RGLS and RLS in Covariance Structure Analysis

RGLS and RLS in Covariance Structure Analysis

Bang Quan Zheng[†]
University of Arizona

Peter M. Bentler
UCLA

2022

Forthcoming in
Structural Equation Modeling: A Multidisciplinary Journal

ABSTRACT

This paper assesses the performance of regularized generalized least squares (RGLS) and reweighted least squares (RLS) methodologies in a confirmatory factor analysis model. Normal theory maximum likelihood (ML) and GLS statistics are based on large sample statistical theory. However, ML and GLS goodness-of-fit tests often make incorrect decisions on the true model, when sample size is small. The novel methods RGLS and RLS aim to correct the over-rejection by ML and under-rejection by GLS. Both methods outperform ML and GLS when samples are small, yet no studies have compared their relative performance. A Monte Carlo simulation study was carried out to examine the statistical performance of these two methods. We find that RLS and RGLS have equivalent performance when $N \geq 70$; whereas when $N < 70$, RLS outperforms RGLS. Both methods clearly outperform ML and GLS with $N \leq 400$. Nonetheless, adopting mean and variance adjusted test for non-normal data, RGLS slightly outperforms RLS. Power analyses found that RLS generally showed small loss in power compared to ML and performed better than RGLS.

Keywords: covariance structure, eigenvalue, estimation method, goodness-of-fit, weight matrix

[†] **Contact** Bang Quan Zheng bangquan@arizona.edu. School of Government and Public Policy, Social Sciences 315, University of Arizona, Tucson, AZ 85721, U.S.A.

Based on the assumption of multivariate normality, maximum likelihood (ML) and generalized least squares (GLS) methods provide the oldest and most widely used estimators and tests in structural equation modeling (SEM) (Bollen 1989, Browne 1974, Hu et al. 1992, Jöreskog 1969, Lee 2007). The behavior of their statistics, especially the chi-square goodness-of-fit tests, is based on asymptotic properties that require sample size to be very large. Simulation research has found that small sample size N is the main contributor to failure of asymptotic theory, but large number of variables p and/or parameters q, small number of indicator loadings per factor, and small ratio of N to degrees of freedom df also contribute to spurious goodness-of-fit model rejections (Arruda and Bentler 2017, Boomsma 1982, Moshagen 2012, Shi et al. 2018, Shi et al. 2019, Yuan and Bentler 1999).

Two methods have recently been developed to correct the false model decision issue. Arruda and Bentler (2017) proposed a regularized GLS (RGLS) that "regularizes" an ill-conditioned sample covariance matrix. Simulations show that RGLS outperforms both ML and GLS across varied sample sizes. Hayakawa (2019) rediscovered a methodology based on both ML and GLS, Reweighted Least Squares (RLS), and showed that it similarly outperforms ML and GLS. However, the comparative performance of RGLS and RLS at various sample sizes is unknown. That is the focus of this paper. ML and GLS are included to provide a historic baseline.

This article proceeds as follows. First, we summarize ML and GLS estimation and tests. Next, we give technical definitions of RGLS and RLS. Then, we introduce the data generation and Monte Carlo simulation methodology to be used in normal and non-normal data. The fourth section discusses simulation results. The fifth section is power analysis of different methods and their robust variants, and a discussion follows.

Classical SEM Test Statistics: ML and GLS

Let $x \in \{x_1, \ldots, x_N\}$ be a random sample with all x_i identically and independently distributed according to a multivariate normal distribution N $\mathbf{[0, \Sigma]}$. A confirmatory factor model $\boldsymbol{x} = \boldsymbol{\Lambda\xi} + \boldsymbol{\varepsilon}$ is used to generate measured variable $\boldsymbol{x}$ under various conditions on a $m \times 1$ vector of common factors $\boldsymbol{\xi}$ and a $p \times 1$ vector of unique measurement errors. In SEM, we further assume that $\boldsymbol{\Sigma}$ is a matrix function of a vector of unknown population parameters $\boldsymbol{\theta}$ $(q \times 1)$, with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The covariance structure of interest here is the confirmatory factor model $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}$, where $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings, $\boldsymbol{\Phi} = cov(\boldsymbol{\xi})$ and $\boldsymbol{\Psi} = cov\ (\boldsymbol{\varepsilon})$. Assuming an identified model, the elements of $\boldsymbol{\theta}$ are the unknown free parameters in these matrices.

The sample covariance matrix, as usual, is $\boldsymbol{S} = \sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})'/(N-1)$, where the sample mean is $\bar{x} = \sum_{i=1}^{N}(x_1, \ldots, x_n)/N$. The maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{ML}$ is obtained at the minimum of

$$F_{ML}(\theta) = \log|\boldsymbol{\Sigma}\ (\boldsymbol{\theta})| - \log|\boldsymbol{S}| + tr(\boldsymbol{S\Sigma}(\boldsymbol{\theta})^{-1}) - p, \tag{1}$$

yielding the ML test statistic $T_{ML} = (N-1)F_{ML}(\widehat{\boldsymbol{\theta}})$ (Jöreskog, 1969). If the structural model is correct and N is sufficiently large, T_{ML} can be referred to a $\chi^2_{p^*-q}$ distribution, where p^* is the number of nonduplicated elements of $\boldsymbol{S}$. In small samples T_{ML} over-rejects the true model.

The normal-distribution GLS function (Browne, 1974) to be minimized is

$$F_{GLS} = 2^{-1}\, tr[\{(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\boldsymbol{V}\}^2] \tag{2}$$

where $\boldsymbol{V}$ is a consistent estimator of $\boldsymbol{\Sigma^{-1}}$. In practice, $\boldsymbol{V} = \boldsymbol{S^{-1}}$. The GLS estimator $\widehat{\boldsymbol{\theta}}_{GLS}$ is obtained at the minimum of (2), with associated GLS test statistic $T_{GLS} = (N-1)F_{GLS}(\widehat{\boldsymbol{\theta}})$ similarly referred to a $\chi^2_{p^*-q}$ distribution. T_{GLS} tends to under-reject the true model, especially with small *N*.

Improved Test Statistics: RGLS and RLS

Regularized GLS

Arruda and Bentler (2017) proposed that the poor performance of GLS might be due to bias in the eigenvalues of $\boldsymbol{S}$, specifically, their excess extremity (too large or too small) as compared to the eigenvalues of $\boldsymbol{\Sigma}$. The condition number (ratio of largest to smallest eigenvalue) of $\boldsymbol{S}$ is larger than that of $\boldsymbol{\Sigma}$ and decreases monotonically with sample size (Yuan and Bentler 2016). Chi and Lange (2014) proposed Maximum a Posteriori (MAP) estimation, which introduces a nuclear norm penalty (CERNN) in the maximum likelihood framework. This function provides a simple non-linear transformation of the sample eigenvalues thereby offering a reliable means of stabilizing covariance estimation. The general idea of covariance estimation regularization is to extract the eigenvalues from an ill-conditioned covariance matrix if not singular and regularize them according to a quadratic function. Through this regularization scheme highest eigenvalues will be pushed down, and lowest eigenvalues will be pulled up.

The method to do this is straightforward. For a given symmetric matrix $\boldsymbol{S}$, it can be decomposed into eigenvectors and eigenvalues through spectral decomposition.

$$\boldsymbol{S} = \boldsymbol{QDQ'} \tag{3}$$

As shown in equation 3, $\boldsymbol{Q}$ is an orthogonal matrix containing the eigenvectors of $\boldsymbol{S}$, and $\boldsymbol{D}$ is a diagonal matrix that contains the eigenvalues of $\boldsymbol{S}$, $diag(d_1, \cdots, d_p)$. Structured estimation of covariance matrices can be evaluated from two perspectives: generalized linear models and regularization (Pourahmadi 2013). Regularized estimation of covariance matrices and their inverses are based on a wide spectrum of structural assumptions, which has been a subject of debate. Covariance matrix regularization schemes are subject specific. For example, banded sample covariance matrices are suitable for time series and spatial data, in which the order of the components is significant (Chi and Lange 2014, Huang et al. 2006, Rohde and Tsybakov 2011). Chi and Lange (2014) do not assume any special prior structure; instead, they adopt the rotationally-invariant estimators proposed by Stein (1975). As they point out, their main purpose is to regularize the eigenvalue structure of the sample covariance matrix. Stein (1956) suggested an alternative unstructured covariance matrix estimator in the form

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{Q} diag(e_1, \cdots, e_p) \boldsymbol{Q}', \tag{4}$$

where $\widehat{\boldsymbol{\Sigma}}$ is a regularized estimation of covariance matrix with improved eigenvalue structure, and e_i is a shrunken estimate of d_i. This method retains the same eigenvectors. The shrunken estimates are obtained by adding a penalty function to a standard function to steer the estimated eigenvalues toward the geometric mean of sample eigenvalues. In MAP (Chi and Lange 2014), this is done by minimizing the objective function

$$f(\boldsymbol{\Sigma}) = \frac{N}{2} ln|\boldsymbol{\Sigma}| + \frac{N}{2} tr(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) + \frac{\lambda}{2}[\alpha\|\boldsymbol{\Sigma}\|_* + (1-\alpha)\|\boldsymbol{\Sigma}^{-1}\|_*]. \tag{5}$$

As shown in equation 5, the first two terms of which are the typical negative log-likelihood function under normality. According to Chi and Lange (2014), the penalty is the term in brackets and is an α-weighted linear combination of nuclear norms, here, simply trace norms. Intuitively, the sums should be as small as possible. λ is a penalty parameter. As $\lambda \longrightarrow 0$, the solution approaches the maximum likelihood solution, and eigenvalues will equal sample eigenvalues. As λ increases, the more aggressively the eigenvalues are shrunk toward the geometric mean. Appropriately, as N$\longrightarrow \infty$, the data will overwhelm the penalty, making it follow a standard chi-square distribution.

The way to minimize the objective function (equation 5) involves the determinations of λ and α. Alpha (α) is a parameter that controls mixture between the trace and inverse trace penalties. Chi and Lange (2014) proposed to compute it as $\hat{\alpha}_r = \left(1 + \bar{d}^2\right)^{-1}$, where $\bar{d}$ is the mean of the d_i, the eigenvalues of $\boldsymbol{S}$. Arruda and Bentler (2017) showed that $\hat{\alpha}_r$ might be susceptible to extreme sample eigenvalues, they chose to $\hat{\alpha}_R = (1 + \hat{d}^2)^{-1}$ where $\hat{d}$ is the median of the d_i. In their work, Arruda and Bentler showed that $\hat{\alpha}_R$ generated smaller condition numbers, and in the subsequent chi-square test, $\hat{\alpha}_R$ outperforms $\hat{\alpha}_r$. Therefore, in this study we only focus on $\hat{\alpha}_R$, and use it to determine penalty parameter λ.

There are different methods to find the penalty parameter λ based on covariance matrix estimation, we followed the same strategy of Chi and Lange (2014) and chose λ in the unsupervised context. That is, we partition the observed data $\boldsymbol{Y} \in \mathbb{R}^{n \times p}$ into k disjoint sets, and employ κ-fold cross-validation, where often $\kappa = 10$ (Pourahmadi 2013). We partition each data set into training and validation sets. The covariance matrix is estimated based on the validation sample. Subsequently, the estimated covariance matrix is evaluated according to the following predictive negative log-likelihood of the estimated covariance matrix of the training set:

$$\ell_k\left(\widehat{\boldsymbol{\Sigma}}_{\lambda}^{(-k)}, \boldsymbol{Y}_k\right) = \frac{n_k}{2}\ln\det\widehat{\boldsymbol{\Sigma}}_{\lambda}^{(-k)} + \frac{n_k}{2}\,\mathrm{tr}\left(\frac{1}{n_k}\boldsymbol{Y}_k^t\boldsymbol{Y}_k\left[\widehat{\boldsymbol{\Sigma}}_{\lambda}^{(-k)}\right]^{-1}\right). \quad (6)$$

Where $\boldsymbol{Y}_k$ denotes the κth subset, and n_k denotes the number of its rows, and $\widehat{\boldsymbol{\Sigma}}_{\lambda}^{(-k)}$ denotes the estimate using all but the κth partition $\boldsymbol{Y}_k$. During these processes, the estimation is based solely on λ, since we adopt the pre-determined $\hat{\alpha}_R$ at the value derived previously. We repeat the procedure κ times for each value of λ that is auditioned, and an empirical average log-likelihood is calculated. Eventually, a series of penalty parameter $\lambda = 0, \cdots, \lambda_{max}$ are tested and an optimal λ is selected, which minimizes the average ℓ_k over the k folds as follows:

$$\hat{\lambda} = \underset{\lambda\in\{0,\ldots,\lambda_{\max}\}}{\arg\min}\frac{1}{n}\sum_{k=1}^{k}\ell_k\left(\widehat{\boldsymbol{\Sigma}}_{\lambda}^{(-k)}, \boldsymbol{Y}_k\right). \quad (7)$$

Following the abovementioned procedures, the optimal values of λ and α will be selected and incorporated into a quadratic equation (equation 8). The original eigenvalues d_i will be shrunk in conforming to the quadratic equation, and "regularized" eigenvalues e_i will be produced.

$$e_i = \left(-N + \sqrt{N^2 + 4\lambda\alpha[Nd_i + \lambda(1-\alpha)]}\right)/2\lambda\alpha. \quad (8)$$

As shown in equation 8, the values under the square root are nonnegative, so that the covariance matrix must be positive definite. The results are incorporated into equation 4, which derive a regularized covariance matrix $\hat{\Sigma}$. Similar to Arruda and Bentler's (2017), we used Chi and Lange's (2014) MAP function to shrink (move toward their median value) the eigenvalues of $\boldsymbol{S}$ and used

the resulting "regularized" sample covariance matrix, say $\hat{\Sigma}_R$, to replace the GLS weight matrix. Hence, RGLS is simply GLS in (2) with $\mathbf{V} = \hat{\mathbf{\Sigma}}_{\mathbf{R}}^{-1}$. The associated test statistic is T_{RGLS}, here denoted T_R for simplicity. Arruda and Bentler showed that T_R outperforms T_{ML} and T_{GLS}, and produces highly stable results across different sample sizes.

Reweighted least squares

Reweighted least squares is even simpler. The first step is to compute the ML estimator $\widehat{\boldsymbol{\theta}}_{ML}$ and the associated $\hat{\mathbf{\Sigma}}_{ML}$. Then, also using (2), $T_{RLS} = \frac{n}{2}\, tr\{(\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}}_{ML})\, \widehat{\boldsymbol{\Sigma}}_{ML}^{-1}\}^2$. Hence, the estimator is ML, but the GLS function (2) is evaluated with weight matrix $\boldsymbol{V} = \widehat{\boldsymbol{\Sigma}}_{ML}^{-1}$. Hayakawa (2019) reported that RLS avoids the over-rejection problem of ML in the context of a confirmatory factor model, a panel autoregressive model, and a cross-lagged panel model. Zheng and Bentler (2022) also show that RLS outperforms ML and GLS in mean and covariance structure.

From a practical perspective, the test statistic T_{RLS} has been available in EQS for decades (Bentler 2006) and in LISREL after Version 8.52 (Joreskog et al. 2001). RGLS and its test statistic T_R are available in EQS 6.4. All computations and test statistics reported in this study were done with the R package '*lavaan*' (Rosseel 2012) along with original R code for computing the above test statistics developed by, and available from, the senior author.

Data Generation and Simulation Design

In this study, the population follows a traditional confirmatory factor model

$$\boldsymbol{X_i} = \boldsymbol{\Lambda \xi_i} + \boldsymbol{\epsilon_i}, \tag{9}$$

where $\boldsymbol{X_i} = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ is a vector of p observations on person i in a population, and $i = 1,2,\ldots,N$. Under the usual assumptions, this leads to the covariance structure $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}$. Specifically, we take

$$\boldsymbol{\Lambda}' = \begin{bmatrix} .7 & .7 & .75 & .8 & .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .7 & .7 & .75 & .8 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & .7 & .75 & .8 & .8 \end{bmatrix},$$

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & & \\ .3 & 1 & \\ .4 & .5 & 1 \end{bmatrix}$$

We take the diagonal of $\boldsymbol{\Sigma} = \boldsymbol{I}$, so the unique variances are given by $\boldsymbol{\Psi} = \boldsymbol{I} - \text{diag}(\boldsymbol{\Lambda\Phi\Lambda}')$. This population model, previously used by Hu et al. (1992), Huang and Bentler (2015), Arruda and Bentler (2017), Jalal and Bentler (2018), and Zheng and Bentler (2022) was adopted to allow comparison to previous research. Regarding methodology, sample size, and testing criteria, we follow Arruda and Bentler (2017). With p =15, and 3 latent factors, there 33 free parameters, and model tests have 87 degrees of freedom. Under the assumed χ^2_{87} distribution, the expected value of a test statistic is 87 and its expected standard deviation is $\sqrt{2df} \approx 13.19$. In normal distribution, $\boldsymbol{\xi} = \boldsymbol{\Phi}^{1/2}\boldsymbol{Z}_\xi$ and $\boldsymbol{\varepsilon} = \boldsymbol{\Psi}^{1/2}\boldsymbol{Z}_\varepsilon$ where $\boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2} = \boldsymbol{\Phi}$, $\boldsymbol{\Psi}^{1/2}\boldsymbol{\Psi}^{1/2} = \boldsymbol{\Psi}$, and both $\boldsymbol{Z}_\xi$ and $\boldsymbol{Z}_\varepsilon$ followed a standard normal distribution $\mathcal{N}(0, 1)$.

The data generating process consists of two steps. For a given *N*, a sample $\boldsymbol{\xi_i}$ is drawn from a multivariate normal distribution with covariance matrix $\boldsymbol{\Phi}$, while the unique factors $\boldsymbol{\epsilon_i}$ are drawn from a multivariate normal distribution with covariance $\boldsymbol{\Psi}$. These are used to generate the observed $\boldsymbol{X_i}$ using (7). This procedure generates one normal sample from the population structure $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}$; this is repeated 1000 times. In each sample, the parameters are estimated, and various

statistics related to T_{ML}, T_{GLS}, T_R, and T_{RLS} are computed. In addition, all this was repeated at varied sample sizes from 50 to 100,000.

Performance of the various methods across the 1000 replications at each *N* is summarized and presented below. Because RGLS aims to reduce the extremity of eigenvalues of $\hat{\Sigma}_R$ as compared to $\boldsymbol{S}$, we first compute and present the condition numbers of those matrices. Then we summarize results on the performance of the various test statistics, such as their means, biases, standard deviations, and model rejection rates with α=.05 at various sample sizes.

Normal Data Results

Condition Numbers

In each sample, we computed the condition number, the ratio of largest to smallest eigenvalue. The empirical averages of these condition numbers across 1,000 replications at each *N* is shown in Table 1. Given that the population covariance matrix has a condition number of 15.35, at *N* = 100,000 we would expect the condition numbers of $\boldsymbol{S}$ and $\hat{\Sigma}_R$ to be close to 15.35 on average. This occurs, thus validating the simulation results. As *N* increases from 50 to 100,000, the mean condition numbers of $\boldsymbol{S}$ monotonically decrease; as expected, eigenvalues are more extreme as *N* decreases. In contrast, the mean condition numbers of $\hat{\Sigma}_R$ are remarkably stable across *N* and are always close to 15.35. RGLS regularization has achieved its objective.

The variability (standard deviations, SDs) of condition numbers across replications within a given *N* also show, as expected, larger SDs with smaller Ns. However, these SDs vary widely from .09 to 14.87 for $\boldsymbol{S}$, but are much more stable with SDs of .08 to 2.71 for $\hat{\Sigma}_R$.

Table 1

Average Condition Number and Standard Deviations by Sample Size

Sample Size	**S** Cond	$\hat{\Sigma}_R$	**S** Cond SD	$\hat{\Sigma}_R$ SD
50	51.53	15.60	14.87	2.71
60	42.73	15.34	10.57	2.56
70	37.08	15.29	8.18	2.29
80	34.06	15.10	7.00	2.13
90	31.92	15.23	6.34	2.18
100	29.90	15.32	5.48	2.12
110	28.41	15.23	4.89	2.05
120	27.14	15.25	4.50	1.89
250	21.35	15.31	2.35	1.45
400	19.27	15.52	1.72	1.22
500	18.67	15.54	1.46	1.15
1000	17.29	15.60	0.99	0.80
2000	16.53	15.66	0.67	0.59
2500	16.36	15.66	0.57	0.54
5000	15.99	15.65	0.40	0.38
100,000	15.46	15.44	0.09	0.08

Performance of Test Statistics

Table 2 shows the mean values, across 1,000 replications of T_{ML}, T_{GLS}, T_R, and T_{RLS} at each sample size. The expected mean test statistic for each estimation method is 87. The table also shows the percent bias of each of these means. It is obvious that the mean T_{ML} is always too large (positive bias), except at the largest sample sizes; the mean T_{GLS} is typically too small (negative bias), except at the largest *N*. These results are consistent with previous simulation research, such as Yuan and Bentler (1999) for ML results with smaller *N*, and normal theory condition 1 for ML and GLS in Hu et al. (1992) for larger *N*. The mean T_R shows a small positive bias at the smallest *N*s, but the mean T_{RLS} shows virtually no bias – less than 1% at all but one sample size.

Table 2

Mean Test Statistics and Bias by Sample Size

Sample Size	T_{ML}	% Bias	T_{GLS}	% Bias	T_R	% Bias	T_{RLS}	% Bias
50	101.435	16.592	76.332	-12.262	92.915	6.799	87.584	0.671
60	99.846	14.766	79.105	-9.074	90.145	3.615	87.695	0.799
70	96.878	11.354	80.499	-7.472	88.280	1.472	87.780	0.897
80	95.914	10.246	81.099	-6.783	87.753	0.865	87.196	0.225
90	95.007	9.203	81.092	-6.791	87.050	0.058	87.653	0.751
100	94.324	8.418	82.716	-4.924	87.678	0.779	87.595	0.684
110	92.811	6.679	82.764	-4.869	87.397	0.456	87.675	0.776
120	93.271	7.208	83.725	-3.765	87.779	0.895	87.436	0.501
250	89.481	2.852	85.451	-1.781	87.672	0.773	87.722	0.830
400	88.911	2.196	85.783	-1.398	87.266	0.306	87.675	0.776
500	88.627	1.870	85.411	-1.827	87.534	0.614	87.428	0.492
1000	87.817	0.939	85.938	-1.220	87.196	0.225	87.060	0.069
2000	87.041	0.048	86.034	-1.110	86.935	0.074	87.941	1.082
2500	86.623	0.433	86.293	-0.812	86.980	0.022	87.284	0.327
5000	87.195	0.224	86.413	-0.675	87.204	0.235	87.084	0.096
100,000	87.486	0.559	86.941	-0.067	87.037	0.043	87.376	0.432

Note: Target for bias calculations=87

Table 3 shows the SDs of the test statistics across replications within each sample size, expected to be about 13.19. All methods' SDs meet our expectations when N>400. The SDs of T_R and T_{RLS} are generally more stable than those of T_{ML} and T_{GLS}. With smallest *N*s, the SDs of T_{ML} and T_{GLS}, and especially that of T_R at *N*=50, deviate from 13.19.

Table 3

Standard Deviation of Test Statistics across Replications

Sample Size	T_{ML}	T_{GLS}	T_R	T_{RLS}
50	15.844	10.342	24.648	12.081
60	15.446	11.101	14.461	13.060
70	14.922	11.935	13.444	12.983
80	14.467	11.842	13.059	12.669
90	14.083	11.961	13.215	13.002
100	14.163	11.933	13.360	13.320
110	14.170	12.506	12.823	13.423
120	13.943	12.418	13.296	13.082
250	14.120	12.881	13.061	13.363
400	13.521	13.302	12.935	13.073
500	12.904	13.476	13.445	13.107
1000	13.297	13.046	13.009	13.470
2000	12.685	12.714	13.655	12.994
2500	12.665	13.380	13.810	13.165
5000	13.307	12.842	13.692	13.087
100,000	13.296	13.197	12.744	12.907

Note: Target standard deviation is 13.19.

Mean P-values and Empirical Rejection Rates

Next, we turn to performance of average p-values and empirical rejection frequencies. Since this simulation was done under the null hypothesis, the distribution of p-values should be approximately uniform with a mean of .5, and with the chosen significance level $\alpha = .05$, the expected empirical rejection rates of the correct model should be about .05.

Table 4 presents, for each sample size and for each statistic, the mean p-value across the 1,000 replications as well as the number and proportions of p-values less than .05. The average p-values are given in the left part of the table. When $N \geq 1{,}000$, the mean p-values of all methods are close to .5, while with N=50, mean p-values of ML and GLS deviate substantially from .5 in opposite directions. The mean p-values of RGLS and RLS are marginally less than .5, with RLS being more stable. The mean p-values of RGLS range from .421 to .507, while those of RLS vary from .483 to .498.

Table 4

Simulation Results on P-values and Empirical Rejection Rates

	Average P-values				Rejection Rates			
Sample Size	ML	GLS	RGLS	RLS	ML	GLS	RGLS	RLS
50	0.242	0.739	0.421	0.483	0.284	0.003	0.118	0.031
60	0.264	0.675	0.442	0.486	0.252	0.006	0.084	0.051
70	0.311	0.640	0.475	0.483	0.177	0.009	0.058	0.048
80	0.322	0.629	0.487	0.494	0.174	0.010	0.064	0.045
90	0.337	0.628	0.502	0.485	0.151	0.011	0.053	0.050
100	0.353	0.592	0.487	0.487	0.137	0.010	0.059	0.063
110	0.385	0.595	0.494	0.484	0.116	0.025	0.051	0.058
120	0.370	0.573	0.506	0.488	0.119	0.029	0.061	0.056
250	0.450	0.534	0.486	0.484	0.076	0.034	0.050	0.060
400	0.460	0.524	0.497	0.485	0.073	0.042	0.053	0.056
500	0.464	0.534	0.507	0.492	0.059	0.041	0.062	0.054
1000	0.480	0.525	0.496	0.498	0.055	0.045	0.054	0.056
2000	0.497	0.520	0.503	0.482	0.039	0.033	0.059	0.054
2500	0.506	0.519	0.504	0.494	0.039	0.047	0.063	0.055
5000	0.490	0.511	0.501	0.498	0.050	0.041	0.050	0.046
100,000	0.493	0.505	0.493	0.490	0.064	0.059	0.038	0.039

The rejection rates out of 1000 replications are shown on the right of Table 4. The results reveal that the empirical rejection rates are excessively large at the smallest of sample sizes for ML; while GLS over-accepts the null hypothesis, with rejection rates $\leq .01$ when $N \leq 100$. When $N=500$, there rejection rates are more reasonable for all methods. The RGLS empirical rejection rates are close to the nominal level, ranging from .038 to .118, but a bit too large at $N<70$. In contrast, the RLS empirical rejection rates are very stable across N, ranging from .031 to .063.

RLS and RGLS in Non-Normal Data

Non-normally distributed data are ubiquitous in real world data analysis, and goodness-of-fit tests which work in normal data may not work equally well in non-normal data. In this section, we will evaluate the performances of RLS and RGLS in the context of non-normal data distributions. We use three different distributional conditions: a normal distribution, an elliptical distribution, and a

skewed distribution. Data generation procedure of normal distribution is the same as what we have discussed in the previous section, where $\boldsymbol{x}$ is simulated from a confirmatory factor analysis (CFA) based on $\boldsymbol{x} = \boldsymbol{\Lambda\xi} + \boldsymbol{\varepsilon}$, and the population covariance matrix is $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}$. In the elliptical distribution condition (symmetric distributions with heavy tails), $\xi = r\boldsymbol{\Phi}^{1/2}Z_{\xi}$, and $\boldsymbol{\varepsilon} = r\boldsymbol{\Psi}^{1/2}Z_{\boldsymbol{\varepsilon}}$ with r $\sim (3/\chi_5^2)^{1/2}$, $\boldsymbol{\Phi} = cov(\xi)$ and $\boldsymbol{\Psi} = cov(\boldsymbol{\varepsilon})$. In the skewed distribution condition, $\xi = r\boldsymbol{\Phi}^{1/2}Z_{\xi}$ and $\boldsymbol{\varepsilon} = r\boldsymbol{\Psi}^{1/2}Z_{\boldsymbol{\varepsilon}}$ where $Z_{\xi} \sim$ standardized (χ_1^2). For each condition, we simulated 1,000 samples. This method of generating elliptical and skewed distributions has been used by Hu et al. (1992), Yuan and Bentler (1998), and Du and Bentler (2022). The descriptive statistics about skew and kurtosis of the variables are include in Table A1 in Appendix.

In this study, we propose to examine three robust test statistics: the scaled test, the adjusted test, and the adjusted test with a df correction on ML, RLS and RGLS. These robust tests are defined as

$$\dot{T}_{ML} = \frac{df}{\hat{a}_1} T_{ML} \tag{10}$$

$$\ddot{T}_{ML} = \frac{\hat{a}_1}{\hat{a}_2} T_{ML} \tag{11}$$

$$\ddot{T}^{c}_{ML} = \frac{\hat{a}_1}{\hat{a}^{c}_2} T_{ML} \tag{12}$$

In equation 10, $\dot{T}_{ML}$ is referred to a chi-square distribution with df degrees of freedom. In these equations, $\hat{a}_1 = tr(\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Gamma}})$, and $\hat{a}_2 = tr[(\boldsymbol{U\Gamma})^2]$. $\widehat{\boldsymbol{U}} = \widehat{\mathbf{W}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}\,(\widehat{\mathbf{G}}'\widehat{\mathbf{W}}\widehat{\mathbf{G}})^{-1}\widehat{\mathbf{G}}'\widehat{\mathbf{W}}$, $\widehat{\mathbf{W}} = \frac{1}{2}\mathrm{D}_p'(\hat{\Sigma}^{-1}\otimes\hat{\Sigma}^{-1})\mathrm{D}_p$, $\widehat{\mathbf{G}} = \mathbf{G}\big(\widehat{\boldsymbol{\theta}}_{ML}\big)$, and $\mathbf{G}(\widehat{\boldsymbol{\theta}}_{ML}) = \frac{\partial\boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}}_{ML})}{\partial\widehat{\boldsymbol{\theta}}'_{ML}}$. $\widehat{\boldsymbol{\Omega}} = (N-1)^{-1}\sum_{i=1}^{N}(s_i - \bar{s})(s_i - \bar{s})'$, and $\bar{s} = N^{-1}\sum_{i=1}^{N} s_i$. Satorra and Bentler (1988) proposed scaling the test statistics $T_{SB} = T_{ML}/k$, where $k = tr(\boldsymbol{U\Gamma})/df$ is a scaling factor that corrects T_{ML} so that the sampling distribution of T_{SB} at least matches the first moment of the nominal chi-square distribution. The scaling factor k is an estimate of the average of the nonzero eigenvalues of $\boldsymbol{U\Gamma}$. Tong and Bentler (2013) find that when

$N < df$, equation 10 will not be a correct formula because there will not be the same number of eigenvalues that match the degrees of freedom. Moreover, Satorra and Bentler (1994) proposed adjusting the test statistic $\ddot{T}_{ML}$ so that the resulting statistic has the same mean and variance as the chi-square distribution. Hence, in equation 12 $\ddot{T}_{ML}$ is referred to as a chi-square distribution with df' degrees of freedom, and $df' = \frac{\hat{a}_1}{\hat{a}_2}$.

Recently Hayakawa (2019) introduced a new adjusted test with a correction, $\ddot{T}^c_{ML}$, as shown in equation 12, which is referred to as a chi-square distribution with df' degrees of freedom. In $\ddot{T}^c_{ML}$ the unbiased estimator is $\hat{a}_2$, which was proposed by Srivastava et al. (2014) and Himeno and Yamada (2014). Du and Bentler (2022) also proposed to use the same unbiased asymptotic distribution free (ADF) estimator $\hat{a}_2$ in the study of robust test statistics. This new $\hat{a}_2$ is different from that of equation 11 in that s_i is defined as

$$s_i = \boldsymbol{\sigma} + \boldsymbol{\Omega^{1/2} u_i}$$

where $s_i = vech\{(x_i - \bar{x})(x_i - \bar{x})'\}$, $E(\boldsymbol{u_i}) = 0$, $E(s_i) = \boldsymbol{\sigma}$, $var(s_i) = \boldsymbol{\Omega}$, $w_i = \boldsymbol{U}^{1/2} s_i$, $E(w_i) = \boldsymbol{U}^{1/2}\boldsymbol{\sigma}$, and $var(w_i) = \mathrm{U}^{1/2}\Omega\mathrm{U}^{1/2}$. Therefore, Hayakawa (2019) proposed a new correction of unbiased estimator $\hat{a}_2^c$ in the adjusted test as follows:

$$\hat{a}_2^c = \frac{1}{N(N-1)(N-2)(N-3)}\{(N-2)(N-1)tr((\boldsymbol{Y'Y})^2) - N(N-1)tr((\boldsymbol{D})^2) + [tr(\boldsymbol{Y'Y})]^2\},$$

where $\boldsymbol{Y} = (y_1, y_2, \dots, y_N)$, $\boldsymbol{D} = diag(y_1'y_1, \dots, y_N'y_N)$, $y_i = \widehat{w}_i - \bar{\widehat{w}}$, $\bar{\widehat{w}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{w}_i$.

The robust tests in equations 10, 11, and 12 are based on T_{ML}. In the following analyses, we replace T_{ML} with the corresponding test statistics of RLS and RGLS, T_{RLS} and T_R. Specifically, for the scaled test, we denote the ML, RLS and RGLS as $\dot{T}_{ML}$, $\dot{T}_{RLS}$ and $\dot{T}_R$. Similarly, we use $\ddot{T}_{ML}$, $\ddot{T}_{RLS}$ and $\ddot{T}_R$ to indicate the adjusted tests of ML, RLS and RGLS. For their corresponding adjusted tests with a correction, we denote them as $\ddot{T}^c_{ML}$, $\ddot{T}^c_{RLS}$ and $\ddot{T}^c_R$.

Non-normal Data Test Results

The evaluation of the test statistics is based on empirical type I error rates. The results reported in Table 5 are empirical Type I error rates ×1000 with $\alpha = 0.05$ across all tests. Based on criteria in Bradley (1978), the acceptable empirical Type I error rates (i.e., 0.025 to 0.075) are in bold font.

Table 5
Type I error rates × 1000 with $\alpha = 0.05$

Normal Distrubution									
	Scaled test			Adjusted test			Adjusted test with a correction		
N	$\dot{T}_{ML}$	$\dot{T}_{RLS}$	$\dot{T}_{R}$	$\ddot{T}_{ML}$	$\ddot{T}_{RLS}$	$\ddot{T}_{R}$	$\ddot{T}^{c}_{ML}$	$\ddot{T}^{c}_{RLS}$	$\ddot{T}^{c}_{R}$
50	376	129	352	**50**	**67**	**52**	390	153	78
80	196	86	122	**59**	105	**53**	209	102	**43**
100	160	89	**59**	**39**	104	23	147	83	**57**
200	104	**56**	**50**	**25**	**54**	**27**	99	**62**	**45**
300	81	**70**	**34**	**37**	**47**	**27**	90	**50**	**39**
400	**51**	**74**	**69**	**51**	**59**	**35**	82	**54**	**45**
500	**73**	**55**	**37**	**43**	**44**	**26**	**49**	**35**	**37**
800	**56**	**52**	**45**	**52**	**26**	**25**	**54**	**62**	**45**
1,000	**70**	**68**	**59**	**74**	12	**42**	**47**	**38**	**48**
2,000	**54**	**50**	**50**	**69**	**45**	**72**	**62**	**41**	**57**
5,000	**36**	**52**	**48**	**63**	**71**	83	**54**	**56**	**61**
Elliptical Distrubution									
	Scaled test			Adjusted test			Adjusted test with a correction		
N	$\dot{T}_{ML}$	$\dot{T}_{RLS}$	$\dot{T}_{R}$	$\ddot{T}_{ML}$	$\ddot{T}_{RLS}$	$\ddot{T}_{R}$	$\ddot{T}^{c}_{ML}$	$\ddot{T}^{c}_{RLS}$	$\ddot{T}^{c}_{R}$
50	458	154	17	18	4	0	446	112	7
80	220	102	17	9	7	0	234	91	16
100	186	82	20	7	3	0	213	82	**27**
200	134	**71**	14	6	5	0	120	**75**	**26**
300	103	**62**	14	11	10	6	81	**74**	**33**
400	**72**	**66**	**53**	10	9	4	86	**71**	**45**
500	84	**63**	**33**	**39**	18	12	79	**69**	**39**
800	**64**	**73**	**46**	14	**28**	13	**60**	**65**	**45**
1,000	**70**	**58**	**52**	**45**	**29**	**34**	**55**	79	**42**
2,000	**72**	**67**	**56**	**25**	**32**	**41**	**75**	**67**	**39**
5,000	**56**	**71**	**40**	**54**	**45**	**43**	**74**	**46**	**63**

	Skewed Distrubution								
	Scaled test			Adjusted test			Adjusted test with a correction		
N	$\dot{T}_{ML}$	$\dot{T}_{RLS}$	$\dot{T}_R$	$\ddot{T}_{ML}$	$\ddot{T}_{RLS}$	$\ddot{T}_R$	$\ddot{T}^c_{ML}$	$\ddot{T}^c_{RLS}$	$\ddot{T}^c_R$
50	244	146	166	8	8	0	469	97	**62**
80	191	108	122	11	9	2	280	78	**58**
100	184	100	116	8	4	0	205	**65**	**69**
200	125	84	85	7	3	0	164	**66**	**52**
300	107	80	83	5	11	2	117	**63**	**50**
400	**63**	91	82	17	9	6	90	**49**	**62**
500	78	90	82	8	8	0	83	77	**58**
800	80	**59**	**75**	15	6	4	**42**	**75**	**55**
1,000	**62**	**62**	**74**	21	17	6	**71**	**71**	**55**
2,000	**56**	**55**	**69**	7	**28**	**36**	**66**	**64**	**61**
5,000	**68**	**70**	**75**	**39**	**43**	21	23	**49**	**55**

Note: The acceptable empirical Type I error rates ×1000 (i.e., 25 to 75) are in bold font.

Table 5 shows that in normal distribution, and when sample size is small, $\dot{T}_{ML}$, $\dot{T}_{RLS}$, and $\dot{T}_R$ start to experience more empirical Type I errors. The explanation is that when sample size is smaller than 100, $\widehat{\boldsymbol{\Omega}}$ becomes an inefficient estimator for $var(s_i)$, because it contains the fourth-order moments (Hayakawa 2019). Nonetheless, when sample size is larger, their empirical Type I error rates increasingly become closer to the nominal level (i.e., 0.05). In normal distribution case, $\dot{T}_{RLS}$, and $\dot{T}_R$ have similar performances. In the adjusted test, the empirical Type I error rates of $\ddot{T}_{ML}$ and $\ddot{T}_{RLS}$ are close to nominal level, while $\ddot{T}_R$ is consistently deviated from it. In the adjusted test with a correction, the empirical Type I error rates of both $\ddot{T}^c_{ML}$ and $\ddot{T}^c_{RLS}$ are inflated when sample size is smaller 100. In contrast, $\ddot{T}^c_R$ delivers the most consistent Type I error rates that are within the acceptable range from 25 to 75.

In non-normal cases, both elliptical and skewed distributions, $\dot{T}_{ML}$, $\dot{T}_{RLS}$ and $\dot{T}_R$ have good performances in empirical Type I error rates when N>500, whereas they tend to increasingly over-reject the null hypothesis as sample size becomes smaller and inflate the Type I error rates. $\dot{T}_R$ has an interesting behavior in the scaled test. $\dot{T}_R$ tends to over-reject the null hypothesis when sample sizes are smaller than 100 in both normal and skewed distributions, whereas it tends to under-reject

the null hypothesis and deflate the Type I error rates in elliptical distribution case. In the adjusted test, all these robust test statistics tend to deflate the Type I error rates. This indicates that in real data analysis, the adjusted test seems to be biased towards the null hypothesis, making it the least ideal robust test of all that we have examined in this study. As for the adjusted test with a correction, $\ddot{T}_{RLS}^{c}$ and $\ddot{T}_{R}^{c}$ outperform $\ddot{T}_{ML}^{c}$ in both elliptical and skewed distribution cases, because they both reduce the Type I error bias when sample sizes are smaller than 400. By and large, $\ddot{T}_{RLS}^{c}$ avoids over-rejection problem unless the sample sizes are less than100 in both normal and non-normal distribution cases. Whereas $\ddot{T}_{R}^{c}$ is advantaged over $\ddot{T}_{RLS}^{c}$ in delivering the acceptable empirical Type I error rates in normal and non-normal distribution cases, which are negligibly deviated from .05, especially when sample sizes become smaller than 100.

Power Analysis

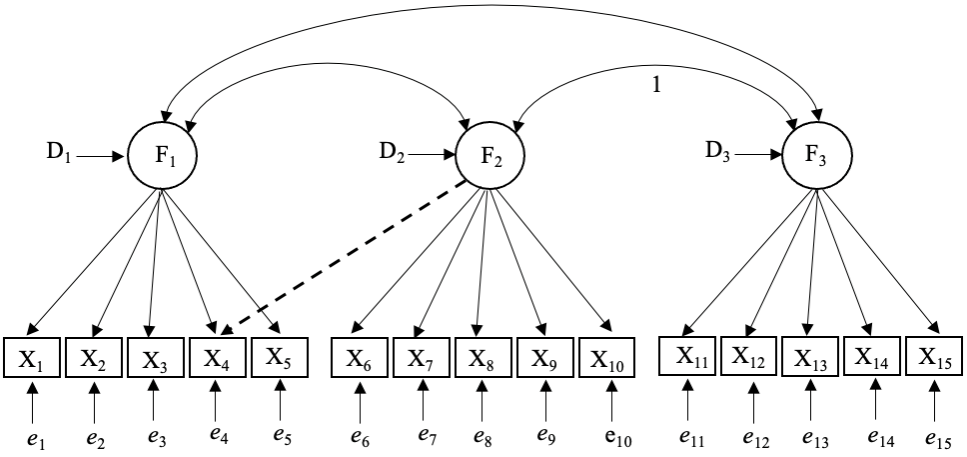
To better examine the performances of ML, RLS and RGLS, we need to compare the performances in which the models are misspecified, and examine which method is better to handle Type II error. If a test statistic requires smaller sample size to reject models with misspecification, then the power of that test is stronger. The power analysis requires that $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\boldsymbol{\theta})$. This is done in two ways: Modifying our population model; and modifying our analysis model.

Condition 1 consists of a modified population model and the original analysis model. In the population model we added two extra parameters to the original population model. We connect the second factor with the first manifest variable and the third factor with the sixth manifest variable and set the factor loadings at the values of .2 and .3 respectively. Thus, the new factor loading matrix is defined as:

$$\boldsymbol{\Lambda}' = \begin{bmatrix} 0.7 & 0.7 & 0.75 & 0.8 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0.7 & 0.7 & 0.75 & 0.8 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0.7 & 0.7 & 0.75 & 0.8 & 0.8 \end{bmatrix}.$$

For condition 2, we modified the original analysis model as indicated in Figure 1 by creating an extra path connecting F_2 to X_4 and holding the correlation between F_2 and F_3 fixed. Under this condition it has a larger misspecification, thus we expect smaller sample sizes to reject the model. The samples are simulated from the original population model.

Figure 1
Diagram of the misspecified analysis model



In both conditions, 1,000 replicated samples were drawn from a population with covariance structure. Because the hypothesized models are incorrect, we expect to reject them, and the rejection rate shows an estimate of the power of the test under these model misspecifications. The mean p-value and rejection rate for each replicated sample are computed. Table 6 reports mean p-values and rejection rates of regular T_{ML}, T_{RLS}, and T_R in normal distribution. Table 7 reports that scaled test, adjusted test, and adjusted test with a correction of ML, RLS and RGLS in both normal and non-normal cases.

Table 6
Power analysis in normal distribution

Normal Distribution

	Condition 1							**Condition 2**						
	T_{ML}		T_{RLS}		T_R			T_{ML}		T_{RLS}		T_R	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.192	0.354	0.378	0.108	0.415	0.110	50	0.075	0.678	0.258	0.220	0.345	0.154
80	0.245	0.265	0.366	0.111	0.457	0.077	80	0.063	0.720	0.132	0.493	0.292	0.215
100	0.279	0.203	0.350	0.135	0.447	0.073	100	0.039	0.816	0.074	0.682	0.222	0.241
200	0.255	0.273	0.291	0.194	0.350	0.124	200	0.004	0.974	0.002	0.990	0.069	0.664
300	0.195	0.348	0.218	0.301	0.271	0.207	300	0.001	0.995	0.000	1.000	0.012	0.942
400	0.156	0.425	0.159	0.439	0.199	0.337	400	0.000	1.000	0.000	1.000	0.002	1.000
500	0.107	0.574	0.109	0.563	0.147	0.445							
800	0.033	0.831	0.032	0.839	0.052	0.761							
1,000	0.012	0.920	0.014	0.936	0.021	0.886							
2,000	0.000	0.999	0.000	1.000	0.000	1.000							

As we can see in Table 6, the empirical power of different estimators increases with larger sample sizes. In condition 1, when the sample size is about 2,000 and in condition 2 when sample size is about 400, all estimators completely reject the chi-square test statistics. Nonetheless, the rejection rates vary with the estimators within these sample sizes. In condition 1 when sample sizes N<800, and in condition 2 when N<200 the ML method produces relatively smaller mean p-values. As a result, it can reject in both conditions with smaller sample sizes as compared to RLS and RGLS. Therefore, the ML method produces the most power, although this is not meaningful since ML does not control Type I errors well. Comparing RLS and RGLS, in condition 1 when N<1,000, and in condition when N<200, RLS starts to outperform RGLS by producing smaller mean p-values, thus RLS produces more power than RGLS.

This performance is consistent in both normal and non-normal cases for scaled test, adjusted test, and adjusted test with a correction based on ML, RLS and RGLS as shown in Table 7. As documented by Zheng and Bentler (2022), GLS has less power than ML and RLS, and this relationship translates to RGLS.

Table 7
Power analysis of robust estimators in condition 1

Condition 1

Normal Distribution

	Scaled test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_R$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_R$		$\ddot{T}^c_{ML}$		$\ddot{T}^c_{RLS}$		$\ddot{T}^c_R$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.009	0.959	0.033	0.829	0.134	0.446	0.053	0.642	0.033	0.815	0.219	0.080	0.009	0.962	0.025	0.866	0.158	0.360
80	0.001	0.993	0.003	0.995	0.078	0.698	0.012	0.943	0.007	0.964	0.146	0.315	0.001	0.998	0.003	0.987	0.081	0.706
100	0.001	0.995	0.001	0.997	0.046	0.784	0.004	0.990	0.001	0.999	0.082	0.520	0.000	1.000	0.001	0.996	0.037	0.759
200	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.003	0.995	0.000	1.000	0.000	1.000	0.000	1.000

Elliptical Distribution

	Scaled Test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_R$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_R$		$\ddot{T}^c_{ML}$		$\ddot{T}^c_{RLS}$		$\ddot{T}^c_R$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.020	0.897	0.019	0.918	0.376	0.045	0.104	0.250	0.119	0.202	0.385	0.000	0.016	0.923	0.061	0.686	0.375	0.058
80	0.008	0.962	0.008	0.966	0.176	0.412	0.059	0.537	0.067	0.526	0.289	0.006	0.007	0.976	0.025	0.858	0.205	0.217
100	0.002	0.992	0.003	0.979	0.111	0.629	0.037	0.770	0.041	0.733	0.241	0.027	0.003	0.990	0.007	0.957	0.141	0.522
200	0.000	1.000	0.000	1.000	0.009	0.953	0.000	1.000	0.001	0.999	0.038	0.738	0.000	1.000	0.000	1.000	0.007	0.982
300	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.004	0.996	0.000	1.000	0.000	1.000	0.000	1.000
400	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Skewed Distribution

	Scaled Test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_R$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_R$		$\ddot{T}^c_{ML}$		$\ddot{T}^c_{RLS}$		$\ddot{T}^c_R$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.118	0.599	0.252	0.294	0.509	0.036	0.241	0.017	0.251	0.011	0.435	0.000	0.120	0.423	0.260	0.378	0.574	0.000
80	0.166	0.475	0.244	0.324	0.546	0.041	0.270	0.025	0.267	0.019	0.446	0.000	0.169	0.485	0.235	0.370	0.500	0.085
100	0.190	0.421	0.234	0.365	0.521	0.104	0.269	0.035	0.278	0.028	0.440	0.001	0.154	0.494	0.213	0.415	0.483	0.139
200	0.112	0.602	0.121	0.601	0.304	0.191	0.210	0.096	0.205	0.092	0.338	0.005	0.108	0.599	0.110	0.594	0.326	0.194
300	0.049	0.794	0.055	0.791	0.170	0.365	0.126	0.332	0.117	0.347	0.242	0.050	0.051	0.767	0.048	0.811	0.160	0.461
400	0.021	0.907	0.023	0.861	0.079	0.613	0.072	0.555	0.077	0.506	0.151	0.194	0.028	0.878	0.020	0.897	0.081	0.601
500	0.008	0.944	0.007	0.952	0.038	0.810	0.033	0.783	0.030	0.815	0.088	0.440	0.005	0.984	0.007	0.967	0.034	0.829
800	0.000	0.998	0.000	1.000	0.002	0.984	0.004	0.988	0.004	0.980	0.017	0.926	0.000	1.000	0.000	1.000	0.004	0.990

Table 7 shows the robust versions of ML, RLS and RGLS methods in condition 1. In normal distribution, $\dot{T}_{ML}$, $\ddot{T}_{ML}$ and $\ddot{T}^c_{ML}$ tend to have the most power, when N=50, both $\dot{T}_{ML}$ and $\ddot{T}^c_{ML}$ can reject 96 percent of the test statistics. $\dot{T}_{RLS}$, $\ddot{T}_{RLS}$ and $\ddot{T}^c_{RLS}$ tend to have similar power, except when N=50, for which its reject rates is about 82 percent. In contrast, $\dot{T}_R$, $\ddot{T}_R$ and $\ddot{T}^c_R$ tend to have the least power when N<200. In elliptical distribution, $\dot{T}_{ML}$ and $\ddot{T}^c_{ML}$ have similar performance as in normal condition, but the rejection rates of $\ddot{T}_{ML}$ and $\ddot{T}_{RLS}$ reduce a lot when N<200. In contrast, the power of $\dot{T}_R$, $\ddot{T}_R$ and $\ddot{T}^c_R$ reduce a lot in elliptical distribution. In skewed distribution, all variants of robust estimators reduce statistical power to reject the misspecified model, as compared to

normal and elliptical distributions. Still, the performance of $\dot{T}_{ML}$, $\ddot{T}_{ML}$ and $\ddot{T}^{c}_{ML}$ is better than other robust variants of RLS and RGLS.

Table 8
Power analysis of robust estimators in condition 2

Condition 2

Normal Distribution

	Scaled test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_{R}$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_{R}$		$\ddot{T}^{c}_{ML}$		$\ddot{T}^{c}_{RLS}$		$\ddot{T}^{c}_{R}$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.085	0.614	0.244	0.242	0.324	0.167	0.182	0.123	0.147	0.290	0.372	0.018	0.094	0.596	0.241	0.235	0.365	0.108
80	0.127	0.459	0.242	0.239	0.283	0.198	0.177	0.210	0.151	0.356	0.326	0.031	0.113	0.535	0.244	0.255	0.258	0.243
100	0.114	0.516	0.193	0.303	0.246	0.174	0.171	0.218	0.137	0.435	0.293	0.056	0.112	0.560	0.193	0.317	0.200	0.327
200	0.045	0.799	0.081	0.599	0.041	0.739	0.073	0.619	0.065	0.636	0.070	0.566	0.039	0.809	0.094	0.595	0.051	0.741
300	0.011	0.948	0.030	0.844	0.012	0.940	0.020	0.898	0.021	0.887	0.021	0.902	0.012	0.950	0.030	0.840	0.009	0.938
400	0.001	1.000	0.007	0.986	0.001	0.994	0.000	1.000	0.004	0.982	0.003	0.992	0.002	0.993	0.009	0.970	0.001	1.000
500	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.001	0.995	0.000	1.000

Elliptical Distribution

	Scaled Test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_{R}$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_{R}$		$\ddot{T}^{c}_{ML}$		$\ddot{T}^{c}_{RLS}$		$\ddot{T}^{c}_{R}$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.127	0.541	0.132	0.511	0.724	0.000	0.238	0.011	0.245	0.009	0.508	0.000	0.139	0.527	0.287	0.208	0.633	0.034
80	0.199	0.332	0.214	0.309	0.569	0.012	0.316	0.020	0.293	0.014	0.503	0.000	0.210	0.337	0.316	0.205	0.514	0.062
100	0.232	0.307	0.226	0.319	0.509	0.057	0.295	0.017	0.295	0.033	0.475	0.000	0.218	0.348	0.337	0.152	0.578	0.068
200	0.226	0.308	0.213	0.339	0.384	0.118	0.243	0.094	0.278	0.089	0.363	0.045	0.194	0.358	0.273	0.277	0.360	0.176
300	0.166	0.424	0.174	0.385	0.226	0.284	0.224	0.184	0.205	0.178	0.260	0.099	0.163	0.407	0.217	0.315	0.233	0.293
400	0.121	0.544	0.108	0.558	0.159	0.435	0.176	0.262	0.175	0.273	0.197	0.215	0.118	0.553	0.173	0.416	0.170	0.451
500	0.077	0.645	0.061	0.691	0.100	0.556	0.133	0.460	0.110	0.448	0.116	0.392	0.061	0.712	0.120	0.477	0.098	0.547
800	0.012	0.930	0.037	0.893	0.022	0.88095	0.041	0.800	0.061	0.706	0.047	0.765	0.039	0.829	0.030	0.828	0.020	0.897
1,000	0.008	0.939	0.003	1.000	0.008	0.947	0.013	0.917	0.006	0.970	0.015	0.923	0.010	0.953	0.018	0.919	0.002	1.000

Skewed Distribution

	Scaled Test						Adjusted test						Adjusted test with a correction					
	$\dot{T}_{ML}$		$\dot{T}_{RLS}$		$\dot{T}_{R}$		$\ddot{T}_{ML}$		$\ddot{T}_{RLS}$		$\ddot{T}_{R}$		$\ddot{T}^{c}_{ML}$		$\ddot{T}^{c}_{RLS}$		$\ddot{T}^{c}_{R}$	
N	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate	P-value	Rej rate
50	0.149	0.474	0.325	0.197	0.605	0.037	0.272	0.013	0.271	0.014	0.498	0.000	0.143	0.574	0.304	0.310	0.647	0.000
80	0.241	0.351	0.340	0.216	0.575	0.046	0.304	0.009	0.299	0.017	0.493	0.000	0.220	0.415	0.325	0.311	0.618	0.040
100	0.244	0.312	0.335	0.216	0.548	0.051	0.322	0.005	0.328	0.012	0.475	0.000	0.241	0.384	0.347	0.232	0.509	0.055
200	0.245	0.295	0.309	0.194	0.367	0.135	0.309	0.025	0.321	0.024	0.409	0.008	0.246	0.348	0.313	0.244	0.409	0.129
300	0.201	0.397	0.234	0.291	0.277	0.215	0.266	0.046	0.260	0.045	0.316	0.038	0.199	0.364	0.222	0.344	0.324	0.209
400	0.151	0.456	0.188	0.369	0.180	0.378	0.206	0.131	0.211	0.137	0.266	0.065	0.132	0.507	0.181	0.422	0.186	0.400
500	0.094	0.604	0.147	0.470	0.115	0.529	0.154	0.289	0.169	0.260	0.188	0.173	0.101	0.598	0.136	0.537	0.125	0.562
800	0.030	0.828	0.051	0.775	0.037	0.820	0.068	0.620	0.069	0.593	0.096	0.510	0.031	0.845	0.042	0.814	0.018	0.871
1,000	0.013	0.935	0.023	0.859	0.007	0.958	0.031	0.803	0.025	0.830	0.024	0.836	0.011	0.932	0.020	0.904	0.023	0.918

Table 8 shows the performances of robust variants of different estimators in condition 2, under which the analysis model is incorrect. In the normal condition, $\dot{T}_{ML}$ and $\ddot{T}^{c}_{ML}$ have similar rejection rates, while adjusted test $\ddot{T}_{ML}$ has the least power as compared to $\dot{T}_{ML}$ and $\ddot{T}^{c}_{ML}$. Similar performances are also shown in $\dot{T}_{RLS}$, $\ddot{T}_{RLS}$, $\ddot{T}^{c}_{RLS}$, $\dot{T}_{R}$, $\ddot{T}_{R}$ and $\ddot{T}^{c}_{R}$. With both elliptical and skewed distributions, $\dot{T}_{ML}$, $\ddot{T}_{ML}$ and $\ddot{T}^{c}_{ML}$ have similar powers, and consistently less than those in normal

distributions. $\dot{T}_{RLS}$ and $\ddot{T}_{RLS}$ have more power in elliptical condition, while $\ddot{T}^{c}_{RLS}$ has more power in skewed condition when N<300. $\dot{T}_{R}$ tends to have slightly more power in the skewed distribution than in the elliptical distribution. In contrast, $\ddot{T}_{R}$ and $\ddot{T}^{c}_{R}$ have more power in the elliptical distribution than in the skewed one.

Discussion

The most important results of these Monte Carlo simulations involve the rejection rates of the four test statistics at various sample sizes in normal data, and the Type I error rates in non-normal data. We found that T_{RLS} and T_{R} perform equally well when the samples are sufficiently large (N>70 in this study), although the behavior of RLS is near-ideal at all sample sizes. Consistent with prior research, and hence not surprisingly, both methods clearly outperform ML and GLS at intermediate to small sample sizes (N≤400). In non-normal cases, particularly the adjusted test with a correction, both $\ddot{T}^{c}_{RLS}$ and $\ddot{T}^{c}_{R}$ are superior to $\ddot{T}^{c}_{ML}$ in overcoming the over-rejecting problem, and $\ddot{T}^{c}_{R}$ has more consistent performance in delivering p-values that are within the acceptable range than $\ddot{T}^{c}_{RLS}$.

These results are consistent with the separate results of Arruda and Bentler (2017) and Hayakawa (2019), but go beyond earlier work by showing that RLS is superior to RGLS at N=50, a condition not considered by Arruda and Bentler (2017). In the case of non-normal distribution, the results of this study really advanced our understanding of RGLS. However, there is some conflict with RGLS at N=60, with Arruda-Bentler showing a rejection rate of .065, while this study found the less acceptable rate of .084. In order to clarify the performance of these statistics at smaller sample sizes, an additional simulation study was done using the same model and methodology as earlier, but with a greater range of small sample size (N=50, 55, 60, 65, 70, 75).

For better stability, the number of replications was increased to 2000. The results are shown in Table 9 in similar format as in previous tables.

Table 9
Simulation Results for Small Samples

	Test Statistics (SD)				Average P-values		Rejection Rates	
N	T_R	(SD)	T_{RLS}	(SD)	RGLS	RLS	RGLS	RLS
50	91.57	20.81	88.02	12.46	0.42	0.48	0.11	0.05
55	90.45	21.06	87.71	12.84	0.44	0.48	0.10	0.05
60	89.30	13.97	87.83	12.69	0.46	0.48	0.08	0.05
65	88.27	13.59	87.59	13.05	0.48	0.49	0.06	0.05
70	88.48	13.16	87.50	12.42	0.47	0.49	0.06	0.04
75	88.55	13.27	87.29	12.88	0.47	0.49	0.06	0.05

Clearly, RGLS marginally over-rejects the true model at *N*=50, 55, 60. In contrast, the rejection rate of RLS is basically perfect. These findings suggest that among the methods considered here, T_{RLS} is the best choice for general SEM practice. It is advantaged over T_R of simplicity in that it is much easier to program and requires less computational power. However, to get a more complete understanding of their comparative characteristics and advantages, further research could compare the performances of T_{RLS} and T_R with a greater range of number of factors and indicators, varying sizes of factor loadings, and an extended range of unique variances. Of course, it would be of interest to see whether any advantage of T_{RLS} over T_R also occurs in models that include a mean structure, such as in growth curve models.

Another question for further research is whether the marginally problematic small sample behavior of T_R can be further improved. As shown in Table 7, the mean of T_R was a bit too high, but its SD was especially high, suggesting that perhaps the regularization used to obtain the weight matrix $\boldsymbol{V}$ in the GLS function (eq. 2) was not always effective enough. RGLS has so far been based on a default Chi and Lange (2014) methodology for shrinking the eigenvalues of the sample

covariance matrix. With small N, perhaps its tuning parameters need to be adjusted, or a completely different shrinkage method considered.

Regarding the power to reject false models, ML and its robust variants tend to outperform RLS and RGLS, along with their robust variants in normal and non-normal cases. However, the inability of ML to control the alpha level under the null hypothesis makes its results less meaningful; since it overrejects the true model, rejecting a false model is not much of an accomplishment. At the same time, the loss in empirical power of RLS – which controls Type I errors –is small compared to ML. Yet, RLS exceedingly well controls alpha level with a correct model. In general, RLS and its robust variants outperform those of RGLS in both normal and non-normal cases.

Finally, the current paper only considered RLS and RGLS in the context of normal theory GLS and GLS with robust corrections. It would be interesting if there were parallels to these methods in the asymptotically distribution free (ADF) method (Browne 1982, Browne 1984). However, we know of no way that RLS can be generalized to ADF, since there is no ML estimator of the ADF weight matrix. However, as noted by Arruda and Bentler (2017), since the estimated ADF weight matrix can be computed as the inverse of a type of sample covariance matrix (Huang and Bentler 2015, Satorra 1992), the eigenvalues of this weight matrix also may be too extreme in small samples. Hence, in theory, weight matrix regularization may also improve the performance of ADF, but whether it actually does so, remains to be studied.

References

Arruda, E. H. and P. Bentler. 2017. "A Regularized Gls for Structural Equation Modeling." *Structural Equation Modeling: A Multidisplinary Journal* (24):657-65.

Bentler, P. 2006. *Eqs 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Boomsma, A. 1982. "The Robustness of Lisrel against Small Sample Sizes in Factor Analysis Models. ." Pp. 149-73 in *Systems under Indirect Observation: Causality, Structure, Prediction: Part 1*, edited by K. G. Joreskog and H. Wold. Amsterdam, The Netherlands: North-Holland.

Bradley, J. V. 1978. "Robustness?". *British Journal of Mathematical and Statistical Psychology* 31(2):144-52. doi: doi: 10.1111/j.2044-8317.1978.tb00581.x.

Browne, M. 1974. "Generalized Least Squares Estimators in the Analysis of Covariance Structures." *South African Statistical Journal* 8:1-24.

Browne, M. 1982. "Covariance Structures." Pp. 72-141 in *Topics in Applied Multivariate Analysis*, edited by D. M. Hawkins. Cambridge, UK: Cambridge University Press.

Browne, M. 1984. "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37:62-83. doi: doi:10.1111/bmsp.1984.37.issue-1.

Chi, E. C. and K. Lange. 2014. "Stable Estimation of a Covariance Matrix Guided by Nuclear Norm Penalties." *Computional Statistics and Data Analysis* (80):117-28.

Du, H. and P. M. Bentler. 2022. "40-Year Old Unbiased Distribution Free Estimator Reliably Improves Sem Statistics for Nonnormal Data." *Structural Equation Modeling: A Multidisplinary Journal*:1-16. doi: https://doi.org/10.1080/10705511.2022.2063870.

Hair, J. F., G. T. M. Hult, C. M. Ringle and M. Sarstedt. 2017. *A Primer on Partial Least Squares Structural Equation Modeling (Pls-Sem). 2nd Edition*. Thousand Oaks, CA: Sage Publications Inc.

Hayakawa, K. 2019. "Corrected Goodness-of-Fit Test in Covariance Structure Analysis." *Psychological Methods* 24(3):371-89.

Himeno, T. and T. Yamada. 2014. "Estimations for Some Functions of Covariance Matrix in High Dimension under Non-Normality and Its Applications." *Journal of Multivariate Analysis* 130:27-44.

Hu, L.-T., P. Bentler and Y. Kano. 1992. "Can Test Statistics in Covariance Structure Analysis Be Trusted?". *Psychological Bulletin* 112(2):351-62.

Huang, J. Z., N. Liu, M. Pourahmadi and L. Liu. 2006. "Covariance Matrix Selection and Estimation Via Penalised Normal Likelihood." *Biometrika* 93(1):85–98.

Huang, Y. and P. Bentler. 2015. "Behavior of Asymptotic Distribution Free Test Statistics in Covariance Versus Correlation Structure Analysis." *Structural Equation Modeling: A Multidisplinary Journal* 22:489-503.

Jalal, S. and P. Bentler. 2018. "Using Monte Carlo Normal Distribution to Evaluate Structural Models with Nonnormal Data." *Structural Equation Modeling: A Multidisplinary Journal* 25:541-57.

Joreskog, K. G., D. Sorbom, S. du Toit and M. du Toit. 2001. *Lisrel 8: New Statistical Features*. Illinois, U.S.: Scientific Software International.

Jöreskog, K. G. 1969. "Some Contribution to Maximum Likelihood Factor Analysis." *Psychometrika* 34:183-202.
Jorgensen, T. D., S. Pornprasertmanit, A. M. Schoemann and Y. Rosseel. 2022. "R Package' Semtools'."
Lee, S.-Y. 2007. *Structural Equation Modeling: A Bayesian Approach*. England: John Wiley & Sons Ltd.
Moshagen, M. 2012. "The Model Size Effect in Sem: Inflated Goodness-of-Fit Statistics Are Due to the Size of the Covariance Matrix." *Structural Equation Modeling: A Multidisplinary Journal* 19:86-98.
Pourahmadi, M. 2013. *High-Dimensional Covariance Estimation*. Hoboken, NJ: Wiley.
Rohde, A. and A. B. Tsybakov. 2011. "Estimation of High-Dimensional Low-Rank Matrices." *The Annals of Statistics* 39(2):887–930.
Rosseel, Y. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48(2):1-36.
Satorra, A. and P. M. Bentler. 1988. "Scaling Corrections for Chi-Square Statistics in Covariance Structure Analysis." Pp. 308-13 in *ASA Proceedings of the Business and Economic Section*: American Statistical Association.
Satorra, A. 1992. "Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures." *Sociological Methodology* 22:249-78. doi: 10.2307/270998.
Satorra, A. and P. Bentler. 1994. *Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis*: Sage Publications, Inc.
Shi, D., T. Lee and R. A. Terry. 2018. "Revisiting the Model Size Effect in Structural Equation Modeling." *Structural Equation Modeling: A Multidisplinary Journal* 25:21-40. doi: https://doi.org/10.1080/10705511.2017.1369088.
Shi, D., T. Lee and A. Maydeu-Olivares. 2019. "Understanding the Model Size Effect on Sem Fit Indices." *Educational Psychological Measurement* 79(2):310-34. doi: DOI: 10.1177/0013164418783530.
Srivastava, M. S., H. Yanagihara and T. Kubokawa. 2014. "Tests for Covariance Matrices in High Dimension with Less Sample Size." *Journal of Multivariate Analysis* 130:289-309.
Stein, C. 1956. "Some Problems in Multivariate Analysis: Part 1 (Report No. 6)." Vol. Stanford, CA.
Stein, C. 1975. "Estimation of a Covariance Matrix." Paper presented at the 39th A. Meet. Institute of Mathematical Statistics, Atlanta.
Tong, X. and P. M. Bentler. 2013. "Evaluation of a New Mean Scaled and Moment Adjusted Test Statistics for Sem." *Structural Equation Modeling: A Multidisplinary Journal* 20:148-56.
Yuan, K. H. and P. M. Bentler. 1998. "Normal Theory Based Test Statistics in Structural Equation Modeling." *British Journal of Mathematical and Statistical Psychology* 51:289-309.
Yuan, K. H. and P. Bentler. 1999. "On Asymptotic Distribution S of Normal Theory Mle in Covariance Structure Analysis under Some Nonnormal Distributions." *Statistics and probability Letters* (42):107-13.
Yuan, K. H. and P. Bentler. 2016. "Improving the Convergence Rate and Speed of Fisher-Scoring Algorithm: Ridge and Anti-Ridge Methods in Structural Equation Modeling." *Annals of the Institute of Statistical Mathematics* 69:571-97.
Zheng, B. Q. and P. M. Bentler. 2022. "Testing Mean and Covariance Structures with Reweighted Least Squares." *Structural Equation Modeling: A Multidisplinary Journal* 29(2):259-66. doi: https://doi.org/10.1080/10705511.2021.1977649.

Appendix

Table A1 provides the descriptive statistics of empirical kurtosis and skewness. A sample with *N*=10,000 is based on a simulation of a fixed covariance structure of the same population that is used in study. Table A1 contains three conditions: Normal, elliptical, and skewed cases. The kurtosis and skewness tests are conducted using R package "*semTools*" (Jorgensen et al. 2022). The measure of excessive kurtosis is computed by the fourth standardized moment of the empirical distribution of a variable, which is known as G2. The measure of skewness is computed by the third standardized moment of the empirical distribution of a variable, which is also known as G1.

Table A1
Univariate Kurtosis-Skewness Normality Tests Under Three Distributional Conditions

	Normal		Elliptical		Skewed	
Variable	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness
X1	0.002	0.041	3.206	-0.064	3.721	1.495
X2	-0.023	0.016	3.359	-0.031	4.085	1.713
X3	0.039	-0.011	3.789	0.055	3.094	1.385
X4	0.044	-0.014	4.786	-0.019	3.901	1.412
X5	-0.036	0.021	3.501	0.058	4.365	1.182
X6	-0.023	0.025	3.856	0.069	4.538	1.246
X7	-0.088	0.014	3.378	-0.020	3.278	1.049
X8	0.022	-0.046	3.740	-0.012	3.070	1.454
X9	0.081	0.000	4.573	0.035	2.881	1.041
X10	0.033	-0.028	2.404	-0.055	4.084	1.029
X11	-0.069	-0.013	4.841	-0.076	2.943	1.363
X12	-0.054	0.046	4.525	0.022	3.485	1.317
X13	-0.085	-0.025	4.706	-0.015	2.069	1.172
X14	0.015	0.004	5.537	0.053	3.803	1.404
X15	-0.001	0.034	4.899	0.047	2.978	1.299

Note: If a skewness or kurtosis is 0, the data are perfectly normally distributed; whereas a skewness or kurtosis is between -.5 and .5 indicates that the data are still approximately normal. A negative kurtosis indicates that the distribution has lighter tails than the normal distribution. If a skewness or kurtosis is between -1 and -.5 or between 1 and .5, the distribution is moderately kurtotic or skewed. If a skewness or kurtosis is less than -1 or greater than 1, the distribution is considered highly kurtotic or skewed (Hair et al. 2017).

Table A1 shows that in the normal distribution simulation, the kurtosis and skewness statistics of all simulated variables are close to zero, with the largest value of -0.088. For the elliptical distribution, the kurtosis statistics range from 2.404 to 5.537, meaning that all simulated variables are kurtotic. Skew under the elliptical distribution is near zero, with -.076 as the largest discrepancy from zero. In contrast, for the skewed distribution, the statistics of skewness range from 1.029 to 1.713; whereas it may be noted that skew also induces kurtosis, with values somewhat smaller than under the elliptical distribution.