Online Supplement for: "Confidence Intervals for Population Reliability Coefficients: Evaluation of Methods, Recommendations, and Software for Homogeneous Composite Measures" (Kelley & Pornprasertmanit, 2015)

Ken Kelley

University of Notre Dame

Notre Dame, Indiana 46556

Sunthud Pornprasertmanit

Texas Tech University

Lubbock, TX 79409

Online Supplement for: "Confidence Intervals for Population Reliability Coefficients: Evaluation of Methods, Recommendations, and Software for Homogeneous Composite Measures" (Kelley & Pornprasertmanit, 2015)

This online supplement accompanies Kelley & Pornprasertmanit (2015), which under review at *Psychological Methods*. Should Kelley & Pornprasertmanit (2015) be accepted, this document will be placed on the University of Notre Dame digital archives so that it will be accessible online in perpetuity.

Online Appendix B: Additional Results

Because of the space limitation in the published document, we provide additional results to support the conclusions from Kelley & Pornprasertmanit (2015).

Study 1

Point Estimation. In this section, we will investigate whether coefficients alpha and omega accurately estimated the population reliability. We investigated the coefficient omega from both maximum likelihood (ML) and asymptotically distribution free methods. The bias in parameter estimates could be the primary reason why confidence interval methods cannot bracket the population reliability. The absolute bias in the parameter estimate is investigated:

$$\text{Absolute Bias}(\hat{\rho}) = \bar{\rho} - \rho, \tag{1}$$

where $\bar{\rho}$ is the average reliability estimate across replications. Table 1 shows η^2 for all main and interaction effects of the design conditions on the absolute bias. For coefficient alpha, the interaction effect between the number of items and factor loading distribution had η^2 greater than .05. Table 2 shows the pattern of this interaction. Coefficient alpha underestimates population reliability in a higher magnitude when factor loadings are not equal, especially when the number of items is low. Also, the main effect of sample size was impactful. The biases of coefficient alpha decreased when sample size increased: -.024, -.017, -.014, -.011, -.009 for sample sizes of 50, 100, 200, 400, and 1000, respectively.

The biases of coefficient omega estimated from ML were smaller than coefficient alpha, but not necessarily much smaller, on average. The biases of coefficient omega from ML decreased when sample size increased: -.008, -.004, -.003, -.001, and .0002 for sample sizes of 50, 100, 200, 400, and 1000, respectively. Notice that the magnitudes of biases of coefficient omega from ML were less than they were from coefficient alpha. The mean of the absolute difference between the bias of coefficient alpha and coefficient omega was .012. The biases of coefficient omega from ML were also influenced by the number of items and the population reliability. The bias of coefficient omega from ML was -.001, -.003, -.003, -.004, and -.004 for the number of items of 4, 8, 12, 16, and 20, respectively. The bias of coefficient omega from ML was -.004, -.003, and -.002 for the population reliability of .7, .8, and .9, respectively.

The biases of coefficient omega estimated from ADF were higher than when estimated from ML. The interaction effect between sample size and the number of items was impactful, in that η^2 was above our benchmark of .05 (it was .054). As shown in Table 3, the biases were high when sample size was low and the number of items were high–these conditions provided low to zero convergence rates. In sum, the biases of coefficient omega from ADF was worse than one from ML; however, they were still better than the biases from coefficient alpha.

Coverage Rates of Confidence Interval of Coefficient Alpha. Table 4 shows η^2 for all main and interaction effects of each design condition on the coverage rates of the confidence intervals for coefficient alpha on population reliability. Note that the results from this study are different from previous studies (e.g., Cui & Li, 2012, Duhachek & Iacobucci, 2004, Padilla, Divers, & Newton, 2012, Romano, Kromrey, & Hibbard, 2010, and Maydeu-Olivares, Coffman, & Hartmann, 2007) that find the coverage rates of coefficient alpha on the population coefficient alpha.

All of the coverage rates were mostly influenced by the interaction effects between the number of items and factor loading distribution or the main effects of each. We would like to only emphasize this interaction effect that is shown in Table 5. The coverage rates were better when the number of items increased and factor loadings were equal. However, the coverage rates were not acceptable in the conditions with low number of items and unequal loadings. The low coverage rates resulted from the biases in population reliability estimation of coefficient alpha in these conditions.

Study 2

Point Estimation. Table 6 shows η^2 for all main and interaction effects of the design conditions on the absolute bias of coefficient omega and hierarchical omega. For coefficient omega, the main effects of sample size and the number of items had η^2 greater than .05. The bias of coefficient omega decreased when sample size increased (.013, .006, .003, .002, and .002 for sample sizes of 50, 100, 200, 400, and 1000, respectively). Also, the bias of coefficient omega decreased when the number of items was higher (.007, .005, and .003 for 4, 8, and 12 items, respectively). It is unclear of how much bias is "large" but any bias in an estimation procedure, holding everything else constant, is not ideal. From a practical perspective in this context, in the sample size of 50 condition the bias is .013, which is a 1.5% bias for a situation in which the population reliability coefficient is .85, as an example.

For hierarchical omega, the interaction effects between (a) sample size and the number of items and (b) the number of items and population coefficient omega were not negligible. As shown in Table 7, the biases of hierarchical omega were highest when sample size was low and the number of items was high. As shown in Table 8, the biases of hierarchical omega were highest when population coefficient omega was low and the number of items was high. However, the magnitude of biases of hierarchical omega was still lower than compared to coefficient omega.

Coverage Rates of Confidence Interval of Coefficient Omega. Table 9 shows η^2 for all main and interaction effects of each design condition on the coverage rates of the confidence intervals for coefficient omega on population reliability. All of the coverage rates were mostly influenced by the interaction effects between sample size and the population coefficient omega. Table 10 revealed that all of the confidence interval methods had poor coverage rates when sample size was small and the population coefficient omega was .9. For lower values of population coefficient omega, the logistic-transformed bootstrap standard error method and percentile method performed better than the bootstrap standard error method and the bias-corrected and accelerated method. In addition, the coverage rate of logistic-transformed bootstrap standard error was influenced by the number of items. The coverage rates were higher with more items (.926, .930, and .941 for 4, 8, and 12 items respectively).

Study 3

The additional result for Study 3 is only the bias in parameter estimate. The coverage rate of hierarchical omega is included in the main article. Table 11 shows η^2 for all main and interaction effects of the design conditions on the absolute bias of hierarchical omega and categorical omega. Both hierarchical and categorical omega were influenced by the interaction between the number of categories and threshold patterns, which is shown in Table 12. In general, categorical omega has less biases than hierarchical omega. The largest differences were in dichotomous items with threshold patterns 4 and 5, which is the conditions that the threshold patterns are not the same across items.

References

Cui, Y., & Li, J. (2012). Evaluating the performance of different procedures for constructing confidence intervals of coefficient alpha: A simulation study. *British Journal of Mathematical and Statistical Psychology*, *65*, 467–498.

Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and percise confidence interval estimate. *Journal of Applied Psychology*, *89*, 792–808.

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*, 157–176.

Padilla, M. A., Divers, J., & Newton, M. (2012). Coefficient alpha bootstrap confidence interval under nonnormality. *Applied Psychological Measurement*, *36*, 331–348.

Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, 376–393.

Author Note

Part of this research was presented at the 2011 Joint Statistics Meeting in Miami, Florida. Correspondence concerning this article should be addressed to Ken Kelley, Department of Management–Analytics, Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana 46556. Electronic mail may be sent to KKelley@ND.Edu. This research benefited from the Center for Research Computing at the University of Notre Dame and from a grant to the first author from the Notre Dame Deloitte Center for Ethical Leadership.

Table 1

The η^2 of the Effects of the Design Conditions on the Biases of Coefficients Alpha and Omega (estimated from maximum likelihood and asymptotically distribution free method) in Population Reliability Estimation for Study 1

Factors	Alpha	ML Omega	ADF Omega
N	**.055**	**.265**	**.134**
J	**.214**	**.053**	.038
LOAD	**.307**	.000	.001
RELIA	.003	**.067**	.015
DIST	.006	.016	.003
N:J	.000	.028	**.054**
N:LOAD	.000	.000	.001
N:RELIA	.006	.040	.010
N:DIST	.003	.012	.046
J:LOAD	**.192**	.000	.000
J:RELIA	.001	.045	.007
J:DIST	.001	.004	.014
LOAD:RELIA	.003	.002	.000
LOAD:DIST	.000	.001	.001
RELIA:DIST	.000	.003	.000

Note: The boldface numbers represent the $\eta^2 \geq .05$. All interactions higher than two ways are not presented here because their $\eta^2 < .05$. N = Sample size. J = The number of items. LOAD = Factor loading distribution. RELIA = Population reliability. DIST = Item distributions. ML = Maximum likelihood. ADF = Asymptotically distribution free.

Table 2

The Bias of Coefficient Alpha in Estimating Population Reliability Classified by the Number of Items and Factor Loading Distribution in Study 1

Number of Items	Factor Loading Distribution	
	Equal	Unequal
4	-.007	-.060
8	-.006	-.024
12	-.006	-.016
16	-.006	-.013
20	-.006	-.011

Table 3

The Bias of Coefficient Omega Estimated by Asymptotic Distribution Free Method in Estimating Population Reliability Classified by the Number of Items and Sample Size for Study 1

Number of Items	Sample Size				
	50	100	200	400	1000
4	.005	.000	-.001	.000	.000
8	.031	.003	-.006	-.005	-.003
12	NA	.032	.002	-.007	-.007
16	NA	NA	.026	-.001	-.008
20	NA	NA	NA	.016	-.007

Note: NA represents conditions in which there was not sufficient convergence to use the results from the Monte Carlo simulation.

Table 4

The η^2 of the Effects of the Design Conditions on the Coverage Rates of Coefficient Alpha for Study 1

Factors	Feldt	NT	NT-L	ADF	ADF-L	Fisher	Bonett	HW	LL	BSE	BSE-L	PER	BC*a*
N	**.065**	**.079**	**.059**	.026	.019	**.051**	**.058**	**.068**	.045	.047	.036	.014	.016
J	**.168**	**.148**	**.161**	**.158**	**.161**	**.210**	**.170**	**.164**	**.180**	**.147**	**.153**	**.157**	**.166**
LOAD	**.140**	**.096**	**.128**	**.166**	**.198**	**.130**	**.157**	**.136**	**.180**	**.159**	**.190**	**.237**	**.199**
RELIA	**.080**	**.074**	**.097**	**.076**	**.092**	.037	**.084**	**.080**	**.092**	**.069**	**.087**	**.087**	**.079**
DIST	.026	.032	.032	.001	.001	.015	.025	.026	.026	.000	.000	.001	.001
N:J	.024	.034	.025	.020	.015	.031	.022	.026	.015	.023	.017	.014	.016
N:LOAD	**.052**	**.063**	.049	**.052**	.043	.045	.048	**.054**	.039	**.055**	.045	.036	.038
N:RELIA	.008	.010	.005	.006	.003	.007	.007	.009	.005	.007	.004	.003	.004
N:DIST	.000	.000	.000	.004	.004	.000	.000	.000	.000	.003	.003	.003	.004
J:LOAD	**.119**	**.103**	**.113**	**.143**	**.145**	**.130**	**.121**	**.116**	**.130**	**.137**	**.142**	**.154**	**.154**
J:RELIA	.024	.029	.030	.032	.029	.025	.023	.024	.022	.031	.029	.024	.028
J:DIST	.000	.000	.000	.001	.001	.001	.000	.000	.000	.001	.001	.001	.001
LOAD:RELIA	.048	.043	**.053**	**.065**	**.072**	.038	**.052**	.048	**.058**	**.062**	**.069**	**.073**	**.068**
LOAD:DIST	.001	.001	.001	.002	.002	.001	.002	.001	.001	.002	.002	.002	.003
RELIA:DIST	.001	.001	.001	.000	.000	.000	.001	.001	.001	.000	.000	.000	.000

Note: The boldface numbers represent the $\eta^2 \geq .05$. All interactions higher than two ways are not presented here because their $\eta^2 < .05$. NT = Normal-theory method. L = Logistic-transformed Wald confidence interval. ADF = Asymptotic distribution free. HW = Hakstian-Whalen method. LL = Likelihood-based method. BSE = Bootstrap standard error. PER = Percentile bootstrap. BCa = Bias-corrected-and-accelerated bootstrap. N = Sample size. J = The number of items. LOAD = Factor loading distribution. RELIA = Population reliability. DIST = Item distributions.

Table 5

The Coverage Rates of Different Confidence Interval of Coefficient Alpha in Estimating Population Reliability Classified by the Number of Items and Factor Loading Distribution in Study 1

LOAD	J	Feldt	NT	NT-L	ADF	ADF-L	Fisher	Bonett	HW	LL	BSE	BSE-L	PER	BCa
	4	.870	.866	.913	.910	.929	.926	.914	.910	.863	.898	.865	.868	.868
	8	.889	.885	.916	.913	.932	.929	.911	.912	.883	.936	.886	.888	.888
Equal	12	.899	.896	.921	.918	.935	.931	.912	.916	.893	.948	.896	.898	.898
	16	.905	.901	.923	.921	.936	.933	.914	.919	.900	.954	.902	.903	.904
	20	.908	.906	.923	.921	.936	.934	.914	.918	.904	.958	.906	.908	.908
	4	.461	.411	.422	.386	.448	.409	.357	.388	.336	.420	.387	.402	.412
	8	.764	.730	.742	.713	.764	.732	.682	.716	.673	.788	.707	.724	.727
Unequal	12	.848	.825	.841	.819	.858	.835	.796	.823	.791	.887	.810	.822	.824
	16	.883	.864	.880	.861	.894	.877	.840	.864	.840	.925	.853	.863	.864
	20	.898	.883	.899	.884	.913	.899	.866	.885	.865	.940	.874	.883	.883

Note: NT = Normal-theory method. L = Logistic-transformed Wald confidence interval. ADF = Asymptotic distribution free. HW = Hakstian-Whalen method. LL = Likelihood-based method. BSE = Bootstrap standard error. PER = Percentile bootstrap. BCa = Bias-corrected-and-accelerated bootstrap. J = The number of items. LOAD = Factor loading distribution.

Table 6

The η^2 of the Effects of the Design Conditions on the Biases of Coefficient Omega and Hierarchical Omega in Population Reliability Estimation for Study 2

Factors	Omega	Hierarchical Omega
N	**.289**	**.218**
J	**.090**	**.114**
RELIA	.023	**.062**
RMSEA	.020	.005
N:J	.014	**.052**
N:RELIA	.021	.035
N:RMSEA	.000	.002
J:RELIA	.039	**.082**
J:RMSEA	.008	.004
RELIA:RMSEA	.002	.003

Note: The boldface numbers represent the $\eta^2 \geq .05$. All interactions higher than two ways are not presented here because their $\eta^2 < .05$. N = Sample size. J = The number of items. RELIA = Population coefficient omega. RMSEA = Root mean square error of approximation.

Table 7

The Bias of Hierarchical Omega in Estimating Population Reliability Classified by the Number of Items and Sample Size in Study 2

Sample Size	Number of Items		
	4	8	12
50	-.001	-.008	-.011
100	-.002	-.004	-.005
200	-.001	-.002	-.003
400	.000	-.001	-.001
1000	.000	.000	.000

Table 8

The Bias of Hierarchical Omega in Estimating Population Reliability Classified by the Number of Items and Population Coefficient Omega in Study 2

Population Coefficient Omega	Number of Items		
	4	8	12
.7	.000	-.004	-.007
.8	-.001	-.003	-.004
.9	-.001	-.001	-.002

Table 9

The η^2 of the Effects of the Design Conditions on the Coverage Rates of Coefficient Omega on Population Reliability for Study 2

Factors	BSE	BSE-L	PER	BCa
N	**.230**	**.131**	**.186**	**.258**
J	.049	**.082**	.041	.030
RELIA	**.241**	**.281**	**.266**	**.202**
RMSEA	.009	.009	.008	.011
N:J	.002	.004	.000	.000
N:RELIA	**.079**	**.098**	**.094**	**.061**
N:RMSEA	.001	.005	.004	.003
J:RELIA	.006	.011	.000	.004
J:RMSEA	.002	.003	.001	.001
RELIA:RMSEA	.001	.003	.001	.003

Note: The boldface numbers represent the $\eta^2 \geq .05$. All interactions higher than two ways are not presented here because their $\eta^2 < .05$. L = Logistic-transformed Wald confidence interval. BSE = Bootstrap standard error. PER = Percentile bootstrap. BCa = Bias-corrected-and-accelerated bootstrap. N = Sample size. J = The number of items. RELIA = Population coefficient omega. RMSEA = Root mean square error of approximation.

Table 10

The Coverage Rates of Different Confidence Interval of Coefficient Omega in Estimating Population Reliability Classified by Sample Size and Population Coefficient Omega in Study 2

RELIA	N	BSE	BSE-L	PER	BCa
	50	.916	.942	.940	.925
	100	.929	.944	.940	.935
.7	200	.931	.938	.935	.934
	400	.945	.948	.948	.948
	1000	.941	.943	.943	.943
	50	.898	.930	.925	.919
	100	.916	.934	.931	.930
.8	200	.925	.935	.932	.932
	400	.943	.947	.948	.946
	1000	.941	.945	.945	.945
	50	.812	.879	.878	.875
	100	.871	.902	.904	.903
.9	200	.898	.916	.917	.916
	400	.929	.942	.943	.940
	1000	.932	.937	.938	.938

Note: L = Logistic-transformed Wald confidence interval. BSE = Bootstrap standard error. PER = Percentile bootstrap. BCa = Bias-corrected-and-accelerated bootstrap. N = Sample size. RELIA = Population coefficient omega.

Table 11

The η^2 of the Effects of the Design Conditions on the Biases of Hierarchical Omega and Categorical Omega in Population Reliability Estimation for Study 3

Factors	Hierarchical Omega	Categorical Omega
N	.015	.001
J	.017	.044
RELIA	.024	.000
NCAT	**.085**	.002
THRES	**.225**	**.075**
N:J	.001	.007
N:RELIA	.000	.000
N:NCAT	.000	.003
N:THRES	.000	.039
J:RELIA	.001	.001
J:NCAT	.006	.004
J:THRES	.029	.001
RELIA:NCAT	.018	.000
RELIA:THRES	.045	.000
NCAT:THRES	**.131**	**.093**

Note: The boldface numbers represent the $\eta^2 \geq .05$. All interactions higher than two ways are not presented here because their $\eta^2 < .05$. N = Sample size. J = The number of items. RELIA = Population categorical omega for perfect fitting model. NCAT = The number of categories. THRES = Threshold pattern.

Table 12

The Biases of Hierarchical Omega and Categorical Omega in Population Reliability Estimation Classified by the Number of Categories and Threshold Patterns in Study 3

NCAT	THRES	Hierarchical Omega	Categorical Omega
	1	-.010	.014
	2	-.010	.015
2	3	-.016	.022
	4	-.026	.012
	5	-.170	-.040
	1	-.008	.005
	2	-.008	.006
5	3	-.010	.007
	4	-.012	.006
	5	-.029	.008

Note: NCAT = The number of categories. THRES = Threshold pattern.