Please avoid the standardized alpha and the ordinal alpha

❊

Jonas Moss 📷
Department of Mathematics, University of Oslo
PB 1053, Blindern, NO-0316, Oslo, Norway
jonasmgj@math.uio.no

Summary

Coefficient alpha is the most widely used measure of reliability. A cousin of coefficient alpha is the standardized alpha, which is calculated from the correlation matrix instead of the covariance matrix. This paper makes the case that standardized alpha should never be used. It is worse than coefficient alpha as an estimator of the congeneric reliability, there are superior estimators of the standardized reliability, and standardization should be avoided to begin with. Ordinal alpha, the standardized alpha based on the polychoric correlation matrix, should also be avoided, as it has no concrete interpretation. We propose a concrete variant of the ordinal alpha and show how its related to the ordinal alpha.

1 Introduction

The sample coefficient alpha is the most popular estimator of the congeneric reliability, and so it has been for a long time. A close cousin of coefficient alpha is the standardized alpha. The difference between these two alphas is easy to state: To calculate coefficient alpha you use the covariance matrix, to calculate the standardized alpha you use the correlation matrix.

While not as popular as coefficient alpha, standardized alpha is frequently used. It is discussed in the popular textbook of Furr and Bacharach (2013, pp. 139–141). It is readily computed using software such as SAS, SPSS, and the popular R (R Core Team, 2020) package `psych` (Revelle, 2019a). Ordinal alpha, the standardized alpha calculated using the polychoric correlation matrix, is definitely popular. The practical guide of Gadermann, Guhn, and Zumbo (2012) has amassed over 900 Google Scholar citations.

There are three situations where the use of standardized alpha appears to be justified. But in each situation there are alternative estimators of reliability that are demonstrably better than standardized alpha.

(A) **The parallel model.** Standardized alpha can be understood as an estimator of the congeneric reliability. As is well known, standardized alpha equals the congeneric reliability under the parallel measurement model, but coefficient alpha also equals the congeneric reliability in this case. And coefficient alpha should always be preferred to standardized alpha as an estimator of the congeneric reliability for four reasons. (i) It is a consistent estimator of the congeneric reliability under the τ-equivalent model, which contains the parallel model as a special case. (ii) It is the maximum

likelihood estimator of the congeneric reliability assuming normality. (iii) Coefficient alpha has the same asymptotic variance as standardized alpha assuming finite fourth moments. (iv) Coefficient alpha performs about as well as standardized alpha in a small simulation study.

(B) **Standardized scores.** The congeneric reliability is arguably the wrong reliability to use for standardized scores. Standardized alpha appears to be a more reasonable choice than its direct competitor coefficient alpha since it is designed to work for standardized scores. But there are strong reasons not to standardize your items, and there are superior standardization methods. Even if you decide to standardize your items, please avoid standardized alpha and use a plug-in estimate of the standardized reliability instead.

(C) **The ordinal alpha.** A common model for Likert scale items is the normal-ogive model. In this model the covariance matrix is not identifiable, only the correlation matrix is. In this case, standardized alpha can be used, but with the sample polychoric correlation matrix in place of the sample correlation matrix. This variant of standardized alpha is called the *ordinal alpha* and was introduced by Zumbo, Gadermann, and Zeisser (2007). As we shall see, ordinal alpha might not be considered a proper reliability coefficient, as it has no concrete interpretation. Luckily, there is a viable alternative ordinal reliability which has a concrete interpretation.

The meat of the arguments are in Sections 3 to 5, while the next section provides background information.

2   RELIABILITY, COEFFICIENT ALPHA, AND STANDARDIZED ALPHA

Let X be random variable in $\mathbb{R}^k$ with finite variance. Let $\Sigma = \operatorname{Cov} X$ be the covariance matrix of X and $\Phi = \operatorname{Cor} X$ be the correlation matrix of X. Let $\mathbf{1} = (1, 1, \ldots, 1)$ by a vector of ones of the appropriate size and tr be the trace operator. The population coefficient alpha (Cronbach, 1951, eq. 2) is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\operatorname{tr}\Sigma}{\mathbf{1}^T\Sigma\mathbf{1}}\right) \tag{1}$$

and the population standardized alpha (Falk & Savalei, 2011, eq. 2) is

$$\alpha_S = \frac{k}{k-1}\left(1 - \frac{k}{\mathbf{1}^T\Phi\mathbf{1}}\right) \tag{2}$$

We care about these alphas since they sometimes coincide with the reliability coefficient under the congeneric model, a special case of the linear one-factor model. Let Z be a mean zero latent variable with unit variance and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ be a mean zero vector of uncorrelated errors with unit variance. In addition, assume Z and ε are uncorrelated. Define the congeneric model

$$X = \lambda Z + \Psi^{1/2}\varepsilon, \tag{3}$$

where λ is a vector of loadings, and $\Psi^{1/2}$ is a diagonal matrix with positive diagonal elements σ_i. Since $\Psi^{1/2}$ is diagonal, the errors $\Psi^{1/2}\varepsilon$ are uncorrelated. If $\Psi^{1/2}$ is allowed to be any positive definite matrix, the model (3) is a linear one-factor model, and if the dimension of Z is larger than one it is a linear factor model. The random variables Z and ε can have any distribution whatsoever. Since λ, Ψ are parameters of the model, we are dealing with a semi-parametric model.

The congeneric model is most commonly used to model Likert items. But it is not is not necessarily a realistic model. The assumption of one latent factor Z and the assumption of uncorrelated errors are unrealistic in most psychometric settings (Green & Yang, 2009, Section 1.2 – 1.3). We restrict our attention to the congeneric model since it is commonly used, standardized alpha only makes sense on specific cases of the congeneric models, and it makes the presentation smoother. Nevertheless, the case against standardized alpha as a reliability coefficient would be even stronger in a general linear factor model.

Two special cases of the congeneric model are important to us. A congeneric model (3) is *τ-equivalent* if $\lambda_i = \lambda_j$ for all i, j. That is, when all unstandardized factor loadings are equal. Second, a congeneric model is *parallel* if it is a *τ*-equivalent model where every error has the same standard deviation, i.e., $\sigma_i = \sigma_j$. (Lord & Novick, 1968, Section 2.13).

It is time to define reliability. Let Z be an unobserved variable and $\hat{Z}$ a predictor, or measurement, of Z. The reliability coefficient is the squared correlation between a latent variable Z and its predictor $\hat{Z}$ (Lord & Novick, 1968, p. 61).

Definition 1. The reliability of $\hat{Z}$ as a predictor of Z is $R = \text{Cor}^2(\hat{Z}, Z)$.

The squared correlation between two variables X and Y quantifies the degree of linear relationship between X and Y, with 0 being no linear relationship and 1 a perfect linear relationship. In the same way, the reliability coefficient quantifies how the strength of the linear relationship between $\hat{Z}$ and Z.

There is well-known concrete interpretation of the squared correlation as the R^2 between Z and $\hat{Z}$. To state this relationship, we will make use of the notation $\text{MSE}_Z(X) = E[(X - Z)^2]$, the mean squared error of X as predictor of Z when X is a random variable.

Lemma 1. *Assume Z, $\hat{Z}$ have finite variance. Then*

$$\text{Cor}^2(\hat{Z}, Z) = 1 - \frac{\text{MSE}_Z(\alpha + \beta\hat{Z})}{\text{Var}(Z)}, \tag{4}$$

where $\alpha = E(Z) - \beta E(\hat{Z})$ and $\beta = \text{Cov}(\hat{Z}, Z)/\text{Var}(\hat{Z})$ are the theoretical regression coefficients minimizing the mean squared error $\text{MSE}_Z(\alpha + \beta\hat{Z})$.

Proof. See Appendix 1, page 18. □

Suppose we decide not to use any random predictor $\hat{Z}$. Then we will minimize the mean squared error by guessing $E(Z)$, and in this case $\text{MSE}_Z[E(Z)] = \text{Var}(Z)$. Therefore, eq. (4) is a statement about the relative reduction in mean squared error when we use the best linear predictor based on $\hat{Z}$ instead of $E(Z)$. This interpretation of the reliability is concrete, as it numerically answers the question "how good can I expected my best possible linear prediction of Z based on $\hat{Z}$ to be?".

We are concerned with predictors $\hat{Z}$ that are linear functions of X. In other words, there are some real weights $w_i, i = 1, \ldots, k$ such that

$$\hat{Z} = \sum_{i=1}^{k} w_i X_i = \sum_{i=1}^{k} w_i (Z\lambda_i + \sigma_i \varepsilon_i).$$

The second equation follows from the definition of X in the congeneric model (3).

We will use the notation $\bar{x}$ to denote the mean of the vector x. Moreover, when x and y are vectors of the same length, xy is the pointwise product of x and y.

Proposition 2. *Assume the congeneric model (3). Let $\hat{Z} = \sum_{i=1}^{k} w_i X$ for some vector of weights w with elements $w_i, i = 1, \ldots, k$. Then the reliability equals*

$$R_w = \frac{(w^T\lambda)^2}{(w^T\lambda)^2 + w^T\Psi w}, \tag{5}$$

$$= 1 - \text{MSE}_Z(w_0\hat{Z}), \tag{6}$$

where $\text{MSE}_Z(w_0\hat{Z}) = E[(Z - w_0\hat{Z})^2]$ is the mean squared error, and $w_0 = \overline{w\lambda}/(k\overline{w\lambda}^2 + \overline{w^2\sigma^2})$ is the constant minimizing the mean squared error.

Proof. See Appendix 1, page 18 for a proof. □

Equation (5) is the definition Jöreskog (1971, p. 112) gave of the composite reliability under the congeneric model (3). The role of w_0 in Proposition 2 is to make sure $\hat{Z}$ predicts Z on he right scale, and in this paper the scale of Z is fixed by $\operatorname{Var} Z = 1$. Since Z is latent we would usually not care about this scale, but it is handy to fix it to 1 since it allows us to compare different predictors of $\hat{Z}$. The formulation $1 - \operatorname{MSE}_Z(w_0 \hat{Z})$ is more concrete than the correlation formulation, as it explicitly references the predictor $w_0 \hat{Z}$, while the correlation is scale-free and more abstract. The mean squared error definition is amenable to generalizations as well, as explored by Korn and Simon (1991).

A modestly popular choice for w are the weights minimizing the mean squared error, known as the Thurstone weights (Thurstone, 1935), $w_i = \lambda_i / [\sigma_i^2 (1 + k \overline{\lambda^2 \sigma^{-2}})]$, $i = 1, \ldots, k$. Using these weights lead to coefficient H (Hancock & Mueller, 2001), also known as the maximal reliability (Li, 1997),

$$R_H = \frac{k \overline{\lambda^2 \sigma^{-2}}}{k \overline{\lambda^2 \sigma^{-2}} + 1}, \tag{7}$$

where $\overline{\lambda^2 \sigma^{-2}} = k^{-1} \sum_{i=1}^k \lambda_i^2 / \sigma_i^2$. By Proposition 2, the weights are equivalent to $w_i = \lambda_i / \sigma_i^2$ and $w_0 = \sigma_i^2 (1 + k \overline{\lambda^2 \sigma^{-2}})$, which are much easier to interpret.

By far the most popular choice is to give the same weight to each X_i, that is, $w = \mathbf{1} = (1, 1, \ldots, 1)$. This leads to the congeneric reliability, which is the estimand of coefficient alpha under the τ-equivalent model. Its definition is

$$R = \frac{k \overline{\lambda}^2 / \overline{\sigma^2}}{k \overline{\lambda}^2 / \overline{\sigma^2} + 1} \tag{8}$$

where $\overline{\lambda} = k^{-1} \sum_{i=1}^k \lambda_i$ and $\overline{\sigma^2} = k^{-1} \sum_{i=1}^k \sigma_i^2$. Finally, the weights $w_i = (\lambda_i^2 + \sigma_i^2)^{-1/2}$ yield the standardized reliability. The standardized reliability is the estimand of the standardized alpha under conditions soon to be made clear, and can be written as

$$R_S = \frac{k \overline{\lambda (\lambda^2 + \sigma^2)^{-1/2}}^2 / \overline{\sigma^2 (\lambda^2 + \sigma^2)^{-1}}}{k \overline{\lambda (\lambda^2 + \sigma^2)^{-1/2}}^2 / \overline{\sigma^2 (\lambda^2 + \sigma^2)^{-1}} + 1}. \tag{9}$$

Proposition 2 connects all reliabilities on the form of eq. (5) to their unique optimal predictor $\hat{Z}$. Given a vector of weights w, the only linear predictor minimizing the mean squared error $\operatorname{MSE}_Z(\hat{Z})$ while simultaneously satisfying $1 - \operatorname{MSE}_Z(\hat{Z}) = R_w$ is $\hat{Z} = w_0 \sum_{i=1}^k w_i X_i$. This relationship also holds for the well-known reliabilities in eqs. (7) to (9).

Coefficient alpha equals the congeneric reliability under the τ-equivalent model, while the standardized alpha equals the congeneric reliability under the parallel model. It is handy to generalize relationships of this sort. To do this, define the generalization of the coefficient alpha and the standardized alpha, the *weighted alpha*.

Definition 3. Let w be a vector of weights and Σ a covariance matrix. The *weighted alpha* is

$$\alpha_w = \frac{k}{k-1} \left(1 - \frac{w^T (\operatorname{diag} \Sigma) w}{w^T \Sigma w} \right) \tag{10}$$

where $\operatorname{diag} \Sigma$ is the diagonal matrix with diagonal equal to the diagonal of Σ.

The ordinary coefficient alpha (1) has weights $w = \mathbf{1}$ while the standardized alpha (2) has weights $w_i = (\lambda_i^2 + \sigma_i^2)^{1/2}$. Coefficient alpha and standardized alpha are the most interesting cases of the weighted alpha, as you can calculate them using only the covariance matrix $\Sigma = \operatorname{Cov} X$. When w are for instance the

Thurstone weights, a_w is not a function of Σ, as they require knowledge of the factor loadings λ and the residual errors σ. Still, the definition makes sense for any weight w.

The following proposition summarizes the basic relationship between the reliability R_w and the weighted α_w.

Proposition 4. *Let w be a vector of weights and assume X follows the congeneric model* (3)*. Then*

$$\alpha_w = R_w - \frac{k}{k-1} \frac{\overline{w^2\lambda^2} - \overline{w\lambda}^2}{k\overline{w\lambda}^2 + \overline{w^2\sigma^2}} \tag{11}$$

and $\alpha_w \leq R_w$ with equality if and only if wX is τ-equivalent, i.e., $w_i\lambda_i = w_j\lambda_j$ for all i, j.

Proof. See Appendix 1, page 18 for a proof. The inequality $\alpha \leq R$ is classical (Lord & Novick, 1968, Theorem 4.4.3). The discrepancy term in (11) for the special case of coefficient alpha is extensively discussed by Raykov (1997). □

Proposition 4 is a minor generalization of the two most frequently mentioned fact about coefficient alpha: (i) That $\alpha = R$ under the τ-equivalent model. (ii) Coefficient alpha is a lower bound for the congeneric reliability under the congeneric model.

When $w_i = (\lambda_i^2 + \sigma_i^2)^{1/2}$, wX is τ-equivalent if and only if the standardized factor loadings are equal, and Proposition 4 tells us that standardized alpha equals the standardized reliabiltiy (that is, $\alpha_S = R_S$) in this case.

The next proposition explores the relationship between coefficient alpha, the standardized alpha, and the congeneric reliability coefficient.

Proposition 5. *Assume the congeneric model* (3)*. Let α be coefficient alpha* (1)*, α_S be the standardized alpha* (2)*, and R be the congeneric reliability* (8)*.*

  (i) *Assume $\lambda_i = \lambda_j$ for all i, j, i.e., the τ-equivalent model. Then $\alpha_S \geq \alpha = R$, with equality if and only if $\sigma_i = \sigma_j$ for all i, j or $\lambda_i = 0$ for all i.*

  (ii) *If $k = 2$, $\alpha_S \geq \alpha$, with equality if and only if $\lambda_1^2 + \sigma_1^2 = \lambda_2^2 + \sigma_2^2$ or $\lambda_1 = 0$ or $\lambda_2 = 0$. But if $k > 2$, both $\alpha_S > \alpha$ and $\alpha_S < \alpha$ are possible.*

  (iii) *All of $\alpha_S = R$, $\alpha_S > R$ and $\alpha_S < R$ are possible when $\lambda_i \neq \lambda_j$ for some j. In particular, $\alpha_S \geq R$ when $\lambda_i = \sigma_i$ for all i, and $\alpha_S \leq R$ when $\lambda_i^2 + \sigma_i^2$ is constant. Both inequalities are equalities if and only if all λs are equal.*

Proof. See page 19 of Appendix 1 for a proof. □

Remark 6. While α_S is an upper bound for α when $k = 2$ and under the τ-equivalent model, it does not bound α in either direction otherwise, see Proposition 5. This is contrary to claims in the literature such as that of Osburn (2000, p.348) "When the components of a composite measure are congeneric, standardized alpha will always exceed the true reliability". Falk and Savalei (2011, p.450) has a clearly laid out example of $\alpha > \alpha_S$.

By Proposition 4, coefficient alpha equals the congeneric reliability if and only if all the factor loadings λ are equal. But there is no simple necessary and sufficient criterion for $\alpha_S = R$, as Proposition 4 (ii) entails that the parallel model is merely sufficient. Still, these cases are like lucky coincidences, and it would be preposterous to assume $\alpha_S = R$ for unknown parameter values without assuming the parallel model.

Let us proceed to the sample variants of coefficient alpha and standardized alpha. Let S be a sample covariance matrix based on n samples of X and P the sample correlation matrix. The sample coefficient alpha is

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\operatorname{tr} S}{\mathbf{1}^T S \mathbf{1}}\right) \tag{12}$$

while the sample standardized alpha is

$$\hat{\alpha}_s = \frac{k}{k-1}\left(1 - \frac{k}{\mathbf{1}^T P \mathbf{1}}\right) \tag{13}$$

The congeneric reliability (8) has no simple sample version, as it depends on estimates of λ and σ^2, which usually do not have closed forms.

That $\alpha = R$ under the τ-equivalent model and $\alpha_S = R$ under the parallel model show us a way to estimate the congeneric reliability R using a simple function of the sample covariance S, which is readily available. On the other hand. the intuitive estimate of R is the plug-in estimator, which needs estimates of λ and σ. But to estimate these we need to use iterative algorithms and complex software. That α_S and α are easy to calculate is not a decisive advantage these days though, as it is easy to estimate λ and σ using for instance the R (R Core Team, 2020) package `lavaan` (Rosseel, 2012). Estimating R using the plug-in estimator allows us to forego the unrealistic assumptions of τ-equivalence and parallel test items altogether, and is recommended by e.g. McNeish and Wolf (2020).

3   Situation A. The parallel model

How does sample coefficient alpha (12) compare to sample standardized alpha (13) as estimators of the congeneric reliability (8)? From Proposition 5 we know the parallel model is the natural setting to make comparisons, as this is the only case when coefficient alpha and the standardized alpha equals the congeneric reliability with certainty.

The parallel model is

$$X = \lambda Z + \sigma \varepsilon, \tag{14}$$

where λ and σ are positive scalars and $\varepsilon \in \mathbb{R}^k$ and $Z \in \mathbb{R}$ are uncorrelated random variables with unit variance and zero mean. Model (14) is clearly a special case of the congeneric model (3). Our goal is to estimate the congeneric reliability (8), which for the parallel model equals

$$R = \frac{k\lambda^2/\sigma^2}{k\lambda^2/\sigma^2 + 1}. \tag{15}$$

Now we have a model, a parameter, and two competing estimators of that parameter, namely $\hat{\alpha}$ and $\hat{\alpha}_s$. Which should we choose?

Choosing between estimators is a classical statistical problem. Some well-established methods to approach such problems are:

(A) Check the relative asymptotic efficiency of the estimators.

(B) Is any of the competing estimators a maximum likelihood estimator under the appropriate parametric restrictions?

(C) Compare finite-sample performance of the estimators in terms of for instance mean squared error.

(D) Investigate the robustness of the estimators against misspecification of the model.

Point (D) about robustness is an easy one, as we have already established that coefficient alpha is more robust than the standardized alpha. For by Proposition 5, coefficient alpha equals the congeneric reliability under the τ-equivalent model and standardized alpha does not. Thus we can check (D) off our list as an easy win for coefficient alpha.

<center>*3.1   Maximum likelihood*</center>

In this section we will work with the parallel model (14) where Z and ε are independent standard normals. Call this model the normal parallel model.

Coefficient alpha the maximum likelihood estimator of R in the normal parallel model provided it is non-negative, see Theorem 7 below for details. Why does this matter? For one thing, maximum likelihood estimators are guaranteed to be efficient under quite weak conditions (Lehmann, 2004, Section 7.3). But maximum likelihood also has an heuristic justification; maximizing likelihood is an intuitive procedure and is the most popular recipe for making estimators.

Theorem 7. *Assume the normal parallel model.*

(i) *The maximum likelihood estimates of λ^2 and σ^2 are*

$$\hat{\lambda}^2_{ML} = \begin{cases} \frac{1}{(k-1)k}(\mathbf{1}^T S \mathbf{1} - \operatorname{tr} S) & \mathbf{1}^T S \mathbf{1} \geq \operatorname{tr} S, \\ 0 & \text{otherwise}, \end{cases}$$

and

$$\hat{\sigma}^2_{ML} = \begin{cases} \frac{1}{k-1}(\operatorname{tr} S - \mathbf{1}^T S \mathbf{1}/k) & \mathbf{1}^T S \mathbf{1} \geq \operatorname{tr} S, \\ \frac{\operatorname{tr} S}{k} & \text{otherwise}. \end{cases}$$

(ii) *The maximum likelihood estimator of the congeneric reliability (8) equals the sample coefficient alpha when $\mathbf{1}^T S \mathbf{1} \geq \operatorname{tr} S$ and 0 otherwise. That is,*

$$\hat{R}_{ML} = \begin{cases} \hat{\alpha} & \mathbf{1}^T S \mathbf{1} \geq \operatorname{tr} S, \\ 0 & \text{otherwise}. \end{cases}$$

Proof. See page 21 of Appendix 1 for a proof.   □

<center>*3.2   Asymptotic efficiency*</center>

Comparing efficiencies is the most popular method to compare estimators. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators converging at the $n^{1/2}$-rate to normal distributions centered on the same parameter θ. The asymptotic variance of the first distribution is ϕ_1^2 and the asymptotic variance of the second distribution is ϕ_2^2. Then θ_1 is more efficient than θ_2 if $\phi_1 < \phi_2$.

An estimator $\hat{\theta}$ is *efficient* (Lehmann, 2004, Section 4.3) if its asymptotic variance is smaller than any other estimator. Maximum likelihood estimators are efficient under weak conditions, hence α is efficient under the normal parallel model by Theorem 7.

Theorem 8. *Assume X has finite variance.*

(i) *Both sample coefficient alpha and sample standardized alpha converge to their population counterparts with probability 1. Under the parallel model they are strongly consistent and converge to the congeneric reliability. That is, $\hat{\alpha} \to \alpha$ and $\hat{\alpha}_s \to \alpha_S$ with probability 1 and $\alpha = \alpha_S = R$.*

(ii) *Under the parallel model (14), coefficient alpha and standardized alpha share a normal limit distribution when Z and ε have finite fourth moments, i.e., $n^{1/2}(\hat{\alpha} - \alpha)$ and $n^{1/2}(\hat{\alpha}_s - \alpha_S)$ both converge to the same normal variable Z.*

(iii) *Under the normal parallel model, both α and α_S are efficient. Their common limit is normal with asymptotic variance*

$$\sigma^2(\alpha) = \sigma^2(\alpha_S) = 2\frac{k}{k-1}(1 - \alpha_S)^2.$$

Proof. See page 22 of Appendix 1 for a proof of (i) and (iii). Appendix 2 (p. 25) contains the proof of (ii) together with an expression of the common variance. The proof of (ii) relies heavily on the work of van Zyl, Neudecker, and Nel (2000) and Hayashi and Kamata (2005). The proof of (iii) relies on van Zyl et al. (2000, eq. 13) and Kristof (1963, eq. 58), who, among others, derived the limit distribution of coefficient alpha under the normal parallel model.   □

Remark 9. Item (i) states that both $\hat{\alpha}_s$ and $\hat{\alpha}$ estimate what they should. Raykov (2019) provides an introduction to convergence with probability 1 in the psychometry context and a proof of (i) for coefficient alpha. Item (ii) is the main argument of this section. It shows that $\hat{\alpha}_s$ and $\hat{\alpha}$ will perform equally well in the limit. It does not matter what λ, σ and the distributions of Z, ε are; we only require that the models have finite fourth moments. Regarding item (iii), a straightforward consequence of Theorem 7 is that $\hat{\alpha}$ is efficient under the normal parallel model, and in this sense (iii) is an extension of said theorem. The shared limit distribution is simple enough to be usable in practice.

3.3 A small simulation

There is no reason to choose standardized alpha over coefficient alpha from an efficiency standpoint. But it might still be the case that the standardized alpha outperforms coefficient alpha in small samples. This section contains a small simulation study comparing the mean squared error of the estimators under the parallel model. Since the coefficient alpha is the maximum likelihood estimator for the normal parallel model, there is a heuristic reason to believe it will perform better than the standardized estimator in this context. Instead of normal errors, we will use three non-normal errors:

1. Scaled t-distribution with 5 degrees of freedom. The t-distribution has heavier tails than the normal distribution but is symmetric and nearly bell-shaped.

2. Scaled and centered Beta distribution with $\alpha = \beta = 1/10$. This is a bimodal distribution with much weight placed on the modes at 0 and 1.

3. Scaled and centered Gamma distribution with $\alpha = \beta = 1/100$. This is asymmetric and with heavy tails.

All of the distributions are extreme, and should be able to force out the difference between the estimators. We use 5 degrees of freedom for the t-distribution since it is the first integer degree of freedom where t has finite fourth moment. The Open Science Foundation repository for this paper (`https://osf.io/s356h/`) contains the code for the simulations.

Table 1: Simulations of $100 \times \mathrm{MSE}_Z(\alpha)/\mathrm{MSE}_Z(\alpha_s)$ in the parallel model

	$t(5)$	$t(5)$	$t(5)$	Beta	Beta	Beta	Gamma	Gamma	Gamma
	$\sigma = 2$	$\sigma = 1$	$\sigma = .5$	$\sigma = 2$	$\sigma = 1$	$\sigma = .5$	$\sigma = 2$	$\sigma = 1$	$\sigma = .5$
$k = 5, n = 50$	112	95.6	99.2	102	137	125	109	102	112
$k = 20, n = 50$	106	107	105	94.9	89.6	99.8	97.4	96	94.8
$k = 5, n = 200$	99.5	99	103	101	100	100	71.6	53.6	68.9
$k = 20, n = 200$	36.5	163	154	93.7	62.7	525	278	168	253

Table 1 contains $100 \times \mathrm{MSE}_Z(\alpha)/\mathrm{MSE}_Z(\alpha_S)$, where MSE_Z denotes the mean squared error. Values higher than 100 are in favor of $\hat{\alpha}_S$, values lower than 100 in favor of $\hat{\alpha}$. The table shows that α_S tends to outperform α when both k and n are large and σ is small, but with no clear trends otherwise. In this simulation $\hat{\alpha}_s$ does slightly better than $\hat{\alpha}$ – the geometric mean of $\mathrm{MSE}_Z(\alpha)/\mathrm{MSE}_Z(\alpha_S)$ is 1.08, giving an 8% advantage to $\hat{\alpha}_s$. Of course, this is a small simulation study and it is hard to predict what would happen with other distributions. Still, the result of these simulations should be counted in favor of sample standardized alpha.

To sum up, sample alpha won the maximum likelihood and robustness competitions, drew the asymptotic efficiency competition, and lost the finite sample competition. The tally is coefficient alpha $2\frac{1}{2}$ - $1\frac{1}{2}$ standardized alpha.

4   SITUATION B. STANDARDIZED SCORES

Standardized alpha is explicitly linked to the standardized reliability (9) but not to the congeneric reliability (8). Recall the definition of reliability from Section 2. Given a predictor $\hat{Z}$ of Z, the reliability R equals the square correlation between $\hat{Z}$ and Z, or $R = \text{Cor}^2(\hat{Z}, Z)$. Assume we know the values of λ and σ. When we use unstandardized scores, our predictor is

$$\hat{Z} = w_0 \sum_{i=1}^{k} X_i \tag{16}$$

where $w_0 = \overline{\lambda}/(k\overline{\lambda}^2 + \overline{\sigma^2})$ scales the predictions to match $\text{Var}\, Z = 1$, as stated in Proposition 2. When we use standardized scores, our predictor is

$$\hat{Z}^* = w_0 \sum_{i=1}^{k} (\lambda^2 + \sigma^2)^{-1/2} X_i \tag{17}$$

The predictors $\hat{Z}$ and $\hat{Z}^*$ are not the same in general, and their reliabilities are not the same either.

This difference matters. Reliability coefficients are useful because they tell you how good you can expected your prediction of Z based on X to be. Provided you use the mean squared error as a measure of prediction success, this is exactly what the mean squared error-based reliability concept in Proposition 2 says. For $1 - \text{MSE}_Z(w_0\hat{Z})$ puts a number between 0 and 1 on the expected quality of your prediction. If you do not use the predictor $\hat{Z}$, but some other predictor $\hat{Z}^*$, this concrete interpretation of the reliability disappears.

An important motivation for using the squared correlation between Z and $\hat{Z}$ as the definition of a reliability coefficient is that it allows us to do correction for attenuation. Let $\hat{Z}_1, \hat{Z}_2$ be two linear predictors of two latent variables Z_1, Z_2 defined on two separate congeneric models. Assume the errors ε_1 and ε_2 of these models are uncorrelated. Our goal is to find the correlation $\text{Cor}(Z_1, Z_2)$ between the latent variables Z_1 and Z_2 using only the reliabilities $R(\hat{Z}_1)$, $R(\hat{Z}_2)$, and the observed correlation $\text{Cor}(\hat{Z}_1, \hat{Z}_2)$. Then

$$
\begin{aligned}
\text{Cor}(\hat{Z}_1, \hat{Z}_2) &= \frac{\text{Cov}(Z_1, \hat{Z}_1)\,\text{Cov}(Z_2, \hat{Z}_2)}{\text{sd}(\hat{Z}_1)\,\text{sd}(\hat{Z}_2)} \text{Cor}(Z_1, Z_2), \\
&= (R(\hat{Z}_1))^{1/2}(R(\hat{Z}_2))^{1/2}\,\text{Cor}(Z_1, Z_2),
\end{aligned}
$$

and $\text{Cor}(Z_1, Z_2) = \text{Cor}(\hat{Z}_1, \hat{Z}_2)/(R(\hat{Z}_1)R(\hat{Z}_2))^{1/2}$, the famous Spearman (1904) attenuation formula.

Assume you use the standardized sum scores $\hat{Z}_1^*$, a predictor on the form of eq. (17), to make predictions for Z_1. To predict Z_2 you use the $\hat{Z}_2$, a predictor on the form of eq. (16). Then the correlation between your predictions is $\text{Cov}(\hat{Z}_1^*, \hat{Z}_2)$. It would be wrong to use the congeneric reliability for Z_1 in Spearman's formula, as it based on the sum scores $\hat{Z}_1$. If you used the congeneric reliability, the correlation between Z_1 and Z_2 would incorrectly be calculated according to $\text{Cor}(\hat{Z}_1^*, \hat{Z}_2) = (R(\hat{Z}_1)R(\hat{Z}_2))^{1/2}\,\text{Cor}(Z_1, Z_2)$ instead of the correct $\text{Cor}(\hat{Z}_1^*, \hat{Z}_2) = (R(\hat{Z}_1^*)R(\hat{Z}_2))^{1/2}\,\text{Cor}(Z_1, Z_2)$. You would be off by a factor of $(R(\hat{Z}_1^*)/R(\hat{Z}_1))^{1/2}$.

This connection between reliabilities and the choice of predictor weights is not new. It has been observed by e.g. Jöreskog (1971, p. 112) and McNeish (2018), who repeatedly insists on taking this connection into account. Several authors recommend you use standardized alpha with standardized sum scores. For instance, Falk and Savalei (2011, p. 451) states "If researchers intend to sum standardized scores, the correlation matrix is more appropriate for determining internal consistency," while Cortina (1993, p. 99) writes "Standardized

Table 2: Reliabilities, weights, and predictors

Reliability	Definition	Weights (w_i)	Predictor name
Congeneric reliability (R)	$\frac{k\overline{\lambda}^2/\overline{\sigma^2}}{k\overline{\lambda}^2/\overline{\sigma^2}+1}$	1	Sum score
Coefficient H (R_H)	$\frac{k\overline{\lambda^2\sigma^{-2}}}{k\overline{\lambda^2\sigma^{-2}}+1}$	$\lambda_i\sigma_i^{-2}$	Thurstone score
Standardized reliability (R_S)	$\frac{k\overline{\lambda(\lambda^2+\sigma^2)^{-1/2}}/\overline{\sigma^2(\lambda^2+\sigma^2)^{-1}}}{k\overline{\lambda(\lambda^2+\sigma^2)^{-1/2}}/\overline{\sigma^2(\lambda^2+\sigma^2)^{-1}}+1}$	$(\lambda_i^2+\sigma_i^2)^{-1/2}$	Standardized score
Sigma reliability (R_σ)	$\frac{k\overline{\lambda\sigma^{-1}}^2}{k\overline{\lambda\sigma^{-1}}^2+1}$	σ_i^{-1}	Sigma-standardized score

Note: The score is $\hat{Z} = w_0 \sum_{i=1}^k w_i X_i$, where $w_0 = \overline{w\lambda}/(k\overline{w\lambda}^2 + \overline{w^2\sigma^2})$.

alpha is appropriate if item standard scores are summed to form scale scores." But the connection is not always appreciated. For instance, Osburn (2000, p.348) writes "[…] it is important to note that the true reliability of a measure is the same for both standardized and unstandardized observed scores."

You should use standardized reliability if you use standardized scores, and standardization looks reasonable on the face of it. For instance, standardization appears reasonable if you have some items on 5-point scale and others on a 7-point scale (Furr & Bacharach, 2013, p. 139). But there are good reasons to never standardize scores, that is, to never use weights on the form $w_i = (\lambda_i^2 + \sigma_i^2)^{-1/2}$ when you form your linear predictor. Using some $w \neq 1$ seems reasonable, for you would not want items with large variance to overshadow the other contributions to $\hat{Z}$. But there is something strange about $w_i = (\sigma_i^2 + \lambda_i^2)^{1/2}$. Dividing by this particular w_i seems wrong, as you want items with large loadings, i.e., λ_is, to have a large influence on your prediction of Z. But dividing by $(\sigma_i^2 + \lambda_i^2)^{1/2}$ makes the influence of items with large loadings smaller!

The weights $w_i = \sigma_i^{-1}$ is an intuitive alternative to standardized weights. The modified model wX will have all residual standard deviations equal to 1 and factor loadings equal to the standardized factor loadings $\lambda_i^* = \lambda_i/\sigma_i$. Since the λ_i^*s are on the same scale, they are directly comparable. Since the weights do not involve λ, they do not penalize items with large loadings.

Call the reliability with weights $w = \sigma^{-1}$ the *sigma reliability* and denote it R_σ. Using Proposition 2, it can be written as

$$R_\sigma = \frac{k\overline{\lambda\sigma^{-1}}^2}{k\overline{\lambda\sigma^{-1}}^2 + 1}. \tag{18}$$

Another option is to use the Thurstone weights, yielding coefficient H and the Thurstone scores instead of the standardized alpha and standardized sum scores. Since the unstandardized Thurstone scores equals λ/σ^2, they are easy too interpret too. A strong case for the Thurstone scores is made by McNeish and Wolf (2020).

Table 4 summarizes the four reliability coefficients discussed up to this point and their associated weights. The next theorem describes how they are related to each other.

Theorem 10. *Let R be the congeneric reliability, R_S the standardized reliability R_σ the sigma reliability and R_H coefficient H. The definitions of these reliabilities are in Table 4.*

(i) *Assume the τ-equivalent model, i.e., the congeneric model with $\lambda_i = \lambda > 0$ for all i but possibly different residual errors $\sigma_i > 0$. Then*

$$R_H \geq R_\sigma \geq R_S \geq R,$$

with all inequalities strict unless $\sigma_i = \sigma$ for all i. In that case, $R_H = R_\sigma = R_S = R$.

(ii) *Assume the congeneric model with $\sigma_i = \sigma > 0$ for all i but possibly different loadings λ_i. Then*

$$R_H \geq R_\sigma = R \geq R_S,$$

and the inequalities are strict unless $\lambda_i = \lambda$ for all i. In that case, $R_H = R_\sigma = R_S = R$.

Proof. See Appendix 1, page 23.   □

Since the τ-equivalent model is such a common assumption among psychometricians, point (i) of the theorem above gives a strong reason to prefer R_σ to R_S and R. Point (ii) is an argument against using the standardized reliability (9) when $\sigma_i^2 = \sigma_j^2$ for all i, j.

You can interpret Theorem 10 as a ranking factor scores too. Under the τ-equivalent model, Thurstone scores are better than σ-standardized sum scores, which in turn are better than the standardized sum scores, which are better than the ordinary sum scores.

Theorem 10 shows how the four reliabilities relate to each other only under the τ-equivalent model and the model where all σ_is are equal. If we the fully general congeneric model, the nice inequalities of Theorem 10 disappear. Since coefficient H uses the optimal weights, it is guaranteed to be larger than the three other reliabilities mentioned above. The ordering of the other three reliabilities are harder to predict, but R_σ will usually be the largest. In a short simulation (10^5 repetitions) of $k = 5$ items with $\lambda_i \sim \text{Exp}(1)$ and $\sigma_i \sim \text{Exp}(1)$ independently with $i = 1, \ldots, 5$, R_σ was the largest 98% of the time, R largest 2% of the time, but R_S was never the largest. The average reliabilities where $\overline{R}_\sigma = 0.91$, $\overline{R}_s = 0.76$, and $\overline{R} = 0.70$. The congeneric reliability R dominated the standardized R_S in 4/10 iterations. You could argue this simulation study is somewhat unrealistic, but the slightly different simulation with λ_i and $\sigma_i \sim \text{HalfNormal}(0, 1)$, yielded virtually the same results. In a simulation where λ and σ where uniform on $[0, 1]$, yet again R_σ was the largest 98% of the time and R the largest the remaining 2%. Moreover, in every iteration of every simulations $R_\sigma > R_S$. The take-away message from these simulations is that R_σ generally does better than R_S and R, while it is harder to predict the winner among R and R_S. See the Open Science Foundation repository for this paper (https://osf.io/s356h/) for the code to run these informal simulations.

Table 3: Comparison of reliability coefficients on personality data

	A	C	E	N	O
Congeneric reliability (R)	0.712	0.733	0.767	0.813	0.610
Coefficient H (R_H)	0.765	0.739	0.778	0.850	0.651
Standardized reliability (R_S)	0.724	0.734	0.763	0.815	0.618
Sigma reliability (R_σ)	0.744	0.736	0.770	0.836	0.630

A, agreeableness; C, conscientiousness; E, extraversion; N, neuroticism; O, openness to experience

Example 11. In this example we will take a look at the data set `bfi` from the R (R Core Team, 2020) package `psychTools` (Revelle, 2019b). This data set contains the responses on 25 personality self report items from the International Personality Item Pool (Goldberg, 1999) collected from the Synthetic Aperture Personality Assessment project (Revelle et al., 2017). The 25 items are evenly distributed across the big five traits: Agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience.

For each trait, `lavaan` (Rosseel, 2012) estimated the factor loadings λ and residual standard deviations σ in the congeneric measurement model (3). Table 3 contains the plug-in estimators for all the four reliability coefficients, i.e., each $\hat{R}$ calculated by substituting $\hat{\lambda}, \hat{\sigma}$ for λ, σ. As expected, R_σ is larger than both R_S and R for all five traits. Moreover, $R_S > R$ for all traits except extraversion, where the difference is marginal.

The Open Science Foundation repository for this paper (https://osf.io/s356h/) contains the code for this example.

Suppose you want to use the standardized weights despite the counterarguments above. By Proposition 4, the standardized coefficient alpha equals the standardized reliability only when the modified model

$$X_i^* = \frac{\lambda_i}{(\lambda_i^2 + \sigma_i^2)^{1/2}} Z + \frac{\sigma_i}{(\lambda_i^2 + \sigma_i^2)^{1/2}} \varepsilon_i, \quad i = 1, \ldots, k \tag{19}$$

is τ-equivalent. This occurs if and only if each $\lambda_i/(\lambda_i^2 + \sigma_i^2) = \lambda^*$ for some λ^*. Of course, this would be an exceptional coincidence, but the standardized alpha is just a lower bound for the true standardized reliability when this condition fails. Luckily, there is a natural way to consistently estimate R_w without making the unwarranted assumption that $\lambda_i/(\lambda_i^2 + \sigma_i^2) = \lambda^*$. Just to use the plug-in estimator for R_S (9). This is not hard, as estimation of λ, σ with structural equation modeling programs is routine.

Table 4: $n \times$ MSE for sample R_S and sample α_S

	$n = 50$	$n = 200$	$n = 1000$	$n = 5000$
R_S	0.17	0.18	0.18	0.17
α_S	0.24	0.23	0.33	0.71

Example 12. `lavaan` estimated λ and σ for the agreeableness factor on the the `bfi` data from `psychTools` (Revelle, 2019b), the data used in Example 11. The estimates are $\hat{\lambda} = (0.53, -0.77, -0.99, -0.72, -0.79)$ and $\hat{\sigma} = (1.3, 0.89, 0.85, 1.3, 0.98)$. Using formula (2) and (9) and flipping the sign of reverse-scored item (9), we find $\hat{R}_s = 0.72$ and $\hat{\alpha}_s = 0.71$. The difference $\hat{\alpha}_s - \hat{R}_s = -0.011$ is not alarmingly high, but keep in mind that the estimator is inconsistent. Table 4 contains mean squared error calculations for different ns when with data simulated from a congeneric model with $\lambda = \hat{\lambda}$ and $\sigma = \hat{\sigma}$. The mean squared error is uniformly lower for $\hat{R}_s$ than $\hat{\alpha}_s$. The results are similar for the other traits in the `bfi` data.

To recap: (i) You should neither avoid standardization altogether nor standardize with $(\lambda_i^2 + \sigma_i^2)^{-1/2}$. You should standardize with σ_i^{-1} or use the optimal weights, as this yields better results in the most plausible scenarios. (ii) Even if you decide to standardize with $(\lambda_i^2 + \sigma_i^2)^{-1/2}$, please avoid standardized alpha, as it will not equal R_S under any reasonable assumptions.

5   SITUATION C: THE ORDINAL ALPHA

Recall the congeneric model from Section 2,

$$X = \lambda Z + \Psi^{1/2} \varepsilon. \tag{3}$$

Consider the scenario when we do not observe the X_is directly, but rather X_is discretized into m_i categories according to

$$Y_i = j1[\tau_{i(j-1)} \leq X_i \leq \tau_{ij}], \quad j = 1, \ldots, m_i. \tag{20}$$

Here $-\infty = \tau_{i0} < \tau_{i1} < \ldots < \tau_{im_i} = \infty$ are thresholds for each i and $1[A]$ is the indicator function of A. This is a common model for Likert scales with m_i levels for the ith item.

If X is multivariate normal, model (20) is a congeneric *normal-ogive model* (McDonald, 1997), also known as a discretized factor analysis model (Takane & De Leeuw, 1987). Under multivariate normality, the correlation matrix $\Phi = \text{Cor}\, X$ is point-identified from the distribution of Y, and is called the *polychoric correlation matrix*. It can be estimated by maximum likelihood directly or by a two-step procedure (Olsson, 1979). To make the model identified, usual practice is to assume that $\text{Cov}\, X = \text{Cor}\, X = \Phi$.

Zumbo et al. (2007) proposed what we will call the *theoretical ordinal alpha*

$$\alpha = \frac{k}{k-1}\left(1 - \frac{k}{\mathbf{1}^T \Phi \mathbf{1}}\right), \tag{21}$$

where $\Phi = \text{Cor}\,X$ is polychoric correlation matrix of Y. Theoretical ordinal alpha is a standardized alpha, but it is based on the latent correlation matrix between the unobserved X instead of the correlation matrix of the observed Y.

Just as the standardized alpha is an estimator of the standardized reliability eq. (9), the theoretical ordinal alpha is an estimator of the *theoretical ordinal congeneric reliability*. In full generality, define the theoretical weighted ordinal reliability, the ordinal analogue of the weighted reliability (5), by

$$\frac{(w^T \lambda^*)^2}{(w^T \lambda^*)^2 + (w^2)^T \sigma^{*2}}. \tag{22}$$

Here $\lambda_i^* = \lambda_i/(\lambda_i^2 + \sigma_i^2)^{1/2}$ and $\sigma_i^* = \sigma_i/(\lambda_i^2 + \sigma_i^2)^{1/2}$ are the standardized factor loadings and error standard deviations. Similarly, we can define the ordinal analogue of the weighed alpha (10) as the weighted alpha based on the polychoric correlation coefficient. Under the assumption that $\text{Cov}\,X$ is a correlation matrix, the ordinal variants of the standardized alpha and coefficient alpha are the same, and likewise for the ordinal standardized reliability and ordinal congeneric reliability. This makes it unnecessary talk about quantities such as the theoretical standardized ordinal alpha, as it equals the theoretical ordinal alpha.

There are several reasons to avoid the theoretical ordinal alpha. If your goal is to estimate the congeneric reliability or the standardized reliability under the congeneric model, the theoretical ordinal alpha is inappropriate since it does not equal the congeneric reliability, even under the parallel model (Chalmers, 2018, p. 1062, "Misconception 2"), a point conceded by Zumbo and Kroc (2019). But this is not the worst problem. Even if you are interested in estimating a reliability coefficient for the multivariate normal-ogive model, you should avoid the theoretical alpha.

Theoretical ordinal alpha is not connected to any predictor of Z that depends on the data Y, only the squared correlation $\text{Cor}^2(\hat{Z}, Z)$ where $\hat{Z} = \sum_{i=1}^{k} w_i X_i$. This correlation is barely interesting though, as we do not observe the latent X_is. The theoretical alpha does not have the concrete interpretation of the congeneric alpha mentioned in Section 2. The question "how good can I expected my prediction of Z based on X to be?" does not make practical sense to ask, as X is unknown and you cannot use it for your prediction. This lack of knowledge about X also makes correction for attenuation impossible. Since X is unknown, it is impossible to directly calculate the correlation between a function of X and a function of some other quantity X'. Due to this, the correction for attenuation formula will not work.

Moreover, if you want the standardized reliability of the unobserved X, the ordinal reliability is a better choice. For the conditions required for standardized alpha to equal the standardized reliability are still needed, i.e., the weighted τ-equivalent model (19). The theoretical ordinal reliability can be estimated using e.g. `lavaan` without much added complexity. Using the same arguments as in the previous section, the theoretical ordinal reliability with optimal weights (corresponding to coefficient H) or weights $w_i = \sigma_i$ (corresponding to the sigma reliability) should be preferred to the standardized reliability too.

Finally, theoretical ordinal alpha only makes sense in relation to the congeneric multivariate normal-ogive model. But this model depends on the fundamentally unverifiable assumption of multivariate normality of X. As Foldnes and Grønneberg (2019) argue, the polychoric correlation coefficient is not robust against normality in general.

5.1   A concrete ordinal reliability

The main argument of Zumbo and Kroc (2019)'s rebuttal of Chalmers (2018) is that the congeneric normal-ogive model (20) is a better approximation to reality than the congeneric model (3). Assume we believe

in the discretized model and can be reasonably sure that X is multivariate normal. Then the theoretical ordinal reliability would still not be appropriate, as it is theoretical and not concrete. However, it is possible to create a class of concrete reliability coefficients for the congeneric normal-ogive model.

Let $\hat{X}_i$ be the conditional mean of X_i given Y_i. Then

$$\hat{X}_i = -\frac{\phi(\tau_{iY_i}) - \phi(\tau_{i(Y_i-1)})}{F(\tau_{iY_i}) - F(\tau_{i(Y_i-1)})}, \tag{23}$$

where ϕ is the density of a standard normal and F is the cumulative distribution function of a standard normal (Johnson, Kotz, & Balakrishnan, 1994, Section 10.1). Since $\hat{X}_i$ is the conditional expectation of X_i given Y_i, it is the optimal predictor of X_i given Y_i in the sense of minimizing the mean squared error. A natural class of predictors of Z given Y is are the linear functions of $\hat{X}_i$, i.e., $\hat{Z} = \sum_{i=1}^k w_i\hat{X}_i$ for some weights w_i. Due to the connection between reliabilities and predictors in Definition 1, this gives rise to the analogue of the weighted reliability (5) in the context of the congeneric normal-ogive model (20).

Definition 13. Let w be a vector of weights and $\hat{Z} = \sum_{i=1}^k w_i\hat{X}_i$, with $\hat{X}_i$ as in (23). The *concrete ordinal reliability* is $R'_w = \mathrm{Cor}^2(\hat{Z}, Z)$.

Now define the natural analogues of the reliabilities in Section 2: The concrete congeneric reliability is the weighted ordinal reliability with unit weights, while the concrete ordinal H is the weighted ordinal reliability with optimal weights.

The following theorem shows how to calculate the concrete ordinal alpha and some of its properties under the assumption that Φ is a correlation matrix.

Theorem 14. *Let $\hat{Z} = \sum_{i=1}^k w_i\hat{X}_i$ with $\hat{X}$ defined as in (23) and $\Xi = \mathrm{Cov}\,\hat{X}$. Assume X is multivariate normal with standard normal marginals, that is, $(\lambda_i^2 + \sigma_i^2)^{1/2} = 1$ for all i.*

  (i) *Let $v_i = \lambda_i/[\sigma_i^2(1 + k\overline{\lambda^2\sigma^{-2}})]$, $i = 1, \ldots, k$ be the Thurstone weights. Then $\sum_{i=1}^k v_i\hat{X}_i$ is the optimal predictor of Z, i.e., it minimizes the mean squared error $\mathrm{MSE}_Z(\hat{Z})$.*

  (ii) *The concrete ordinal reliability is*

$$\begin{aligned} R'_w &= \frac{(w^T\Xi v)^2}{w^T\Xi w}, \tag{24} \\ &= 1 - \mathrm{MSE}_Z(v_0\hat{Z}), \end{aligned}$$

  where $v_i = \lambda_i/[\sigma_i^2(1 + k\overline{\lambda^2\sigma^{-2}})]$, $i = 1, \ldots, k$ are the Thurstone weights for X and $v_0 = w^T\Xi v$ is the minimizer of $\mathrm{MSE}_Z(v_0\hat{Z})$. Moreover, $R'_v = v^T\Xi v$ when v are the Thurstone weights.

  (iii) *When the underlying congeneric model is parallel and $w = \mathbf{1}$,*

$$\begin{aligned} R'_w &= \frac{\lambda^2}{(k\lambda^2 + \sigma^2)^2}\mathbf{1}^T\Xi\mathbf{1}, \tag{25} \\ &= \alpha_S\frac{\mathbf{1}^T\Xi\mathbf{1}}{\mathbf{1}^T\Psi\mathbf{1}}, \end{aligned}$$

  where α_S equals the theoretical ordinal alpha.

  (iv) *For any weight w, $R'_w \to R_w$ when $\Xi \to \Phi$.*

Proof. See Appendix 1, page 24.   □

Point (i) is a strong justification for using $\hat{Z} = \sum_{i=1}^k v_i\hat{X}_i$, where v_i are the Thurstone weights. This is not only the optimal linear predictor of Z given the data Y, it is the optimal predictor of Z period. The

analogous result for the congeneric model with multivariate normal X is also true, that is, $Z = \sum_{i=1}^{k} v_i X_i$ is the optimal predictor of Z. (This follows from the proof of Theorem 14.) The representation of the ordinal reliability in (ii) is similar to the representation of the weighted reliability in Proposition 2, but the covariance Ξ is used instead of $\Sigma = \operatorname{Cov} X$. Moreover, $v^T \Xi v$ is easy to estimate provided we are able to estimate the Thurstone weights v. Point (iii) demonstrates that the concrete ordinal congeneric reliability (R' with unit weights) can be calculated without running a factor analysis, since the average correlation $\mathbf{1}^T \Xi \mathbf{1}/\mathbf{1}^T \Psi \mathbf{1}$ is a simple function of the polychoric correlation matrix and the covariance matrix of $\hat{X}$. It also provides a correction factor for the theoretical ordinal alpha as an estimator of the concrete ordinal congeneric reliability.

Chalmers (2018, p. 1068) discusses one possible use of the ordinal alpha, namely as the theoretical reliability when the number of thresholds τ goes towards infinity. Point (iv) makes this claim concrete.

The expectation of $\hat{X}_i$ is

$$E\hat{X}_i = -\sum_{k=1}^{m_i} [\phi(\tau_{iY_i}) - \phi(\tau_{i(Y_i-1)})] = 0,$$

and the covariance matrix $\operatorname{Cov} \hat{X} = \Xi$ has elements

$$
\begin{aligned}
E(\hat{X}_i \hat{X}_j) &= E\left(\sum_{k=1}^{m_i} v_{ik} 1[Y_i = k] \sum_{l=1}^{m_j} v_{jl} 1[Y_j = l] \right), \\
&= \sum_{k=1}^{m_i} v_{ik}^2 P(\tau_{i(k-1)} \le Z_1 \le \tau_{ik}), \text{ when } i = j, \quad\quad (26) \\
&= \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} v_{ik} v_{jl} P(\tau_{i(k-1)} \le Z_1 \le \tau_{ik}, \tau_{j(l-1)} \le Z_2 \le \tau_{jl}), \text{ when } i \ne j. \quad (27)
\end{aligned}
$$

Here $Z = (Z_1, Z_2)$ is a bivariate normal with variances equal to 1, and

$$v_{ik} = -\frac{\phi(\tau_{ik}) - \phi(\tau_{i(k-1)})}{F(\tau_{ik}) - F(\tau_{i(k-1)})}.$$

There are two natural estimators of Ξ. The plug-in estimator $\hat{\Xi}_p$ is calculated by plugging the estimated polychoric correlation $\hat{\Phi}$ and thresholds $\hat{\tau}$ into (26) and (27). The sample estimator $\hat{\Xi}_s$ equals the sample covariance matrix of $\hat{X}_i$ when $\hat{X}_i$ is formed using the estimated thresholds $\hat{\tau}$. The estimators tend to give roughly the same results, but the sample estimator is faster to compute, as it foregoes the computationally expensive bivariate normal probabilities.

The R-package `conogive` is the companion R-package of this paper, available at https://github.com/JonasMoss/conogive. It can used to calculate concrete ordinal reliabilities.

1. The function `conogive` estimates a congeneric normal-ogive model using maximum likelihood. It is a wrapper for the `fa` and `polychoric` functions of the `psych` package (Revelle, 2019a).

2. The function `ordinal_r` calculates the concrete ordinal reliability, using either $\hat{\Xi}_s$ or $\hat{\Xi}_p$ as estimators of Ξ. It supports optimal weights and equal weights.

3. The `predict` function predicts the value of Z given $\hat{X}_i$.

4. The functions `theoretical_ordinal_r` and `theoretical_ordinal_alpha` calculate the theoretical ordinal reliability and the theoretical ordinal alpha.

Example 15. We will yet again have a look at 6-level agreeableness factor from the personality data `bfi` from `psychTools`. The calculations in this example were done with `conogive`. The resulting estimates are

$$
\begin{aligned}
\hat{\lambda} &= (0.43, 0.71, 0.80, 0.52, 0.67), \\
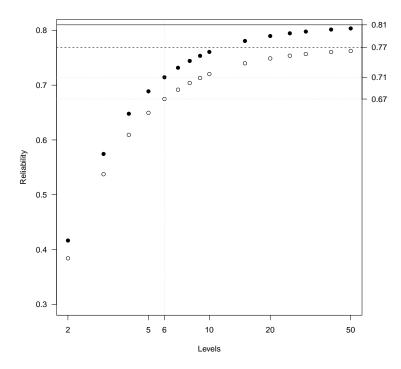\hat{\sigma} &= (0.90, 0.70, 0.59, 0.86, 0.74).
\end{aligned}
$$

Figure 1: Concrete ordinal reliability (○) and concrete ordinal H (●) for a selection of Likert levels. The theoretical coefficient H is 0.81 (solid line), and the theoretical composite reliability is 0.77 (dashed line). All cutoffs are uniform, and every reliability has been calculated with $\lambda = (0.43, 0.71, 0.80, 0.52, 0.67)$ and $\sigma = (0.90, 0.70, 0.59, 0.86, 0.74)$. The original number of levels (6) and their corresponding concrete ordinal reliabilities ($R'_H = 0.71$, $R' = 0.67$) are marked for convenience. Notice the logarithmic x-axis.

The concrete ordinal reliability with optimal weights is $\hat{R}'_H = 0.68$, and with equal weights it is $\hat{R}' = 0.64$. But the theoretical ordinal H is 0.81 and the theoretical ordinal congeneric reliability is 0.77, both missed by large margin. In comparison, the reliabilities calculated under the congeneric model with lavaan (Rosseel, 2012) are $\hat{R}_H = 0.72$, $\hat{R} = 0.71$, and $\hat{R}_S = 0.72$, while $\hat{\alpha} = 0.70$ and $\hat{\alpha}_S = 0.71$.

These reliabilities are calculated under $k = 6$ Likert levels. The theoretical ordinal reliability does not change with k, but the concrete reliability does. More Likert levels should make the reliabilities higher, but by how much? To answer this question, we calculated R'_H and R' for $k = 2, \ldots, 50$, with $\lambda = \hat{\lambda}$ and $\sigma = \hat{\sigma}$. The thresholds τ are $\{F^{-1}[i/(k+1)]\}_{i=1}^{k+1}$, where F^{-1} is the quantile function of a standard normal. These calculations approximate the reliabilities we would have obtained had we used k Likert levels, as the thresholds for $k \neq 6$ cannot be deduced from the data.

Figure 1 shows the results. The difference between the concrete ordinal congeneric reliability and the concrete ordinal H is approximately 0.04 for all k, but it is unclear if this relationship can be expected in other examples. The distance from theoretical ordinal reliability to the concrete is unacceptably large when the number of Likert levels is less than or equal to 6, where it equals about 0.1 for both reliabilities.

6   Conclusion

The definition of reliability (1) entails a strong dependence between the predictor $\hat{Z}$ and the reliability coefficient R. In particular, using standardized reliability (9) is only justified if $\hat{Z}$ is the standardized score. But since the standardized score should not be used, standardized reliability should not be used either.

McNeish and Wolf (2020) uses similar arguments against sum-scores, and makes a strong case for R_H, the coefficient H. Section 4 introduced the sigma-reliability R_σ, but this reliability suffers from the same issue as the congenerifc reliability R: Why bother with R_σ when R_H is superior by definition? The main point of R_σ is to give R_H skeptics an alternative to standardized reliability and congeneric reliability, and Theorem 10 shows that R_σ is superior to both.

The concrete ordinal reliability of Definition 13 takes the dependence between reliabilities and predictors even further. This newly proposed reliability arguably measures what it should measure; how well can we predict a latent variable Z given our observed data? On the other hand, the theoretical ordinal alpha only measures how well we can predict Z given unobserved data. The difference between the concrete ordinal reliability and the theoretical ordinal reliability is larger than 0.1 in our example with six Likert levels. This difference is significant and should not be ignored.

7   Acknowledgements

APPENDIX 1: PROOFS

Proof of Lemma 1. Given β, the optimal α is $E(Z) - \beta E(\hat{Z})$ since $\operatorname{argmin}_\mu E[(\mu - Z)^2]$ is the mean of Z for any Z. Now observe that $\operatorname{MSE}_Z(\alpha + \beta \hat{Z})$ equals $\beta^2 \operatorname{Var}(\hat{Z}) - 2\beta \operatorname{Cov}(\hat{Z}, Z) + \operatorname{Var} Z^2$. Differentiate with respect to β and set equal to zero, so that $2\beta \operatorname{Var}(\hat{Z}) - 2 \operatorname{Cov}(\hat{Z}, Z) = 0$, which implies $\beta = \operatorname{Cov}(\hat{Z}, Z) / \operatorname{Var}(\hat{Z})$. Plug the expression of β into the expression for the MSE_Z to get eq. (4). □

Proof of Proposition 2. Since $\operatorname{Var} Z = 1$, the covariance between $\hat{Z}$ and Z is

$$\operatorname{Cov}(Z, \hat{Z}) = \operatorname{Cov}[Z, \textstyle\sum_{i=1}^k w_i(Z\lambda_i + \sigma_i \varepsilon)] = \sum_{i=1}^k w_i \lambda_i, \tag{28}$$

and the variance of $\hat{Z}$ is

$$\operatorname{Var}(\hat{Z}) = \operatorname{Var}[\textstyle\sum_{i=1}^k w_i(Z\lambda_i + \sigma_i \varepsilon)] = (\sum_{i=1}^k w_i \lambda_i)^2 + \sum_{i=1}^k w_i^2 \sigma_i^2.$$

Combining these two formulas gives us the squared correlation between Z and $\hat{Z}$,

$$R_w = \operatorname{Cor}^2(\hat{Z}, Z) = \frac{(\sum_{i=1}^k w_i \lambda_i)^2}{(\sum_{i=1}^k w_i \lambda_i)^2 + \sum_{i=1}^k w_i^2 \sigma_i^2}. \tag{29}$$

The second inequality is an application of Lemma 1.

The formula for w_0 follows from

$$E\{[Z - \textstyle\sum_{i=1}^k w_0 w_i(\lambda_i Z + \sigma_i \varepsilon_i)]^2\} = w_0^2 \sum_{i=1}^k w_i^2 \sigma_i^2 + (w_0 \sum_{i=1}^k w_i \lambda_i - 1)^2.$$

Differentiate with respect to w_0 to find the minimizer

$$w_0 = \frac{\sum_{i=1}^k w_i \lambda_i}{\sum_{i=1}^k w_i^2 \sigma_i^2 + (\sum_{i=1}^k w_i \lambda_i)^2}.$$

□

Proof of Proposition 4. Clearly, wX is τ-equivalent if and only if $w_i \lambda_i = w_j \lambda_j$ for each i, j, which implies $w^T \lambda = k \lambda_i w_i$. Then

$$
\begin{aligned}
\alpha_w &= \frac{k}{k-1}\left(1 - \frac{w^T \operatorname{diag} \Sigma w}{w^T \Sigma w}\right), \\
&= \frac{k}{k-1} \frac{k^2 \lambda_1^2 w_1^2 - k\lambda_1^2 w_1^2}{k^2 \lambda_1^2 w_1^2 + \sum_{i=1}^k w_i^2 \sigma_i^2}, \\
&= \frac{(w^T \lambda)^2}{(w^T \lambda)^2 + w^T \Psi w}.
\end{aligned}
$$

The bias term for coefficient α_w can be derived as follows:

$$
\begin{aligned}
\alpha_w &= \frac{k}{k-1}\left(1 - \frac{w^T(\operatorname{diag} \Sigma)w}{w^T \Sigma w}\right), \\
&= \frac{k}{k-1} \frac{(w^T \lambda)^2 - (w^2)^T \lambda^2}{(w^T \lambda)^2 + (w^2)^T(\sigma^2)}, \\
&= \frac{k}{k-1} \frac{(w^T \lambda)^2 - (w^T \lambda)^2/k + (w^T \lambda)^2/k - (w^2)^T \lambda^2}{(w^T \lambda)^2 + (w^2)^T(\sigma^2)}, \\
&= R_w - \frac{k}{k-1} \frac{(w^2)^T \lambda^2 - (w^T \lambda)^2/k}{(w^T \lambda)^2 + (w^2)^T(\sigma^2)}, \\
&= R_w - \frac{k}{k-1} \frac{\overline{w^2 \lambda^2} - \overline{w\lambda}^2}{k\overline{w\lambda}^2 + \overline{w^2 \sigma^2}}.
\end{aligned}
$$

Since $(\overline{w^2\lambda^2} - \overline{w\lambda}^2)/(k\overline{w\lambda}^2 + \overline{w^2\sigma^2}) \geq 0$ and equals 0 if and only if all products $w_i\lambda_i$ are equal by Chebyshev's inequality, $\alpha_w = R_w$ iff all $w_i\lambda_i$ are equal and is otherwise an underestimate. $\qquad\square$

This lemma comes in handy in the proof of Proposition 5 as it allows us to use Jensen's inequality.

Lemma 2. *Assume the congeneric model of (3) and let α be as in (1) and α_S be as in (2). Then $\alpha_S \geq \alpha$ if and only if*

$$\sum_{i \neq j} \frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_j^2 + \lambda_j^2)^{1/2}} \geq \sum_{i \neq j} \frac{\lambda_i \lambda_j}{\sum_{i=1}^{k}(\lambda_i^2 + \sigma_i^2)/k}. \tag{30}$$

Proof. From definition (2)

$$\alpha_S = \frac{k}{k-1}\left(1 - \frac{k}{\mathbf{1}^T\boldsymbol{\Phi}\mathbf{1}}\right) = \frac{k}{k-1}\left(1 - \frac{k}{\sum_{i \neq j} \frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_j^2 + \lambda_j^2)^{1/2}} + k}\right)$$

and definition (1)

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\operatorname{tr}\boldsymbol{\Sigma}}{\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}}\right) = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}(\sigma_i^2 + \lambda_i^2)}{\sum_{i \neq j}\lambda_i\lambda_j + \sum_{i=1}^{k}(\sigma_i^2 + \lambda_i^2)}\right)$$

Thus $\alpha_S \geq \alpha$ if and only if

$$\frac{k}{\sum_{i \neq j} \frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_j^2 + \lambda_j^2)^{1/2}} + k} \leq \frac{\sum_{i=1}^{k}(\sigma_i^2 + \lambda_i^2)}{\sum_{i \neq j}\lambda_i\lambda_j + \sum_{i=1}^{k}(\sigma_i^2 + \lambda_i^2)}$$

which happens if and only if

$$\sum_{i \neq j} \frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_j^2 + \lambda_j^2)^{1/2}} \geq \sum_{i \neq j} \frac{\lambda_i \lambda_j}{\sum_{i=1}^{k}(\lambda_i^2 + \sigma_i^2)/k}$$

$$\square$$

Proof of Proposition 5. (i) From eq. (30) and $\lambda_i = \lambda$ for all i, we need to show

$$\sum_{i \neq j} \frac{1}{(\sigma_i^2 + \lambda^2)^{1/2}(\sigma_j^2 + \lambda^2)^{1/2}} \geq \sum_{i \neq j} \frac{1}{\sum_{i=1}^{k}(\lambda^2 + \sigma_i^2)/k} = \frac{k^2(k-1)}{\sum_{i=1}^{k}(\lambda^2 + \sigma_i^2)}$$

Recall Jensen's inequality (Bullen, Mitrinovic, & Vasic, 2013, Theorem 12 of Chapter 1): When f is convex, $E[f(X)] \geq f[E(X)]$ with equality if and only if f is linear or X is constant with probability 1. Define the random variable X with outcomes $(\sigma_1^2, \ldots, \sigma_k^2)$ and $P(X = \sigma_i^2) = 1/k$ for each i, and let X' be identically distributed to X but independent of X. Define the function $f(x, y) = (x + \lambda^2)^{-1/2}(y + \lambda^2)^{-1/2}$. Then f is convex for any λ. Applying the expectation operator first yields,

$$f(E(X, X')) = \frac{1}{\lambda^2 + \frac{1}{k}\sum_{i=1}^{k}\sigma_i^2} = \frac{k}{\sum_{i=1}^{k}(\sigma_i^2 + \lambda^2)}, \tag{31}$$

while applying f first yields

$$E(f(X, X')) = \frac{1}{k^2}\sum_{i \neq j} \frac{1}{(\sigma_i^2 + \lambda^2)^{1/2}(\sigma_j^2 + \lambda^2)^{1/2}} + \frac{1}{k^2}\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \lambda^2}. \tag{32}$$

Using Jensen's inequality on eq. (31) and eq. (32) gives us

$$\frac{1}{k} \sum_{i \neq j} \frac{1}{(\sigma_i + \lambda^2)^{1/2}(\sigma_j + \lambda^2)^{1/2}} \geq k \frac{k}{\sum_{i=1}^{k}(\sigma_i^2 + \lambda^2)} - \frac{1}{k} \sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \lambda^2}.$$

By the harmonic mean–arithmetic mean inequality,

$$\frac{1}{k} \sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \lambda^2} \geq \frac{k}{\sum_{i=1}^{k}(\sigma_i^2 + \lambda^2)},$$

with equality if and only if all σ_is are equal. Hence

$$\sum_{i \neq j} \frac{1}{(\sigma_i + \lambda^2)^{1/2}(\sigma_j + \lambda^2)^{1/2}} \geq \frac{k^2(k-1)}{\sum_{i=1}^{k}(\sigma_i^2 + \lambda^2)},$$

with equality if and only if all σ_is are equal.

(ii) Let $k = 2$. Then (30) simplifies to

$$\frac{1}{\frac{\lambda_1 \lambda_2}{((\sigma_1^2 + \lambda_1^2)(\sigma_2^2 + \lambda_2^2))^{1/2}} + 1} \leq \frac{(\lambda_1^2 + \sigma_1^2) + (\lambda_2^2 + \sigma_2^2)}{2\lambda_1\lambda_2 + (\lambda_1^2 + \sigma_1^2) + (\lambda_2^2 + \sigma_2^2)}$$

and further to

$$\frac{((\sigma_1^2 + \lambda_1^2)(\sigma_2^2 + \lambda_2^2))^{1/2}}{\lambda_1\lambda_2 + ((\sigma_1^2 + \lambda_1^2)(\sigma_2^2 + \lambda_2^2))^{1/2}} \geq \frac{[(\lambda_1^2 + \sigma_1^2) + (\lambda_2^2 + \sigma_2^2)]/2}{\lambda_1\lambda_2 + [(\lambda_1^2 + \sigma_1^2) + (\lambda_2^2 + \sigma_2^2)]/2}$$

By the arithmetic mean–geometric mean inequality, $1/2[(\lambda_1^2 + \sigma_1^2) + (\lambda_2^2 + \sigma_2^2)] \geq ((\sigma_1^2 + \lambda_1^2)(\sigma_2^2 + \lambda_2^2))^{1/2}$ with equality only if $\lambda_1^2 + \sigma_1^2 = \lambda_2^2 + \sigma_2^2$.

Let $k > 2$ and assume all $\lambda_j = 0$ except λ_1, λ_2. By (30), $\alpha_S > \alpha$ if and only if

$$\frac{k}{2\frac{\lambda_1 \lambda_2}{(\sigma_1^2 + \lambda_1^2)^{1/2}(\sigma_1^2 + \lambda_2^2)^{1/2}} + k} < \frac{\lambda_1^2 + \lambda_2^2 + \sum_{i=1}^{k} \sigma_i^2}{2\lambda_1\lambda_2 + \lambda_1^2 + \lambda_2^2 + \sum_{i=1}^{k} \sigma_i^2},$$

or, equivalently,

$$\frac{k(\sigma_1^2 + \lambda_1^2)^{1/2}(\sigma_1^2 + \lambda_2^2)^{1/2}}{2\lambda_1\lambda_2 + k(\sigma_1^2 + \lambda_1^2)^{1/2}(\sigma_1^2 + \lambda_2^2)^{1/2}} < \frac{\lambda_1^2 + \lambda_2^2 + \sum_{i=1}^{k} \sigma_i^2}{2\lambda_1\lambda_2 + \lambda_1^2 + \lambda_2^2 + \sum_{i=1}^{k} \sigma_i^2},$$

Now, $k(\sigma_1^2 + \lambda_1^2)^{1/2}(\sigma_1^2 + \lambda_2^2)^{1/2} > \lambda_1^2 + \lambda_2^2 + \sum_{i=1}^{k} \sigma_i^2$ can be made true by choosing $\sigma_i^2 = 0$ (or very small) for $i > 2$ and choosing a large k. For the other way around, just choose σ_i^2 large enough.

(iii) Observe that $\alpha_S \geq R$ if and only if

$$\frac{k}{k-1} \left(1 - \frac{k}{\sum_{i \neq j} \frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_j^2 + \lambda_j^2)^{1/2}} + k} \right) \leq \frac{(\sum_{i=1}^{k} \lambda_i)^2}{(\sum_{i=1}^{k} \lambda_i)^2 + \sum_{i=1}^{k} \sigma_i^2}. \tag{33}$$

Now we show that $\alpha_S > R$ is possible. First assume that $\lambda_i = \sigma_i$ for each i. Then

$$\frac{\lambda_i \lambda_j}{(\sigma_i^2 + \lambda_i^2)^{1/2}(\sigma_i^2 + \lambda_j^2)^{1/2}} = \frac{\lambda_i \lambda_j}{2\lambda_i\lambda_j} = \frac{1}{2},$$

and the left hand side of eq. (33) equals

$$\frac{k}{k-1} \left(1 - \frac{2}{(k+1)} \right) = \frac{k}{k+1} = \frac{1}{1 + 1/k}.$$

On the other hand, the right hand side becomes

$$\frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \lambda_i^2} = \left(1 + \frac{\sum_{i=1}^k \lambda_i^2}{(\sum_{i=1}^k \lambda_i)^2}\right)^{-1}.$$

By Chebyshev's inequality, $k\sum_{i=1}^k \lambda_i^2 \geq (\sum_{i=1}^k \lambda_i)^2$, hence $(\sum_{i=1}^k \lambda_i^2)/(\sum_{i=1}^k \lambda_i)^2 \geq 1/k$ with equality if and only if all λ_is are equal. It follows that $\alpha_S \geq R$, with strict inequality when the λ_is are unequal.

Now we show that $\alpha_S \leq R$ is possible. Assume there is an a such that $\sigma_i^2 + \lambda_i^2 = a$ for all i, that is, $\sigma_i^2 + \lambda_i^2$ is constant. In this case, eq. (33) becomes

$$\frac{k}{k-1}\left(1 - \frac{ak}{\sum_{i\neq j} \lambda_i\lambda_j + ak}\right) \leq \frac{(\sum_{i=1}^k \lambda_i)^2}{\sum_{i\neq j} \lambda_i\lambda_j + ak},$$

which can be further reduced to

$$\frac{k}{k-1}\frac{\sum_{i\neq j} \lambda_i\lambda_j}{\sum_{i\neq j} \lambda_i\lambda_j + ak} \leq \frac{\sum_{i\neq j} \lambda_i\lambda_j + \sum_{i=1}^k \lambda_i^2}{\sum_{i\neq j} \lambda_i\lambda_j + ak}.$$

This inequality is true if $(\sum_{i=1}^k \lambda_i)^2 \leq k\sum_{i=1}^k \lambda_i^2$, which always holds by Chebyshev's inequality, with equality only if all λs are equal.

That $\alpha_S = R$ is possible when some λs fail to be equal follows from the intermediate value theorem. Let $A = \{(\lambda, R) \mid \lambda_i \neq \lambda_j \text{ for some } j\}$. Then A is connected, and by the intermediate value theorem there is a (λ, R) where $[\alpha_S - R](\lambda, \sigma) = 0$, for there are elements in A where $\alpha_S - R$ takes on negative values and elements where $\alpha_S - R$ takes on positive values. □

Proof of Theorem 7. (i) Assume the parallel model. Define ψ^2 and ρ by $\sigma^2 + \lambda^2 = \psi^2$ and $\lambda^2 = \psi^2\rho$. Then, by exercise 3.9 of Muirhead (1982, p. 114), the maximum likelihood estimates of $\hat{\psi}^2$ and $\hat{\rho}$ (Kristof (1963, Section B) also worked out the maximum likelihood estimator of ρ) are

$$\hat{\psi}^2 = \operatorname{tr} S/k, \quad \hat{\rho} = \frac{\mathbf{1}^T S\mathbf{1} - \operatorname{tr} S}{(k-1)\operatorname{tr} S}1\{\mathbf{1}^T S\mathbf{1} \geq \operatorname{tr} S\}.$$

That $\hat{\rho} = 0$ when $\mathbf{1}^T S\mathbf{1} < \operatorname{tr} S$ is not explicitly stated in the exercise, but is required since $\rho \geq 0$ is needed for Σ to be positive definite. We find, using the invariance principle for maximum likelihood,

$$\hat{\lambda}^2 = \frac{\mathbf{1}^T S\mathbf{1} - \operatorname{tr} S}{k(k-1)}1\{\mathbf{1}^T S\mathbf{1} \geq \operatorname{tr} S\}.$$

Assume $\mathbf{1}^T S\mathbf{1} < \operatorname{tr} S$. Then $\hat{\sigma}^2 = \hat{\psi}^2 = \operatorname{tr} S/k$, as claimed in the statement of this theorem. If $\mathbf{1}^T S\mathbf{1} \geq \operatorname{tr} S$,

$$\hat{\sigma}^2 = \hat{\psi}^2(1 - \hat{\rho}) = \frac{\operatorname{tr} S}{k}\left(1 - \frac{\mathbf{1}^T S\mathbf{1} - \operatorname{tr} S}{(k-1)\operatorname{tr} S}\right) = \frac{\operatorname{tr} S - \mathbf{1}^T S\mathbf{1}/k}{(k-1)}.$$

As for (ii), first assume that $\hat{\lambda}^2 \neq 0$. By the invariance property of maximum likelihood and the definition of the congeneric reliability of eq. (8), the maximum likelihood estimator of the congeneric reliability is

$$
\begin{aligned}
\frac{(\mathbf{1}^T\hat{\lambda})^2}{(\mathbf{1}^T\hat{\lambda})^2 + \mathbf{1}^T\Psi\mathbf{1}} &= \frac{k(\mathbf{1}^T S\mathbf{1} - \operatorname{tr}S)}{k(\mathbf{1}^T S\mathbf{1} - \operatorname{tr}S) + (k\operatorname{tr}S - \mathbf{1}^T S\mathbf{1})}, \\
&= \frac{k(\mathbf{1}^T S\mathbf{1} - \operatorname{tr}S)}{(k-1)\mathbf{1}^T S\mathbf{1}}, \\
&= \frac{k}{k-1}\left(1 - \frac{\operatorname{tr}S}{\mathbf{1}^T S\mathbf{1}}\right).
\end{aligned}
$$

If the maximum likelihood estimator of λ^2 is 0, then $(\mathbf{1}^T\hat{\lambda})^2 = 0$, and consequently the maximum likelihood estimator of the reliability is 0. □

Proof of Theorem 8. (i) Let R_n, S_n be the sample correlation and covariance matrices and Φ, Σ be their population counterparts. By the Law of Large Numbers and Slutsky's theorem (van der Vaart, 2000, Lemma 2.8, p. 11), $R_n \to \Phi$ and $S_n \to \Sigma$ with probability 1 in the entry-wise matrix norm $\|A\|_1 = \sum_{j=1}^{k} \sum_{i=1}^{k} |a_{ij}|$. The mapping

$$f(A) = \frac{k}{k-1}\left(1 - \frac{\mathbf{1}^T(\operatorname{diag}A)\mathbf{1}}{\mathbf{1}^T A\mathbf{1}}\right)$$

is continuous with respect to the norm $\|\cdot\|_1$, and by the continuous mapping theorem (van der Vaart, 2000, Theorem 2.3, p. 7) we have $f(S_n) \to f(\Sigma)$ and $f(R_n) \to f(\Phi)$ with probability 1.

(ii) See Appendix 2 (p. 25).

(iii) van Zyl et al. (2000, eq. 22) showed that the asymptotic variance is

$$\sigma^2(\alpha) = \frac{2k}{(k-1)[1+\rho(k-1)]^3}[(k-1)\rho^3 - (2k-3)\rho^2 + (k-3)\rho + 1].$$

This can be simplified to

$$\sigma^2(\alpha) = \frac{2k(\rho-1)^2}{(k-1)[1+\rho(k-1)]^2}.$$

Hayashi and Kamata (2005, equation 10) found the asymptotic variance

$$\sigma^2(\alpha_S) = \frac{1}{(k-1)^2}\frac{1}{(1+\rho(k-1))^4}2\mathbf{1}^T Q_p(P \otimes P)Q_P^T\mathbf{1}.$$

The expression $2\mathbf{1}^T Q_p(P \otimes P)Q_P^T\mathbf{1}$ can be rewritten to (Hayashi & Kamata, 2005, Appendix 3)

$$2k(k-1)\{(k-1)^2\rho^4 - 2(k-1)(k-2)\rho^3 + (k^2 - 6k + 6)\rho^2 + 2(k-2)\rho + 1\},$$

which can be further simplified to $(\rho-1)^2(1+\rho(k-1)^2)^2$. Thus

$$\sigma^2(\alpha_S) = \frac{2k}{(k-1)}\frac{2k(\rho-1)^2}{(k-1)[1+\rho(k-1)]^2},$$

and the limits are equal.

Since

$$\alpha_S = \frac{k}{k-1}\left(1 - \frac{k}{k(k-1)\rho + k}\right) = \frac{k\rho}{(k-1)\rho + 1},$$

we get that $1 - \alpha_S = (1-\rho)/((k-1)\rho + 1)$. Taken together, the asymptotic variance is

$$\sigma^2(\alpha_S) = 2\frac{k}{k-1}(1-\alpha_S)^2.$$

Since $\hat{\alpha}$ is the maximum likelihood estimator of R when n is large it is efficient. And since the asymptotic variances of $\hat{\alpha}$ and $\hat{\alpha_S}$ coincide they are both efficient. $\qquad\square$

The proof of Theorem 10 uses the theory of power means. Let p be real number and $x = \{x_i\}_{i=1}^{k}$ be a sequence of positive reals. The *power mean* with exponent p is

$$M_p(x) = \begin{cases} (\frac{1}{k}\sum_{i=1}^{k} x_i^p)^{1/p} & p \neq 0, \\ (\prod_{i=1}^{k} x_i)^{1/k} & p = 0. \end{cases}$$

We will make use of a generalization of the arithmetic mean–geometric mean and harmonic mean–arithmetic mean inequalities. The *power mean inequality* (Bullen et al., 2013, Chapter III) is

$$M_p(x) \leq M_q(x), \quad p < q \qquad (34)$$

with equality if and only if $x_1 = \cdots = x_k$.

Moreover, a generalized Minkowski inequality holds for the power means. See (Bullen et al., 2013, Section 2.4, Theorem 9.) for proofs of this result.

Theorem 16 (Minkowski inequality). *Let x, y be k-ary sequences of non-negative reals. Then*

$$M_p^p(x) + M_p^p(y) \geq M_p^p(x+y), \ p > 1, \tag{35}$$

$$M_p(x) + M_p(y) = M_1(x+y), \ p = 1, \tag{36}$$

$$M_p^p(x) + M_p^p(y) \leq M_p^p(x+y), \ p < 1. \tag{37}$$

Here eq. (35) *and eq.* (37) *are equalities if and only there are some* $\lambda, \mu \in \mathbb{R}$, *not both 0, such that* $\lambda x + \mu y = 0$.

This lemma is instrumental to the proof of Theorem 10.

Lemma 3. *Let* $\{x_i\}_{i=1}^k$ *be some positive real numbers,* $a > 0$, *and* $M_p(x)$ *the power mean with exponent* p.

$$M_p(x+a)M_q(x/(x+a)) \leq \bar{x} \tag{38}$$

when $p, q \leq 1$, *with equality if and only if* $x_1 = x_2 = \cdots = x_k$. *If* $-1 \leq p < 0$,

$$M_p(x) \leq M_p(x+a)M_p(x/(x+a)) \leq \bar{x} \tag{39}$$

holds too, again with equality if and only if $x_1 = x_2 = \cdots = x_k$.

Proof. Consider eq. (38) first. By the power mean inequality (eq. (34)) we have $M_p(x+a) \leq \bar{x} + a$ and $M_q(x/(x+a)) \leq \overline{x/(x+a)}$, hence it suffices to show $(\bar{x}+a)\overline{x/(x+a)} \leq \bar{x}$, or equivalently, $\overline{x/(x+a)} \leq \bar{x}/(\bar{x}+a)$. The function $x \mapsto x/(a+x)$ is concave, hence $E[f(X)] \leq f[E(X)]$ by Jensen's inequality. But then $\overline{x/(x+a)} \leq \bar{x}/(\bar{x}+a)$ follows.

Now let us take on eq. (39). Since $-1 \leq p < 0$, the function $f(x) = x^p/(x^p + a)$ is convex, as can be verified by differentation. Hence $f(E(x)) \leq Ef(x)$ by Jensen's inequality (here x is regarded as a uniform random variable). Since

$$E[f(x)] = k^{-1} \sum_{i=1}^k x_i^p/(x_i^p + a) = M_p^p(x/(x+a)),$$

and

$$f[E(x)] = \overline{x^p}/(\overline{x^p} + a) = M_p^p(x)/(M_p^p(x) + a),$$

we get that $M_p^p(x)/[M_p^p(x)+a] \leq M_p^p(x/(x+a))$. Moreover, $M_p^p(x) + a = M_p^p(x) + M_p^p(a) \leq M_p^p(x+a)$ by the Minkowski inequality (eq. (37)), hence $M_p(x)/M_p(x+a) \leq M_p(x/(x+a))$. This can be rearranged to get the left-most inequality. Now notice that the right-most inequality is a special case of eq. (38), and we are done. □

Proof of Theorem 10. (i) The equality $R_H \geq R_\sigma$ trivially true, so we will focus on the rest. Each reliability is on the form $a/(a+1)$ for some $a > 0$, and $a/(a+1) \geq b/(b+1)$ iff $a > b$. Let's take on $R_S > R$. From eq. (8) and eq. (9) we see that we must verify

$$k\overline{\lambda(\lambda^2 + \sigma^2)^{-1/2}}^2/\overline{\sigma^2(\lambda^2 + \sigma^2)^{-1}} \geq k\overline{\lambda}^2/\overline{\sigma^2}.$$

Since all λ_is are equal by assumption, this is equivalent to

$$\overline{\sigma^2/(\lambda^2 + \sigma^2)}/\overline{(\lambda^2 + \sigma^2)^{-1/2}}^2 \leq \overline{\sigma^2}$$

which is true by Lemma 3, eq. (38) using exponent $p = -1/2, q = 1$ and $a = \lambda^2$.

To prove $R_\sigma > R_S$, refer to eq. (18) and eq. (9) to verify that we need to show

$$\overline{\sigma^{-1}}^{-2} \leq \overline{\sigma^2/(\lambda^2 + \sigma^2)}/\overline{(\lambda^2 + \sigma^2)^{-1/2}}^2,$$

or equivalently,

$$M_{-1/2}(\sigma^2) \leq \overline{\sigma^2/\lambda^2 + \sigma^2}M_{-1/2}(\lambda^2 + \sigma^2).$$

This is a case of Lemma 3, eq. (39), as $\overline{\sigma^2/\lambda^2 + \sigma^2} \geq M_{-1/2}(\sigma^2/\lambda^2 + \sigma^2)$ by the power mean inequality.

(ii) That $R_{\hat{\sigma}} = R$ follows from Proposition 2, and $R_H \geq R_{\hat{\sigma}}$ since the weights of R_H are optimal. It suffices to show $R \geq R_S$. From the proof of (i), we must verify

$$\overline{\sigma^2(\lambda^2 + \sigma^2)}/\overline{(\lambda^2 + \sigma^2)^{-1/2}}^2 \geq \sigma^2$$

Since $\overline{\sigma^2(\lambda^2 + \sigma^2)} = \sigma^2\overline{(\lambda^2 + \sigma^2)}$, this is equivalent to

$$\overline{(\lambda^2 + \sigma^2)} \geq \overline{(\lambda^2 + \sigma^2)^{-1/2}}^2,$$

but this follows from the power mean inequality, eq. (34).

$\square$

Proof of Theorem (14). (ii) That $R' = 1 - \text{MSE}_Z(v_0\hat{Z})$ follows from Lemma 1 and the formula for the covariance, which we now show. By the law of total covariance, the covariance between $\sum_{i=1}^k w_i\hat{X}_i$ and Z is

$$\text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, Z) = E[\text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, Z \mid \{Y_i\})] + \text{Cov}(E[\sum_{i=1}^k w_i\hat{X}_i \mid \{Y_i\}], E[Z \mid \{Y_i\}]),$$
$$= \text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, E[Z \mid \{Y_i\}]).$$

The second equality holds since $\text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, Z \mid \{Y_i\}) = 0$, as X_i is a constant function of Y_i, and $E[\sum_{i=1}^k w_i\hat{X}_i|\{Y_i\}] = \sum_{i=1}^k w_i\hat{X}_i$ (slightly abusing notation). Since Y_i are functions of X, the expectation $E[Z \mid \{Y_i\}]$ equals $E[E[Z \mid X] \mid \{Y_i\}]$ by the conditional variant of the law of total expectation (Billingsley, 1995, Theorem 34.4). The conditional expectation $E(Z \mid X) = \sum_{i=1}^k v_iX_i$ under multivariate normality, where $v_i = \lambda_i/[\sigma_i^2(1 + k\overline{\lambda^2\sigma^{-2}})]$ are the Thurstone weights, see Tong (1990, Theorem 3.3.4). Consequently, $E[E(Z \mid X) \mid \{Y_i\}] = \sum_{i=1}^k v_i\hat{X}_i$ and

$$\text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, Z) = \text{Cov}(\sum_{i=1}^k w_i\hat{X}_i, \sum_{i=1}^k v_i\hat{X}_i) = w^T\Xi v.$$

Standardize to find the correlation $w^T\Xi v/(w^T\Xi w)^{1/2}$, and the squared correlation is $R' = (w^T\Xi v)^2/w^T\Xi w$. This completes point (i).

As for (iii), under the parallel model, $v = w_0\mathbf{1}$, where $w_0 = \lambda/[\sigma^2 + k\lambda^2]$ and the squared correlation is $w_0^2\mathbf{1}^T\Xi\mathbf{1} = \lambda^2\mathbf{1}^T\Xi\mathbf{1}/(k\lambda^2 + \sigma^2)^2$. For the second equation, first notice that

$$\alpha_S = \frac{k}{k-1}\frac{i^T\Phi i - k}{i^T\Phi i} = k^2\frac{\overline{\rho}}{i^T\Phi i},$$

where $\overline{\rho}$ is the unique correlation. Since $(k\lambda^2 + \sigma^2)^2 = (i^T\Phi i/k)^2$,

$$\frac{\lambda^2}{(k\lambda^2 + \sigma^2)^2}\mathbf{1}^T\Xi\mathbf{1} = k^2\frac{\overline{\rho}}{(i^T\Phi i)^2}\mathbf{1}^T\Xi\mathbf{1} = \alpha_S\frac{\mathbf{1}^T\Xi\mathbf{1}}{\mathbf{1}^T\Psi\mathbf{1}},$$

as claimed.

(i) Follows from the representation $R_w = (w^T\Xi v)^2/(w^T\Xi w)$. We use the derivatives $\frac{d}{dw}w^T\Xi w = 2w^T\Xi$ and $\frac{d}{dw}w^T\Xi v = v^T\Xi$, so that $\frac{d}{dw}(w^T\Xi v)^2 = 2(w^T\Xi v)v^T\Xi$. Now apply the quotient rule

$$\frac{d}{dx}\frac{(x^T\Xi y)^2}{x^T\Xi x} = \frac{2(w^T\Xi v)v^T\Xi(w^T\Xi w) - (w^T\Xi v)^2 2w^T\Xi}{(w^T\Xi w)^2}$$

This is equal to 0 when $w = v$, for $v^T\Xi(v^T\Xi v)^2 - (v^T\Xi v)^2 v^T\Xi = 0$.

$\square$

<center>APPENDIX 2: PROOF OF THEOREM 8 (II)</center>

In this appendix we prove a more detailed variant of Theorem 8 (ii).

Proposition 17. *Assume the parallel model with finite fourth moments, $k \times k$ covariance matrix Σ, variances ϕ^2, and correlation ρ. Then the asymptotic variances of standardized alpha and coefficient alpha are both equal to*

$$\sigma^2 = \sigma_s^2 = \frac{1}{(k-1)^2} \frac{1}{[(k-1)\rho + 1]^4} \frac{1}{\phi^4} v^T \Gamma v, \tag{40}$$

where

$$\Gamma = E[XX^T \otimes XX^T] - (\operatorname{vec}\Sigma)(\operatorname{vec}\Sigma)^T,$$

and

$$v = [(k-1)\rho + 1]\operatorname{vec}(I_p) - \operatorname{vec}(\mathbf{1}\mathbf{1}^T).$$

The operation $\otimes$ denotes the Kronecker product. The proof and statement of Proposition (17) makes use of three kinds of *vectorizations*, operations that map matrices into vectors. The first operation is full vectorization, vec A, which maps the $m \times n$ matrix A to the mn column vector $(A_{.1}, A_{.2}, \ldots, A_{.n})$ (Magnus & Neudecker, 2019, Chapter 2.4). The second operation is half-vectoriazation, where vech A maps the symmetric $n \times n$ matrix A to the $n(n+1)/2$ vector obtained by stacking the columns of the lower triangular part of A (Magnus & Neudecker, 2019, Chapter 3.8). The third w(A) operation maps the symmetric $n \times n$ matrix with zero-diagonal to the $n(n-1)/2$ vector obtained by stacking the columns of the infradiagonal elements of A; that is, it the variant of half-vectorization that ignores the main diagonal (Neudecker & Satorra, 1996).

Define the transition matrices L_k by vec $A = L_k^T \operatorname{vech}(A)$ and L_k^- by $\operatorname{vech}(A) = L_k^{-T} \operatorname{vec} A$ when the $k \times k$ matrix A is symmetric. These matrices are called K_k and K_k^- by van Zyl et al. (2000). The matrix D satisfies vec $A = D\operatorname{w}(A)$ and $\operatorname{w}(A) = \frac{1}{2}D^T \operatorname{vec}(A)$ when A is symmetric with only 0s on the diagonal (Neudecker & Satorra, 1996). The matrix K is the *commutation matrix*, making $K \operatorname{vec} X = \operatorname{vec} X^T$ (see Magnus & Neudecker, 2019, Chapter 3.8), and K_d is the diagonal of K. In general, if A is a square matrix, A_d is the matrix with the same diagonal as A and off-diagonal elements set to 0.

Lemma 4. *Assume the parallel model with variance ϕ^2 and correlation ρ, covariance matrix Σ, correlation matrix Φ. Then the asymptotic variance for coefficient alpha is as in Proposition 17. The asymptotic variance for standardized alpha is*

$$\sigma^2 = \frac{1}{(k-1)^2} \frac{1}{[\{(k-1)\rho + 1]^4} \frac{1}{\phi^4} \mathbf{1}_*^T D^T \Psi_* D \mathbf{1}_*,$$

where $\mathbf{1}_*$ has $\frac{1}{2}k(k-1)$ elements and

$$\Psi_* = [I - (I \otimes \Phi)K_d]\Gamma[I - K_d(I \otimes \Phi)].$$

Proof. We take on coefficient alpha first. In the case of multivariate normality, $\sigma^2 = h^T L_k^T 2(\Sigma \otimes \Sigma)L_k h$, where h is (van Zyl et al., 2000, eq. 19)

$$h = (\mathbf{1}^T\Sigma\mathbf{1})^{-2}[(\mathbf{1}^T\Sigma\mathbf{1}^T \operatorname{vec} I - \operatorname{tr}\Sigma\operatorname{vec}(\mathbf{1}\mathbf{1}^T)]L_k^-.$$

Now we use Theorem 1 of Neudecker and Wesselman (1990) to generalize the formula for σ^2 to non-normal variables; the substitution $2(\Sigma \otimes \Sigma) = \Gamma$ gives the general formula $\sigma^2 = h^T L_k^T \Gamma L_k h$.

Using the identities

$$\begin{aligned}
\mathbf{1}^T\Sigma\mathbf{1} &= \phi^2 k\{(k-1)\rho + 1, \\
\operatorname{tr}\Sigma &= \phi^2 k,
\end{aligned}$$

we obtain

$$
\begin{aligned}
b &= \frac{[\phi^2 k\{(k-1)\rho + 1\}]\operatorname{vec}(I_p) - \phi^2 k \operatorname{vec}(\mathbf{1}\mathbf{1}^T)}{[\phi^2 k\{(k-1)\rho + 1\}]^2} L_k^-, \\
&= \frac{[\{(k-1)\rho + 1\}\operatorname{vec}(I_p) - \operatorname{vec}(\mathbf{1}\mathbf{1}^T)}{\phi^2 k[(k-1)\rho + 1]^2} L_k^-.
\end{aligned}
$$

Define

$$
g = \{[(k-1)\rho + 1]\operatorname{vec}(I_p) - \operatorname{vec}(\mathbf{1}\mathbf{1}^T)\}L_k^-,
$$

so that

$$
\sigma^2 = \frac{1}{(k-1)^2}\frac{1}{[\{(k-1)\rho + 1\}]^4}\frac{1}{\phi^4} g^T L_k^T \Gamma L_k g.
$$

Moreover, $L_k L_k^- = M_p$, which satisfies $M_p \operatorname{vec} S = \operatorname{vec} S$ whenever S is symmetric, see van Zyl et al. (2000, p. 275). Thus

$$
\begin{aligned}
\sigma^2 &= \frac{1}{(k-1)^2}\frac{1}{[\{(k-1)\rho + 1\}]^4}\frac{1}{\phi^4} v^T M_p^T \Gamma M_p^T v, \\
&= \frac{1}{(k-1)^2}\frac{1}{[\{(k-1)\rho + 1\}]^4}\frac{1}{\phi^4} v^T \Gamma v.
\end{aligned}
$$

where v is defined in Proposition 17.

Now it is time for standardized alpha. Combining equations (8) and (9) of Hayashi and Kamata (2005), we find that the asymptotic variance for standardized alpha is

$$
\sigma_s^2 = \frac{1}{(k-1)^2}\left(\frac{1}{1+\rho(k-1)}\right)^4 \mathbf{1}_*^T D^T \Psi D \mathbf{1}_*,
$$

where

$$
\Psi = [I - (I \otimes \Phi)K_d](\Sigma_d^{-1/2} \otimes \Sigma_d^{-12})\Gamma(\Sigma_d^{-1/2} \otimes \Sigma_d^{-12})[I - K_d(I \otimes \Phi)]
$$

Since $\Sigma_d^{-1/2} = \phi^{-1}I$,

$$
\Sigma_d^{-1/2} \otimes \Sigma_d^{-12} = \phi^{-1}I \otimes \phi^{-1}I = \phi^{-2}I \otimes I = \phi^{-2}I.
$$

hence

$$
\Psi = \phi^{-4}[I - (I \otimes \Phi)K_d]\Gamma[I - K_d(I \otimes \Phi)].
$$

<div style="text-align: right">□</div>

Propositon 17 now follows from this lemma.

Lemma 5. *The following equalities are true*

$$
\{I - K_d(I \otimes \Phi)\}D\mathbf{1}_* = \{I - K_d(I \otimes \Phi)\}[\operatorname{vec}(\mathbf{1}\mathbf{1}^T) - \operatorname{vec} I] = \operatorname{vec}(\mathbf{1}\mathbf{1}^T) - [(k-1)\rho + 1]\operatorname{vec}(I_p).
$$

Proof. First observe that $D\mathbf{1}_* = \operatorname{vec}(\mathbf{1}\mathbf{1}^T) - \operatorname{vec} I$ by the definition of D, proving the first inequality.

Now we show that $(I - K_d(I \otimes \Phi))\operatorname{vec} I = 0$. First observe that $(I \otimes \Phi)\operatorname{vec} I = \operatorname{vec}\Phi$ by Theorem 2.2 of Magnus and Neudecker (2019). The operation $K_d \operatorname{vec} A$ selects the "diagonal" elements of $\operatorname{vec} A$, i.e., $K_d \operatorname{vec} A = \operatorname{vec} A_d$. Since $\Phi_d = I$, $K_p \operatorname{vec}\Phi = K_p \operatorname{vec} I = \operatorname{vec} I$, and therefore $(I - K_d(I \otimes \Phi))\operatorname{vec} I = 0$ as claimed.

The next step is $K_d(I \otimes \Phi)\operatorname{vec}(\mathbf{1}\mathbf{1}^T) = (1 + (k-1)\rho)\mathbf{1}\mathbf{1}^T$. First, by Theorem 2.2 of Magnus and Neudecker (2019), $K_d(I \otimes \Phi)\operatorname{vec}(\mathbf{1}\mathbf{1}^T) = K_d \operatorname{vec}(\Phi\mathbf{1}\mathbf{1}^T)$. Since $\Phi\mathbf{1}\mathbf{1}^T = [1 + (k-1)\rho]\mathbf{1}\mathbf{1}^T$, $K_d \operatorname{vec}(\Phi\mathbf{1}\mathbf{1}^T) = [1 + (k-1)\rho]K_d \operatorname{vec}(\mathbf{1}\mathbf{1}^T)$, but since $K_d \operatorname{vec}(\mathbf{1}\mathbf{1}^T) = \operatorname{vec} I$, we get $K_d(I \otimes \Phi)\operatorname{vec}(\mathbf{1}\mathbf{1}^T) = [1 + (k-1)\rho]\operatorname{vec} I$. □

References

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons. [24]

Bullen, P. S., Mitrinovic, D. S., & Vasic, M. (2013). *Means and their inequalities*. Springer Science & Business Media. [19, 22]

Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement*, *78*(6), 1056–1071. https://doi.org/10.1177/0013164417727036 [13, 15]

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98 [9]

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. [2]

Falk, C. F., & Savalei, V. (2011). The relationship between unstandardized and standardized alpha, true reliability, and the underlying measurement model. *Journal of Personality Assessment*, *93*(5), 445–453. https://doi.org/10.1080/00223891.2011.594129 [2, 5, 9]

Foldnes, N., & Grønneberg, S. (2019). Pernicious polychorics: The impact and detection of underlying non-normality. *Structural Equation Modeling*, 1–19. https://doi.org/10.1080/10705511.2019.1673168 [13]

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction* (2nd ed.). Sage Publications. [1, 10]

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, *17*(1), 1–13. https://doi.org/10.7275/n560-j767 [1]

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in europe* (Vol. 7, pp. 7–28). Tilburg University Press. [11]

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121–135. https://doi.org/10.1007/s11336-008-9098-4 [2]

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. C. S. Du Toit Dag Sörbom (Ed.), *Structural equation modeling: Present and future* (pp. 195–216). Scientific Software International. [4]

Hayashi, K., & Kamata, A. (2005). A note on the estimator of the alpha coefficient for standardized variables under normality. *Psychometrika*, *70*(3), 579–586. https://doi.org/10.1007/s11336-001-0888-1 [8, 22, 26]

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions*. Wiley. [14]

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. https://doi.org/10.1007/BF02291393 [4, 9]

Korn, E. L., & Simon, R. (1991). Explained residual variation, explained risk, and goodness of fit. *The American Statistician*, *45*(3), 201–206. https://doi.org/10.1080/00031305.1991.10475802 [4]

Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, *28*(3), 221–238. https://doi.org/10.1007/BF02289571 [8, 21]

Lehmann, E. L. (2004). *Elements of large-sample theory* (2nd ed.). Springer Science & Business Media. [7]

Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika*, *62*(2), 245–249. https://doi.org/10.1007/BF02295278 [4]

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley. [3, 5]

Magnus, J. R., & Neudecker, H. (2019). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons. https://doi.org/10.1002/9781119541219 [25, 26]

McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). Springer. https://doi.org/10.1007/978-1-4757-2691-6_15 [12]

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144 [9]

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01398-0 [6, 10, 17]

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons. [21]

Neudecker, H., & Satorra, A. (1996). The algebraic equality of two asymptotic tests for the hypothesis that a normal distribution has a specified correlation matrix. *Statistics & Probability Letters*, *30*(2), 99–103. https://doi.org/10.1016/0167-7152(95)00206-5 [25]

Neudecker, H., & Wesselman, A. M. (1990). The asymptotic variance matrix of the sample correlation matrix. *Linear Algebra and its Applications*, *127*, 589–599. https://doi.org/10.1016/0024-3795(90)90363-h [25]

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207 [12]

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*(3), 343–355. https://doi.org/10.1037/1082-989x.5.3.343 [5, 10]

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. [1, 6, 11]

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. https://doi.org/10.1207/s15327906mbr3204_2 [5]

Raykov, T. (2019). Strong convergence of the coefficient alpha estimator for reliability of multiple-component measuring instruments. *Structural Equation Modeling*, *26*(3), 430–436. https://doi.org/10.1080/10705511.2018.1515019 [8]

Revelle, W. (2019a). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=psych (R package version 1.9.12) [1, 15]

Revelle, W. (2019b). psychtools: Tools to accompany the 'psych' package for psychological research [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=psychTools (R package version 1.9.12) [11, 12]

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2017). Web and phone based data collection using planned missing designs. In R. M. L. . G. B. Nigel G. Fielding (Ed.), *The SAGE handbook of online research methods* (2nd ed., pp. 578–595). London: SAGE Publications. https://doi.org/10.4135/9781473957992.n33 [11]

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02 [6, 11, 16]

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159 [9]

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*. https://doi.org/10.1007/BF02294363 [12]

Thurstone, L. (1935). *The vectors of mind*. University of Chicago Press. [4]

Tong, Y. L. (1990). *The multivariate normal distribution*. Springer-Verlag. [24]

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press. [22]

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, *65*(3), 271–280. https://doi.org/10.1007/BF02296146 [8, 22, 25, 26]

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29. https://doi.org/10.22237/jmasm/1177992180 [2, 12]

Zumbo, B. D., & Kroc, E. (2019). A measurement is a choice and Stevens' scales of measurement do not help make it: A response to Chalmers. *Educational and Psychological Measurement*, *79*(6), 1184–1197. https://doi.org/10.1177/0013164419844305 [13]