STRUCTURAL EQUATION MODELING
2024, VOL. 31, NO. 6
ISSN: 1070-5511 | E-ISSN: 1532-8007

OPEN ACCESS

Estimating the Weight Matrix in Distributionally Weighted Least Squares Estimation: An Empirical Bayesian Solution

Han Du[a*] and Hao Wu[b*]

[a]University of California; [b]Vanderbilt University

ABSTRACT

Real data are unlikely to be exactly normally distributed. Ignoring non-normality will cause misleading and unreliable parameter estimates, standard error estimates, and model fit statistics. For non-normal data, researchers have proposed a distributionally-weighted least squares (*DLS*) estimator to combines the normal theory based generalized least squares estimation (GLS_N) and *WLS*. The key in *DLS* is to select an optimal weight a_s to compute a weighted average of GLS_N and *WLS*. To better estimate a_s in *DLS*, we propose a method based on the delta method and the empirical Bayesian method. When data were normal, *DLS* and GLS_N provided similar root mean square errors (RMSEs) and biases of the standard error estimates, and were smaller than those from *WLS*. When the data were elliptical or skewed, *DLS* generally provided the smallest RMSEs and biases of the standard error estimates. Additionally, the Type I error rates of Jiang-Yuan rank adjusted test statistic (T_{JY}) using *DLS* were generally around the nominal level.

KEYWORDS
Non-normal data; robust statistics; structural equation modeling

Structural Equation Modeling (SEM) is a commonly used statistical technique in social and behavioral research. Widely used estimation methods in SEM such as maximum likelihood (*ML*) and normal theory based generalized least squares estimation (GLS_N) assume that data are normally distributed, which is often not the case in reality (e.g., Cain et al., 2017). When data are not normally distributed, the consistent *ML* and GLS_N parameter estimates are biased for small and moderate sample sizes, and their standard errors and model fit statistics are generally incorrect. Although robust standard errors and rescaled-and-adjusted model fit test statistics have been proposed to correct for non-normality, they can still lead to inflated Type I error rates. (e.g., Jalal & Bentler, 2018; Satorra & Bentler, 1988; Yuan & Chan, 2016).

To relax the normality assumption, Browne (1984) proposed an asymptotically distribution-free (ADF) estimator for the covariance of sample variance-covariance. framework. The generalized least squares estimation that uses the ADF estimator is known as weighted least squares (*WLS*). The ADF estimator and *WLS* do not rely on any distribution assumption. *WLS* provides asymptotically the most efficient estimates but is not optimal for small and modest sample sizes as the ADF estimator needs to estimate more components (i.e., sample fourth-order moments) compared to methods that rely on the normality assumption. With a large number of variables or a small sample size, *WLS* may suffer from non-convergence. Moreover, *WLS* is unstable and has a large variance unless sample size is huge since it involves inverting a sample fourth-order moment matrix.

To address the issue of small sample sizes of *WLS* while retaining its desirable properties, Du and Bentler (2022a) recently introduced distributionally weighted least squares (*DLS*) estimation as a hybrid of normal theory based and ADF based generalized least squares methods. Their research on confirmatory factor analysis and latent growth curve modeling indicates that *DLS* performs better than *WLS* in terms of root mean square errors (RMSE), relative biases of the SE estimates, and model fit Type I error rates (Du & Bentler, 2022a; Du et al., 2022). Additionally, *DLS* outperforms normal theory-based estimators (GLS_N[1] and *ML*) when data are non-normal. By balancing information from the data and the normality assumption, *DLS* provides more accurate and efficient parameter estimates for SEM models.

DLS can be viewed as a compromise between GLS_N and *WLS* and an optimal weight, a, has to be chosen. The optimal a is chosen based on the efficiency and accuracy of parameter estimates (i.e., minimum RMSE), and is denoted as a_s. Currently, there are two ways to determine a_s. The first approach involves bootstrapping, in which the data are first transformed to fit the model of interest perfectly with the ML estimates as the model parameters. Then, Bootstrap samples are drawn from the transformed data to compute RMSE for varied values of a, and a_s is selected based on the minimum RMSE. In the end, we apply this optimal weight a_s to the original data. A problem of this approach is that we have to "create" true population parameters to compute RMSE. We can choose *ML* parameter estimates or other estimates, such as GLS_N estimates, but we have known that GLS_N and *ML* estimates are biased in small samples. Using biased estimates to create bootstrap samples may lead to biased estimates of a_s. Furthermore, since real data are never exactly normally distributed, bootstrapping from real data tend to be more liberal (giving more weight to WLS) than needed. A third shortcoming of this approach is its heavy computational load since all simulated bootstrap samples need to be fitted to the model. The second approach is creating a mapping function between a_s and all data/model features by an extensive simulation. In real data, this mapping function can be used to estimate a_s based on the data/model information. However, creating the mapping function is time-consuming and hard to consider all features in the simulation.

To solve the aforementioned problems in *DLS*, we propose an empirical Bayesian method to estimate a_s, in which the balance between GLS_N and *WLS* is automatically chosen without resampling. The outline of this paper is as follows: in the "*DLS* Estimator" section, we review the distributionally-weighted least squares (*DLS*) estimation. In the "a_s Estimation" section, we present the proposed algorithm of estimating a_s. In the "Simulation Study" section, the performance of the proposed a_s in *DLS* is thoroughly examined via simulations and compared with GLS_N and *WLS*. In the "Real Data Example" section, a real data example is provided to illustrate to estimate a_s in practice. We end the paper with some concluding remarks in the "Conclusion" section.

1. *DLS* Estimator

The distributionally-weighted least squares (*DLS*) estimator is within the generalized least squares (*GLS*) framework. Let $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, …, $\boldsymbol{x_N}$ be a multivariate random sample of size N from a p-variate population with $\mathrm{E}(\boldsymbol{x_i}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\boldsymbol{x_i}) = \boldsymbol{\Sigma}$ for $i = 1, \ldots, N$. Let vector $\boldsymbol{\theta}$ be a $q \times 1$ vector containing the free parameters in the structural equation model and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ be the population covariance matrix implied by the model. In practice, we can use the model implied covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ or the unstructured sample covariance matrix $\mathbf{S}$ to estimate the population covariance matrix $\boldsymbol{\Sigma}$, where

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})', \tag{1}$$

and $\bar{\boldsymbol{x}}$ is the sample mean. Note that $\Sigma(\theta)$ may or may not be correctly specified in practice. Let $\boldsymbol{s} = \mathrm{vech}(\mathbf{S})$ be a $p^* \times 1$ vector with the $p^* = p(p+1)/2$ non-duplicated elements in $\mathbf{S}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \mathrm{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]$ be a $p^* \times 1$ vector with the non-duplicated elements in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. By the multivariate central limit theorem, the asymptotic distribution of $\boldsymbol{s}$ is

$$\sqrt{N-1}(\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))\xrightarrow{D}\mathrm{N}(0, \boldsymbol{\Gamma}), \tag{2}$$

where $\boldsymbol{\Gamma}$ is the $p^* \times p^*$ asymptotic covariance matrix and D indicates convergence in distribution.

The general equation of generalized least squares (*GLS*) loss functions (Browne, 1974) is

$$F_{GLS}(\boldsymbol{\theta}) = [\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\hat{\mathbf{W}}[\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \tag{3}$$

where $\hat{\mathbf{W}}$ is a weight matrix of size $p^* \times p^*$ that can take various forms. The most ideal $\mathbf{W}$ is $\boldsymbol{\Gamma}^{-1}$, but $\boldsymbol{\Gamma}$ is unknown. In practice, different estimators adopt different $\hat{\mathbf{W}}$. For example, the least squares (LS) estimator uses $\hat{\mathbf{W}} = \mathbf{I}$. Although all weight matrices provide consistent estimates, some can be more efficient and less biased for finite sample sizes. The most widely used $\hat{\mathbf{W}}$ is the normal theory based estimator $\hat{\boldsymbol{\Gamma}}_N^{-1}$, and the estimated is referred as GLS_N. When the data are normal, $\boldsymbol{\Gamma}$ reduces to $\boldsymbol{\Gamma}_N$ as follows.

$$\{\boldsymbol{\Gamma}_N\}_{hj,kl} = \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk} \tag{4}$$

In samples, $\hat{\boldsymbol{\Gamma}}_N$ can be either estimated by sample covariances.

$$\{\hat{\boldsymbol{\Gamma}}_{N.S}\}_{hj,kl} = w_{hk}w_{jl} + w_{hl}w_{jk} \tag{5}$$

where $w_{hj} = \frac{1}{N}\Sigma_{i=1}^{N}(\{x_{h,i} - \bar{x}_h\}\{x_{j,i} - \bar{x}_j\})$ (or $s_{hk}s_{jl} + s_{hl}s_{jk}$ where $s_{hk} = \frac{1}{N-1}\Sigma_{i=1}^{N}(\{x_{h,i} - \bar{x}_h\}\{x_{k,i} - \bar{x}_k\})$) or iteratively updated by the model implied covariances $\{\hat{\boldsymbol{\Gamma}}_{N.M}\}_{hj,kl} = \hat{\sigma}_{hk}\hat{\sigma}_{jl} + \hat{\sigma}_{hl}\hat{\sigma}_{jk}$. Du and Bentler (2022a) and Du et al. (2022) found that $\hat{\boldsymbol{\Gamma}}_{N.M}$ performs better than $\hat{\boldsymbol{\Gamma}}_{N.S}$ in simulations of factor analysis and growth curve modeling. Since $\hat{\boldsymbol{\Gamma}}_{N.M}$ is iteratively updated during estimation, the method is also called the iteratively reweighted weighted least squares (*IRLS*) estimation by Lee and Jennrich (1979). It produces the same point estimate as *ML* but a different test statistic since *IRLS* and *ML* use different discrepancy functions. Hayakawa (2019) found that with the same parameter estimates, the GLS_N loss function can provide better model fit statistics than the *ML* discrepancy function. Hence, in this paper we focus on GLS_N (i.e., *IRLS*), which theoretically has the same estimates as *ML* but uses the better *GLS* loss function.

Considering that data may be non-normal, Browne (1984) proposed an asymptotically distribution-free (ADF) estimator for $\boldsymbol{\Gamma}$ without relying on any distribution assumption ($\hat{\boldsymbol{\Gamma}}_{ADF}$; Browne's Eq. 3.4).

$$\{\hat{\boldsymbol{\Gamma}}_{ADF}\}_{hj,kl} = w_{hjkl} - w_{hj}w_{kl}, \tag{6}$$

where $w_{hjkl} = \frac{1}{N}\Sigma_{i=1}^{N}(\{x_{h,i} - \bar{x}_h\}\{x_{j,i} - \bar{x}_j\}\{x_{k,i} - \bar{x}_k\}\{x_{l,i} - \bar{x}_l\})$ is the sample fourth order moment. With $\hat{\mathbf{W}} = \hat{\boldsymbol{\Gamma}}_{ADF}^{-1}$, the estimator is the so-called weighted least squares (*WLS*) estimator. Although the ADF estimator is asymptotically efficient, its performance is unstable with finite sample sizes and the empirical standard errors can be much greater than GLS_N even when data are nonnormal (Du et al., 2022; Du & Bentler, 2022a; Foss et al., 2011; Olsson et al., 2003; Yang & Yuan, 2019; Yuan & Bentler, 1997; Yuan & Chan, 2016).

DLS combines $\hat{\boldsymbol{\Gamma}}_N$ and $\hat{\boldsymbol{\Gamma}}_{ADF}$ to stabilize the performance from the ADF estimator and improve its efficiency. The weight function is

$$\hat{\mathbf{W}} = \left((1-a)\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}} + a\hat{\boldsymbol{\Gamma}}_{\mathbf{N}}\right)^{-1} \tag{7}$$

where $a \in [0, 1]$ is a tuning parameter. A larger a provides finite sample stability and a smaller a provides asymptotic efficiency. When $a = 1$, *DLS* becomes GLS_N, and when $a = 0$, *DLS* becomes *WLS*. Since the performance of *DLS* largely relies on a, it is critical to select the optimal a (denoted as a_s). In Du and Bentler (2022a) and Du et al. (2022), this optimal a is selected to minimize the root mean square error (RMSE) which considers both efficiency and accuracy in estimates. However, the calculation of RMSE requests knowing the population parameters. In practice, bootstrapping can be used to create multiple samples and calculate empirical RMSE to select a_s (Yuan & Chan, 2016; Yuan, Jiang, et al., 2017). More specifically, GLS_N or ML can be used to estimate the model and obtain the model implied covariance matrix $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$, based on which Bollen-Stine transformation can be used to obtain bootstrap samples (Bollen-Stine transformation ensures that the covariance matrix of the transformed data is $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$). Then, the tuning parameter a is varied from 0 to 1, and DLS is applied with different a values to each bootstrap sample. The value a that gives the smallest RMSE in the bootstrap samples is chosen as a_s. A dilemma is that we need to create bootstrap samples based on GLS_N or ML even though the purpose of DLS is to avoid solely relying on the normality assumption. In addition, this requires fitting the model 1000×101 times when there are 101 a values, which is a heavy computational burden.

Additionally, DLS uses sandwich type standard errors (SE) (Hardin, 2003; Huber, 1967; White, 1980; 1982), which is the square root of the diagonals of the sandwich covariance matrix,

$$\begin{aligned}&\mathrm{Cov}(\sqrt{N}\hat{\boldsymbol{\theta}}_{GLS})\\&\quad=\left(\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})'\hat{\mathbf{W}}\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})\right)^{-1}\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})'\hat{\mathbf{W}}\hat{\boldsymbol{\Gamma}}_{ADF}\hat{\mathbf{W}}\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})\left(\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})'\hat{\mathbf{W}}\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})\right)^{-1}\end{aligned} \tag{8}$$

where $\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})$ is the matrix of first-order derivative of $\boldsymbol{\sigma}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

2. An Empirical Bayesian Estimate of the Weight a_s

To solve the dilemma that we need to rely on normality theory based estimator to get DLS estimation and shorten computation time, we propose to estimate a_s using an empirical Bayesian method.

The population Γ in Equation (2) without any distributional assumption is $\{\Gamma\}_{hj,kl} = \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk} + \frac{N-1}{N}(\sigma_{hjkl} - \sigma_{hj}\sigma_{kl} - \sigma_{hk}\sigma_{jl} - \sigma_{hl}\sigma_{jk})$ (Browne, 1984; Eq. 2.1). The difference between $N-1$ and N disappears when the sample size tends to infinite. Hence, asymptotically $\{\Gamma\}_{hj,kl} = \sigma_{hjkl} - \sigma_{hj}\sigma_{kl}$. For normal data, the population Γ reduces to $\{\Gamma_N\}_{hj,kl} = \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk}$, which for non-normal data differs from the true population Γ. The difference between the two is the fourth order cumulant in Equation (9).

$$\begin{aligned}\kappa_{hjkl} &= \sigma_{hjkl} - \sigma_{hj}\sigma_{kl} - \sigma_{hk}\sigma_{jl} - \sigma_{hl}\sigma_{jk}\\ &= \{\Gamma\}_{hj,kl} - \{\Gamma_N\}_{hj,kl}\end{aligned} \tag{9}$$

For normal data, $\kappa_{hjkl} = 0$. In a sample, one can estimate $\{\Gamma\}_{hj,kl}$ as in Equation (6) by

$$\{\hat{\Gamma}_{ADF}\}_{hj,kl} = w_{hjkl} - w_{hj}w_{kl} \tag{10}$$

And one can estimate $\{\Gamma_{N.S}\}_{hj,kl}$ by $\{\hat{\Gamma}_{N.S}\}_{hj,kl} = w_{hk}w_{jl} + w_{hl}w_{jk}$ as in Equation (5).[2] The model based estimator $\{\hat{\Gamma}_{N.M}\}_{hj,kl}$ is not considered in estimating a_s because the analytical form will be too complex, but a_s obtained using $\hat{\Gamma}_{N.S}$ will be applied to DLS with $\hat{\Gamma}_{N.M}$ later in the estimation of the covariance structure model. Hence, based on Equations. (5) and (10), the estimate of Equation (9) is as follows.

$$\begin{aligned}\hat{\kappa}_{hjkl} &= w_{hjkl} - w_{hj}w_{kl} - w_{hk}w_{jl} - w_{hl}w_{jk}\\ &= \{\hat{\Gamma}_{ADF}\}_{hj,kl} - \{\hat{\Gamma}_{N.S}\}_{hj,kl}\end{aligned} \tag{11}$$

2.1. The Empirical Bayesian Model

We now propose an empirical Bayesian model for the observed cumulant $\hat{\kappa}_{hjkl}$. This model is similar in rationale to that in Chen (1979) and Wu and Browne (2015a, 2015b, 2016), which concerns empirical Bayesian modeling of a covariance matrix. Asymptotically, the vector of $p^{**} = p(p+1)(p+2)(p+3)/24$ estimated fourth order cumulants $\hat{\kappa}$ is normally distributed around its population value κ with a $p^{**} \times p^{**}$ covariance matrix Π, which leads to the likelihood function:

$$\hat{\kappa}|\kappa \sim \mathrm{N}(\kappa, \Pi) \tag{12}$$

To shrink it towards its value for normal data, the population cumulant vector κ is assumed to follow a multivariate normal prior distribution with a zero mean vector and a covariance matrix proportional to a normal theory based covariance matrix V. V is also a $p^{**} \times p^{**}$ matrix. The difference between V and Π is that Π does not have a distributional assumption while V assumes the data are normal. The scalar factor ν in Equation (13) determines the dispersion of this prior distribution and will be estimated from data.

$$\kappa \sim N(0, V\nu). \tag{13}$$

Based on Equations (12) and (13), the marginal distribution of $\hat{\kappa}$ is given by

$$\hat{\kappa} \sim N(0, \Pi + V\nu) \tag{14}$$

And the posterior distribution of the population cumulants in κ can be calculated as

$$\kappa|\hat{\kappa} \sim \mathrm{N}\left(\tilde{\kappa} = (\nu I + V^{-1}\Pi)^{-1}\nu\hat{\kappa}, \left(\Pi^{-1} + (V\nu)^{-1}\right)^{-1}\right) \tag{15}$$

where the posterior mean vector $\tilde{\kappa}$ is a shrinkage estimate of the population cumulant.

Our goal is to find the optimal a_s in $\hat{\mathbf{W}} = ((1-a)\hat{\mathbf{\Gamma}}_{ADF} + a\hat{\mathbf{\Gamma}}_N)^{-1}$ or $\hat{\mathbf{\Gamma}} = (1-a)\hat{\mathbf{\Gamma}}_{ADF} + a\hat{\mathbf{\Gamma}}_N$. Based on Equation (9), we have $\{\Gamma\}_{hj,kl} = \kappa_{hjkl} + \{\Gamma_N\}_{hj,kl}$, and we can estimate Γ using $\tilde{\kappa} + \hat{\Gamma}_{N.S}$:

$$\begin{aligned}\hat{\mathbf{\Gamma}} &= \tilde{\kappa} + \hat{\Gamma}_{N.S}\\ &= (\nu I + V^{-1}\Pi)^{-1}\nu\hat{\kappa} + \hat{\Gamma}_{N.S}\end{aligned} \tag{16}$$

$V^{-1}\Pi$ is not a scalar, making weight computation more complex. As one typically prefers a simple weighted average

[2]We can use unbiased estimation by $\{\hat{\Gamma}_{N.S}\}_{hj,kl} = s_{hk}s_{jl} + s_{hl}s_{jk}$, but it did not result in much better performance in our pilot simulation.

between $\hat{\Gamma}_{ADF}$ and $\hat{\Gamma}_{N.S}$, we approximate the matrix $V^{-1}\Pi$ by $\text{rank}(V^{-1}\Pi) \times I$ and approximate Equation (16) using Equation (17):

$$\hat{\Gamma} = \frac{\nu \hat{\Gamma}_{ADF} + \text{rank}\ (V^{-1}\Pi)\hat{\Gamma}_{N.S}}{\nu + \text{rank}\ (V^{-1}\Pi)} \tag{17}$$

where $\text{rank}\ (V^{-1}\Pi) = \text{rank}(\Pi) = \min(n-1, p^{**})$. By setting Equation (17) equal to Equation (7), we can solve the optimal tuning parameter as

$$a_s = \frac{\text{rank}\ (V^{-1}\Pi)}{\text{rank}\ (V^{-1}\Pi) + \nu} \tag{18}$$

2.2. The Estimation of ν

From Equation (14), we can create a quadratic form as follows:

$$Q = \hat{\kappa}'(\Pi + V\nu)^{-1}\hat{\kappa} \tag{19}$$

which satisfies $\text{E}(Q) = p^{**}$ where $p^{**} = p(p+1)(p+2)(p+3)/24$. Unfortunately, as the covariance matrix of the fourth order cumulants, an estimate of Π is likely unstable and even not positive definite. A simpler matrix with lower order moments can be employed to replace Π to stabilize the performance. We choose to use $V = \text{Cov}(\hat{\kappa})$ with the normality assumption, which leads to

$$Q_2 = \hat{\kappa}' V^{-1}\hat{\kappa} \tag{20}$$

Based on Scharf (1991), Q_2 is a weighted sum of χ^2 variates with $\text{E}(Q_2) = \text{tr}\{V^{-1}\Pi\} + \nu p^{**}$. Setting this expected value to the observed Q_2, we obtain the estimate of ν.

$$\hat{\nu} = \left(\hat{\kappa}' V^{-1}\hat{\kappa} - \text{tr}\{V^{-1}\Pi\}\right)/p^{**} \tag{21}$$

2.3. The Estimation of Π

Since $a_s = \frac{\text{rank}\ (V^{-1}\Pi)}{\text{rank}\ (V^{-1}\Pi)+\nu}$ and ν depends on Π and V (Equation 21), Π and V have to be estimated. The $p^{**} \times p^{**}$ matrix Π is the covariance matrix of the fourth order cumulants in $\hat{\kappa}$. We now use the delta method to calculate this matrix. Similar to the distribution-free estimation of the covariance matrix Γ of the sample covariances s_{ij}, we will construct new "variables" from the existing variables and calculate the sample covariance matrix of these new variables.

To demonstrate this process, we first calculate the covariance matrix of the first term of Equation (11): the fourth order central moment w_{hjkl}. It can be expressed as a function of raw first, second, third, and fourth order moments:

$$\begin{aligned} w_{hjkl} = {} & \overline{x_h x_j x_k x_l} - \overline{x_h} \times \overline{x_j x_k x_l} - \overline{x_j} \times \overline{x_h x_k x_l} - \overline{x_k} \times \overline{x_h x_j x_l} - \overline{x_l} \times \overline{x_h x_j x_k} \\ & + \overline{x_h} \times \overline{x_j} \times \overline{x_k x_l} + \overline{x_h} \times \overline{x_k} \times \overline{x_j x_l} + \overline{x_h} \times \overline{x_l} \times \overline{x_j x_k} \\ & + \overline{x_j} \times \overline{x_k} \times \overline{x_h x_l} + \overline{x_j} \times \overline{x_l} \times \overline{x_h x_k} + \overline{x_k} \times \overline{x_l} \times \overline{x_h x_j} \\ & - 3\overline{x_h} \times \overline{x_j} \times \overline{x_k} \times \overline{x_l} \end{aligned} \tag{22}$$

where $\overline{x_h x_j x_k x_l}$ is the sample average of $x_{h,i} x_{j,i} x_{k,i} x_{l,i}$, $\overline{x_j x_k x_l}$ is the sample average of $x_{j,i} x_{k,i} x_{l,i}$, and $\overline{x_k x_l}$ is the sample average of $x_{k,i} x_{l,i}$. Following the central limit theorem and assuming the eighth order central moments are finite, the aforementioned raw moments jointly follow a multivariate normal distribution in a large sample, so the covariance matrix of the fourth order central moments can be calculated through the delta method. More specifically, the differential of w_{hjkl} is given by

$$\begin{aligned} dw_{hjkl} = {} & d\overline{x_h x_j x_k x_l} \\ & - \left\{\overline{x_h} \times d\overline{x_j x_k x_l} + \overline{x_j} \times d\overline{x_h x_k x_l} + \overline{x_k} \times d\overline{x_h x_j x_l} + \overline{x_l} \times d\overline{x_h x_j x_k}\right\} \\ & - \left\{d(\overline{x_h}) \times \overline{x_j x_k x_l} + d(\overline{x_j}) \times \overline{x_h x_k x_l} + d(\overline{x_k}) \times \overline{x_h x_j x_l} + d(\overline{x_l}) \times \overline{x_h x_j x_k}\right\} \\ & + \left\{\begin{array}{l} \overline{x_h} \times \overline{x_j} \times d\overline{x_k x_l} + \overline{x_h} \times \overline{x_k} \times d\overline{x_j x_l} + \overline{x_h} \times \overline{x_l} \times d\overline{x_j x_k} \\ \quad + \overline{x_j} \times \overline{x_k} \times d\overline{x_h x_l} + \overline{x_j} \times \overline{x_l} \times d\overline{x_h x_k} + \overline{x_k} \times \overline{x_l} \times d\overline{x_h x_j} \end{array}\right\} \\ & + \left\{d(\overline{x_h}) \times \overline{x_j} \times \overline{x_k x_l} + \cdots + d(\overline{x_l}) \times \overline{x_k} \times \overline{x_h x_j}\right\} \\ & - 3d(\overline{x_h}) \times \overline{x_j} \times \overline{x_k} \times \overline{x_l} - 3\overline{x_h} \times d(\overline{x_j}) \times \overline{x_k} \times \overline{x_l} - 3\overline{x_h} \times \overline{x_j} \times d(\overline{x_k}) \times \overline{x_l} \\ & - 3\overline{x_h} \times \overline{x_j} \times \overline{x_k} \times d(\overline{x_l}) \end{aligned} \tag{23}$$

Replacing the differentials $d\overline{x_h x_j x_k x_l}$, $d\overline{x_j x_k x_l}$, $d\overline{x_k x_l}$ and $d\overline{x_l}$, respectively, by the random variables $x_h x_j x_k x_l$, $x_j x_k x_l$, $x_k x_l$ and x_l, and replacing the non-differential terms (i.e., the partial derivatives) by their population values, we obtain new "variables" $\tilde{y}_{hjkl}$:

$$\begin{aligned} \tilde{y}_{hjkl} = {} & x_h x_j x_k x_l - \left\{\mu_h \times (x_j x_k x_l) + \cdots + \mu_l \times (x_h x_j x_k)\right\} \\ & - \left\{x_h \times \mu_{jkl} + \cdots + x_l \times \mu_{hjk}\right\} \\ & + \left\{\mu_h \times \mu_j \times (x_k x_l) + \cdots + \mu_k \times \mu_l \times (x_h x_j)\right\} \\ & + \left\{x_h \times \mu_j \times \mu_{kl} + \cdots + \mu_k \times x_l \times \mu_{hj}\right\} \\ & - 3\left\{\begin{array}{l} x_h \times \mu_j \times \mu_k \times \mu_l + \mu_h \times x_j \times \mu_k \times \mu_l \\ \quad + \mu_h \times \mu_j \times x_k \times \mu_l + \mu_h \times \mu_j \times \mu_k \times x_l \end{array}\right\} \end{aligned} \tag{24}$$

where the μ's denote the population raw (i.e., uncentered) moments. Equation (24) is equivalent to y_{hjkl} below in Equation (25) plus a constant, and the constant does not change the y_{hjkl}'s variance.

$$\begin{aligned} y_{hjkl} = {} & \{x_h - \mu_h\}\{x_j - \mu_j\}\{x_k - \mu_k\}\{x_l - \mu_l\} \\ & - (x_h - \mu_h) \times \sigma_{jkl} - (x_j - \mu_j) \times \sigma_{hkl} \\ & - (x_k - \mu_k) \times \sigma_{hjl} - (x_l - \mu_l) \times \sigma_{hjk} \end{aligned} \tag{25}$$

where the σ's are the population central moments. The asymptotic covariance matrix of fourth order central moments w_{hjkl} can be estimated as the covariance matrix of y_{hjkl} divided by $N-1$.

The remaining three terms of $\hat{\kappa}_{hjkl}$ in Equation (11) $(-w_{hj}w_{kl} - w_{hk}w_{jl} - w_{hl}w_{jk})$ can be handled similarly using the delta method, yielding new variables z_{hjkl} :

$$\begin{aligned} z_{hjkl} = {} & -\sigma_{kl}\{x_h - \mu_h\}\{x_j - \mu_j\} \\ & - \sigma_{hj}\{x_k - \mu_k\}\{x_l - \mu_l\} - \sigma_{jl}\{x_h - \mu_h\}\{x_k - \mu_k\} \\ & - \sigma_{hk}\{x_l - \mu_l\}\{x_j - \mu_j\} - \sigma_{hl}\{x_j - \mu_j\}\{x_k - \mu_k\} \\ & - \sigma_{jk}\{x_l - \mu_l\}\{x_h - \mu_h\} \end{aligned} \tag{26}$$

The covariance matrix of z_{hjkl} divided by $N-1$ is an estimate of the asymptotic covariance matrix of $-w_{hj}w_{kl} - w_{hk}w_{jl} - w_{hl}w_{jk}$. Combining the two sets of new variables y_{hjkl} and z_{hjkl}, we create new variables t_{hjkl}:

$$\begin{aligned} t_{hjkl} & = y_{hjkl} + z_{hjkl} \\ & = \{x_h - \mu_h\}\{x_j - \mu_j\}\{x_k - \mu_k\}\{x_l - \mu_l\} \\ & \quad - \{x_h - \mu_h\} \times \sigma_{jkl} - \{x_j - \mu_j\} \times \sigma_{hkl} - \{x_k - \mu_k\} \times \sigma_{hjl} - \{x_l - \mu_l\} \times \sigma_{hjk} \\ & \quad - \sigma_{kl}\{x_h - \mu_h\}\{x_j - \mu_j\} - \sigma_{hj}\{x_k - \mu_k\}\{x_l - \mu_l\} - \sigma_{jl}\{x_h - \mu_h\}\{x_k - \mu_k\} \\ & \quad - \sigma_{hk}\{x_l - \mu_l\}\{x_j - \mu_j\} - \sigma_{hl}\{x_j - \mu_j\}\{x_k - \mu_k\} - \sigma_{jk}\{x_l - \mu_l\}\{x_h - \mu_h\} \end{aligned} \tag{27}$$

whose covariance matrix divided by $N-1$ yields the asymptotic covariance matrix of $\hat{\kappa}_{hjkl}$. To estimate the covariance matrix of t_{hjkl} in a sample, for each case i we compute the variables $\tilde{t}_{hjkl,i}$ below.

$$\begin{aligned}\tilde{t}_{hj,kl,i} = &\left(\tfrac{n}{n-1}\right)^4\{x_{h,i}-\bar{x}_h\}\{x_{j,i}-\bar{x}_j\}\{x_{k,i}-\bar{x}_k\}\{x_{l,i}-\bar{x}_l\}\\ &-\frac{n}{n-1}[x_{h,i}\times w_{jkl}+x_{j,i}\times w_{hkl}+x_{k,i}\times w_{hjl}+x_{l,i}\times w_{hjk}]\\ &-\left(\tfrac{n}{n-1}\right)^2[w_{kl}\{x_{h,i}-\bar{x}_h\}\{x_{j,i}-\bar{x}_j\}+w_{hj}\{x_{k,i}-\bar{x}_k\}\{x_{l,i}-\bar{x}_l\}\\ &+w_{jl}\{x_{h,i}-\bar{x}_h\}\{x_{k,i}-\bar{x}_k\}+w_{hk}\{x_{l,i}-\bar{x}_l\}\{x_{j,i}-\bar{x}_j\}\\ &+w_{hl}\{x_{j,i}-\bar{x}_j\}\{x_{k,i}-\bar{x}_k\}+w_{jk}\{x_{l,i}-\bar{x}_l\}\{x_{h,i}-\bar{x}_h\}]\end{aligned} \tag{28}$$

where $w_{jkl}=\frac{1}{N}\sum_{i=1}^{N}(\{x_{j,i}-\bar{x}_j\}\{x_{k,i}-\bar{x}_k\}\{x_{l,i}-\bar{x}_l\})$. Note that $\frac{n}{n-1}\{x_{h,i}-\bar{x}_h\}=x_{h,i}-\bar{x}_{h,-i}$, where $\bar{x}_{h,-i}$ is the average of all observations except case i. Denote their sample covariance matrix by S_t. Then $\frac{S_t}{N-1}$ can be used to estimate Π.

A problem of $\Pi=S_t/(N-1)$ is that it relies on large sample size approximation and does not perform well in small samples. Based on the James-Stein estimator (James & Stein, 1992), a small sample size adjustment $\frac{1}{(N-3)}$ was used. We adopt this idea and provide an alternative way to compute Π as

$$\Pi=\frac{S_t}{N-3} \tag{29}$$

Based on our pilot simulation [3], $\frac{S_t}{N-3}$ performed much better than $\frac{S_t}{N-1}$ in small samples, and hence we focus on $\frac{S_t}{N-3}$ in the current paper.

2.4. The Estimation of V

We choose to specify V as $V=\mathrm{Cov}(\hat{\kappa})$ with the normality assumption. We have derived new variables T_{hjkl} for estimating Π without any distributional assumption in the previous section. Following the same idea, now assuming normality of the observed data, we can derive the V matrix. Since with normal data, the third order moments are zero, t_{hjkl} reduces to

$$\begin{aligned}r_{hjkl} &= \{x_h-\mu_h\}\{x_j-\mu_j\}\{x_k-\mu_k\}\{x_l-\mu_l\}\\ &-\sigma_{kl}\{x_h-\mu_h\}\{x_j-\mu_j\}-\sigma_{hj}\{x_k-\mu_k\}\{x_l-\mu_l\}-\sigma_{jl}\{x_h-\mu_h\}\{x_k-\mu_k\}\\ &-\sigma_{hk}\{x_l-\mu_l\}\{x_j-\mu_j\}-\sigma_{hl}\{x_j-\mu_j\}\{x_k-\mu_k\}-\sigma_{jk}\{x_l-\mu_l\}\{x_h-\mu_h\}\end{aligned} \tag{30}$$

The covariance of r_{hjkl} and r_{mnpq} is $\mathrm{Cov}(r_{hjkl},r_{mnpq})=\mathrm{E}(r_{hjkl}\times r_{mnpq})$-$\mathrm{E}(r_{hjkl})\times\mathrm{E}(r_{mnpq})$. The expected value in the second term is given by

$$\begin{aligned}\mathrm{E}(r_{hjkl}) &= \sigma_{hjkl}-2(\sigma_{kl}\sigma_{hj}+\sigma_{jl}\sigma_{hk}+\sigma_{jk}\sigma_{hl})\\ &= -(\sigma_{kl}\sigma_{hj}+\sigma_{jl}\sigma_{hk}+\sigma_{jk}\sigma_{hl})\end{aligned} \tag{31}$$

The first term $\mathrm{E}(r_{hjkl}\times r_{mnpq})$ involves the 8th, 6th and 4th order central moments, which for a multivariate normal distribution can all be reduced to products of population covariances. After some tedious derivation (see the Appendix), $\mathrm{Cov}(r_{hjkl},r_{mnpq})$ can be simplified to

$$\mathrm{Cov}(r_{hjkl},r_{mnpq})=\sum_{h',j',k',l'}\sigma_{hh'}\sigma_{jj'}\sigma_{kk'}\sigma_{ll'} \tag{32}$$

where h', j', k', and l' can be any combination of m, n, p, and q (e.g., $\sigma_{hm}\sigma_{jn}\sigma_{kp}\sigma_{lq}$). Hence, the summation in Equation (32) is taken over all 24 permutations of (m,n,p,q). From the sample, V can be estimated by replacing the population covariances by sample covariances.

$$[V]_{hjkl,mnpq}=\sum_{h',j',k',l'}s_{hh'}s_{jj'}s_{kk'}s_{ll'}/(N-1) \tag{33}$$

where the form of the denominator is not crucial as it will cancel out from the numerator and denominator of Equation (21).

2.5. Summary of the Proposed Method

Overall, the proposed method contains 4 steps.

1. Create new variables $\tilde{t}_{hj,kl,i}$ to compute their covariance matrix S_t and $\Pi=S_t/(N-3)$ (Eq. 29).
2. Compute V (Eq. 33).
3. Based on V and Π to estimate ν by $\hat{\nu}=(\hat{\kappa}'V^{-1}\hat{\kappa}-tr\{V^{-1}\Pi\})/p^{**}$ (Eq. 21).
4. The optimal a is computed based on the estimated ν, $a_s=\frac{\mathrm{rank}(V^{-1}\Pi)}{\mathrm{rank}(V^{-1}\Pi)+\nu}$ (Eq. 18).

3. Simulation Study

3.1. Simulation Design

This simulation study examines the performance of the proposed method for estimating the optimal tuning parameter (a_s) in *DLS*. For computing $\hat{\mathbf{W}}=((1-a)\hat{\mathbf{\Gamma}}_{\mathbf{ADF}}+a\hat{\mathbf{\Gamma}}_{\mathbf{N}})^{-1}$, since $\hat{\mathbf{\Gamma}}_{N.M}$ performs better than $\hat{\mathbf{\Gamma}}_{N.S}$ in Du and Bentler (2022a) and Du et al. (2022), we compute the weight matrix as $\hat{\mathbf{W}}=((1-a)\hat{\mathbf{\Gamma}}_{\mathbf{ADF}}+a\hat{\mathbf{\Gamma}}_{\mathrm{N.M}})^{-1}$ in the current paper. But in estimating a_s, we still use $\hat{\mathbf{\Gamma}}_{N.S}$ for the sake of computational convenience.

We varied the values of the following four factors: the total number of variables ($p=5$, 15, and 21), the number of factors ($m=1$ and 3), the sample size (N ranging from 40 to 1000), and the distributional conditions (a normal distribution, an elliptical distribution, and a skewed distribution). There are 19 sample size conditions (Table 1). We used a confirmatory factor analysis (CFA) model to simulate data.

$$\boldsymbol{x}=\mathbf{\Lambda}\boldsymbol{\xi}+\boldsymbol{\varepsilon}, \tag{34}$$

where $\mathbf{\Lambda}$ is a $p\times m$ vector of factor loadings, $\boldsymbol{\xi}$ is a $m\times 1$ vector of factor scores, and $\boldsymbol{\varepsilon}$ is a $p\times 1$ vector of independent measurement errors for p variables. Let $\mathbf{\Phi}=cov(\boldsymbol{\xi})$ and $\mathbf{\Psi}=cov(\boldsymbol{\varepsilon})$. The population covariance matrix of x is $\mathbf{\Sigma}=\mathbf{\Lambda\Phi\Lambda}'+\mathbf{\Psi}$. Each factor had the same number of free nonzero loadings, which were randomly sampled from .70 to .95 with an interval of .05. The correlations/covariances between the factors were specified as 0.50 and the variances of the factors were specified as 1. $\mathbf{\Psi}$ was calculated to ensure that the diagonal elements of $\mathbf{\Sigma}$ were 1. In the normal distribution condition, $\xi=\mathbf{\Phi}^{1/2}Z_\xi$ and $\varepsilon=\mathbf{\Psi}^{1/2}Z_\varepsilon$ where $\mathbf{\Phi}^{1/2}\mathbf{\Phi}^{1/2}=\mathbf{\Phi}$, $\mathbf{\Psi}^{1/2}\mathbf{\Psi}^{1/2}=\mathbf{\Psi}$, and both Z_ξ and Z_ε followed a standard normal distribution $N(0,1)$. Following the simulation strategy in Yuan and Chan (2016), we simulated elliptical and skewed data. In the elliptical distribution condition (symmetric distributions with heavy tails), $\xi=r\mathbf{\Phi}^{1/2}Z_\xi$ and $\varepsilon=r\mathbf{\Psi}^{1/2}Z_\varepsilon$ with

[3]We tried multiple options in the pilot simulations. For example, $\frac{S_t}{N-3}$ vs. $\frac{S_t}{N-1}$, and Equation (11) vs. the use of the unbiased estimator of κ_{hjkl} as given by Equation (3.5) in Browne (1984).

Table 1. Simulation conditions of N, p, and m.

19 Conditions of N, p, and m		
m	p	N
1	5	40, 60, 100, 200, 300, 500, 1000
3	15	40, 60, 100, 200, 300, 500, 1000
3	21	100, 200, 300, 500, 1000

4 Distributional Conditions			
	Z_ξ	Z_ε	r
Normal	N (0, 1)	N (0, 1)	–
Elliptical	N (0, 1)	N (0, 1)	$(3/\chi^2_5)^{1/2}$
Skewed	*standardized* (χ^2_1)	*standardized* (χ^2_1)	$(3/\chi^2_5)^{1/2}$

$r \sim (3/\chi^2_5)^{1/2}$. In the skewed distribution condition, $\xi = r\boldsymbol{\Phi}^{1/2}Z_\xi$ and $\varepsilon = r\boldsymbol{\Psi}^{1/2}Z_\varepsilon$ where $Z_\xi \sim standardized(\chi^2_1)$ and $Z_\varepsilon \sim standardized(\chi^2_1)$. In addition, we generated nonnormal data using the third-order polynomial transformations (Vale & Maurelli, 1983). Following Falk (2018), we considered two levels of nonnormality: univariate skewness = 2 and kurtosis = 7 (VM1), or univariate skewness = 2 and kurtosis = 15 (VM2). We use *semTools* (Jorgensen et al., 2021) to implement Vale and Maurelli's method. For each condition, we simulated 1000 datasets. We conducted the simulation with R (version 3.6.1). The R code of for the proposed method is provided on https://github.com/hduquant/DLS-with-the-delta-method.git.

We compared 3 estimators in the simulation: weighted least squares estimation (*WLS*), normal theory based *GLS* (GLS_N or *IRLS* since $\hat{\boldsymbol{\Gamma}}_{N.M}$ is iteratively updated during estimation), and the distributionally-weighted least squares (*DLS*) with the proposed algorithm in testing a_s. These 3 estimators were compared in terms of the root mean square error (RMSE), the relative biases of the SE estimates, and Type I error rates. The root mean square error (RMSE) is widely used to investigate both the efficiency and accuracy of parameter estimates (e.g., Du et al., 2022; Du & Bentler, 2022a; Yuan, Yang, et al., 2017; Yuan & Chan, 2016). Let $\hat{\theta}_{ij}$ be the estimate of the ith parameter in the jth replication and θ_i be the true value for the ith parameter. For each condition, the relative RMSE was averaged over all parameters.

$$RMSE = \frac{1}{q}\sum_{i=1}^{i=q}\frac{1}{\theta_i}\left(\frac{1}{1000}\sum_{j=1}^{j=1000}\left(\hat{\theta}_{ij}-\theta_i\right)^2\right)^{1/2} \tag{35}$$

The relative biases of the SE estimates is to investigate the performance of the SE estimates. The true SE is unknown, so we calculated the standard deviation for each parameter estimate across the converged replications as the empirical SE of each parameter. Let $\widehat{SE}_{ij}$ be the SE estimate of ith parameter in the jth replication and SE_i be the empirical SE of ith parameter. The relative biases of the SE estimates were averaged over all parameters.

$$Relative\ Bias = \frac{1}{q}\sum_{i=1}^{i=q}\left(\left|\frac{\left(\frac{1}{1000}\sum_{j=1}^{j=1000}\sum\widehat{SE}_{ij}\right)-SE_i}{SE_i}\right|\right)$$

Besides RMSE and relative biases of the SE estimates, a good estimator should provide reliable model fit inference. The standard model fit statistic under the normality assumption is $T = (N-1)F(\hat{\boldsymbol{\theta}})$, where $F(\hat{\boldsymbol{\theta}})$ is the loss function in Equation (3). T asymptotically follows χ^2_{df} where $df = p^* - q$. When the data are not normal, the asymptotic distribution of T is not χ^2_{df}. Hence, researchers have proposed adjusted test statistics so the adjusted statistics can approximate a χ^2 reference distribution. The most widely used adjusted statistics is the Satorra–Bentler statistic T_{SB} (Satorra & Bentler, 1986, 1988, 1994) that $T_{SB} = T/c_{SB}$. $c_{SB} = \text{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}})/df$, and $\hat{\mathbf{U}} = \hat{\mathbf{W}} - \hat{\mathbf{W}}\dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}})(\dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}}))^{-1}\dot{\boldsymbol{\sigma}}(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}$. T_{SB} is referred to χ^2_{df} as its reference distribution. In practice, the rank of $\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}}$ can be smaller than *df*. Hence, Jiang and Yuan (2017) proposed to estimate the average eigenvalues of $\mathbf{U\Gamma}$ by replacing *df* with $\text{rank}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}})$. $T_{JY} = T/c_{JY}$ with $c_{JY} = \text{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}})/\text{rank}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}})$. T_{JY} is referred to $\chi_{\text{rank}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}})}{}^2$ as its reference distribution (Du & Bentler, 2022b; Du, 2023). When $\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}_{\mathbf{ADF}}$ is not rank-deficient (i.e., the sample size is not too small), T_{SB} and T_{JY} are essentially the same. We will use T_{JY} in the simulation. Besides T_{JY}, we also presented T_{SB} and the mean-and-variance adjusted test statistic T_{MVA} (Satorra & Bentler, 1988) in the https://github.com/hduquant/DLS-with-the-delta-method.git, given that they performed worse than T_{JY} in our simulation.

3.2. Structure of Simulation Results

The convergence rates of GLS_N and *DLS*'s were above 0.98, whereas *WLS* did not converge at all with large p and small N. We kept only the converged solutions among the 1000 replications. In the following sections, we first present the estimated a_s in *DLS* from the proposed method. Second, we compare RMSEs of parameter estimates and biases of SE estimates across *WLS*, GLS_N, and *DLS*. We present all detailed results in the https://github.com/hduquant/DLS-with-the-delta-method.git.

Estimated a_s in DLS

The estimated a_s in *DLS* under difference condition is summarized in Table 2. When the data are normal, we expect that a_s is 1 or close to 1 so we can rely on the normality assumption (e.g., $\hat{\boldsymbol{\Gamma}}_{N.M}$) and have more stable estimation. The simulation results were consistent with our anticipation. Across replications, a_s based on the proposed method was very close to 1 or at 1.

When data are non-normal, we expect that a_s lies in the middle between 0 and 1 so the estimation does not solely rely on the normality assumption. But we also do not expect a_s to be too close to 0 since $\hat{\boldsymbol{\Gamma}}_{ADF}$ tends to be unstable. With the elliptical, skewed, VM1, and VM2 data, the minimums, medians, and means of a_s deviated more from 1 compared to normal data. With the same p, a larger sample size generally decreased a_s. It implies that when the sample size was small, *DLS* tended to rely more on $\hat{\boldsymbol{\Gamma}}_{N.M}$ to stabilize the performance. But with a larger sample size, *DLS* could weight more on $\hat{\Gamma}_{ADF}$ to capture the non-normality feature in the data. While the difference between a_s and 1 was not substantial when $p = 15$ and 21, this distance played a crucial role in reducing RMSE and biases of SE estimates, and improving Type I error rates. We will illustrate it in the upcoming sections.

3.3. RMSE Comparisons

We compared RMSE from *WLS*, GLS_N, and *DLS* to explore both their efficiency and accuracy. The plots of RMSEs from *WLS*, GLS_N, and *DLS* with normal, elliptical, skewed, VM1, and VM2 data are in Figure 1. For all three estimators, when N increased, the RMSEs decreased, and the difference between estimators became smaller. When data were normal, the RMSEs of GLS_N and *DLS* were almost the same since a_s in *DLS* was very close to 1. And the RMSEs of GLS_N and *DLS* were smaller than those from *WLS*.

When data were non-normal, *DLS* generally yielded smaller RMSE than GLS_N and *WLS*. When data were elliptical, *DLS* demonstrated a reduction in RMSE ranging from 9.1% to 17.7% compared to GLS_N. When data were skewed, *DLS* reduced RMSE ranging from 10.3% to 32.1% compared to GLS_N. When data were VM1, the superiority of *DLS* was not significantly pronounced: *DLS* reduced RMSE ranging from 1.4% to 7.2% compared to GLS_N. When data were VM2, *DLS* reduced RMSE ranging from 0.9% to 19.6% compared to GLS_N.

3.4. Relative Biases of SE Estimates across Methods

We compared *WLS*, GLS_N, and *DLS* in terms of their relative biases of the SE estimates in Figure 2. With normal data, the relative biases of GLS_N and *DLS* were almost the same since a_s under this condition was close to 1. The relative biases of *WLS* were larger than the ones from GLS_N and *DLS*.

When data were non-normal, *DLS* generally had the smallest biases of SE estimates and the biases were much smaller than the ones from *WLS*. When data were elliptical, *DLS* demonstrated a reduction in the biases of SE estimates ranging from 18.8% to 67.3% compared to GLS_N. When data were skewed, *DLS* reduced the biases of SE estimates

Table 2. The estimated a_s in *DLS* under difference conditions.

			Normal				Elliptical				Skewed			
p	*m*	*N*	min	mean	median	max	min	mean	median	max	min	mean	median	max
5	1	40	0.983	0.998	0.999	1	0.605	0.975	0.983	1	0.455	0.931	0.944	0.992
5	1	60	0.991	0.999	0.999	1	0.731	0.978	0.986	1	0.412	0.903	0.923	0.995
5	1	100	0.991	0.999	0.999	1	0.661	0.967	0.979	1	0.065	0.804	0.835	0.975
5	1	200	0.989	0.999	1	1	0.29	0.919	0.94	1	0.038	0.561	0.585	0.912
5	1	300	0.99	0.999	1	1	0.073	0.85	0.884	0.991	0.032	0.43	0.436	0.797
5	1	500	0.989	0.999	1	1	0.034	0.741	0.775	0.961	0.016	0.252	0.249	0.572
5	1	1000	0.992	0.999	1	1	0.089	0.529	0.545	0.86	0.007	0.103	0.097	0.286
15	3	40	0.997	0.998	0.999	1	0.959	0.994	0.995	0.998	0.965	0.991	0.992	0.997
15	3	60	0.998	0.999	0.999	1	0.986	0.995	0.995	0.999	0.962	0.991	0.991	0.997
15	3	100	0.999	1	1	1	0.982	0.995	0.996	0.999	0.974	0.989	0.99	0.997
15	3	200	0.999	1	1	1	0.98	0.995	0.996	0.999	0.961	0.985	0.986	0.996
15	3	300	1	1	1	1	0.964	0.995	0.996	0.999	0.945	0.98	0.981	0.993
15	3	500	1	1	1	1	0.923	0.994	0.995	0.999	0.921	0.971	0.973	0.99
15	3	1000	1	1	1	1	0.901	0.991	0.993	0.997	0.704	0.951	0.956	0.982
21	3	100	0.999	1	1	1	0.992	0.997	0.997	0.999	0.99	0.995	0.995	0.998
21	3	200	1	1	1	1	0.991	0.997	0.998	0.999	0.984	0.993	0.994	0.997
21	3	300	1	1	1	1	0.989	0.997	0.998	0.999	0.978	0.992	0.992	0.997
21	3	500	1	1	1	1	0.986	0.997	0.998	0.999	0.969	0.988	0.989	0.994
21	3	1000	1	1	1	1	0.962	0.996	0.996	0.999	0.944	0.981	0.981	0.99

			VM1				VM2			
p	*m*	*N*	min	mean	median	max	min	mean	median	max
5	1	40	0.791	0.966	0.972	1	0.68	0.933	0.94	1
5	1	60	0.794	0.958	0.966	1	0.707	0.923	0.933	0.994
5	1	100	0.699	0.923	0.933	0.993	0.583	0.865	0.878	0.984
5	1	200	0.446	0.812	0.826	0.96	0.347	0.655	0.664	0.909
5	1	300	0.324	0.704	0.716	0.933	0.153	0.461	0.456	0.776
5	1	500	0.206	0.512	0.513	0.796	0.086	0.259	0.254	0.582
5	1	1000	0.116	0.294	0.293	0.546	0.026	0.105	0.102	0.244
15	3	40	0.986	0.994	0.994	0.998	0.823	0.991	0.991	0.997
15	3	60	0.985	0.994	0.994	0.998	0.983	0.992	0.992	0.997
15	3	100	0.985	0.994	0.994	0.998	0.976	0.991	0.991	0.998
15	3	200	0.983	0.994	0.994	0.998	0.962	0.987	0.987	0.994
15	3	300	0.981	0.993	0.993	0.997	0.946	0.981	0.982	0.994
15	3	500	0.977	0.99	0.991	0.996	0.937	0.972	0.973	0.988
15	3	1000	0.963	0.987	0.988	0.994	0.904	0.956	0.958	0.978
21	3	100	0.993	0.996	0.997	0.999	0.991	0.995	0.995	0.998
21	3	200	0.994	0.997	0.997	0.998	0.988	0.994	0.994	0.997
21	3	300	0.992	0.997	0.997	0.999	0.983	0.992	0.993	0.997
21	3	500	0.992	0.996	0.996	0.998	0.98	0.989	0.989	0.995
21	3	1000	0.99	0.994	0.994	0.997	0.96	0.983	0.983	0.992

a_s is the optimal a in $(1-a)\hat{\Gamma}_{\mathbf{ADF}}+a\hat{\Gamma}_{\mathbf{N}}$.

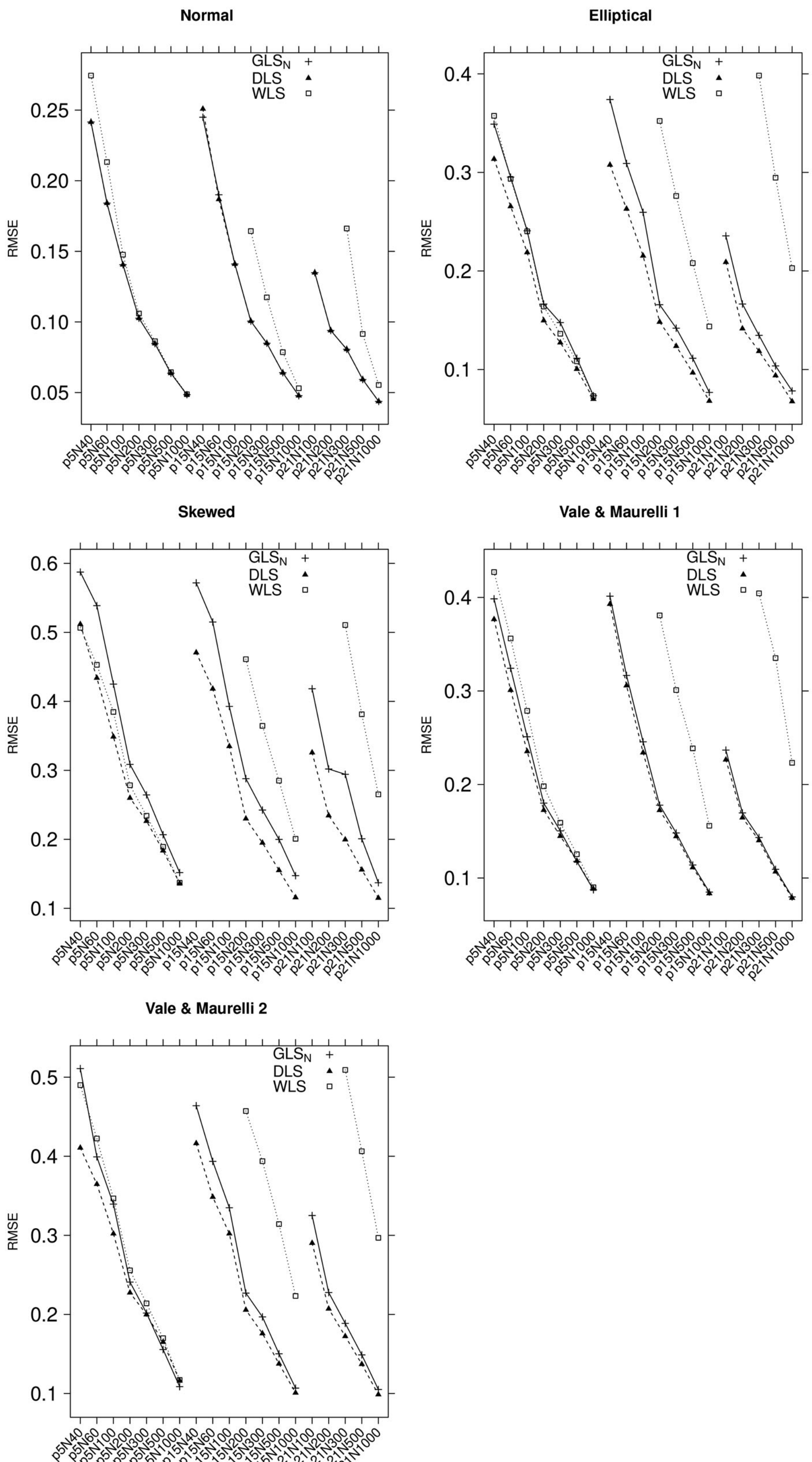

Figure 1. RMSEs From *WLS*, GLS_N, and *DLS*.

ranging from 2.4% to 63.3% compared to GLS_N. When data were VM1, the superiority of *DLS* was not obvious: *DLS* reduced the biases of SE estimates ranging from 1.4% to 9.6% compared to GLS_N. And GLS_N could have smaller biases of SE estimates in some cases with $p = 5$. When data were VM2, *DLS* reduced the biases of SE estimates ranging from 4.9% to 19% compared to GLS_N, but when $p = 5$, GLS_N could have smaller biases.

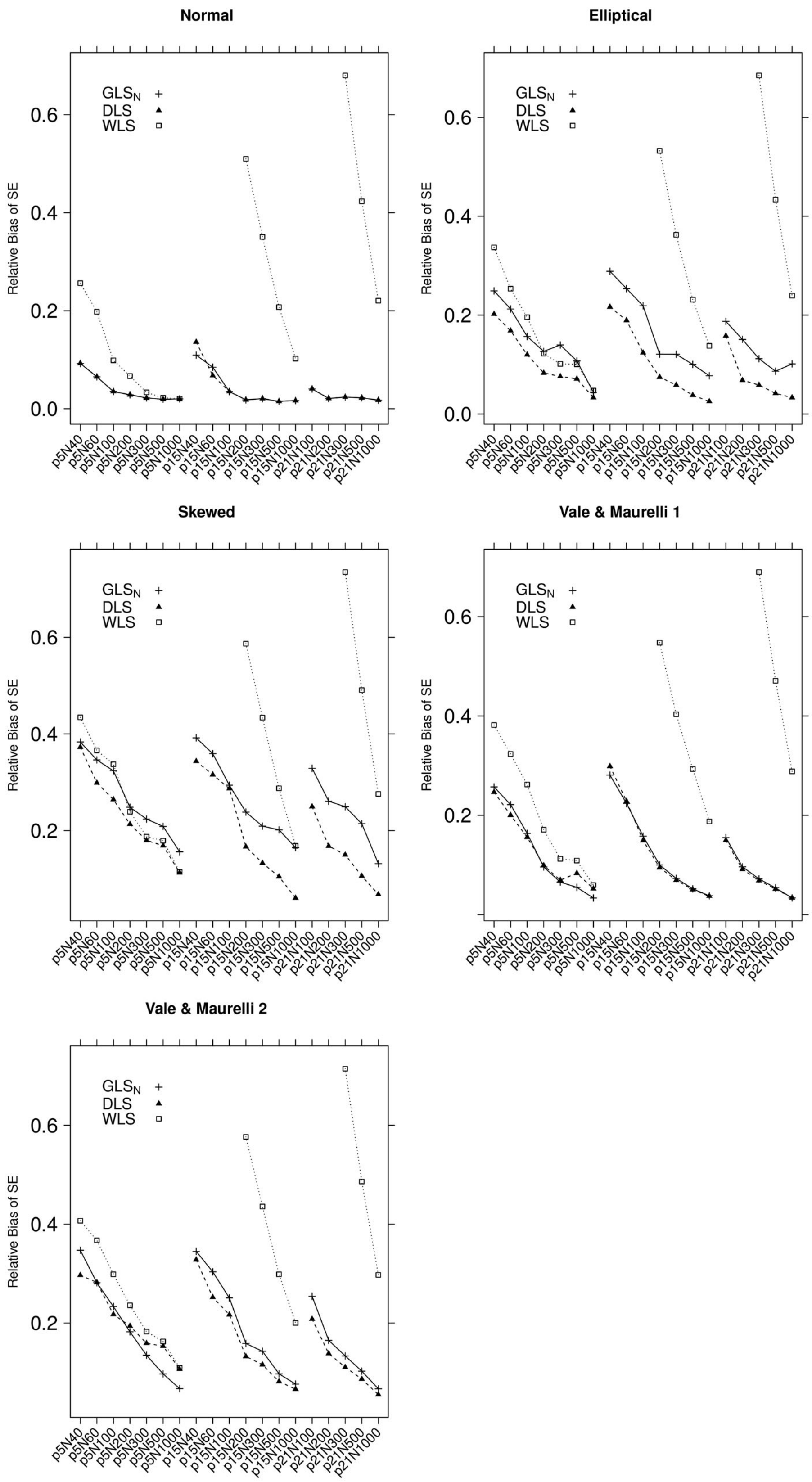

Figure 2. Relative biases of SE estimates from *WLS*, GLS_N, and *DLS*.

3.5. Type I Error Rates

We examined the Type I error rates of the Jiang-Yuan rank adjusted test statistic (T_{JY}) from *WLS*, GLS_N, and *DLS* (see Figure 3). When the data were normal, both GLS_N and *DLS* with T_{JY} produced Type I error rates close to 0.05, while *WLS* had substantial inflated Type I error rates (almost 1) except when $p = 5$. With non-normal data, *DLS* generally provided Type I error rates closer to 0.05 than GLS_N and *WLS*. When $p = 5$, *WLS* could provided Type I error rates

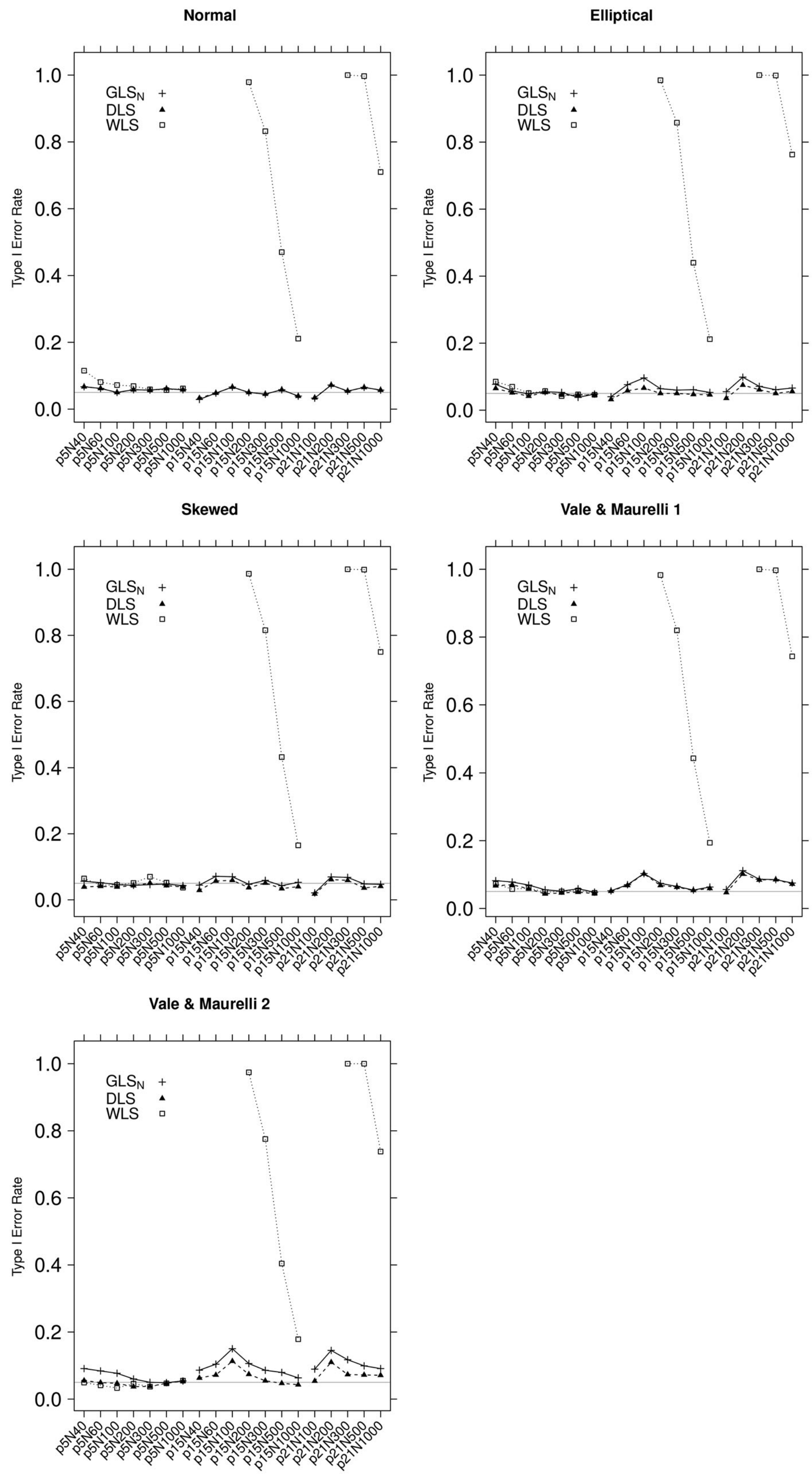

Figure 3. Type I error rates from *WLS*, GLS_N, and *DLS*.

closer to 0.05 than *DLS* and GLS_N. But in other cases, *WLS*'s Type I error rates were substantially inflated.

We also examined the Type I error rates of the standard model fit statistic *T*, Satorra–Bentler statistic T_{SB}, and the mean-and-variance adjusted test statistic T_{MVA} (see https://github.com/hduquant/DLS-with-the-delta-method.git for detailed information). *T* produced inflated Type I error rates with non-normal data. T_{SB} was different from T_{JY} only when the sample size was small. Hence, in the majority examined cases, T_{SB} was equivalent as T_{JY}. With small sample size, T_{SB} produced inflated Type I error rate regardless of the distribution. The Type I error rates from T_{MVA} were too small and almost 0 in many cases.

4. An Artificial Data Example

In this section, we illustrate the difference between GLS_N (i.e., *IRLS*) and *DLS* in a public dataset, which is available in the R package, *lavaan (version 0.6-5)* (Rosseel, 2012). The original dataset has mental ability test scores of 26 tests for the 7th and 8th grade children from two different schools (Pasteur and Grant-White; Holzinger & Swineford, 1939). The Grant-White dataset has 145 students. There are three dimensions/factors: spatial ability, verbal ability, and ability related to speed. The 9th variable is the speeded discrimination of straight and curved capitals. This variable measures both a spatial ability and an ability related to speed, therefore it has loadings on both factors. The factor model is Equation (34) with $\mathbf{\Psi} = diag(\psi_{11}, \psi_{22}, \psi_{33}, \psi_{44}, \psi_{55}, \psi_{66}, \psi_{77}, \psi_{88}, \psi_{99})$,

$$\mathbf{\Phi} = \begin{pmatrix} 1 & \phi_{12} & \phi_{13} \\ \phi_{12} & 1 & \phi_{23} \\ \phi_{13} & \phi_{23} & 1 \end{pmatrix},$$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & 0 & 0 & 0 & 0 & 0 & \lambda_{19} \\ 0 & 0 & 0 & \lambda_{24} & \lambda_{25} & \lambda_{26} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{37} & \lambda_{38} & \lambda_{39} \end{pmatrix}'.$$

The data and model were used in Du and Bentler (2022a) to demonstrate *DLS* estimation with bootstrap. With bootstrap, a_s was estimated to be 0.75 in Du and Bentler (2022a). As mentioned earlier, there are three concerns of bootstrap procedures. First, it is relatively time consuming. Second, we have to rely on biased estimates to create bootstrap samples, which will also influence the estimation of a_s. And third, the resultant a_s tends to be more liberal. Using the current proposed delta method, a_s was estimate to be 0.9995, so it heavily relied on $\hat{\mathbf{\Gamma}}_{N.M}$. As a result, the estimates and inferences from GLS_N and *DLS* were similar. It also implies the proposed method is more conservative than the bootstrap procedure that it tends to exclude $\hat{\mathbf{\Gamma}}_{ADF}$ unless the data substantively deviate from normality.

To better illustrate the difference between GLS_N and *DLS*, we add non-normality to the data (the artificial data are on https://github.com/hduquant/DLS-with-the-delta-method.git). Based on $\hat{\boldsymbol{\theta}}_{ML}$ in the original data, we use the method of simulating the skewed distribution in the simulation to simulate a new set of data with 45 participants. We calculate the multivariate skewness and kurtosis by Yuan et al. (2017) with $\frac{1}{Np(p+1)(p+2)}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(\mathbf{x_i}-\bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x_j}-\bar{\mathbf{x}})\right]^3$ and $\frac{1}{Np(p+2)}\sum_{i=1}^{N}\left[(\mathbf{x_i}-\bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x_i}-\bar{\mathbf{x}})\right]^2$, respectively. The population multivariate skewness and kurtosis should be 1 in the normal case. This artificial dataset has a multivariate kurtosis value of 1.80 and a multivariate skewness value of 5.06.

With this artificial dataset, a_s was estimate to be 0.993. In terms of model fit, since $\hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\mathbf{ADF}}$ is not rank-deficient, T_{SB} and T_{JY} are the same in this dataset. Since we simulated data from the true model, we expect a good model fit. *DLS* with T_{SB} showed a good model fit ($T_{SB} = 24.23,\ p = 0.391$) whereas GLS_N with T_{SB} showed a poor model fit ($T_{SB} = 36.26,\ p = .039$). We present parameter estimates, the SE estimates, and the z scores of GLS_N and *DLS* in Table 3. Additionally, we also present $\hat{\boldsymbol{\theta}}_{ML}$ in the original data as a reference since we simulated data with treating $\hat{\boldsymbol{\theta}}_{ML}$ as the true values. Although a_s was close to 1, GLS_N and *DLS* showed huge differences on multiple coefficients (e.g., λ_{24}, λ_{25}, λ_{26}, ψ_{77}, ψ_{88}). Given $n = 45$, we did not expect that either GLS_N or *DLS* could provide estimates close to $\hat{\boldsymbol{\theta}}_{ML}$ in the original data. But *DLS* still provided estimates closer to $\hat{\boldsymbol{\theta}}_{ML}$ compared to GLS_N. For example, λ_{26} was estimated to be 0.83 in *DLS* and 2.049 in GLS_N, while $\hat{\boldsymbol{\theta}}_{ML}$ was 0.934. We also computed RMSE across these 22 coefficients with treating $ML_{E.M}$ in the original data as the true parameters. *DLS* had RMSE of 0.28, and GLS_N had RMSE of 0.76. Hence, the estimates from *DLS* were closer to $ML_{E.M}$.

Table 3. Artificial data example.

	$ML_{E.M}$ in the original data			GLS_N			*DLS*		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
λ_{11}	0.817	0.099	8.263	0.565	0.154	3.679	0.599	0.109	5.482
λ_{12}	0.541	0.1	5.426	1.108	0.499	2.223	0.162	0.084	1.929
λ_{13}	0.686	0.09	7.642	1.245	0.58	2.148	0.429	0.078	5.496
λ_{19}	0.458	0.089	5.126	0.237	0.104	2.277	0.021	0.186	0.111
λ_{24}	0.972	0.078	12.383	1.756	0.547	3.207	1.026	0.194	5.297
λ_{25}	0.96	0.083	11.631	1.77	0.559	3.165	0.981	0.168	5.835
λ_{26}	0.934	0.081	11.553	2.049	0.813	2.522	0.83	0.135	6.136
λ_{37}	0.705	0.09	7.853	0.31	0.369	0.84	0.617	0.118	5.234
λ_{38}	0.897	0.093	9.598	1.781	1.91	0.933	0.884	0.142	6.226
λ_{39}	0.45	0.09	5.032	0.075	0.127	0.594	0.381	0.13	2.935
ψ_{11}	0.652	0.117	5.551	0.35	0.076	4.617	0.126	0.085	1.471
ψ_{22}	0.933	0.122	7.634	0.677	0.321	2.109	0.538	0.17	3.165
ψ_{33}	0.603	0.096	6.254	0.559	0.205	2.731	0.349	0.111	3.136
ψ_{44}	0.313	0.065	6.4	0.048	0.044	1.082	0.023	0.012	1.904
ψ_{55}	0.419	0.072	4.847	0.099	0.056	1.781	0.1	0.032	3.083
ψ_{66}	0.408	0.069	5.824	0.169	0.062	2.723	0.023	0.016	1.454
ψ_{77}	0.565	0.096	5.913	0.633	0.31	2.043	0.292	0.104	2.812
ψ_{88}	0.289	0.118	5.865	−2.443	6.858	−0.356	−0.077	0.115	−0.671
ψ_{99}	0.476	0.065	2.448	1.055	0.734	1.437	0.667	0.439	1.517
ϕ_{12}	0.554	0.081	6.86	0.944	0.051	18.575	0.534	0.121	4.394
ϕ_{13}	0.393	0.103	3.804	0.189	0.265	0.714	0.645	0.167	3.864
ϕ_{23}	0.239	0.095	2.511	0.158	0.218	0.727	0.525	0.162	3.252

5. Summary and Conclusion

Real data are unlikely to be exactly normally distributed. Ignoring non-normality will cause misleading and unreliable parameter estimates, standard error estimates, and model fit statistics. For non-normal data, researchers have proposed a generalized least squares estimation called the weighted least squares (*WLS*) based on the asymptotically distribution free (ADF) estimator. However, *WLS*'s advantage does not exist with small and modest sample sizes To improve *WLS*, Du and Bentler (2022a) and Du et al. (2022) proposed a distributionally-weighted least squares (*DLS*) estimator to combines the normal theory based generalized least squares estimation (GLS_N) and *WLS*. Their simulations on confirmatory factor analysis and latent growth curve modeling demonstrated that *DLS* performs better than *WLS* and GLS_N (and *ML*) in terms of root mean square errors (RMSE), relative biases of the SE estimates, and model fit Type I error rates when the data were non-normal. The key in *DLS* is to select an optimal weight a_s to compute a weighted average of GLS_N and *WLS*. Du and Bentler (2022a) and Du et al. (2022) used bootstrapping to estimate a_s (a gave the minimum RMSE). There are two concerns of this bootstrap method. First, in real data, we need to treat some estimates such as *ML* parameter estimates as the true parameters to compute RMSE, but all estimators can be biased in small samples with nonnormal data. Another concern is that bootstrapping is time consuming.

To better estimate a_s in *DLS*, we propose a method based on the delta method and the empirical Bayesian method. In the simulation, we found that the selection of a_s was based on sample size, model complexity, and distribution. When the data were normal, a_s was 1 or close to 1. When the data were elliptical or skewed, a_s deviated from 1. A larger sample size generally decreased a_s and increased the weight of $\hat{\boldsymbol{\Gamma}}_{ADF}$. In the real data example, we compared the estimated a_s with bootstrapping in Du and Bentler (2022a) and the estimated a_s from the current proposed algorithm, it seems that the proposed algorithm tends to be more conservative than the bootstrap procedure regarding weighting on $\hat{\boldsymbol{\Gamma}}_{ADF}$.

When data were normal, since a_s was close to 1, *DLS* and GLS_N provided similar RMSEs and biases of the SE estimates, and were smaller than those from *WLS*. When the data were non-normal, *DLS* generally provided the smallest RMSEs and biases of the SE estimates. Additionally, the Type I error rates of Jiang-Yuan rank adjusted test statistic (T_{JY}) using *DLS* were generally around the nominal level (0.05), which were closer to 0.05 than those from GLS_N and *WLS*.

We did not compare *DLS* using the bootstrap procedure and using the proposed algorithm in the simulation. The main reason is simulation time. Suppose we construct 1000 bootstrap samples for each replication and repeat for 1000 replications, and we also vary the tuning parameter a from 0 to 1 with an equal interval of .01 (i.e., 0, 0.01, … , 0.99, 1). Just for one condition, there are 10^8 analyses. Hence, it is not feasible to conduct the simulation using the bootstrap procedure. Furthermore, while Du et al. (2022) has expanded *DLS* to accommodate models with mean structures, the currently proposed algorithm for estimating a_s requires adjustment to be suitable for models with mean structures. Additionally, it does not yet address missing data or categorical data.

In sum, our paper provides an alternative way to compute the distributionally-weighted least squares estimator. The algorithm is much quicker than the original bootstrap procedure in Du and Bentler (2022a) and Du et al. (2022). With this algorithm, *DLS* still provides more accurate and efficient estimates than GLS_N and *WLS*.

References

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1–24. https://doi.org/10.1002/j.2333-8504.1973.tb00197.x

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37*(Pt 1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*, 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Chen, C.-F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*, 235–248. https://doi.org/10.1111/j.2517-6161.1979.tb01078.x

Du, H. (2023). Extended unbiased distribution free estimator with mean structures. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*, 412–428. https://doi.org/10.1080/10705511.2022.2135515

Du, H., & Bentler, P. M. (2022a). Distributionally weighted least squares in structural equation modeling. *Psychological Methods*, *27*, 519–540. https://doi.org/10.1037/met0000388

Du, H., & Bentler, P. M. (2022b). 40-Year old unbiased distribution free estimator reliably improves SEM statistics for nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 872–887.

Du, H., Bentler, P. M., & Rosseel, Y. (2022). Distributionally-weighted least squares in growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 1–22. https://doi.org/10.1080/10705511.2021.1931870

Falk, C. F. (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 244–266. https://doi.org/10.1080/10705511.2017.1367254

Foss, T., Jöreskog, K. G., & Olsson, U. H. (2011). Testing structural equation models: The effect of kurtosis. *Computational Statistics & Data Analysis*, *55*, 2263–2275. https://doi.org/10.1016/j.csda.2011.01.012

Hardin, J. W. (2003). The sandwich estimate of variance. In T. B. Fomby & R. C. Hill (Eds.), *Maximum likelihood estimation of misspecified models: Twenty years later* (pp. 45–73). Elsevier. https://doi.org/10.1016/s0731-9053(03)17003-x

Hayakawa, K. (2019). Corrected goodness-of-fit test in covariance structure analysis. *Psychological Methods*, *24*, 371–389. https://doi.org/10.1037/met0000180

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* (supplementary educational monograph no. 48). University of Chicago Press.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. Lecam & J. Neyman (Eds.), *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 221–233). University of California Press.

Jalal, S., & Bentler, P. M. (2018). Using monte carlo normal distributions to evaluate structural models with nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 541–557. https://doi.org/10.1080/10705511.2017.1390753

James, W., & Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in statistics: Foundations and basic theory* (pp. 443–460). Springer.

Jiang, G., & Yuan, K.-H. (2017). Four new corrected statistics for SEM with small samples and nonnormally distributed data. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 479–494. https://doi.org/10.1080/10705511.2016.1277726

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). semTools: Useful tools for structural equation modeling [Computer software manual]. https://CRAN.R-project.org/package=semTools (R package version 0.5-5)

Lee, S.-Y., & Jennrich, R. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, *44*, 99–113. https://doi.org/10.1007/BF02293789

Olsson, U. H., Foss, T., & Troye, S. V. (2003). Does the adf fit function decrease when the kurtosis increases? *The British Journal of Mathematical and Statistical Psychology*, *56*, 289–303. https://doi.org/10.1348/000711003770480057

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. version 0.5–12 (beta). *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. In *American Statistical Association: Proceedings of the Business and Economic Statistics Section* (pp. 549–554). American Statistical Association.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *American Statistical Association 1988: Proceedings of Business and Economics Sections*. (pp. 308–313). American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 339–419). Sage.

Scharf, L. L. (1991). *Statistical signal processing: detection, estimation, and time series analysis*. Pearson.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471. https://doi.org/10.1007/BF02293687

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25. https://doi.org/10.2307/1912526

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838. https://doi.org/10.2307/1912934

Wu, H., & Browne, M. W. (2015a). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*, 571–600. https://doi.org/10.1007/s11336-015-9451-3

Wu, H., & Browne, M. W. (2015b). Random model discrepancy: Interpretations and technicalities (a rejoinder). *Psychometrika*, *80*, 619–624. https://doi.org/10.1007/s11336-015-9456-y

Wu, H., & Browne, M. W. (2016). Erratum to: Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *81*, 1168–1168. https://doi.org/10.1007/s11336-016-9542-9

Yang, M., & Yuan, K.-H. (2019). Optimizing ridge generalized least squares for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 24–38. https://doi.org/10.1080/10705511.2018.1479853

Yuan, K.-H. Bentler, & P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, *26*, 177–198. https://doi.org/10.1016/S0167-9473(97)00025-X

Yuan, K.-H., & Chan, W. (2016). Structural equation modeling with unknown population distributions: Ridge generalized least squares. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 163–179. https://doi.org/10.1080/10705511.2015.1077335

Yuan, K.-H., Jiang, G., & Cheng, Y. (2017). More efficient parameter estimates for factor analysis of ordinal variables by ridge generalized least squares. *The British Journal of Mathematical and Statistical Psychology*, *70*, 525–564. https://doi.org/10.1111/bmsp.12098

Yuan, K.-H., Yang, M., & Jiang, G. (2017). Empirically corrected rescaled statistics for SEM with small N and large p. *Multivariate Behavioral Research*, *52*, 673–698.https://doi.org/10.1080/00273171.2017.1354759

Appendix

Derivation of the V Matrix

$$[V]_{hjkl,\,mnpq} = \frac{\mathrm{Cov}(R_{hjkl}, R_{mnpq})}{N-1} = \frac{\mathrm{E}(R_{hjkl} \times R_{mnpq}) - \mathrm{E}(R_{hjkl}) \times \mathrm{E}(R_{mnpq})}{N-1}$$

The first term $\mathrm{E}(R_{hjkl} \times R_{mnpq})$ in the V matrix involves the 8th, 6th and 4th central moments.

$$\mathrm{E}(R_{hjkl} \times R_{mnpq}) = \sigma_{hjklmnpq} - \sum_{m',n',p',q'} \sigma_{m'n'}\sigma_{mnpqp'q'} - \sum_{h',j',k',l'} \sigma_{h'j'}\sigma_{hjklk'l'} + \sum_{h',j',k',l',m',n',p',q'} \sigma_{h'j'}\sigma_{m'n'}\sigma_{k'l'p'q'}$$

where $\sigma_{hjklmnpq}$ is the 8th central moments, $\sigma_{mnpqp'q'}$ and $\sigma_{hjklk'l'}$ are the 6th central moments, and σ_{mnpq} and σ_{hjkl} are the 4th central moments. In $\sum_{m',n',p',q'} \sigma_{m'n'}\sigma_{mnpqp'q'}$, m', n', p', and q' can be any combination of h, j, k, and l (e.g., $\sigma_{hj}\sigma_{mnpqkl'}$), and in $\sum_{h',j',k',l'} \sigma_{h'j'}\sigma_{hjklk'l'}$, h', j', k', and l' can be any combination of m, n, p, and q (e.g., $\sigma_{mn}\sigma_{hjklpq}$). Hence, each of $\sum_{m',n',p',q'} \sigma_{m'n'}\sigma_{mnpqp'q'}$ and $\sum_{h',j',k',l'} \sigma_{h'j'}\sigma_{hjklk'l'}$ has 6 terms. In $\sum_{h',j',k',l',m',n',p',q'} \sigma_{h'j'}\sigma_{m'n'}\sigma_{k'l'p'q'}$, h', j', k', and l' can be any combination of h, j, k, and l, and m', n', p', and q' can be any combination of m, n, p, and q (e.g., $\sigma_{hl}\sigma_{np}\sigma_{jkmq}$); in total, it has 36 terms. With a multivariate normal distribution, $\mathrm{E}(R_{hjkl} \times R_{mnpq})$ can all be reduced to products of population covariances.

$$\begin{aligned}\mathrm{E}(R_{hjkl} \times R_{mnpq}) = & \sum_{h',j',k',l'} \sigma_{hh'}\sigma_{jj'}\sigma_{kk'}\sigma_{ll'} \\ & +\sigma_{hj}\sigma_{kl}\sigma_{mn}\sigma_{pq} + \sigma_{hk}\sigma_{jl}\sigma_{mn}\sigma_{pq} + \sigma_{hl}\sigma_{jk}\sigma_{mn}\sigma_{pq} \\ & +\sigma_{hj}\sigma_{kl}\sigma_{mp}\sigma_{nq} + \sigma_{hk}\sigma_{jl}\sigma_{mp}\sigma_{nq} + \sigma_{hl}\sigma_{jk}\sigma_{mp}\sigma_{nq} \\ & +\sigma_{hj}\sigma_{kl}\sigma_{mq}\sigma_{np} + \sigma_{hk}\sigma_{jl}\sigma_{mq}\sigma_{np} + \sigma_{hl}\sigma_{jk}\sigma_{mq}\sigma_{np}\end{aligned}$$

where h', j', k', and l' can be any combination of m, n, p, and q (e.g., $\sigma_{hm}\sigma_{jn}\sigma_{kp}\sigma_{lq}$) with a total 24 combinations. $\mathrm{E}(R_{hjkl}) \times \mathrm{E}(R_{mnpq})$ is equivalent to the later 9 terms in $\mathrm{E}(R_{hjkl} \times R_{mnpq})$.

$$\begin{aligned}\mathrm{E}(R_{hjkl}) \times \mathrm{E}(R_{mnpq}) = & \sigma_{hj}\sigma_{kl}\sigma_{mn}\sigma_{pq} + \sigma_{hk}\sigma_{jl}\sigma_{mn}\sigma_{pq} + \sigma_{hl}\sigma_{jk}\sigma_{mn}\sigma_{pq} \\ & +\sigma_{hj}\sigma_{kl}\sigma_{mp}\sigma_{nq} + \sigma_{hk}\sigma_{jl}\sigma_{mp}\sigma_{nq} + \sigma_{hl}\sigma_{jk}\sigma_{mp}\sigma_{nq} \\ & +\sigma_{hj}\sigma_{kl}\sigma_{mq}\sigma_{np} + \sigma_{hk}\sigma_{jl}\sigma_{mq}\sigma_{np} + \sigma_{hl}\sigma_{jk}\sigma_{mq}\sigma_{np}\end{aligned}$$

Hence, $[V]_{hjkl,\,mnpq} = \sum_{h',j',k',l'} \sigma_{hh'}\sigma_{jj'}\sigma_{kk'}\sigma_{ll'}/(N-1)$.