATISTICS IN PRINCIPAL CANONICAL CORRELATION ANALYSIS

Takakazu Sugiyama*, Toru Ogura*, Fumitake Sakaori** and Tomoya Yamada***

We investigate the canonical correlation of the principal components from two populations, and attain the limiting distribution using the perturbation expansion of the canonical correlation estimate. We discuss the numerical accuracy of the limiting distribution.

Key words and phrases: Canonical correlation analysis, perturbation method, principal component.

1. Introduction

We consider two sets of variables with a joint distribution and analyze the canonical correlations between the variables in the two sets. One of the analyses used is the canonical correlation analysis, which finds linear combinations of variables in the sets that have the maximum correlation, and these linear combinations are the first coordinates in new systems. Then, a second linear combination in each set is obtained such that the linear combination is uncorrelated with the first linear combination. The procedure is continued until two new coordinate systems are specified completely. This theory was developed by Hotelling (1935, 1936).

In this paper, we first determine the principal components of the two sets and then calculate the canonical correlation between the two principal components. Principal components analysis is a procedure used for analyzing multivariate data that transforms the original variables into new ones that are uncorrelated and account for decreasing proportions of the variance in the data. This analysis attempts to characterize or explain the variability in a vector variable by replacing it with a new variable with fewer components with large variance.

We know that the interpretation of principal components is easier than the canonical variate. Therefore, comparing canonical correlation analysis with principal component analysis, we can say that the canonical correlations of two principal components are more useful for understanding the relationships of the given data sets. This paper derives the limiting distribution of the canonical correlation

Received May 8, 2006. Revised June 17, 2006. Accepted September 3, 2006.

*Department of Mathematics, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.
**Faculty of Sociology, Rikkyo University, 3-34-1, Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan.
***Faculty of Economics, Sapporo Gakuin University, 11-banchi Bunkyodai, Ebetsu, Hokkaido 069-8555, Japan.

of the principal components from two populations. In Section 2, we derive the canonical correlation of the principal components from two populations. In Section 3, we find the limiting distribution of the canonical correlation. Finally, we compare the results of the canonical correlation using an example, a simulation study, and bootstrapping.

2. Canonical correlation of the principal components from two populations

2.1. Principal component canonical correlation in the population

Suppose the random vector $\boldsymbol{Z}$ of $p+q$ components has covariance matrix $\boldsymbol{\Sigma}$, which is assumed to be positive definite. Since we are only interested in the variance and covariance in this section, we assume $E(\boldsymbol{Z}) = \boldsymbol{0}$ without loss of generality. We partition $\boldsymbol{Z}$ into two subvectors of p and q components, $\boldsymbol{X}$ and $\boldsymbol{Y}$, such that:

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix}. \tag{2.1}$$

Similarly, the covariance matrix is partitioned into p and q rows and columns,

$$\mathrm{Cov}\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}, \tag{2.2}$$

where $\boldsymbol{\Sigma}_{xx}$ is $p\times p$, $\boldsymbol{\Sigma}_{xy}$ is $p\times q$, $\boldsymbol{\Sigma}_{yx}$ is $q\times p$, and $\boldsymbol{\Sigma}_{yy}$ is $q\times q$. Let $\lambda_{1x} \geq \cdots \geq \lambda_{px}$ be the ordered latent roots of $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\gamma}_{1x}, \ldots, \boldsymbol{\gamma}_{px}$ be the corresponding latent vectors; similarly, let $\lambda_{1y} \geq \cdots \geq \lambda_{qy}$ be the ordered latent roots of $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\gamma}_{1y}, \ldots, \boldsymbol{\gamma}_{qy}$ be the corresponding latent vectors. For $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$, we may decompose this as:

$$\boldsymbol{\Gamma}'_{x,p}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}_{x,p} = \boldsymbol{\Lambda}_{x,p}, \qquad \boldsymbol{\Gamma}'_{y,q}\boldsymbol{\Sigma}_{yy}\boldsymbol{\Gamma}_{y,q} = \boldsymbol{\Lambda}_{y,q}, \tag{2.3}$$

where $\boldsymbol{\Lambda}_{x,p} = \mathrm{diag}(\lambda_{1x}, \ldots, \lambda_{px})$ and $\boldsymbol{\Lambda}_{y,q} = \mathrm{diag}(\lambda_{1y}, \ldots, \lambda_{qy})$ are the diagonal matrices, and the orthogonal matrix is denoted $\boldsymbol{\Gamma}_{x,p} = (\boldsymbol{\gamma}_{1x}, \ldots, \boldsymbol{\gamma}_{px})$ and $\boldsymbol{\Gamma}_{y,q} = (\boldsymbol{\gamma}_{1y}, \ldots, \boldsymbol{\gamma}_{qy})$. Then, we obtain the i-th principal component of $\boldsymbol{X}$, $U_i = \boldsymbol{\gamma}'_{ix}\boldsymbol{X}$ and the j-th principal component of $\boldsymbol{Y}$, $V_j = \boldsymbol{\gamma}'_{jy}\boldsymbol{Y}$. Furthermore, we obtain $\mathrm{Var}(U_i) = \lambda_{ix}$, $\mathrm{Cov}(U_i, U_j) = 0$ $(i \neq j)$, $\mathrm{Var}(V_i) = \lambda_{iy}$, $\mathrm{Cov}(V_i, V_j) = 0$ $(i \neq j)$. Let

$$\boldsymbol{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_{p_1} \end{pmatrix}, \qquad \boldsymbol{V} = \begin{pmatrix} V_1 \\ \vdots \\ V_{q_1} \end{pmatrix}, \tag{2.4}$$

where $p_1 \leq p$, $q_1 \leq q$, $p_1 \leq q_1$. The covariance matrix of $(\boldsymbol{U}, \boldsymbol{V})'$ is:

$$\begin{aligned} \mathrm{Cov}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uv} \\ \boldsymbol{\Sigma}_{vu} & \boldsymbol{\Sigma}_{vv} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}'_x\boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}_x & \boldsymbol{\Gamma}'_x\boldsymbol{\Sigma}_{xy}\boldsymbol{\Gamma}_y \\ \boldsymbol{\Gamma}'_y\boldsymbol{\Sigma}_{yx}\boldsymbol{\Gamma}_x & \boldsymbol{\Gamma}'_y\boldsymbol{\Sigma}_{yy}\boldsymbol{\Gamma}_y \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Lambda}_x & \boldsymbol{\Gamma}'_x\boldsymbol{\Sigma}_{xy}\boldsymbol{\Gamma}_y \\ \boldsymbol{\Gamma}'_y\boldsymbol{\Sigma}_{yx}\boldsymbol{\Gamma}_x & \boldsymbol{\Lambda}_y \end{pmatrix}, \end{aligned} \tag{2.5}$$

where we denote the matrices as $\mathbf{\Gamma}_x = (\boldsymbol{\gamma}_{1x}, \dots, \boldsymbol{\gamma}_{p_1x})$, $\mathbf{\Gamma}_y = (\boldsymbol{\gamma}_{1y}, \dots, \boldsymbol{\gamma}_{q_1y})$, $\mathbf{\Lambda}_x = \mathrm{diag}(\lambda_{1x}, \dots, \lambda_{p_1x})$, and $\mathbf{\Lambda}_y = \mathrm{diag}(\lambda_{1y}, \dots, \lambda_{q_1y})$. The quantities of the canonical correlation coefficient $\nu_1^{2*} \geq \cdots \geq \nu_{p_1}^{2*} \geq 0$ satisfy:

$$|\mathbf{\Sigma}_{uv}\mathbf{\Sigma}_{vv}^{-1}\mathbf{\Sigma}_{vu} - \nu^{2*}\mathbf{\Sigma}_{uu}| = 0, \tag{2.6}$$

or

$$|\mathbf{\Sigma}_{vu}\mathbf{\Sigma}_{uu}^{-1}\mathbf{\Sigma}_{uv} - \nu^{2*}\mathbf{\Sigma}_{vv}| = 0, \tag{2.7}$$

and $\boldsymbol{\alpha}_i = (\alpha_{1i} \cdots \alpha_{p_1i})'$ and $\boldsymbol{\beta}_i = (\beta_{1i} \cdots \beta_{q_1i})'$ satisfy

$$\mathbf{\Sigma}_{uv}\mathbf{\Sigma}_{vv}^{-1}\mathbf{\Sigma}_{vu}\boldsymbol{\alpha}_i = \nu_i^{2*}\mathbf{\Sigma}_{uu}\boldsymbol{\alpha}_i, \qquad \boldsymbol{\alpha}_i\mathbf{\Sigma}_{uu}\boldsymbol{\alpha}_j = \delta_{ij}, \tag{2.8}$$

$$\mathbf{\Sigma}_{vu}\mathbf{\Sigma}_{uu}^{-1}\mathbf{\Sigma}_{uv}\boldsymbol{\beta}_i = \nu_i^{2*}\mathbf{\Sigma}_{vv}\boldsymbol{\beta}_i, \qquad \boldsymbol{\beta}_i'\mathbf{\Sigma}_{vv}\boldsymbol{\beta}_j = \delta_{ij}, \tag{2.9}$$

where δ_{ij} is a Kronecker's delta. Substituting (2.6) into (2.5) gives:

$$\begin{aligned}
&|\mathbf{\Sigma}_{uv}\mathbf{\Sigma}_{vv}^{-1}\mathbf{\Sigma}_{vu} - \nu^{2*}\mathbf{\Sigma}_{uu}| = 0 \\
&|\mathbf{\Sigma}_{uu}^{-1/2}\mathbf{\Sigma}_{uv}\mathbf{\Sigma}_{vv}^{-1}\mathbf{\Sigma}_{vu}\mathbf{\Sigma}_{uu}^{-1/2} - \nu^{2*}\boldsymbol{I}| = 0 \\
&|\mathbf{\Lambda}_x^{-1/2}\mathbf{\Gamma}_x'\mathbf{\Sigma}_{xy}\mathbf{\Gamma}_y\mathbf{\Lambda}_y^{-1}\mathbf{\Gamma}_y'\mathbf{\Sigma}_{yx}\mathbf{\Gamma}_x\mathbf{\Lambda}_x^{-1/2} - \nu^{2*}\boldsymbol{I}| = 0,
\end{aligned} \tag{2.10}$$

and therefore

$$\boldsymbol{O}'\mathbf{\Lambda}_x^{-1/2}\mathbf{\Gamma}_x'\mathbf{\Sigma}_{xy}\mathbf{\Gamma}_y\mathbf{\Lambda}_y^{-1}\mathbf{\Gamma}_y'\mathbf{\Sigma}_{yx}\mathbf{\Gamma}_x\mathbf{\Lambda}_x^{-1/2}\boldsymbol{O} = \begin{pmatrix} \nu_1^{2*} & 0 & \cdots & 0 \\ 0 & \nu_2^{2*} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \nu_{p_1}^{2*} \end{pmatrix}, \tag{2.11}$$

where $\boldsymbol{O}$ is the orthogonalization matrix. ρ^{2*} is defined:

$$\rho^{2*} = \nu_1^{2*} + \cdots + \nu_{p_1}^{2*}. \tag{2.12}$$

Then we get:

$$\begin{aligned}
\rho^{2*} &= \mathrm{tr}(\mathbf{\Lambda}_x^{-1/2}\mathbf{\Gamma}_x'\mathbf{\Sigma}_{xy}\mathbf{\Gamma}_y\mathbf{\Lambda}_y^{-1}\mathbf{\Gamma}_y'\mathbf{\Sigma}_{yx}\mathbf{\Gamma}_x\mathbf{\Lambda}_x^{-1/2}) \\
&= \sum_{i=1}^{p_1}\sum_{j=1}^{q_1}\frac{(\boldsymbol{\gamma}_{ix}'\mathbf{\Sigma}_{xy}\boldsymbol{\gamma}_{jy})^2}{\lambda_{ix}\lambda_{jy}}.
\end{aligned} \tag{2.13}$$

Above ρ^{2*} means the total sum of the canonical correlations based on p_1 principal components of $\boldsymbol{X}$ and q_1 principal components of $\boldsymbol{Y}$.

2.2. Estimation of the principal canonical component correlation

Let $\boldsymbol{z}_1, \dots, \boldsymbol{z}_N$ be N observations from $N(\mu, \Sigma)$ and $\boldsymbol{z}_\iota$ be partitioned into two subvectors of p and q components, respectively,

$$\boldsymbol{z}_\iota = \begin{pmatrix} \boldsymbol{x}_\iota \\ \boldsymbol{y}_\iota \end{pmatrix}, \qquad \iota = 1, \dots, N.$$

Let the sample covariance matrix $\boldsymbol{S}$ be written as:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{xx} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{S}_{yy} \end{pmatrix} = \frac{1}{n}\sum_{\iota=1}^{N}(\boldsymbol{z}_\iota - \bar{\boldsymbol{z}})(\boldsymbol{z}_\iota - \bar{\boldsymbol{z}})' = \frac{1}{n}\begin{pmatrix} \sum_{\iota=1}^{N}(\boldsymbol{x}_\iota - \bar{\boldsymbol{x}})(\boldsymbol{x}_\iota - \bar{\boldsymbol{x}})' & \sum_{\iota=1}^{N}(\boldsymbol{x}_\iota - \bar{\boldsymbol{x}})(\boldsymbol{y}_\iota - \bar{\boldsymbol{y}})' \\ \sum_{\iota=1}^{N}(\boldsymbol{y}_\iota - \bar{\boldsymbol{y}})(\boldsymbol{x}_\iota - \bar{\boldsymbol{x}})' & \sum_{\iota=1}^{N}(\boldsymbol{y}_\iota - \bar{\boldsymbol{y}})(\boldsymbol{y}_\iota - \bar{\boldsymbol{y}})' \end{pmatrix}, \tag{2.14}$$

where $n = N - 1$. Let $l_{1x} \geq \cdots \geq l_{px}$ be the ordered latent roots of $\boldsymbol{S}_{xx}$, and $\boldsymbol{h}_{1x}, \ldots, \boldsymbol{h}_{px}$ be the corresponding latent vectors; similarly, let $l_{1y} \geq \cdots \geq l_{qy}$ be the ordered latent roots of $\boldsymbol{S}_{yy}$, and $\boldsymbol{h}_{1y}, \ldots, \boldsymbol{h}_{qy}$ be the corresponding latent vectors. For $\boldsymbol{S}_{xx}$ and $\boldsymbol{S}_{yy}$, we may decompose this as:

$$\boldsymbol{H}'_{x,p}\boldsymbol{S}_{xx}\boldsymbol{H}_{x,p} = \boldsymbol{D}_{x,p}, \qquad \boldsymbol{H}'_{y,q}\boldsymbol{S}_{yy}\boldsymbol{H}_{y,q} = \boldsymbol{D}_{y,q}, \tag{2.15}$$

where $\boldsymbol{D}_{x,p} = \mathrm{diag}(l_{1x}, \ldots, l_{px})$ and $\boldsymbol{D}_{y,q} = \mathrm{diag}(l_{1y}, \ldots, l_{qy})$ are the diagonal matrices, and the orthogonal matrix is denoted by $\boldsymbol{H}_{x,p} = (\boldsymbol{h}_{1x}, \ldots, \boldsymbol{h}_{px})$ and $\boldsymbol{H}_{y,q} = (\boldsymbol{h}_{1y}, \ldots, \boldsymbol{h}_{qy})$. Let

$$\begin{aligned} \boldsymbol{S}_{uu} &= \boldsymbol{H}'_x\boldsymbol{S}_{xx}\boldsymbol{H}_x, & \boldsymbol{S}_{uv} &= \boldsymbol{H}'_x\boldsymbol{S}_{xy}\boldsymbol{H}_y, \\ \boldsymbol{S}_{vu} &= \boldsymbol{H}'_x\boldsymbol{S}_{xy}\boldsymbol{H}_y, & \boldsymbol{S}_{vv} &= \boldsymbol{H}'_y\boldsymbol{S}_{yy}\boldsymbol{H}_y, \end{aligned}$$

where we denote the orthogonal matrix as $\boldsymbol{H}_x = (\boldsymbol{h}_{1x}, \ldots, \boldsymbol{h}_{p_1x})$ and $\boldsymbol{H}_y = (\boldsymbol{h}_{1y}, \ldots, \boldsymbol{h}_{q_1y})$ and $p_1 \leq p$, $q_1 \leq q$ and $p_1 \leq q_1$. The estimate quantities of the canonical correlation coefficient $f_1^{2*} \geq \cdots \geq f_{p_1}^{2*} \geq 0$ satisfy:

$$|\boldsymbol{S}_{uv}\boldsymbol{S}_{vv}^{-1}\boldsymbol{S}_{vu} - f^{2*}\boldsymbol{S}_{uu}| = 0, \tag{2.16}$$

or

$$|\boldsymbol{S}_{vu}\boldsymbol{S}_{uu}^{-1}\boldsymbol{S}_{uv} - f^{2*}\boldsymbol{S}_{vv}| = 0, \tag{2.17}$$

and $\boldsymbol{a}_i = (a_{1i}, \ldots, a_{p_1i})'$ and $\boldsymbol{b}_i = (b_{1i}, \ldots, b_{q_{1i}})'$ satisfy

$$\boldsymbol{S}_{uv}\boldsymbol{S}_{vv}^{-1}\boldsymbol{S}_{vu}\boldsymbol{a}_i = f_i^{2*}\boldsymbol{S}_{uu}\boldsymbol{a}_i, \qquad \boldsymbol{a}'_i\boldsymbol{S}_{uu}\boldsymbol{a}_j = \delta_{ij}, \tag{2.18}$$

$$\boldsymbol{S}_{vu}\boldsymbol{S}_{uu}^{-1}\boldsymbol{S}_{uv}\boldsymbol{b}_i = f_i^{2*}\boldsymbol{S}_{vv}\boldsymbol{b}_i, \qquad \boldsymbol{b}'_i\boldsymbol{S}_{vv}\boldsymbol{b}_j = \delta_{ij}. \tag{2.19}$$

Then, we may estimate ρ^{2*} by:

$$\begin{aligned} r^{2*} &= f_1^{2*} + \cdots + f_{p_1}^{2*} \\ &= \mathrm{tr}(\boldsymbol{D}_x^{-1/2}\boldsymbol{H}'_x\boldsymbol{S}_{xy}\boldsymbol{H}_y\boldsymbol{D}_y^{-1}\boldsymbol{H}'_y\boldsymbol{S}_{yx}\boldsymbol{H}_x\boldsymbol{D}_x^{-1/2}) \\ &= \sum_{i=1}^{p_1}\sum_{j=1}^{q_1}\frac{(\boldsymbol{h}'_{ix}\boldsymbol{S}_{xy}\boldsymbol{h}_{jy})^2}{l_{ix}l_{jy}}. \end{aligned} \tag{2.20}$$

3. Limiting distributions of estimates of the sum of canonical correlation coefficients based on principal components

From Appendix A, we have the following theorem.

THEOREM 1. *The limiting distribution of* $\sqrt{n}(r^{2*} - \rho^{2*})$ *is normal with mean* 0 *and variance* σ_*^2:

$$\sigma_*^2 = \sum_{k=1}^{5} \sum_{k'=1}^{5} \sum_{i=1}^{p_1} \sum_{j=1}^{q_1} \sum_{l=1}^{p_1} \sum_{m=1}^{q_1} \sigma^*_{kijk'lm}, \tag{3.1}$$

where $\sigma^*_{kijk'lm} = E(K_{kij}K_{k'lm})$, $k, k' = 1, \ldots, 5$, $i, l = 1, \ldots, p_1$, $j, m = 1, \ldots, q_1$. *Appendix C gives* $\sigma^*_{kijk'lm}$. *This proof is in Appendix B.*

4. Example

We investigate the correlations between the academic record in senior high school and the score on the common first-stage university entrance examination. From the academic records and scores for 147 students, we obtained the following correlation matrix:

$$\begin{pmatrix} \boldsymbol{R}_{xx} & \boldsymbol{R}_{xy} \\ \boldsymbol{R}_{yx} & \boldsymbol{R}_{yy} \end{pmatrix} \tag{4.1}$$

$$= \begin{array}{c} \text{English} \\ \text{Japanese} \\ \text{Mathematics} \\ \text{Science} \\ \text{Social Studies} \\ \\ \text{English} \\ \text{Japanese} \\ \text{Mathematics} \\ \text{Science} \\ \text{Social Studies} \end{array} \left(\begin{array}{cccccccccc} 1 & & & & & & & & & \\ .257 & 1 & & & & & & & & \\ .330 & -.200 & 1 & & & & & & & \\ .103 & -.083 & .474 & 1 & & & & & & \\ .409 & .337 & .007 & .271 & 1 & & & & & \\ & & & & & & & & & \\ .537 & .169 & .101 & .083 & .347 & 1 & & & & \\ .375 & .448 & -.069 & -.004 & .434 & .638 & 1 & & & \\ .155 & -.124 & .571 & .437 & .096 & .474 & .193 & 1 & & \\ .083 & -.017 & .395 & .496 & .205 & .353 & .295 & .612 & 1 & \\ .279 & .365 & -.015 & .092 & .461 & .502 & .631 & .221 & .415 & 1 \end{array} \right),$$

where $\boldsymbol{R}_{xx}$ is the correlation of the common first-stage university entrance examination and $\boldsymbol{R}_{yy}$ is the correlation of the academic records in senior high school. The variances of $\boldsymbol{x}$ and $\boldsymbol{y}$ are $\mathrm{diag}(554.1, 474.4, 536.4, 463.5, 459.2)$ and $\mathrm{diag}(354.6, 328.3, 391.6, 452.0, 327.2)$, respectively. Now, we apply principal component analysis to the covariance matrices $\boldsymbol{S}_{xx}$ and $\boldsymbol{S}_{yy}$, respectively, and obtain the latent roots and the corresponding latent vectors as follows:

$$\begin{pmatrix} 856.1 \\ 771.6 \\ 358.8 \\ 296.1 \\ 205.0 \end{pmatrix}, \quad \begin{pmatrix} 0.625 & 0.142 & -0.731 & 0.121 & 0.200 \\ 0.378 & 0.417 & 0.552 & 0.587 & 0.188 \\ 0.161 & -0.712 & -0.021 & 0.551 & -0.405 \\ 0.326 & -0.542 & 0.298 & -0.312 & 0.643 \\ 0.578 & 0.078 & 0.267 & -0.490 & -0.590 \end{pmatrix},$$

and

$$\begin{pmatrix} 1014.7 \\ 413.4 \\ 223.9 \\ 113.0 \\ 88.8 \end{pmatrix}, \quad \begin{pmatrix} 0.465 & 0.261 & 0.563 & 0.139 & -0.616 \\ 0.400 & 0.518 & 0.004 & 0.430 & 0.622 \\ 0.441 & -0.536 & 0.442 & -0.380 & 0.423 \\ 0.515 & -0.480 & -0.537 & 0.413 & -0.213 \\ 0.406 & 0.381 & -0.447 & -0.693 & -0.096 \end{pmatrix}.$$

In this example we shall use the first and second principal components in each group for explaining the scores of the common first-stage university entrance examination and the academic records in senior high school. We compute the canonical component analysis using those principal components, and obtain the following canonical correlation coefficients:

$$r_1^{2*} = 0.50774, \qquad r_2^{2*} = 0.30770.$$

When we use all variables, we have the following canonical correlation coefficients:

$$r_1^2 = 0.52547, \qquad r_2^2 = 0.32277, \qquad r_3^2 = 0.21834,$$
$$r_4^2 = 0.05387, \qquad r_5^2 = 0.00992.$$

In general, the sum of the all canonical correlation coefficients between $\boldsymbol{x}$ and $\boldsymbol{y}$ is equal to that between $\boldsymbol{u}$ and $\boldsymbol{v}$. Additionally, as given in Fujikoshi (1982), the k-th canonical correlation coefficient of selected variables is less than or equal to that of all variables. In our case, the two principal components have almost the same information about the relationships between $\boldsymbol{x}$ and $\boldsymbol{y}$ as all the variables, because the difference of the two canonical correlation coefficients is small; furthermore, it is easier to interpret the canonical variables, as they are written using the uncorrected principal components.

With the score on the common first-stage university entrance examination, the first and second principal components keep 65% of the information, while 77% of the information from the academic records in senior high school is kept.

From the values of the characteristic vectors corresponding to the largest and second largest characteristic roots, we know that the first principal component is the factor with overall ability in five subjects, and the second principal component is the factor with ability in science and language.

From the canonical correlation analysis based on the first and second principal components in each group, we obtain the first canonical correlation coefficient of 0.511, and the second canonical correlation coefficient of 0.308. In this paper r^2 is the sum of 0.508 and 0.308, namely 0.815, i.e., the total sum of the canonical correlations based on the largest and second largest principal components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. By contrast, the largest and second largest canonical correlations on all variables are 0.525 and 0.323, and their sum is 0.848. Since the difference between 0.848 and 0.815 is small, based on the analysis proposed here, we lose only a slight amount of information contained in the correlations.

Let two principal components be u_1 and u_2 on the common first-stage university entrance examination, and v_1 and v_2 on the academic records in senior high school. The first canonical variables given by $0.326u_1 - 1.203u_2$ and $0.106v_1 - 1.045v_2$ are the variables mainly concerned with the second principal components. The second canonical variables given by $1.275u_1 + 0.307u_2$ and $1.653v_1 + 0.067v_2$ are the variables concerned with the first principal components. This means that the second principal component is higher correlated than the first principal components. We may obtain similar results in canonical correlation analysis, however the canonical correlation analysis based on principal components gives us clear results.

5. Simulation study

A simulation study is examined by generating random samples from a normal distribution having the population covariance matrix obtained in Section 4. In this case, $\sigma_*^2 = 1.03713$ from (3.1) by substituting the covariance matrix (4.1), its latent roots and the corresponding latent vectors. We compare the actual percentile point of $\sqrt{n}(r^{2*} - \rho^{2*})/\sigma_*$ with that of the standard normal distribution. The results of simulation studies for $N = 50, 100$ and 300 are listed in Table 1 by 100,000 replicates.

Table 1. Principal canonical components correlation.

Prob.	percentile	$N = 50$	100	300
0.95	1.645	1.6959	1.6776	1.6592
0.975	1.960	2.0138	1.9971	1.9722

From the results of the simulation study, we may find that actual percentile points are close to 1.64 or 1.96.

6. Bootstrapping

Developed by Efron in 1979, the bootstrap method can estimate measures of variability and bias. It can be used in nonparametric or parametric modes. The basic steps in the bootstrap procedure are as follows:

Step 1. Construct an empirical probability distribution, Ω, from the sample by setting a probability of $1/n$ for each point, $z_1, \ldots, z_n$ of the sample. This is the empirical distribution function of the sample, which is the nonparametric maximum likelihood estimate of the population distribution, ω; now, each sample element has the same probability of being drawn.

Step 2. From the empirical distribution function, Ω, draw a random sample of size n with replacement. This is a "resample".

Step 3. Calculate the statistic of interest, z, for this resample, yielding, z^*.

Step 4. Repeat Steps 2 and 3 B times, where B is a large number, in order to create B resamples. The practical size of B depends on the tests to be run on the data. Typically, B is at least 1000 when an estimate of the confidence interval around Γ is required.

Step 5. Construct the relative frequency histogram from the B number of z^*s by placing a probability of $1/B$ at each point, $z_1^*, \ldots, z_B^*$. The distribution obtained is the bootstrapped estimate of the sampling distribution of z.

We investigate the correlations between the academic record in senior high school and the score on the common first-stage university entrance examination using bootstrapping. The results of the bootstrap simulation study for $N = 100$, 150, and 300 are listed in Table 2 for $B = 1000$ replicates, where $\hat{r}^{2*} = \hat{r}_1^{2*} + \hat{r}_2^{2*}$ is the sum of the sample canonical correlation coefficient of the sample of size N, generated from the population having the covariance matrix (4.1), $\bar{r}^{2*}$ is the bootstrap mean of the sum of the sample canonical correlation coefficients, s_*^2 is the plug-in estimator of σ_*^2, and $z_{(1)}^* \leq \cdots \leq z_{(950)}^* \leq \cdots \leq z_{(1000)}^*$.

Table 2. Bootstrap $B = 1000$ times.

N	$\hat{r}^{2*}$	$\bar{r}^{2*}$	s_*^2	mean of z^*	variance of z^*	$0.95(z_{(950)}^*)$
100	0.8230	0.8197	1.0591	−0.03140	1.2125	1.7898
150	0.8434	0.8407	1.0225	−0.03108	1.1977	1.6981
300	0.8270	0.8284	1.0373	−0.02313	1.0252	1.6236

From the bootstrap results, we find that the value using the correlation is close to the real value.

7. Conclusion

We find that the limiting distribution of the canonical correlation of the principal components from two populations is the normal distribution by expansion of the correlation. From the results of a simulation study and bootstrapping, we get a value close to the real value, and one can easily determine the meanings of the principal components for both the academic records in senior high school and the scores on the common first-stage university entrance examination. Comparing the correlation from the principal components to the canonical correlation, it is clear that the canonical correlation is larger than the correlation from the principal components, but the difference between the methods is not large. Therefore, it is worth considering to use the correlation from the principal components. The limiting distribution of the canonical correlation of the principal components becomes a complex expression. Then, the bootstrap method is used. It is understood that we can analyze reliability, even if we do not perform a complex calculation by using the bootstrap method. However, the mathematical result in the first half of the paper is a result of worthy to research the characteristic of the amount of presumption.

Appendix A: Asymptotic expansion and limiting distribution of the correlation coefficient r^{2*}

We consider an asymptotic expansion of latent roots and vectors. The derivation is in Siotani *et al.* (1985). Let $\boldsymbol{W}$ be distributed as the $(p+q)\times(p+q)$ variate Wishart distribution with $p+q$ degrees of freedom, $W_{p+q}(\boldsymbol{\Sigma}, n)$, where

$$\boldsymbol{S} = \frac{1}{n}\boldsymbol{W} = \boldsymbol{\Sigma} + \frac{1}{\sqrt{n}}\boldsymbol{G}, \tag{A.1}$$

and $\boldsymbol{\Sigma}$ is the population covariate matrix. Then, it follows that the limiting distribution of $\boldsymbol{G} = (g_{ij})$ is normal with mean zero and covariance:

$$\mathrm{cov}(g_{ij}g_{kl}) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}. \tag{A.2}$$

The matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are partitioned in a similar to (2.2) and (2.14). Similarly, $\boldsymbol{G}$ is partitioned into p rows and q columns:

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_{xx} & \boldsymbol{G}_{xy} \\ \boldsymbol{G}_{yx} & \boldsymbol{G}_{yy} \end{pmatrix}. \tag{A.3}$$

From the perturbation method, the latent roots and vectors of $\boldsymbol{S}_{xx}$ and $\boldsymbol{S}_{yy}$ are expanded as follows:

$$l_{ix} = \lambda_{ix} + \frac{1}{\sqrt{n}}\lambda_{ix}^{(1)} + O_p(n^{-1}), \qquad \boldsymbol{h}_{ix} = \boldsymbol{\gamma}_{ix} + \frac{1}{\sqrt{n}}\boldsymbol{\gamma}_{ix}^{(1)} + O_p(n^{-1}), \tag{A.4}$$

$$l_{jy} = \lambda_{jy} + \frac{1}{\sqrt{n}}\lambda_{jy}^{(1)} + O_p(n^{-1}), \qquad \boldsymbol{h}_{jy} = \boldsymbol{\gamma}_{jy} + \frac{1}{\sqrt{n}}\boldsymbol{\gamma}_{jy}^{(1)} + O_p(n^{-1}), \tag{A.5}$$

where

$$\begin{aligned} \lambda_{ix}^{(1)} &= \boldsymbol{\gamma}_{ix}'\boldsymbol{G}_{xx}\boldsymbol{\gamma}_{ix}, \qquad \lambda_{jy}^{(1)} = \boldsymbol{\gamma}_{jy}'\boldsymbol{G}_{yy}\boldsymbol{\gamma}_{jy}, \\ \boldsymbol{\gamma}_{ix}^{(1)} &= -\boldsymbol{\Sigma}_{ix}^{*}(\boldsymbol{G}_{xx} - \lambda_{ix}^{(1)}\boldsymbol{I})\boldsymbol{\gamma}_{ix}, \qquad \boldsymbol{\gamma}_{jy}^{(1)} = -\boldsymbol{\Sigma}_{jy}^{*}(\boldsymbol{G}_{yy} - \lambda_{jy}^{(1)}\boldsymbol{I})\boldsymbol{\gamma}_{jy}, \\ \boldsymbol{\Sigma}_{ix}^{*} &= (\boldsymbol{\Sigma}_{xx} - \lambda_{ix}\boldsymbol{I})^{-}, \qquad \boldsymbol{\Sigma}_{jy}^{*} = (\boldsymbol{\Sigma}_{yy} - \lambda_{jy}\boldsymbol{I})^{-}, \end{aligned}$$

and $(\boldsymbol{\Sigma}_{xx} - \lambda_{ix}\boldsymbol{I})^{-}$ and $(\boldsymbol{\Sigma}_{yy} - \lambda_{jy}\boldsymbol{I})^{-}$ denote the generalized inverse matrixes of $\boldsymbol{\Sigma}_{xx} - \lambda_{ix}\boldsymbol{I}$ and $\boldsymbol{\Sigma}_{yy} - \lambda_{jy}\boldsymbol{I}$, respectively. The sample correlation of the principal components from the two populations is written as:

$$r^{2*} = \sum_{i=1}^{p_1} r_i^{2*} = \sum_{i=1}^{p_1}\sum_{j=1}^{q_1} l_{ix}^{-1} l_{jy}^{-1} (\boldsymbol{h}_{ix}'\boldsymbol{S}_{xy}\boldsymbol{h}_{jy})^2. \tag{A.6}$$

Substituting these expansions into r^{2*}, we obtain the asymptotic expansion of

$$r^{2*} = \rho^{2*} + \frac{1}{\sqrt{n}} \sum_{k=1}^{5} \sum_{i=1}^{p_1} \sum_{j=1}^{q_1} K_{kij} + O_p(n^{-1}), \tag{A.7}$$

where

$$K_{1ij} = 2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\tau_{ij}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}^{(1)}, \qquad K_{2ij} = 2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\tau_{ij}\boldsymbol{\gamma}_{ix}'\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{jy},$$
$$K_{3ij} = 2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\tau_{ij}\boldsymbol{\gamma}_{ix}^{(1)'}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}, \qquad K_{4ij} = -\rho_{ij}^{2}\lambda_{jy}^{-1}\lambda_{jy}^{(1)},$$
$$K_{5ij} = -\rho_{ij}^{2}\lambda_{ix}^{-1}\lambda_{ix}^{(1)}$$

and

$$\tau_{ij} = \boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}, \qquad \rho_{ij}^2 = \lambda_{ix}^{-1}\lambda_{jy}^{-1}\tau_{ij}^2, \qquad \rho^{2*} = \sum_{i=1}^{p_1}\sum_{j=1}^{q_1}\rho_{ij}^2.$$

The limiting normality of r_i^{2*} is assured by Theorem 2.1 of Seo *et al.* (1994) if $(\boldsymbol{U}, \boldsymbol{V})$ has finite sixth moments. Thus the limiting distribtuion of $r^{2*} = \sum_{i=1}^{p_1} r_i^{2*}$ is normal.

Appendix B: Proof of Theorem 1

We give the following Lemma before proving Theorem 1.

LEMMA. *The matrices* $\boldsymbol{G}$, $\boldsymbol{W}$ *and* $\boldsymbol{\Sigma}$ *are defined as before. Let*

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} = (g_{ab}), \tag{B.1}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = (\sigma_{ab}). \tag{B.2}$$

Then, we obtain

$$E(\boldsymbol{G}_{ij}\boldsymbol{A}\boldsymbol{G}_{kl}) = \boldsymbol{\Sigma}_{ik}\boldsymbol{A}'\boldsymbol{\Sigma}_{jl} + \boldsymbol{\Sigma}_{il}\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{kj}), \tag{B.3}$$
$$E(\boldsymbol{\alpha}'\boldsymbol{G}_{ij}\boldsymbol{\alpha}\boldsymbol{G}_{kl}) = \boldsymbol{\Sigma}_{kj}\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{il} + \boldsymbol{\Sigma}_{ki}\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{jl}, \tag{B.4}$$

where $\boldsymbol{A}$ *and* $\boldsymbol{\alpha}$ *are a real matrix and a real vector, respectively, and* i, j, k *and* l *are* x *or* y.

PROOF OF THE LEMMA. Let g_{ab}^{ij} denote (a, b)-th element of $\boldsymbol{G}_{ij}$, where i and j are x or y. Then, we have $g_{ab}^{ij} = u_{a+c_i,b+c_j}$, where $c_x = 0$, $c_y = p_1$. Therefore, we obtain:

$$E(\boldsymbol{G}_{ij}\boldsymbol{A}\boldsymbol{G}_{kl}) = \boldsymbol{\Sigma}_{ik}\boldsymbol{A}'\boldsymbol{\Sigma}_{jl} + \boldsymbol{\Sigma}_{il}\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_{kj}),$$
$$E(\boldsymbol{\alpha}'\boldsymbol{G}_{ij}\boldsymbol{\alpha}\boldsymbol{G}_{kl}) = \boldsymbol{\Sigma}_{kj}\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{il} + \boldsymbol{\Sigma}_{ki}\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{jl},$$

which gives the Lemma.

In order to get the parameter $\sigma^*_{kijk'lm}$ in Theorem 1, the following calculation is convenient:

$$\text{(B.5)}\qquad E(\lambda^{(1)}_{ix}\lambda^{(1)}_{lx}) = 2\delta_x^2,$$

$$\text{(B.6)}\qquad E(\lambda^{(1)}_{ix}\lambda^{(1)}_{jy}) = 2\tau_{ij}^2,$$

$$\text{(B.7)}\qquad E(\lambda^{(1)}_{ix}\boldsymbol{\gamma}^{(1)}_{lx}) = -2\delta_x\boldsymbol{\Sigma}^*_{lx}(\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{ix} - \delta_x\boldsymbol{\gamma}_{lx}),$$

$$\text{(B.8)}\qquad E(\lambda^{(1)}_{ix}\boldsymbol{\gamma}^{(1)}_{jy}) = -2\tau_{ij}\boldsymbol{\Sigma}^*_{jy}(\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{ix} - \tau_{ij}\boldsymbol{\gamma}_{jy}),$$

$$\begin{aligned}\text{(B.9)}\qquad E(\boldsymbol{\gamma}^{(1)}_{ix}\boldsymbol{\gamma}^{(1)\prime}_{lx}) = \boldsymbol{\Sigma}^*_{ix}(&\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xx} + \delta_x\boldsymbol{\Sigma}_{xx} - 2\delta_x\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{lx}\\ &- 2\delta_x\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xx} + 2\delta_x^2\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}'_{lx})\boldsymbol{\Sigma}^*_{lx},\end{aligned}$$

$$\begin{aligned}\text{(B.10)}\qquad E(\boldsymbol{\gamma}^{(1)}_{ix}\boldsymbol{\gamma}^{(1)\prime}_{jy}) = \boldsymbol{\Sigma}^*_{ix}(&\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy} + \tau_{ij}\boldsymbol{\Sigma}_{xy} - 2\tau_{ij}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{jy}\\ &- 2\tau_{ij}\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy} + 2\tau_{ij}^2\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}'_{jy})\boldsymbol{\Sigma}^*_{jy}\end{aligned}$$

where

$$\delta_x = \boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}, \qquad \delta_y = \boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}.$$

The limiting distribution of K_{kij} in Theorem 1 is normal because K_{kij} is a linear function of $\boldsymbol{G}_{ij}$, and the limiting distribution of $\boldsymbol{G}$ is normal. The variance of $\sqrt{n}(r^{2*} - \rho^{2*})$ is given by:

$$\begin{aligned}\operatorname{Var}(\sqrt{n}(r^{2*} - \rho^{2*})) &= \operatorname{Var}\left(\sum_{k=1}^{5}\sum_{i=1}^{p_1}\sum_{j=1}^{q_1} K_{kij} + O_p(n^{-1/2})\right)\\ &= \sum_{k=1}^{5}\sum_{k'=1}^{5}\sum_{i=1}^{p_1}\sum_{j=1}^{q_1}\sum_{l=1}^{p_1}\sum_{m=1}^{q_1} E(K_{kij}K_{k'lm}) + O_p(n^{-1/2}),\end{aligned}$$

where

$$\begin{aligned}E(K_{kij}) &= 0,\\ E(K_{1ij}K_{1lm}) &= 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}E(\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}^{(1)}_{jy}\boldsymbol{\gamma}^{(1)\prime}_{my}\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx})\\ &= 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^*_{jy}\\ &\quad\times(\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yy} + \delta_y\boldsymbol{\Sigma}_{yy} - 2\delta_y\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}\boldsymbol{\gamma}'_{my}\\ &\quad - 2\delta_y\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yy} + 2\delta_y^2\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{my})\boldsymbol{\Sigma}^*_{my}\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx},\\ E(K_{2ij}K_{2lm}) &= 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}E(\boldsymbol{\gamma}'_{ix}\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{lx}\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{my})\\ &= 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix} = 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix}\\ &\quad\times E(\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{lx}\boldsymbol{G}_{xy})\boldsymbol{\gamma}_{my}\\ &\quad\times(\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yy} + \boldsymbol{\Sigma}_{xy}\operatorname{tr}(\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{lx}\boldsymbol{\Sigma}_{xy}))\boldsymbol{\gamma}_{my}\\ &= 4\lambda^{-1}_{ix}\lambda^{-1}_{lx}\lambda^{-1}_{jy}\lambda^{-1}_{my}\tau_{ij}\tau_{lm}(\delta_x\delta_y + \tau_{im}\tau_{lj}),\\ E(K_{4ij}K_{4lm}) &= \lambda^{-1}_{jy}\lambda^{-1}_{my}\rho^2_{ij}\rho^2_{lm}E(\lambda^{(1)}_{jy}\lambda^{(1)}_{my}) = 2\lambda^{-1}_{jy}\lambda^{-1}_{my}\rho^2_{ij}\rho^2_{lm}\delta_y^2,\end{aligned}$$

$$
\begin{aligned}
E(K_{1ij}K_{2lm}) &= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}E(\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}^{(1)}\boldsymbol{\gamma}_{lx}'\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{my}) \\
&= -4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*} \\
&\quad \times (E(\boldsymbol{G}_{yy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{lx}'\boldsymbol{G}_{xy}) - \boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{lx}'E(\lambda_{jy}^{(1)}\boldsymbol{G}_{xy}))\boldsymbol{\gamma}_{my} \\
&= -4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*} \\
&\quad \times ((\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yy} + \boldsymbol{\Sigma}_{yy}\,\mathrm{tr}(\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{lx}'\boldsymbol{\Sigma}_{xy})) \\
&\quad - \boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{lx}'E(\boldsymbol{\gamma}_{jy}'\boldsymbol{G}_{yy}\boldsymbol{\gamma}_{jy}\boldsymbol{G}_{xy}))\boldsymbol{\gamma}_{my} \\
&= -4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*} \\
&\quad \times (\delta_y\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx} + \tau_{lj}\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my} - 2\tau_{lj}\delta_y\boldsymbol{\gamma}_{jy}), \\
E(K_{1ij}K_{3lm}) &= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}E(\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{jy}^{(1)}\boldsymbol{\gamma}_{lx}^{(1)'}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my}) \\
&= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}E(\boldsymbol{\gamma}_{jy}^{(1)}\boldsymbol{\gamma}_{lx}^{(1)'})\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my} \\
&= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*} \\
&\quad \times (\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yx} + \tau_{lj}\boldsymbol{\Sigma}_{yx} - 2\tau_{lj}\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}_{lx}' \\
&\quad - 2\tau_{lj}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yx} + 2\tau_{lj}^2\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{lx}')\boldsymbol{\Sigma}_{lx}^{*}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my}, \\
E(K_{1ij}K_{4lm}) &= -2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}\tau_{ij}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}E(\lambda_{my}^{(1)}\boldsymbol{\gamma}_{jy}^{(1)}) \\
&= 4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}\tau_{ij}\delta_y\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*}(\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my} - \delta_y\boldsymbol{\gamma}_{jy}), \\
E(K_{2ij}K_{4lm}) &= -2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\boldsymbol{\gamma}_{ix}'E(\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{jy}\lambda_{my}^{(1)}) \\
&= -2\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\boldsymbol{\gamma}_{ix}'E(\boldsymbol{G}_{xy}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{my}'\boldsymbol{G}_{yy})\boldsymbol{\gamma}_{my} \\
&= -4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\tau_{im}\delta_y, \\
E(K_{4ij}K_{5lm}) &= \lambda_{jy}^{-1}\lambda_{lx}^{-1}\rho_{ij}^2\rho_{lm}^2E(\lambda_{lx}^{(1)}\lambda_{jy}^{(1)}) = 2\lambda_{jy}^{-1}\lambda_{lx}^{-1}\rho_{ij}\rho_{lm}\tau_{lj}^2,
\end{aligned}
$$

and the others are obtained using the same calculation.

Appendix C: We may obtain the parameter $\sigma^{*}_{kijk'lm} = E(K_{kij}K_{k'lm})$ in Theorem 1 from Appendix B as follows:

$$
\begin{aligned}
\sigma^{*}_{1ij1lm} &= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{jy}^{*} \\
&\quad \times (\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yy} + \delta_y\boldsymbol{\Sigma}_{yy} - 2\delta_y\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}\boldsymbol{\gamma}_{my}' \\
&\quad - 2\delta_y\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yy} + 2\delta_y^2\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{my}')\boldsymbol{\Sigma}_{my}^{*}\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}, \\
\sigma^{*}_{2ij2lm} &= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}(\delta_x\delta_y + \tau_{im}\tau_{lj}), \\
\sigma^{*}_{3ij3lm} &= 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}_{jy}'\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{ix}^{*} \\
&\quad \times (\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xx} + \delta_x\boldsymbol{\Sigma}_{xx} - 2\delta_x\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}_{lx}' \\
&\quad - 2\delta_x\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}_{ix}'\boldsymbol{\Sigma}_{xx} + 2\delta_x^2\boldsymbol{\gamma}_{ix}\boldsymbol{\gamma}_{lx}')\boldsymbol{\Sigma}_{lx}^{*}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my}, \\
\sigma^{*}_{4ij4lm} &= 2\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{ij}^2\rho_{lm}^2\delta_y^2, \\
\sigma^{*}_{5ij5lm} &= 2\lambda_{ix}^{-1}\lambda_{lx}^{-1}\rho_{ij}^2\rho_{lm}^2\delta_x^2,
\end{aligned}
$$

$$\sigma^*_{1ij2lm} = -4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^*_{jy} \times(\delta_y\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}+\tau_{lj}\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}-2\tau_{lj}\delta_y\boldsymbol{\gamma}_{jy}),$$

$$\sigma^*_{1ij3lm} = 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^*_{jy}(\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yx}+\tau_{lj}\boldsymbol{\Sigma}_{yx} -2\tau_{lj}\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{lx}-2\tau_{lj}\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yx}+2\tau_{lj}^2\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}'_{lx})\boldsymbol{\Sigma}^*_{lx}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my},$$

$$\sigma^*_{1ij4lm} = 4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\delta_y\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^*_{jy}(\boldsymbol{\Sigma}_{yy}\boldsymbol{\gamma}_{my}-\delta_y\boldsymbol{\gamma}_{jy}),$$

$$\sigma^*_{1ij5lm} = 4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{lx}^{-1}\rho_{lm}^2\tau_{ij}\tau_{lj}\boldsymbol{\gamma}'_{ix}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}^*_{jy}(\boldsymbol{\Sigma}_{yx}\boldsymbol{\gamma}_{lx}-\tau_{lj}\boldsymbol{\gamma}_{jy}),$$

$$\sigma^*_{2ij3lm} = 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\tau_{ij}\tau_{lm}\boldsymbol{\gamma}'_{ix} \times(\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yx}+\tau_{lj}\boldsymbol{\Sigma}_{xx}-2\tau_{lj}\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}\boldsymbol{\gamma}'_{lx})\boldsymbol{\Sigma}^*_{lx}\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my},$$

$$\sigma^*_{2ij4lm} = -4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\tau_{im}\delta_y,$$

$$\sigma^*_{2ij5lm} = -4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\rho_{lm}^2\tau_{ij}\tau_{lj}\delta_x,$$

$$\sigma^*_{3ij4lm} = 4\lambda_{ix}^{-1}\lambda_{jy}^{-1}\lambda_{my}^{-1}\rho_{lm}^2\tau_{ij}\tau_{im}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yx}\Sigma^*_{ix}(\boldsymbol{\Sigma}_{xy}\boldsymbol{\gamma}_{my}-\tau_{im}\boldsymbol{\gamma}_{ix}),$$

$$\sigma^*_{3ij5lm} = 4\lambda_{ix}^{-1}\lambda_{lx}^{-1}\lambda_{jy}^{-1}\rho_{lm}^2\delta_x\tau_{ij}\boldsymbol{\gamma}'_{jy}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}^*_{ix}(\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_{lx}-\delta_x\boldsymbol{\gamma}_{ix}),$$

$$\sigma^*_{4ij5lm} = 2\lambda_{jy}^{-1}\lambda_{lx}^{-1}\rho_{ij}^2\rho_{lm}^2\tau_{lj}^2.$$

Acknowledgements

Authors are grateful to referees who carefully read the manuscript and gave valuable comments on improving this paper.

References

Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis, *Ann. Inst. Statist. Math.*, **34**, 523–530.

Hotelling, H. (1935). The most predictable criterion, *Journal of Educational Psychology*, **26**, 139–142.

Hotelling, H. (1936). Relations between two sets of variates, *Biometrika*, **28**, 321–377.

Seo, T., Kanda, T. and Fujikoshi, Y. (1994). The effects on the distributions of sample canonical correlations under nonnormality, *Commun. Statist. - Theory and Methods*, **23**, 2615–2628.

Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Sciences Press.