Alpha and Omega: Simulating the Impact of Tau-Equivalence Violations on Reliability Estimates.

June 9, 2025

Version: 1.1
Last updated: 2024-11-18
Template based on the paper:
Siepe, B. S., Bartoš, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., & Pawel, S. (2024). Simulation Studies for Methodological Research in Psychology: A Standardized Template for Planning, Preregistration, and Reporting. *Psychological Methods*.
https://doi.org/10.1037/met0000695, https://doi.org/10.31234/osf.io/ufgy6

1 General Information

1.1 What is the title of the project?

Alpha and Omega: Simulating the Impact of Tau-Equivalence Violations on Reliability Estimates.

1.2 Who are the current and future project contributors?

Alina Apel, Dr. Julien Patrick Irmer

1.3 Provide a description of the project.

Cronbach's alpha (Cronbach, 1951) and McDonald's omega (McDonald, 1999) are commonly used reliability estimates. I will investigate their performance under varying degrees of violation of the tau-equivalence assumption. Using a simulation study, I systematically vary the item loading structure (including a tau-equivalent condition and three levels of increasing loading variability), the number of items (6, 12, 18, 24), and the sample size (50, 200, 500).

For each simulated dataset, I compute observed Cronbach's Alpha along with bootstrapped confidence intervals and compare it to the true Omega (calculated based on the population parameters used in the simulation). Additionally, I estimate Omega from the observed data using confirmatory factor analysis (CFA) implemented in the lavaan package.

I evaluate the performance of Alpha and Omega in terms of coverage (i.e., whether the estimator's confidence interval contains the true Omega), bias, and confidence interval width. The goal is to quantify how violations of tau-equivalence and other characteristics (such as small sample size) influence the validity and precision of Alpha compared to Omega. This investigates the conditions under which Alpha may be sufficient and when Omega should be preferred as a reliability estimate.

1.4 Did any of the contributors already conduct related simulation studies on this specific question?

I did not conduct previous simulation studies.

2 Aims

2.1 What is the aim of the simulation study?

The aim of the simulation study is to evaluate the performance of Cronbach's Alpha as an estimator of scale reliability by comparing it to McDonald's Omega, examining bias, confidence interval coverage, and precision across different loading structures, item numbers, and sample sizes.

3 Data-Generating Mechanism

3.1 How will the parameters for the data-generating mechanism (DGM) be specified?

The parameters for the data-generating mechanism (DGM) are specified to simulate responses to items measuring a single latent trait. In each simulation repetition, I generate responses for n individuals on n_{items} items using a one-factor model with continuous latent trait scores drawn from a standard normal distribution.

Specifically, the latent variable $\xi \sim \mathcal{N}(0,1)$ is generated for each individual. Item responses are generated as

$$Y_{ij} = \lambda_j \xi_i + \epsilon_{ij}$$

where λ_j are the factor loadings for item j, and $\epsilon_{ij} \sim \mathcal{N}(0, \theta_j)$ are independent error terms with variances θ_j. The factor loadings λ_j are specified according to four experimental conditions reflecting different degrees of tau-equivalence violation:

- **Tau-equivalent:** All loadings equal ($\lambda_j = 0.7\ \forall j$).
- **Small variance:** Loadings sampled uniformly from 0.6 to 0.8.
- **Medium variance:** Loadings sampled uniformly from 0.4 to 0.95.
- **High variance:** Loadings sampled uniformly from 0.2 to 1.0.

Error variances θ_j are sampled independently from a uniform distribution on $[0.2, 1.2]$ to introduce realistic variability in measurement error across items.

Simulated continuous responses Y_{ij} are then transformed to discrete ordinal responses on a Likert-type scale from 1 to 6 by standardizing each item response, scaling to the 1–6 range, and rounding to the nearest integer.

3.2 What will be the different factors of the data-generating mechanism?

The factors varied in the data-generating mechanism are:

- **Loading structure:** Four levels reflecting the degree of tau-equivalence violation: Tau-equivalent (all loadings equal), small variance, medium variance, and high variance of factor loadings.
- **Number of items:** Four levels (6, 12, 18, 24 items).
- **Sample size:** Three levels (50, 200, 500 participants).

3.3 If possible, provide specific factor values for the DGM as well as additional simulation settings.

I will use the following values for the data-generating mechanism:

- **Loading structure (degree of tau-equivalence violation):**

$$\text{loading\_type} \in \{\text{tau\_equivalent}, \text{small\_var}, \text{medium\_var}, \text{high\_var}\}$$

  where:

  - *tau_equivalent*: all loadings set to 0.7
  - *small_var*: loadings drawn uniformly from [0.6, 0.8]
  - *medium_var*: loadings drawn uniformly from [0.4, 0.95]
  - *high_var*: loadings drawn uniformly from [0.2, 1]

- **Number of items:**

$$n_{\text{items}} \in \{6, 12, 18, 24\}$$

- **Sample size:**

$$n \in \{50, 200, 500\}$$

- **Error variances:** Error variances are drawn uniformly from [0.2, 1.2] for each item in every condition.

3.4 If there is more than one factor: How will the factor levels be combined and how many simulation conditions will this create?

The data will be simulated in a fully factorial design. This will result in 4 (loading structure) $\times$ 4 (number of items) $\times$ 3 (sample sizes) $=$ 48 simulation conditions.

If computational time becomes excessive, I plan to omit the 18-item level of the n_{items} factor as well as the *medium_var* level on the factor structure condition. This reduces the number of simulation conditions to 3 (loading structure) $\times$ 3 (number of items) $\times$ 3 (sample sizes) $=$ 27 simulation conditions.

4 Estimands and Targets

4.1 What will be the estimands and/or targets of the simulation study?

My primary estimands are the true population values of Cronbach's alpha and McDonald's omega as estimators of reliability under different degrees of tau-equivalence violation, sample sizes, and test lengths. The primary target is the bias and the width of the bootstrapped confidence intervals.

5 Methods

5.1 How many and which methods will be included and which quantities will be extracted?

I will compare the following methods for estimating reliability:

- **Cronbach's Alpha:** Calculated using the `psych::alpha()` function with default settings. This classical measure assumes tau-equivalence and is used to estimate internal consistency of scales.
- **McDonald's Omega:** Estimated via confirmatory factor analysis (CFA) implemented in the `lavaan` package, specifying a one-factor model with standardized loadings and residual variances estimated freely.

From each simulated dataset, I will extract the following quantities:

- The observed Cronbach's alpha estimate.
- The estimated omega coefficient from the CFA.
- The lower and upper bounds of the 95% confidence interval for alpha and omega obtained via bootstrapping using the `boot` package.
- The true omega value calculated directly from the known population loadings and error variances via

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

  where λ_i are the factor loadings for each item i, θ_i are the corresponding error variances, and k is the number of items.
- The bias of alpha relative to true omega.
- The coverage of the alpha confidence interval with respect to the true omega.

Cronbach's alpha assumes tau-equivalence, while McDonald's omega allows for varying loadings.

6 Performance Measures

6.1 Which performance measures will be used?

7.1 Which performance measures will be used?

I will use the following performance measures to evaluate how well Cronbach's alpha and McDonald's omega estimate the true reliability under various degrees of tau-equivalence violation.

- **Bias**
  The average difference between the estimated reliability and the true population omega:
$$\text{Bias} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)$$
  where $\hat{\theta}_i$ is the estimated reliability coefficient in simulation i, and θ is the true omega.

- **Monte Carlo Standard Deviation (MCSD)**
  The standard deviation of the estimated reliability across simulations:
$$\text{MCSD} = \sqrt{\frac{1}{n_{\text{sim}} - 1} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \bar{\hat{\theta}})^2}$$
  where $\bar{\hat{\theta}}$ is the mean of the estimates over simulations.

- **Confidence Interval Coverage**
  The proportion of bootstrap-based confidence intervals for Cronbach's alpha and the observed omega that include the true omega:
$$\text{Coverage} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \mathbb{I}[\theta \in \text{CI}_i]$$
  where $\mathbb{I}$ is the indicator function and CI_i is the confidence interval in simulation i.

- **Confidence Interval Width (CI Width)**
  The average width of the confidence intervals:
$$\text{CI Width} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} (\text{CI}_{\text{upper},i} - \text{CI}_{\text{lower},i})$$

6.2 How will Monte Carlo uncertainty of the estimated performance measures be calculated and reported?

Monte Carlo uncertainty of the estimated performance measures will be quantified and reported using Monte Carlo Standard Deviation (MCD).

6.3 How many simulation repetitions will be used for each condition?

At this stage, I do not yet know how many simulation repetitions will be feasible given the computational demands of the study. However, I aim to run 2,000 repetitions per condition. If computational limitations change during execution, I will adjust the number of repetitions accordingly and report the final number used in each condition.

6.4 How will missing values due to non-convergence or other reasons be handled?

I do not expect missing values or non-convergence. If I observe any non-convergence, I exclude the non-converged cases case-wise (keeping the converged values from the other methods in the same repetition) and report the number of non-converged cases per method and condition.

6.5 How do you plan on interpreting the performance measures? (optional)

Bias will be considered problematic if it is consistently large across conditions or if it changes substantially with increasing violations of tau-equivalence. Coverage rates of confidence intervals will be compared against the nominal level of 95%. CI width will be evaluated as an indicator of precision, with wider intervals interpreted as less desirable.

7 Other

7.1 Which statistical software/packages do you plan to use?

I will conduct the simulation study using **R version 4.5.0** (R Core Team, 2025). The following R packages (in their most recent CRAN versions at the time of implementation) will be used:

- `psych` (Revelle, 2025): for computing Cronbach's alpha.
- `boot` (Canty & Ripley, 2021): for nonparametric bootstrapping to obtain confidence intervals.
- `lavaan` (Rosseel, 2012): for computing McDonald's omega.
- `tidyverse` (Wickham et al., 2019): for data manipulation and visualization (especially `dplyr` and `ggplot2`).

7.2 Which computational environment do you plan to use?

I will run the simulation study on a Windows 11 machine. The complete output of `sessionInfo()` will be reported.

7.3 Which other steps will you undertake to make simulation results reproducible? (optional)

I will set a fixed random seed using `set.seed(2025)` in the R script. Additionally, I plan to publicly share my R code on `https://github.com//apelalina`.

7.4 Is there anything else you want to preregister? (optional)

No.

References

Canty, A., & Ripley, B. (2021). *Boot: Bootstrap r (s-plus) functions* [R package version 1.3-28]. https://CRAN.R-project.org/package=boot

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.

R Core Team. (2025). *R: A language and environment for statistical computing* [Version 4.5.0]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Revelle, W. (2025). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.5.3]. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

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

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686