Title
Estimating Mean and Covariance Structure with Reweighted Least Squares

Author
Zheng, Bang Quan

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Estimating Mean and Covariance Structure with Reweighted Least Squares

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Bang Quan Zheng

2020

ABSTRACT OF THE THESIS

Estimating Mean and Covariance Structure with Reweighted Least Squares

by

Bang Quan Zheng

Master of Science in Statistics

University of California, Los Angeles

Professor Peter M. Bentler, Chair

Does Reweighted Least Squares (RLS) perform better in small samples than maximum likelihood (ML) for mean and covariance structure? ML statistics in covariance structure analysis are based on the asymptotic normality assumption; however, actual applications of structural equation modeling (SEM) in social and behavioral science research usually involve small samples. It has been found that chi-square tests often incorrectly over-reject the null hypothesis: $\Sigma = \Sigma(\theta)$, because when sample is small the sample covariance matrix becomes ill-conditioned and entails unstable estimates. In certain SEM models, the vector of parameter must contain both means, variances and covariances. Yet, whether RLS also works in mean and covariance structure remains unexamined. This research is an extended examination of

reweighted least squares in mean and covariance structure. Specifically, we replace biased covariance matrix in traditional GLS function (Browne, 1974) with the unbiased sample covariance matrix that derives from ML estimation. Moreover, under the assumption of multivariate normality, a Monte Carlo simulation study was carried out to examine the statistical performance as compared with ML methods in different sample sizes. Based on empirical rejection frequencies and empirical averages of test statistic, this study shows that RLS performs much better than ML in mean and covariance structure models when sample sizes are small.

The thesis of Bang Quan Zheng is approved.

Minjeong Jeon

Li Cai

Chad J. Hazlett

Peter M. Bentler, Committee Chair

University of California, Los Angeles

2020

To my parents,

鄭仕麟 & 許少媚

For their unconditional love and support

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Acknowledgement

Earning a Master of Science degree in statistics along the way to obtain my Ph.D. degree in political science fulfilled my passion for mathematical and statistical modeling. There are many people who helped me achieve this milestone, I owe them a deep debt of gratitude. The first person is my political science advisor Prof. John Zaller, who encouraged me to pursue a master's degree in statistics and suggested me to take a course in multivariate analysis with latent models with Prof. Peter Bentler. John would never have realized that I took his advice way too far. For me, covariance structure analysis is really fascinating; I was captivated by its complexity and versatility. In order to learn more about it, I also earned a graduate certificate in advanced quantitative methodology in educational research from UCLA Graduate School of Education and Information Studies (GSEIS), where I was fortunate to learn from MJ, Li Cai, Matt Madison (now at Clemson University), Noreen Webb, Michael Seltzer, and Mark Hansen. Particularly, MJ inspired me a lot in generalized item response tree model and cross-classified multilevel model. The most important person for my statistical training is my statistics advisor Prof. Peter Bentler, without his patient guidance this thesis would never have been realized. More profoundly, Peter's passion for teaching, meticulous attitude toward research, and genuine caring for students' learning left a deep imprint on my perception of academic role model whom I look up to.

1. Introduction

It is well documented that structural equation modeling is based on asymptotic properties, in which sample size and the observed variables are assumed to be large and multivariate normally distributed to follow a χ^2 distribution. Yet, in actual applications of covariance structure analysis in social science research, particularly in analyzing longitudinal data with growth curve modeling (GCM), this asymptotic assumption can be violated easily. This is because most social and behavioral data contain small or modest sample sizes and are not necessarily multivariate normally distributed (Hu, Bentler, & Kano, 1992; Yuan & Bentler, 1997). As a result, normal-theory methods such as maximum likelihood (ML) put forth by (Jöreskog, 1969) too often incorrectly reject the null hypothesis in χ^2 goodness-of-fit tests in the cases where model specifications are correct. The high frequent rejection rate renders a structural bias against small sample sizes (Arruda & Bentler, 2017; Hayakawa, 2019; Jalal & Bentler, 2018). Other scholars also pointed out that the primary reason of over-rejection problem is due to the number of the manifest variables; as a result when the size of covariance matrix is large, it is more likely to reject the null hypothesis (Moshagen, 2012; Shi, Lee, & Terry, 2018). Along with this perspective, some scholars also argue that the over-rejection problem is also due to the number of free parameters or degrees of freedom of the model (Herzog, Boomsma, & Reinecke, 2007; Hoogland & Boomsma, 1998; Jackson, 2003).

Over the last decades, scholars have proposed new corrected goodness-of-fit tests such as reweighted least squares (RLS) and regularized GLS (Arruda & Bentler, 2017). Lee (2007) also proposed that the sampling-based Bayesian methods may produce more reliable parameter

estimates when the sample size is small, because they rely less on asymptotic theory. Although these methods can effectively reduce the bias when sample sizes are small, the extent to which these methods can also be applied to mean and covariance structure[1] remain unexamined. Therefore, building on the existing reweighted least squares that introduced by Browne (1974), this research aims to test the performances of the RLS on mean and covariance structure.

The research method undertaking in this research is quite straightforward. It relies on Monte Carlo simulation to draw different sample sizes from N=50 to 10,000, to compare the performances of chi-square model fit statistics from different estimators of interest in both covariance structure, as well as mean and covariance structure. The chi-square tests that will be compared are from maximum likelihood (ML), generalized least squares (GLS), reweighted least squares (RLS), maximum likelihood mean and covariance structure (ML_{MS}) and reweighted least squares mean and covariance structure (RLS_{MS}) estimation. Based on Monte Carlo simulation and 1,000 replications on different sample sizes, we find that reweighted least squares method outperforms conventional ML on mean and covariance structure and offers highly consistent goodness-of-fit chi-square model tests across different sample sizes.

This study proceeds as follows. It first reviews the theories in covariance structure analysis and traditional estimators: Maximum Likelihood and Generalized Least Squares (GLS) method. The next section is the literature review of covariance structure in tandem with mean structure. Within this context we will introduce the reweighted least squares (RLS) and its relationship with GLS, as well as the application of RLS to estimate mean and covariance structure. The fourth section discusses data generation and the simulation procedures. The fifth section is about testing methods and results. The last section is discussion and conclusion.

[1] In this paper, I use the terms "mean structure" and "structured means" to convey the same meaning.

1.1 Bias and Sample Size

To demonstrate how sample size affects the covariance structure, one must understand how sample sizes affect eigenvalues. Following the example of Chi and Lange (2014), we conduct a series of Monte Carlo simulations by drawing 100×10 multivariate normal distributions on varying sample sizes, n=5, 10, 50, 250 and 500, and setting the population eigenvalues to be 1 for all samples. Based on these samples, we extract the sample covariance matrix $\boldsymbol{S}$ and compute the eigenvalues on each sample size. Because these covariance structure $\boldsymbol{S}$ are 10-dimensional (p=10), there are 10 eigenvalues in each simulated sample. To visualize the effects, the boxplots in Figure 1 are sorted from biggest eigenvalues to smallest. Across different sample sizes, we can see that large eigenvalues are increasingly inflated as the sample size decreases, while in the same processes the small eigenvalues are correspondingly deflated. When the sample size increases to 500, the eigenvalues are asymptotically collapsed to the expected value of 1.

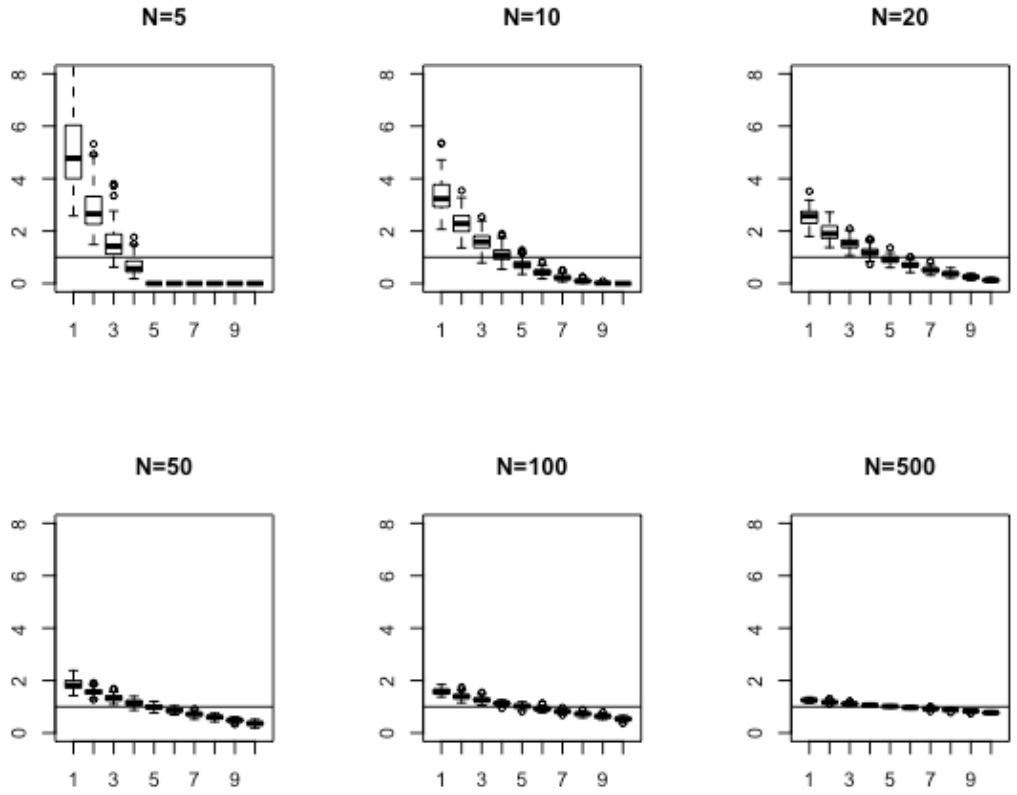

Figure 1. Multivariate simulation and sorted eigenvalues

This example illustrates that sample covariance structure is sensitive to sample size. Scholars have documented that when the number of variables is greater than the number of observations, that is, $p > N$, the sample covariance matrix is not invertible, in part because the covariance matrix may not be positive definite. Moreover, when the ratio of $\frac{p}{N}$ is less than 1 by slight amount, the sample covariance matrix can be ill conditioned, and generates a large condition number.[2] As a result, the parameter estimates can be unstable in covariance structure analysis (Arruda & Bentler, 2017; Chi & Lange, 2014; Hayakawa, 2019), as well as in mean and

[2] The condition number refers to the ratio between largest and smallest eigenvalues.

covariance structure (Bentler, 1989). In what follows we briefly introduce the statistical procedures that aims to correct these biases caused by small sample sizes.

2. Covariance Structure

In this section, we will first review the traditional covariance structure and the existing methods to derive the parameter estimates and model fit tests. Let $\{x_1, \ldots, x_n\}$ be a random sample of x, and all x_i are identically and independently distributed according to a multivariate normal distribution N $\mathbf{[0, \Sigma}_0]$. That is, we assume a random vector x $(p \times 1)$ has a mean 0 and a positive definite covariance matrix $\mathbf{\Sigma}_0$. Also, we assume that $\mathbf{\Sigma}_0$ is a matrix function of an unknown population parameters $\boldsymbol{\theta}_0$ $(q \times 1)$, and $\mathbf{\Sigma}_0 = \mathbf{\Sigma}(\boldsymbol{\theta}_0)$. Therefore, the sample covariance matrix is

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

where the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_1, \ldots, x_n)$. When the sample size n is significantly large the difference between $\frac{1}{n}$ and $\frac{1}{n-1}$ can be neglected. According to the Multivariate Central Limit Theorem (Anderson, 1984), the sample covariance matrix $\boldsymbol{S}$ is positive definite with probability 1, and $\boldsymbol{S}$ is an unbiased estimator of population covariance structure $\mathbf{\Sigma}_0$, and also converges to $\mathbf{\Sigma}_0$ in probability. That is,

$$n^{1/2}\{\mathrm{vec}(\boldsymbol{s} - \boldsymbol{\sigma})\} \xrightarrow{L} N\,[\mathbf{0}, 2\boldsymbol{K}_p'(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0)\boldsymbol{K}_p] \tag{2}$$

where '$\xrightarrow{L}$' denotes convergence in distribution. As $\boldsymbol{S}$ $(p \times p)$ is a symmetric matrix, the $\frac{p(p+1)}{2} \times 1$ vec($\boldsymbol{s}$) may be expressed in terms of the $p^2 \times 1$ vec($\boldsymbol{S}$), and similarly for $\sigma(\theta)$ and $\boldsymbol{\Sigma_0}(\theta)$. Thus $\boldsymbol{K}_p$ is of order $p^2 \times p(p+1)/2$. As Lee (2007) points out, the quadratic functions of $\{x_1, \dots, x_n\}$ can be considered as the covariances and variances of elements of the sample covariance matrix $\boldsymbol{S}$, which can be expressed in terms of $\boldsymbol{\Sigma}_0$. This is the most fundamental result for building the covariance structure analysis theory.

Moreover, in a confirmatory factor analysis (CFA) model

$$\boldsymbol{x_i} = \boldsymbol{\Lambda \xi_i} + \boldsymbol{\epsilon_i}, \qquad i = 1, \dots, n$$

where x_i is a random sample, $\boldsymbol{\Lambda}$ $(p \times q)$ is a matrix of factor loadings, $\boldsymbol{\xi_i}$ $(p \times 1)$ is a vector of latent factors, and $\boldsymbol{\epsilon_i}$ $(p \times 1)$ is a vector of residuals. In SEM, the parameters involved in the model are contained in the covariance matrix of the observed variable $\boldsymbol{\Sigma}$. $\boldsymbol{\Sigma = \Lambda \Phi \Lambda' + \Psi}$, where $\boldsymbol{\Lambda}$ again is a $p \times q$ factor loading matrix, and $\boldsymbol{\Phi}$ is a $q \times q$ covariance matrix of the factors, and $\boldsymbol{\Psi}$ is a $q \times q$ covariance matrix of unique scores. In SEM, the population covariance matrix $\boldsymbol{\Sigma}$ has a hypothesized structure $\boldsymbol{\Sigma = \Sigma(\theta)}$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a model implied covariance matrix, and $\boldsymbol{\theta}$ contains a vector of free parameters. In the following text $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ will be denoted by $\boldsymbol{\Sigma}$, should the context be clear. Since the sample covariance matrix $\boldsymbol{S}$ is an unbiased estimator of the population covariance matrix, an objective function $F[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{S}]$ measures the discrepancy between $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{S}$ (Browne, 1974; Jöreskog, 1969). Supposed that $\boldsymbol{\theta}' = \{\theta_1, \dots, \theta_q\}$ is an unknown parameter vector and we want to estimate $\boldsymbol{\theta}$ by minimizing a real objective function $F[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{S}]$. The major purpose of nonlinear programming is to locate $\widehat{\boldsymbol{\theta}}$, the best estimated value of $\boldsymbol{\theta}$ for which

$F[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{S}]$ is minimized. That is, we start with an initial and arbitrary guess $\theta_i \in \{\theta_1, \ldots, \theta_q\}$, and iteratively generate a sequence of $\theta_1, \ldots, \theta_q$ until it converges to $\widehat{\boldsymbol{\theta}}$, assuming that the second order partial derivatives of $F[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{S}]$ with respect to $\boldsymbol{\theta}$ exist and are continuous.

Furthermore, estimation of a discrepancy function $F[\boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{S}]$ involves the choice of iterative procedures. The Newton-Raphson and the Fletcher-Powell algorithms can be used for minimizing any arbitrary objective functions. Whereas the scoring algorithm is most appropriate for minimizing the likelihood function, while the Gauss-Newton algorithm is most appropriate for minimizing the GLS function (Lee, 2007).

In classical covariance structure analysis and based on multivariate normally distributed variables, the maximum likelihood (ML) and generalized least squares (GLS) are the most common methods to obtain the test statistics for evaluating the goodness-of-fit. In this study we use ML discrepancy function T_{ML} (Jöreskog, 1969) to derive the goodness-of-fit test statistic. As equation 3 shows, the model implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is fitted to the sample covariance matrix $\boldsymbol{S}$ using the Wishart likelihood function.

$$F_{ML}(\theta) = \log|\boldsymbol{\Sigma}\,(\boldsymbol{\theta}) - \log|\boldsymbol{S}| + tr(\boldsymbol{S}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}) - p \tag{3}$$

$$\widehat{\boldsymbol{\theta}}_{ML} = argmin\ F_{ML}(\boldsymbol{\theta}) \tag{4}$$

As equation 4 shows, at the minimum of the fit function $F_{ML}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{\theta}}_{ML}$ contains parameter estimates $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{\Psi}}$, where $\widehat{\boldsymbol{\Lambda}}$ is a matrix of estimated factor loadings, $\widehat{\boldsymbol{\Phi}}$ is a estimated factor covariance, and $\widehat{\boldsymbol{\Psi}}$ is the covariance matrix of error variables. Through these parameter estimates, we can reproduce the covariance matrix of the observed variables, that is, $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Lambda}}' + \widehat{\boldsymbol{\Psi}}$.

Moreover, the goodness-of-fit test statistic is defined by

$$T_{ML} = (N-1)F_{ML}(\widehat{\boldsymbol{\theta}}), \tag{5}$$

where T_{ML} is a test statistic, which is evaluated at the final parameter estimates, and multiplied by (N-1). As $N \to \infty$, T_{ML} should asymptotically follow a chi-square distribution with degrees of freedom $df = p^* - q$, where $p^* = p(p+1)/2$ and q is the number of free parameters. Chi-square test is a likelihood ratio test statistic for testing the model against the alternative model, in which the $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is unconstrained.

The function minimized by GLS is given by

$$\boldsymbol{Q} = \big(\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\big)' \boldsymbol{W} (\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))\,, \tag{6}$$

where $\boldsymbol{s}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ were defined previously, and $\boldsymbol{W}$ is a weight matrix. When $\boldsymbol{\theta}$ is a vector of parameter estimates, the optimal parameter estimates contained in $\widehat{\boldsymbol{\theta}}$ are those that minimize the function $\boldsymbol{Q}$. The test statistic *T* is usually computed as $(N-1)\boldsymbol{Q}\big(\widehat{\boldsymbol{\theta}}\big) \to \chi^2_{p^*-q}$, that is, when the sample size is large, the test statistic *T* will follow a chi-square distribution.

In covariance structure analysis, H_o: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\,(\boldsymbol{\theta})$, that is, population covariance matrix is equal to model implied covariance matrix. The critical value is given by the chi-square test with $p^* - q$ degrees of freedom. If the proposed model implied covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not rejected by $(N-1)\boldsymbol{Q}\big(\widehat{\boldsymbol{\theta}}\big)$, then it is considered a plausible model.

2.1 Mean and Covariance Structure

$$T^* = T_\Sigma + T_\mu \tag{7}$$

Let T_Σ be either T_{ML} or T. Then, the mean and covariance structure test statistic T^* consists of two parts T_Σ and T_μ, where T_μ derives from $(\bar{x} - \mu)'\hat{\Sigma}^{-1}(\bar{x} - \mu)$. We can use different estimating methods to derive mean and covariance structure parameter estimates. When modeling means and covariances, data and model vectors must take both of them into account (Bentler, 1989; Yuan, Zhang, & Deng, 2019). That is, if the covariance structure function T_Σ is inflated, T^* will be inflated, even if the structured means fit perfectly. On the flip side, a good covariance structure with badly modeled means will behave badly as well. Nonetheless, most SEM-based fit indices tend to concentrate on factor covariances and factor loadings, while ignoring mean structures (Wu & West, 2010).

Bollen & Curran (2005) put that, the overall goodness-of-fit test of the mean and covariance structure can be defined using ML fitting function estimator by setting the simultaneous null hypotheses

$$H_o: \boldsymbol{\mu} = \boldsymbol{\mu}(\theta) \text{ and } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$$

Nonetheless, according to Bollen and Curran (2006), there are two problems: 1) the observed variables might derive from multivariate distributions that exhibit excess kurtosis, which would cause the test statistic to be too high. 2) The power of the chi-square test, which can be sensitive with sample sizes. Yuan et al. (2019) proposed that we can use hybrid methods to model covariance structure and structured means separately, because the method that works well

in covariance structure part may not necessarily work efficiently for structured means part, and *vice versa*. Despite we agree with Yuan et al. (2019), the detailed discussions of fit indices for the structured means is beyond the scope of this paper. In this study, we only focus on the reweighted least squares method to test both mean and covariance structure in the context of multivariate normal data.

2.2 Reweighted Least Squares

In this section we propose to test the both mean and covariance structure jointly using reweighted least squares (RLS) and compare the results produced by maximum likelihood (ML). Reweighted least squares method is not novel to SEM; whereas its performance has not been tested and thoroughly compared with ML until recently by Hayakawa (2019). Using Monte Carlo simulation, Hayakawa (2019) finds that reweighted least squares (RLS) overcome the over-rejection problem and performs better than ML in the context of a confirmatory factor model, a panel autoregressive model and cross-lagged panel model. However, whether RLS also efficiently work in mean and covariance structure models remains unexamined in the field of SEM.

Reweighted least squares method was built based on the generalized least squares (GLS) fit function. As equation 6 shows, it is specialized to multivariate normal populations. The definition of the GLS function is motivated from the residual quadratic form:

$$\{\mathrm{vec}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\}'\{2n^{-1}\boldsymbol{K}_p'(\boldsymbol{\Sigma}_0\otimes\boldsymbol{\Sigma}_0)\boldsymbol{K}_p\}^{-1}\{\mathrm{vecs}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\} \quad (8)$$

where $\boldsymbol{\Sigma}_0$ is population covariance matrix, and $\boldsymbol{K}_p$ is of order $p^2 \times p(p+1)/2$. The inverse of $\boldsymbol{K}_p'(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\boldsymbol{K}_p$ is $\boldsymbol{K}_p^-(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)^{-1}\boldsymbol{K}_p^{-\prime}$, hence substituting it into equation 8, we have

$$n2^{-1}\{\text{vecs}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\}'\boldsymbol{K}_p^-(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)^{-1}\boldsymbol{K}_p^{-\prime}\{\text{vecs}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\}$$
$$= n2^{-1}\{\text{vec}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\}'\,(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\{\text{vec}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\}. \quad (9)$$

This residual quadratic form is a function of $\boldsymbol{\Sigma}_0^{-1}$. As $\boldsymbol{\Sigma}_0$ is unknown, $\boldsymbol{\Sigma}_0^{-1}$ should be replaced by $\boldsymbol{V}$ $(p \times p)$, which is a positive definite matrix. This GLS function (equation 10) was proposed by (Browne, 1974),

$$F_{GLS} = 2^{-1}[\text{vec}\,(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))'(\boldsymbol{V} \otimes \boldsymbol{V})\text{vec}(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))]$$
$$= 2^{-1}\,tr[\{(\boldsymbol{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\boldsymbol{V}\}^2] \quad (10)$$

where vec() is a matrix operator that transforms a matrix into a vector by staking rows of the matrix, and $\otimes$ is the Kronecker product. $\boldsymbol{V}$ can be constant definite matrix or a stochastic matrix that converges to a positive definite matrix (Lee, 2007), and $\boldsymbol{V}$ is a consistent estimator of $\boldsymbol{\Sigma}_0^{-1}$.

Following equation 10, at the minimum of $\hat{F}_{GLS}$, we will obtain parameter estimates $\hat{\boldsymbol{\theta}}_{GLS}$. The method of reweighted least squares is based on this platform and modify only $\boldsymbol{V}$. When $\boldsymbol{V}$ becomes an identity matrix, the GLS function reduces to a least squares function (Lee, 2007). Browne (1974) showed that when $\boldsymbol{V}=\boldsymbol{\Sigma}_0^{-1}$, the weighted least squares estimator shares most of the nice asymptotic properties of ML estimate such as consistency, asymptotic normality and efficiency. Lee and Jennrich (1979) show that the standard Gauss-Newton algorithm produces weighted least squares estimates that can produce maximum likelihood estimates in iteratively

reweighted form as well. Therefore, the test statistic of the reweighted least squares T_{RLS} can be expressed as

$$T_{RLS} = \frac{n}{2} tr\left\{\left(\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}}_{ML}\right)\widehat{\boldsymbol{\Sigma}}_{ML}^{-1}\right\}^2$$

$$= \frac{n}{2}\{\boldsymbol{S} - \widehat{\boldsymbol{\sigma}}\}' D_p'\left(\widehat{\boldsymbol{\Sigma}}_{ML}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}_{ML}^{-1}\right) D_p\{\boldsymbol{S} - \widehat{\boldsymbol{\sigma}}\} \quad (11)$$

Where $\widehat{\boldsymbol{\Sigma}}_{ML}^{-1}$ is an unbiased sample covariance matrix derived from maximum likelihood fit function as illustrated in equation 3, and is updated iteratively. $\widehat{\boldsymbol{\sigma}}$ =vech($\hat{\Sigma}_{ML}$), and $\boldsymbol{S}$ =vech(s), D_p is a $p^2 \times \frac{p(p+1)}{2}$ duplication matrix. That is, after we run the ML function, we extract the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}_{ML}^{-1}$ and replace the biased weight matrix $\boldsymbol{V}$ in equation 10, which becomes equation 11.

One of the most important properties of T_{RLS} is that it can efficiently estimate the test statistic across various sample sizes, and also asymptotically follow a chi-square distribution, that is, $T_{RLS} \xrightarrow{d} \chi^2_{df}$ as $N \to \infty$. Moreover, similar to T_{ML}, T_{RLS} also demonstrates the asymptotic robustness to non-normality (Amemiya & Anderson, 1990; Hayakawa, 2019). All these flexible properties are due to the endogenous relationship between T_{ML} and T_{RLS}, that is,

$$T_{ML} = T_{RLS} + B$$

$$B = n\sum_{k=3}^{\infty} \frac{1}{k} tr\left\{I_p - S\hat{\Sigma}^{-1}\right\}^k \quad (12)$$

Browne (1974) has shown that T_{ML} can be decomposed into T_{RLS} and a vanishing term B as indicated in equation 12, and both T_{ML} and T_{RLS} will follow a chi-square distribution as $N \rightarrow \infty$. As illustrated in Figure 1, the ratios of the dimensions of the estimated matrix and sample size, p/n, affect the behavior of eigenvalues, and the same ratios can also affect the magnitude of the second term B. Hayakawa (2019) points out that the second term B can be positive and negative values; however, when the dimension p is large, the vanishing term B is mostly positive. As p gets larger, the relative magnitude of B to the degrees of freedom also increases; however, the test statistic of T_{RLS} is closer to its expected value. Simply put, when sample size is small, the second term B will be positive, which means that T_{ML} will be too large, while T_{RLS} is about right. When sample size is sufficiently large, the B part will vanish, the best GLS estimator is asymptotically equivalent to the ML estimator. Therefore, equation 12 precisely explains why T_{RLS} can generate highly consistent test statistics across different sample sizes, and have a well behaved asymptotic distribution of the goodness-of-fit test statistics (Lee, 2007).

2.3 Mean Structure

Compared to covariance structure, mean structure is less intuitive. To illustrate the concept of mean structure, we can consider a simple regression model:

$$y = \alpha + \beta x + \epsilon.$$

According to Bentler (2006b), α is an intercept parameter, which helps define the mean of y; whereas it is not equal to the mean. Taking expectations of both sides, and assuming that $E(\epsilon) = 0$, we have

$$\mu_y = \alpha + \beta \mu_x,$$

where μ_y is the mean of y, and μ_x is the mean of x. In this case, μ_y can be considered in terms of the model parameters α, β, and μ_x. This decomposition of the mean of the dependent variable provides an illustration of structured means, in which the means of dependent variables are structured in terms of the means of independent variables μ_x and structural coefficients β. Hence, adding means extends a basic idea used in covariance structure analysis. Specifically, when incorporating a mean structure into the SEM models, we introduce the new parameters to α and μ_x into the same model.[3] Extension of this idea leads to the development of latent growth curve modeling, that is, parameterizing the mean and covariance structure of repeated measures of latent factors, in which the growth trajectories measured by latent factors represent the polynomial influences of time (Bentler, 2018; Yuan et al., 2019), whereas this is beyond the scope of the present study.

The structured means function $(\bar{\mathrm{x}}-\mu)'\hat{\Sigma}^{-1}(\bar{\mathrm{x}}-\mu)$ is the generalized least squares function, where $\bar{x}$ is a vector of sample means and μ is a vector of means of the observed variables that depends on parameters, and $\hat{\Sigma}^{-1}$ is a weight function. Similar to covariance structure part, we can minimize $(\bar{\mathrm{x}}-\mu)'\hat{\Sigma}^{-1}(\bar{\mathrm{x}}-\mu)$ for the parameters, and as $N \to \infty$ the test statistic will be close to the degrees of freedom. Because the test statistics of both the mean and covariance structure can independently follow a chi-square distribution when variables x are multivariate normal, we can treat the test statistic of mean and covariance structure as the sum of the test statistics of the mean and covariance structure parts.

To estimate μ, we need a vector of factors ξ. Like other random variables, each factor has a mean and variance, and may correlate with other factors. $E(\xi)$=μ_ξ is a vector of model-implied latent factor means. Therefore, taking the expectation of $X = \mu_x + \Lambda\xi + \epsilon$, we have

[3] For details, see *EQS Structural Equations Program Manual* (Bentler 1989: p165).

$$X = \mu_x + E(\Lambda\xi) + E(\epsilon). \tag{13}$$

Suppose that $\mu_x = 0$ and $E(\epsilon) = 0$, then $\mu = \Lambda\mu_\xi$. Based on this equation, we can reconstruct estimated $\hat{\mu}$ of the observed variables, which is vector of p x1, the dimension of Λ is a p x q matrix, and ξ is a $q \times 1$ vector. Thus we can rewrite the structured means function $(\bar{\mathrm{x}} - \mu)'\hat{\Sigma}^{-1}(\bar{\mathrm{x}} - \mu)$ as

$$\left(\bar{x} - \Lambda\mu_\xi\right)'\hat{\Sigma}^{-1}\left(\bar{x} - \Lambda\mu_\xi\right) \tag{14}$$

The evaluation of the structured mean model fit is based on residual moments. The chi-square test statistic offers an overall assessment of the degree to which the model implied moments fit the moments of the observed variables. That is, structured mean models reflects how closely the sample mean vector $\bar{x}$ is reproduced by the estimated population mean vector $\hat{\mu}$, that is, $\left(\bar{x} - \Lambda\mu_\xi\right)$, as well as how closely the sample covariance matrix $\boldsymbol{S}$ is reproduced by the estimated model covariance matrix $\widehat{\boldsymbol{\Sigma}}$, that is, $\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The ways in which we derive the test statistic is to minimize equation 14, where $\widehat{\boldsymbol{\Sigma}}^{-1}$ is obtained from the ML fit function as shown in equation 3. Therefore, the asymptotic distribution of $F\left(\widehat{\boldsymbol{\theta}}\right)$ is a chi-square test with $p^* - q$ degrees of freedom, that is,

$$(N-1)F(\widehat{\boldsymbol{\theta}}) \rightarrow \chi^2_{p^*-q} \tag{15}$$

3. Data Generation and Simulation

In this study, the data generation scheme follows a traditional confirmatory factor model

$$\boldsymbol{X_i} = \boldsymbol{\Lambda\xi_i} + \boldsymbol{\epsilon_i} \tag{16}$$

where $\boldsymbol{X_i} = \left(X_{i1}, X_{i2}, \ldots X_{ip}\right)'$ is a vector of p observations on person i in a population, and $i = 1,2,\ldots n$. $\boldsymbol{\Lambda}$ is a matrix of factor loadings, and $\boldsymbol{\epsilon_i} = (\epsilon_{i1,}\epsilon_{i2,} \ldots . \epsilon_{ip})'$ is a vector of error terms, and $\text{var}(\epsilon) = \boldsymbol{\Psi}$. $\boldsymbol{\xi_i} = (\xi_{i1,}\xi_{i2,\ldots}\xi_{im})'$ is a vector of latent factors, and $\text{var}(\xi) = \boldsymbol{\Phi}$. Each latent factor $\boldsymbol{\xi_i}$ has a mean and a variance, and may correlate with other latent factors $\boldsymbol{\xi_i}$; whereas $\boldsymbol{\xi_i}$ and $\boldsymbol{\epsilon_i}$ are uncorrelated, so that $E(\xi) = \mu_\xi$, which is the mean of the factors. Moreover, using the notation of the covariance structure, the population matrix is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi} \tag{17}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, and $\boldsymbol{\theta}$ contains a vector of parameters μ_ξ, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$.

Moreover, the means of observed variables can be simulated according to equation 16. Suppose that we take the expectation on both sides of the equation, and assume $E(\epsilon)$=0, we derive

$$E(x) = E(\Lambda\xi) + E(\epsilon) \tag{18}$$
$$E(x) = \mu = \Lambda\mu_\xi$$

where $\mu = \Lambda\mu_\xi$ indicates that the means of the observed variables are functions of the means of the latent variable. In classical covariance structure models, structured means are not of interest, latent factor means are usually assumed to be zero, $\mu_\xi = 0$, and μ does not have structure as equation 18. Hence then $\bar{X} = \hat{\mu}$.

This implies that the ML test statistic in equation 5 can add the mean structure part:

$$T_{ML^*} = T_{\Sigma_{ML}} + (n-1)\left(\bar{x} - \Lambda\mu_{\xi}\right)' \hat{\Sigma}^{-1}\left(\bar{x} - \Lambda\mu_{\xi}\right) \tag{19}$$

where T_{ML^*} is the test statistics of mean and covariance structure, and $T_{\Sigma_{ML}}$ is the covariance structure that derives from the ML function. The structured means part is a quadratic function that is formed by residual matrix of $\left(\bar{x} - \Lambda\mu_{\xi}\right)$ and weight function is inverse of the model covariance matrix $\boldsymbol{\Sigma}$. Therefore, when we have simulated data, we will derive sample means and sample covariances. At the minimum of above ML fit function F_{ML} (See Equation 3), $\widehat{\boldsymbol{\theta}}_{ML}$ will contain parameter estimates $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\xi}}$, $\widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{\Sigma}}$.

Following the above data generation scheme, we can determine our population from which we draw samples using Monte Carlo simulation. We have the following factor loading and factor correlation matrix. Where there are three latent factors, and each comes with five indicators, which yield 15 variables. The factor loading $\boldsymbol{\Lambda}'$ is defined as

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

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & & \\ 0.3 & 1 & \\ 0.4 & 0.5 & 1 \end{bmatrix}$$

When diag(Σ)=I, the unique variances can be determined as suggested by $\boldsymbol{\Psi} = \mathrm{I}_{15\mathrm{x}15} - \mathrm{diag}(\boldsymbol{\Lambda\Phi\Lambda}')$. We also set $\mu's$=(0, 0, 0). The data generating process consists of two steps. 1) We draw from a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Phi}$, as indicated in equation 17. Unique factors $\boldsymbol{\epsilon_i}$ are drawn from a multivariate normal distribution

with zero mean and covariance $\boldsymbol{\Psi}$. This procedure generates multivariate normal observations with mean and covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

This population model has been used by Hu et al. (1992), Huang and Bentler (2015), Arruda and Bentler (2017), and Jalal and Bentler (2018) in their research of covariance structure analysis. Adopting the same population model in existing research by other scholars offer a means of validating results from simulations. Nevertheless, the choice of other population models will not change the statistical performance.

Data generation was accomplished with '*lavaan*' package (Rosseel, 2012) in R, based on the aforementioned population, under the assumption of multivariate normality. The simulation studies include the numbers of sample sizes range from 50 to 10,000, which is enough to examine the performance of different estimators. The results from R programming, parameter estimates and test statistics were verified by EQS software Version 6.4 (Bentler, 2006a)

For the covariance structure part, our testing models have 15 observed variables (p=15), 3 latent factors. Thus, the covariance component $p^* = \frac{15(15+1)}{2} = 120$, with 33 free parameters to estimate. Our testing models, therefore, have 87 degrees of freedom. For structured means part, there are 15 sample means derived from 15 observed variables, and 3 factor means. Thus, it has 12 degrees of freedom. Altogether the mean and covariance structure have 135 data points and 36 free parameters, which yields 99 degrees of freedom. Under the asymptotic properties and multivariate normality, the expected test statistic of covariance structure should be about 87, while the expected test statistic of the mean and covariance structure should be around 99.

4. Testing Method

The Monte Carlo simulation will be conducted to compare the reweighted least squares and ML at varying sample sizes range from 50 to 10,000. Both 'lavaan' and EQS program (Bentler, 2006a) can estimate ML mean and covariance structure jointly, whereas there is no statistical software or R packages that directly estimate reweighted least squares in mean and covariance structure. Therefore, we have to write the R code for computing RLS test statistics.

We will focus on two tests: Test statistic and empirical rejection frequency by different sample sizes. Each of 1,000 independent normal samples of each sample size will be analyzed by employing maximum likelihood (T_{ML}), generalized least squares (T_{GLS}), reweighted least squares (T_{RLS}), ML in mean and covariance structure (T_{ML_MS}), and reweight least squares in mean and covariance structure (T_{RLS_MS}) estimator methods. In this study, T_{GLS} is included for covariance structure analysis and for reference purpose only, and we will not test mean and covariance structure employing T_{GLS}.

For covariance structure models, we will conduct 1,000 replications, and calculate the means and standard deviations on all test statistics. The correct mean test statistic reference is 87, because the degrees of freedom is 87 and $p^* = \frac{15\times16}{2} = 120$. Thus, the average of 1,000 replications, the test statistics across different estimators will follow the asymptotic normality and a chi-square distribution, $T \xrightarrow{d} \chi^2_{87}$, when sample size is large. The expected mean standard deviation of 1,000 replications is $\sqrt{2df} \approx 13.19$. Similarly, for mean and covariance structure model, the average test statistics for T_{ML_MS} and T_{RLS_MS} should be about 99, because $df = 99$ and $p^* = 135$, and the expected mean standard deviation of 1,000 replications is $\sqrt{2df} \approx 14.07$.

Moreover, p-values are the criteria by which the null hypothesis is being rejected at α=0.05. Each replication will generate a corresponding p-value of the fitted model. The average

p-values of all models will be calculated. Also, considering the variability in p-values, we will also use empirical rejection frequencies as one of the benchmarks by which to evaluate the performances of the estimation methods. Empirical rejection frequencies are the ratio of the number of p-values that are less than 0.05 with total number of replication (1,000). Specifically, it is calculated by summing number of replications for which the model is rejected based on significance level $\alpha = 0.05$, and divided by the number of replications. Across all estimators, if the models perform correctly and follow asymptotic properties, their mean rejection rates should be around 0.05 when the sample sizes are sufficiently large. Any deviations far from α level 0.05 indicate that chi-square distribution is not an adequate reference distribution for evaluating model fit.

4.1 Results

Test Statistics

	Test Statistics					Standard Deviation				
N	ML	GLS	RLS	ML MS	RLS MS	ML	GLS	RLS	ML MS	RLS MS
50	102.32	76.71	87.65	115.9225	97.13	15.53	10.51	12.04	16.58	12.43
80	96.39	80.70	87.22	108.4681	97.34	14.71	11.78	13.36	15.31	12.38
100	93.66	82.21	87.10	106.1221	97.16	14.77	12.32	12.55	15.18	13.67
200	89.86	83.86	87.42	103.0728	98.97	13.99	12.97	13.71	15.04	13.81
300	89.55	85.50	87.40	101.5591	98.71	14.06	12.65	12.81	14.66	14.68
400	88.59	85.89	87.31	100.414	98.69	13.10	12.62	13.16	13.79	13.93
500	88.54	86.27	87.23	100.4166	99.19	13.76	13.04	12.79	14.08	14.27
800	87.45	86.20	87.14	100.5286	98.43	13.57	12.68	13.52	14.18	13.72
1,000	88.06	86.21	86.99	99.83	99.68	13.51	12.87	13.06	13.94	14.03
2,000	87.40	87.07	87.13	98.99	99.34	12.86	12.96	12.87	13.72	13.82
5,000	87.38	87.12	87.07	98.94	99.04	13.44	12.94	12.96	14.28	14.46
10,000	86.58	87.01	86.54	98.75	98.56	12.95	12.71	13.00	14.61	13.93

Note: GLS contains 37 non-convergence when N=50. 1 non-convergence when N=80

Table 1. Mean test statistics and standard deviations by sample size

Table 1 shows that T_{ML} tends to work efficiently and follows two asymptotic properties when the samples are greater than 400. First, the test statistics converge to the expected value of 87 as the

sample size becomes large. Whereas when sample sizes are less than 400, the test statistics start to deviate from the expected value. When sample sizes are less than 100, the rate of deviation increased radically, as evident that when N=50, the mean test statistics increased to 102.32. The test statistics of T_{GLS} also follow asymptotic behavior when sample sizes are large; however, when sample sizes are smaller than 500, they are increasingly biased. Quite opposite from T_{ML}, when N=50, the mean test statistics of T_{GLS} decreased to 76.71. In sharp contrast, the mean test statistics of T_{RLS} are highly consistent across all samples. As Table 1 shows, when the sample size varies from 50 to 10,000, the corresponding mean test statistics are still very close to the expected value of 87. These findings are consistent with those of Hayakawa (2019) in his study.

The mean and covariance structure models follow similar patterns. The mean of test statistics T_{ML_MS} asymptotically converge to the expected value of 99 when sample sizes are greater than 1,000. When sample sizes are smaller, the mean estimates become increasingly inaccurate. Specifically, when N=50, the mean test statistic increased to 115.92. The rate of deviation from the expected value is slightly more than that of T_{ML}. In contrast, the mean test statistics of T_{RLS_MS} are highly stable across all sample sizes, which are very close to the expected value of 99.

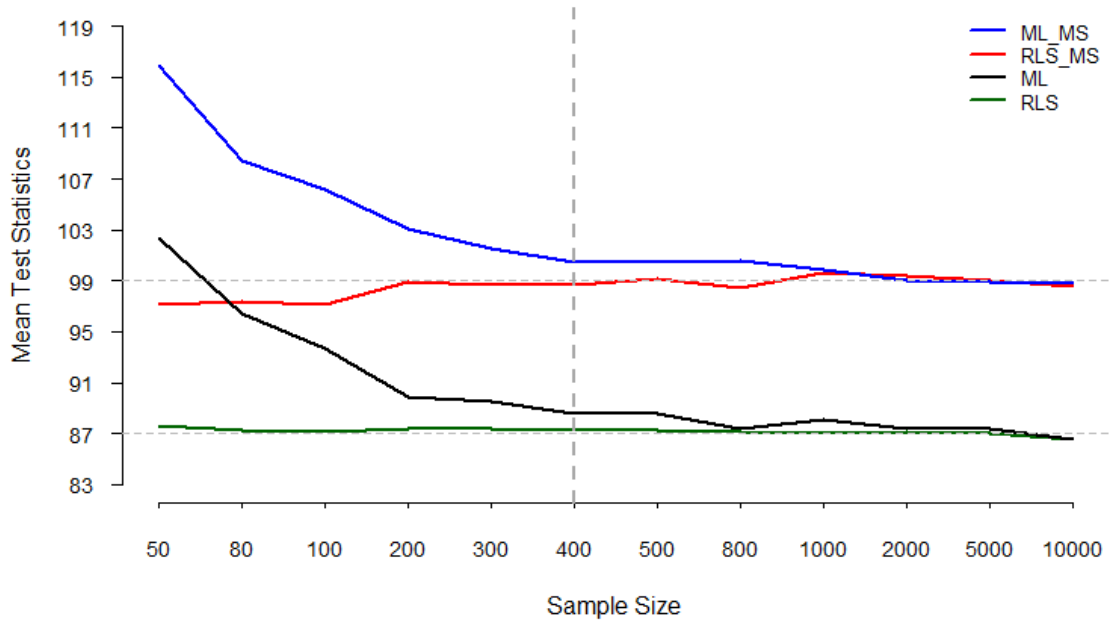

Figure 2. The effect of sample size on mean test statistics

Figure 2 summarizes mean the test statistics of all the estimation methods in Table 1, except for T_{GLS}. At a glance, the test statistics of T_{ML} and T_{ML_MS} are highly parallel to each other across all samples. The test statistics of T_{RLS} and T_{RLS_MS} are also similar to each other, and T_{RLS_MS} tends to be slightly less consistent when sample sizes are than 200, as compared to T_{RLS}.

In terms of mean standard deviation derived from the 1,000 replications, we expect them to be about 13.19 and 14.07 for covariance structure alone and mean and covariance structure respectively. Table 1 shows that all models meet our expectations on standard deviations when N>400. However, when sample sizes are smaller, the mean standard deviations of T_{ML} and T_{ML_MS} start to become larger than those of T_{GLS}, T_{RLS}, and T_{RLS_MS}. In sharp contrast, the mean standard deviations of T_{RLS} and T_{RLS_MS} tend to be relatively consistent across all sample sizes. This is to say, test statistic of T_{RLS} and T_{RLS_MS} produce very stable estimates.

4.2 Average P-values and Empirical Rejection Rates

Another perspective to examine the performances is average p-values and empirical rejection frequencies. Since we use significance level $\alpha = 0.05$ for hypothesis tests, our expected value for the empirical rejection frequencies should be about 0.05. To validate the performance of statistical simulations, we use the largest sample size (N=10,000) as a baseline to evaluate different estimators and sample sizes. For each sample size, we calculate the average mean p-values on the 1,000 replications and calculate the mean empirical rejection frequencies of the p-values that are less than 0.05. That is, we calculate the percentages on each 1,000 replications and see how often these models reject the null hypothesis. As Table 2 shows, when N=10,000 the mean empirical rejection rates of all models are in the ballpark of 0.05.

	Average P-values					**Number of P< 0.05**					**Empirical Rejection Frequency**				
N	ML	GLS	RLS	ML MS	RLS MS	ML	GLS	RLS	ML MS	RLS MS	ML	GLS	RLS	ML MS	RLS MS
50	0.226	0.726	0.485	0.218	0.537	310	0	50	309	27	0.310	0.000	0.050	0.309	0.027
80	0.316	0.638	0.499	0.327	0.531	188	9	53	165	25	0.188	0.009	0.053	0.165	0.025
100	0.370	0.607	0.497	0.367	0.536	140	22	43	129	37	0.140	0.022	0.043	0.129	0.037
200	0.441	0.570	0.495	0.420	0.501	82	34	60	90	50	0.083	0.034	0.060	0.090	0.050
300	0.448	0.531	0.493	0.450	0.504	83	36	54	77	45	0.071	0.036	0.054	0.077	0.045
400	0.465	0.526	0.495	0.471	0.506	62	40	56	53	48	0.062	0.040	0.056	0.053	0.048
500	0.470	0.515	0.492	0.471	0.498	75	44	44	61	64	0.075	0.044	0.044	0.061	0.064
800	0.494	0.515	0.496	0.471	0.512	63	39	50	66	44	0.063	0.039	0.050	0.066	0.044
1,000	0.479	0.517	0.499	0.481	0.487	61	42	48	50	54	0.061	0.042	0.048	0.050	0.054
2,000	0.494	0.495	0.494	0.497	0.494	56	46	44	47	48	0.056	0.046	0.044	0.047	0.048
5,000	0.495	0.497	0.499	0.503	0.500	63	49	51	51	56	0.063	0.049	0.051	0.051	0.056
10,000	0.512	0.499	0.507	0.508	0.509	53	44	47	59	49	0.053	0.044	0.047	0.059	0.049

Note: GLS had 37 non-convergences when N=50. 1 non-convergence when N=80

Table 2. Simulation concerning mean empirical rejection rates

Table 2 shows that T_{ML} tends to have large variations in average p-values across different samples. When N=10,000, the mean p-values is about 0.5, while when N=50, it decreased to 0.226. This pattern is highly identical to T_{ML_MS} model. In contrast, T_{RLS} and T_{RLS_MS} also have similar mean p-values across all samples, whereas T_{RLS_MS} tend to have slightly large mean p-values when N<100. T_{GLS} has 0.5 average p-value when N=10,000; however, when N<400, the

number increased rapidly. In terms of mean empirical rejection frequency, T_{ML} and T_{ML_MS} share identical patterns. When N is large, they both have about 5 percent mean rejection rates; while when N<400, the rejection rates start to increase rapidly. When N=50, the mean rejection rate increased to about 0.31. In contrast, both T_{RLS} and T_{RLS_MS} have very consistent rejection rates almost across all samples. However, when N<200 T_{RLS_MS} tends to slight under-reject the true model.

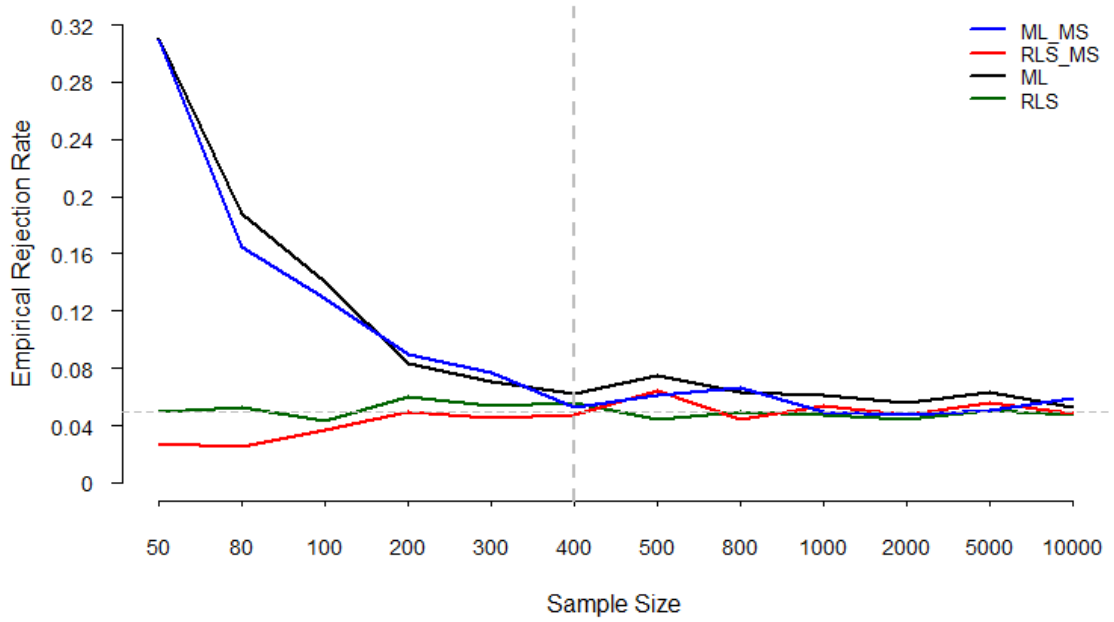

Figure 3. The effect of sample size on empirical rejection frequency

Figure 3 visualizes the different performances between these methods. In terms of empirical rejection rates, the different performances of T_{ML} and T_{RLS} are parallel to what we have discussed. That is, when sample sizes become less than 400, we start to see T_{ML} and T_{ML_MS} tend to incorrectly over-reject the null hypothesis frequently.

5. Discussion & Conclusion

Scholars in the field of SEM have documented that sample covariance matrix $\boldsymbol{S}$ can be ill-conditioned when sample sizes are small. This has an effect on T_{ML}, specifically its behavior is not χ^2 when sample size is small. Two solutions have been proposed to remedy this problem: Regularized GLS function and reweighted least squares. Regularized GLS is based on Chi and Lange's (2014) MAP covariance matrix estimator. The basic idea is to extract eigenvalues from a poorly conditioned covariance matrix and correct the inflated eigenvalues by adding a penalty function to a standard function to steer the estimated eigenvalues toward the geometric mean of sample covariances. Arruda and Bentler (2017) have shown that regularized GLS method can produce consistent test statistics in small samples. This method can be easily extended to estimate the mean and covariance structure models, whereas the programing of the regularized GLS is relatively complicated.

In contrast, the T_{RLS} function is much easier to program and requires less computational power. As we have discussed, T_{RLS} relies on replacing the sample covariance $\widehat{\boldsymbol{\Sigma}}^{-1}$ in both mean and covariance structure with $\widehat{\boldsymbol{\Sigma}}_{ML}^{-1}$ that derives from the F_{ML} function. As Harlow (1985) has found, the covariance structure T_{RLS} and T_{ML} perform similarly well when sample sizes are large. Moreover, recently Hayakawa (2019) also finds that when the sample sizes are small, T_{RLS} obviously outperforms T_{ML} in a confirmatory factor model, a panel autoregressive model, and a cross-lagged panel model. This study not only affirms the previous findings in the literature, but also extends the statistical power of T_{RLS}. We find the similar patterns hold with the mean and covariance structure models. That is, T_{RLS_MS} and T_{ML_MS} perform equally well when the samples are large enough, and both of these methods follow the asymptotic properties; whereas in the context of small samples and under the assumption of multivariate normality, T_{RLS_MS} performs

better than T_{ML_MS} in terms of chi-square test statistics as shown by empirical rejection frequencies. A finding in this report is that the near-ideal performance of RLS in covariance structures is not fully maintained in mean and covariance structures. That is, at N<200, we found a slight under-rejection in T_{RLS_MS}. This implies some over-acceptance of the mean structure in equation 18. At this time, we do not have a proposal on how to avoid this problem.

For our future study in covariance structure analysis, we have yet to test the mean structure part with different estimators. It would be of interest to investigate the performances of different estimators on mean structure and covariance structure respectively. As mentioned earlier, regularized GLS performs equally efficient in reducing the overrejection problem of T_{ML}. It is thus important to thoroughly compare the performances of T_{RLS} and T_{RGLS} in various sample sizes, and examine their comparative characteristics and advantages. Finally, no studies have tested mean and convariance structure analysis using T_{RGLS}. As such, the extent to which the performance of T_{RLS_MS} vis-à-vis r T_{RGLS} in mean and covariance structure models remain unexamined. In particular, the applications of reweighted least squares with other estimation in SEM require further examinations of these two new methods.

References

Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics, 18*(3), 1453-1463.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis (2nd Edition)*. New York: John Wiley & Sons, Inc.

Arruda, E. H., & Bentler, P. (2017). A regularized GLS for structural equation modeling. *Structural Equation Modeling: A Multidisplinary Journal*(24), 657-665.

Bentler, P. (1989). *EQS Structural Equations Program Manual*. Los Angeles: BMDP Statistical Software.

Bentler, P. (2006a). EQS 6 Structural Equations program (Version 6.4). Encino, CA: Multivariate Software, Inc.

Bentler, P. (2006b). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc. .

Bentler, P. (2018). Number of factors in growth curve modeling. *Structural Equation Modeling: A Multidisplinary Journal, 25*(6), 961-964.

Bollen, K. A., & Curran, P. J. (2005). *Latent Curve Models: A Structural Equation Perspective*: Wiley-Interscience.

Browne, M. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*, 1-24.

Chi, E. C., & Lange, K. (2014). Stable estimation of a covariance matrix guided by nuclear norm penalties. *Computional Statistics and Data Analysis*(80), 117-128.

Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structure framework: A Monte Carlo Study.* (PhD Dissertation). UCLA,

Hayakawa, K. (2019). Corrected Goodness-of-fit test in covariance structure analysis. *Psychol Methods, 24*(3), 371-389.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traidtional and modified tests of covariance structures. *Structural Equation Modeling: A Multidisplinary Journal, 14*, 361-390.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling *Sociological Methods & Research, 26*, 329-367.

Hu, L.-T., Bentler, P., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*(2), 351-362.

Huang, Y., & Bentler, P. (2015). Behavior of asymptotic distribution free test statistics in covariance versus correlation structure analysis. *Structural Equation Modeling: A Multidisplinary Journal, 22*, 489-503.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis. *Structural Equation Modeling: A Multidisplinary Journal, 10*, 128-141.

Jalal, S., & Bentler, P. (2018). Using Monte Carlo normal distribution to evaluate structural models with nonnormal data. *Structural Equation Modeling: A Multidisplinary Journal, 25*, 541-557.

Jöreskog, K. G. (1969). Some contribution to maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Lee, S.-Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. England: John Wiley & Sons Ltd.

Lee, S.-Y., & Jennrich, R. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika, 44*(1), 99-133.

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling: A Multidisplinary Journal, 19*, 86-98.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling: A Multidisplinary Journal, 25*, 21-40. doi:https://doi.org/10.1080/10705511.2017.1369088

Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research, 45*, 420-452.

Yuan, K. H., & Bentler, P. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association, 92*, 767-774.

Yuan, K. H., Zhang, Z., & Deng, L. (2019). Fit indices for mean structures with growth curve models. *Psychological Methods, 24*(1), 36-53.