An Entropy-Based Tool to Help the Interpretation of Common-Factor Spaces in Factor Analysis

Nobuoki Eshima [1,*], Claudio Giovanni Borroni [2], Minoru Tabata [3] and Takeshi Kurosawa [4]

1 Center for Educational Outreach and Admissions, Kyoto University, Yoshida-machi, Sakyoku, Kyoto 660-8501, Japan
2 Department of Statistics and Quantitative Methods, University of Milano Bicocca, 20126 Milano, Italy; claudio.borroni@unimib.it
3 Department of Mathematical Sciences, Osaka Prefecture University, Osaka 599-8532, Japan; mnrtabata@luck.ocn.ne.jp
4 Department of Applied Mathematics, Tokyo University of Science, Kagurazaka, Shinzyukuku, Tokyo 162-0825, Japan; tkuro@rs.tus.ac.jp
* Correspondence: eshima.nobuoki.2z@kyoto-u.ac.jp; Tel.: +81-75-753-2517

Abstract: This paper proposes a method for deriving interpretable common factors based on canonical correlation analysis applied to the vectors of common factors and manifest variables in the factor analysis model. First, an entropy-based method for measuring factor contributions is reviewed. Second, the entropy-based contribution measure of the common-factor vector is decomposed into those of canonical common factors, and it is also shown that the importance order of factors is that of their canonical correlation coefficients. Third, the method is applied to derive interpretable common factors. Numerical examples are provided to demonstrate the usefulness of the present approach.

1. Introduction

In factor analysis, extracting interpretable factors is important for practical data analysis. In order to carry it out, methods for factor rotation have been studied, e.g., varimax [1] and orthomax [2] for orthogonal rotations and oblimin [3] and orthoblique [4] for oblique rotations. The basic idea for factor rotation in factor analysis is owed to the criteria of simple structures of factor analysis models by Thurstone [5], and the methods of factor rotation are constructed with respect to maximizations of variations of the squared factor loadings in order to derive simple structures of factor analysis models. Let X_i be manifest variables, let ξ_j be latent variables (common factors), let ε_i be unique factors related to X_i, and finally, let λ_{ij} be factor loadings that are weights of common factors ξ_j to explain X_i. Then, the factor analysis model is given as follows:

$$X_i = \sum_{j=1}^m \lambda_{ij}\xi_j + \varepsilon_i, \quad i = 1, 2, \ldots, p, \tag{1}$$

where

$$\begin{cases} \mathrm{E}(X_i) = \mathrm{E}(\varepsilon_i) = 0, \ i = 1, 2, \ldots, p; \\ \mathrm{E}(\xi_j) = 0, \ j = 1, 2, \ldots, m; \\ \mathrm{Var}(\xi_j) = 1, \ j = 1, 2, \ldots, m; \ \mathrm{Cov}(\xi_k, \xi_l) = \phi_{kl}; \\ \mathrm{Var}(\varepsilon_i) = \omega_i^2 > 0, \ i = 1, 2, \ldots, p; \\ \mathrm{Cov}(\varepsilon_k, \varepsilon_l) = 0, \ k \neq l. \end{cases}$$

To derive simple structures of factor analysis models, for example, in the varimax method, the following variation function of squared factor loadings is maximized with respect to factor loadings:

$$V = \sum_{i=1}^{p} \sum_{j=1}^{m} \left(\lambda_{ij}^2 - \overline{\lambda^2} \right)^2, \tag{2}$$

where $\overline{\lambda^2} = \frac{1}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_{ij}^2$. In this sense, the basic factor rotation methods can be viewed as those for exploratively analyzing multidimensional common-factor spaces. The interpretation of factors is made according to manifest variables with large weights in common factors. As far as we know, novel methods for factor rotation have not been investigated except for rotation methods similar to the above basic ones. In real data analyses, manifest variables are usually classified into some groups of variables in advance that may have common factors and concepts for themselves. For example, suppose we have a test battery including the following five subjects: Japanese, English, Social Science, Mathematics, and Natural Science. It is then reasonable to classify the five subjects into two groups, {Japanese, English, Social Science} and {Mathematics, Natural Science}. In such cases, it is meaningful to determine common factors related to the two manifest variable groups. For this objective, it is useful to develop a novel method to derive the common factors based on a factor contribution measure. In conventional methods of factor rotation, for example, as mentioned above, variation function (2) for the varimax method is not related to factor contribution.

An entropy-based method for measuring factor contribution was proposed by [6], and the method can measure factor contributions to manifest variables vectors and can decompose the factor contributions into those of manifest subvectors and individual manifest variables. By using the method, we can derive important common factors related to the manifest subvectors and the manifest variables. The aim of the present paper is to propose a new method for deriving simple structures based on entropy, that is, extracting common factors easy to interpret. In Section 2, an entropy-based method for measuring factor contribution [6] is reviewed to apply its properties for deriving simple structures in factor analysis models. Section 3 discusses canonical correlation analysis between common factors and manifest variables, and the contributions of common factors to the manifest variables are decomposed into components related to the extracted pairs of canonical variables. A numerical example is given to demonstrate the approach. In Section 4, canonical correlation analysis is applied to obtain common factors easy to interpret, and the contributions of the extracted factors are measured. Numerical examples are given to illustrate the present approach, and finally, Section 5 provides discussions and conclusions to summarize the present approach.

2. Entropy-Based Method for Measuring Factor Contributions

First, in order to derive factor contributions, factor analysis model (1) with error terms ε_i, $i = 1, 2, \ldots, p$, which are normally distributed, can be discussed in the framework of generalized linear models (GLMs) [7]. A general path diagram among manifest variables X_i, $i = 1, 2, \ldots, p$ and common factors ξ_j, $j = 1, 2, \ldots, m$ in the factor analysis model is illustrated in Figure 1. The conditional density functions of manifest variables of X_i, $i = 1, 2, \ldots, p$, given the factors ξ_j, $j = 1, 2, \ldots, m$, are expressed as follows:

$$f_i(x_i|\boldsymbol{\xi}) = \frac{1}{\sqrt{2\pi\omega_i^2}} \exp\left(-\frac{\left(x_i - \sum_{j=1}^{m} \lambda_{ij}\xi_j \right)^2}{2\omega_i^2} \right)$$

$$= \exp\left(\frac{x_i \sum_{j=1}^{m} \lambda_{ij}\xi_j - \frac{1}{2}\left(\sum_{j=1}^{m} \lambda_{ij}\xi_j \right)^2}{\omega_i^2} - \frac{x_i^2}{2\omega_i^2} - \log\sqrt{2\pi\omega_i^2} \right), \quad i = 1, 2, \ldots, p.$$

Let $\theta_i = \sum_{j=1}^{m} \lambda_{ij} \xi_j$ and $d(x_i, \omega_i^2) = -\frac{x_i^2}{2\omega_i^2} - \log \sqrt{2\pi\omega_i^2}$. Then, the above density function is described in a GLM framework as

$$f_i(x_i|\boldsymbol{\xi}) = \exp\left(\frac{x_i\theta_i - \frac{1}{2}\theta_i^2}{\omega_i^2} + d\left(x_i, \omega_i^2\right) \right), \, i = 1, 2, \ldots, p. \tag{3}$$

According to the local independence of the manifest variables in factor analysis model (1), the conditional density function of $X = (X_1, X_2, \ldots, X_p)^T$ given $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$ is expressed as

$$f(\boldsymbol{x}|\boldsymbol{\xi}) = \prod_{i=1}^{p} \exp\left(\frac{x_i\theta_i - \frac{1}{2}\theta_i^2}{\omega_i^2} + d\left(x_i, \omega_i^2\right) \right) = \exp\left(\sum_{i=1}^{p} \frac{x_i\theta_i - \frac{1}{2}\theta_i^2}{\omega_i^2} + \sum_{i=1}^{p} d\left(x_i, \omega_i^2\right) \right). \tag{4}$$
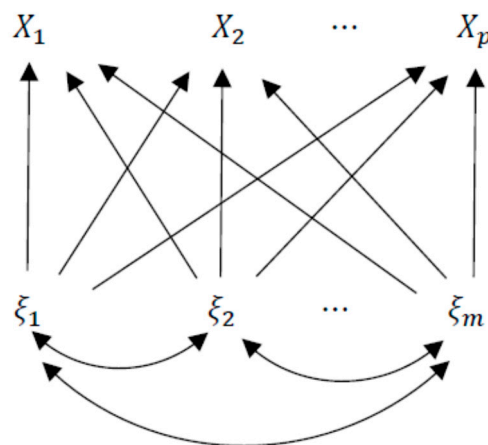

Figure 1. Path diagram of a general factor analysis model.

Let $g(\boldsymbol{\xi})$ be the joint density function of common-factor vector $= (\xi_1, \xi_2, \ldots, \xi_m)^T$; let $f_i(x_i)$ be the marginal density functions of X_i, $i = 1, 2, \ldots, p$; and let us set

$$\mathrm{KL}(X, \boldsymbol{\xi}) = \iint f(\boldsymbol{x}|\boldsymbol{\xi})g(\boldsymbol{\xi}) \log \frac{f(\boldsymbol{x}|\boldsymbol{\xi})}{f(\boldsymbol{x})} d\boldsymbol{x}d\boldsymbol{\xi} + \iint f(\boldsymbol{x})g(\boldsymbol{\xi}) \log \frac{f(\boldsymbol{x})}{f(\boldsymbol{x}|\boldsymbol{\xi})} d\boldsymbol{x}d\boldsymbol{\xi}, \tag{5}$$

$$c\mathrm{KL}(X_i, \boldsymbol{\xi}) = \iint f_i(x_i|\boldsymbol{\xi})g(\boldsymbol{\xi}) \log \frac{f_i(x_i|\boldsymbol{\xi})}{f_i(x_i)} dx_id\boldsymbol{\xi} + \iint f_i(x_i)g(\boldsymbol{\xi}) \log \frac{f_i(x_i)}{f_i(x_i|\boldsymbol{\xi})} dx_id\boldsymbol{\xi}, \tag{6}$$
$$i = 1, 2, \ldots, p.$$

where "KL" stands for "Kullback–Leibler information" [8]. From (3) and (4), we have

$$\mathrm{KL}(X_i, \boldsymbol{\xi}) = \frac{\mathrm{Cov}(X_i, \theta_i)}{\omega_i^2} = \sum_{j=1}^{m} \frac{\lambda_{ij}\mathrm{Cov}\left(X_i, \xi_j\right)}{\omega_i^2}, i = 1, 2, \ldots, p; \tag{7}$$

$$\mathrm{KL}(X, \boldsymbol{\xi}) = \sum_{i=1}^{p} \frac{\mathrm{Cov}(X_i, \theta_i)}{\omega_i^2} = \sum_{i=1}^{p} \sum_{j=1}^{m} \frac{\lambda_{ij}\mathrm{Cov}\left(X_i, \xi_j\right)}{\omega_i^2}. \tag{8}$$

The above quantities (7) and (8) are interpreted as the signal-to-noise ratios for dependent variables X_i and predictors θ_i; and the signal-to-noise ratio for dependent-variable vectors X and common-factor vector $\boldsymbol{\xi}$, respectively.

From (7) and (8), the following theorem can be derived [6]:

Theorem 1. *In factor analysis model (1), let* $X = (X_1, \ X_2, \ldots, X_p)^T$ *and* $\xi = (\xi_1, \ \xi_2, \ldots, \xi_m)^T$. *Then,*

$$\mathrm{KL}(X, \xi) = \sum_{i=1}^{p} \mathrm{KL}(X_i, \xi).$$

Consistently, the following theorem, which is actually an extended version of Corollary 1 in [6], can be also obtained:

Theorem 2. *Let manifest variable subvectors* $X^{(a)}$, $a = 1, 2, \ldots, A$ *be any decomposition of manifest variable vector* $X = (X_1, \ X_2, \ldots, X_p)^T$. *Then,*

$$\mathrm{KL}(X, \xi) = \sum_{a=1}^{A} \mathrm{KL}\left(X^{(a)}, \xi\right). \tag{9}$$

Following Eshima et al. [6], the contribution of factor vector $\xi = (\xi_1, \ \xi_2, \ldots, \xi_m)^T$ to manifest variable vector $X = (X_1, \ X_2, \ldots, X_p)^T$ is thus defined as

$$C(\xi \to X) = \mathrm{KL}(X, \xi),$$

so that, in Theorem 2, the contributions of factor vector $\xi = (\xi_1, \ \xi_2, \ldots, \xi_m)^T$ to manifest variable vectors $X^{(a)}$, $a = 1, 2, \ldots, A$ are defined by

$$C\left(\xi \to X^{(a)}\right) = \mathrm{KL}\left(X^{(a)}, \xi\right), \ a = 1, 2, \ldots, A.$$

Let $\xi^{\backslash j}$ be subvectors of all variables ξ_i except ξ_j from $\xi = (\xi_1, \ \xi_2, \ldots, \xi_m)^T$, i.e.,

$$\xi^{\backslash j} = (\xi_1, \ \xi_2, \ldots, \xi_{j-1}, \xi_{j+1}, \ldots, \xi_m)^T, \ j = 1, 2, \ldots, m;$$

and let $\mathrm{KL}\left(X, \xi^{\backslash j} \middle| \xi_j\right)$ and $\mathrm{KL}\left(X^{(a)}, \xi^{\backslash j} \middle| \xi_j\right)$ be the conditional Kullback–Leibler information as defined in (5) and (6). The contributions of common factors ξ_j are defined by

$$C(\xi_j \to X) = \mathrm{KL}(X, \xi) - \mathrm{KL}\left(X, \xi^{\backslash j} \middle| \xi_j\right),$$

$$C\left(\xi_j \to X^{(a)}\right) = \mathrm{KL}\left(X^{(a)}, \xi\right) - \mathrm{KL}\left(X^{(a)}, \xi^{\backslash j} \middle| \xi_j\right), \ j = 1, 2, \ldots, m.$$

Remark 1. *Information* $\mathrm{KL}\left(X, \xi^{\backslash j} \middle| \xi_j\right)$ *and* $\mathrm{KL}\left(X^{(a)}, \xi^{\backslash j} \middle| \xi_j\right)$ *can be expressed by using the conditional covariances* $\mathrm{Cov}\left(X_i, \theta_i \middle| \xi_j\right)$. *For example,*

$$\mathrm{KL}\left(X, \xi^{\backslash j} \middle| \xi_j\right) = \sum_{i=1}^{p} \frac{\mathrm{Cov}\left(X_i, \theta_i \middle| \xi_j\right)}{\omega_i^2}.$$

Finally, the following decomposition of $\mathrm{KL}(X, \xi)$ holds for orthogonal factors ([6], Theorem 3):

Theorem 3. *If the common factors are mutually independent, it follows that*

$$C(\xi \to X) = \sum_{j=1}^{m} \sum_{a=1}^{A} C\left(\xi_j \to X^{(a)}\right) = \sum_{j=1}^{m} \sum_{i=1}^{p} C(\xi_j \to X_i).$$

The entropy coefficient of determination (ECD) [9] between ξ and X is defined by

$$\mathrm{ECD}(\xi, X) = \frac{\mathrm{KL}(\xi, X)}{\mathrm{KL}(\xi, X) + 1},$$

so that the total relative contribution of factor vector ξ to manifest variable vector X in entropy can be defined as

$$\widetilde{RC}(\boldsymbol{\xi} \to \boldsymbol{X}) = \text{ECD}(\boldsymbol{\xi}, \boldsymbol{X}) = \frac{C(\boldsymbol{\xi} \to \boldsymbol{X})}{C(\boldsymbol{\xi} \to \boldsymbol{X}) + 1},$$

while, for a single factor ξ_j, two relative contribution ratios can be defined:

$$\text{RC}(\xi_j \to \boldsymbol{X}) = \frac{C(\xi_j \to \boldsymbol{X})}{C(\boldsymbol{\xi} \to \boldsymbol{X})} = \frac{\text{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \text{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{\backslash j} \middle| \xi_j\right)}{\text{KL}(\boldsymbol{\xi}, \boldsymbol{X})},$$

$$\widetilde{\text{RC}}(\xi_j \to \boldsymbol{X}) = \frac{C(\xi_j \to \boldsymbol{X})}{\text{KL}(\boldsymbol{\xi}, \boldsymbol{X}) + 1} = \frac{\text{KL}(\boldsymbol{X}, \boldsymbol{\xi}) - \text{KL}\left(\boldsymbol{X}, \boldsymbol{\xi}^{\backslash j} \middle| \xi_j\right)}{\text{KL}(\boldsymbol{\xi}, \boldsymbol{X}) + 1}$$

(see [6] for details).

Second, factor analysis model (1) in a general case is discussed. Let $\boldsymbol{\Sigma}$ be the variance–covariance matrix of manifest variable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$; let $\boldsymbol{\Omega}$ be the $p \times p$ variance–covariance matrix of unique factor vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$; let $\boldsymbol{\Lambda}$ be the $p \times m$ factor loading matrix of λ_{ij}; and let $\boldsymbol{\Phi}$ be the correlation matrix of common-factor vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$. Then, model (1) can be expressed as

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon}$$

and we have

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Omega}.$$

Now, the above discussion is extended in a general factor analysis model (1) with the following variance–covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$:

$$\begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Omega} & \boldsymbol{\Lambda}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Lambda}^T & \boldsymbol{\Phi} \end{pmatrix}. \tag{10}$$

Let $\boldsymbol{\theta} = \boldsymbol{\Lambda}\boldsymbol{\xi}$ be the predictor vector of manifest variable vector $\boldsymbol{X}^T = (X_1, X_2, \ldots, X_p)$. Then, the contribution of common-factor vector $\boldsymbol{\xi}$ to manifest variable vector X is defined by the following generalized signal-to-noise ratio:

$$E\left(\boldsymbol{X}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}\right) = \frac{E\left(\boldsymbol{X}^T\tilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\xi}\right)}{|\boldsymbol{\Omega}|} = \frac{\text{tr}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T}{|\boldsymbol{\Omega}|}, \tag{11}$$

where $\tilde{\boldsymbol{\Omega}}$ is the cofactor matrix of $\boldsymbol{\Omega}$. The signal is $\text{tr}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T$ and the noise $|\boldsymbol{\Omega}|$, and both are positive. Hence, the above quantity is defined as the explained entropy with the factor analysis model, and the same notation $KL(\boldsymbol{X}, \boldsymbol{\xi})$ as above is used, having to do with the Kullback–Leibler information for the factor analysis model with normal distribution errors (4). Similarly, in the general model, as in (9), signal-to-noise ratio (11) is decomposed into

$$\frac{\text{tr}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T}{|\boldsymbol{\Omega}|} = \sum_{i=1}^{p} \frac{\text{Cov}(X_i, \theta_i)}{\omega_i^2} = \sum_{i=1}^{p}\sum_{j=1}^{m} \frac{\lambda_{ij}\text{Cov}(X_i, \xi_j)}{\omega_i^2},$$

so the above theorems hold true as well. Thus, the results mentioned above are applicable to factor analysis models with error terms with non-normal distributions.

3. Canonical Factor Analysis

In order to derive interpretable factors from the common-factor space, we propose taking advantage of the results of canonical correlation analysis applied to manifest variables and common factors. This approach can be referred to as "canonical factor analysis" [10]. In the factor analysis model (1), the variance–covariance matrix of $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)^T$ is given by (10). Then, we have the following theorem:

Theorem 4. *For canonical correlation coefficients ρ_k, $k = 1, 2, \ldots, m$ between X and ξ in factor analysis model (1) with (10), it follows that*

$$\mathrm{KL}(X, \xi) = \sum_{j=1}^{m} \frac{\rho_j^2}{1 - \rho_j^2}.$$

Proof. Let $B^{(1)}$, $B^{(2)}$, and F be $m \times p$, $(p - m) \times p$, and $m \times m$ matrices, respectively; let $V^{(1)} = (V_1, V_2, \ldots, V_m)^T = B^{(1)} X$, $V^{(2)} = B^{(2)} X$, and $\eta = (\eta_1, \eta_2, \ldots, \eta_m)^T = F\xi$. It is assumed that (V_j, η_j) are the pairs of canonical variables with correlation coefficients ρ_j, $j = 1, 2, \ldots, m$; that matrices $\begin{pmatrix} B^{(1)} \\ B^{(2)} \end{pmatrix}$ and F are nonsingular; and that $V^{(1)}$ and $V^{(2)}$ are statistically independent. Since all pairs of canonical variables (V_j, η_j) and $V^{(2)}$ are mutually independent, we have

$$\mathrm{KL}\left(V^{(2)}, \eta\right) = 0, \ \mathrm{KL}\left(V^{(1)}, \eta_j\right) = \mathrm{KL}(V_j, \eta_j), \ j = 1, 2, \ldots, m.$$

From Theorem 2, it follows that

$$\mathrm{KL}(X, \xi) = \mathrm{KL}(V, F\xi) = \mathrm{KL}\left(\begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}, \eta\right) = \mathrm{KL}\left(V^{(1)}, \eta\right) + \mathrm{KL}\left(V^{(2)}, \eta\right) = \mathrm{KL}\left(V^{(1)}, \eta\right) = \sum_{j=1}^{m} \mathrm{KL}\left(V^{(1)}, \eta_j\right)$$

$$= \sum_{j=1}^{m} \mathrm{KL}(V_j, \eta_j) = \sum_{j=1}^{m} \frac{\rho_j^2}{1 - \rho_j^2}.$$

This completes the theorem. $\square$

In the proof of the above theorem, we have

$$\mathrm{KL}(X, \eta_j) = \mathrm{KL}(V_j, \eta_j) = \frac{\rho_j^2}{1 - \rho_j^2}, \ j = 1, 2, \ldots, m. \tag{12}$$

It implies that

$$\mathrm{C}(\eta_j \to X) = \mathrm{C}(\eta_j \to V_j) = \frac{\rho_j^2}{1 - \rho_j^2};$$

$$\widetilde{\mathrm{RC}}(\eta_j \to X) = \frac{\mathrm{KL}(X, \eta_j)}{\mathrm{KL}(X, \xi) + 1} = \frac{\mathrm{KL}(X, \eta_j)}{\mathrm{KL}(\eta, V) + 1} = \frac{\dfrac{\rho_j^2}{1 - \rho_j^2}}{\sum_{a=1}^{m} \dfrac{\rho_a^2}{1 - \rho_a^2} + 1};$$

$$\mathrm{RC}(\eta_j \to X) = \frac{\mathrm{KL}(X, \eta_j)}{\mathrm{KL}(\xi, X)} = \frac{\mathrm{KL}(V_j, \eta_j)}{\mathrm{KL}(V, \eta)} = \frac{\dfrac{\rho_j^2}{1 - \rho_j^2}}{\sum_{a=1}^{m} \dfrac{\rho_a^2}{1 - \rho_a^2}}, \ j = 1, 2, \ldots, m.$$

Theorem 4 shows that the contribution of common-factor vector ξ to manifest variable vector X is decomposed into those of canonical common factors η_j, i.e.,

$$\mathrm{KL}(X, \xi) = \sum_{j=1}^{m} \mathrm{KL}(X, \eta_j) = \sum_{j=1}^{m} \mathrm{KL}(V_j, \eta_j), \ j = 1, 2, \ldots, m.$$

Let us assume

$$1 > \rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_m^2 \geq 0. \tag{13}$$

According to the entropy-based criterion in Theorem 4, the order of importance of canonical common factors is that of canonical correlation coefficients. The interpretation of factors η_j can be made with the corresponding manifest canonical variables V_j and the factor loading matrix of canonical common factors $\boldsymbol{\eta} = \boldsymbol{F}\boldsymbol{\xi}$. For the canonical common factors, the factor loading matrix can be obtained as $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{F}^{-1}$. We refer to the canonical correlation analysis in Theorem 4 as canonical factor analysis [10].

Theorem 5. *In factor analysis model (1), for any $p \times p$ and $m \times m$ nonsingular matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, the canonical factor analysis between manifest variable vector $\boldsymbol{PX}$ and common-factor vector $\boldsymbol{Q}\boldsymbol{\xi}$ is invariant.*

Proof. Since the variance–covariance matrix of $\boldsymbol{PX}$ and $\boldsymbol{Q}\boldsymbol{\xi}$ is given by

$$
\begin{pmatrix} \boldsymbol{P} & 0 \\ 0 & \boldsymbol{Q} \end{pmatrix}
\begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Lambda}^{\mathrm{T}} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_m \end{pmatrix}
\begin{pmatrix} \boldsymbol{P} & 0 \\ 0 & \boldsymbol{Q} \end{pmatrix}^{T},
$$

the theorem follows. $\square$

Notice that we also have

$$
\mathrm{KL}(\boldsymbol{PX}, \boldsymbol{Q}\boldsymbol{\xi}) = \mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}).
$$

From the above theorem, the results of the canonical factor analysis do not depend on the initial common factors $\tilde{\xi}_j$ in factor analysis model (1). For factor analysis model (1), it follows that

$$
\mathrm{KL}(\boldsymbol{X}, \boldsymbol{\xi}) = \sum_{j=1}^{m} \mathrm{KL}(V_j, \eta_j) = \sum_{i=1}^{p} \mathrm{KL}(X_i, \boldsymbol{\xi}),
$$

implying that

$$
\frac{\mathrm{tr}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T}{|\boldsymbol{\Omega}|} = \sum_{j=1}^{m} \frac{\rho_j^2}{1-\rho_j^2} = \sum_{i=1}^{p} \frac{R_i^2}{1-R_i^2},
$$

where R_i are the multiple correlation coefficients between manifest variables X_i and factor vector $\boldsymbol{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_m)$, $i = 1, 2, \ldots, p$.

Numerical Example 1

Table 1 shows the results of orthogonal factor analysis (varimax method by S-PLUS ver. 8.2) as reported in [6]; the same example is used here to demonstrate the canonical factor analysis mentioned above. In Table 1, manifest variables X_1, X_2, and X_3 are scores in some subjects in the liberal arts, while variables X_4 and X_5 are those in the sciences. We refer to the factors as the initial common factors. In this example, from Table 1, the variance–covariance matrices in (10) are given as follows:

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
1 & 0.54 & 0.39 & 0.42 & 0.36 \\
0.54 & 1 & 0.49 & 0.38 & 0.22 \\
0.39 & 0.49 & 1 & 0.21 & 0 \\
0.42 & 0.38 & 0.21 & 1 & 0.54 \\
0.36 & 0.22 & 0 & 0.54 & 1
\end{pmatrix},
$$

$$
\boldsymbol{\Phi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
$$

where covariance matrix $\boldsymbol{\Lambda}^T$ is given in Table 1.

Table 1. Factor loadings of orthogonal (varimax) factor analysis.

	X_1	X_2	X_3	X_4	X_5
ξ_1	0.60	0.75	0.65	0.32	0.00
ξ_2	0.39	0.24	0.00	0.59	0.92
uniqueness	0.50	0.38	0.58	0.55	0.16

Uniqueness is the proportion of unique factor ε_i related to manifest variable X_i.

From the above matrices, to obtain the pairs of canonical variables, linear transformation matrices $\boldsymbol{B}^{(1)}$ and $\boldsymbol{F}$ in Theorem 4 are as follows:

$$\boldsymbol{B}^{(1)} = \begin{pmatrix} 0.19 & 0.20 & 0.06 & 0.20 & 0.94 \\ 0.32 & 0.58 & 0.37 & 0.00 & -0.65 \end{pmatrix},$$

and

$$\boldsymbol{F} = \begin{pmatrix} 0.32 & 0.95 \\ 0.95 & -0.32 \end{pmatrix}.$$

By the above matrices, we have the following pairs of canonical variables (V_i, η_i) and their squared canonical correlation coefficients ρ_i^2:

$$\begin{cases} V_1 = 0.19X_1 + 0.20X_2 + 0.06X_3 + 0.20X_4 + 0.94X_5, \\ \qquad \eta_1 = 0.32\xi_1 + 0.95\xi_2, \\ \qquad\qquad \rho_1^2 = 0.88, \end{cases}$$

$$\begin{cases} V_2 = 0.32X_1 + 0.58X_2 + 0.37X_3 + 0.07X_4 - 0.65X_5, \\ \qquad \eta_2 = 0.95\xi_1 - 0.32\xi_2, \\ \qquad\qquad \rho_2^2 = 0.73. \end{cases}$$

According to the above canonical variables, the factor loading for canonical factors η_i, $i = 1, 2$ is calculated with the initial loading matrix $\boldsymbol{\Lambda}$ and the rotation matrix $\boldsymbol{F}$, and we have

$$\begin{aligned} \boldsymbol{\Lambda}^{*T} = \left(\boldsymbol{\Lambda F}^{-1}\right)^T &= \begin{pmatrix} 0.32 & 0.95 \\ 0.95 & -0.32 \end{pmatrix}^{-1} \begin{pmatrix} 0.6 & 0.75 & 0.65 & 0.32 & 0.00 \\ 0.39 & 0.24 & 0.00 & 0.59 & 0.92 \end{pmatrix} \\ &= \begin{pmatrix} 0.56 & 0.47 & 0.21 & 0.66 & 0.87 \\ 0.45 & 0.64 & 0.62 & 0.12 & -0.29 \end{pmatrix}. \end{aligned}$$

From the above results, the first canonical factor η_1 can be viewed as a general common ability (factor) to solve all five subjects. The second factor η_2 can be regarded as a factor related to subjects in the liberal arts, which is independent of the first canonical factor. In the canonical correlation analysis, the contributions of canonical factors are calculated. Since the multiple correlation coefficient between η_1 and $\boldsymbol{X} = (X_1, X_2, \ldots, X_5)^T$ is $\rho_1^2 = 0.88$ and that between η_2 and $\boldsymbol{X}$ is $\rho_2^2 = 0.73$, we have

$$\mathrm{C}(\eta_1 \to \boldsymbol{X}) = \frac{\rho_1^2}{1 - \rho_1^2} = 7.06, \quad \mathrm{C}(\eta_2 \to \boldsymbol{X}) = \frac{\rho_2^2}{1 - \rho_2^2} = 2.70.$$

Let $\boldsymbol{\xi} = (\xi_1, \xi_2)$. From the above results, we have

$$\mathrm{C}(\boldsymbol{\xi} \to \boldsymbol{X}) = \mathrm{KL}(\boldsymbol{\xi}, \boldsymbol{X}) = \mathrm{C}(\eta_1 \to \boldsymbol{X}) + \mathrm{C}(\eta_2 \to \boldsymbol{X}) = 9.86,$$
$$\widetilde{\mathrm{CR}}(\boldsymbol{\xi} \to \boldsymbol{X}) = \frac{\mathrm{KL}(\boldsymbol{\xi}, \boldsymbol{X})}{\mathrm{KL}(\boldsymbol{\xi}, \boldsymbol{X}) + 1} = 0.91 (= \mathrm{ECD}(\boldsymbol{\xi}, \boldsymbol{X})).$$

From this, 91% of the variation of manifest random vector $\boldsymbol{X}$ in entropy is explained by the common latent factors $\boldsymbol{\xi}$. The contribution ratios of canonical common factors are calculated as follows:

$$\mathrm{CR}(\eta_1 \to \boldsymbol{X}) = \frac{7.06}{7.06 + 2.70} = 0.72, \quad \mathrm{CR}(\eta_2 \to \boldsymbol{X}) = 2.70.$$

The contribution of the first canonical factor is about 2.6 times greater than that of the second one.

4. Deriving Important Common Factors Based on Decomposition of Manifest Variables into Subsets

From (9) in Theorem 2, $KL(X, \xi)$ is decomposed into those for manifest variable subvectors $X^{(a)}$, $KL\left(X^{(a)}, \xi\right)$, $a = 1, 2, \ldots, A$. Thus, we have the following theorem:

Theorem 6. *Let manifest variable vector X be decomposed into subvectors $X^{(a)}$, $a = 1, 2, \ldots, A$. Let $\rho_{(a)j}$, $j = 1, 2, \ldots, m_{(a)}$ be the canonical correlation coefficients between manifest variable subvector $X^{(a)}$ and common-factor vector ξ, $a = 1, 2, \ldots, A$ in the factor analysis model (1), where $m_{(a)} \leq \min\left\{ \text{dimension of } X^{(a)}, m \right\}$. Then, $KL(X, \xi)$ is decomposed into canonical components as follows:*

$$KL(X, \xi) = \sum_{a=1}^{A} \sum_{j=1}^{m_{(a)}} \frac{\rho_{(a)j}^2}{1 - \rho_{(a)j}^2}.$$

Proof. For manifest variable vector $X^{(a)}$ and common-factor vector ξ, applying canonical correlation analysis, we have $m_{(a)}$ pairs of canonical variables $\left(V_j^{(\alpha)}, \eta_j^{(\alpha)} \right)$ with squared canonical correlation coefficients $\rho_{(a)j}^2$, $j = 1, 2, \ldots, m_{(a)}$. Then, applying Theorem 4 to $KL\left(X^{(a)}, \xi\right)$ it follows that

$$KL\left(X^{(a)}, \xi\right) = \sum_{j=1}^{m_{(a)}} KL\left(V_j^{(\alpha)}, \eta_j^{(\alpha)}\right) = \sum_{j=1}^{m_{(a)}} \frac{\rho_{(a)j}^2}{1 - \rho_{(a)j}^2}, \quad a = 1, 2, \ldots, A.$$

From Theorem 2, the theorem follows. □

Remark 2. *As shown in the above theorem, the following relations hold:*

$$KL\left(X^{(a)}, \eta_j^{(\alpha)}\right) = KL\left(V_j^{(\alpha)}, \eta_j^{(\alpha)}\right) = \frac{\rho_{(a)j}^2}{1 - \rho_{(a)j}^2}, \quad j = 1, 2, \ldots, m_{(a)}; \ a = 1, 2, \ldots, A.$$

In this sense,

$$C\left(\eta_j^{(a)} \to X^{(a)}\right) = \frac{\rho_{(a)j}^2}{1 - \rho_{(a)j}^2}, \quad j = 1, 2, \ldots, m_{(a)}; \ a = 1, 2, \ldots, A.$$

To derive important common factors, the above theorem can be used. In many of the data in factor analysis, manifest variables can be classified into subsets that have common concepts (factors) to be measured. For example, in the data used for Table 1, it is meaningful to classify the five variables into two subsets $X^{(1)} = (X_1, X_2, X_3)$ and $X^{(2)} = (X_4, X_5)$, where the first subset is related to the liberal arts and the second one is related to the sciences. In $\left(X^{(1)}, \xi\right)$ and $\left(X^{(2)}, \xi\right)$, it is possible to derive the latent ability for the liberal arts and that for the sciences, respectively.

4.1. Numerical Example 1 (Continued)

For $\left(X^{(1)}, \xi\right)$ and $\left(X^{(2)}, \xi\right)$, two sets of canonical variables are obtained, respectively, as follows:

$$
\begin{cases}
\eta_1^{(1)} = 0.95\xi_1 + 0.32\xi_2, \; V_1^{(1)} = 0.52X_1 + 0.76X_2 + 0.39X_3, \; \rho_{(1)1}^2 = 0.77, \\
\eta_2^{(1)} = 0.32\xi_1 - 0.95\xi_2, \; V_2^{(1)} = 0.71X_1 - 0.07X_2 - 0.71X_3, \; \rho_{(1)2}^2 = 0.12,
\end{cases}
$$

$$
\begin{cases}
\eta_1^{(2)} = 0.06\xi_1 + 1.00\xi_2, \; V_1^{(2)} = 0.18X_4 + 0.98X_5, \; \rho_{(2)1}^2 = 0.97, \\
\eta_2^{(2)} = 1.00\xi_1 - 0.06\xi_2, \; V_2^{(2)} = 0.83X_4 - 0.55X_5, \; \rho_{(2)2}^2 = 0.03.
\end{cases}
$$

According to the above canonical variables, we have the following factor contributions:

$$
\begin{cases}
C\left(\eta_1^{(1)} \to X^{(1)}\right) = C\left(\eta_1^{(1)} \to V_1^{(1)}\right) = \dfrac{0.77}{1-0.77} = 3.27, \; CR\left(\eta_1^{(1)} \to X^{(1)}\right) = 0.96, \\
C\left(\eta_2^{(1)} \to X^{(1)}\right) = \dfrac{0.12}{1-0.12} = 0.14. \; CR\left(\eta_2^{(1)} \to X^{(1)}\right) = 0.04;
\end{cases}
$$

$$
\begin{cases}
C\left(\eta_1^{(2)} \to X^{(2)}\right) = 6.14, \; CR\left(\eta_1^{(2)} \to X^{(2)}\right) = 0.97, \\
C\left(\eta_2^{(2)} \to X^{(2)}\right) = 0.17, \; CR\left(\eta_2^{(2)} \to X^{(2)}\right) = 0.03.
\end{cases}
$$

From the above results, canonical factors $\eta_1^{(1)}$ and $\eta_1^{(2)}$ can be interpreted as general common factors for the liberal arts and for the sciences, respectively. By using the factors, the factor loadings are given in Table 2. In this case, Table 2 is similar to Table 1; however, the factor analysis model is oblique and the correlation coefficient between $\eta_1^{(1)}$ and $\eta_1^{(2)}$ is 0.374. The contributions of the factors to manifest variable vector $X = (X_1, X_2, X_3, X_4, X_5) = \left(X^{(1)}, X^{(2)}\right)$ are calculated as follows:

$$
\begin{cases}
C\left(\eta_1^{(1)} \to X\right) = 6.563, \; CR\left(\eta_1^{(1)} \to X\right) = 0.687, \; \widetilde{CR}\left(\eta_1^{(1)} \to X\right) = 0.60, \\
C\left(\eta_1^{(2)} \to X\right) = 4.223. \; CR\left(\eta_1^{(2)} \to X\right) = 0.442, \; \widetilde{CR}\left(\eta_1^{(2)} \to X\right) = 0.39.
\end{cases}
$$

In this case, factors $\eta_1^{(1)}$ and $\eta_1^{(2)}$ are correlated, so it follows that

$$
CR\left(\eta_1^{(1)} \to X\right) + CR\left(\eta_1^{(2)} \to X\right) = 1.129 > 1.
$$

Table 2. Factor loadings by using canonical common factors $\eta_1^{(1)}$ and $\eta_1^{(2)}$.

	X_1	X_2	X_3	X_4	X_5
$\eta_1^{(1)}$	0.62	0.80	0.70	0.31	−0.06
$\eta_1^{(2)}$	0.19	−0.02	−0.22	0.49	0.94
uniqueness	0.50	0.38	0.58	0.55	0.16

4.2. Numerical Example 2

Table 3 shows the results of the maximum likelihood factor analysis (orthogonal) for six scores X_i, $i = 1, 2, \ldots, 6$ ([11], pp. 61–65); such results are treated as the initial estimates in the present analysis. In this example, variables are classified into the following three groups: variable X_1 is related to the Spearman's g factor; variables X_2, X_3, and X_4 account for problem-solving ability; and variables X_5 and X_6 are associated with verbal ability [11]; however, it is difficult to explain the three factors by using Table 3. In this example, the

present approach is employed for deriving the three factors. From (10) and Table 3, the correlation matrix of the manifest variables is given as follows:

$$
\hat{\Sigma} = \begin{pmatrix}
1 & 0.417 & 0.576 & 0.312 & 0.576 & 0.514 \\
0.417 & 1 & 0.567 & 0.306 & 0.265 & 0.263 \\
0.576 & 0.567 & 1 & 0.427 & 0.355 & 0.354 \\
0.312 & 0.306 & 0.427 & 1 & 0.193 & 0.193 \\
0.576 & 0.265 & 0.355 & 0.193 & 1 & 0.799 \\
0.514 & 0.263 & 0.354 & 0.193 & 0.799 & 1
\end{pmatrix}.
$$

Let $X^{(2)} = (X_2, X_3, X_4)$, let $X^{(3)} = (X_5, X_6)$, and let $\xi = (\xi_1, \xi_2)$. Canonical correlation analysis is carried out for (X_1, ξ), $\left(X^{(2)}, \xi\right)$, and $\left(X^{(3)}, \xi\right)$, and we have the following canonical variables, respectively:

$$
\eta_1^{(1)} = \frac{1}{\sqrt{0.67^2 + 0.37^2}}(0.64\xi_1 + 0.37\xi_2) = 0.87\xi_1 + 0.50\xi_2, \ V_1^{(1)} = X_1, \ \rho_{(1)1}^2 = 0.55
$$

$$
\begin{cases}
\eta_1^{(2)} = 0.52\xi_1 + 0.85\xi_2, \ V_1^{(2)} = 0.24X_2 + 0.96X_3 + 0.14X_4, \ \rho_{(2)1}^2 = 0.83, \\
\eta_2^{(2)} = 0.85\xi_1 - 0.52\xi_2, \ V_2^{(2)} = 0.81X_2 - 0.59X_3 - 0.02X_4, \ \rho_{(1)2}^2 = 0.00,
\end{cases}
$$

$$
\begin{cases}
\eta_1^{(3)} = 0.99\xi_1 - 0.12\xi_2, \ V_1^{(3)} = 0.99X_5 + 0.11X_6, \ \rho_{(3)1}^2 = 0.96, \\
\eta_2^{(3)} = 0.12\xi_1 + 0.99\xi_2, \ V_2^{(3)} = 0.64X_5 - 0.77X_6, \ \rho_{(3)2}^2 = 0.01.
\end{cases}
$$

The contributions of canonical factors $\eta_i^{(k)}$, $i = 1, 2; k = 2.3$ are calculated as follows:

$$
\begin{cases}
C\left(\eta_1^{(2)} \to X^{(2)}\right) = C\left(\eta_1^{(2)} \to V_1^{(2)}\right) = \dfrac{0.83}{1 - 0.83} = 4.88, \ CR\left(\eta_1^{(2)} \to X^{(2)}\right) = 1.00, \\
C\left(\eta_2^{(2)} \to X^{(2)}\right) = \dfrac{0.00}{1 - 0.00} = 0.00. \ CR\left(\eta_2^{(1)} \to X^{(1)}\right) = 0.00;
\end{cases}
$$

$$
\begin{cases}
C\left(\eta_1^{(3)} \to X^{(3)}\right) = \dfrac{0.96}{1 - 0.96} = 24.00, \ CR\left(\eta_1^{(3)} \to X^{(3)}\right) = 0.99, \\
C\left(\eta_2^{(2)} \to X^{(2)}\right) = 0.01, \ CR\left(\eta_2^{(2)} \to X^{(2)}\right) = 0.01.
\end{cases}
$$

The common factor $\eta_1^{(1)} (= g)$ can be interpreted as the Spearman's g factor (general intelligence) and canonical common factors $\eta_1^{(2)}$ and $\eta_1^{(3)}$ can be interpreted as problem-solving ability and verbal ability, respectively. The correlation coefficients between the three factors are given by

$$
\mathrm{Corr}\left(g, \eta_1^{(2)}\right) = 0.88, \ \mathrm{Corr}\left(g, \eta_1^{(3)}\right) = 0.80, \ \mathrm{Corr}\left(\eta_1^{(2)}, \eta_1^{(3)}\right) = 0.42.
$$

The contributions of the above three factors to manifest variable vector $X = (X_1, X_2, X_3, X_4, X_5, X_6)$ are computed as follows:

$$
\begin{cases}
C(g \to X) = 19.93, & CR(g \to X) = 0.68 & \widetilde{CR}(g \to X) = 0.66 \\
C\left(\eta_1^{(2)} \to X\right) = 9.77, & CR\left(\eta_1^{(2)} \to X\right) = 0.33, & \widetilde{CR}\left(\eta_1^{(2)} \to X\right) = 0.32, \\
C\left(\eta_1^{(3)} \to X\right) = 25.02, & CR\left(\eta_1^{(3)} \to X\right) = 0.85, & \widetilde{CR}\left(\eta_1^{(3)} \to X\right) = 0.82.
\end{cases}
$$

The common-factor space is two-dimensional, and the factor loadings with common factors $\eta_1^{(2)}$ and $\eta_1^{(3)}$ are calculated as in Table 4. The table shows a clear interpretation of the common factors. Thus, the present method is effective for deriving interpretable factors in situations such as that of this example. The expressions of the factor analysis model can

also be given by factor vectors $\left(g, \eta_1^{(2)}\right)$ and $\left(g, \eta_1^{(3)}\right)$, respectively. The present method is applicable for any subsets of manifest variables.

Table 3. The initial maximum likelihood estimates of factor loadings (varimax).

	X_1	X_2	X_3	X_4	X_5	X_6
ξ_1	0.64	0.34	0.46	0.25	0.97	0.82
ξ_2	0.37	0.54	0.76	0.41	−0.12	−0.03
uniqueness	0.45	0.59	0.21	0.77	0.04	0.33

Table 4. The factor loadings with common factors $\eta_1^{(2)}$ and $\eta_1^{(3)}$.

	X_1	X_2	X_3	X_4	X_5	X_6
$\eta_1^{(2)}$	0.49	0.63	0.89	0.48	−0.01	0.07
$\eta_1^{(3)}$	0.39	0.01	0.00	0.00	0.98	0.79
uniqueness	0.45	0.59	0.21	0.77	0.04	0.33

5. Discussion

In order to find interpretable common factors in factor analysis models, methods of factor rotation are often used. The methods are based on maximizations of variation functions of squares of factor loadings, and orthogonal or oblique factors are applied. The factors derived by the conventional methods may be interpretable; however, it may be more useful to propose a method for detecting interpretable common factors based on factor contribution measurement, i.e., importance of common factors. An entropy-based method for measuring factor contribution [6] can measure the contribution of the common-factor vector to the manifest variable vector, and one can decompose such a contribution into those of single manifest variables (Theorem 1) and into that of some manifest variable subvectors as well (Theorem 2). A characterization in the case of orthogonal factors can be also given (Theorem 3). The paper shows that the most important common factor with respect to entropy can be identified by using canonical correlation analysis between the factor vector and the manifest variable vector (Theorem 4). Theorem 4 shows that the contribution of the common-factor vector to the manifest variable vector can be decomposed into those of canonical factors and that the order of canonical correlation coefficients is that of factor contributions. In most multivariate data, manifest variables can be naturally classified into subsets according to common concepts as in Examples 1 and 2. By using Theorems 2 and 5, canonical correlation analysis can also be applied to derive canonical common factors from subsets of manifest variables and the initial common-factor vector (Theorem 6). According to the analysis, interpretable common factors can be obtained easily, as demonstrated in Examples 1 and 2. In Example 1, Tables 1 and 2 have similar factor patterns; however, the derived factors in Table 1 are orthogonal and those in Table 2 are oblique. In Example 2, it may be difficult to interpret the factors in Table 3 produced by the varimax method. On the other hand, Table 4, obtained by using the present method, can be interpreted clearly. Finally, according to Theorem 5, the present method produces results that are invariant with respect to linear transformations of common factors, so that the method is independent of the initial common factors. The present method is the first one to derive interpretable factors based on a factor contribution measure, and the interpretable factors can be obtained easily through canonical correlation analysis between manifest variable subvectors and the factor vectors.

Author Contributions: Conceptualization, N.E.; methodology, N.E., C.G.B., M.T., T.K.; formal analysis, N.E.; writing—original draft preparation, N.E.; writing—review and editing, C.G.B.; funding acquisition, T.K. All authors have read and agreed to the published version of the manuscript.

References

1. Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200. [CrossRef]
2. Ten Berge, J.M.F. A joint treatment of VARIMAX rotation and the problem of diagonalizing symmetric matrices simultaneously in the least-squares sense. *Psychometrika* **1984**, *49*, 347–358. [CrossRef]
3. Jennrich, R.I.; Sampson, P.F. Rotation for simple loadings. *Psychometrika* **1966**, *31*, 313–323. [CrossRef] [PubMed]
4. Harris, C.W.; Kaiser, H.F. Oblique factor analytic solutions by orthogonal transformation. *Psychometrika* **1964**, *29*, 347–362. [CrossRef]
5. Thurstone, L.L. *Vector of Mind: Multiple Factor Analysis for the Isolation of Primary Traits*; University of Chicago Press: Chicago, IL, USA, 1935.
6. Eshima, N.; Tabata, M.; Borroni, C.G. An entropy-based approach for measuring factor contributions in factor analysis models. *Entropy* **2018**, *20*, 634. [CrossRef] [PubMed]
7. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear model. *J. R. Stat. Soc. A* **1972**, *135*, 370–384. [CrossRef]
8. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
9. Eshima, N.; Tabata, M. Entropy coefficient of determination for generalized linear models. *Comput. Stat. Data Anal.* **2010**, *54*, 1381–1389. [CrossRef]
10. Rao, C.R. Estimation and tests of significance in factor analysis. *Psychometrika* **1955**, *20*, 93–111. [CrossRef]
11. Bartholomew, D.J. *Latent Variable Models and Factor Analysis*; Oxford University Press: New York, NY, USA, 1987.