

Running head: DISTRIBUTIONALLY-WEIGHTED LEAST SQUARES IN STRUCTURAL EQUATION MODELING

Distributionally-Weighted Least Squares in Structural Equation Modeling

Han Du and Peter M. Bentler

Department of Psychology, University of California, Los Angeles

Correspondence should be addressed to Han Du, Pritzker Hall, 502 Portola Plaza, Los Angeles, CA 90095.

Email: [hdu@psych.ucla.edu](mailto:hdu@psych.ucla.edu).

This paper has been presented at the 85th annual Meeting of the Psychometric Society and the 2020

Meeting of the International Society for Data Science and Analytics

©American Psychological Association, 2021. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: <https://doi.org/10.1037/met0000388>

**Abstract**

In real data analysis with structural equation modeling, data are unlikely to be exactly normally distributed. If we ignore the non-normality reality, the parameter estimates, standard error estimates, and model fit statistics from normal theory based methods such as maximum likelihood (ML) and normal theory based generalized least squares estimation (GLS) are unreliable. On the other hand, the asymptotically distribution free (ADF) estimator does not rely on any distribution assumption but cannot demonstrate its efficiency advantage with small and modest sample sizes. The methods which adopt misspecified loss functions including ridge GLS (RGLS) can provide better estimates and inferences than the normal theory based methods and the ADF estimator in some cases. We propose a distributionally-weighted least squares (DLS) estimator, and expect that it can perform better than the existing generalized least squares, because it combines normal theory based and ADF based generalized least squares estimation. Computer simulation results suggest that model-implied covariance based DLS ( $DLS_M$ ) provided relatively accurate and efficient estimates in terms of RMSE. In addition, the empirical standard errors, the relative biases of standard error estimates, and the Type I error rates of the Jiang-Yuan rank adjusted model fit test statistic ( $T_{JY}$ ) in  $DLS_M$  were competitive with the classical methods including ML, GLS, and RGLS. The performance of  $DLS_M$  depends on its tuning parameter  $a$ . We illustrate how to implement  $DLS_M$  and select the optimal  $a$  by a bootstrap procedure in a real data example.

## Distributionally-Weighted Least Squares in Structural Equation Modeling

Structural equation modeling (SEM) is widely used in social and behavioral research, but its statistical methodology remains marginally capable of dealing with empirical data encountered in many psychological and behavioral studies. First, statistics in SEM rely on large sample size approximation. That is, their use relies on asymptotic properties as sample size  $N$  becomes extremely large ( $N \rightarrow \infty$ ). However in real data analysis, sample sizes are usually moderate or even small. Although SEM methods usually provide consistent parameter estimates and consistent standard error (SE) estimates, the estimates are not necessarily unbiased with finite sample size. Second, although data are typically nonnormally distributed (e.g., Cain et al., 2017), the mainstream estimators for SEM are still based on normal theory, such as maximum likelihood (ML) and normal theory based generalized least squares estimation (GLS). With nonnormally distributed data, ML and GLS still provide consistent parameter estimates, however their standard errors and model fit statistics generally are incorrect. Even when robust standard errors and rescaled-and-adjusted model fit test statistics are used to correct for non-normality, their performance in terms of efficiency and Type I error rates with finite sample sizes has been proved inadequate in a vast literature (e.g., Jalal & Bentler, 2018; Satorra & Bentler, 1988; Yuan & Chan, 2016).<sup>1</sup>

To relax the normality assumption, Browne (1984) proposed an asymptotically distribution free (ADF) estimator for nonnormal data within the generalized least squares framework. This method is sometimes called weighted least squares (WLS). The ADF estimator adopts a completely distribution-free estimate of the asymptotic covariance matrix of sample covariances (we will expand on this later) and provides the most asymptotically efficient estimates. However, the efficiency advantage of the ADF estimator cannot be realized with small and modest sample sizes. When sample sizes are rather small, there may be serious convergence problems especially when the number of variables is large, because the ADF estimator needs to estimate more components (i.e., sample fourth-order moments) compared to the methods that rely on the normality assumption. When sample sizes are modest, the ADF estimator is

unstable because it involves inverting a sample fourth-order moment matrix. As a consequence, the sampling distribution of ADF estimates has a large variance and thus standard errors are large compared to the normal theory based methods. Some researcher have directly used a (multivariate) t-distribution instead of a (multivariate) normal distribution to handle data with long tails and tolerate outliers (Song et al., 2007; Tong & Zhang, 2012, 2020; Z. Zhang et al., 2013). In this way, the influence of outliers and distributional deviation is down-weighted. Rather than using normal distributions, some researchers used a mixture of distributions which is empirically determined by the data (Lee & Xia, 2006; Muthén & Shedden, 1999).

Besides ML, GLS, and WLS, there are other estimators within the generalized least squares estimation framework, especially, least squares (LS), diagonally weighted least squares (DWLS), as well as ridge GLS estimation. We can classify the existing methods into three categories. In the first category, the methods including ML and GLS rely on a normality assumption. When the distribution is nonnormal, statistics inference is not correct. In the second category, WLS makes no distribution assumption. It enjoys the asymptotic efficiency of ADF but requires a large sample size to be stable. In the third category, the methods including LS, DWLS, and ridge GLS use a misspecified loss function. A misspecified loss function still provides consistent parameter estimates. In some contexts, such as with ordinal variables, these estimates can be more accurate and efficient than normal theory based methods (e.g., Li, 2016; Yuan et al., 2019).

We propose a new generalized least squares method, distributionally-weighted least squares (DLS) estimation. DLS estimation is a combination of normal theory based and ADF based generalized least squares estimation, in which the weight matrix of the loss function is the inverse of a combination of the ADF based and normal theory based estimators of the covariance matrix of sample covariances ( $cov(\mathbf{s})$ ). DLS has three advantages. First, with finite sample sizes, it yields more efficient estimates than those from the ADF estimator. Second, DLS is partially normal theory dependent, which helps to stabilize the performance of DLS compared to WLS. Third, it balances the information from the data and the normality

assumption. Our context is that of the typical situation in which sample size substantially exceeds the number of variables. For the contrary cases, see Deng et al. (2018) and Yuan et al. (2019).

The outline of this paper is as follows: in the “Estimators in SEM” section, an overview is given of some widely used estimators in SEM and estimators that are related to the current paper. In the “Distributionally-Weighted Least Squares” section, we present the proposed distributionally-weighted least squares (DLS) estimation. In the “Model Fit Evaluation” section, we introduce the model fit statistics that we will explore in the simulation. In the “Simulation Study” section, the performance of DLS is thoroughly examined via simulations. In the “Real Data Example” section, a real data example is provided to illustrate the implementation of DLS with bootstrapping in practice. We end the paper with some concluding remarks in the “Conclusion” section.

### Estimators in SEM

We briefly introduce some widely used estimators and some estimators that are related to our proposed distributionally-weighted least squares (DLS) method, including maximum likelihood (ML), generalized least squares (GLS), least squares (LS), weighted least squares (WLS), and ridge GLS methods. Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  be a multivariate random sample of size  $N$  from a  $p$ -variate population with  $E(\mathbf{x}_i) = \boldsymbol{\mu}$  and  $cov(\mathbf{x}_i) = \boldsymbol{\Sigma}$  for  $i = 1, \dots, N$ . Let a vector  $\boldsymbol{\theta}$  be a  $q \times 1$  vector containing free parameters in the structural equation model. The population covariance is assumed to be a function of  $\boldsymbol{\theta}$ , therefore  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  is proposed to fit the data. The population covariance matrix is unknown. In a sample, we can calculate the model implied covariance matrix  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$  or the sample covariance matrix  $\mathbf{S}$  to estimate the population covariance matrix  $\boldsymbol{\Sigma}$ ,

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

where  $\bar{\mathbf{x}}$  is the sample mean.

Let  $\mathbf{s} = \text{vech}(\mathbf{S})$  be a  $p^* \times 1$  vector with the  $p^* = p * (p + 1)/2$  nonduplicated elements in  $\mathbf{S}$  and  $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \text{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]$  be a  $p^* \times 1$  vector with the nonduplicated elements in  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ . By the multivariate central limit theorem,  $\mathbf{s}$  converges in distribution to a normal distribution:

$$\sqrt{N}(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})) \xrightarrow{D} N(0, \boldsymbol{\Gamma}), \quad (2)$$

where  $\boldsymbol{\Gamma}$  is the  $p^* \times p^*$  asymptotic covariance matrix of  $\sqrt{N}(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))$ . An element of  $\boldsymbol{\Gamma}$  is

$$\{\boldsymbol{\Gamma}\}_{hj,kl} = \text{cov}(\sqrt{N}(\mathbf{S}_{hj} - \boldsymbol{\Sigma}_{hj}), \sqrt{N}(\mathbf{S}_{kl} - \boldsymbol{\Sigma}_{kl})) = \sigma_{hjdkl} - \sigma_{hj}\sigma_{kl} \text{ with}$$

$$\sigma_{hjdkl} = E(\{x_h - E(x_h)\}\{x_j - E(x_j)\}\{x_k - E(x_k)\}\{x_l - E(x_l)\}) \text{ and}$$

$$\sigma_{hj} = E(\{x_h - E(x_h)\}\{x_j - E(x_j)\}) = \boldsymbol{\Sigma}_{hj} \text{ (Browne, 1984)}. \text{ In practice, the ADF estimator uses a}$$

consistent estimator of  $\boldsymbol{\Gamma}$  with elements

$$\{\hat{\boldsymbol{\Gamma}}_{ADF}\}_{hj,kl} = s_{hjdkl} - s_{hj}s_{kl}, \quad (3)$$

where  $s_{hjdkl} = \frac{1}{N} \sum_{i=1}^N (\{x_{h,i} - \bar{x}_h\}\{x_{j,i} - \bar{x}_j\}\{x_{k,i} - \bar{x}_k\}\{x_{l,i} - \bar{x}_l\})$  is the sample fourth moment and

$$s_{hj} = \frac{1}{N} \sum_{i=1}^N (\{x_{h,i} - \bar{x}_h\}\{x_{j,i} - \bar{x}_j\}) \text{ is an element of } \mathbf{S}.$$

### Maximum Likelihood Estimation

Maximum likelihood (ML) estimation minimizes a function which measures the discrepancy between  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  and  $\mathbf{S}$  (Browne, 1974, 1984; Jöreskog, 1967, 1969),

$$F_{ML}(\boldsymbol{\theta}) = \text{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] - \log|\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})| - p. \quad (4)$$

As a variant of ML, Yuan and Chan (2008) proposed a ridge ML estimation by replacing  $\mathbf{S}$  in the discrepancy function with  $\mathbf{S}_a = \mathbf{S} + a\mathbf{I}$ , where  $a \in [0, \infty)$  is a ridge tuning parameter and  $\mathbf{I}$  is a  $p \times p$  identity matrix. Yuan and Chan (2008) summarized that ridge ML estimation does not appropriately handle

the variance of  $\mathbf{S}$  because the variance of  $\mathbf{S}$  requires the fourth-order moments information as in Equation (3). Both ML and ridge ML use only the sample covariance matrix and do not involve sample fourth-order moments in parameter estimation.

Let  $\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})$  be the first-order derivative of  $\boldsymbol{\sigma}(\boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$  and be a matrix with dimension  $p^* \times q$  where  $q$  is the number of free parameters (i.e.,  $\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ ; Jacobian matrix). The standard error (SE) is the square root of the diagonal elements of  $(\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{I}_\sigma \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}))^{-1} / N$  where  $\mathbf{I}_\sigma$  is the information matrix for the structured model. One can use the expected or observed information matrix to estimate the standard errors. Additionally, to calculate the expected/observed information matrix, one can choose to use the sample covariance matrix  $\mathbf{S}$  or the model implied covariance matrix  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ . But with the sample covariance matrix  $\mathbf{S}$ , the expected and observed information matrices become the same. Hence, there are three combinations:  $ML_S(\mathbf{S})$ ,  $ML_{O.M}$  (observed information and  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ ), and  $ML_{E.M}$  (expected information and  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ ), which yield different sets of standard error estimates.

Robust statistics, such as sandwich type standard errors (Hardin, 2003; Huber et al., 1967; White et al., 1980; White, 1982) and rescaled-and-adjusted model fit test statistics (Satorra & Bentler, 1986, 1988, 1994; Jiang & Yuan, 2017; Yuan, Yang, & Jiang, 2017), make use of the sample fourth-order moment information. The sandwich standard error is obtained from the diagonals of the sandwich covariance matrix,

$$cov(\sqrt{N}\hat{\boldsymbol{\theta}}_{ML}) = (\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{I}_\sigma \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}))^{-1} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{I}_\sigma \boldsymbol{\Gamma}_{ADF} \mathbf{I}_\sigma \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) (\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{I}_\sigma \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}))^{-1}. \quad (5)$$

We use the estimator  $\hat{\boldsymbol{\Gamma}}_{ADF}$  to replace  $\boldsymbol{\Gamma}_{ADF}$  in practice.

### Generalized Least Squares Estimation

A class of generalized least squares (GLS) loss functions (Browne, 1974) is

$$F_{GLS}(\boldsymbol{\theta}) = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]' \hat{\mathbf{W}} [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \quad (6)$$

where  $\hat{\mathbf{W}}$  is a weight matrix of size  $p^* \times p^*$  that can take various forms. The best population weight matrix  $\mathbf{W}$  is the inverse of the asymptotic covariance matrix of  $\sqrt{N}(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))$ ,  $\boldsymbol{\Gamma}^{-1}$ . The loss function in Equation (6) is misspecified whenever  $\hat{\mathbf{W}}^{-1}$  is not a consistent estimate of  $\boldsymbol{\Gamma}$  (Savalei, 2014). When  $\hat{\mathbf{W}} = \mathbf{I}$ , estimation becomes least squares (LS) estimation. It is very unlikely that the asymptotic covariance matrix of sample covariances is an identity matrix. Due to the misspecification, sandwich standard errors usually accompany LS estimates.

The ADF estimator in Browne (1984) can be used to specify the weight matrix  $\hat{\mathbf{W}} = \hat{\boldsymbol{\Gamma}}_{ADF}^{-1}$ , as illustrated in Equation (3). Although the ADF estimator is known to be asymptotically efficient when  $N \rightarrow \infty$  and  $p/N \rightarrow 0$ , with finite sample sizes, it is unstable and the empirical standard errors can be much greater than those of normal theory based estimators even when data are nonnormally distributed (Yang & Yuan, 2019; Yuan & Bentler, 1997; Yuan & Chan, 2016). We refer to this method as the weighted least squares estimation (WLS; Rosseel, 2012).

Under the normal theory assumption,  $\mathbf{s}$  has an asymptotic covariance matrix that has a relatively simple form, that is the asymptotic covariance matrix is  $\{\boldsymbol{\Gamma}_N\}_{hj,kl} = \sigma_{hk}\sigma_{jl} + \sigma_{hl}\sigma_{jk}$ . Hence, with the normal theory assumption,  $\hat{\mathbf{W}}$  is specified to be  $\hat{\boldsymbol{\Gamma}}_N^{-1}$ . In samples,  $\hat{\boldsymbol{\Gamma}}_N$  can be either estimated by sample covariances  $\{\hat{\boldsymbol{\Gamma}}_{N.S}\}_{hj,kl} = s_{hk}s_{jl} + s_{hl}s_{jk}$  or estimated by the model implied covariances  $\{\hat{\boldsymbol{\Gamma}}_{N.M}\}_{hj,kl} = \hat{\sigma}_{hk}\hat{\sigma}_{jl} + \hat{\sigma}_{hl}\hat{\sigma}_{jk}$ . We refer to the estimation using sample covariances as  $GLS_S$  and the estimation using model implied covariances as  $GLS_M$ .

The loss functions of the aforementioned GLS methods plus some to be explained below are



summarized in Table 1. The sandwich standard error is obtained from the square root of the diagonals of the sandwich covariance matrix,

$$\text{cov} \left( \sqrt{N} \hat{\boldsymbol{\theta}}_{GLS} \right) = \left( \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{W} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) \right)^{-1} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{W} \boldsymbol{\Gamma}_{ADF} \mathbf{W} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) \left( \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \mathbf{W} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) \right)^{-1}. \quad (7)$$

We use  $\hat{\boldsymbol{\Gamma}}_{ADF}$  to replace  $\boldsymbol{\Gamma}_{ADF}$  and  $\hat{\mathbf{W}}$  to replace  $\mathbf{W}$  in practice. For WLS, Equation (7) simplifies to

$$\text{cov} \left( \sqrt{N} \hat{\boldsymbol{\theta}}_{GLS} \right) = \left( \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta})' \hat{\mathbf{W}}_{ADF} \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) \right)^{-1}.$$

### *Ridge Generalized Least Squares Estimation*

When the population distribution is unknown, the assumption of normality is unlikely to be supported. However, using a distribution free estimator leads to unstable performance. Yuan and Chan (2016) and Yuan, Jiang, and Cheng (2017) proposed two types of ridge GLS method (RGLS) for continuous and ordinal variables to stabilize the performance of the ADF estimator. To increase the efficiency of the ADF estimator (i.e., the WLS method), Yuan and Chan (2016) and Yuan, Jiang, and Cheng (2017) added components to the diagonals of  $\hat{\mathbf{W}}$ . More specifically,  $\hat{\mathbf{W}}$  is constructed as  $\left( (1-a)\hat{\boldsymbol{\Gamma}}_{ADF} + a\mathbf{I} \right)^{-1}$  or  $\left( (1-a)\hat{\boldsymbol{\Gamma}}_{ADF} + a \times \text{diag}(\hat{\boldsymbol{\Gamma}}_{ADF}) \right)^{-1}$  where  $a \in [0, 1]$  is a ridge tuning parameter. The former one is referred to as  $RGLS_I$  and the latter one is referred to as  $RGLS_D$ . RGLS gains the stability of employing simple weight matrices (i.e.,  $\mathbf{I}$  and  $\text{diag}(\hat{\boldsymbol{\Gamma}}_{ADF})$ ) and the asymptotic efficiency of the ADF estimator. For both  $RGLS_I$  and  $RGLS_D$ , the empirical performance depends on the value of the ridge tuning parameter, although the estimates remain consistent. When  $a = 1$  in  $RGLS_I$ ,  $RGLS_I$  becomes  $LS$ . When  $a = 0$  in  $RGLS_I$  or  $RGLS_D$ ,  $RGLS_I$  or  $RGLS_D$  becomes  $WLS$ . Yang and Yuan (2019), Yuan and Chan (2016), and Yuan, Jiang, and Cheng (2017) suggest that one can select the optimal  $a$  based on the efficiency and accuracy of parameter estimates. The  $a$  with the minimum root mean square error (RMSE) is the optimal  $a$  (denoted as  $a_s$ ). RMSE considers both efficiency and accuracy in estimation. The reason for considering both efficiency and accuracy is that estimates are not necessarily

unbiased with a small or moderate sample size, and the variance of SE estimates is not a good index for biased estimates.

Yang and Yuan (2019), Yuan and Chan (2016), and Yuan, Jiang, and Cheng (2017) found that  $a_s$  depend on all aspects of the data and model, including the number of variables, the number of factors, and the population distribution. In practice,  $a_s$  is unknown and needs to be estimated. Currently, there are two ways to estimate  $a_s$ . First, one can use the bootstrap procedure to create multiple samples and calculate empirical RMSE to select  $a_s$  (Yuan & Chan, 2016; Yuan, Jiang, & Cheng, 2017). Second, one can obtain a mapping function between  $a_s$  and all data/model features by an extensive simulation. In a real data analysis, such a mapping function can be adopted to estimate  $a_s$  based on the data/model information in that real data set (Jiang, 2018; Yang, 2018; Yang & Yuan, 2019).

Simulations from Yuan and Chan (2016) showed that with  $a_s$ ,  $RGLS_I$  performed better than  $RGLS_D$ ,  $LS$ , and  $WLS$  in terms of the efficiency and accuracy of parameter estimates. The sandwich SEs from  $RGLS_I$  were close to the empirical SEs across replications. Additionally, the convergence rates of  $RGLS_I$  and  $RGLS_D$  were higher than the ones of  $WLS$ .

### **Distributionally-Weighted Least Squares**

We introduce the procedure of distributionally-weighted least squares (DLS) in this section. The DLS method falls in the generalized least squares estimation framework. Similar to the RGLS methods, DLS balances performance under finite sample size with asymptotic performance. The research on RGLS by Yuan and Chan (2016) and Yuan, Jiang, and Cheng (2017) sheds light on the current study.  $\hat{\Gamma}_{ADF}$  has large variability and cannot demonstrate its merits with small to moderate sample sizes. By adding information from normal theory based estimated  $\Gamma$  ( $\hat{\Gamma}_N$ ), we can stabilize the performance from the ADF estimator and improve efficiency with small to moderate sample sizes. Using the loss function of general

GLS in Equation (6), we propose to specify the weight function as

$$\hat{\mathbf{W}} = \left( (1 - a)\hat{\Gamma}_{ADF} + a\hat{\Gamma}_{\mathbf{N}} \right)^{-1}, \quad (8)$$

where  $a$  is a tuning parameter. DLS provides consistent estimates of  $\theta$  regardless of the value of  $a$ , and DLS is more efficient (i.e., smaller standard errors) than WLS with  $\hat{\mathbf{W}} = \hat{\Gamma}_{ADF}$  unless  $a = 0$ . Although  $\hat{\theta}$  needs to be solved iteratively,  $\hat{\theta}$  is asymptotically equivalent to  $\left( \hat{\sigma}'\hat{\mathbf{W}}\hat{\sigma} \right)^{-1} \hat{\sigma}'\hat{\mathbf{W}}(s - \hat{\sigma})$  with  $\dot{\sigma}(\theta) = \frac{\partial \sigma(\theta)}{\partial \theta'}$ . As mentioned above,  $\hat{\Gamma}_{\mathbf{N}}$  can be estimated by sample covariances  $\left\{ \hat{\Gamma}_{N.S} \right\}_{ij,kl} = s_{ik}s_{jl} + s_{il}s_{jk}$  based on  $\mathbf{S}$  or the model implied covariances  $\left\{ \hat{\Gamma}_{N.M} \right\}_{ij,kl} = \hat{\sigma}_{ik}\hat{\sigma}_{jl} + \hat{\sigma}_{il}\hat{\sigma}_{jk}$  based on  $\Sigma(\hat{\theta})$ . Thus, there are sample covariance based DLS ( $DLS_S$ ) and model-implied covariance based DLS ( $DLS_M$ ) depending on how we calculate  $\hat{\Gamma}_{\mathbf{N}}$ . Given an  $a$ , we can use a Newton method to minimize the loss function and estimate the parameters. When  $a$  is not 1, the SE estimate is calculated as the sandwich standard error based on Equation (7). When  $a$  is 1,  $\Gamma_{\mathbf{N}}$  is selected and the normal assumption is used in the weight function, therefore we use the standard non-robust SE estimate.

The empirical performance of DLS depends on  $a$ . A larger  $a$  provides finite sample stability, whereas a smaller  $a$  provides asymptotic efficiency. When  $a = 1$  in DLS,  $\hat{\mathbf{W}}$  in Equation (8) simplifies to be  $\hat{\Gamma}_{\mathbf{N}}^{-1}$  and leads to normal theory based generalized least squares estimation. More specifically,  $DLS_S$  simplifies to the sample covariance based GLS ( $GLS_S$ ) and  $DLS_M$  simplifies to the model-implied covariance based GLS ( $GLS_M$ ). When  $a = 0$  in DLS,  $\hat{\mathbf{W}}$  simplifies to be  $\hat{\Gamma}_{ADF}^{-1}$  and leads to weighted least squares (WLS). Because DLS combines the strengths of the normal theory based GLS and distribution free (i.e., ADF) based GLS, we expect it to yield more efficient and accurate parameter estimates than ML, WLS, normal theory based GLS ( $GLS_M$  and  $GLS_S$ ), and LS when data are nonnormal, and to yield similar estimates as ML and normal theory based GLS when data are normal.

Following Yuan and Chan (2016), Yuan, Jiang, and Cheng (2017), and Yang and Yuan (2019), we select the optimal  $a$  ( $a_s$ ) as the one corresponding to the most efficient and accurate parameter estimates. We can quantify both the efficiency and accuracy of parameter estimates by the root mean square error (RMSE). When data are normal, we expect that  $a_s$  is 1 or close to 1, because the information in  $\hat{\Gamma}_{ADF}^{-1}$  should not be able to improve the efficiency and accuracy of the parameter estimates. When data are nonnormal,  $\hat{\Gamma}_N$  is misspecified, therefore we expect that  $a_s$  is not 1 and  $\hat{\Gamma}_{ADF}^{-1}$  steps in to provide more efficient and accurate parameter estimates.

### Model Fit Evaluation

In SEM, researchers usually evaluate whether the model fits the data well. The standard model fit statistic under the normality assumption is  $T = (N - 1) F(\hat{\theta})$ , where  $F(\hat{\theta})$  is the discrepancy ( $F_{ML}(\theta)$ ) or loss function ( $F_{GLS}(\theta)$ ).  $T$  asymptotically follows  $\chi_{df}^2$  where  $df = p^* - q$ . When the distribution of data is not normal, the asymptotic distribution of  $T$  is a weighted sum of  $df$  independent  $\chi_1^2$ . Let  $\mathbf{U} = \mathbf{W} - \mathbf{W}\dot{\sigma}(\theta) (\dot{\sigma}(\theta)' \mathbf{W}\dot{\sigma}(\theta))^{-1} \dot{\sigma}(\theta)' \mathbf{W}$ .  $\mathbf{W}$  is the population counter part of the  $\hat{\mathbf{W}}$  in the loss function for the GLS related methods, or the population counter part of the information matrix  $\mathbf{I}_\sigma$  for the ML methods. The mean of the asymptotic distribution is  $tr(\mathbf{U}\Gamma)$  (Satorra & Bentler, 1988). The key idea of the rescaled-and-adjusted test statistics is to adjust test statistics so that we can more closely approximate a reference  $\chi^2$  distribution. The general form is  $T_R = T/c$  with  $c$  as the adjustment constant. There are several widely used rescaled-and-adjusted test statistics in the literature. The most widely used one is the Satorra–Bentler statistic  $T_{SB}$  (Satorra & Bentler, 1986, 1988, 1994) with

$$c_{SB} = tr(\hat{\mathbf{U}}\hat{\Gamma}_{ADF})/df \quad (9)$$

where  $\hat{\mathbf{U}}$  and  $\hat{\Gamma}_{ADF}$  are consistent estimates of  $\mathbf{U}$  and  $\Gamma$ , respectively.  $c_{SB}$  rescales the asymptotic distribution to have a mean of  $df$ .  $T_{SB} = T/c_{SB}$  is referred to  $\chi_{df}^2$ . Satorra and Bentler (1988) proposed

another corrected statistic that has both the mean and variance of the test statistic adjusted,

$T_{MVA} = T/c_{MVA}$  with

$$c_{MVA} = tr \left( \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right)^2 \right) / tr \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right). \quad (10)$$

$T_{MVA}$  is referred to  $\chi_{df}^2$  where  $df^* = \left[ tr \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right) \right]^2 / tr \left( \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right)^2 \right)$ . In practice,  $\hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}}$  can be rank-deficient. If the rank of  $\hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}}$  is smaller than  $df$ , Equations (9) and (10) are not valid. Hence, Jiang and Yuan (2017) proposed to estimate the average eigenvalues of  $\mathbf{U}\mathbf{\Gamma}$  by replacing  $df$  with

$rank \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right)$ ,  $T_{JY} = T/c_{JY}$  with

$$c_{JY} = tr \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right) / rank \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right). \quad (11)$$

$T_{JY}$  is compared to  $\chi_{df}^2$  in Jiang and Yuan (2017). However, we propose to compare  $T_{JY}$  to  $\chi_{rank(\hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}})}^2$  because  $T_{JY}$  is re-centered to have a mean of  $rank \left( \hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}} \right)$ . In the simulation, we will explore the performance of  $T_{JY}$  compared to both  $\chi_{df}^2$  and  $\chi_{rank(\hat{\mathbf{U}}\hat{\mathbf{\Gamma}}_{\text{ADF}})}^2$ .

## Simulation Study

### *Simulation Design*

This simulation study examines the performance of the proposed distributionally-weighted least squares estimation,  $DLS_M$  and  $DLS_S$ . We varied the values of the following four factors: the total number of variables ( $p = 5, 15, \text{ and } 30$ ), the number of factors ( $m = 1 \text{ and } 3$ ), the sample size ( $N$  ranging from 40 to 1000), and the distributional conditions (a normal distribution, an elliptical distribution, and two skewed distributions due to either skewed factors or skewed errors). Specifically, in a model with a larger  $p$ , the sample size  $N$  needs to be greater in order to obtain a set of converged parameter estimates, therefore the conditions of  $N$  were nested within  $p$ . In total, we considered 18 conditions of  $N$ ,  $p$ , and  $m$  (Table 2

illustrates all the conditions).  $x$  was simulated from a confirmatory factor analysis (CFA) model

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad (12)$$

where  $\boldsymbol{\mu}$  is a  $p \times 1$  vector of means (fixed at 0 when generating the data and not estimated),  $\boldsymbol{\Lambda}$  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  $\boldsymbol{\Phi} = \text{cov}(\boldsymbol{\xi})$  and  $\boldsymbol{\Psi} = \text{cov}(\boldsymbol{\varepsilon})$ , then the corresponding population covariance matrix of  $x$  is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}. \quad (13)$$

We specified a simple cluster structure, with each factor having the same number of free non-zero loadings. For example, when  $p = 15$  and  $m = 3$ , each factor had 5 nonzero loadings. Following Yang and Yuan (2019), the population values of the factor loadings 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.  $\boldsymbol{\Psi}$  was calculated to ensure that the diagonal elements of  $\boldsymbol{\Sigma}$  were 1. Following Yuan and Chan (2016) and Yang and Yuan (2019), when the distributional condition was normal,  $\boldsymbol{\xi} = \boldsymbol{\Phi}^{1/2}Z_{\xi}$  and  $\boldsymbol{\varepsilon} = \boldsymbol{\Psi}^{1/2}Z_{\varepsilon}$  where  $\boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2} = \boldsymbol{\Phi}$ ,  $\boldsymbol{\Psi}^{1/2}\boldsymbol{\Psi}^{1/2} = \boldsymbol{\Psi}$ , and both  $Z_{\xi}$  and  $Z_{\varepsilon}$  followed a standard normal distribution  $N(0, 1)$ . When the distributional condition was elliptical (symmetric distributions with heavy tails),  $\boldsymbol{\xi} = r\boldsymbol{\Phi}^{1/2}Z_{\xi}$  and  $\boldsymbol{\varepsilon} = r\boldsymbol{\Psi}^{1/2}Z_{\varepsilon}$  with  $r \sim (3/\chi_5^2)^{1/2}$ . Because  $E(r^2) = 1$ ,  $E(Z_{\xi}) = 0$ , and  $E(Z_{\varepsilon}) = 0$ , Equation (13) is still applicable. When the factors were skewed, we also considered the heavy-tail feature in generating data. In other words, we used  $r \sim (3/\chi_5^2)^{1/2}$  to add heavy tails to the skewed distribution:  $\boldsymbol{\xi} = r\boldsymbol{\Phi}^{1/2}Z_{\xi}$  and  $\boldsymbol{\varepsilon} = r\boldsymbol{\Psi}^{1/2}Z_{\varepsilon}$  where  $Z_{\xi} \sim \text{standardized}(\chi_1^2)$  and  $Z_{\varepsilon} \sim N(0, 1)$ . When the error were skewed,  $\boldsymbol{\xi} = r\boldsymbol{\Phi}^{1/2}Z_{\xi}$  and  $\boldsymbol{\varepsilon} = r\boldsymbol{\Psi}^{1/2}Z_{\varepsilon}$  where  $Z_{\xi} \sim N(0, 1)$  and  $Z_{\varepsilon} \sim \text{standardized}(\chi_1^2)$ . For each condition, we simulated 1000 datasets. We conducted the simulation with R (version 3.6.1). The R code of DLS is provided on

([https://github.com/hduquant/lab\\_code\\_collection/blob/master/DLS/DLS\\_code.R](https://github.com/hduquant/lab_code_collection/blob/master/DLS/DLS_code.R)) and the ML estimation is implemented by an R package, *lavaan* (version 0.6-5) (Rosseel, 2012). DLS will soon be available in the forthcoming version of *lavaan*.

We considered the values for the tuning parameter  $a$  from 0 to 1 with equal interval of 0.01 (i.e., 0, 0.01, ..., 0.99, 1). Hence, there were 101  $a$  values adopted for each simulated dataset. In estimating the CFA model in Equation (12), all the diagonal elements of  $\Phi$  were fixed at 1 and all non-zero factor loadings were freely estimated. Therefore, the  $q \times 1$  vector  $\theta$  contains all the free parameters in the SEM model:  $p$  free factor loadings,  $m(m-1)/2$  factor covariances, and  $p$  error variances ( $q = 2p + m(m-1)/2$ ).

There were 11 methods considered in the simulation (see Table 1). The distributionally-weighted least squares (DLS) estimation can rely on either the sample covariances ( $S$ ) or the model-implied covariances ( $\Sigma(\hat{\theta})$ ) to obtain the estimated normal theory based asymptotic covariance matrix of  $s(\hat{\Gamma}_N)$ , therefore there were sample covariance based DLS ( $DLS_S$ ) and model-implied covariance based DLS ( $DLS_M$ ). When  $a = 1$  in DLS, the sample covariance and normal theory based GLS ( $GLS_S$ ) and model-implied covariance and normal theory based GLS ( $GLS_M$ ) become special cases of DLS. When  $a = 0$  in DLS, weighted least squares estimation (WLS) becomes a special case of DLS. We also considered  $RGLS_D$  and  $RGLS_I$  from Yuan and Chan (2016). In  $RGLS_I$ , when  $a = 1$ ,  $\hat{W}$  simplifies to be  $I$  and leads to least squares estimation (LS). In addition, we considered normal theory based maximum likelihood (ML) estimation. Among ML procedures, the combination of different information matrix (observed or expected information) and different covariance matrix (sample or model-implied covariance) yields three methods ( $ML_{O.M}$ ,  $ML_S$ , and  $ML_{E.M}$ ) with different sets of standard error (SE) estimates but the same parameter estimates.

To investigate both the efficiency and accuracy of parameter estimates, the root mean square error (RMSE) is a widely used index (e.g., Yuan & Chan, 2016; Yuan, Yang, & Jiang, 2017; Yang & Yuan, 2019). Let  $\hat{\theta}_{ij}$  be the estimate of the  $i$ th parameter in the  $j$ th replication. The RMSE for each condition was

averaged over all parameters,

$$RMSE = \frac{1}{q} \sum_{i=1}^{i=q} \left( \frac{1}{1000} \sum_{j=1}^{j=1000} (\hat{\theta}_{ij} - \theta_i)^2 \right)^{1/2}, \quad (14)$$

where  $\theta_i$  was the true value for the  $i$ th parameter.

To investigate the performance of the SE estimates of different methods, we calculated the relative biases of the SE estimates. The true SE is unknown, therefore we calculated the standard deviation for each parameter estimate across 1000 replications as the empirical SE of each parameter. Let  $\hat{SE}_{ij}$  be the SE estimate of  $i$ th parameter in the  $j$ th replication and  $SE_i$  be the empirical SE of  $i$ th parameter. The relative biases of the SE estimates were averaged over all parameters that we are interested in,

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

For  $RGLS_D$ ,  $RGLS_I$ ,  $LS$ ,  $WLS$ ,  $GLS_M$ ,  $GLS_S$ ,  $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$ , the SE estimates were the sandwich SE estimates. For  $DLS_M$  and  $DLS_S$ , the SE estimates were the sandwich SE estimates when  $a$  was not 1. When  $a$  was 1, the standard SE estimates were adopted. In terms of model fit evaluation, we considered 4 model fit statistics: the standard model fit statistic ( $T$ ), the Satorra–Bentler test statistic ( $T_{SB}$ ), the mean and variance adjusted test statistic ( $T_{MVA}$ ), and the Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ).

### Structure of Simulation Results

In the results sections, we first summarize convergence issues of the 11 methods. Second, we present the influence of  $a$  on the performance of  $DLS$ . Third, we compared RMSEs of parameter estimates, empirical SEs, biases of SE estimates, and Type I error rates of model fit statistics across methods,



respectively. In the end, we provide our conclusions from the simulation results. We present all the detailed results in the supplemental material.

### *Convergence Issue*

The convergence rates of  $LS$ ,  $GLS_M$ ,  $ML_{O.M}$ ,  $ML_{E.M}$ ,  $MLS$ ,  $RGLS_D$  (with the optimal  $a$ ) and  $RGLS_I$  (with the optimal  $a$ ) were all almost 1 across conditions (i.e.,  $> 0.98$ ). When the sample size  $N$  was too small relative to the model complexity (e.g.,  $N = 40$ ,  $p = 15$ ,  $m = 3$ ),  $GLS_S$  and  $DLS_S$  could have convergence rates lower than 0.8 but higher than 0.7, and  $WLS$  has no converged results at all because only the ADF estimator was used.  $DLS_M$  with the optimal  $a$  always had a convergence rate near 1; with a larger sample size,  $DLS_S$  went above 0.9. The influence of different  $a$  values on the convergence of  $DLS$  will be expanded upon in next section. We kept only the converged solutions among the 1000 replications.

### *Influence of $a$ on $DLS$*

The value of  $a$  determines the performance of  $DLS_M$  and  $DLS_S$ . We discuss the performance in terms of convergence rates, RMSE, relative biases of SE, and Type I error rates of test statistics.

In terms of root mean square errors (RMSE) of parameter estimates, the optimal  $a$  ( $a_s$ ) which minimized the average RMSE across all parameters depended on the distribution and the complexity of the model. Within each model,  $N$  did not obviously change the trajectories of RMSE along with  $a$ .  $a_s$  was consistent between  $DLS_M$  and  $DLS_S$ , and the minimal RMSE was smaller in  $DLS_M$ . When the distribution was normal,  $a = 1$  (or almost 1) in both  $DLS_M$  and  $DLS_S$  provided the smallest RMSE, which indicated that the algorithm heavily weighted on  $\hat{\Gamma}_N$ . We illustrate the plots of RMSE of all the parameter estimates for  $DLS_M$ ,  $DLS_S$ ,  $RGLS_I$ , and  $RGLS_D$ , when  $N = 300$ ,  $p = 5$  and  $m = 1$  in Figure 1 and  $N = 300$ ,  $p = 30$  and  $m = 3$  in Figure 2 respectively, as examples of simple and complex models with a moderate sample size. When the distribution was nonnormal,  $a_s$  depended on the complexity of the model. With a simpler model,  $a_s$  was smaller in  $DLS_M$  and  $DLS_S$  (e.g., can be about

0.7; see Figure 1). In a more complex model,  $a_s$  was close to 1 in  $DLS_M$  and  $DLS_S$  (see Figure 2). For  $RGLS_I$  and  $RGLS_D$ ,  $a_s$  was smaller in a simpler model. This indicated that  $RGLS$  employed simple weight matrices (i.e.,  $\mathbf{I}$  and  $diag(\hat{\Gamma}_{ADF})$ ) more strongly in a simpler model. Compared to  $RGLS$ ,  $DLS_M$  and  $DLS_S$  were more sensitive to the selection of  $a$ .

In terms of convergence rates, when  $a$  was large,  $DLS_M$  and  $DLS_S$  had no convergence issues. With a smaller  $a$ ,  $DLS_M$  and  $DLS_S$  relied more on the ADF estimator. The ADF estimator had convergence issues when the sample size  $N$  was small relative to the number of variables  $p$ , therefore  $DLS_M$  and  $DLS_S$  with small  $a$  values could have low convergence rates. For example, when  $N = 60$ ,  $p = 15$ ,  $m = 3$ , and  $a = 0.1$ , the convergence rates ranged from 0.4 to 0.8 across all distributional conditions. With a large enough  $N$ , the convergence rates were almost 1 even when  $a = 0$ . For example, when  $N = 200$ ,  $p = 15$ ,  $m = 3$ , and  $a = 0$ , the convergence rates were above 0.95 across all distributional conditions. With  $a_s$ , the convergence rates of  $DLS_M$  were larger than 0.98 across all conditions, and the convergence rates of  $DLS_S$  were larger than 0.91 (only when  $p = 15$ ,  $m = 3$ ,  $N = 40$  with skewed factors, the convergence rate in  $DLS_S$  was 0.735).

The relative biases of the SE estimates were also influenced by  $a$  in  $DLS_M$  and  $DLS_S$ . Because factor loadings and covariances between factors are usually the focus of research questions, we focus on the relative biases of the SE estimates of factor loadings and factor covariances. The SE estimates of residual variances were slightly higher than those of factor loadings and factor covariances. Surprisingly, when  $p = 5$  and  $m = 1$ ,  $N \geq 300$ , and the data were normal,  $a = 1$  yielded the largest average relative biases for factor loadings and factor covariances but smallest average relative biases for all parameters, whereas  $N < 300$ ,  $a = 1$  yielded the smallest average relative biases for factor loadings and factor covariances (see Figure 3 for  $N = 300$ ,  $p = 5$  and  $m = 1$ ). When the data were normal with other models,  $a = 1$  yielded the smallest average relative biases for factor loadings and factor covariances and smallest average relative biases for all parameters (see Figure 4 for  $N = 300$ ,  $p = 30$  and  $m = 3$ ). When the data

were nonnormal, similar to RMSE, the  $a$  which gave the minimal biases was smaller with a simpler model (see Figures 3-4). Note that  $a_s$  was not guaranteed to provide the smallest biases. For example, when  $N = 300$ ,  $p = 5$ ,  $m = 1$ , and the data were elliptical, the  $a_s$  that provided the smallest average RMSE in  $DLS_M$  was 0.27, but the  $a$  that provided the smallest average bias was 0.66.

In addition,  $a$  influenced Type I error rates from the standard model fit statistic ( $T$ ), the Satorra–Bentler test statistic ( $T_{SB}$ ), the mean and variance adjusted test statistic ( $T_{MVA}$ ), the Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ). The influential pattern depended on the sample size, the model complexity, and the distribution.

In general,  $a$  influenced convergence rates (when  $N$  was small), RMSEs of parameter estimates, biases of standard error estimates, and Type I error rates of model fit statistics in  $DLS_M$  and  $DLS_S$ .  $a$ 's influence depended on the sample size, the model complexity, and the distribution. With the  $a_s$  from the minimal RMSE, the convergence rates of  $DLS_M$  and  $DLS_S$  were acceptable. When we select  $a_s$  based on the smallest RMSE,  $a_s$  was not guaranteed to provide the smallest biases of SE estimates.

#### *RMSE across Methods*

We compared the 11 methods ( $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$ ,  $RGLS_I$ ,  $LS$ ,  $WLS$ ,  $GLS_S$ ,  $GLS_M$ ,  $ML_S$ ,  $ML_{O.M}$ , and  $ML_{E.M}$ ) in terms of their efficiency and accuracy using the RMSE across all sample sizes and models, and separately by distributional conditions. Among  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$ , and  $RGLS_I$ , the minimal RMSEs were selected given each method, and  $a$  corresponding to the smallest RMSE is referred to as  $a_s$ . We present the RMSEs from all methods and all conditions in the supplemental material. When the data were normal, we illustrate the RMSEs from the 11 methods in Figure 5 with different  $N$  and models. One overall pattern was that when  $N$  was larger, the RMSEs from the 11 methods became smaller. The RMSEs from  $DLS_M$  were the smallest among  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$  and  $RGLS_I$ , followed by  $RGLS_I$  (see the upper left panel of Figure 5). There were almost no differences of

RMSEs when the model was simple ( $p = 5$  and  $m = 1$ ). When the model became more complex and  $N$  was small, the sample covariance based DLS ( $DLS_S$ ) had a large RMSE which indicated inefficient and inaccurate estimation, probably due to sample covariances not being stable with a small  $N$ . Additionally,  $RGLS_I$  performed better than  $RGLS_D$ , consistent with the findings of Yuan and Chan (2016). Among  $LS$ ,  $WLS$ ,  $GLS_S$ , and  $GLS_M$ ,  $GLS_M$  had the smallest RMSEs (see the upper right panel of Figure 5). Similar to  $DLS_M$  and  $DLS_S$ , the model-implied covariance based GLS ( $GLS_M$ ) outperformed the sample covariance based GLS ( $GLS_S$ ). The point estimates from  $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$  were the same and hence had the same RMSEs (see the lower left panel of Figure 5). We select the methods that provided the smallest RMSEs from the upper left panel, upper right panel, and lower left panel, and plot them again in the lower right panel of Figure 5:  $DLS_M$ ,  $RGLS_I$ ,  $GLS_M$ , and  $ML$  ( $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$  had the same point estimates and RMSEs). Although the RMSEs in  $DLS_M$  were also larger with a more complex model and a smaller  $N$ , they were still smaller than the other methods (e.g.,  $RGLS_I$ ) or equivalent to the normal theory based methods ( $GLS_M$  and  $ML$ ), because with normal data,  $a_s$  in  $DLS_M$  was usually almost 1.

When the distributional condition was elliptical, the errors were skewed, or the factors were skewed, the patterns of RMSEs from the 11 methods were similar across the distributional conditions, therefore we present the elliptical condition as an example. Among  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$ , and  $RGLS_I$ ,  $DLS_M$  had the smallest RMSEs, followed by  $RGLS_I$  (see the upper left panel of Figure 6). Among  $LS$ ,  $WLS$ ,  $GLS_S$ , and  $GLS_M$ ,  $GLS_M$  had the smallest RMSEs (see the upper right panel of Figure 6). Due to the instability of the sample covariances with small  $N$ s,  $DLS_S$  and  $GLS_S$  could yield large RMSEs. We select the methods that provided the smallest RMSEs and plot them in the the lower right panel of Figure 6 again:  $DLS_M$ ,  $RGLS_I$ ,  $GLS_M$ , and  $ML$  (the types of information matrix and covariance matrix did not matter).  $DLS_M$  provided the smallest RMSEs, followed by  $RGLS_I$ . The normal theory based methods, such as  $GLS_M$  and  $ML$ , had larger RMSEs compared to  $DLS_M$ . Such a difference of RMSEs was larger

when the error were skewed or the factors were skewed (e.g., 0.08). This indicated that the ADF component in  $DLS_M$  improved the efficiency and accuracy of parameter estimation when data were nonnormal, while the completely normal theory based methods provided somewhat less accurate estimates due to assumption violations.

#### *Empirical SEs across Methods*

We focus on the empirical SEs of the estimates of factor loadings and covariances between factors. The average empirical SEs of residual variances were generally lower than those of factor loadings and factor covariances. We present the empirical SEs from all methods and all conditions in the supplemental material. When the data were normal, the average empirical SEs of  $DLS_M$  were the relatively smallest and equivalent to the normal theory based methods ( $GLS_M$  and  $ML$ ). When the data were nonnormal,  $DLS_M$  and  $RGLS_I$  had the smallest average empirical SEs. Depending on the distributional condition and  $N$ ,  $DLS_M$  or  $RGLS_I$  could be smaller than the other (see Figure 7 for an example for the skewed factor case). Especially, when  $N$  was small,  $DLS_M$  could have smaller SEs than  $RGLS_I$ .

#### *Relative Biases of SE Estimates across Methods*

We compared the 11 methods in terms of their relative biases of the SE estimates, averaging over factor loadings and factor covariances. The presented  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$ , and  $RGLS_I$  used the  $a_s$  which provided the smallest RMSE. We present the relative biases of SE estimates from all methods and all conditions in the supplemental material. When the data were normal, the relative biases of the SE estimates from the 11 methods with different  $N$  and models are presented in Figure 8. Similar to the RMSEs, the overall pattern was that when  $N$  increased, the relative biases from the 11 methods generally became smaller. The biases of SE estimates from  $DLS_M$  were the smallest among  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$  and  $RGLS_I$ , followed by  $RGLS_I$  (see the upper left panel of Figure 8). By our definition, the SE estimates of  $DLS_M$  and  $DLS_S$  depended on  $a_s$ . With  $a_s = 1$ , the SEs were standard SE estimates; otherwise the SEs

were sandwich SE estimates. Among  $LS$ ,  $WLS$ ,  $GLS_S$ , and  $GLS_M$ , the sandwich SE estimates of  $LS$  had the smallest biases (see the upper right panel of Figure 8). Although  $GLS_S$  and  $GLS_M$  had the correctly specified normal assumption, the sandwich SE estimates were calculated which could increase the biases of SE estimates. The  $ML$  sandwich SE estimates based on the expected information and model implied covariance  $ML_{E.M}$  provided the smallest biases among the  $ML$  methods (see the lower left panel of Figure 8). We select the methods that provided the smallest biases of SE estimates in the upper left panel, upper right panel, and lower left panel, and plot them in the lower right panel of Figure 8:  $DLS_M$ ,  $RGLS_I$ ,  $LS$ , and  $ML_{E.M}$ . The SE estimates of  $DLS_M$  had the smallest biases.

When the distributional condition was elliptical, errors were skewed, and factors were skewed, the patterns of the relative biases of SE estimate from the 11 methods were similar across the distributional conditions, therefore we present the elliptical condition as an example. Among  $DLS_M$ ,  $DLS_S$ ,  $RGLS_D$ , and  $RGLS_I$ ,  $RGLS_I$  had the smallest biases of SE estimates (see the upper left panel of Figure 9). When  $N$  was not extremely small, the SE estimates of  $DLS_M$  were similar to those of  $RGLS_I$ . Among  $LS$ ,  $WLS$ ,  $GLS_S$ , and  $GLS_M$ , the sandwich SE estimates of  $GLS_M$  and  $LS$  had the similar biases (see the upper right panel of Figure 9). The  $ML$  sandwich SE estimates based on the expected information and sample covariance  $ML_{E.M}$  provided the smallest biases among the  $ML$  methods (see the lower left panel of Figure 9). We select the methods that provided the smallest biases of SE estimates and plot them in the lower right panel of Figure 9:  $DLS_M$ ,  $RGLS_I$ ,  $GLS_M$ , and  $ML_{E.M}$ .  $DLS_M$  or  $RGLS_I$  generally had the smallest biases of SE estimates. The SE estimates of  $RGLS_I$  could be less biased compared to  $DLS_M$  when  $N$  was small.

#### *Type I Error Rates across Methods*

We examined the Type I error rates of the standard model fit statistic ( $T$ ), the Satorra–Bentler test statistic ( $T_{SB}$ ), the mean and variance adjusted test statistic ( $T_{MVA}$ ), and the Jiang-Yuan rank adjusted test

statistic ( $T_{JY}$ ) from the 11 methods from the four distributional conditions. We consider a Type I error rate between 0.025 and 0.075 as satisfactory (Bradley, 1978). We present the Type I error rates from all methods and all conditions in the supplemental material.  $T$ ,  $T_{SB}$ , and  $T_{MVA}$  yielded either too small (e.g., 0) or too larger Type I error rates (e.g., 1) with all 11 methods. We failed to find a method which uniformly provided acceptable Type I error rates using  $T$ ,  $T_{SB}$ , or  $T_{MVA}$ . We examined the performance of  $T_{JY}$  compared to both  $\chi_{df}^2$  and  $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ . With referring to  $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ , the performance of  $T_{JY}$  was better, whereas with referring to  $\chi_{df}^2$ , the Type I error rates in most methods generally were 0 when  $N$  was small. Hence, we report the results of  $T_{JY}$  with  $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ . There were 4 methods that performed relatively better than the others using  $T_{JY}$ :  $DLS_M$ ,  $DLS_S$ ,  $GLS_M$ , and  $GLS_S$ . Their Type I error rates of  $T_{JY}$  across distributional conditions, models, and  $N$  are presented in Figure 10. When  $N$  was too small,  $DLS_M$ ,  $DLS_S$ ,  $GLS_M$ , and  $GLS_S$  deviated from the nominal level. As  $N$  became larger, the Type I error rates were more acceptable.  $DLS_M$  almost always provided acceptable Type I error rates unless  $N$  was too small relative to the model complexity. However, the ML methods with  $T_{JY}$  ( $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ ) could have too high Type I error rates.

### *Conclusions from the Simulation Study*

In summary,  $DLS_M$  provided the smallest RMSEs regardless of the distributions. When data were normal,  $GLS_M$  and  $ML$  provided similar RMSEs as  $DLS_M$ ; when data were nonnormal,  $RGLS_I$  provided the second smallest RMSEs. In terms of the empirical SEs, when data were normal, the empirical SE estimates of  $DLS_M$ ,  $GLS_M$ , and  $ML$  were the smallest, and the SE estimates of  $DLS_M$  had the smallest biases. When data were nonnormal, the empirical SE estimates of  $DLS_M$  and  $RGLS_I$  were the smallest, and the SE estimates of  $DLS_M$  and  $RGLS_I$  had similar small biases while the SE estimates of  $RGLS_I$  could be less biased when  $N$  was small. Additionally, the Type I error rates of Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ) using  $DLS_M$  were generally around the nominal level (0.05). Overall, we

recommend  $DLS_M$  given its performance.

### *Model Misspecification Simulation*

Since  $DLS_s$  was inferior to  $DLS_M$  in the simulation above, we conducted a small-scale simulation study to explore whether  $DLS_s$  outperformed  $DLS_M$  when the model was misspecified. We considered the case where  $p = 30$ ,  $m = 3$ ,  $N$  varied as 100, 300, 500, and 1000, and the distributional condition was normal or elliptical. We generated data the same as in the previous section, but we assumed (1) the factor correlations were 0 or (2) all factor loadings were equal. All the detailed results are presented in the supplemental material. The patterns from the two types of misspecification did not differ much.  $DLS_M$  outperformed  $DLS_s$  in terms of RMSEs, the relative biases of SE estimates, and the empirical SEs. When data were normal, as the sample size ( $N$ ) increased, the difference between  $DLS_s$  and  $DLS_M$  became smaller. When data were elliptical, with a larger sample size, the difference between  $DLS_s$  and  $DLS_M$  in terms of RMSEs and empirical SEs became smaller but the difference regarding the relative biases of SE estimates did not get smaller (see Figure 11 for elliptical data and equal factor loading assumption as an example). We found that  $T_{JY}$  with  $DLS_M$  generally indicated a poor model fit across all sample sizes. It indicated that consistent with the Type I error rate simulation results,  $T_{JY}$  is an appropriate test statistic for  $DLS_M$  for model fit evaluation.

### **Real Data Example**

In real data analyses,  $a_s$  is unknown and needs to be estimated. In this section, we illustrate how to apply the proposed distributionally-weighted least squares estimation using a bootstrap procedure. We considered  $DLS_M$  and  $RGLS_I$  that performed relatively well in the simulation and two normal theory based methods,  $MLS$  and  $GLS_M$ . We used a public dataset which is available in the R package, *lavaan* (version 0.6-5) (Rosseel, 2012). The original dataset from Holzinger and Swineford (1939) has mental ability test scores of 26 tests for the 7th and 8th grade children from two different schools (Pasteur and



Grant-White). We focused on a subset of 9 variables and 145 children from the Grant-White school only. This subset is widely used in the SEM literature (e.g., Jöreskog, 1969; Yuan & Chan, 2016). 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 (13) with  $\Psi = \text{diag}(\psi_{11}, \psi_{22}, \psi_{33}, \psi_{44}, \psi_{55}, \psi_{66}, \psi_{77}, \psi_{88}, \psi_{99})$ ,

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

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

Following Yuan and Chan (2016), we conducted a bootstrap study to evaluate the empirical RMSE using the estimated parameters from the bootstrap samples. Let  $\mathbf{x}_i$  be a  $9 \times 1$  vector of test scores for individual  $i$ ,  $\mathbf{S}_x$  be the sample covariance matrix for  $\mathbf{x}_i$ , and  $\hat{\Sigma} = \Sigma(\hat{\theta}_{ML})$  be the model-implied covariance matrix based on the ML estimates using the raw data. First, we adopted the Bollen-Stine transformation for bootstrapping (Bollen & Stine, 1992),

$$\mathbf{x}_i^{(0)} = \hat{\Sigma}^{1/2} \mathbf{S}_x^{-1/2} \mathbf{x}_i. \quad (15)$$

Equation (15) is to create a new sample covariance matrix. The sample covariance of  $\mathbf{x}_i$  is  $\mathbf{S}_x$ , while the sample covariance of  $\mathbf{x}_i^{(0)}$  is  $\hat{\Sigma}$ . After transforming, the null hypothesis (the factor model above) is true and  $\hat{\theta}_{ML}$  gives the true population parameters for  $\mathbf{x}_i^{(0)}$ . Second, we drew with replacements of  $\mathbf{x}_i^{(0)}$  to construct

1000 bootstrap samples. We varied the tuning parameter  $a$  from 0 to 1 with an equal interval of .01 (i.e., 0, 0.01, ..., 0.99, 1) and applied  $DLS_M$  and  $RGLS_I$  (with each  $a$  value) to each bootstrap sample. We applied the same RMSE equation as Equation (14) where  $\hat{\theta}_{ML}$  from the raw data  $x_i$  ( $i = 1, \dots, 145$ ) was treated as  $\theta$  to calculate an empirical RMSE. Under the null hypothesis, the empirical RMSE is a consistent estimator of the true RMSE (Yuan & Chan, 2016).

We plot the empirical average RMSE of all the parameter estimates for  $DLS_M$  and  $RGLS_I$  along with  $a$  in Figure 12.  $a_s$  was 0.75 in  $DLS_M$  with the smallest RMSE at 0.094, and  $a_s$  was 0.36 in  $RGLS_I$  with the smallest RMSE at 0.097. Consistent with the simulation results,  $DLS_M$  yielded a smaller RMSE compared to  $RGLS_I$ . Yuan and Chan (2016) reported  $a_s$  at 0.35 but they considered  $a$  from 0 to 1 with an equal interval of .05. The difference between our replication of  $RGLS_I$  and the result in Yuan and Chan (2016) is trivial. We conducted a sensitivity test with respect to the number of bootstrap samples. Besides 1000 samples, we also estimated  $a_s$  using 500, 2000, and 5000 samples. The estimated  $a_s$  was always 0.36 in  $RGLS_I$  and the estimated  $a_s$  was 0.74 or 0.75 in  $DLS_M$ , which supported the validation of the selected  $a_s$  value.

We applied  $DLS_M$  ( $a_s = 0.75$ ),  $RGLS_I$  ( $a_s = 0.36$ ),  $MLE_M$ , and  $GLS_M$  to the raw data. The Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ) with  $\chi^2_{rank(\hat{U}\hat{\Gamma}_{ADF})}$  for  $DLS_M$  ( $a_s = 0.75$ ),  $RGLS_I$  ( $a_s = 0.36$ ),  $MLE_M$  and  $GLS_M$  were 28.631 ( $p = 0.193$ ), 31.759 ( $p = 0.105$ ), 27.833 ( $p = 0.222$ ), and 27.255 ( $p = 0.245$ ), respectively.  $DLS_M$ ,  $RGLS_I$ ,  $MLE_M$ , and  $GLS_M$  were shown to have good model fits. The parameter estimates, the SE estimates, and the z scores are in Table 3. There was little difference between the results of the two normal theory based methods,  $MLE_M$  and  $GLS_M$ .  $DLS_M$  ( $a_s = 0.75$ ) had similar results as  $MLE_M$  and  $GLS_M$  except that  $\psi_{88}$  is not statistically significantly different from 0.  $RGLS_I$  ( $a_s = 0.36$ ) reached the similar significance conclusion as  $DLS_M$  ( $a_s = 0.75$ ) although  $\hat{\theta}$ s and SEs were slightly different.

### Conclusion

In real data analysis, data are unlikely to be exactly normally distributed. If we ignore the non-normality reality, the parameter estimates, standard error estimates, and model fit statistics from normal theory based methods (e.g., ML and GLS) are unreliable. Even with the help of robust statistics, the normal theory based methods' performances are not adequate with finite sample sizes. On the other hand, the asymptotically distribution free (ADF) estimator (i.e., WLS) does not rely on any distribution assumption but cannot demonstrate its efficiency advantage with small and modest sample sizes. We propose a distributionally-weighted least squares (DLS) estimator, and expect that it can perform better than the existing generalized least squares, because it combines normal theory based and ADF based generalized least squares estimation. And there are sample covariance based DLS ( $DLS_S$ ) and model-implied covariance based DLS ( $DLS_M$ ).

Computer simulation results suggest that  $DLS_M$  provides relatively accurate and efficient estimates. Compared to ML estimators ( $ML_S$ ,  $ML_{O.M}$ , and  $ML_{E.M}$ ), LS, WLS,  $RGLS_D$ ,  $RGLS_I$ ,  $GLS_M$ , and  $GLS_S$ ,  $DLS_M$  provided the smallest RMSEs regardless of the distributions. With normal data,  $DLS_M$  had the relatively smaller empirical SE estimates and smallest biases of SE estimates, which were similar to those of  $ML_{E.M}$  and LS; with nonnormal data,  $DLS_M$  and  $RGLS_I$  had relatively smallest empirical SE estimates and smallest biases, while  $RGLS_I$ 's SE estimates could be smaller and less biased when  $N$  was small. When DLS coupled with the Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ),  $DLS_M$  generally provided Type I error rates close to the nominal level unless  $N$  was too small relative to the model complexity. However,  $DLS_S$  did not perform well in terms of RMSEs, biases of SE estimates, and model fit Type I error rates. Overall,  $DLS_M$  is competitive with the existing methods in different aspects. The simulation findings echo our anticipations at the beginning of the paper: (1)  $DLS_M$  yields more accurate and efficient estimates than those from the ADF estimator (WLS), (2)  $DLS_M$  boosts convergence rate compared to WLS, and (3) using data information while holding the normality assumption to some degree

enhances  $DLS_M$ 's performance. But compared to  $RGLS_I$ ,  $DLS_M$  was shown to be more sensitive to the selection of  $a$  (Figures 2 and 4), which can be a reason for one to use  $RGLS_I$  instead of  $DLS_M$ .

In the simulation, we explored the performance of  $T_{JY}$  by referring to  $\chi_{df}^2$  and  $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ . We found when referring to  $\chi_{df}^2$ , the Type I error rates in most methods generally were 0 when  $N$  was small. When referring to  $\chi_{rank(\hat{U}\hat{\Gamma}_{ADF})}^2$ , multiple methods including  $DLS_M$  provided acceptable Type I error rates, whereas the ML methods could be too liberal. It should be recognized that general methods for statistical model evaluation, such as the Monte Carlo approach of Jalal and Bentler (2018), can also be adapted to evaluate DLS results.

A consequence of the availability of a method (i.e.,  $T_{JY}$ ) with an acceptable Type I error rate is that the noncentral  $\chi^2$  distribution is a good candidate to describe the behavior of DLS tests under conditions of not-too-large misspecification. Hence methods of describing model adequacy based on noncentrality-based fit indices such as RMSEA (Steiger & Lind, see Steiger, 2016) and CFI (Bentler, 1990) should be able to be utilized. Illustrative recent research on these indices is Lai (2019; 2020), Lai and Green (2016), Moshagen and Auerswald (2018), and Zhang and Savalei (2016; 2020). It also makes sense to consider new descriptive indices (e.g., Gomer et al., 2019), or even traditional ones such as NFI and SRMR (e.g., Bentler, 2006; see also Maydeu-Olivares, 2017) for use with DLS.

The value of the tuning parameter  $a$  determines the performance of  $DLS_M$ . In the simulation, we selected the  $a$  which yielded the smallest RMSEs in  $DLS_M$  and referred to it as  $a_s$ . We were able to calculate RMSEs because we knew the population parameters in the simulation. In practice, as we illustrated in the real data example, we can use a bootstrap procedure to calculate an empirical RMSEs to select the  $a_s$ .  $a$  influenced convergence rates. With a small  $a$ ,  $DLS_M$  could have nonconvergence when  $N$  was small relative to the model complexity. But  $a_s$  was not selected to be small based on RMSEs.  $a$  also influenced RMSEs, biases of SE estimates, and model fit Type I error rates of  $DLS_M$ . The influence depended on the distribution and the complexity of the model.

In the simulation, we found that the selection of  $a_s$  was based on various factors such as the sample size, model complexity, and distribution. In practice, we can estimate  $a_s$  using the bootstrap procedure by assuming the null hypothesis and the model implied covariance are true. The selection of  $a_s$  and the DLS inferences are based on this assumption. In other words, the accuracy of estimated  $a_s$  depends on how well the model fits the data. If the null hypothesis is true, the bootstrapping results should be the same as our simulation results. If the null hypothesis deviates from the true data generating model, the estimated  $a_s$  is not the best  $a$  for the true model but the best  $a$  for the assumed model. As illustrated in the misspecification simulation, with a misspecified model,  $DLS_M$  with the estimated  $a_s$  generally performed better or was equivalent to the normal theory based methods,  $GLS_M$  and  $GLS_S$ , in terms of RMSE, the biases of SE estimates, and the empirical SEs. We think it is the best we can do given our model assumption. Additionally, when analyzing real data, we suggest conducting a sensitivity analysis as we did in the real data example. The estimated  $a_s$  should not change once the number of bootstrap samples is large enough. Another option to select  $a_s$  is to follow the work by Jiang (2018), Yang (2018), Yang and Yuan (2019), which constructs a mapping function between  $a_s$  and all data/model features. The selected  $a_s$  should be more accurate than that from the bootstrap procedure. But this approach requires an extensive simulation to consider a variety of data/model features.

Due to the scope and word limitation of this paper, we did not explore the performance of the proposed method with missing data (e.g., multiple imputation; Du et al., under review; Enders et al., 2020). Future work could look into investigating the performance of the proposed method with missing data.

In sum, our paper outlines a new distributionally-weighted least squares estimator,  $DLS_M$ , which works well with both normal and nonnormal data.  $DLS_M$  can provide more accurate and efficient estimates than classical methods, and the combination of  $T_{JY}$  and  $DLS_M$  provides acceptable Type I error rates.

### References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.  
doi: 10.1037/0033-2909.107.2.238
- Bentler, P. M. (2006). *Eqs 6 structural equations program manual*. Encino, CA: Multivariate Software.
- Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, *21*(2), 205–229. doi: 10.1177/0049124192021002004
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. doi: 10.1111/j.2044-8317.1978.tb00581.x
- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*(1), 1–24. doi: 10.1002/j.2333-8504.1973.tb00197.x
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. doi: 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*(5), 1716–1735. doi: 10.3758/s13428-016-0814-1
- Deng, L., Yang, M., & Marcoulides, K. M. (2018). Structural equation modeling with many variables: A systematic review of issues and developments. *Frontiers in Psychology*, *9*, 580. doi: 10.3389/fpsyg.2018.00580
- Du, H., Enders, C. K., Keller, B. T., Bradbury, T. N., & Karney, B. R. (under review). A bayesian latent variable selection model for nonignorable missingness.

- Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, 25(1), 88–112. doi: 10.1037/met0000228
- Gomer, B., Jiang, G., & Yuan, K.-H. (2019). New effect size measures for structural equation modeling. *Structural Equation Modeling*, 26(3), 371–389. doi: 10.1080/10705511.2018.1545231
- 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). New York: Elsevier. doi: 10.1016/s0731-9053(03)17003-x
- Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution (supplementary educational monograph no. 48)*. Chicago, IL: University of Chicago Press.
- Huber, P. J., et al. (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). Berkeley: 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*, 25(4), 541–557. doi: 10.1080/10705511.2017.1390753
- Jiang, G. (2018). *Ridge methods for confirmatory factor analysis of ordinal variables* (Unpublished doctoral dissertation). University Of Notre Dame.
- Jiang, G., & Yuan, K.-H. (2017). Four new corrected statistics for sem with small samples and nonnormally distributed data. *Structural Equation Modeling*, 24(4), 479–494. doi: 10.1080/10705511.2016.1277726
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4), 443–482. doi: 10.1007/bf02289658

- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202. doi: 10.1007/bf02289343
- Lai, K. (2019). A simple analytic confidence interval for cfi given nonnormal data. *Structural Equation Modeling*, 26(5), 757–777. doi: 10.1080/10705511.2018.1562351
- Lai, K. (2020). Confidence interval for RMSEA or CFI difference between nonnested models. *Structural Equation Modeling*, 27(1), 16–32. doi: 10.1080/10705511.2019.1631704
- Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when rmsea and cfi disagree. *Multivariate Behavioral Research*, 51(2-3), 220–239. doi: 10.1080/00273171.2015.1134306
- Lee, S.-Y., & Xia, Y.-M. (2006). Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data. *Psychometrika*, 71(3), 565–585.
- Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21(3), 369–389. doi: 10.1037/met0000093
- Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, 82(3), 533–558. doi: 10.1007/s11336-016-9552-7
- Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, 23(2), 318. doi: 10.1037/met0000122
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55(2), 463–469. doi: 10.1111/j.0006-341x.1999.00463.x



- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling and more. version 0.5–12 (beta). *Journal of Statistical Software*, 48(2), 1–36. doi: 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).
- 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).
- 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). Thousand Oaks, CA: Sage.
- Savalei, V. (2014). Understanding robust corrections in structural equation modeling. *Structural Equation Modeling*, 21(1), 149–160. doi: 10.1080/10705511.2013.824793
- Song, P. X.-K., Zhang, P., & Qu, A. (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions. *Statistica Sinica*, 929–943.
- Steiger, J. H. (2016). Notes on the steiger–lind (1980) handout. *Structural Equation Modeling*, 23(6), 777–781. doi: 10.1080/10705511.2016.1217487
- Tong, X., & Zhang, Z. (2012). Diagnostics of robust growth curve modeling using student’s t distribution. *Multivariate Behavioral Research*, 47(4), 493–518. doi: 10.1080/00273171.2012.692614
- Tong, X., & Zhang, Z. (2020). Robust Bayesian approaches in growth curve modeling: Using Student’s t distributions versus a semiparametric method. *Structural Equation Modeling*, 27(4), 544–560. doi: 10.1080/10705511.2019.1683014

- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.  
doi: 10.2307/1912526
- White, H., et al. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. doi: 10.2307/1912934
- Yang, M. (2018). *Optimizing ridge generalized least squares for structural equation modeling* (Unpublished doctoral dissertation). University Of Notre Dame.
- Yang, M., & Yuan, K.-H. (2019). Optimizing ridge generalized least squares for structural equation modeling. *Structural Equation Modeling*, 26(1), 24–38. doi: 10.1080/10705511.2018.1479853
- Yuan, K.-H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, 26(2), 177–198. doi: 10.1016/s0167-9473(97)00025-x
- Yuan, K.-H., & Bentler, P. M. (1998). Structural equation modeling with robust covariances. *Sociological Methodology*, 28(1), 363–396. doi: 10.1111/0081-1750.00052
- Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis*, 52(10), 4842–4858. doi: 10.1016/j.csda.2008.03.030
- Yuan, K.-H., & Chan, W. (2016). Structural equation modeling with unknown population distributions: Ridge generalized least squares. *Structural Equation Modeling*, 23(2), 163–179. doi: 10.1080/10705511.2015.1077335
- Yuan, K.-H., Fan, C., & Zhao, Y. (2019). What causes the mean bias of the likelihood ratio statistic with many variables? *Multivariate Behavioral Research*, 54(6), 840–855. doi: 10.1080/00273171.2019.1596060

- Yuan, K.-H., Jiang, G., & Cheng, Y. (2017). More efficient parameter estimates for factor analysis of ordinal variables by ridge generalized least squares. *British Journal of Mathematical and Statistical Psychology*, *70*(3), 525–564. doi: 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*(6), 673–698. doi: 10.1080/00273171.2017.1354759
- Yuan, K.-H., & Zhang, Z. (2012). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, *77*(4), 803–826. doi: 10.1007/s11336-012-9282-4
- Zhang, X., & Savalei, V. (2016). Bootstrapping confidence intervals for fit indexes in structural equation modeling. *Structural Equation Modeling*, *23*(3), 392–408. doi: 10.1080/10705511.2015.1118692
- Zhang, X., & Savalei, V. (2020). Examining the effect of missing data on RMSEA and CFI under normal theory full-information maximum likelihood. *Structural Equation Modeling*, *27*(2), 219–239. doi: 10.1080/10705511.2019.1642111
- Zhang, Z., Lai, K., Lu, Z., & Tong, X. (2013). Bayesian inference and application of robust growth curve models using Student's t distribution. *Structural Equation Modeling*, *20*(1), 47–78. doi: 10.1080/10705511.2013.742382

**Footnotes**

<sup>1</sup>Another type of robust procedure is to weight each observation by its distance from the center of the data to obtain parameter estimates and standard errors (Yuan & Bentler, 1998; Yuan & Zhang, 2012).

Table 1: 11 methods in the simulation

Method	Discrepancy/Loss function	SE Estimates
$DLS_S$	$[s - \sigma(\theta)]' \left( (1-a)\hat{\Gamma}_{ADF} + a\hat{\Gamma}_{N.S} \right)^{-1} [s - \sigma(\theta)]$	$a = 1$ Standard, $a \neq 1$ Sandwich
$DLS_M$	$[s - \sigma(\theta)]' \left( (1-a)\hat{\Gamma}_{ADF} + a\hat{\Gamma}_{N.M} \right)^{-1} [s - \sigma(\theta)]$	
$GLS_S$	$[s - \sigma(\theta)]' \left( \hat{\Gamma}_{N.S} \right)^{-1} [s - \sigma(\theta)]$	Sandwich
$GLS_M$	$[s - \sigma(\theta)]' \left( \hat{\Gamma}_{N.M} \right)^{-1} [s - \sigma(\theta)]$	
$WLS$	$[s - \sigma(\theta)]' \left( \hat{\Gamma}_{ADF} \right)^{-1} [s - \sigma(\theta)]$	
$RGLS_D$	$[s - \sigma(\theta)]' \left( (1-a)\hat{\Gamma}_{ADF} + a \times \text{diag}(\hat{\Gamma}_{ADF}) \right)^{-1} [s - \sigma(\theta)]$	
$RGLS_I$	$[s - \sigma(\theta)]' \left( (1-a)\hat{\Gamma}_{ADF} + a\mathbf{I} \right)^{-1} [s - \sigma(\theta)]$	
$LS$	$[s - \sigma(\theta)]' \mathbf{I} [s - \sigma(\theta)]$	
$MLS$		
$ML_{O.M}$	$tr [\mathbf{S}\Sigma^{-1}(\theta)] - \log  \mathbf{S}\Sigma^{-1}(\theta)  - p$	
$ML_{E.M}$		

Table 2: Simulation Conditions of  $N$ ,  $p$ , and  $m$ 

18 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	30	100, 300, 500, 1000				
4 Distributional Conditions						
	$Z_\xi$	$Z_\epsilon$	$r$	Skewness	Kurtosis	
Normal	$N(0, 1)$	$N(0, 1)$	-	0.995	0.999	
Elliptical	$N(0, 1)$	$N(0, 1)$	$(3/\chi_{df}^2)^{1/2}$	21.947	2.270	
Skewed Factor	<i>standardized</i> $(\chi_1^2)$	$N(0, 1)$	$(3/\chi_{df}^2)^{1/2}$	76.535	2.749	
Skewed Error	$N(0, 1)$	<i>standardized</i> $(\chi_1^2)$	$(3/\chi_{df}^2)^{1/2}$	208.130	4.068	

Note: The average skewness and kurtosis are calculated from the simplest model ( $m = 1$  and  $p = 5$ ) and the largest sample size ( $N = 1000$ ), based on 1000 replications. Following Yuan, Yang, & Jiang (2017), the multivariate skewness is calculated as  $\frac{1}{Np(p+1)(p+2)} \sum_{i=1}^N \sum_{j=1}^N [(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})]^3$  and the multivariate kurtosis is calculated as  $\frac{1}{Np(p+2)} \sum_{i=1}^N [(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^2$ . The population multivariate skewness and kurtosis should be 1 in the normal case.

Table 3: Real Data Example

	$MLE_M$			$GLS_M$			$DLS_M (a_s = 0.75)$			$RGLS_I (a_s = 0.36)$		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
$\lambda_{11}$	0.817	0.099	8.263	0.817	0.109	7.482	0.811	0.1	8.148	0.797	0.098	8.145
$\lambda_{12}$	0.541	0.1	5.426	0.541	0.094	5.778	0.531	0.088	6.059	0.516	0.085	6.073
$\lambda_{13}$	0.686	0.09	7.642	0.686	0.088	7.808	0.685	0.083	8.235	0.699	0.083	8.392
$\lambda_{19}$	0.458	0.089	5.126	0.458	0.103	4.444	0.503	0.095	5.278	0.507	0.088	5.778
$\lambda_{24}$	0.972	0.078	12.383	0.972	0.084	11.596	0.958	0.08	11.988	0.948	0.078	12.123
$\lambda_{25}$	0.96	0.083	11.631	0.96	0.083	11.541	0.956	0.079	12.092	0.94	0.078	12.031
$\lambda_{26}$	0.934	0.081	11.553	0.934	0.083	11.272	0.928	0.08	11.591	0.916	0.083	11.092
$\lambda_{37}$	0.705	0.09	7.853	0.705	0.084	8.383	0.669	0.078	8.549	0.639	0.081	7.917
$\lambda_{38}$	0.897	0.093	9.598	0.897	0.098	9.138	0.919	0.094	9.77	0.925	0.097	9.506
$\lambda_{39}$	0.45	0.09	5.032	0.451	0.094	4.812	0.401	0.088	4.54	0.405	0.082	4.91
$\psi_{11}$	0.652	0.117	5.551	0.652	0.16	4.078	0.615	0.137	4.494	0.598	0.129	4.619
$\psi_{22}$	0.933	0.122	7.634	0.933	0.142	6.586	0.933	0.131	7.111	0.929	0.124	7.478
$\psi_{33}$	0.603	0.096	6.254	0.603	0.096	6.251	0.601	0.092	6.565	0.568	0.094	6.046
$\psi_{44}$	0.313	0.065	6.4	0.313	0.066	4.735	0.318	0.063	5.05	0.31	0.067	4.62
$\psi_{55}$	0.419	0.072	4.847	0.419	0.072	5.83	0.395	0.063	6.247	0.387	0.066	5.9
$\psi_{66}$	0.408	0.069	5.824	0.408	0.077	5.329	0.399	0.073	5.476	0.402	0.075	5.37
$\psi_{77}$	0.565	0.096	5.913	0.565	0.083	6.778	0.575	0.078	7.343	0.577	0.08	7.245
$\psi_{88}$	0.289	0.118	5.865	0.289	0.13	2.224	0.179	0.121	1.477	0.104	0.134	0.773
$\psi_{99}$	0.476	0.065	2.448	0.476	0.076	6.274	0.471	0.073	6.485	0.473	0.071	6.683
$\phi_{12}$	0.554	0.081	6.86	0.554	0.092	6.026	0.583	0.085	6.844	0.603	0.081	7.477
$\phi_{13}$	0.393	0.103	3.804	0.393	0.113	3.488	0.411	0.105	3.923	0.362	0.105	3.465
$\phi_{23}$	0.239	0.095	2.511	0.239	0.118	2.023	0.233	0.115	2.02	0.244	0.113	2.16

### Figure Captions

*Figure 1.* Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  when  $N = 300$ ,  $p = 5$  and  $m = 1$

figure.1 *Figure 2.* Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  when  $N = 300$ ,  $p = 30$  and  $m = 3$

figure.2 *Figure 3.* Plot of average relative biases of the SE estimates of factor loadings and factor covariances depending on  $a$  when  $N = 300$ ,  $p = 5$  and  $m = 1$

figure.3 *Figure 4.* Plot of average relative biases of the SE estimates of factor loadings and factor covariances depending on  $a$  when  $N = 300$ ,  $p = 30$  and  $m = 3$

figure.4 *Figure 5.* RMSEs from the 11 methods when data are normal

figure.5 *Figure 6.* RMSEs from the 11 methods when data are elliptical

figure.6 *Figure 7.* Empirical SEs of factor loadings and factor covariances from the 11 methods when the factor are skewed

figure.7 *Figure 8.* Average relative biases of the SE estimates of factor loadings and factor covariances from the 11 methods when data are normal

figure.8 *Figure 9.* Average relative biases of the SE estimates of factor loadings and factor covariances from the 11 methods when data are elliptical

figure.9 *Figure 10.* Type I error rates of the Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ) from  $DLS_M$ ,  $DLS_S$ ,  $GLS_M$ , and  $GLS_S$

figure.10 *Figure 11.* RMSEs, relative biases of the SE estimates, and empirical SEs for elliptical data and equal factor loading assumption



figure.11 *Figure 12*. Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  in the real data example

figure.12

Figure 1: Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  when  $N = 300$ ,  $p = 5$  and  $m = 1$

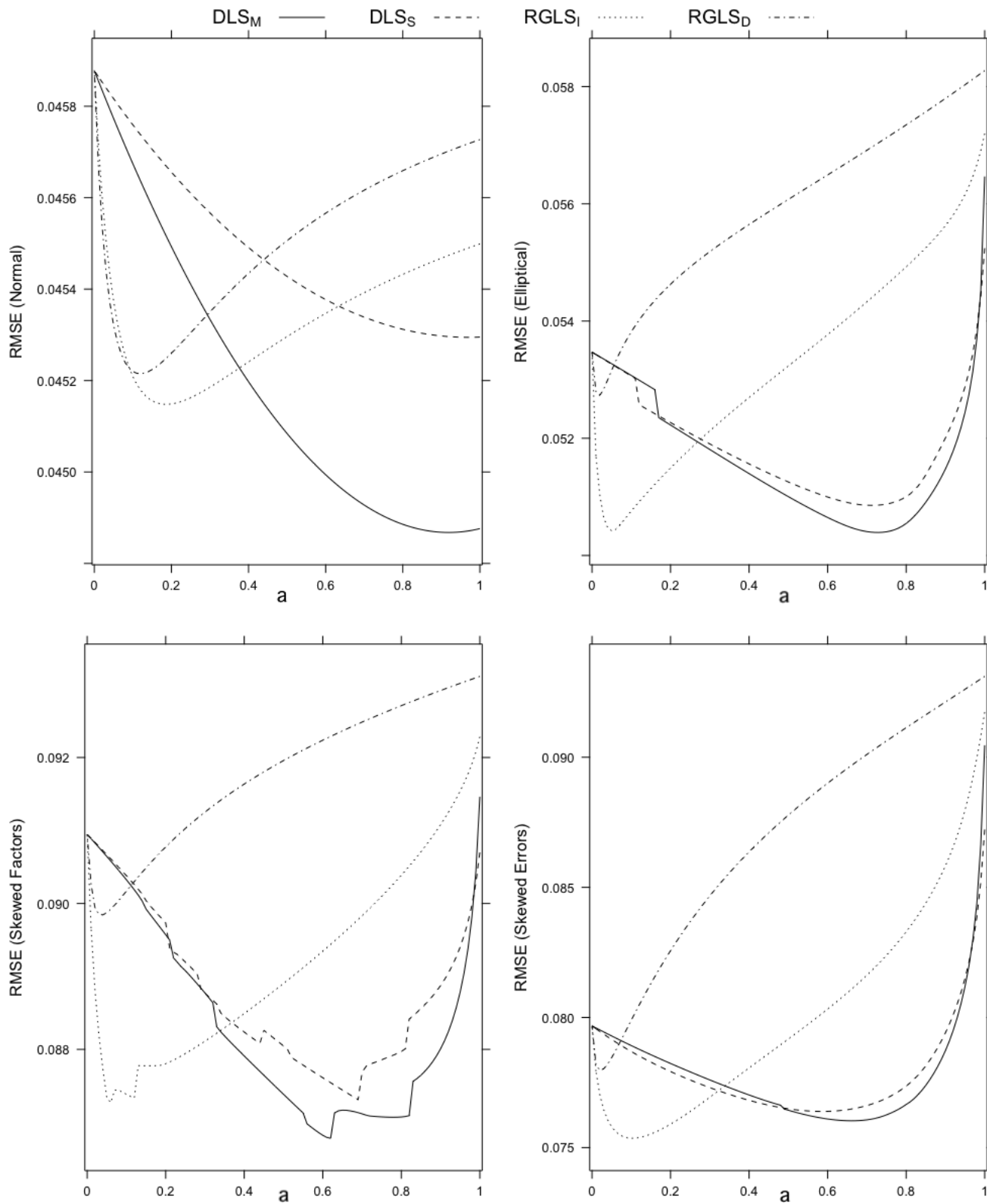


Figure 2: Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  when  $N = 300$ ,  $p = 30$  and  $m = 3$

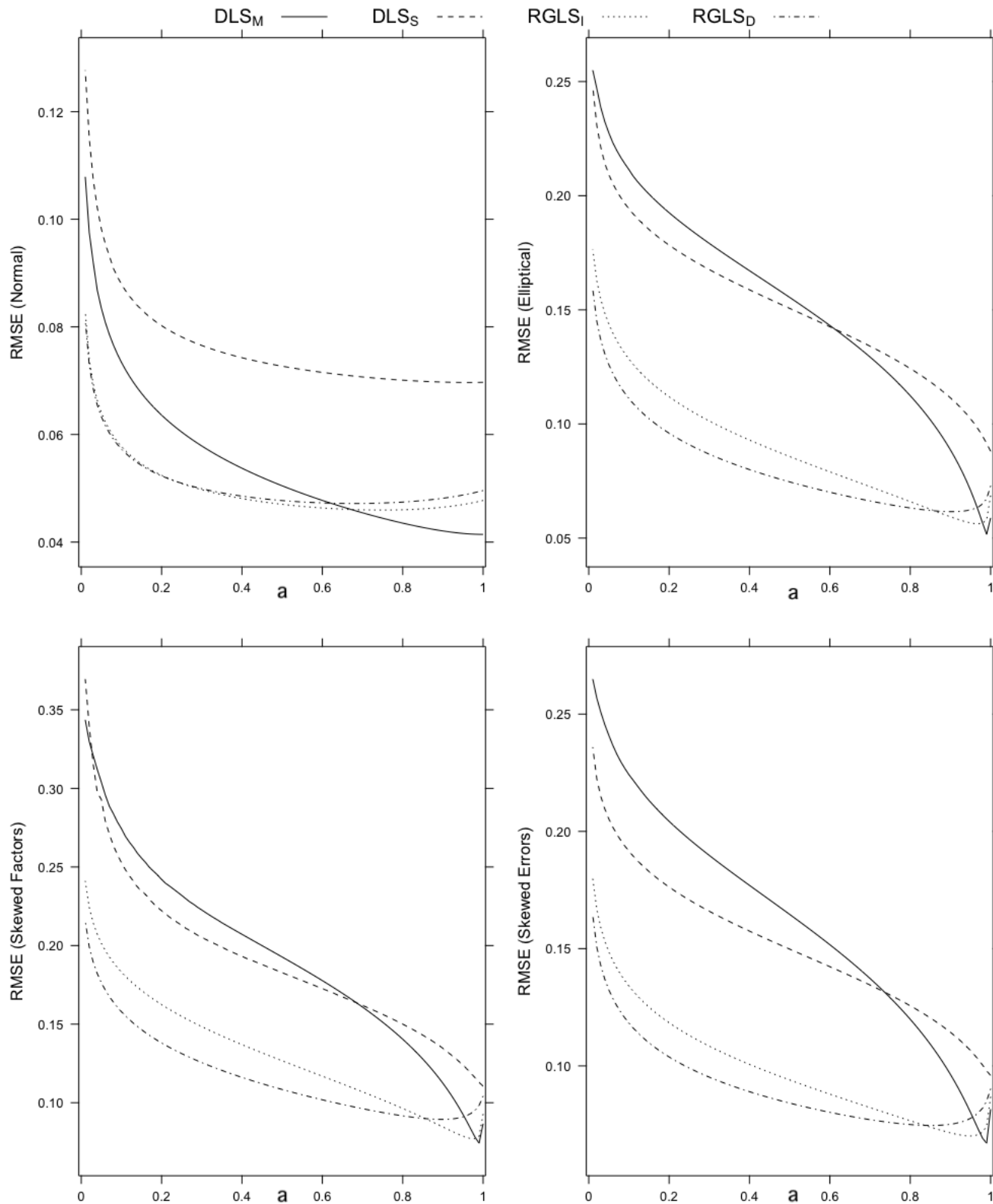


Figure 3: Plot of average relative biases of the SE estimates of factor loadings and factor covariances depending on  $a$  when  $N = 300$ ,  $p = 5$  and  $m = 1$

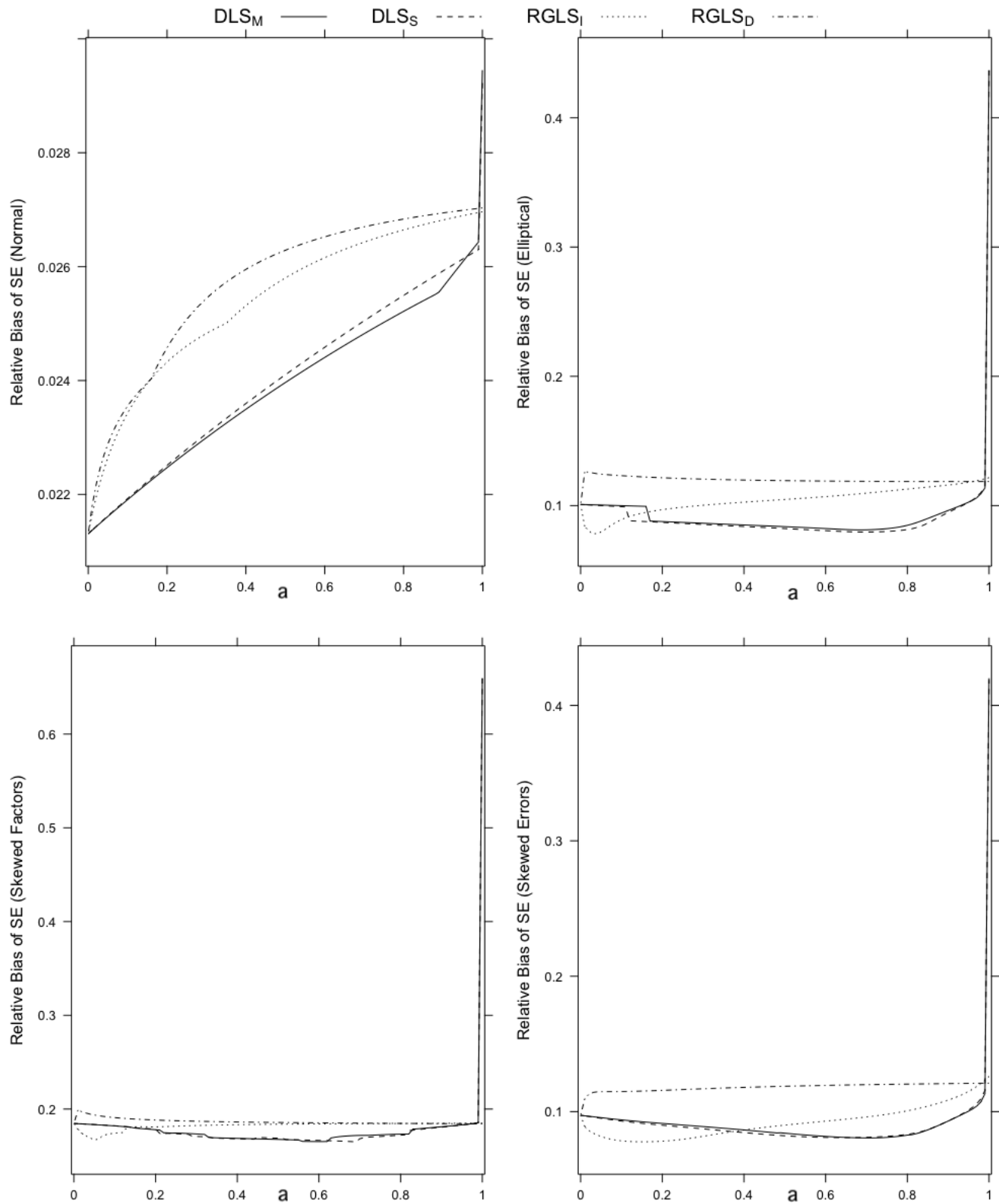


Figure 4: Plot of average relative biases of the SE estimates of factor loadings and factor covariances depending on  $a$  when  $N = 300$ ,  $p = 30$  and  $m = 3$

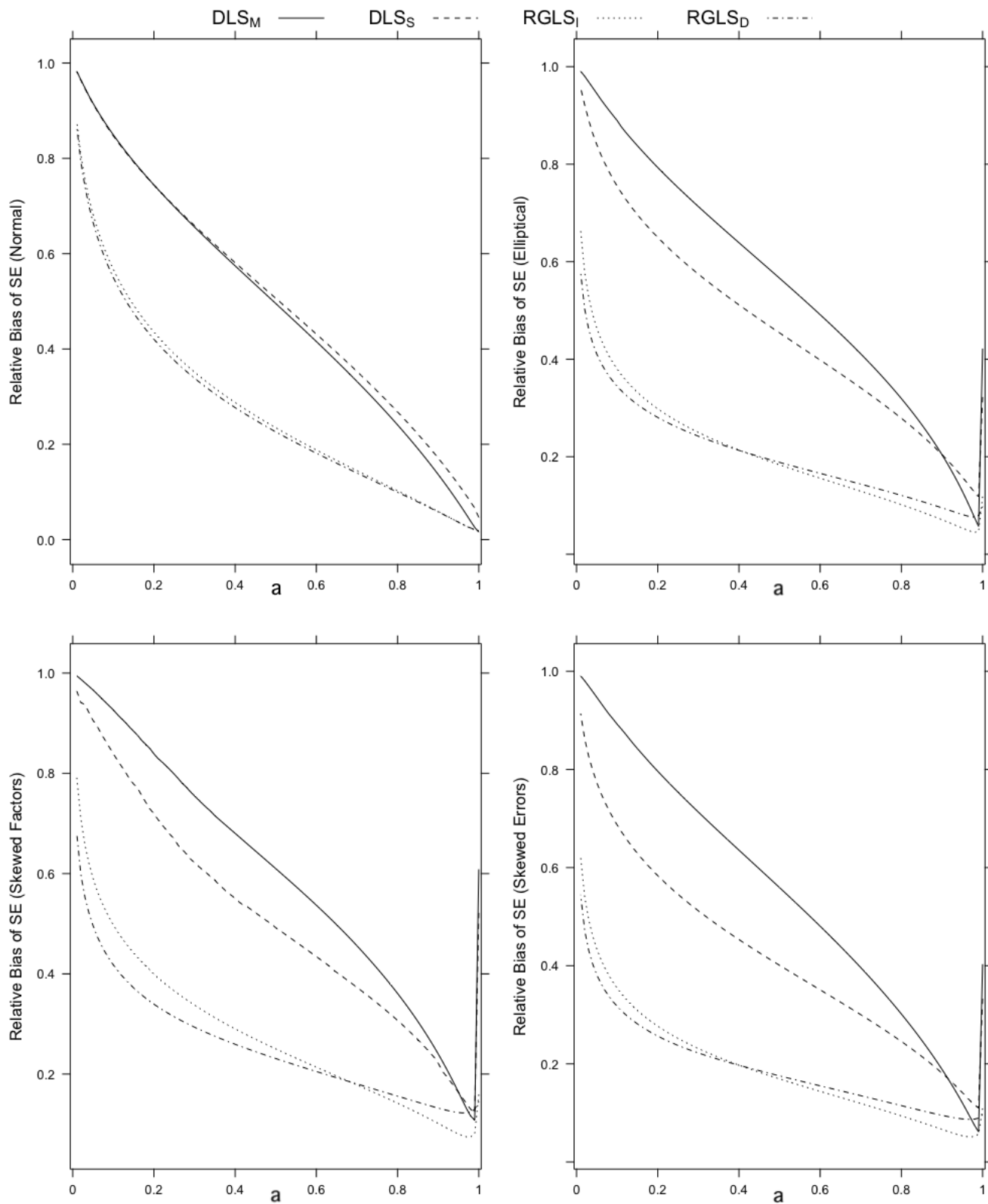
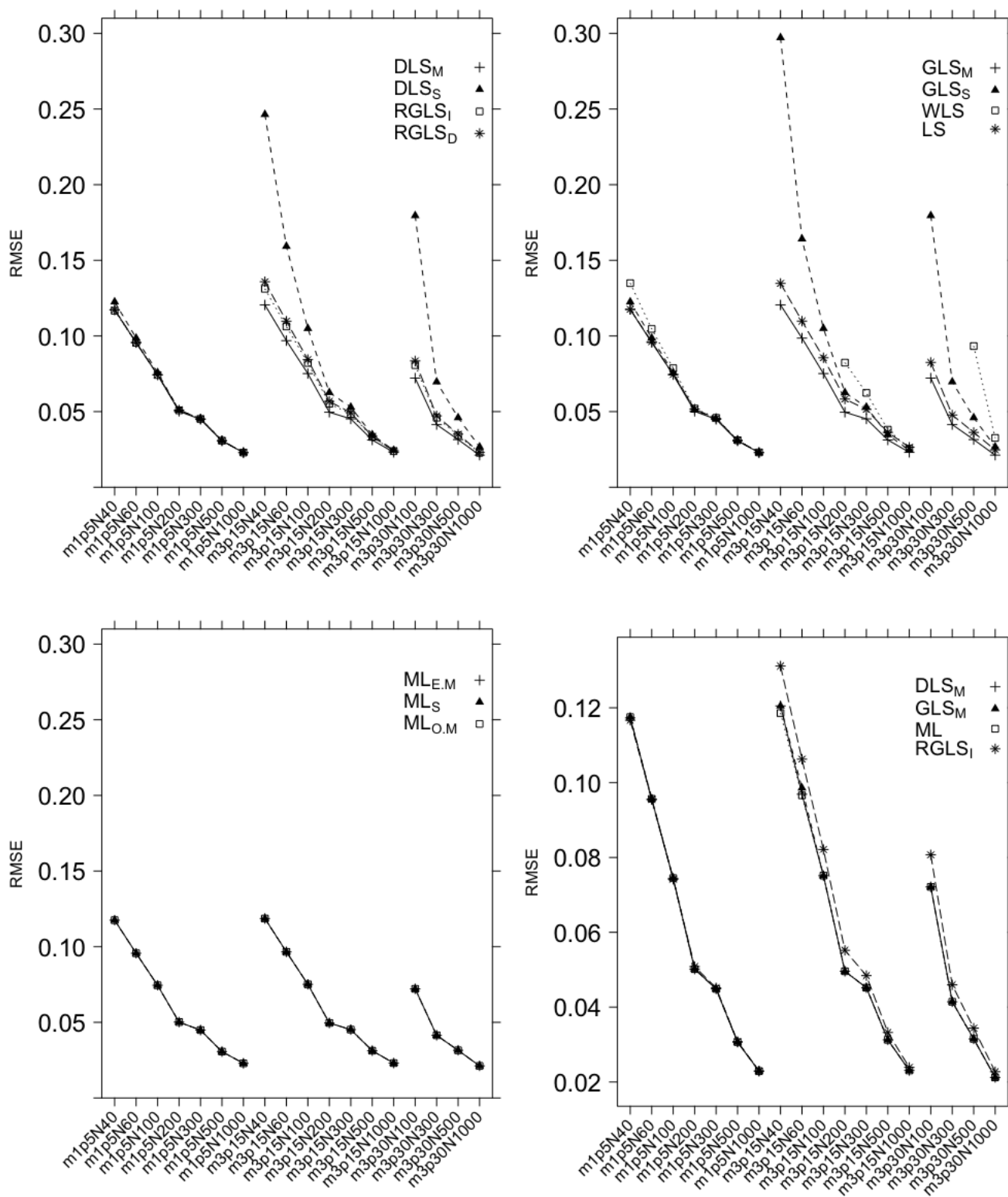
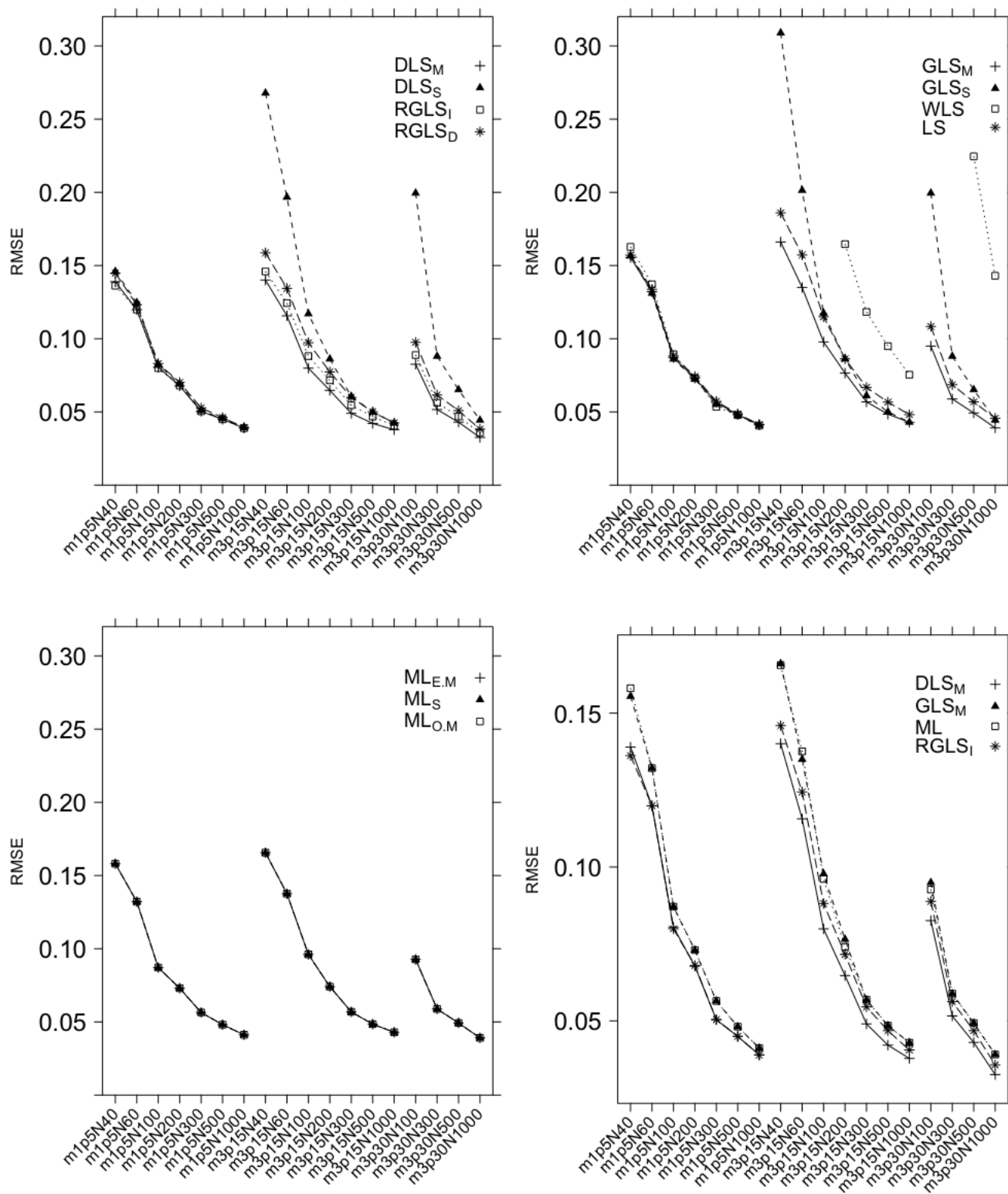


Figure 5: RMSEs from the 11 methods when data are normal



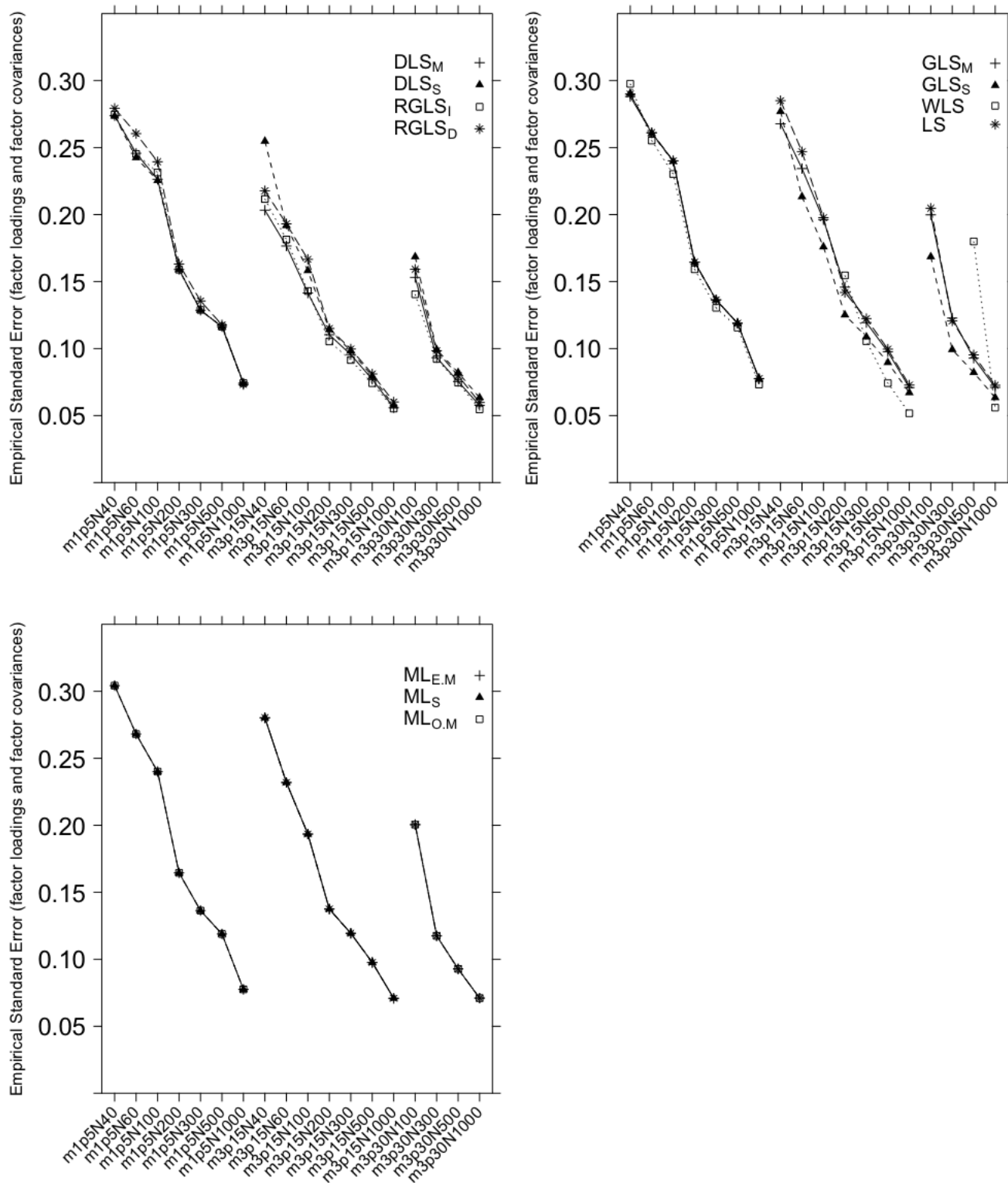
Note: The  $ML$  methods ( $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$ ) have the same parameter estimates but different standard errors, therefore their RMSEs are the same. The y-axis of the lower right panel is different from the other three panels.

Figure 6: RMSEs from the 11 methods when data are elliptical



Note: The  $ML$  methods ( $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$ ) have the same parameter estimates but different standard errors, therefore their RMSEs are the same. The y-axis of the lower right panel is different from the other three panels.

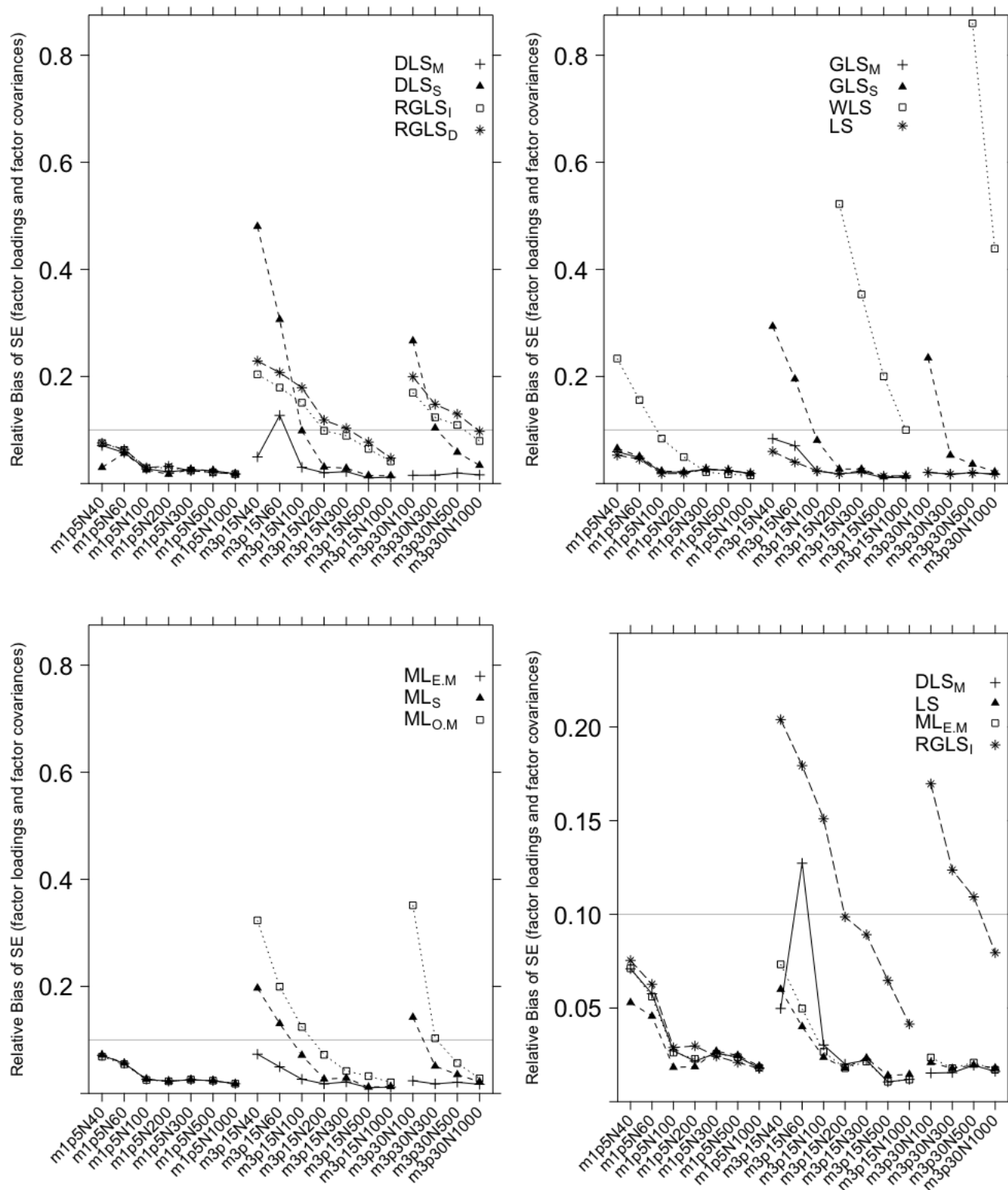
Figure 7: Empirical SEs of the SE estimates of factor loadings and factor covariances from the 11 methods when the factor are skewed



Note: The  $ML$  methods ( $ML_{O.M}$ ,  $ML_{E.M}$ , and  $ML_S$ ) have the same parameter estimates but different standard errors, therefore their empirical standard errors (SE) are the same. In this condition, the methods providing the minimal empirical SEs are  $DLS_M$ ,  $DLS_S$ ,  $RGLS_I$ , and  $RGLS_D$ . Because they are all in the upper left panel, we do not create a lower right panel to summarize these 4 methods again.

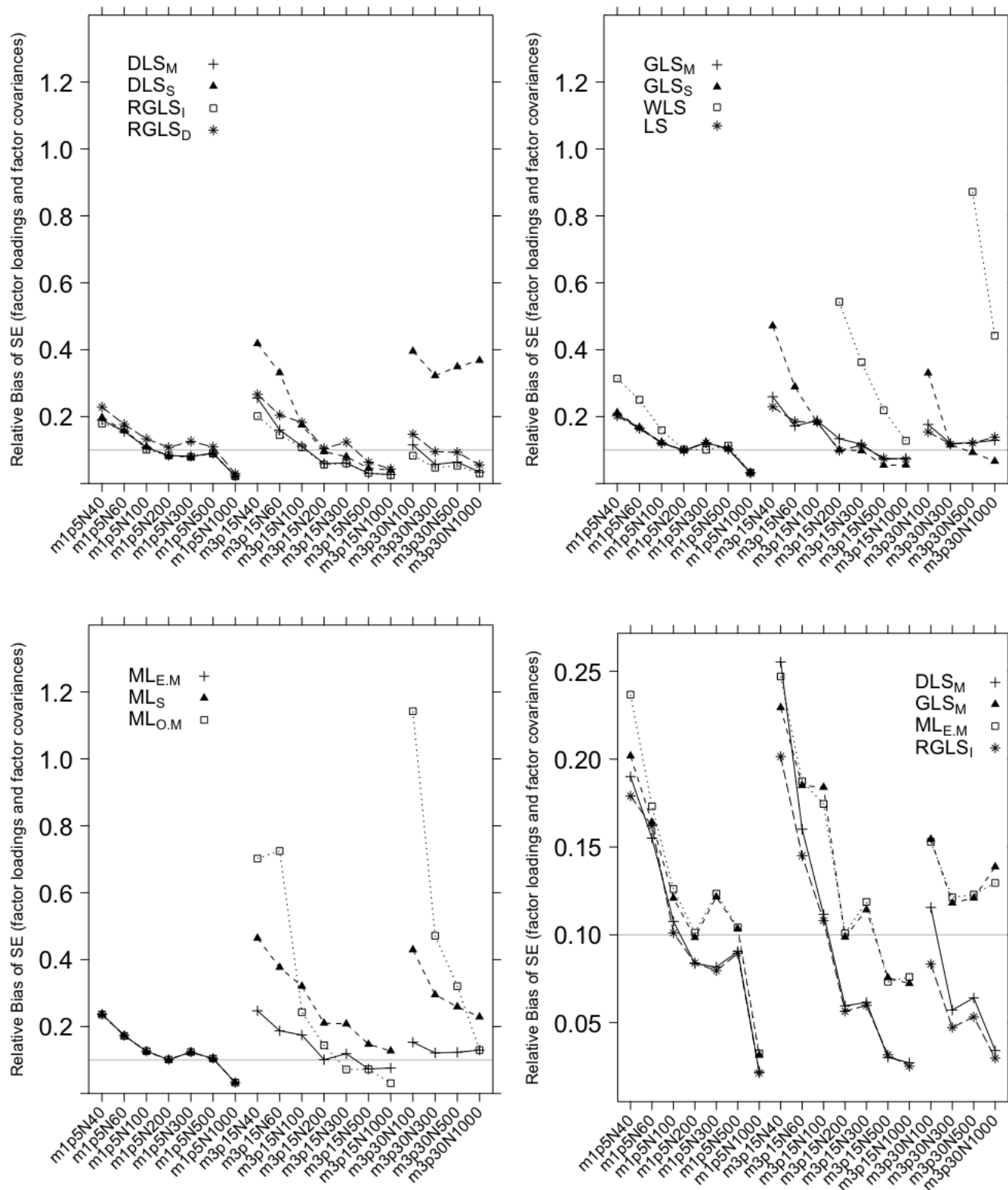


Figure 8: Average relative biases of the SE estimates of factor loadings and factor covariances from the 11 methods when data are normal



Note: The y-axis of the lower right panel is different from the other three panels. The grey line indicates a 10% relative bias threshold.

Figure 9: Average relative biases of the SE estimates of factor loadings and factor covariances from the 11 methods when data are elliptical



Note: The y-axis of the lower right panel is different from the other three panels. The grey line indicates a 10% relative bias threshold.

Figure 10: Type I error rates of the Jiang-Yuan rank adjusted test statistic ( $T_{JY}$ ) from  $DLS_M$ ,  $DLS_S$ ,  $GLS_M$ , and  $GLS_S$

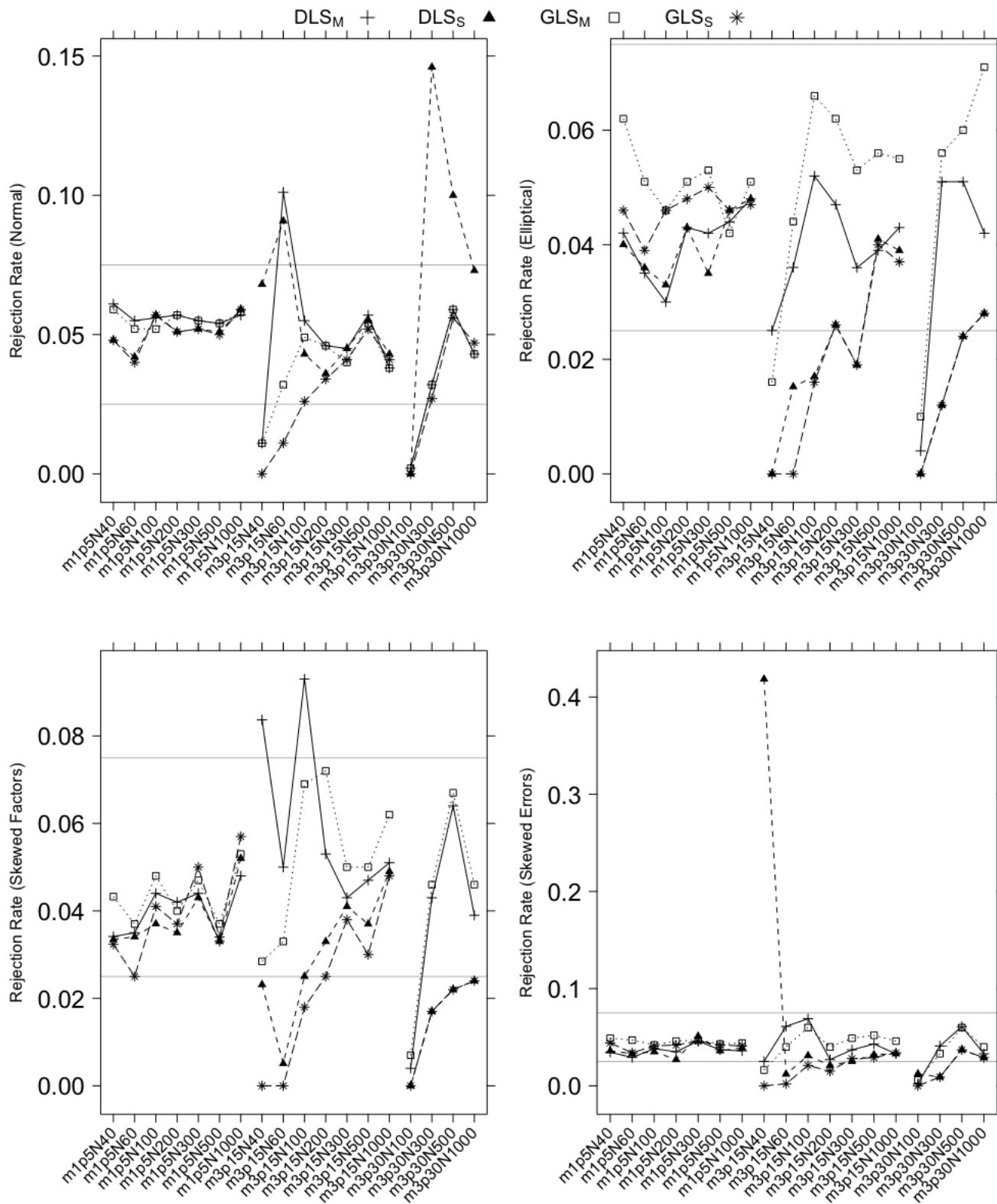


Figure 11: RMSEs, relative biases of the SE estimates, and empirical SEs for elliptical data and equal factor loading assumption

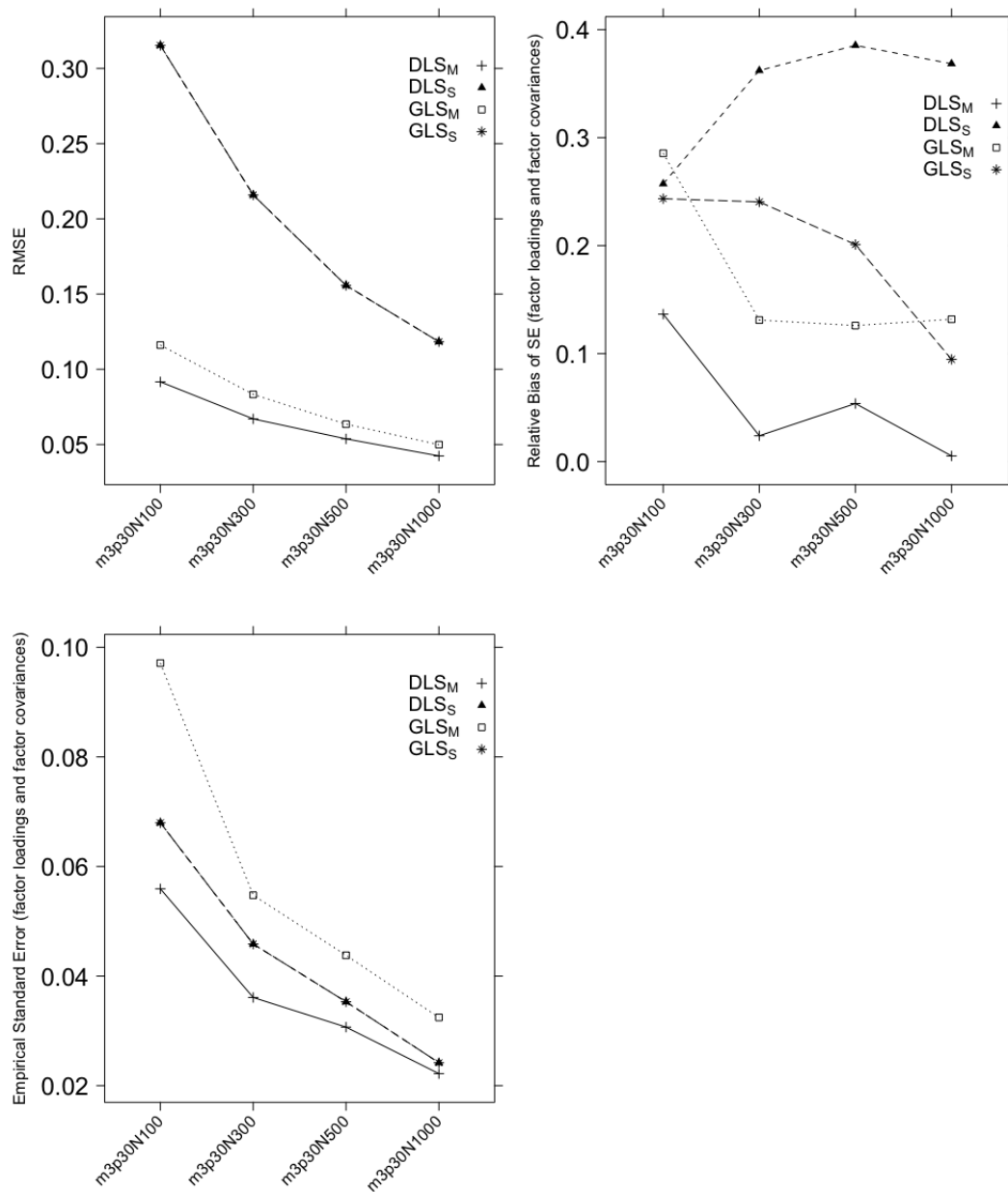


Figure 12: Plot of root mean square error (RMSE) of model parameter estimates depending on  $a$  in the real data example

