An Entropy-Based Approach for Measuring Factor Contributions in Factor Analysis Models

Nobuoki Eshima 1,*, Minoru Tabata 2 and Claudio Giovanni Borroni 3

1 Center for Educational Outreach and Admissions, Kyoto University, Kyoto 606-8501, Japan
2 Department of Mathematical Sciences, Osaka Prefecture University, Osaka 599-8532, Japan; mnrtabata@luck.ocn.ne.jp
3 Department of Statistics and Quantitative Methods, University of Milano Bicocca, 20126 Milano, Italy; claudio.borroni@unimib.it

* Correspondence: eshima.nobuoki.2z@kyoto-u.ac.jp; Tel.: +81-75-753-2517

Received: 12 July 2018; Accepted: 20 August 2018; Published: 24 August 2018

Abstract: In factor analysis, factor contributions of latent variables are assessed conventionally by the sums of the squared factor loadings related to the variables. First, the present paper considers issues in the conventional method. Second, an alternative entropy-based approach for measuring factor contributions is proposed. The method measures the contribution of the common factor vector to the manifest variable vector and decomposes it into contributions of factors. A numerical example is also provided to demonstrate the present approach.

Keywords: entropy coefficient of determination; factor contribution; factor loading; path analysis

1. Introduction

Factor analysis is a statistical method for extracting simple structures to explain inter-relations between manifest and latent variables. The origin dates back to the works of [1], and the single factor model was extended to the multiple factor model [2]. These days, factor analysis is widely applied in behavioral sciences [3]; hence, it is important to interpret the extracted factors and is critical to explain how such factors influence manifest variables, that is, measurement of factor contribution. Let \( X_i \) be manifest variables; \( \xi_j \) latent variables (common factors); \( \varepsilon_i \) unique factors related to \( X_i \); and let \( \lambda_{ij} \) be factor loadings that are weights of factors \( \xi_j \) 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

\[
E(X_i) = E(\xi_j) = E(\varepsilon_i) = 0, \quad \text{var}(\xi_j) = 1, \quad \text{cov}(\xi_j, \varepsilon_i) = 0, \quad \text{cov}(\varepsilon_i, \varepsilon_k) = 0 \text{ for } i \neq k \quad \text{and} \quad \text{var}(\varepsilon_i) = \sigma_i^2 > 0.
\]

For the simplicity of discussion, common factors \( \xi_j \) are assumed to be mutually independent in this section, that is, we first consider an orthogonal factor analysis model. In the conventional approach, the contribution of factor \( \xi_j \) to all manifest variables \( X_i, C_j \), is defined as follows:

\[
C_j = \sum_{i=1}^{p} \text{cov}(X_i, \xi_j)^2 = \sum_{i=1}^{p} \lambda_{ij}^2 \tag{2}
\]

The above definition of factor contributions is based on the following decomposition of the total of variances of the observed variables \( X_i \) [4] (p. 59):

\[
\sum_{i=1}^{p} \text{var}(X_i) = \sum_{j=1}^{m} \sum_{i=1}^{p} \lambda_{ij}^2 + \sum_{i=1}^{p} \sigma_i^2
\]
What physical meaning does the above quantity have? Applying it to the manifest variables observed, however, such a decomposition leads to scale-variant results. For this reason, factor contribution is usually considered on the standardized versions of manifest variables $X_i$. What does it mean to measure factor contributions by (2)? For standardized manifest variables $X_i$, we have

$$\lambda_{ij} = \text{cor}(X_i, \xi_j)$$

(3)

Then, (2) is the sum of the coefficients of determination for all standardized manifest variables $X_i$ with respect to a single latent variable $\xi_j$. The squared correlation coefficients (3), that is, $\text{cor}(X_i, \xi_j)^2$, are the ratios of explained variances of a manifest variable $X_i$, and in this sense, they can be interpreted as the contributions (effects) of factors $\xi_j$ to the manifest variable $X_i$. Although, what does the sum of these with respect to all manifest variables $X_i$, that is, (2), mean? The conventional method may be intuitively reasonable for measuring factor contributions; however, we think it is sensible to propose a method measuring factor contributions as the effects of factors on the manifest variable vector $X = (X_1, X_2, \ldots, X_p)$, which are interpretable and have a theoretical basis. There is no research on this topic as far as we have searched. The present paper provides an entropy-based solution to the problem. Entropy is a useful concept to measure the uncertainty in the systems of random variables and sample spaces [5] and it can be applied to measure multivariate dependences of random variables [6,7].

This paper proposes an entropy-based method for measuring factor contributions of $\xi_j$ to the manifest variable vector $X = (X_1, X_2, \ldots, X_p)$ concerned, which can treat not only orthogonal factors, but also oblique cases. The present paper has five sections in addition to this section. In Section 2, the conventional method for measuring factor contributions is reviewed. Section 3 considers the factor analysis model in view of entropy and makes a preliminary discussion on measurement of factor contribution. In Section 4, an entropy-based path analysis is applied as a tool to measure factor contributions. Contributions of factors $\xi_j$ are defined by the total effects of the factors on the manifest variable vector, and the contributions are decomposed into those to manifest variables and subsets of manifest variables. Section 5 illustrates the present method using a numerical example. Finally, in Section 6, some conclusions are provided.

### 2. Relative Factor Contributions in the Conventional Method

In the conventional approach, for the orthogonal factor model (1), the contribution ratio of $\xi_j$ is defined by

$$\text{RC}_j = \frac{C_j}{\sum_{i=1}^{m} C_i} = \frac{\sum_{i=1}^{p} \lambda_{ij}^2}{\sum_{i=1}^{m} \sum_{k=1}^{p} \lambda_{ik}^2}$$

(4)

The above measure is referred to as the factor contribution ratio in the common factor space. Let $R_i$ be the multiple correlation coefficient of latent variable vector $\xi = (\xi_1, \xi_2, \ldots, \xi_m)^T$ and manifest variable $X_i$. Then, for standardized manifest variable $X_i$, we have

$$R_i^2 = \sum_{j=1}^{m} \lambda_{ij}^2$$

(5)

The above quantity can be interpreted as the effect (explanatory power) of latent variable vector $\xi = (\xi_j)$ on manifest variable $X_j$; however, the denominator of (4) is the sum of those effects (5) and there is no theoretical basis to interpret it. Another contribution ratio of $\xi_j$ is referred to as that in the whole space of $X = (X_j)$, and is defined by

$$\text{RC}\_C_j = \frac{C_j}{\sum_{i=1}^{p} \text{var}(X_i)} = \frac{\sum_{i=1}^{p} \lambda_{ij}^2}{\sum_{k=1}^{p} [(\sum_{i=1}^{m} \lambda_{ik}^2 + \sigma_k^2)]}$$

(6)

If the manifest variables are standardized, we have
where \( \tilde{X} \) = \((X_i)\). Indeed, it is the sum of the variances of manifest variables and does not include covariances between them. In the next section, the factor analysis model (1) is reconsidered in the framework of generalized linear models (GLMs), and the effects (contributions) of latent variables \( \xi_j \) on the manifest variable vector \( X = (X_i) \), that is, factor contributions, are discussed through entropy [8].

3. Factor Analysis Model and Entropy

It is assumed that factors \( \varepsilon_i \) and \( \xi_j \) are normally distributed, and the factor analysis model (1) is reconsidered in the GLM framework. Let \( \Lambda = (\lambda_{ij}) \) be a \( p \times m \) factor loading matrix; let \( \Phi \) be an \( m \times m \) correlation matrix of common factor vector \( \xi = (\xi_1, \xi_2, \ldots, \xi_m)^T \); and let \( \Omega \) be the \( p \times p \) variance-covariance matrix of unique factor vector \( \varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T \). The conditional density function of \( X \) given \( \xi \), \( f(x|\xi) \), is normal with mean \( \Lambda \xi \) and variance matrix \( \Omega \), and is given as follows:

\[
f(x|\xi) = \frac{1}{(2\pi)^{p/2} |\Omega|^{1/2}} \exp \left( \frac{X^T \Omega \Lambda \xi - \frac{1}{2} \xi^T \Omega^T \Lambda^2 \xi}{|\Omega|} - \frac{1}{2} X^T \Omega X \right)
\]

where \( \hat{\Omega} \) is the cofactor matrix of \( \Omega \). Let \( f(x) \) and \( g(\xi) \) be the marginal density functions of \( X \) and \( \xi \), respectively. Then, a basic predictive power measure for GLMs [9] is based on the Kullback–Leibler information [6], and applying it to the above model, we have

\[
\text{KL}(X, \xi) = \iint f(x|\xi)g(\xi) \log \frac{f(x|\xi)}{f(x)} \, dx \, d\xi + \iint f(x)g(\xi) \log \frac{f(x)}{f(x|\xi)} \, dx \, d\xi = \frac{\text{tr} \hat{\Omega} \Lambda \Phi \Lambda^T}{|\Omega|} \tag{7}
\]

The above measure was derived from a discussion on log odds ratios in GLMs [9], and is scale-invariant with respect to manifest variables \( X_i \). The numerator of (7) is the explained entropy of \( X \) by \( \xi \), and the denominator is the dispersion of the unique factors in entropy, that is, the generalized variance of \( \varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T \). Thus, (7) expresses the total effect (contribution) of factor vector \( \xi = (\xi_j) \) on manifest variable vector \( X = (X_i) \) in entropy, and is denoted by \( \text{C}(\xi \rightarrow X) \) in the present paper. The entropy coefficient of determination (ECD) is calculated as follows [9]:

\[
\text{ECD}(X, \xi) = \frac{\text{tr} \hat{\Omega} \Lambda \Phi \Lambda^T}{\text{tr} \hat{\Omega} \Lambda \Phi \Lambda^T + |\Omega|} \tag{8}
\]

The denominator of the above measure is interpreted as the variation of manifest variable vector \( X = (X_i) \) in entropy and the numerator is the explained variation of random vector \( X \) in entropy. In this sense, ECD (8) is the factor contribution ratio of \( \xi = (\xi_j) \) for the whole entropy space of \( X = (X_i) \), and it expresses the standardized total effect of \( \xi = (\xi_1, \xi_2, \ldots, \xi_m)^T \) on the manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \), which is denoted by \( \varepsilon_T(\xi \rightarrow X) \) [8,10]. As for (6), in the present paper, the ECD is denoted by \( \text{RC}(\xi \rightarrow X) \), that is, the relative contribution of factor vector \( \xi \) for the whole space of manifest variable \( X \) in entropy.

Remark 1. Let \( \Sigma \) be the \( p \times p \) variance-covariance matrix of manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \) and let \( \Phi \) be the \( m \times m \) correlation matrix of \( \xi \). Then, we have

\[
\Sigma = \Lambda \Phi \Lambda^T + \Omega \tag{9}
\]
For assessing the goodness-of-fit of the models, the following overall coefficient of determination (OCD) is suggested ([11], p. 60) on the basis of (9):

\[
\text{OCD}(X, \xi) = 1 - \left| \frac{\Omega}{\Sigma} \right| = \left| \frac{\Sigma - \Omega}{\Sigma} \right|
\]

Determinant \(|\Omega|\) is the generalized variance of unique factor vector \(\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T\) and \(|\Sigma|\) is that of manifest variable vector \(X = (X_1, X_2, \ldots, X_p)^T\). Then, OCD is interpreted as the ratio of the explained generalized variance of manifest variable vector \(X = (X_1, X_2, \ldots, X_p)^T\) by common factor vector \(\xi = (\xi_1, \xi_2, \ldots, \xi_m)^T\) in the \(p\)-dimensional Euclidian space. On the other hand, from (8), it follows that

\[
\text{ECD}(X, \xi) = 1 - \frac{|\Omega| \text{tr} \tilde{\Omega} \Lambda \Phi \Lambda^T + |\Omega|}{|\Omega|}
\]

Hence, ECD is viewed as the ratio of the explained variation of the manifest variable vector in entropy.

Cofactor matrix \(\tilde{\Omega}\) is diagonal and the \((i, i)\) elements are \(\prod_{k \neq i} \sigma_k^2\), \(i = 1, 2, \ldots, p\). If common factors are statistically independent, it follows that

\[
\text{tr} \tilde{\Omega} \Lambda \Phi \Lambda^T = \sum_{i=1}^{p} \prod_{k \neq i} \sigma_k^2 \sum_{j=1}^{m} \lambda_{ij}^2 = \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_{ij}^2 \prod_{k \neq i} \sigma_k^2
\]

Thus, (7) is decomposed as

\[
\text{KL}(X, \xi) = \sum_{i=1}^{p} \sum_{j=1}^{m} \frac{\lambda_{ij}^2}{\sigma_i^2}
\]

As detailed below, in the present paper, the contribution of factor \(\xi_j\) to \(X\), \(C(\xi_j \rightarrow X)\), is defined by

\[
C(\xi_j \rightarrow X) = \sum_{i=1}^{p} \frac{\lambda_{ij}^2}{\sigma_i^2}
\]

**Remark 2.** The above contribution is different from the conventional definition of factor contribution (2); unless \(\sigma_i^2 = 1\), \(i = 1, 2, \ldots, p\). In this sense, we may say that the standardization of manifest variables in entropy is obtained by setting all the unique factor variances to one.

In the next section, the contributions (effects) of factors \(\xi_j\) to manifest variable vector \(X\) are discussed in a general framework through an entropy-based path analysis [8].

### 4. Measurement of Factor Contribution Based on Entropy

A path diagram for the factor analysis model is given in Figure 1, in which the single-headed arrows imply the directions of effects of factors and the double-headed curved arrows indicate the associations between the related variables. In this section, common factors are assumed to be correlated, that is, we consider an oblique case, and an entropy-based path analysis [8] is applied to make a general discussion in the measurement of factor contributions.
In the factor analysis model (1),

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

**Proof.** Let $f_i(x_i|\xi)$ be the conditional density functions of manifest variables $X_i$, given factor vector $\xi$; let $f_i(x_i)$ be the marginal density functions of $X_i$; let $f(x)$ be the marginal density function of $X$; and let $g(\xi)$ be the marginal density function of common factor vector $\xi$. As the manifest variables are conditionally independent, given factor vector $\xi$, the conditional density function of $X$ is

$$f(x|\xi) = \prod_{i=1}^{P} f_i(x_i|\xi)$$

From (7), we have

$$KL(X, \xi) = \iiint \prod_{i=1}^{P} f_i(x_i|\xi)g(\xi) \log \frac{\prod_{k=1}^{P} f_k(x_k|\xi)}{f(x)} dx d\xi + \iint f(x)g(\xi) \log \frac{f(x)}{\prod_{k=1}^{P} f_k(x_k|\xi)} dx d\xi$$

$$= \iint \left( \prod_{i=1}^{P} f_i(x_i|\xi) - f(x) \right)g(\xi) \log \prod_{k=1}^{P} f_k(x_k|\xi) dx d\xi$$

$$= \iint \prod_{i=1}^{P} f_i(x_i|\xi)g(\xi) \log \frac{\prod_{k=1}^{P} f_k(x_k|\xi)}{f(x)} dx d\xi + \iint f(x)g(\xi) \log \frac{f(x)}{\prod_{k=1}^{P} f_k(x_k|\xi)} dx d\xi$$

$$= \sum_{k=1}^{P} \int f_k(x_k|\xi)g(\xi) \log \frac{f_k(x_k|\xi)}{f_k(x_k)} dx_k d\xi + \sum_{k=1}^{P} \int f_k(x_k|\xi)g(\xi) \log \frac{f_k(x_k|\xi)}{f_k(x_k)} dx_k d\xi$$

$$= \sum_{i=1}^{P} \left( \int f_i(x_i|\xi)g(\xi) \log \frac{f_i(x_i|\xi)}{f_i(x_i)} dx_i d\xi + \int f_i(x_i|\xi)g(\xi) \log \frac{f_i(x_i)}{f_i(x_i|\xi)} dx_i d\xi \right) = \sum_{i=1}^{P} KL(X_i, \xi)$$

\[\square\]

In model (1) with correlation matrix $\Phi = (\varphi_{ij})$, we have

$$KL(X_i, \xi) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \varphi_{ij} \lambda_j}{\sigma_i^2}$$

The above quantity is referred to as the contribution of $\xi$ to $X_i$, and is denoted as $C(\xi \rightarrow X_i)$. Let $R_i$ be the multiple correlation coefficient of $X_i$ and $\xi = (\xi_i)$. Then,
\[ C(\xi \to X_i) = \frac{R_i^2}{1 - R_i^2} (= \text{KL}(X_i, \xi)) \]  
(11)

From Theorem 1, we then have

\[ C(\xi \to X) = \sum_{i=1}^{p} \frac{R_i^2}{1 - R_i^2} (= \text{KL}(X, \xi)) \]  
(12)

Hence, Theorem 1 gives the following decomposition of the contribution of \( \xi \) on \( X \) into those on the single manifest variables \( X_i \) (11):

\[ C(\xi \to X) = \sum_{i=1}^{p} C(\xi \to X_i) \]  
(13)

**Remark 3.** Notice that in the denominator of (4), the total contribution of all factors \( \xi_i \) is simply defined as the total sum assessed:

\[ \sum_{i=1}^{m} C_i = \sum_{i=1}^{p} R_i^2 \]

On the other hand, in the present approach, the total effect (contribution) of factor vector \( \xi \) on manifest variable vector \( X \) is decomposed into those of manifest variables \( X_i \), (12) and (13).

Let \( X_{sub} \) be any sub-vector of manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \). Then, the contribution of factor vector \( \xi \) to \( X_{sub} \) is defined by

\[ C(\xi \to X_{sub}) = \text{KL}(X_{sub}, \xi) \]

From Theorem 1, we have the following corollary.

**Corollary 1.** Let \( X_{(1)} = (X_{i_1}, X_{i_2}, \ldots, X_{i_q})^T \) and \( X_{(2)} = (X_{j_1}, X_{j_2}, \ldots, X_{j_{p-q}})^T \) be a decomposition of manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \), where \( q < p \). Then, for factor analysis model (1), it follows that

\[ C(\xi \to X) = C(\xi \to X_{(1)}) + C(\xi \to X_{(2)}) \]

\[ C(\xi \to X_{(1)}) = \sum_{k=1}^{q} C(\xi \to X_{k}), \quad C(\xi \to X_{(2)}) = \sum_{k=q+1}^{p} C(\xi \to X_{k}) \]

**Proof:** From a similar discussion to the proof of Theorem 1, we have

\[ \text{KL}(X, \xi) = \text{KL}(X_{(1)}, \xi) + \text{KL}(X_{(2)}, \xi) \]

\[ \text{KL}(X_{(1)}, \xi) = \sum_{k=1}^{q} \text{KL}(X_{k}, \xi), \quad \text{KL}(X_{(2)}, \xi) = \sum_{k=q+1}^{p} \text{KL}(X_{k}, \xi) \]

Hence, the corollary follows.

Next, the standardized total effects of single factors \( \xi_j \) on manifest variable vector \( X \), that is, \( e_T(\xi_j \to X) \), are calculated \([8,10]\). Let \( \xi^{/j} = (\xi_1, \xi_2, \ldots, \xi_{j-1}, \xi_{j+1}, \ldots, \xi_m)^T \); \( f(x, \xi^{/j} | \xi_j) \) be the conditional density function of \( X \) and \( \xi^{/j} \) given \( \xi_j \); \( f(x | \xi_j) \) be the conditional density function of \( X \) given \( \xi_j \); \( g^{/j}(\xi^{/j} | \xi_j) \) be the conditional density function of \( \xi^{/j} \) given \( \xi_j \); and \( g_j(\xi_j) \) be the marginal density function of \( \xi_j \). Then, we have

\[
\text{KL}(X, \xi^{/j} | \xi_j) = \int \int f(x, \xi) \log \frac{f(x, \xi^{/j} | \xi_j)}{f(x | \xi_j) g^{/j}(\xi^{/j} | \xi_j)} \, dx \, d\xi^{/j} / d\xi_j \\
+ \int \int f(x | \xi_j) g^{/j}(\xi^{/j} | \xi_j) g_j(\xi_j) \log \frac{f(x | \xi_j) g^{/j}(\xi^{/j} | \xi_j)}{f(x, \xi^{/j} | \xi_j)} \, dx \, d\xi^{/j} / d\xi_j = \frac{\text{tr}(\Omega \Lambda^{\text{cov}}(X \xi | \xi_j))}{|\Omega|} 
\]
where \( \text{cov}(\xi, X|\tilde{\xi}_j) \) is a \( m \times p \) covariance matrix given \( \tilde{\xi}_j \), of which the \((k, i)\) elements are \( \text{cov}(\tilde{\xi}_k, X_i|\tilde{\xi}_j) \).

The standardized total effect \( e_T(\tilde{\xi}_j \rightarrow X) \) is given by

\[
e_T(\tilde{\xi}_j \rightarrow X) = \frac{\text{KL}(X, \xi) - \text{KL}(X, \xi^{j/}|\tilde{\xi}_j)}{\text{KL}(X, \xi) + 1} = \frac{\text{tr}\Omega \Lambda (\text{cov}(\xi, X) - \text{cov}(\xi, X|\tilde{\xi}_j))}{\text{tr}\Omega \Lambda \text{cov}(\xi, X) + |\Omega|}
\]

The standardized total effect \( e_T(\tilde{\xi}_j \rightarrow X) \) [8] is interpreted as the contribution ratio of factor \( \tilde{\xi}_j \) in the whole entropy space of \( X \), and in the present paper, it is denoted by \( \tilde{R}C(\tilde{\xi}_j \rightarrow X) \). The contribution of factor \( \tilde{\xi}_j \) measured in entropy is defined by

\[
C(\tilde{\xi}_j \rightarrow X) = \frac{\tilde{R}C(\tilde{\xi}_j \rightarrow X)}{\tilde{R}C(\tilde{\xi} \rightarrow X)} = \frac{C(\tilde{\xi}_j \rightarrow X)}{C(\tilde{\xi} \rightarrow X)}
\]

Concerning factor contributions of \( \tilde{\xi}_j \) on the single manifest variables \( X_i \), that is, \( C(\tilde{\xi}_j \rightarrow X_i) \), the following theorem can be stated.

**Theorem 2.** In the factor analysis model (1),

\[
C(\tilde{\xi}_j \rightarrow X) = \sum_{i=1}^p C(\tilde{\xi}_j \rightarrow X_i)
\]

**Proof:** From Theorem 1, it follows that

\[
\text{KL}(X, \xi^{j/}|\tilde{\xi}_j) = \sum_{i=1}^p \text{KL}(X_i, \xi^{j/}|\tilde{\xi}_j)
\]

Then, we have

\[
C(\tilde{\xi}_j \rightarrow X_i) = \text{KL}(X_i, \xi) - \text{KL}(X_i, \xi^{j/}|\tilde{\xi}_j)
\]

and,

\[
C(\tilde{\xi}_j \rightarrow X) = \text{KL}(X, \xi) - \text{KL}(X, \xi^{j/}|\tilde{\xi}_j)
\]

\[
= \sum_{i=1}^p \text{KL}(X_i, \xi) - \sum_{i=1}^p \text{KL}(X_i, \xi^{j/}|\tilde{\xi}_j)
\]

\[
= \sum_{i=1}^p \left( \text{KL}(X_i, \xi) - \text{KL}(X_i, \xi^{j/}|\tilde{\xi}_j) \right) = \sum_{i=1}^p C(\tilde{\xi}_j \rightarrow X_i)
\]

From the above theorem, we have the following corollary.

**Corollary 2.** Let \( X_{(1)} = (X_{i_1}, X_{i_2}, \ldots, X_{i_q})^T \) and \( X_{(2)} = (X_{i_{q+1}}, X_{i_{q+2}}, \ldots, X_{i_p})^T \) be decomposition of manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \), where \( q < p \).

\[
C(\tilde{\xi}_j \rightarrow X) = C(\tilde{\xi}_j \rightarrow X_{(1)}) + C(\tilde{\xi}_j \rightarrow X_{(2)})
\]

\[
C(\tilde{\xi}_j \rightarrow X_{(1)}) = \sum_{k=1}^{q} C(\tilde{\xi}_j \rightarrow X_{i_k}), \quad C(\tilde{\xi}_j \rightarrow X_{(2)}) = \sum_{k=q+1}^{p} C(\tilde{\xi}_j \rightarrow X_{i_k})
\]

**Proof:** From a similar discussion in the proof of Theorem 2, the corollary follows. \( \square \)

**Remark 4.** Let \( X_{\text{sub}} \) be any sub-vector of manifest variable vector \( X = (X_1, X_2, \ldots, X_p)^T \). By substituting \( X \) for \( X_{\text{sub}} \) in the above discussion, \( C(\xi \rightarrow X_{\text{sub}}) \), \( C(\tilde{\xi}_j \rightarrow X_{\text{sub}}) \), \( \tilde{R}C(\tilde{\xi}_j \rightarrow X_{\text{sub}}) \), and \( \tilde{R}C(\tilde{\xi}_j \rightarrow X_{\text{sub}}) \) can be defined.
For orthogonal factor analysis models, the following theorem holds true.

**Theorem 3.** In factor analysis model (1), if common factors $\xi_j$ are statistically independent, then

$$C(\xi \rightarrow X) = \sum_{j=1}^{m} \sum_{i=1}^{p} C(\xi_j \rightarrow X_i).$$

**Proof**: From model (1), we have

$$C(\xi_j \rightarrow X_i) = KL(X_i, \xi_j) - KL(X_i, \xi_j/|\xi_j|) = \frac{\lambda_{ij}^2}{\sigma_i^2}$$

This completes the theorem. □

From the above discussion, if common factors $\xi_j$ are statistically independent, (10) is derived. Moreover, we have

$$RC(\xi_j \rightarrow X) = \frac{KL(X, \xi_j) - KL(X, \xi_j/|\xi_j|)}{KL(X, \xi_j) + 1} = \frac{\sum_{i=1}^{p} \lambda_{ij}^2}{\sigma_j^2}$$

This measure is the relative contribution ratio of $\xi_j$ for the variation of $X$ in entropy. The relative contributions of $\xi_j$ on $X$ in entropy are calculated as follows:

$$RC(\xi_j \rightarrow X) = \frac{C(\xi_j \rightarrow X)}{C(\xi \rightarrow X)} = \frac{\sum_{i=1}^{p} \lambda_{ij}^2}{\sum_{j=1}^{m} \sum_{i=1}^{p} \lambda_{ij}^2}$$

**Remark 5.** It is difficult to use OCD for assessing factor contributions, because $|\Sigma|$ cannot be decomposed as in the above discussion.

**5. Numerical Example**

In order to illustrate the present method, we use the data shown in Table 1 [12]. In this table, manifest variables $X_1$, $X_2$, and $X_3$ are subjects in liberal arts and variables $X_4$ and $X_5$ are those in sciences. First, orthogonal factor analysis (varimax method by S-PLUS ver. 8.2) is applied to the data and the results are illustrated in Table 2. From the estimated factor loadings, the first factor is interpreted as an ability relating to liberal arts, and the second factor as that for sciences. According to the factor contributions $C(\xi_j \rightarrow X)$ shown in Table 3, the contribution of factor $\xi_2$ is about twice as big than that of factor $\xi_1$ from a view point of entropy, and from the relative contributions $RC(\xi_j \rightarrow X)$, about 30% of variation of manifest variable vector $X$ in entropy is explained by factor $\xi_1$ and about 60% by factor $\xi_2$. The relative contribution $RC(\xi \rightarrow X)$ in Table 3 implies about 90% of the entropy of manifest variable vector $X$ is explained by the two factors. On the other hand, in the conventional method, the measured factor contributions of $\xi_1$ and $\xi_2$, that is, $C_j$, are almost equal (Table 4). As discussed in the present paper, the conventional method is intuitive and does not have any logical foundation for multidimensionally measuring contributions of factors to manifest variable vectors. Table 5 decomposes “the contribution of $\xi$ to $X$” into components $C(\xi_j \rightarrow X_i)$. The contribution of $\xi_2$ to $X_5$ is prominent compared with the other contributions.
Table 1. Data for illustrating factor analysis.

| Subject | Japanese $X_1$ | English $X_2$ | Social $X_3$ | Mathematics $X_4$ | Science $X_5$ |
|---------|----------------|--------------|--------------|-------------------|--------------|
| 1       | 64             | 65           | 83           | 69                | 70           |
| 2       | 54             | 56           | 53           | 40                | 32           |
| 3       | 80             | 68           | 75           | 74                | 84           |
| 4       | 71             | 65           | 40           | 41                | 68           |
| 5       | 63             | 61           | 60           | 56                | 80           |
| 6       | 47             | 62           | 33           | 57                | 87           |
| 7       | 42             | 53           | 50           | 38                | 23           |
| 8       | 54             | 17           | 46           | 58                | 58           |
| 9       | 57             | 48           | 59           | 26                | 17           |
| 10      | 54             | 72           | 58           | 55                | 30           |
| 11      | 67             | 82           | 54           | 67                | 28           |
| 12      | 71             | 82           | 74           | 75                | 53           |
| 13      | 53             | 67           | 74           | 75                | 53           |
| 14      | 90             | 96           | 63           | 87                | 100          |
| 15      | 71             | 69           | 74           | 76                | 42           |
| 16      | 61             | 100          | 92           | 53                | 58           |
| 17      | 61             | 69           | 63           | 63                | 71           |
| 18      | 87             | 84           | 64           | 65                | 53           |
| 19      | 77             | 75           | 78           | 37                | 44           |
| 20      | 57             | 27           | 41           | 54                | 30           |

Table 2. Factor loadings of orthogonal factor analysis ($\chi^2 = 0.55$, $df = 1$, $P = 0.45$).

| $\xi_1$ | $\xi_2$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---------|---------|-------|-------|-------|-------|-------|
| 0.60    | 0.39    | 0.75  | 0.65  | 0.32  | 0.00  |
| 0.50    | 0.24    | 0.38  | 0.58  | 0.55  | 0.16  |

Table 3. Factor contributions based on entropy (orthogonal case).

| $\xi_1$ | $\xi_2$ | Total |
|---------|---------|-------|
| C($\xi_j \rightarrow X$) | 3.11 | 6.23 | 9.34 = C($\xi \rightarrow X$) |
| RC($\xi_j \rightarrow X$) | 0.30 | 0.60 | 0.90 = RC($\xi \rightarrow X$) |
| RC($\xi_j \rightarrow X$) | 0.33 | 0.67 | 1 |

Table 4. Factor contributions with the conventional method.

| $\xi_1$ | $\xi_2$ | Total |
|---------|---------|-------|
| C_j     | 1.44    | 1.39  | 2.83 |
| RC_j    | 0.29    | 0.28  | 0.57 |
| RC_j    | 0.51    | 0.49  | 1 |

Table 5. Decomposition of factor contribution $C(\xi \rightarrow X_j)$ into $C(\xi_j \rightarrow X_j)$.

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Total = $C(\xi_j \rightarrow X)$ |
|-------|-------|-------|-------|-------|----------------------------------|
| $\xi_1$ | 0.72  | 1.49  | 0.72  | 0.19  | 0.00 | 3.11 |
| $\xi_2$ | 0.30  | 0.15  | 0     | 0.63  | 5.14 | 6.23 |
| total = $C(\xi \rightarrow X_j)$ | 1.01  | 1.64  | 0.72  | 0.82  | 5.14 | 9.34 |

From the discussion in the previous section, the contributions of factors are flexibly calculated. For example, it is reasonable to divide the manifest variable vector into $X_{(1)} = (X_1, X_2, X_3)$ and
\( X_2 = (X_4, X_5) \), because the first sub-vector is related to the liberal arts and the second one to the sciences. First, the contributions of \( \xi_1 \) and \( \xi_2 \) to \( X_1 \) are calculated according to the present method, and the details are given as follows:

\[
C\left(\xi_1 \rightarrow X_1(1)\right) = C(\xi_1 \rightarrow X_1) + C(\xi_1 \rightarrow X_2) + C(\xi_1 \rightarrow X_3) = 0.72 + 1.49 + 0.72 = 2.93
\]
\[
C\left(\xi_2 \rightarrow X_1(1)\right) = 0.30 + 0.15 + 0.00 = 0.45
\]
\[
C\left(\xi \rightarrow X_1\right) = C(\xi_1 \rightarrow X_1) + C(\xi_2 \rightarrow X_1(1)) = 2.93 + 0.45 = 3.38
\]
\[
\text{RC}\left(\xi \rightarrow X_1\right) = \frac{C(\xi_1 \rightarrow X_1)}{C(\xi \rightarrow X_1)} + 1 = \frac{3.38}{3.38 + 1} = 0.77
\]
\[
\text{RC}\left(\xi_1 \rightarrow X_1\right) = \frac{C(\xi_1 \rightarrow X_1)}{C(\xi \rightarrow X_1)} + 1 = \frac{2.93}{3.38 + 1} = 0.67
\]
\[
\text{RC}\left(\xi_2 \rightarrow X_1\right) = \frac{0.45}{3.38 + 1} = 0.10
\]
\[
\text{RC}\left(\xi_1 \rightarrow X_1\right) = \frac{C(\xi_1 \rightarrow X_1)}{C(\xi \rightarrow X_1)} + 1 = \frac{2.93}{2.93 + 0.45} = 0.87
\]
\[
\text{RC}\left(\xi_2 \rightarrow X_1\right) = \frac{0.45}{2.93 + 0.45} = 0.13
\]

From (14), 77% of the entropy of manifest variable sub-vector \( X_1 \) are explained by the two factors, in which 67% of that are explained by factor \( \xi_1 \) (15) and 10% by factor \( \xi_2 \) (16). From the relative contributions (17) and (18), 87% of the total contribution of the two factors are made by factor \( \xi_1 \) and 13% by factor \( \xi_2 \).

On the other hand, the contributions of \( \xi_1 \) and \( \xi_2 \) on \( X_2 = (X_4, X_5) \) are calculated as follows:

\[
C\left(\xi_1 \rightarrow X_2\right) = C(\xi_1 \rightarrow X_4) + C(\xi_2 \rightarrow X_3) = 0.19 + 0.00 = 0.19
\]
\[
C\left(\xi_2 \rightarrow X_2\right) = 0.63 + 5.14 = 5.77
\]
\[
C\left(\xi \rightarrow X_2\right) = C(\xi_1 \rightarrow X_2) + C(\xi_2 \rightarrow X_2) = 0.19 + 5.77 = 5.96
\]
\[
\text{RC}\left(\xi \rightarrow X_2\right) = \frac{C(\xi_1 \rightarrow X_2)}{C(\xi \rightarrow X_2)} + 1 = \frac{5.96}{5.96 + 1} = 0.86
\]
\[
\text{RC}\left(\xi_1 \rightarrow X_2\right) = \frac{C(\xi_1 \rightarrow X_2)}{C(\xi \rightarrow X_2)} + 1 = \frac{0.19}{5.96 + 1} = 0.03
\]
\[
\text{RC}\left(\xi_2 \rightarrow X_2\right) = \frac{5.77}{5.96 + 1} = 0.83
\]
\[
\text{RC}\left(\xi_1 \rightarrow X_2\right) = \frac{C(\xi_1 \rightarrow X_2)}{C(\xi \rightarrow X_2)} + 1 = \frac{0.19}{5.96} = 0.03
\]
\[
\text{RC}\left(\xi_2 \rightarrow X_2\right) = \frac{5.77}{5.96} = 0.97
\]

From (19), 86% of entropy of manifest variable sub-vector \( X_2 \) is explained by the two factors, in which 3% of the entropy are explained by factor \( \xi_1 \) (20) and 83% by factor \( \xi_2 \) (21). The contribution ratios of the factors to sub-vector \( X_2 \) are calculated in (22) and (23). Ninety-seven percent of the entropy was made by factor \( \xi_2 \).
Second, factor contributions in an oblique case are calculated. The estimated factor loadings and the correlation matrix of factors based on the covarimin method are shown in Tables 6 and 7, respectively. Based on factor loadings in Table 6, factor $\xi_1$ is interpreted as an ability for subjects in the liberal arts and factor $\xi_2$ as an ability for subjects in sciences. The results are similar to those in the orthogonal case mentioned above, because the correlation between the factors is not strong. Table 8 shows the decomposition of $C(\xi \rightarrow X)$ based on Theorems 1 and 2. In this case, it is noted that $C(\xi \rightarrow X) \neq C(\xi_1 \rightarrow X) + C(\xi_2 \rightarrow X)$; however, $C(\xi \rightarrow X) = \sum_{i=1}^{5} C(\xi_i \rightarrow X)$. According to the table, the contributions of $\xi_1$ and $\xi_2$ to sub-vectors of manifest variable vector $X$ can also be calculated as in the above orthogonal factor analysis. Table 9 illustrates the contributions of factors on manifest variable vector $X$. Factor $\xi_1$ explains 42% of the entropy of $X$ and factor $\xi_2$ explains 71%.

### Table 6. Factor loadings of oblique factor analysis ($\chi^2 = 0.55, df = 1, P = 0.45$).

|       | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|-------|-------|-------|-------|-------|-------|
| $\xi_1$ | 0.59  | 0.77  | 0.68  | 0.29  | 0     |
| $\xi_2$ | 0.24  | 0.00  | $-0.12$ | 0.52  | 0.92  |
| uniqueness | 0.50  | 0.41  | 0.58  | 0.55  | 0.16  |

### Table 7. Correlation matrix of factors.

|       | $\xi_1$ | $\xi_2$ |
|-------|---------|---------|
| $\xi_1$ | 1       | 0.315  |
| $\xi_2$ | 0.315  | 1       |

### Table 8. Decomposition of factor contribution $C(\xi \rightarrow X)$ into $C(\xi_j \rightarrow X_i)$ (oblique case).

|       | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Total = $C(\xi_j \rightarrow X)$ |
|-------|-------|-------|-------|-------|-------|-------------------------------|
| $\xi_1$ | 0.90  | 1.44  | 0.70  | 0.37  | 0.54  | 3.95                          |
| $\xi_2$ | 0.37  | 0.14  | 0.01  | 0.68  | 5.43  | 6.65                          |
| $C(\xi \rightarrow X)$ | 1.01  | 1.44  | 0.73  | 0.82  | 5.43  | $C(\xi \rightarrow X)$ = 9.43 |

### Table 9. Factor contributions based on entropy (oblique case).

|       | $\xi_1$ | $\xi_2$ | Effect of $\xi$ on $X$ |
|-------|---------|---------|------------------------|
| $C(\xi_j \rightarrow X)$ | 3.95  | 6.65  | $C(\xi \rightarrow X)$ = 9.43 |
| $RC(\xi_j \rightarrow X)$ | 0.38  | 0.64  | $RC(\xi \rightarrow X)$ = 0.90 |
| $RC(\xi_j \rightarrow X)$ | 0.42  | 0.71  |                        |

### 6. Discussion

For orthogonal factor analysis models, the conventional method measures factor contributions (effects) by the sums (totals) of squared factor loadings related to the factors (2); however, there is no logical foundation for how they can be interpreted. It is reasonable to measure factor contributions as the effects of factors on the manifest variable vector concerned. The present paper has proposed a method of measuring factor contributions through entropy, that is, applying an entropy-based path analysis approach. The method measures the contribution of factor vector $\xi$ to manifest variable vector $X$ and decomposes it into those of factors $\xi_j$ to manifest variables $X_i$ and/or those to sub-vectors of $X$. Comparing (2) and (10), for standardization of unique factor variances $\sigma^2_i = 1$, the present method equals to the conventional method. As discussed in this paper, the present method can be employed in oblique factor analysis as well, and it has been illustrated in a numerical example. The present method has a theoretical basis for measuring factor contributions in a framework of entropy, and it is a
A novel approach for factor analysis. The present paper confines itself to the usual factor analysis model. A more complicated model with a mixture of normal factor analysis models [13] is excluded, and a further study is needed to apply the entropy-based method to the model.

Author Contributions: N.E. conceived the study; N.E., M.T., and C.G.B. discussed the idea for measuring factor contribution; N.E. and M.T. proven the theorems in the paper; N.E. and C.G.B. computed the numerical example; N.E. wrote the paper, and the coauthors reviewed it.

Funding: Grant-in-aid for Scientific Research 18993038, Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Acknowledgments: The authors would like to thank the referees for their useful comments and suggestions to improve the first version of the present paper.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Spearman, S. “General-intelligence”, objectively determined and measured. Am. J. Psychol. 1904, 15, 201–293. [CrossRef]
2. Thurstone, L.L. Vector of Mind: Multiple Factor Analysis for the Isolation of Primary Traits; The University of Chicago Press: Chicago, IL, USA, 1935.
3. Young, A.G.; Pearce, S. A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. Quant. Methods Psychol. 2013, 9, 79–94. [CrossRef]
4. Bartholomew, D.J. Latent Variable Models and Factor Analysis; Oxford University Press: New York, NY, USA, 1987.
5. Shannon, C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1946, 27, 379–423. [CrossRef]
6. Kullback, S.; Leibler, R.A. On information and sufficiency. Ann. Math. Stat. 1951, 22, 79–86. [CrossRef]
7. Harry, J. Relative entropy measures of multivariate dependence. J. Am. Stat. Assoc. 1989, 84, 157–164.
8. Eshima, N.; Tabata, M.; Borroni, G.C.; Kano, Y. An entropy-based approach to path analysis of structural generalized linear models: A basic idea. Entropy 2015, 17, 5117–5132. [CrossRef]
9. Eshima, N.; Tabata, M. Entropy coefficient of determination for generalized linear models. Comput. Stat. Data Anal. 2010, 54, 1381–1389. [CrossRef]
10. Eshima, N.; Borroni, C.G.; Tabata, M. Relative importance assessment of explanatory variables in generalized linear models: An entropy-based approach. Stat. Appl. 2016, 16, 107–122.
11. Everitt, B.S. An Introduction to Latent Variable Models; Chapman and Hall: London, UK, 1984.
12. Adachi, K.; Trendafilov, N.T. Some mathematical properties of the matrix decomposition solution in factor analysis. Psychometrika 2018, 83, 407–424. [CrossRef] [PubMed]
13. Attias, H. Independent factor analysis. Neural Comput. 1999, 11, 803–851. [CrossRef] [PubMed]