

Tau-Proportionate Form of Congeneric Covariance Structure

Saran Ishika Maiti^{1,2}

¹ Indiana University-Purdue University Indianapolis, USA

² Department of Statistics, Visva-Bharati University, Santiniketan, India

Correspondence: Saran Ishika Maiti, Indiana University-Purdue University Indianapolis, USA. E-mail: saran.ishika@gmail.com

Received: May 15, 2016 Accepted: June 14, 2016 Online Published: August 9, 2016

doi:10.5539/ijsp.v5n5p1 URL: http://dx.doi.org/10.5539/ijsp.v5n5p1

Abstract

From the view point of constructing a composite possessing the largest possible common covariability with all of its constituent tests, the concept of the best equi-covariable composite (BEC) has been introduced. When the covariance matrix of the constituent tests are so structured that its row (column) totals are all equal, the corresponding BEC is shown to be reducible to simple aggregated (or averaged) test. Under the purview of congeneric tests, it is shown that there exists a special form termed tau-proportionate congeneric form in which case the simple aggregated (or averaged) test is not only the BEC but also the most reliable composite. Statistical techniques of the estimation of parameters, testing and goodness-of-fit of the such structure are considered with the aid of an illustrative example.

Keywords: congeneric tests, most reliable composite, Cronbach's alpha, goodness-of-fit indices, structural equation modeling

1. Introduction

In classical test theory, it is assumed that the score obtained by an examinee in a test is the sum of a true score and error score where those are uncorrelated. For a set of p tests purportedly aiming at measuring the same trait, the scores x_1, x_2, \dots, x_p with corresponding true scores $\tau_1, \tau_2, \dots, \tau_p$ is said to be congeneric if every pair of true scores τ_g and τ_h has unit correlation. Such a set of observed test scores is represented in a vector form as (Jöreskog, 1971, Sec.2)

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\beta}\tau + \mathbf{e} \quad (1)$$

where $\mathbf{X} = (x_1, x_2, \dots, x_p)'$ is the vector of p scores, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ is a vector of regression coefficients, $\mathbf{e} = (e_1, e_2, \dots, e_p)'$ is the vector of error scores, τ is a random variable without any loss of generality scaled to zero mean and unit variance, providing $\beta_g\tau$ as the true score for any typical test g . The error scores e_1, e_2, \dots, e_p are assumed random variables having zero expectations and zero correlations with τ and as well as within themselves. They may have possibly different variances, say $\psi_{11}, \psi_{22}, \dots, \psi_{pp}$. Thus, τ and the components of \mathbf{X} and \mathbf{e} are all regarded as random variables for a populations of examinees (Zimmerman, 1975). The p -component vector $\boldsymbol{\mu}$ is treated as the mean score vector of \mathbf{X} . Typically for any two tests g and h , $\tau_g = \beta_g\tau$, $\tau_h = \beta_h\tau$, $v(\tau_g) = \beta_g^2 = cov(x_g, \tau)$, $v(\tau_h) = \beta_h^2 = cov(x_h, \tau)$ and $cov(\tau_g, \tau_h) = \beta_g\beta_h$ implying $corr(\tau_g, \tau_h) = 1$ (as assumed for congenericity of tests). The covariance matrix $\boldsymbol{\Sigma} = ((\sigma_{gh}))p \times p$ under congeneric model (1) is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\psi} \quad (2)$$

where $\boldsymbol{\psi} = diag(\psi_{11}, \psi_{22}, \dots, \psi_{pp})$. The precision, more clearly the reproducible capacity of the test g , is quantified by reliability coefficient (ρ_{gg}) being defined as the proportion of the observed score variance 'accounted for' by the true score variance (Miller, 1995):

$$\rho_{gg} = \frac{\beta_g^2}{\beta_g^2 + \psi_{gg}} \quad (3)$$

As the true score variance remains unknown, ρ_{gg} is obtained by constructing another test having the form "parallel" to the test g (say, test h) and applying upon the same group of individuals and finally correlating the two series of scores due to the fact that the correlation between x_g and x_h has the same expression as (3). The notion of parallel forms of a test aims at "making no difference which is used". The statistical criterion of parallelity is to have "equality of means, equality of variances and equality of covariances of the tests" (cf. Gulliksen, 1950, Ch.14). Parallel tests can be used interchangeably. They have equal reliabilities and equal validities in predicting a given criterion. The requirement of equal means will not be considered in this paper. The covariance matrix ($\boldsymbol{\Sigma}$) arising from a group of p parallel forms of the typical test g would take the form

$$\boldsymbol{\Sigma} = \beta_0^2 \mathbf{E}_p + \psi_{00} \mathbf{I}_p \quad (4)$$

Such structure (4) is otherwise called “uniform” or “intraclass” structure (ICS) (Geisser, 1964; Srivastava, 1965). Here, $V(x_g) = \beta_0^2 + \psi_{00}$ and $Cov(x_g, x_h) = \beta_0^2, g \neq h$. E_p is the $(p \times p)$ matrix with all elements unity. I_p is the identity matrix of order p . (4) is a positive definite matrix involving only two parameters β_0^2 and ψ_{00} . The reliability of the test g under parallel form of tests (PFT) is thus

$$\rho_{gg/PFT} = \frac{\beta_0^2}{\beta_0^2 + \psi_{00}} \tag{5}$$

Sometimes the forms of the tests are considered less restrictive compared to parallel forms, e.g., tau-equivalent tests as well as essentially tau-equivalent tests have equal true score variances but possibly different error variances. Thus, here β_g 's are equal while ψ_{gg} 's are different. Such tests cannot be used interchangeably. The test with smallest error variance (smallest ψ_{gg}) is the most reliable test and each test however indicates a different validity in predicting a given criterion. The corresponding covariance matrix for tau-equivalent tests would have the structure

$$\Sigma = \beta_0^2 E_p + \psi \tag{6}$$

The congeneric model (1) may be looked upon as an equivalence to the factor analytic model arising from Spearman’s unifactor theory. The true score (τ) is treated as an unknown latent variable (f), called “common” factor (trait) assumed to be a random variable. The vector of error scores (e) has the components called “unique” or “specific” factor — another random variable associated with the individual test score, but uncorrelated with f . The coefficients $\beta_1, \beta_2, \dots, \beta_p$ are rather called “factor loadings” amounting the importance of f to explain the test scores. Thus the unifactor model is

$$X = \mu + \beta f + e \tag{7}$$

with the assumptions on f and e analogous to those as stated in (1), providing with the same covariance matrix Σ as expressed by (2). The reliability of any test g as defined in (3) is rather interpreted as the common factor’s contribution to the test variance being termed communality of the test g (Jöreskog, 1971, Sec.2).

As the reliability of a test cannot be determined from a single test administration, it requires the use of a parallel form. More often than not, parallel forms are not available. In such situations, a common practice is to obtain indirect information by calculating the lower bound of the reliability with an implication that when the lower bound is high enough, the reliability is adequate (Ten Berge & Socan, 2004). Keeping in view the unifactor theory, when a battery of tests (items) focuses on a single idea or construct (factor), instead of assessing reliability of a test, the quality of the composite formed by the tests conforming the battery is measured by what is called “internal consistency reliability”. Cronbach’s alpha coefficient(1951) is generally used for this purpose indicating indirectly the degree of interrelatedness among the tests (items) measuring different substantive areas within a single construct (unifactor) (Cortina, 1993; Zinbarg, Yovel, Revelle, & McDonald, 2006).

$$\text{Cronbach's alpha}(\alpha) = \frac{p}{p-1} \sum_{g \neq h} \sigma_{gh} / \sum_g \sum_h \sigma_{gh} \tag{8}$$

where σ_{gh} is the (g, h) element in the covariance matrix (Σ) by p test (item) scores x_1, x_2, \dots, x_p .

Instead of a single test, when a number of tests is administered to a group of students, the main problem is to have an appropriate method of combining their scores in terms of a linear composite. The weights used in constructing a linear composite depend upon the type of judgment or criterion fixed up beforehand. As a common practice, however, simple aggregation or average of the test scores(AVT) is used as a composite possessing two simple properties. One, it attaches no special weightage to any particular test among p tests. Two, it does not take into account the structure of covariability among p tests.

Optimum property of the AVT such as minimum variance with unbiasedness, is found to hold so long as the PFT is undertaken. As the associated covariance matrix under PFT has the intraclass covariance structure (ICS), the variance of the AVT would be equally affected no matter which one of the p constituent tests is removed from the composite.

The tests belonging to a battery are most often combined into a suitably weighted composite aiming at increasing the reliability. The AVT under PFT set-up is the “most reliable” composite (MRC) attaining the reliability

$$\rho_{\bar{x}\bar{x}/PFT} = \frac{\beta_0^2}{\beta_0^2 + (\psi_{00}/p)} \tag{9}$$

while under the congeneric model (2),

$$\rho_{\bar{x}\bar{x}/CON} = \frac{\bar{\beta}^2}{\bar{\beta}^2 + (\bar{\psi}/p)} \tag{10}$$

where $\bar{\beta} = \sum_g \beta_g / p$ and $\bar{\psi} = \sum_g \psi_{gg} / p$.

The present investigation aims at searching the situation where the AVT (\bar{x}) may be claimed to be the “most reliable” composite amongst all linear composites.

In Section the present investigation, the concept of equi-covariability as well as of the best equi-covariable composite (BEC) is introduced. It is established that the AVT may be BEC when (2) is restricted to equal row (column) sums(ERS). Another restricted congeneric model, termed tau-proportionate congeneric one, has been developed. The AVT under such model is shown to be not only the most reliable composite (MRC) but also the best equi-covariable one (BEC). The performance of the internal consistency reliability as measured by alpha is compared with that of the reliability of the MRC within the purview of ERS congeneric tests. A statistical test has been forwarded to check a ERS Σ from a sample data set. The maximum likelihood estimation of the parameters involved in tau-proportionate congeneric structure of Σ , is considered along with some related results. An illustrative example, based on real data, is put forward showing the compatibility of tau-proportionate congeneric structure , followed by some discussion finally.

2. Equi-covariable Composite

Definition: 1. A linear composite $y = \gamma'X$ is said to have the property of equi-covariability with respect to a random vector $X = (x_1, x_2, \dots, x_p)'$ if $Cov(y, x_1) = Cov(y, x_2) = \dots = Cov(y, x_p)$.

Construction: The coefficient vector γ of $y = \gamma'X$ is the solution to the system of linear equations $\Sigma\gamma = cJ$ where the scalar c is the common covariance of y with the variables and Σ is the covariance matrix of the random vector X . J is the p -component vector of ones. Under the assumption of positive definiteness of Σ , $\gamma = c\Sigma^{-1}J$. A class of equi-covariable composites may be demarcated for various choices of the scalar $c(> 0)$. The “best” equi-covariable composite (BEC) is a member belonging to this class, indicating the largest covariability which happens when c is chosen as the variance of y . On fixing $c = Var[(c\Sigma^{-1}J)'X]$, the optimum value of c would be $(J'\Sigma^{-1}J)^{-1}$ providing the expression of BEC as

$$y_0 = \frac{J'\Sigma^{-1}}{J'\Sigma^{-1}J}X \tag{11}$$

Once the entries of Σ are so structured that its row (or column) sums are all equal (ERS), $\Sigma J \propto J$ as well as $\Sigma^{-1}J \propto J$ giving rise to the BEC composite (y_0) being simply reducible to \bar{x} . Thus, the averaged test would be the BEC if and only if Σ is ERS. For instance, for an ICS Σ (vide (4)), the corresponding BEC is no other than \bar{x} .

It appears that for a congeneric ERS covariance matrix,

$$\beta_g = \frac{T_0 - \psi_{gg}}{\sqrt{p(T_0 - \bar{\psi})}}, \quad g = 1(1)p \tag{12}$$

where $\bar{\psi} =$ the average of $\psi_{11}, \dots, \psi_{pp}$ and $T_0 =$ common value of row or column total. More explicitly,

$$\Sigma = \frac{1}{p(T_0 - \bar{\psi})} (T_0 - \psi_{11}, \dots, T_0 - \psi_{pp})' (T_0 - \psi_{11}, \dots, T_0 - \psi_{pp}) + \Psi \tag{13}$$

involving only $(p + 1)$ parameters. It is interesting to note further that if the test score variances (σ_{gg} 's as the diagonal elements of Σ) along with T_0 are numerically specified, ψ_{gg} 's can be determined by a convergent iterative scheme [vide Appendix]. For such congeneric model (13), BEC composite (11) simplifies to averaged test (\bar{x}).

Note: In the theory of statistical estimation where there is a number of competent estimators estimating the same parameter unbiasedly, their optimum unbiased linear combination having the least variance is termed best linear unbiased estimator (BLUE). Eventually, the expression (11) would be identical to BLUE when all the test means are equal.

Noting that $V(y_0) = (J'\Sigma^{-1}J)^{-1}$ and $V(\bar{x}) = (J'\Sigma J)/p^2$, the extended Cauchy-Schwarz's inequality (cf. Johnson & Wichern, 1996, Ch.2) would establish that $V(\bar{x}) \geq V(y_0)$. The equality holds when Σ is a ERS. Thus unless Σ satisfies ERS condition, y_0 would not be reducible to \bar{x} . Conversely, unless the test means are all equal, a BEC \bar{x} would no longer become BLUE.

Reliability of BEC under Congeneric Set-up

Under congeneric set-up, the variance and the reliability of any typical (linear) composite $y = \gamma'X$ are given by

$$\begin{aligned} V(y) &= (\gamma'\beta)^2 + \gamma'\psi\gamma \\ \rho_{yy} &= \frac{(\gamma'\beta)^2}{(\gamma'\beta)^2 + \gamma'\psi\gamma} \end{aligned} \tag{14}$$

As the ratio $\frac{(\gamma'\beta)^2}{\gamma'\psi\gamma}$ would attain its maximum when $\gamma \propto \psi^{-1}\beta$, the “most reliable” composite (MRC) would be expressible as $y_m = k(\psi^{-1}\beta)'X$ where k is a positive constant. Consequently,

$$V(y_m) = k^2[(\beta'\psi^{-1}\beta)^2 + \beta'\psi^{-1}\beta] \tag{15}$$

and $\rho_{y_m y_m} = \frac{\beta'\psi^{-1}\beta}{1 + \beta'\psi^{-1}\beta}$.

In the context of equi-covariability criterion, the best equi-covariable composite (BEC) y_0 as expressed in (11), would have

$$V(y_0) = \frac{1}{\mathbf{J}'\psi^{-1}\mathbf{J} - \frac{(\mathbf{J}'\psi^{-1}\beta)^2}{(1+\beta'\psi^{-1}\beta)}} \tag{16}$$

and $\rho_{y_0 y_0} = \frac{(\mathbf{J}'\psi^{-1}\beta)^2}{(1 + \beta'\psi^{-1}\beta)[(1 + \beta'\psi^{-1}\beta)(\mathbf{J}'\psi^{-1}\mathbf{J}) - (\mathbf{J}'\psi^{-1}\beta)^2]}$.

Clearly, $\rho_{y_0 y_0} \leq \rho_{y_m y_m}$. The equality holds if and only if $\beta \propto \mathbf{J}$, i.e., iff β_g 's are all equal (say, β_0) in which case the corresponding composite would be the most-reliable-cum-best equi-covariable composite (MRBEC) being expressible as

$$y_{m_0} = \frac{\mathbf{J}'\psi^{-1}\mathbf{X}}{\mathbf{J}'\psi^{-1}\mathbf{J}} \tag{17}$$

having $V(y_{m_0}) = \beta_0^2 + \frac{1}{\mathbf{J}'\psi^{-1}\mathbf{J}}$ and $\rho_{y_{m_0} y_{m_0}} = \frac{\beta_0^2(\mathbf{J}'\psi^{-1}\mathbf{J})}{1+\beta_0^2(\mathbf{J}'\psi^{-1}\mathbf{J})}$.

β_0^2 signifies the common true score variances of each of p tests, indicating a special situation, called essentially tau-equivalent congeneric tests in the sense of Lord & Novick(1968). Such composite (17) is obtainable when each component of \mathbf{X} is given a weight inversely proportional to its error variances (ψ_{gg} 's) that are usually different (Jöreskog, 1971, Sec.2).

Now, as a natural query, what situation could make the AVT(\bar{x}) MRC? On equalizing the reliability formulae corresponding to \bar{x} and y_m (vide (10) and (15)).

$$\beta'\psi^{-1}\beta = \frac{\beta'\mathbf{J}\mathbf{J}'\beta}{\mathbf{J}'\psi\mathbf{J}}. \tag{18}$$

Applying the generalized Cauchy-Schwartz inequality (Johnson and Wichern, 1996, Ch.2), (18) would hold when $\beta \propto \psi\mathbf{J}$. Consequently, the covariance matrix would have the following structure.

$$\Sigma = k_0\psi\mathbf{J}\mathbf{J}'\psi + \psi \tag{19}$$

Thus the structure (19) may be revealed to stem from such congeneric tests as would have the key feature that the covariance of any constituent test (x_g) with the averaged test (\bar{x}) is proportional to its error variance (ψ_{gg}). Such tests may be termed “tau-proportionate” congeneric tests.

As the reliability (ρ_{gg}) of any g^{th} test is $\frac{k_0\psi_{gg}}{1+k_0\psi_{gg}}$, the function $\frac{\rho_{gg}}{(1-\rho_{gg})\psi_{gg}}$ (being equal to k_0) will remain invariant in respect of any g^{th} test belonging to tau-proportionate congeneric tests.

If additionally, (19) possesses the ERS feature, corresponding MRC averaged test would be lifted to MRBEC averaged test. Clearly, the MRBEC averaged test would be available when $(\psi^{-1}\mathbf{J}) \propto \mathbf{J}$ (vide (17)) in which case (19) is reducible to a ICS structure indicating the parallelity of the congeneric tests.

3. Internal Consistency Reliability and MRC under ERS Congeneric Set-up

Recalling Cronbach’s alpha (vide (8)) as a measure of internal consistency reliability,

$$\alpha = \frac{pb}{a + (p - 1)b} \tag{20}$$

where a = average of the variances and b = average of the covariances.

In the case of ERS congeneric tests (vide(13)), a and b would take the expression as follows.

$$a = \bar{\psi} + \frac{T_0 - \bar{\psi}}{p} + \frac{s_{\bar{\psi}}^2}{p(T_0 - \bar{\psi})} \tag{21}$$

$$b = \frac{T_0 - \bar{\psi}}{p} - \frac{s_{\bar{\psi}}^2}{p(T_0 - \bar{\psi})} \tag{22}$$

where $s_{\bar{\psi}}^2 = \frac{1}{p} \sum_{g=1}^p (\psi_{gg} - \bar{\psi})^2$. Correspondingly,

$$\alpha = \frac{1 - \bar{\psi}}{T_0} - \frac{s_{\bar{\psi}}^2}{(p - 1)T_0(T_0 - \bar{\psi})} \tag{23}$$

Clearly, for more variations among ψ_{gg} 's, a would be expected larger while b and, in effect, α would be expected smaller. In view of maximum possible reliability of any composite (MRC) under ERS congeneric tests, let us recall (15) and obtain

$$\rho_{y_m y_m} = \frac{1 - \bar{\psi}}{T_0} + \frac{\bar{\psi} - \bar{\psi}_h}{T_0 - \bar{\psi}} \tag{24}$$

where $\bar{\psi}_h$ denotes $p / \sum_{g=1}^p (1 / \psi_{gg})$, the harmonic mean of ψ_{gg} 's. It is to be noted that $\bar{\psi} \geq \bar{\psi}_h$ and that more the variations among ψ_{gg} 's, more would be the quantities $\bar{\psi} - \bar{\psi}_h$ and $s_{\bar{\psi}}^2$ producing larger $\rho_{y_m y_m}$ but smaller α due to second terms of (24) and (23) respectively. Thus, for a fixed value of $\bar{\psi}$, α , as a lower bound indicator of reliability, would never exceed $\frac{1 - \bar{\psi}}{T_0}$ while $\rho_{y_m y_m}$ of the related MRC would be at least $\frac{1 - \bar{\psi}}{T_0}$. Clearly, their equality would be possible when ψ_{gg} 's are all equal implying trivially a ICS structure of Σ .

4. A Statistical Test for the Tenability of ERS Σ

A statistical test procedure may now be forwarded to test the tenability of ERS structure of Σ under the assumption of multinormality of the score vector $X(p \times 1)$ comprising p scores obtained by an individual when treated by p congeneric forms. Let S be the data covariance matrix based on the $(p \times n)$ data set comprising the score vectors of n individuals.

As J / \sqrt{p} is necessarily an eigenvector of any ERS matrix, the tenability of ERS structure of Σ may be assured equivalently by the tenability of J / \sqrt{p} as an eigenvector of Σ . Following the testing procedure by Mallows(1961), an F-statistic is computed by the expression as follows.

$$F = \left(\frac{n - p}{p - 1} \right) \left[\frac{(\sum_g \sum_h s_{gh})(\sum_g \sum_h s^{gh})}{p^2} - 1 \right] \tag{25}$$

where s_{gh} and s^{gh} are the typical elements of S and S^{-1} respectively. If the computed value of F is less than $F_{\alpha}(p - 1, n - p)$, the tenability of a ERS Σ is asserted at 100 α % level of significance.

Consequently, the common row total (T_0) of ERS Σ may be estimated by noticing that the variable $(J / \sqrt{p})' X$ is distributed as univariate normal with variance T_0 . Considering the joint distribution of n such variables in respect of n individuals, the maximum likelihood estimate (MLE) of T_0 may be obtained as

$$\hat{T}_0 = \frac{p}{n} \sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}})^2 \tag{26}$$

where \bar{x}_i is the average score obtained by i^{th} individual while $\bar{\bar{x}}$ is the overall averaged score ($\frac{1}{n} \sum_{i=1}^n \bar{x}_i$).

Noticeably, the least squares estimate of T_0 based on p row totals of S would be found as $\frac{1}{p} \sum_g \sum_h s_{gh}$ which is reducible to \hat{T}_0 .

ML Estimation of Tau-proportionate Congeneric Structure and Some Related Results

Under usual multinormality assumptions, the maximum likelihood estimates (MLE) of $(p + 1)$ parameters, e.g. k_0 and $\psi_{11}, \dots, \psi_{pp}$ as involved in (19) are obtained by solving the following likelihood equations.

$$|\Sigma|^{-1} \frac{\partial |\Sigma|}{\partial k_0} + \text{tr} \left(S \frac{\partial \Sigma^{-1}}{\partial k_0} \right) = 0 \tag{27}$$

and

$$|\Sigma|^{-1} \frac{\partial |\Sigma|}{\partial \psi_{gg}} + \text{tr} \left(S \frac{\partial \Sigma^{-1}}{\partial \psi_{gg}} \right) = 0, \quad g = 1, 2, \dots, p \tag{28}$$

where S is the sample covariance matrix with the divisor n , the sample size. Noting that

$$\Sigma^{-1} = \Psi^{-1} - \frac{k_0 \mathbf{J} \mathbf{J}'}{1 + k_0 \mathbf{J}' \Psi \mathbf{J}}$$

and $|\Sigma| = |\Psi| (1 + k_0 \mathbf{J}' \Psi \mathbf{J})$,

(27) and (28) are reducible to following implicit equations.

$$k_0 = \frac{\sum_{g=1}^p (c_g - \psi_{gg})}{(\sum_{g=1}^p \psi_{gg})^2} \tag{29}$$

$$\text{and } \psi_{gg} = \frac{\sqrt{1 + 4k_0 s_{gg}} - 1}{2k_0}, \quad g = 1, 2, \dots, p \tag{30}$$

where c_g is the g^{th} column total in S . Starting with suitable trial solutions for k_0 and ψ_{gg} 's, (29) and (30) may be solved by iterations. Keeping \wedge mark over any parameter as a generic notation to indicate its MLE, some straight forward results are furnished below.

- i) $\text{diag}(\hat{\Sigma}) = \text{diag}(S)$, $\text{diag}(\hat{\Sigma}^{-1} S) = \text{diag}(1 + \hat{k}_0(\hat{\psi}_{gg} - \frac{c_g}{\bar{c}} \hat{\psi}))$, $g = 1, 2, \dots, p$ where \bar{c} = average of c_g 's.
- ii) $\text{tr}(\hat{\Sigma}) = \text{tr}(S)$, $\text{tr}(\hat{\Sigma} E) = \text{tr}(SE)$ and $\text{tr}(\hat{\Sigma}^{-1} S) = p$.
- iii) $\hat{\alpha} = \frac{p}{p-1} (1 - \frac{\text{tr} S}{\text{tr}(SE)})$, $\hat{\rho}_{\bar{x}\bar{x}} = \hat{\rho}_{y_m y_m} = 1 - \frac{\text{tr} \hat{\Psi}}{\text{tr}(SE)}$, $\hat{\rho}_{\bar{x}\bar{x}} \geq \hat{\alpha}$.
- iv) $\hat{V}ar(\bar{x}) = \frac{\text{tr}(SE)}{p^2}$, $\hat{C}ov(\bar{x}, x_g) = [\frac{\text{tr}(SE)}{p^2}] \frac{\hat{\psi}_{gg}}{\hat{\psi}}$, $g = 1, 2, \dots, p$ where $\hat{\psi}$ is the average of $\hat{\psi}_{gg}$'s.
- v) $\frac{\mathbf{J}}{\sqrt{p}}$ is a characteristic vector of $\Psi^{-1} \Sigma$ with corresponding characteristic root $1 + pk_0 \bar{\psi}$ whose MLE is $1 + p\hat{k}_0 \hat{\psi}$.

It may be noted that the above results are rather trivial when $\hat{\Sigma} = S$ which happens for a non-structured Σ . But in the case of a structured Σ , as the number of independent parameters is less than the number of non-duplicated elements, $\hat{\Sigma}$ would no longer be equal to S . As such, the findings as furnished above are quite structure-dependent. Some one or more may likely to hold for any other structure(s) also.

In practice, sample covariance matrix (6) is used in the computation instead of its population counterpart, which is essentially never available. It is well-known that model-based estimated covariance matrix such as $\hat{\Sigma}$, can be more efficient compared to S (Bentler, 2009).

5. Analysis of Bock's Vocabulary Data

To study the evolution of the vocabulary of children, the relevant data were drawn by Bock (1975) from the test results on file in the Records Office of the Laboratory School of the Chicago. They consist of scores, obtained from a cohort of pupils from the 8th through 11th grade levels, on alternative forms of the vocabulary section of the Co-operative Reading test. For a sample of 64 (n) pupils scaled scores (after suitably changed origin and unit) are shown in the form of (64 × 4) array in Hårdley and Hlávka (2007). Table 1 shows the computed covariance matrix (S). As the scores under consideration are concerned with the measurements of the same trait (vocabulary) through various grade levels, the present analysis aims at studying the nature of congenericity of the tests.

Table 1. Covariance matrix of Bock’s Vocabulary data

	X_1	X_2	X_3	X_4
X_1	3.5654			
X_2	3.3082	4.7980		
X_3	3.5503	3.6659	4.7035	
X_4	2.8523	3.1248	3.3886	3.7075

Recalling Mallows test (vide Section 5), computed F statistic has the value 2.2184 which is less than $F_{.05}(3, 60)$ having the value 2.7580 leading to acceptance (tenability) of ERS general structure of the covariance matrix of Bock’s vocabulary data. However, for a rigorous study on ERS structure in a congeneric reference formula, we consider various hypothesized structures of Σ in connection with congenericity, e.g., intraclass (ICS) reflecting the parallelity of the tests, tau-equivalent congeneric, tau-proportionate congeneric, Equal-row-sum congeneric (ERS congeneric) being selected as the variants of congenericity. It may be noted that ERS-congeneric one is also a special case of equal-row-sum structure of the covariance matrix (ERS general, vide Sec.2). In Table 2, these structures are arranged in accordance with the number of involved parameters in increasing order. The involved parameters are all estimated by the method of maximum likelihood (ML) under the assumption of multinormality of the data. The most of the likelihood equations are not explicitly solvable. However by iterative procedure (vide Appendix), the convergent ML solutions are obtained. For testing the tenability of the hypothesized structures, the conventional likelihood ratio criterion provides the test statistic $(n - 1)\log(\frac{|\hat{\Sigma}|}{|S|})$ being distributed asymptotically as a chi-square with appropriate degrees of freedom (d.f.) (ν), $\hat{\Sigma}$ and S being respectively the MLE of Σ and the sample covariance matrix. The columns 3, 4 and 5 of Table 2, show respectively the computed chi-square, d.f. and P-value. On comparing the P-values with 5% level of significance, it is clear that except intraclass structure (ICS), all other structures are tenable.

In order to critically judge the competency of the structures, the assessment of their goodness-of-fit is now to be used to supplement the chi-square test (Hu & Bentler, 1998).

A high goodness-of-fit index value may be an encouraging sign that the structure is useful even when it fails to fit exactly on statistical ground and/or there stands a number of competent structures tenable.

Table 2. Goodness-of-fit measures of different hypothesized structures for Bock’s data

	Hypothesized Structure	No.of parameters	Obs. χ^2	d.f. (ν)	P value	GFI	AGFI	AD	RMR	SRMR
1.	Intraclass	2	15.890	8	0.044	0.867	0.834	0.304	0.411	0.102
2.	τ -equi.cong.	5	9.254	5	0.099	0.915	0.830	0.594	0.686	0.180
3.	τ -prop.cong.	5	9.954	5	0.077	0.995	0.991	0.100	0.133	0.032
4.	ERS cong	5	7.506	5	0.186	0.888	0.776	0.259	0.382	0.094
5.	ERS general	7	6.627	3	0.085	0.868	0.560	0.267	0.378	0.092
6.	Cong.general	8	1.600	2	0.449	0.999	0.998	0.036	0.040	0.010

Among many, the most widely used goodness-of-fit indices proposed by Jöreskog and Sörbom (1981) are computed by the following formulae.

$$GFI = 1 - \frac{tr(\hat{\Sigma}^{-1}S - I_p)^2}{tr(\hat{\Sigma}^{-1}S)^2}$$

and adjusting the degrees of freedom (ν) of the hypothesized structure modified version of GFI,

$$AGFI = 1 - \frac{p(p + 1)}{2\nu}(1 - GFI).$$

According to a recommended thumb rule (Hu & Bentler, 1998), computed GFI and AGFI both exceed the “cut-off” point 0.90 (vide Table 2) for only two hypothesized structures, e.g. congeneric and tau-proportionate congeneric indicating their closeness of fit to a greater extent compared to other structures.

We further consider the measures of goodness-of-fit based on the deviations between the elements $\hat{\Sigma}$ and \mathbf{S} . Those are Average Deviation (AD) by Werts, Pike, Rock and Grandy (1981), Root-mean-square Residual (unstandardized) (RMR) by Jöreskog and Sörbom (1981) and Standardized Root-mean-square Residual (SRMR) by Bentler (1995) having the following expressions.

$$AD = \sum_{i=1}^p \sum_{j=1}^p |s_{ij} - \hat{\sigma}_{ij}| / p^2$$

$$RMR = [2 \sum_{i=1}^p \sum_{\substack{j=1 \\ j \geq i}}^p (s_{ij} - \hat{\sigma}_{ij})^2 / p(p+1)]^{\frac{1}{2}}$$

$$SRMR = [2 \sum_{i=1}^p \sum_{\substack{j=1 \\ j \geq i}}^p \left(\frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii}s_{jj}}} \right)^2 / p(p+1)]^{\frac{1}{2}}$$

where $\hat{\sigma}_{ij}$ and s_{ij} are the (i, j) elements of MLE $\hat{\Sigma}$ and \mathbf{S} respectively. The corresponding computed figures are shown in the last three columns of Table 2, amongst which too low values lie along the rows of congeneric and tau-proportionate congeneric structures. In particular, a cut-off value close to 0.08 (as recommended by Hu & Bentler, 1998) for SRMR selects out both the structures. Although both the structures fit the data equally well, to break the tie, the tau-proportionate structure is preferable as it is more restricted with larger degrees of freedom (Forster, 1998; Graham, 2006; Kenny, Kaniskar & McCoach, 2011).

Table 3. ML Estimation of Alpha and Reliability of the Averaged Test

	Hypothesized Structure	Alpha($\hat{\alpha}$)	Reliability of the Averaged Tests ($\hat{\rho}_{\bar{x}\bar{x}}$)	Maximum Possible Reliability of a Composite($\hat{\rho}_{y_m y_m}$)
1.	Intraclass	0.93786	0.93786	0.93786
2.	τ -equi. Cong.	0.94869	0.94869	0.95548
3.	τ -prop. Cong.	0.93779	0.93954	0.93954
4.	ERS Cong	0.93782	0.93800	0.94657
5.	ERS General	0.93786	-	-
6.	Cong. General	0.93768	0.93935	0.94784

Using the ML estimates of the parameters of the hypothesized structures, the ML estimates of the Cronbach’s alpha, $\hat{\rho}_{\bar{x}\bar{x}}$ and $\hat{\rho}_{y_m y_m}$ are shown in the columns of Table 3. All the entries are recorded to five places of decimal after computing them up to six places for a vis-a-vis comparison. The column of alpha indicates the lower bound estimate of the internal consistency reliability for any structure. For the tau-equivalent congeneric structure, the averaged test has the reliability equal to corresponding alpha and thus is the “least reliable composite”. Contrarily, for tau-proportionate congeneric structure, the situation is reverse indicating that the averaged test is the “most reliable composite”. For parallel form of tests providing the intraclass structure, the averaged test is trivially a unique composite as its reliability meets the maximum and alpha both.

6. Discussion

The main concern in this article is to search for constrained congeneric test/item set due to which the aggregated/averaged test (AVT) (sometimes termed scaled score (Kano & Azuma, 2003) or, a composite with unit weights (Bentler, 2004) formed from the set would be the most reliable composite (MRC). Determination is indeed exclusively within the framework of classical test theory (CTT). In the case of intraclass covariance structure arising from parallel form of tests (PFT), Cronbach’s alpha estimates exactly the reliability of the AVT with an attainment of the maximum possible reliability for any composite. However, in the case of (essential) tau-equivalent structure arising from ‘weak’ PFT, corresponding alpha although estimates the reliability of the AVT exactly, that tends to underestimate the test reliability as such (Novick & Lewis, 1967). Present investigation indicates that there may exist a composite which is most reliable but different from AVT (e.g., y_m of (15)). The set of tau-proportionate congeneric tests is, of course, a set of constrained congeneric tests establishing AVT as not only the MRC but also as the best equi-covariable composite (BEC).

Present investigation is confined to the classical test theoretic reliability estimation based on measurement models such as “parallel” and “(essential) tau-equivalent” stemming from congeneric set-up. Keeping in view the single factor analytic model, congeneric set-up would be more meaningful when one considers subject-centered scalable tests/items (Drewes, 2009). When the tests/items are so demonstrated that the corresponding scores achieved by the individuals are continuous and a linear function of factor scores, the situation would be quite conducive to assessing reliability by coefficient alpha (Graham, 2006).

Although reliability analysis is an old topic, it is still given much attention, particularly after the invention of Cronbach's alpha (1951). Limitations of CTT with reference to reliability are studied by many authors. The violation of CTT assumptions highly affects alpha (Green & Hershberger, 2000; Zimmerman, Zambo, & Lalonde, 1993). Raykov (1997) studied the violation of essential tau-equivalence condition that is necessary and sufficient for alpha to equal scale reliability by quantifying alpha's discrepancy (Slippage) from scale reliability with fixed congeneric measures when errors are uncorrelated (Lord & Novick, 1968). As an alternative to coefficient alpha, structural equation modeling (SEM) method to estimate reliability is currently in use (Bentler, 2009). In correlated error situation, alpha is no longer a lower bound of the true reliability. As an alternative, use of Bollen's formula (Bollen, 1980), on reliability is highly recommended. Kano and Azuma (2003) developed a web-based program termed STERA aiming at STEPwise Reliability Analysis with the help of factor analysis and SEM. Accurate estimate of reliability using SEM is more likely under well-constructed scales and critically assessing model fit (Yang & Green, 2010).

References

- Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA:Multivariate Software.
- Bentler, P. M. (2004). Maximal reliability for unit-weighted composites. *UCLA Statistics Preprint No.405*, Oct.7.
- Bentler, P. M. (2009). Alpha, dimension-free and model based internal consistency reliability. *Psychometrika*, 74, 137-143. <http://dx.doi.org/10.1007/s11336-008-9100-1>
- Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill Book Company, New York.
- Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Social Review*, 45, 370-390. <http://dx.doi.org/10.2307/2095172>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Cronbach, L. J.(1951). Coefficient alpha and internal structure of tests. *Psychometrika*, 16, 297-334. <http://dx.doi.org/10.1007/BF02310555>
- Drewes, D. W. (2009). Subject-centered scalability:sine qua non of summated ratings. *Psychological Methods*, 14, 258-274. <http://dx.doi.org/10.1037/a0016621>
- Forster, M. R. (1998). <http://philosophy.wis.edu/forster/220/simplicity.html>.
- Geisser, S. (1964). Estimation in the uniform covariance case. *Journal of Royal Statistical Society Ser.B*, 26, 477-483.
- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent estimates of score reliability what they are and how to use them. *Educational and Psychological Measurement*, 6, 930-944. <http://dx.doi.org/10.1177/0013164406288165>
- Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true scores models and their effects on coefficient alpha. *Structural Equation Modeling*, 7, 251-270.DOI:10.1207/S15328007SEM0702_6
- Gulliksen, H. (1950). *Theory of mental tests*, John Wiley and Sons, Inc. New York. <http://dx.doi.org/10.1037/13240-000>
- Härdley, W., & Hlávka, Z. (2007). *Multivariate Statistics:Exercise and Solutions*. Springer.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling:sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453. <http://dx.doi.org/10.1037/1082-989X.3.4.424>
- Johnson, R. A., & Wichern, D. W. (1996). *Applied Multivariate Statistical Analysis*, 3rd Ed., Prentice Hall of India Private Limited, New Delhi.
- Jöreskog, K.G.(1971). Statistical Analysis of sets of congeneric tests. *Psychometrika*, 36(2), 109-133. <http://dx.doi.org/10.1007/BF02291393>
- Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago:National Educational Resources.
- Kano, Y., & Azuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. *In new developments in*

Psychometrics. Springer-verlag:Tokyo.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333-351. http://dx.doi.org/10.1207/S15328007SEM1003_1

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, M.A: Addison-Wesley.

Mallows, C. L. (1961). Latent vectors of random symmetric matrices. *Biometrika*, 48, 133-149. <http://dx.doi.org/10.1093/biomet/48.1-2.133>

Miller, M. B. (1995). Coefficient alpha: a basic introduction from the perspectives of classical test theory and structures equation modelling. *Structure Equation Modeling*, 2, 255-273. <http://dx.doi.org/10.1080/10705519509540013>

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13. DOI:10.1007/BF02289400

Raykov, T. (1997). Scale reliability, Cronbach’s coefficient alpha and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32(4), 329-353. <http://dx.doi.org/10.1007/BF02289400>

Srivastava, M. S. (1965). Some tests for the intra-class correlation model. *Annals of Mathematical Statistics*, 36(6), 1802-1806. <http://dx.doi.org/10.1214/aoms/1177699809>

Ten Berge, J. M. F., & Socan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613-625. <http://dx.doi.org/10.1007/BF02289858>

Werts, C. E., Pike, L. W., Rock, D. A., & Grandy, J. (1981). Applications of Quasi-Markov Simplex Models across Populations. *Educational and Psychological Measurement*, 41(2), 295-307. <http://dx.doi.org/10.1177/001316448104100207>

Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimate of reliability. *Structural Equation Modeling*, 17, 66-81. <http://dx.doi.org/10.1080/10705510903438963>

Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces and the axioms of test theory. *Psychometrika*, 40(3), 395-412. <http://dx.doi.org/10.1007/BF02291765>

Zimmerman, D. W., Zambo, R. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53(1), 33-49. <http://dx.doi.org/10.1177/0013164493053001003>

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale’s indicators: A comparison of estimators for w_h . *Applied Psychological Measurement*, 30(2), 121-144. <http://dx.doi.org/10.1177/0146621605278814>

Appendix

i) τ -equivalence Congeneric Model

$$\Sigma = \beta_0^2 \mathbf{J}\mathbf{J}' + \psi$$

Likelihood equations are given by $(p + 1)$ simultaneous equations as follows.

$$\beta_0^2 = \frac{\sum_h \sum_k (\frac{s_{hk}}{\psi_{hh}\psi_{kk}}) - \sum_h (\frac{1}{\psi_{hh}})}{(\sum_h \frac{1}{\psi_{hh}})}$$

$$\psi_{gg} = s_{gg} + \frac{\beta_0^2}{1 + \beta_0^2 (\sum_h \frac{1}{h})} [1 - 2 \sum_h (\frac{s_{gh}}{\psi_{hh}}) + \beta_0^2 \sum_h (\frac{1}{\psi_{hh}})], \quad g = 1, 2, \dots, p$$

ii) τ -proportionate congeneric model

$$\Sigma = k_0 \psi \mathbf{J}\mathbf{J}' \psi + \psi$$

Likelihood equations are given by $(p + 1)$ simultaneous equations as follows.

$$k_0 = \frac{\sum_h \sum_k s_{hk} - \sum_h \psi_{hh}}{\sum_h \psi_{hh}^2}$$

$$\psi_{gg} = \frac{\sqrt{1 + 4k_0 s_{gg}} - 1}{2k_0}, \quad g = 1, 2, \dots, p.$$

iii) ERS Congeneric model

$$\Sigma = \frac{\delta\delta'}{d} + \Psi$$

where $\delta = (T_0 - \psi_{11} \quad T_0 - \psi_{22} \quad \dots \quad T_0 - \psi_{pp})'$.

$$\hat{T}_0 = \frac{p}{n} \sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}})^2$$

Likelihood equations for ψ_{gg} 's are given by p -simultaneous equations as follows.

$$\psi_{gg} = (s_{gg} - \psi_{gg}^2) + \frac{(\alpha_g + \psi_{gg})^2 - 2(\alpha_g + \psi_{gg}) \sum_{h=1}^p \left(\frac{s_{gh}\alpha_h}{\psi_{hh}}\right)}{d + \sum_{h=1}^p \frac{\alpha_h^2}{\psi_{hh}}} + \left(\frac{\alpha_g + \psi_{gg}}{d + \sum_{h=1}^p \frac{\alpha_h^2}{\psi_{hh}}}\right)^2 \sum_h \sum_k \left(\frac{\alpha_h \alpha_k s_{kh}}{\psi_{hh} \psi_{kk}}\right)$$

where $d = p(T_0 - \bar{\psi})$ and $\alpha_g = T_0 - \psi_{gg}$, $g = 1, 2, \dots, p$.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).