FORMULATION AND DEVELOPMENT OF COMMON FACTOR BATTERY RELIABILITY COEFFICIENTS

We had in passing made a mention of the common factor battery reliability coefficients and also of their necessity in Section 3.1. Some CFBR coefficients have been developed in this chapter. Three of these are based on the ratios of:

(i) two appropriate quadratic distance functions;
(ii) the linear composites of two appropriate matrices; and
(iii) traces of two appropriate matrices.

Two other CFBR coefficients are based on the positive eigenvalues of two appropriate matrices.

In Sections 5.1 and 5.2 we take up approach (i) and approaches (ii) and (iii), respectively. Finally in Section 5.3, we take up the CFBR coefficients based on the positive eigenvalues of two appropriate matrices.

5.1. CFBR Coefficients Based on Ratios of Squared Distance Functions

Let $X_i$ ($i = 1, 2, \ldots, N$) be the $(p \times 1)$ column vector of observed scores on the $i$th individual. Let $C_i, U_i$ ($i = 1, 2, \ldots, N$) denote the corresponding $(p \times 1)$ column vectors of common and unique parts scores of an $i$th individual. Further, we have already
introduced in eq. (2.1.2) the following model:-

\[ X_i = C_i + U_i \]  \hspace{1cm} (5.1.1)

Adopting the usual assumptions of the orthogonal common factor model (cf. Section 2.1), we have the corresponding expression for the correlation matrix as below:-

\[ R = A A' + U^2 \]  \hspace{1cm} (5.1.2)

where the matrices, \( R \), \( U^2 \) and \( A \) have the usual meanings.

Consider the difference vector \( d_{ij} \) between the observed scores of \( i \)th and \( j \)th individuals, i.e.

\[ d_{ij} = X_i - X_j \]

where \( i, j = 1, 2, ..., N; \ i \neq j \).

Using eq. (5.1.1), we further have

\[ d_{ij} = C_i - C_j + U_i - U_j \]  \hspace{1cm} (5.1.3)

Now, consider a metric or a squared distance measure, \( D^2(i,j) \), between the \( i \)th and \( j \)th individuals as below:

\[ D^2(i,j) = d_{ij}^T M d_{ij}, \ i \neq j = 1, 2, ..., N \]  \hspace{1cm} (5.1.4)

where \( M \) is an appropriate weights matrix and is at least positive semidefinite to ensure that \( D^2(i,j) \geq 0 \).
The squared distance measure $D^2(i,j)$ can be interpreted for different types of weight matrices, $M$, as follows:

(i) Let $M = R^{-1}$, where $R^{-1}$ is the inverse of positive definite correlation matrix $R$, so that

$$D^2_1(i,j) = d_{ij}' R^{-1} d_{ij}$$  \hspace{1cm} (5.1.5)

$D^2_1(i,j)$ is analogue of a widely known generalised distance statistic, $D^2$, due to Mahalanobis (1936). Its use has been very widely recommended by Rao (1964).

(ii) Let $M = I$, then

$$D^2_2(i,j) = d_{ij}' d_{ij}$$  \hspace{1cm} (5.1.6)

which is a familiar Euclidean squared distance function. Further, this squared distance function, in fact, is based on the hypothesis of test variables being uncorrelated (Harris, 1955).

In psychometrical literature, these two measures are associated with Cronbach-Gleser (1953) who popularised these as techniques for comparing the individuals on the basis of their vector scores.

(iii) Yet another metric may be chosen by setting $M = H^{-2}$, where $H^{-2}$ is the inverse matrix of the positive definite diagonal matrix of communalities, viz., $H^2$. In this form we have

$$D^2_3(i,j) = d_{ij}' H^{-2} d_{ij}$$  \hspace{1cm} (5.1.7)
Now, we propose a general measure of common factor battery reliability (CFBR) based on the squared distance function defined as in eq. (5.1.4) and also for its particular forms as in eqs. (5.1.5), (5.1.6) and (5.1.7).

**Definition:** The CFBR coefficient is the ratio of the average distance among common factor scores to the average distance among observed scores, i.e.

\[
\psi_{D^2} = \frac{E(D^2(C_i, C_j))}{E(D^2(x_i, x_j))}
\]  

(5.1.8)

where expectation is over all the pairs of individuals \(i\) and \(j\) \((i \neq j)\).

\(\psi_{D^2}\), defined as such, is completely general where the distance function of the form given in eq. (5.1.4) is to be used. \(\psi_{D^2}\) is a genuinely interpretable coefficient i.e. the nearer is the value of \(\psi_{D^2}\) to 1, the greater
is the predictability of common part score vector $C$ from the observed score vector $X$ and the nearer the value of $\rho D^2$ to zero, the smaller is the predictability of $C$ from $X$.

Now for any squared distance function of the form represented in eq. (5.1.4), we have

$$\rho D^2 = \frac{E((C_i - C_j)' M (C_i - C_j))}{E((X_i - X_j)' M (X_i - X_j))} \quad \text{(5.1.9)}$$

However, in eq. (5.1.9), we shall use only the squared distance functions as represented by $D_1^2$, $D_2^2$ and $D_3^2$ in eqs. (5.1.5), (5.1.6) and (5.1.7), respectively.

In order to explore $\rho D^2$ to have a meaningful expression, the result of the following simple lemma will be useful:

**Lemma 1:** Let $Y$ be a $p \times 1$ column vector and $M = (m_{ij})$, $i, j = 1, 2, \ldots, p$ be a $(p \times p)$ symmetric matrix of constants. Then the expected value of the quadratic form $Y' M Y$ is given as follows:

$$E(Y' M Y) = \sum_{i=1}^{p} \sum_{j \neq i}^{p} m_{ij} E(Y_i Y_j)$$

$$= 2 \text{tr.} (MS)'$$

where $S = E(Y Y')$

where $\text{tr.}(A)$ stands for trace of a matrix $A$. 
Now applying Lemma 1 we get the following reduced form of eq.(5.1.9):

\[ \rho_{D^2} = \frac{\text{tr}(M(R-U^2))}{\text{tr}(M \cdot R)} \]  \hspace{1cm} (5.1.10)

where \( R - U^2 = A \) (vide eq.(2.1.8)).

Now setting successively \( M = R^{-1}, I \) and \( H^{-2} \) in eq.(5.1.10), i.e. corresponding to the use of \( D_1^2 (i,j), D_2^2 (i,j) \) and \( D_3^2 (i,j) \), we get \( \rho_{D_1^2}, \rho_{D_2^2} \) and \( \rho_{D_3^2} \) respectively, as below:

(i) Set \( M = R^{-1} \), we get

\[ \rho_{D_1^2} = \frac{\text{tr}(R^{-1}(R-U^2))}{\text{tr}(R^{-1} \cdot R)} \] \hspace{1cm} (5.1.11a)

\[ = \frac{1}{p} \sum_{i=1}^{p} \rho^2 (i) \] \hspace{1cm} (5.1.11b)

where \( \rho^2 (i) \)'s are eigenvalues of the matrix, \( R^{-1}(R-U^2) \).

Incidently, in the context of factor analysis, \( \rho_{D_1^2} \) is nothing but the average of squared canonical correlations (Rao, 1955) or average of \( \alpha_0 \)-coefficients (Bentler, 1968).

(ii) Set \( M = I \), we get
Thus, $\frac{f}{D_2}^2$ is clearly a simple statistical average, viz., the arithmetic mean of communalities.

Alternatively, $\frac{f}{D_2}^2$ can be expressed in terms of the eigenvalues of the matrix, $R-U^2$ i.e.

\[
\frac{f}{D_2}^2 = \frac{\sum_{i=1}^{p} d^{*2}(i)}{p} \tag{5.1.12b}
\]

where $d^{*2}(i)$ is the $i$th eigenvalue of the matrix, $R-U^2$.

(iii) Lastly, set $\bar{H} = H^{-2}$

\[
\frac{f}{D_3}^2 = \frac{\text{tr}(H^{-2}(R-U^2))}{\text{tr}(H^{-2}R)} = \frac{\sum_{i=1}^{p} h^{-2} i}{p} \tag{5.1.13a}
\]

The reliability coefficient, $\frac{f}{D_3}^2$, is a simple statistical average, viz., harmonic mean of communalities.
can also be expressed in terms of eigenvalues of the matrix, \( H^{-2}(R-U^2) \) or equivalently of the matrix \( H^{-1}(R-U^2)H^{-1} \), i.e.

\[
\frac{p}{D_{ij}^2} = \frac{\sum_{i=1}^{p} \theta^2_i}{\sum_{i=1}^{p} h^{-2}_i}
\]

(5.1.13b)

where \( \theta^2_i \) is the \( i \)th eigenvalue of the matrix \( H^{-2}(R-U^2) \) or equivalently that of the matrix, \( H^{-1}(R-U^2)H^{-1} \).

5.2. CFBR Coefficients Based on the Ratios of the Linear Composites and Traces of Two Appropriate Matrices

Let \( \mathbf{X}^* = (\mathbf{X}, \mathbf{F})' \) be a column random vector of \( p \)-test and \( k \)-orthogonal factor variables. Let \( \mathbf{R}^* \) be the correlation matrix of order \( (p + k) \) based on the vector \( \mathbf{X}^* \). Let the matrix \( \mathbf{R}^* \) be partitioned as follows:

\[
\mathbf{R}^* = \begin{pmatrix}
\mathbf{R}_{XX}' & \mathbf{R}_{XF}' \\
\mathbf{R}_{FX}' & \mathbf{R}_{FF}'
\end{pmatrix}
\]

(5.2.14)

Under the assumptions as in eq.(2.1.6), the matrix, \( \mathbf{R}^* \) is expressible as follows:

\[
\mathbf{R}^* = \begin{pmatrix}
\mathbf{R} & \mathbf{A} \\
\mathbf{A}' & \mathbf{I}_k
\end{pmatrix}
\]

(5.2.15)
Further, using the result of Theorem 2.5.1 (Anderson, 1972),

$$R_{XX'} = R_{XX'} - R_{X'f} - R_{f'f} - R_{f'y}$$

(5.2.16)

where $R_{XX'}$ denotes the matrix of residual correlations after removing the effect in $R_{XX}$ due to $k$ (unknown) common factors and matrix, $R_{XX'} . R_{f'f} - R_{f'y}$, denotes the effect removed out of the matrix, $R_{XX}$ attributable to the existence of the hypothesised common factors.

5.2a. CFBR Coefficient Based on the Ratio of Linear Composites of Matrices: A Total Information Approach:

Let $\ell = (\ell_1, \ell_2, \ldots, \ell_p)'$ be a column vector of weights.

Let $\ell' (R_{XX'} \cdot \ell)$ denote the total weighted residual information (attributable to unique part variables in the context of factor analysis) and $\ell' (R_{XX'} . R_{f'f} - R_{f'y})$ denote the total weighted information attributable to the common factors. Both $\ell' (R_{XX'} \cdot \ell)$ and $\ell' (R_{XX'} . R_{f'f} - R_{f'y})$, as such are 'absolute' measures. Let the quantity $\ell' (R_{XX'})$ denote the total weighted correlational information contained in the vector $X$. 
Consider the ratio

\[ \rho_{\mathcal{C}}^2 = \frac{\ell'_{(R_{\mathcal{X}X}^{-1} \cdot R_{\mathcal{X}X}^{-1} \cdot R_{\mathcal{X}X}^{-1}) \ell}}{\ell'_{(R_{\Delta X}^{-1}) \ell}} \]

\[ = 1 - \frac{\ell'_{(R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1}) \ell}}{\ell'_{(R_{\Delta X}^{-1}) \ell}} \]  

(5.2a.17)

Now, the weights \( \ell' \)'s are so chosen that \( \rho_{\mathcal{C}}^2 \) is the least or equivalently the ratio

\[ \frac{\ell'_{(R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1}) \ell}}{\ell'_{(R_{\Delta X}^{-1}) \ell}} = \rho_{\mathcal{C}}^2 \]  

(say)  

(5.2a.18)

is the maximum.

The ratio, \( \rho_{\mathcal{C}}^2 \), denotes the proportion of information due to the common factors and to that in the observed variables.

On comparing eqs. (5.2a.14) and (5.2a.15), the reduced form of the ratio \( \rho_{\mathcal{C}}^2 \) is

\[ \rho_{\mathcal{C}}^2 = \frac{\ell'_{(R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1} \cdot R_{\Delta X}^{-1}) \ell}}{\ell'_{(R) \ell}} \]

(5.2a.19)
Using eq.(2.1.8), we obtain

\[
\beta^*_\tau = \frac{\ell'(R - U^2) \ell}{\ell'(R) \ell}
\]  

(5.2a.20)

The maximization of the coefficient, \(\beta^*_\tau\), with respect to \(\ell\) gives us the following determinental equation:

\[
\left| R^{-1} (R-U^2) - \beta^*_\tau^2 I \right| = 0
\]  

(5.2a.21)

i.e. the coefficients, \(\beta^*_\tau\)'s, are the eigenvalues of the matrix, \(R^{-1}(R-U^2)\). Since the matrix, \(R-U^2\) is assumed to be a positive-semidefinite, the p-values of \(\beta^*_\tau\) would be \(\geq 0\).

In this case, an overall common factor battery reliability coefficient, called the CFBR, can be formulated by taking the average of p eigenvalues i.e. \(\frac{1}{p} \sum_{i=1}^{p} \beta^*_\tau\) and this, in magnitude, is also equivalent to \(\frac{\beta^*\tau^2}{D^2}\).

5.2b. CFBR Coefficient Based on the Ratio of Traces of Two Appropriate Matrices: A Trace Approach

From eq.(5.2.16), we have

\[
\text{tr}(R_{\Delta^k} - \ell) = \text{tr}(R_{\Delta^k}) - \text{tr}(R_{\Delta^k} R^{-1}_{\Delta^k} R_{\Delta^k})
\]  

(5.2b.22)

As such \(\text{tr}(R_{\Delta^k} R^{-1}_{\Delta^k} R_{\Delta^k})\) is an absolute measure of common variance (communality) information extracted from \(\text{tr}(R_{\Delta^k})\).
A relative trace measure may be defined as follows:

\[
\zeta \triangleq \frac{\text{tr} \left( R_{X \mathbf{A}^t} \cdot R_{\mathbf{A}^t}^{-1} \cdot R_{X \mathbf{A}^t} \right)}{\text{tr} \left( R_{X \mathbf{A}^t} \right)} \quad (5.2b.23)
\]

and substituting appropriately from eq. (5.2.15) in eq. (5.2b.23), we have

\[
\zeta \triangleq \frac{\text{tr} \left( A_1 \cdot A_1^t \right)}{p} = \frac{\text{tr} \left( A_1 \cdot A_1^t \right)}{p} = \frac{\text{tr} \left( R - U_2 \right)}{p} \quad \text{from eq. (5.1.8)}
\]

\[
= \frac{\sum_{i=1}^{p} h_i^2}{p} \quad (5.2b.24)
\]

Thus, it is observed that this trace formulation of the reliability coefficient problem leads to a value of \( \varphi \cdot \frac{p}{\sum h_i^2} \), which, in turn, is equivalent to \( \zeta \cdot \frac{p}{2} \).

5.3. A Generalised Variance (GV) Based Approach to Common Factor Battery Reliability (CFBR) Coefficient

Before we proceed to propound the generalised variance based reliability formulations, it seems essential here to give definition of generalised variance.

Generalised Variance: Let \( A \) be at least a positive semidefinite dispersion matrix of random vector \( \mathbf{X} \). A multivariate analogue of
the variance, $\sigma^2$, of a univariate distribution is the scalar $|A|$ (i.e. the determinant of A) and is called the generalised variance (GV) of the random vector $X$. Further, it may be noted that

$$|A| = \prod_{i=1}^{p} \theta_i$$

(5.3.25)

where, $\prod$ denotes the product and $\theta_i$'s ($\theta_i > 0$) are the eigenvalues of the matrix, A, of order $p$, i.e. the generalised variance of X is expressible in terms of product of eigenvalues of the dispersion matrix of X.

Remark 5.3.1:- It may be noted that Roy et al. (1971, Ch.II, Section (III)) recommend that certain functions of the eigenvalues of any dispersion matrix may be used as unidimensional summaries of the size of the dispersion. For this, two prospective bases may be conceived: the arithmetic and geometric means of the eigenvalues. The former has already been discussed in Sections 5.1 and 5.2. In this section, the discussion centres round the geometric mean of eigenvalues of some appropriate matrices of interest and this average may be taken as measure of the CFBR coefficient.

Remark 5.3.2:- In factor analysis the matrix, $R-U^2$, is assumed to be positive semidefinite. In empirical data analysis, the
assumption is found to be of hypothetical significance, for, as
pointed out by Harman (1970), "The reduced correlation matrix
of R (i.e. R-U²) will not be positive semidefinite in practice
and both positive and negative eigenvalues may be expected".

In view of this argument and perfectly sticking to the
recommendations of Gnanadesikan (1977), a natural modification is
to base the geometric mean only on the positive eigenvalues of the
relevant matrix.

CFBR Coefficients Based on Generalised Variance (GV):

We define in what follows two CFBR coefficients based on
generalised variance. They pertain to the positive eigenvalues of
matrices, R-U², and U⁻¹(R-U²)U⁻¹ which, respectively, are bases of
the PFA and AMFA (CFA) procedures (cf. Sections 2.2b, 2.2d and
2.2f).

(i) Let d*² 's, i = 1, 2, ..., p, such that

\[ d*²_1 > d*²_2 > \ldots > d*²_k > 0 > d*²_{k+1} > \ldots > d*²_p \]  

be the eigenvalues of the matrix, R-U². Consider the geometric
mean of the positive eigenvalues d*²_1, ..., d*²_k i.e.

\[ \sqrt[k]{\prod_{i=1}^{k} d*²_i} = \left( \prod_{i=1}^{k} d*²_i \right)^{1/k} \]  

(5.3.27)
which in fact is a CFBR coefficient based on the positive eigenvalues of the matrix, $R-U^2$, involved in the FFA procedure. This is analogous to the generalised variance concept.

(ii) Let $\lambda_1^2, \lambda_2^2, \ldots, \lambda_p^2$ be the eigenvalues of the determinental equation

$$| (R-U^2) - \lambda^2 R | = 0 \quad (5.3.28)$$

which forms basis of the CFA and AMFA procedures and, further, let

$$\lambda_1^2 > \lambda_2^2 > \ldots > \lambda_t^2 > 0 > \lambda_{t+1}^2 > \ldots > \lambda_p^2$$

Therefore, considering the recommendations of Gnanadesikan (1977), we may define the CFBR coefficient in the present case as

$$\bar{\lambda}^{\text{GV}} = \left( \prod_{i=1}^{t} \frac{\lambda_i^2}{1} \right)^{\frac{1}{t}} \quad (5.3.29)$$

The $\lambda_i^2$'s, $i = 1, \ldots, t$ are nothing but the squared canonical correlations due to Rao (1955) and $\alpha$-coefficients of Bentler (1968). In this context, therefore $\bar{\lambda}^{\text{GV}}$ can be interpreted as geometric mean of the canonical predictability coefficients or the internal consistency coefficients of common factors depending upon the context in which the basic matrix,
$R^{-1}(R-U^2)$, has been arrived at. This again is analogous to the generalised variance concept.

**Remark 5.3.3:** The representation of various common factor battery reliability coefficients developed in the current chapter in terms of eigenvalues of the matrices $R-U^2$, $U^{-1}(R-U^2)U^{-1}$ and $H^{-1}(R-U^2)H^{-1}$, shall be taken up again in the forthcoming Chapter VI where we deal with the practical utilization of these CFBR coefficients in order to determine the number of interpretable factors in any factor analytic study.