Title
Applications of Regularization to SEM: Shrinking Eigenvalues to Improve Stability of Covariance Matrices

Author
Arruda, Erin Hilary

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Applications of Regularization to SEM:

Shrinking Eigenvalues to Improve Stability of Covariance Matrices

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Erin Hilary Arruda

2017

ABSTRACT OF THE DISSERTATION

Applications of Regularization to SEM:

Shrinking Eigenvalues to Improve Stability of Covariance Matrices

by

Erin Hilary Arruda

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2017

Professor Peter M. Bentler, Chair

Estimation methods employed in Structural Equation Modeling (SEM) depend on asymptotic theory. When assumptions are violated (e.g., sample sizes are not especially large relative to the number of variables), methods break down, and conclusions are dubious. It has been suggested that ill-conditioned matrices contribute to poor performance (Huang & Bentler, 2015; Yuan & Bentler, 2017). In the present investigation, a Maximum a Posteriori (MAP) estimator was proposed and implemented for two matrices common to SEM to address conditioning: the sample covariance matrix and the asymptotic covariance matrix based on fourth order moments. This MAP estimator improves the condition of matrices by pushing down (pulling up) the over (under) estimated sample eigenvalues of poorly conditioned matrices, and better-conditioned matrices were expected to improve solution propriety as well as global model fit. Three differing implementations were proposed for Generalized Least Squares estimation methods (GLS and ADF) as well as correction methods to Maximum Likelihood solutions. Potential advantages of

the approaches were evaluated using three Monte Carlo simulation studies across a wide range of sample sizes and estimation methods. The results reveal overall solution propriety is improved, and regularization when applied directly to weight matrices is more effective than indirect application (i.e., by modifying an input matrix or using correction methods). Moreover, results were dramatically improved for normal theory GLS even at samples sizes as small as $N = 60$ and greatly improved for ADF/RES methods at samples as small as $N = 150$. Generally, advantages did not carry over to non-normal conditions. Potential reasons for this result are given as well as prospective solutions. An illustrative example demonstrates the use of regularized GLS with real-world data.

The dissertation of Erin Hilary Arruda is approved.

Craig Enders

James Stigler

Noreen Webb

Peter M. Bentler, Committee Chair

University of California, Los Angeles

2017

To Bradd, who sacrificed as much as I did

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

If I have seen further, it is by standing on the shoulders of giants.

I extend thanks and admiration to my committee members, Drs. Craig Enders, Noreen Webb, and Jim Stigler. I am especially grateful to Dr. Peter Bentler for his guidance and patience. I am truly fortunate for the opportunity to work with and learn from him. His generosity in his time and feedback is invaluable. Peter, thank you for opening new doors to exciting avenues of research and scholarship.

I am still, and will always, be indebted to Drs. Allen and Adele Gottfried for wisdom, passion for data, and encouragement.

Thank you also to my parents- for teaching me to question and to be persistent.

Finally, I want to acknowledge the continuous love and support of Bradd Arruda. Thank you for journeying with me through this experience, and all over the world. I look forward to more adventures.

This research was partially supported by the Institute of Education Sciences and the U.S. Department of Education, through Training Grant R305B080016 to the University of California, Los Angeles. The opinions expressed herein are my own and do not reflect the views or policies of this funding agency.

VITA

EDUCATION

2011 Masters of Arts, Experimental Psychology, California State University, Fullerton

2009 Bachelors of Arts, Psychology, California State University, Fullerton

DISTINCTIONS

2011-2014 Advanced Quantitative Methods Fellow- Recipient of Doctoral Training Grant Funded by the Institute of Education Sciences

2011 CSU Chancellor's Doctoral Incentive Program

EMPLOYMENT

2016-2017 Graduate Student Researcher, University of California, Los Angeles

2011-2016 Teaching Associate, University of California, Los Angeles

PUBLICATIONS

Arruda, E. H. & Bentler, P. M. (2017). A regularized GLS for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal.* Advance online publication. doi: 10.1080/10705511.2017.1318392

Fuligni, A., Arruda, E. H., Krull, J. L. & Gonzales, N. (2016). Sleep duration, variability, and peak levels of achievement and mental health. *Child Development.* Advance online publication. doi: 10.1111/cdev.12729

Krull, J. L. & Arruda, E. H. (2015). Growth curve modeling. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology*. Hoboken, NJ: Wiley-Blackwell. doi: 10.1002/9781118625392.wbecp570

CHAPTER 1

INTRODUCTION

1.1 Current Problems in SEM

Structural equation modeling (SEM) is utilized extensively in the behavioral sciences and across disciplines, including psychology and education. Perhaps this is the case because SEM is a flexible multivariate framework with many advantageous benefits ideally suited for the study of human behavior such as the ability to incorporate latent constructs and account for measurement error. One specialized variant for SEM is covariance structure analysis, reflecting a principal objective to compare a sample covariance matrix to a low-parameter model covariance matrix and find parameter estimates that minimize some measure of discrepancy between the data and the model.

Global model evaluation is another advantage of SEM. The specification of a model translates hypotheses and graphical representations of a model into a series of equations, asserting SEM as a confirmatory technique. This is one of the strengths of SEM as the parameters of the full model are not only estimated, but the model (and constraints placed on the parameters of the model) are evaluated. A sample covariance matrix $\mathbf{S}$ is an estimator of the population covariance matrix $\boldsymbol{\Sigma}$. Based on a user-specified model with parameter estimates, a reproduced or "model-implied" covariance matrix is produced. The covariance matrices are compared to test a hypothesis with regards to model adequacy. The test of goodness of fit essentially tests if a model produces an estimated population covariance matrix is that consistent with the observed unrestricted covariance matrix. In other words, does the model "fit" the data? Of course, other models may be consistent with the same data.

In statistical terms, the null hypothesis for model evaluation is

$$H_0: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}),$$

where $\boldsymbol{\Sigma}$ is an unstructured population covariance matrix, and if $\boldsymbol{\theta}$ is a vector that comprises the unknown model parameters, then $\boldsymbol{\Sigma(\theta)}$ represents the model-implied population covariance matrix. A fitting function is used to test the null hypothesis and also used in calculations of many fit indices that address the overall adequacy of the models (though see Barrett (2007), Hayduck et al. (2007), as well as MacCallum, Browne, & Sugawara (1996) for more discussion on relative merits of different types of model evaluation).

A test statistic is computed using the minimized discrepancy function at convergence

$$T = (N\text{-}1)\hat{F}_{\text{min,}}$$

where N is the sample size and $\hat{F}_{\text{min}}$ is the minimal value of the fit function for the parameter estimation used and evaluated based on the model degrees of freedom (*df*)[1]. The test statistic and corresponding *df* permit tests of the null hypothesis. Given the null hypothesis above, a non-significant result is desirable. The question of interest is if the fit function is sufficiently small, given the sample size involved, and the *df* associated with the model. The fewer parameters (*q*), the larger the *df* is but of course this depends on how much "data" or unique sample variances and covariances (*p**) are available. Additional methods of model evaluation exist such as fit indices and these are also typically based on the fit function, however these will not be further considered here.

Although SEM is a flexible technique, there are cases when current methods are not ideal. Estimators rely on asymptotic theory. Beyond correct model specification, asymptotic theory implies a correct functional form and reasonably large samples. When violations of

[1] If .5(*p*) (*p*+1) = *p**, and *q* is the number of estimated parameters, the degrees of freedom (*df*) is calculated as *df* = *p**-*q* + *c* where c is the number of constraints (when no constraints are considered, *c* = 0 and is often dropped from notation).

assumptions occur, such as incorrect distributional forms and small samples, the validity of results are questionable. Estimators differ in that they are each characterized by their own set of assumptions, and the choice of an estimator should be made with theory in mind. However, often a method is chosen that best dampens the consequences of violations illustrated by a body of literature using finite simulation study. It may also be the case that users ride the default method in the computer program without any consideration of asymptotic theory.

Maximum likelihood (ML) is one of the earliest and most common methods of estimation for SEM (Jöreskog, 1969). It is often the case in application that ML estimation based on large sample theory is employed, even with high dimensional, non-normal data, with small or medium sample sizes (e.g., MacCallum & Austin, 2000; Paxton, Curran, Bollen, Kirby, & Chen, 2001). Maximum likelihood also presumes that the joint distribution of variables is multivariate normal (MVN). The effects of non-normality on statistical inference include incorrect test statistic and inflated Type I error (West, Finch, & Curran, 1995; Chou & Bentler, 1995). Asymptotic robustness theory developed in the context of latent variable models originated by Browne (1987) and Anderson & Amemiya, (1988) should not necessarily be applied to every situation. Sometimes conditions are not able to be verified based on sample data (Bentler & Dudgeon, 1996), and simulation results reflecting robustness are dependent upon and limited by the testing conditions. For example, the degree of robustness may depend on the type and degree of nonnormality as well as independence (Satorra & Bentler, 1990; Savalei, 2008). Although assumed, in practice, multivariate normality of observed variables is rarely satisfied (Micceri, 1989). Therefore, an alternative method of estimation should be considered instead for these cases.

One such method is asymptotic-distribution free (ADF) estimation (Browne, 1982; 1984; Chamberlain, 1982). ADF is an extension of generalized least squares (GLS; Browne, 1974), therefore is sometimes referred to as arbitrary distribution GLS (AGLS), or simply weighted least squares (WLS). Due to the fact ADF requires fewer assumptions (requiring only eighth-order moments are finite; Browne, 1984) it should be highly useful in order to avoid making strong assumptions necessary for ML estimation.

Although such a method should be practically useful across many real life data situations, the method requires a large sample size to perform in a trustworthy manner, and breaks down when the number of variables is large. Research reveals the sample size necessary to achieve reasonable approximation for the distribution of the test statistic can be as large as 2500 cases, and in some conditions as large as 5000 (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1985; Huang & Bentler, 2015). These Monte Carlo simulation studies suggest at the smallest sample sizes, the ADF test statistic is inflated and ADF virtually always rejects the true model.

Due to the fact ADF estimation requires the inversion of a large dimension matrix, moderate to large number of variables become computationally problematic, therefore worsens as *p* increases (Muthén & Kaplan, 1985). Convergence and improper solutions are also encountered when employing the classic ADF estimator. For example, in simulation research, Curran, Finch & West (1996) dropped 17% of data due to out of bound parameter estimates and non-convergence across conditions when employing ADF estimation. Small sample size compounds the non-normality problem (Bentler & Yuan, 1999; Hu et al., 1992; Hoogland & Boomsma, 1998). Because differential results are obtained by ADF and GLS under normality when asymptotically results should converge, it is proposed that higher order elements of the

weight matrix are unstable because of large sampling variability given they are estimated given a small amount of information (see Hu et. al., 1992).

Recently, it was suggested that the extremity of eigenvalues of the sample covariance matrix in ML estimation (Yuan & Bentler, 2016) and the weight matrix in ADF estimation (Huang & Bentler, 2015) are related to issues of poor performance described above. It is known that in general larger sample eigenvalues are over estimated, smaller eigenvalues are underestimated, and this distortion is related to sample size as well as the dimension of the estimated matrix. When the matrix dimension p is larger than the number of observations the sample covariance matrix $\mathbf{S}$ is not invertible. When the ratio p/N is less than one, but by a negligible amount, the sample covariance matrix is invertible but numerically ill-conditioned. Inverting such an ill-conditioned matrix amplifies estimation error dramatically (Ledoit & Wolf, 2004) and this in turn leads to unstable estimates. If distortion actually reflects an ill-conditioned matrix, it calls importance toward the development of a well-conditioned estimator particularly in the case of large-dimensional (large p, small N) matrices that can be typical for SEM estimation, especially with regards to weight matrices.

This call for attention to the condition number (κ) of these matrices is further bolstered by recent research by Yuan and Bentler (2017). They found the condition number of an input covariance matrix ($\mathbf{S}$) to be larger than the population condition number, though these decreased monotonically with sample size. Interestingly, condition numbers also increased with violations of non-normality. Huang and Bentler (2015) studied ADF estimation and found substantially large condition numbers of the weight matrix, even under conditions of multivariate normality at sample sizes used in application, ($\kappa \approx 25{,}000$ & 2000 for sample sizes $N = 150$ & $N = 250$, respectively). Condition numbers at these sample sizes worsened greatly under non-normality

and dependence, and in several cases exceeded a condition number "rule of thumb" of 1000 for an ill-conditioned matrix 100 times over. In addition, Huang and Bentler (2015) found the variability of the condition number across replications was also quite large, and worsened under violations of normality and independence.

1.2 Proposed Methodology

Although these contemporary and valuable research findings reveal important relationships between violation of assumptions, condition numbers, and poor performance of estimation methods, very little methodological research addresses tackling ill conditioned matrix problem. Employing regularization techniques is a new avenue of research that could be quite promising. Of course, application of such techniques should be evaluated empirically.

A more thorough literature review on regularization will be provided in Chapter 2. However it is well known that statistical procedures can benefit from shrinkage estimation of symmetric matrices like the covariance matrix based on the general shrinkage principle dating back to Stein (1956). If eigenvalues are distorted, the matrix is ill-conditioned. If the ratio of the maximum to minimum eigenvalue, (termed the condition number) of the matrix is too large it reflects a poorly conditioned covariance matrix. It therefore seems valuable to improve these covariance matrices and to stabilize an estimate of the covariance matrix (and importantly it's inverse) especially since these matrices are foundational to structural equation modeling. I propose to use a regularization method that will directly address the eigenvalue distortion and adjust the sample eigenvalues using a nonlinear transformation. The specific regularization approach I will use in this dissertation is referred to as Covariance Estimation Regularization by Nuclear Norm penalties (CERNN; Chi & Lange, 2014), a Maximum a Posteriori estimator (MAP). I will refer to this method outlined in Chapter 2 as MAP throughout this dissertation.

The purpose of these studies was to develop and investigate applications of the MAP estimator to SEM, and evaluate them relative to one another and to traditional methods under a variety of conditions, including varying both sample size and normality. Details will be provided later but in general terms MAP estimators supplanted matrices in the following ways: (1) In place of the sample covariance matrix (**S**) that is used in calculations of the weight matrix in GLS estimation (2) In place of a sample covariance matrix (**S**) as the input data matrix for the ADF estimator (3) In place of a matrix of fourth order sample moments ($\hat{\mathbf{\Gamma}}$) for two distribution free methods and two correction methods. Moreover, these methods were compared to their traditional counterparts, as well as maximum likelihood (ML), one of the most commonly applied methods.

1.3 Significance of Research

The importance of this study is threefold. First, even though breakdown of methods have been known for some time, there has not been an adequate explanation for why this is the case. This study seeks to shed light on potential reasons. Second, the present study is the first to examine this particular type of regularization of matrices important to SEM and estimators used in SEM. While this research builds on and extends contemporary studies regularization in general has moved very little beyond ML (i.e., Yuan & Chan, 2008, Yuan & Bentler, 2016, Jacobucci et al. 2016). Additionally, existing approaches are ridge-type approaches, which are a small subset of possible regularization methods. Research reveals condition numbers of weight matrices are quite large but they only have been examined under some, but not all, conditions studied here. The wide ranging sample sizes and methods in this study are chosen to further demonstrate the extension of earlier work. Third, beyond theoretical implications, regularization techniques can be implemented to address practical issues such as model convergence and

problems that occur when the number of variables is quite large relative to small sample size, allowing for more trustworthy results.

1.4 Research Questions

The present studies were designed to address the following research questions:

1. Are condition numbers of commonly employed matrices inflated? If so, under what conditions and to what extent?
2. Do MAP estimators improve condition numbers? If so, under what conditions and to what extent?
3. Is solution propriety improved when MAP estimators are employed? If so, are there certain types of problems that are improved, and by how much?
4. Is model fit improved when MAP estimators are employed? In what specific ways and to what extent?

Regarding question one, previous research outlined earlier reveals condition numbers are inflated. Therefore, condition numbers for samples are expected to be larger than population values and worsen when p increases relative to N and as nonnormality increases. Magnitudes of bias of matrices common to SEM have only been studied under a small number of conditions and are generally unknown, which is one of the objectives of this study.

Additionally, solution propriety including non-convergence, negative variances and problem solutions are expected to improve when MAP estimators are employed. When $\mathbf{S}$ is not full rank or near singular the iteration process for obtaining the parameter estimates can be unstable and may require hundreds of iterations to reach convergence (e.g., Boomsma, 1985). Poor conditioning not only is related to inaccuracy in the computation of the inverse of a matrix but also, if the computations include iterations, could cause an algorithm to fail. Additionally,

related research with ridge methods reflects convergence is improved (Yuan & Bentler, 2017, Yuan & Chan, 2008) therefore it seems likely the same improvements will result with this type of regularization.

Moreover, model fit is expected to improve in regards to test statistics and rejection rates. Given that this is the first exploration of a MAP approach to SEM, it is unknown which under what conditions this approach will garner the largest improvements. Under large sample conditions and no misspecification, methods are expected to perform similarly and optimally. ML will be inflated with too many rejection rates with small sample sizes, while GLS test statistics will be underestimated with too few rejection rates. Additionally under normality ADF is expected to perform poorly until *N*s reach at least 2000. Correction methods generally work well under normality except at the smallest of *N*s, however may worsen under nonnormality. If MAP estimators improve matrices it seems this would translate to better, more optimal results given model and the estimator is not misspecified, at small sample sizes and under nonnormality. The remainder of this dissertation is organized as follows:

Chapter 2. Regularization is discussed along with applications to SEM and covariance matrices.

Chapters 3-5. Simulation studies and results are provided.

Chapter 6. Discussion of simulation findings.

Chapter 7. An illustrative example is provided.

CHAPTER 2

REGULARIZATION

2.1 Regularization in Social & Behavioral Sciences

Regularization is a broad term referring to a class of methods, though is often referred to as a single technique applied in a specific context. A layperson may define regularization as "the condition of having been made regular" or "to cause to conform". Similarly, the term regularization is often used in mathematical context to refer to the addition of information to solve an ill-posed problem. These definitions relate to Bickel and Li's (2006) informal definition that reflects modification of a method in order to give reasonable answers in unstable situations. An early use of the term regularization was by Tikhonov (1943) in the context of using a Tikhonov factor gamma > 0 as a regularization parameter, related to ridge regression discussed later.

Regularization methods are applied with regularity in areas such as machine learning and statistics. Most applications address problems of multicollinearity, overfitting, or sparsity- which refers to the abundance of zeros in a matrix. Others regularization methods concern parsimony principles (e.g. smoothing, model selection, or methods to control model complexity). Relatedly, regularization can also be applied for the purpose of inverting a matrix or speeding up slow computations. These types of complications are quite typical with high-dimensional data. High-dimensional data refers to data sets with large number of variables, p given a relatively small sample, N though some reserve this term for the condition for which p exceeds N exclusively (see for example, Pourahmadi, 2013, p. 3).

Applications of regularization that are well-known in the social and behavioral sciences are typically those that are related to regression. Such types of regularization methods

occasionally used in psychological literature are ridge regression (Hoerl & Kennard, 1970, Tikhonov, 1943), and to a lesser but more recent extent, the use of Least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani, 1996) and Elastic Net Regularization (Zou & Hastie, 2005). These methods are used to solve the problems of prediction accuracy and dimensionality, respectively. In a basic sense, the coefficients are regularized. The approach is to penalize the sum of squared residuals that are minimized by use of a penalty term that involves the coefficients. This is equivalent to an inflation of variance by adding small positive quantities to the diagonal entries of the covariance matrix (hence the name "ridge"). The method aims to reduce the problem of multicollinearity for ridge regression and induces sparsity by controlling the length of regression parameters for LASSO. Penalty terms are discussed further in the next section, however resultant effects for ridge and LASSO procedures differ because the penalty differs. The LASSO approach drives coefficients to zero using the squared sum of absolute value of coefficients as a penalty, while ridge regression utilizes the sum of squared regression coefficients.

Additionally, Akaike Information Criterion (AIC; formerly An Information Criterion; Akaike, 1971) and Bayesian Information Criterion (BIC; Schwartz Criterion; Schwartz, 1978) are flavors of regularization referred to as penalized likelihood methods. They use the likelihood function to discourage overfitting from machine learning literature also fall under the term regularization. Both use a penalty term based on the number of parameters in the model (larger for BIC than AIC). These information criteria are often used for model selection in the psychological literature, particularly for model comparison when models are non-nested.

Contemporary research reveals regularized methods applied to regression are starting to cross over to covariance modeling. Specific to SEM methodology, some regularization

techniques have been recently applied and studied (Jacobucci, Grimm, & McArdle, 2016; Yuan & Bentler, 2016; Yuan & Chan, 2008, 2016; Yuan, Wu, & Bentler, 2011). Ridge regression and LASSO techniques for example have been adopted to SEM (Jacobucci, Grimm, & McArdle, 2016; Jung, 2013). Jacobucci et al. suggests penalizing specific parameters in order to penalize model complexity and increase generalizability of models, naming this application regularized structural equation modeling or regSEM. Regularized SEM is an extension of regularization in the regression framework but also borrows from the principal components and exploratory factor analysis (PCA; EFA) literature. Further approaches of regularization to PCA and EFA frameworks are not detailed here and interested readers are referred to Jacobucci et al. or Liang (2016) for such a review.

Furthermore, Jung (2013) introduced a regularized extension of two-stage least squares (2SLS) estimation, particularly for small sample cases. This method incorporates a ridge type of estimation of parameters in order to address problems inherent to 2SLS. Aims of this approach are to find stable and accurate solutions of parameter estimates with (extremely) small samples (*N*s ranging from 5-50). An interesting feature of both approaches by Jung and Jacobucci et al. is the use of cross validation in order to find the ridge parameter. On the other hand, 2SLS is not a full information estimation procedure. Global fit is not able to be assessed with 2SLS which is also a main focus of this paper. In fact, several ridge parameters need to be identified given the way 2SLS treats the model as subsets of model equations rather than globally.

Yuan and Chan (2008) proposed and tested a solution to non-convergence and poor inference when covariance matrices are singular or ill-conditioned. Their estimator for the covariance matrix ($\mathbf{S}_a$) is a linear combination of **S** and $a\mathbf{I}$, where $a = p/N$ was suggested. Basically, this method is a ridge method that introduces bias in order to reduce mean square error

for parameter estimates and aims to reduce discrepancy between $\mathbf{S}_a$ and the estimated structural covariance matrix over **S**. Yuan and Chan employ the widely-used Maximum Likelihood (ML) procedure and supplant the sample covariance matrix with the augmented matrix $\mathbf{S}_a$ which allows for convergence for sample sizes as small as the number of observed variables (in this case, N = 15, p = 15, df = 87). In comparison to ML, this method enjoys more accurate parameter estimates and improved convergence rates, however the distribution of the resulting test statistic is unknown, making inferential decisions complex for practical purposes. For inferential tests of overall model fit, critical values were found by simulation, and two approximations or corrected test statistics paralleling Satorra and Bentler's rescaled and adjusted test statistics are offered, though not studied extensively in this paper. Additionally, resulting error variances are biased by a and need adjustment, although this simply amounts to subtracting the tuning parameter from the estimates.

Yuan and Chan's approach was extended to correlation matrices common to SEM, such as when data are ordinal (Yuan, Wu, & Bentler, 2011). This is an intuitive extension given convergence problems are typical during the popular two stage process of estimating the polychoric/polyserial correlation matrix then performing robust ML, particularly at small sample sizes. Results suggest improvements in convergence for sample sizes under N = 300, with solutions converging at a quicker pace and increasing the overall rate of convergence. Overall, findings suggest the ridge procedure (with proper a) also leads to less biased and more efficient accurate parameter estimates than ML. The ease of implementation is one advantage of this procedure given the correlation matrix is simply supplanted with the augmented matrix. Practical issues of this approach include calculating scaled and adjusted test statistics parallel to those put forth by Yuan & Chan (2008) post-hoc based on the resulting statistics from ML estimation, and

the problem of finding the optimal positive scalar, a. The parameter is not based on characteristics of the observed data, which is typically desirable.

With this goal, another method for obtaining the shrinkage parameter (a in Yuan & Chan's 2008 application) was explored by Kamada (2011) using an information criterion (IC) approach based on the data rather than the constant *p/N*. This approach is closer to finding the penalty based on the out of sample prediction error by cross validation. ICs are used rather than cross validation to lessen the computational burden and perhaps avoid sparse data problems. Although some difference in solution quality is found at the smallest sample size (N = 30), at more typical sample size (Ns of 50 &100) little difference is found. In fact, the average risk function across the 1000 replications were the same as Yuan and Chan's method to the third decimal place. A wider range of conditions should be explored and compared to cross validation methods, to further understand benefits of this method.

Yuan and Chan (2016) extended earlier work to another application of a regularization method. Specifically, Yuan and Chan propose to calculate a linear combination of the identity matrix and the sample fourth order moment matrix $\widehat{\mathbf{\Gamma}}_{\mathrm{I}}$, a symmetric matrix that can be described as covariance matrix of variance and covariances further described in Chapter 4. This linear combination amounts to

$$\widehat{\mathbf{\Gamma}}_{\mathrm{I}} = (1 - a)\widehat{\mathbf{\Gamma}} + a\mathbf{I}.$$

Both matrices are weighted by a tuning parameter (a) in order to harness the stability of $\mathbf{I}$. The objective is to improve stability of the matrix $\widehat{\mathbf{\Gamma}}$ that is used in calculation of the weight matrix. Additionally, Yuan and Chan propose correcting the resulting test statistic as an approximation to the correct distribution. The optimal values of the tuning parameter were selected empirically by simulation across a grid of candidate tuning parameters ranging from

zero to one. Yuan and Chan also introduced a ridge method based on Diagonally Weighted Least Squares though the method did not outperform RGLSI and in general produced less accurate solutions in terms of parameter estimates.

Most recently, Yuan and Bentler (2017) also applied a ridge approach, as well as an anti-ridge approach, to examine and explain a counter-intuitive finding that both increasing the ridge of a matrix and increasing factor loadings can lead to improved convergence rates and speed for the Fisher-scoring algorithm. Anti-ridge refers to modification of other elements in the sample covariance matrix with the focus on strengthening factor loadings and is especially applicable to problems where multiple indicators can be auditioned, such as with Monte Carlo simulation. Simulation results reflect ridge and anti-ridge methods are effective in addressing non-convergence, though anti-ridge might be less practical for some implementation. Importantly, condition numbers of information matrices were reported under a variety of conditions, reflecting huge numbers even under normality and worsening under nonnormal and dependent conditions.

Statistical details of these and some alternative methods of regularization specific to improvement of covariance matrices are treated next.

2.2 Regularized Covariance Matrix: History & Methods

Regularization techniques are applied to estimation of the covariance matrix with the goal of obtaining an accurate estimate of the population covariance matrix and its inverse, the so-called the precision matrix, particularly under non-ideal conditions.

2.2.1 Sample Covariance Matrix

The sample covariance matrix is sometimes referred to as the sample variance-covariance

matrix, the dispersion matrix or the scatter matrix. It is known that the sample covariance is an unbiased estimator of population covariance (σ_{xy})

$$s_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})', \quad (2.1)$$

where x_i and y_i are the *i*th observations of random variables, and

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

is the sample mean. In the special case where $x = y$, the sample covariance reduces to the sample variance, s_x^2. Moreover, in the multivariate case the matrix form of the estimator of the sample covariance matrix of variance and covariance of multiple variables is expressed as

$$\mathbf{S} = \mathbf{Y}'\mathbf{Y}\frac{1}{N-1}, \quad (2.2)$$

where the columns of **Y** are centered such that $\overline{Y}_p = 0$) is an estimator of the population covariance matrix, **Σ**. The correction N -1 in Eq. 2.1 and 2.2 is replaced with N if the distribution of the random variables is normally distributed (Gaussian) such that the estimator is now the maximum likelihood estimator of the covariance matrix. The differences between **S** calculated with and without the bias correction diminishes as N approaches infinity. It is also well known that this estimator performs poorly under high dimension conditions and can be unstable (Stein, 1956).

2.2.2 Problems with High Dimensional Data

The estimators 2.1 and 2.2 above have a number of advantages including the fact the covariance matrix is simple to construct and unbiased such that E(**S**) = **Σ**. The calculations are also intuitive since they are the sample analog of the population covariance matrix. However, there are problems, particularly when $p > N$. In high-dimensional cases such that $p > N$, the

matrix is singular and is non-invertible, even if the population covariance matrix is well conditioned and invertible. The rank of the matrix is no longer p but at most N and the inverse no longer exists which impedes estimation methods for which the inverse of the covariance matrix is necessary.

Other problems occur when N is not much larger than p**,** or in cases of very large p which is sometimes referred to semi-high-dimensionality to qualitatively differentiate this case from the former one where $p > N$. Ultra-high dimensional data is a term that is sometimes reserved for the case when p is much greater than N, ($p >> N$). It is important to recognize high-dimensional does not necessarily mean "big data", but instead reflects the relative size of p and N. In general, increasing N has the effect of improving precision and certainty of inference, whereas increasing p has the opposite impact as this could reduce precision and certainty (Pourahmadi, 2013, p. 4). In other words, covariance estimation improves with increasing N but could deteriorate when p increases. Donoho (2000) refers to this and other tradeoffs with high-dimensionality as a curse and a blessing. If sample size is small and the number of variables is large, the above empirical estimators for a covariance matrix are unstable. Even when N is comparable or larger than p, the sample covariance matrix can have a significant amount of sampling error (for example see Yuan & Bentler, 2017). Other estimators can improve upon the mean squared error (MSE). In addition, $\mathbf{S}$ may not be well conditioned and its inverse is a poor estimator for $\mathbf{\Sigma}^{-1}$. Inversion of an ill-conditioned matrix can induce large noise and can be computationally expensive, especially given large p. Penalized maximum likelihood methods and Stein-type shrinkage are some approaches can improve on the estimators above.

Stein (1956) noted that the sample covariance matrix performs poorly with large p, particularly when p/N is large attributed this to distortion in estimation of eigenvalues. The range

of the sample eigenvalues increases as *p*/*N* grows. A simple simulation following Pourahmadi (2013, p. 27) and Chi and Lange (2014) demonstrates this by drawing independent samples from a 10-dimensional multivariate normal distribution, $y_i \sim N(0, I_{10})$. Figure 1 presents results of this example.

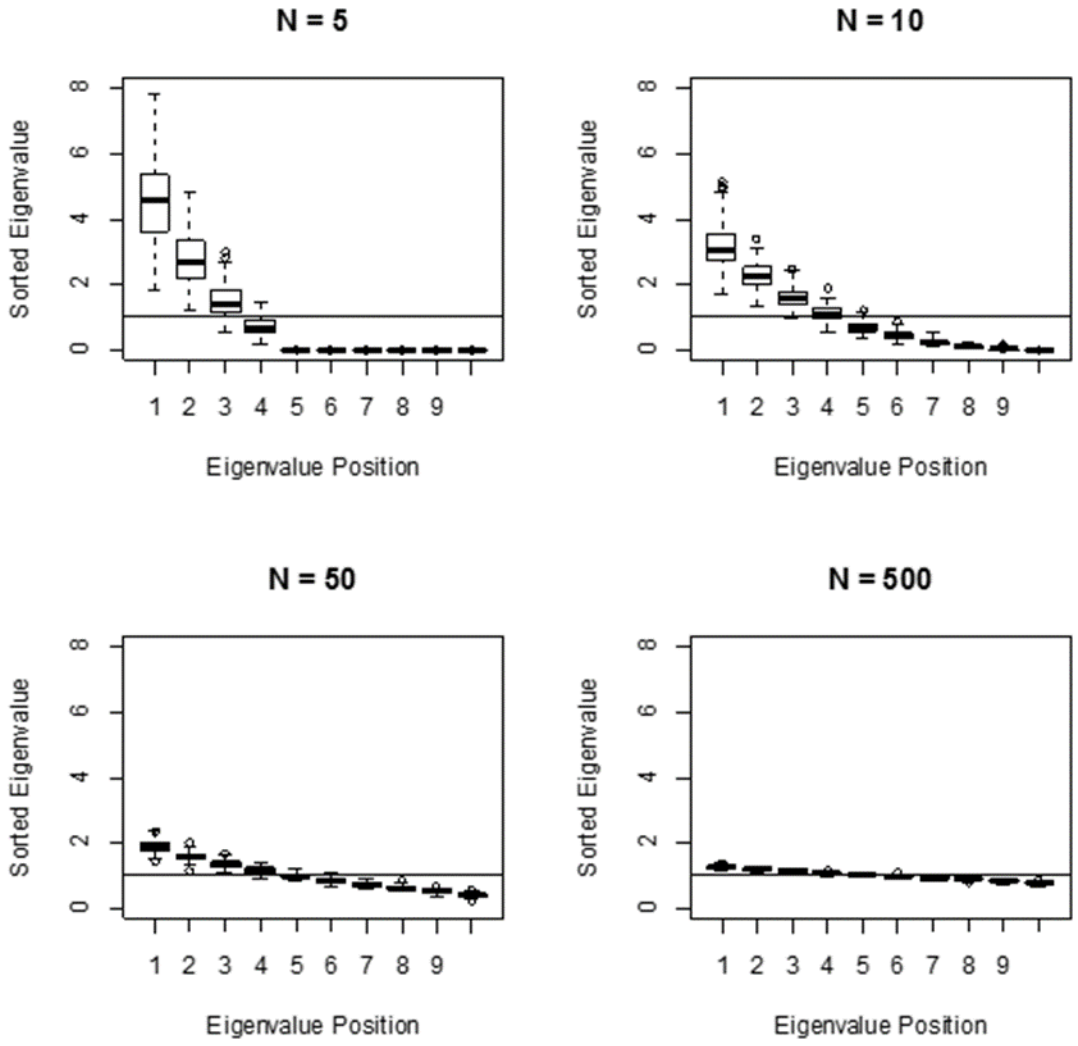

Figure 2.1. Boxplots of sorted eigenvalues of the sample covariance matrix.

The sorted sample eigenvalues of the sample covariance matrix are ordered from largest to smallest for sample sizes *N* drawn from the set {5, 10, 50, 500} over 100 trials. The boxplots depict the highest eigenvalues inflated upwards above the population expected value, (the

reference line $y = 1$), while the lowest eigenvalues are deflated downwards, below the population expected value. In general, if the sample size N and the number of components p approach ∞ in such a way that the ratio p/N approaches $\zeta \in (0, 1)$, then the eigenvalues of $\mathbf{S}$ tend to the Marčenko-Pastur law (Marčenko & Pastur, 1967), such that as p/N increases, the over- and under estimation of eigenvalues worsens in the interval

$$[(1 - \sqrt{\zeta})^2, (1 + \sqrt{\zeta})^2]$$

The condition of the covariance matrix (signified as κ) can be described by the ratio of the maximum eigenvalue to the minimum eigenvalue, where larger condition numbers reflect poor or ill-conditioned matrices. More formally, condition numbers can be expressed as the product of norms such that the condition number of matrix $\mathbf{A}$ amounts to

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A^{-1}}\|,$$

where double bar notation $\| \cdot \|$ reflects any norm defined for matrices. A norm can generally be described as a simple unique scalar number that is always positive and defined for all matrices. It is basically a function such that input information, (say x) is manipulated based on the type of norm such that output is a (nonnegative) number that satisfies certain conditions, where $\| x \|$ describes the function. Often norms grouped together as family of norms and denoted with a symbol or letter subscript outside the double bar (e.g. $\| \cdot \|_{l2}$ is the popular l_2-norm also sometimes referred to as the Euclidean norm). When the l_2-norm is used for calculating $\kappa(\mathbf{A})$, the result will be equivalent to the maximum to minimum eigenvalue ratio.

Norms are often utilized in terms of calculations for penalties. For example, in the case of ridge regression the l_2-norm is used in the calculation and in the case of LASSO the l_1-norm norm is employed. The relationship of different types of l_p norms (where p indicates a subscript designating which norm is employed is more succinctly described by the following equation

$$\| x \|_p = \sqrt[p]{\sum_i |x_i|^p}$$

which amounts to a *p*-th root of a summation of absolute value of elements to the *p*-th power. A simple example of a particular well-known norm that uses single bar notation for scalars is the absolute value.

Condition numbers importantly quantify the trustworthiness of solutions. A well-conditioned matrix has a small condition number (e.g. $\kappa = 1/1 = 1$ for the identity matrix), while ill-conditioned matrices will have a large condition number. Heuristics that might advise benchmarks for condition numbers are simply guidelines. There are no trustworthy benchmarks except a condition number of ∞ indicates a matrix is not invertible, and a condition number equal to or close to one is far from singularity. However, based on sample data the over- and under-estimation of eigenvalues reflects a condition number that is greater than population counterpart, therefore the condition number is biased upward as the simulated example reveals.

2.2.3 Shrinkage Estimators for Covariance Matrices

Improved estimators ($\widehat{\mathbf{\Sigma}}$) can be used instead of **S** based using shrinkage estimators. The shrinkage estimator seeks the optimal tradeoff between bias and estimating variability. Though not directly obvious, methods discussed above such as ridge regression indirectly addresses improving matrix estimation by

$$\widehat{\mathbf{\Sigma}} = \mathbf{S} + a\mathbf{I}_p$$

where *a* is the tuning parameter and I_p is the *p*-dimensional identity matrix.

Techniques under this framework are often referred to as Steinian shrinkage estimators. In fact, though the more general shrinkage principle is credited to Stein due to his classic highly cited symposium paper in 1956 that focuses on means, much of the work was not formally

published until Efron and Morris (1977) published a paper on the James-Stein (1961) shrinkage estimator. Their paper illuminated Stein's paradox—that shrunken estimators of means can perform better than a typical average, with applications surrounding familiar contexts like baseball. Many approaches specific to covariance shrinkage estimation stem a lecture about such estimators by Stein (1975).

More recently, regularization has been further advanced by work under portfolio optimization by Ledoit and Wolf (2003, 2004). Optimization refers maximizing return and minimize risk for investments. Central to optimization is the estimation and inversion of a covariance matrix based on vast amounts of information, with high multicollinearity. The goal is to make good out of sample predictions, therefore this ill-posed problem makes regularization an advantageous tool.

Seminal papers by Ledoit and Wolf (2003, 2004) applied Steinian-type methods. Closed form estimators were derived in order to find optimal values to weight eigenvalues from the sample covariance matrix and a pre-specified positive definite matrix such as the identity matrix is often utilized. This linear combination

$$\widehat{\mathbf{\Sigma}} = (1 - a)\mathbf{S} + a\mathbf{I}_p$$

was not always ideal, especially since this adjustment of sample eigenvalues may not accommodate eigenvalues that are highly dispersed very well.

There are other covariance matrix estimators either inspired by or related to these shrinkage approaches. Different types of shrinkage may consider treating the eigenvalues, the eigenvectors or both. What eigenvalues to shrink and by which method are the subject of a wide literature. One approach involves shrinking the eigenvalues of **S** toward a central value in a non-linear fashion. This approach by Won, Lim, Kim, and Rajaratnam (2013) acknowledges the

condition number in their shrinkage technique, such that they limit the maximum threshold of eigenvalues so that the maximum condition number is not exceeded. The method is ideal in that it subjects a covariance matrix to a better conditioned state overall and non-linearly shrinks the maximum and minimum eigenvalues but may be competitive only in cases when one eigenvalue is much larger than others (Chi & Lange, 2014).

Another approach developed by Chi and Lange (2014) uses nonlinear shrinkage that impacts all eigenvalues. Their estimator is a maximum a posteriori (MAP) estimator; they also refer to it as covariance estimation regularization by nuclear norm penalties (CERNN). An important property is that their MAP estimator is consistent and asymptotically efficient.

2.3 MAP Covariance Matrix Estimator

The general idea is to extract all eigenvalues from a poorly conditioned covariance matrix and adjust them, pushing the highest eigenvalues down and pulling the lowest eigenvalues up. If the spectral decomposition of a symmetric matrix **S** is

$$\mathbf{S} = \mathbf{QDQ}',$$

where **Q** is an orthogonal matrix containing the eigenvectors of **S** and **D** is a diagonal matrix that contains the eigenvalues of **S**, $diag(d_1, \ldots\ d_p)$, then Stein (1956) suggests an alternative unstructured covariance matrix estimator in the form

$$\widehat{\mathbf{\Sigma}} = \mathbf{Q}diag\left(e_1, \ldots, e_p\right)\mathbf{Q}',$$

where e_i is a shrunken estimate of d_i. This method retains the same eigenvectors. The shrunken estimates are obtained by adding a penalty function to a standard function in order to steer the estimated eigenvalues toward the geometric mean of sample eigenvalues. In MAP, this is done by minimization the objective function

$$f(\mathbf{\Sigma}) = \frac{N}{2}\ln|\mathbf{\Sigma}| + \frac{N}{2}tr(\mathbf{S\Sigma}^{-1}) + \frac{\lambda}{2}[\alpha\|\mathbf{\Sigma}\|_* + (1 - \alpha)\|\mathbf{\Sigma}^{-1}\|_*]. \qquad (2.3)$$

The first two terms of Eq. 2.3 are the typical negative log likelihood function under normality. The penalty is the term in brackets and is an α -weighted linear combination of nuclear norms, here, simply trace norms of singular values. For symmetric covariance matrices, a singular value of a matrix is the absolute value of an eigenvalue. Naturally, their sums should be as small as possible given the goal to steer eigenvalues away from infinity and zero. Lambda (λ) is a strength or penalty parameter. As λ approaches zero, the solution approaches the maximum likelihood solution, and eigenvalues will equal sample eigenvalues. As λ increases, the more aggressively the eigenvalues are shrunk toward the geometric mean. Appropriately, as N approaches infinity data overwhelm the penalty.

As is typical for regularization schemes, Chi and Lange (2014) estimate the regularization penalty λ by employing *k*-fold cross-validation where often $k = 10$ (Hastie, Tibshirani, & Friedman, 2009, p. 243). Several penalty values are auditioned and an optimal λ is selected. Excluding a different fixed proportion of data of the data *k* times, the covariance matrix is estimated given λ based on the remainder of the data (the training sample). Next, the estimated covariance matrix is evaluated based on the predictive negative log-likelihood of the estimated covariance matrix that was held out (the validation sample). This is done *k* times for each value of λ that is auditioned, and an average log-likelihood is calculated. The penalty with the smallest average predictive log likelihood is selected.

Alpha (α) is a parameter that controls mixture between the trace and inverse trace penalties. This could be an a priori value such as $\hat{\alpha} = .5$, but Chi and Lange (2014, Eq. 5) propose to compute it as

$$\hat{\alpha}_r = (1 + \bar{d}^2)^{-1}, \tag{2.4}$$

where $\bar{d}$ is the mean of the d_i, the eigenvalues of $\mathbf{S}$. Since $\hat{\alpha}_r$ may be susceptible to extreme sample eigenvalues, other suggestions may be useful such as

$$\hat{\alpha}_R = (1+\hat{d}^2)^{-1}, \tag{2.5}$$

where $\hat{d}$ is the median of the d_i.

Once λ and α are chosen, the eigenvalues are shrunk according to an essentially quadratic equation

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

rejecting the negative root as this is inconsistent with the covariance matrix as positive definite.

Other techniques that fall under the umbrella of regularization should be acknowledged but are not explored further here because in general they do not guarantee positive definiteness of the estimated covariance matrix. These include types of regularization that deal directly with entries of the sample covariance matrix. These techniques might include banding and tapering, or thresholding (see Bickel and Levina, 2008, Pourahmadi, 2013, or Tong, Wang, &Wang, 2014 for reviews of these types of techniques).

Banding or tapering techniques exploit the natural ordering of variables or a specific distance between variables such that estimators can take advantage of the principle that the variables further away from other variables are weakly related relative to those closer in time or proximity (e.g. covariance matrix for longitudinal data). These approaches might discard (banding) or gradually shrink (tapering) off-diagonal elements toward zero in order to create sparsity. Additionally, they might also take advantage of the structure exchanging the more typical sample covariance matrix for the banded structured matrix. Thresholding (soft and hard; see pioneering work by Bickel and Levina, 2008, Donoho and Johnston, 1994, & El Karoui,

2008) is used to improve consistency of the sample covariance matrix, especially in high dimensions. Thresholding applies to every element of the covariance matrix. With this regularization technique, estimation of each entry in the matrix is subject to constraint.

These descriptions of regularization techniques for estimation of covariance matrices are not exhaustive. In fact, many methods are hybrids of multiple types of techniques further blurring any distinctions between categorizes of regularization methods. In sum, the purpose of reviewing the literature is to give the reader a sense of the wide range of regularization methods, and also research that has been conducted specifically in regards to applications to SEM. This dissertation builds on this most recent surge in interest in regularization methods. However, although clearly applications of regularization to SEM is a burgeoning line of research, there is a paucity of literature compared to the large body of literature that already exists in other disciplines that utilize a large variety of regularization techniques under a wide variety of conditions. Most of the research on SEM applications focuses on maximum likelihood estimation and ridge-type regularization. Far fewer studies have explored applications of other regularization methods. The following chapters present three simulation studies that employ a different regularization approach of non-linear shrinkage to eigenvalues via the MAP estimator to various covariance matrices important to SEM under a wide variety of conditions, including wide consideration of sample sizes and estimation methods.

CHAPTER 3

SIMULATION STUDY 1

The next three chapters concern three simulation studies in order to evaluate different proposed applications of the MAP estimator, supplanting matrices common to SEM. The first simulation study investigated regularization in the context of GLS for the sample covariance matrix whose inverse is a fixed weight matrix for this normal theory method. The second simulation addressed the application of regularization to the sample covariance matrix in order to indirectly improve the weight matrix for ADF estimation. The third approach targeted the weight matrix in a more direct fashion, applying regularization directly to the asymptotic covariance matrix of sample fourth order moments ($\hat{\mathbf{\Gamma}}$). This shrinkage estimator was employed for a variety of methods for which $\hat{\mathbf{\Gamma}}$ (or the inverse of $\hat{\mathbf{\Gamma}}$) is used.

3.1 A Normal Theory Regularized GLS Method

This simulation was carried out to illustrate the performance of GLS using the MAP estimator relative to traditional GLS and ML methods for small sample sizes that meet multivariate normality. First, I describe and compare the condition of the covariance matrices of the **S** and the MAP estimator, with the aim to demonstrate MAP improvement over **S.** Secondly, given the improved condition of the MAP estimator for the fixed weight matrix in GLS, the number of iterations, negative variances, and convergence were evaluated to demonstrate expected improvement in the estimation process, especially at small sample sizes. Lastly I aim to demonstrate improvement for overall model evaluation in terms of the test statistics and empirical rejection rates particularly at small sample sizes, therefore test statistics, standard deviations of test statistics, and rejection rates were compared across methods.

3.2 Details of Method and Simulation

Regularization is applied in the manner described in Section 2.3 to **V** in the following GLS fit function (Browne, 1974)

$$.5 \text{ tr}\left[\{(\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\}^2\right], \tag{3.1}$$

where **V** is a weight matrix. With $\hat{F}_{\text{GLS}}$ the minimum of Eq. 3.1 evaluated at $\hat{\boldsymbol{\theta}}_{\text{GLS}}$, Browne showed that $T_{\text{GLS}} = n\hat{F}_{\text{GLS}}$ is an asymptotic χ^2 variate with $df = p(p+1)/2 - q$ when $\mathbf{V}^{-1}$ is a consistent estimator of $\boldsymbol{\Sigma}^{-1}$, that is, **V** estimates **Σ**. In standard GLS, **V = S**, but here **V** is replaced by $\hat{\boldsymbol{\Sigma}}_R$ resulting in a method called here, regularized GLS (RGLS) with test statistic, T_{R}. For comparison purposes, we also study **V** = $\hat{\boldsymbol{\Sigma}}_r$ labeling the method rGLS with test statistic T_{r} (see Eqs. 2.4 & 2.5 regarding differences of the two MAP estimators, $\hat{\boldsymbol{\Sigma}}_R$ and $\hat{\boldsymbol{\Sigma}}_r$).

Details of the Monte Carlo simulation study carried out to illustrate the performance of test statistics associated with GLS and ML methods relative to RGLS and rGLS for small sample sizes under the assumption of multivariate normality are presented next.

A confirmatory factor model (**x** = **Λξ** + **ε,** where **x** is a vector of observed variables, **Λ** is a factor loading matrix**, ξ** is a vector of normally distributed common latent factors, and **ε** is a vector of unique unobserved factors) was employed consisting of three common latent factors, each with five manifest variable indicators, and 15 unique errors independent from factors and one another. The population covariance matrix is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi},$$

where **Φ** and **Ψ** are the covariance matrices of ξ and **ε,** respectively. The model factor loading matrix was

$$\boldsymbol{\Lambda}' = \begin{bmatrix} 0.70 & 0.70 & 0.75 & 0.80 & 0.80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 & 0.70 & 0.75 & 0.80 & 0.80 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.70 & 0.70 & 0.75 & 0.80 & 0.80 \end{bmatrix}.$$

Variances of factors were set to one, with covariances between factors equal to 0.30, 0.40, and 0.50, respectively. Variances of unique factors were set to values such that under normality, variances of observed variables are one. This model has been used in several Monte Carlo simulation studies (e.g., Hu et al., 1992, Huang & Bentler, 2015), and is used here to validate the simulation and as part of a recommended strategy to compare results to existing findings in literature (see Boomsma, 2013).

Data generation was accomplished with R software (Version 3.2.3; R Core Team, 2015) using the simulation function in *lavaan* package (Rosseel, 2012) based on the population model described above with observed variables exhibiting no skewness or kurtosis. This simulation focused particularly on small sample sizes. Selected conditions are representative of small *N* SEM simulation studies, (60, 70, 80, 87, 90, 100, 110, and 120), with specific values related to this model, $df = 87$, $p^* = (15)16/2 = 120$. Also included were a wide range of larger sample sizes (250, 500, 1000, 2000, 2500, 5000, and 100,000). Results across estimators were expected to converge at sizable samples. Sample moments were collected to compare with population moments, including information with respect to matrix eigenvalues.

Each of 1000 independent samples of each sample size was analyzed with the *lavaan* package in R by specifying the correct model with 87 *df*, employing GLS and ML estimation methods. The ability to replace the **V** matrix for the GLS method was not available as an option in this package for our regularized conditions, therefore R script was written and an optimizer (*nlminb* in the *stats* package) was employed to mimic the default in the *lavaan* package. Start

values were set to the *lavaan* simple default such that all parameter values are set to zero, except factor loadings which are unity, and residual variances are started at half the observed variance. The method was not applied to ML and GLS to save computation time, and the R script was verified by simulating conditions and comparing results with unmodified covariance matrices. All results were equivalent.

To review, the basic steps of the simulation are as follows:

1) A true model with population parameters is specified, and a population covariance matrix subject to the model is generated.
2) Data are generated conforming to the particular sample size of the condition and a covariance matrix is estimated.
3) A correct model is specified, and is fit to the sample data using traditional estimators, ML and GLS.
4) Covariance matrices ($\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_R$) are estimated, (see Section 2.3 for detail).
5) A correct model is fit to the sample data, and **V** in GLS is replaced with $\hat{\mathbf{\Sigma}}_R$ (this
6) Step 5 is repeated with **V** replaced with $\hat{\mathbf{\Sigma}}_r$

The set of results from the simulation were recorded and analyzed.

Assessing results and the impact of sample size and method.

For each condition, minimum, maximum, means and standard deviations were calculated for eigenvalues and condition numbers (ratio of the largest to smallest eigenvalue) across 1000 iterations of the following matrices: (a) the population covariance matrix, (b) the sample covariance matrix, and (c) the regularized covariance matrix. Poor solutions were examined including non convergences, the number of flagged iterations, and number of results with at least one negative variance. Each test statistic was averaged across the number of converged

replications and compared to expected values (the degrees of freedom for the model assuming a χ^2 reference distribution, here 87). Standard deviations of test statistics for each method were compared to the square root of twice the degrees of freedom (variance = $2df$, SD = $\sqrt{2df} \approx$ 13.19). Empirical rejection rates were calculated by summing number of replications for which the model was rejected based on the nominal alpha of .05. Since the fitted model is the correct model, this rate should approximate .05.

3.3 Results

3.3.1 Condition of Matrices

The condition number of the population covariance matrix is about 15.35. A finite sample size of 100,000 was first considered to demonstrate asymptotic properties. The average condition number of **S** calculated across 1,000 replications within rounding is 15.45, evidencing validation of the simulation. Next, a condition number for each sample covariance matrix was calculated and averaged across 1,000 replications for each sample size condition, ranging from 60 to 5,000. Table 1 reveals that average condition numbers are biased upward for each sample size condition, and bias decreases as sample size increases.

Table 1.

Average Condition Number of Covariance Matrices and SDs by Sample Size

Samp Size	**S** Cond	$\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}$ Cond	$\hat{\boldsymbol{\Sigma}}_{\mathbf{R}}$ Cond	**S** Cond SD	$\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}$ Cond SD	$\hat{\boldsymbol{\Sigma}}_{\mathbf{R}}$ Cond SD
60	42.71	11.74	15.89	10.22	3.15	4.33
70	37.74	12.35	15.80	8.63	3.00	3.85
80	34.30	12.79	15.71	7.05	2.85	3.44
87	32.47	13.07	15.67	6.46	2.79	3.26
90	31.84	13.18	15.65	6.19	2.72	3.15
100	30.07	13.50	15.64	5.42	2.63	2.97
110	28.62	13.75	15.56	4.96	2.54	2.79
120	27.51	13.95	15.55	4.59	2.51	2.70
250	21.22	15.07	15.35	2.30	1.69	1.67
500	18.66	15.64	15.55	1.43	1.21	1.18
1000	17.27	15.81	15.66	0.97	0.90	0.87
2000	16.54	15.82	15.72	0.68	0.65	0.64
2500	16.36	15.79	15.70	0.58	0.56	0.55
5000	15.99	15.71	15.66	0.41	0.40	0.40
100000	15.46	15.45	15.44	0.09	0.09	0.09

Note. Samp = Sample, Cond = Condition, SD = Standard Deviation, Population condition number is approximately 15.35.

Upward bias of the sample covariance matrix condition number is substantial until *N* reaches at least 500 (ranging up to 178%), and not less than 10% until *N* = 2,000. On the other hand, the expectations of condition numbers for the MAP estimator $\hat{\boldsymbol{\Sigma}}_r$ reflect negative bias at *N*s smaller than 500, though in general bias is smaller in magnitude compared to **S**. Condition numbers change in a non-monotonic fashion as *N* increases. That is, as sample size increases, average condition number of $\hat{\boldsymbol{\Sigma}}_r$ increase until *N* = 2,500, then decrease. Finally, when examining MAP estimator $\hat{\boldsymbol{\Sigma}}_R$, in general condition numbers exhibit some positive bias, yet approach the population condition number as *N* increases. The condition numbers are fairly stable across sample size conditions compared to the other estimators, and increments of change are much smaller.

Considering variability of condition numbers, the standard deviation of **S** condition numbers decreases as *N* increases, and the difference between the spread of sample condition numbers and those of $\hat{\boldsymbol{\Sigma}}_r$ approaches zero as *N* increases. When *N* is less than 100, the spread of **S** condition numbers is about double the spread of condition numbers of $\hat{\boldsymbol{\Sigma}}_r$ and $\hat{\boldsymbol{\Sigma}}_R$, and at the smallest *N* condition, was about three times as large. The magnitudes of *SDs* of $\hat{\boldsymbol{\Sigma}}_r$ and $\hat{\boldsymbol{\Sigma}}_R$ are fairly similar in value although when *N* is less than 250, are more variable, whereas the opposite is true at larger *N*s. These differences are small compared to **S** especially at smaller sample sizes.

3.3.2 Solution Propriety

Poor solutions occurred for the traditional GLS estimator at the smallest of sample sizes ($N < 100$; see summary in Table 2).

Table 2.
Number of Poor Solutions by Sample Size

N	Non-Conv	High Iter	Negative Var
60	7	15	35
70	2	7	13
80	2	2	6*
87	0	1	2
90	0	0	1

Note. No poor solutions for larger *N* conditions,
* = 1 negative variance for RGLS and rGLS each.

Non convergences were fairly rare across the 1,000 replications per sample size, and mostly occurred at the smallest of sample size ($N = 60$). The number of replications resulting in at least one negative variance was calculated. Totals reflected these were less frequent as *N* increased. Results revealed no improper solutions were recorded for ML. Only a single case of a replication with a negative variance was recorded at the $N = 80$ condition for rGLS and RGLS.

3.3.3 Performance of Test Statistics

The target mean test statistic is the degrees of freedom (here, 87), asymptotically. Findings for ML and GLS are similar to previous simulation research, providing evidence for validation of the simulation (see Bentler & Yuan, 1999 for ML results for smaller samples; for larger sample sizes, see the normal theory condition 1 for ML and GLS in Hu et al., 1992). In addition, bias is less than 1% for the N = 100,000 condition, further validating the Monte Carlo procedure. Results in Table 3 reveal at the smallest of sample sizes, the test statistic for ML is inflated, whereas both GLS and rGLS are in general downwardly biased. When sample size was equal to 500, the bias is quite small across estimation methods (only about 6%, or less), and approaches zero as N increases. In comparison, RGLS performs the best of all four methods. Bias of the test statistic is less than 5%, regardless of sample size.

Table 3.

Test Statistic Simulation Results by Sample Size

Samp Size	T_{ML}	%Bias	T_{GLS}	%Bias	T_r	%Bias	T_R	%Bias
60	99.57	14.44	78.61	9.64	55.42	36.29	90.12	3.58
70	97.55	12.13	80.30	7.70	58.40	32.88	90.40	3.91
80	95.88	10.20	80.84	7.06	60.59	30.36	90.06	3.52
87	95.74	10.04	81.77	6.01	62.36	28.32	90.44	3.96
90	95.14	9.36	81.79	5.98	62.66	27.98	90.12	3.59
100	94.57	8.70	82.77	4.86	64.67	25.66	90.43	3.95
110	93.39	7.35	82.85	4.76	65.83	24.34	89.80	3.22
120	93.62	7.61	83.59	3.92	67.35	22.58	90.09	3.56
250	89.24	2.57	84.78	2.55	75.58	13.12	88.29	1.48
500	88.87	2.15	86.64	0.41	81.59	6.22	88.49	1.72
1000	87.80	0.92	86.76	0.28	84.22	3.20	87.72	0.82
2000	87.73	0.84	87.17	0.20	85.88	1.29	87.64	0.74
2500	88.03	1.18	87.61	0.71	86.57	0.50	87.98	1.13
5000	87.11	0.13	86.88	0.13	86.36	0.73	87.06	0.07
100000	87.19	0.21	87.18	0.21	87.15	0.18	87.19	0.21

Note . Target for bias calculations = 87; Replications for GLS N s = 60, 70, & 80, are 986, 995, & 998, respectively.

Variability of Test Statistics

Asymptotically, the standard deviation of test statistics should be about 13.19. Results in Table 4 reveal for ML, standard deviations (*SDs*) are too large at the smallest sample size conditions, while for GLS, they are too small.

Table 4.
Standard Deviations of Test Statistics across Replications

Samp Size	SD T_{ML}	SD T_{GLS}	SD T_r	SD T_R
60	15.48	10.92	8.72	12.84
70	14.98	12.09	9.24	13.22
80	14.26	11.65	9.22	12.73
87	14.40	12.29	9.74	13.17
90	14.51	11.93	9.58	12.93
100	14.68	12.51	10.17	13.35
110	13.97	12.54	10.12	13.00
120	14.58	13.06	10.75	13.73
250	12.83	12.26	10.95	12.65
500	12.98	12.81	12.07	13.05
1000	13.71	13.47	13.09	13.63
2000	12.89	12.68	12.67	12.93
2500	13.88	13.86	13.69	13.91
5000	12.75	12.74	12.66	12.77
100000	12.82	12.82	12.82	12.83

Note . Target is about 13.19; Replications for GLS *N* s = 60, 70, & 80 are 986, 995, & 998, respectively.

The values are in line with those of previous simulation results (see earlier citations), although at larger sample size conditions, SDs are a bit larger in this study, perhaps owing to the larger number of replications in this simulation. The pattern of results (reflecting GLS SDs are smaller than ML) is also similar to previous research concerning larger sample sizes. Results for rGLS reveals test statistics have smaller variability than even GLS, although this difference decreases as *N* increases. In general, SDs for RGLS are larger than rGLS, but smaller than ML. The differences narrow as *N* increases, and overall RGLS SDs are closest to the target SD.

3.3.4 Rejection Rates

Regarding empirical rejection rates given in Table 5, approximately 50 of 1,000 replications would be an ideal number of rejections, given an alpha of .05. The number of rejections is given in the left part of the table, and the rates in the right part.

Table 5.
Simulation Results Concerning Empirical Rejection Rates

Samp Size	ML Sum	GLS Sum	rGLS Sum	RGLS Sum	ML Emp Rej Rate	GLS Emp Rej Rate	rGLS Emp Rej Rate	RGLS Emp Rej Rate
60	235	2	0	65	.235	.002	.000	.065
70	199	15	0	80	.199	.015	.000	.080
80	179	14	0	67	.179	.014	.000	.067
87	172	21	0	79	.172	.021	.000	.079
90	148	22	0	67	.148	.022	.000	.067
100	146	21	0	74	.146	.021	.000	.074
110	131	24	0	62	.131	.024	.000	.062
120	125	26	2	80	.125	.026	.002	.080
250	68	29	4	53	.068	.029	.004	.053
500	60	48	20	63	.060	.048	.020	.063
1000	61	52	33	59	.061	.052	.033	.059
2000	50	44	37	49	.050	.044	.037	.049
2500	63	58	48	62	.063	.058	.048	.062
5000	47	47	41	48	.047	.047	.041	.048
100000	20/500	20/500	20/500	20/500	.040	.040	.040	.040

Note . Alpha = .05; Replications for GLS *N* s = 60, 70, & 80 are 986, 995, & 998, respectively.

Results reveal empirical rejection rates are quite large at the smallest of sample sizes for ML. GLS over accepts, with rejection close to zero at the smallest of *N*s, while rGLS never rejects at the smallest of sample sizes. When $N = 500$, there are more reasonable rejection rates for all methods except rGLS, (which needs closer to $N = 2000$). The RGLS empirical rejection rates are close to the nominal level, ranging from .04 to .08.

3.3.5 Overall Results for Simulation 1

Small *N*s associated with poor performance are also associated with larger, and more variable, condition numbers of **S**. In contrast, even at the smallest *N*s condition numbers for

regularized covariance matrices were closer to population counterparts. Nonetheless, rGLS performed with overly optimistic results, frequently over-accepting the true model. This was the case for GLS as well. Although rGLS performed well asymptotically, the method did not reflect much improvement until $N = 2000$. The empirical findings reflect $\bar{d}$ with α is not optimal for SEM in the small N conditions, therefore rGLS is not recommended at these sample sizes. On the other hand, results clearly reflect the superiority of RGLS, as it performs well across the wide range of N with the mean of test statistics close to the theoretical value and rejection rates close to nominal levels, even at the smallest of sample sizes. RGLS noticeably outperformed ML and GLS at Ns of 120 and less.

CHAPTER 4

SIMULATION STUDY 2

4.1 Regularization of the Sample Data for ADF Estimator

This simulation was carried out to illustrate the performance of regularization applied to the Asymptotically Distribution Free (ADF; Browne, 1982, 1984; Chamberlain, 1982; Dijkstra, 1981) method via the sample covariance matrix (YRADF). This novel method, described shortly, is evaluated relative to the typical ADF method for a variety of sample sizes under multivariate normality. Additionally, nonnormal distributions were considered due to the fact ADF does not require any distributional assumptions yet, as reviewed earlier in the introduction, doesn't perform well until sample sizes are quite large (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Huang & Bentler, 2015). Therefore it is important to examine whether YRADF can improve upon ADF performance. Moreover, this study incorporates the following methods as comparison conditions: Normal theory methods (GLS, ML, and RGLS from Study 1), and a robust method, the Satorra-Bentler mean scaled chi-square test statistic (SB, Satorra & Bentler, 1994).

Another contemporary regularization method, Ridge Generalized Least Squares (Yuan & Chan, 2016), is considered. Yuan and Chan use GLS in a broad sense to refer to a more specific ADF procedure (Browne, 1984). Generalized Least Squares does not directly refer to the GLS method considered in Study 1 that uses a fixed weight matrix estimated using **S** according to the normal distribution assumption. To avoid confusion here, Yuan and Chan's ridge ADF/GLS will be referenced here as RGLSI. The "I" refers to the identity matrix used in the ridge procedure. This method was reviewed in Chapter 2, however a more detailed description is provided. Yuan

and Chan propose to calculate a linear combination of the identity matrix and the sample fourth order moment matrix such that

$$\widehat{\boldsymbol{\Gamma}}_{\mathrm{I}} = (1 - a)\hat{\boldsymbol{\Gamma}} + a\mathbf{I}$$

Both matrices are weighted by a tuning parameter (a) in order to harness the stability of **I**. The objective is to improve stability of the matrix $\hat{\boldsymbol{\Gamma}}$ that is used in calculation of the weight matrix. When $a = 0$ the solution is the typical ADF solution. When $a = 1$ the solution is the Least Squares solution. Their method extends previous work discussed in the introduction on ridge SEM (see for example, Yuan & Chan, 2008).

Additionally, Yuan and Chan propose correcting the resulting test statistic as an approximation to the correct distribution. The RGLSI is a rescaled test statistic that applies a mean correction. The mean and variance adjusted method is referred to as AGLSI, where "A" refers to the fact test statistic and degrees of freedom are also adjusted. These types of corrections are described completely in Yuan and Bentler (2010). The optimal values of the tuning parameter were selected empirically by simulation across a grid of candidate tuning parameters ranging from zero to one. Those identified in the original paper will be employed here, except where noted later. Yuan and Chan also introduced a ridge method based on Diagonally Weighted Least Squares however this method is not studied further due to the fact the method yielded almost exact results as the GLSI method, and under some conditions produced less accurate solutions.

The study sought the following aims: First, to demonstrate improved differences in the condition of the sample fourth-order moment matrix ($\hat{\boldsymbol{\Gamma}}$) after regularization is applied to the sample covariance matrix. The elements of matrix $\hat{\boldsymbol{\Gamma}}$ are functions of the variances and covariances of **S**, as well as fourth-order moments about the mean. Therefore under normality

when those fourth-order moments are essentially zero, we should see improvement in condition numbers and furthermore perhaps some improvement under nonnormality. If this is the case it would be important to understand by what magnitude these matrices are improved.

Secondly, this study also aimed to demonstrate expected improvements in solution propriety. Iterations, negative variances, and convergence typical for ADF under small sample size conditions were expected to improve given the enhanced condition of the data matrix, which will in turn indirectly improve the weight matrix.

Thirdly, the simulation aimed to demonstrate YRADF improvement in model evaluation over ADF, given regularization improvements on **S**. Additionally, though generally ML and GLS perform well under asymptotic robustness theory under some nonnormality conditions such as in Case B, ML (GLS) has been found to over (under) reject at small sample sizes considered in this study. Furthermore, under Case C, these methods should not perform as well, since asymptotic theory is violated. It was unknown by how much YADF might improve or approach normal theory methods, especially given ADFs poor performance in comparison conditions.

Additionally, the Satorra-Bentler (SB) scaled chi-square test statistic was included to compare this well-performing corrected test statistic to regularization methods. This is of interest particularly because the SB method doesn't invert the matrix made up of sample fourth order moments and the objective of regularization methods is to better condition the matrix that is inverted with ADF. The SB scaled test statistic was expected to perform similar to normal theory methods under normality (Case A) and under nonnormal conditions (Cases B & C), the test statistic was expected to perform as well as ADF at larger *N*s and better than ADF at small *N*s given prior research (Bentler & Yuan, 1999; Hu, Bentler, & Kano, 1992, Tong & Bentler, 2013).

Again, it is unknown how much YRADF might improve on ADF therefore it is unknown if the method will perform better than SB.

Moreover, this study aimed to extend Study 1 in order to investigate the performance of RGLS under nonnormality, and in comparison to other normal theory methods. While certainly regularization was not expected to overcome misspecification in nonnormality conditions it was unknown if it might perform similarly to Study 1 under Case B like normal theory methods, and what happens under Case C.

Furthermore, under normality, RGLSI is expected to over reject the true model until sample sizes of at least $N = 1000$, while AGLSI should perform better across the range of sample sizes considered, though may over accept the true model according to simulation performance (Yuan & Chan, 2016). Interestingly, Yuan and Chan did not consider the same nonnormality conditions considered here, so no conclusions can be made with great certainty as to direction of results.

4.2 Details of Method and Simulation

In this section, I first review the data regularized ADF method (YRADF), then the simulation methodology. First, shrink eigenvalues of the sample covariance matrix as described in Chapter 2 (particularly, section 2.3) and use the MAP estimator, denoted $\tilde{\mathbf{S}}$ for these purposes and calculate a data matrix $\mathbf{Y}$ such that $\tilde{\mathbf{S}} = \mathbf{Y}'\mathbf{Y}N^{-1}$**.** This can be accomplished using the Cholesky decomposition as follows:

$$\mathbf{S}^{-1} = \mathbf{S}^{-.5\prime}\mathbf{S}^{-.5}$$

$$\tilde{\mathbf{S}} = \tilde{\mathbf{S}}^{.5}\,\tilde{\mathbf{S}}^{.5\prime},$$

then calculate $\mathbf{Y}$ such that

$$\mathbf{Y} = \mathbf{X}\mathbf{S}^{-.5}\tilde{\mathbf{S}}^{.5\prime}$$

Next, employ ADF estimation using this new input data, **Y.** Consider the general quadratic form of the GLS function

$$F_{ADF} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\hat{\boldsymbol{\Gamma}}^{-1}[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})],$$

where **s** is a column vector of $p^* = p(p+1)/2$ non-duplicated elements of the covariance matrix **S**, and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is a vector of the same size of non-duplicated elements of the model–implied matrix, and $\hat{\boldsymbol{\Gamma}}$ is a $p^* \times p^*$ matrix defined by the asymptotic distribution of the residual

$$\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{\Gamma}),$$

where $n = N - 1$, and $\hat{\boldsymbol{\Gamma}}^{-1}$ is the matrix of optimal weights. The typical element of $\hat{\boldsymbol{\Gamma}}$ consists of estimates of the second- and fourth-order product moments around the mean and is calculated as

$$[\gamma]_{ij,kl} = s_{ijkl} - s_{ij}s_{kl}, i \geq j, k \geq l \qquad (4.1)$$

, where the fourth order moment is calculated as

$$s_{ijkl} = N^{-1}\sum_{n=1}^{N}(y_{ni} - \bar{y}_i)(y_{nj} - \bar{y}_j)(y_{nk} - \bar{y}_k)(y_{nl} - \bar{y}_l)$$

and the second order moment is calculated as

$$s_{ij} = N^{-1}\sum_{n=1}^{N}(y_{ni} - \bar{y}_i)(y_{nj} - \bar{y}_j)$$

while the first order moment is calculated as

$$\bar{y}_i = N^{-1}\sum_{n=1}^{N}(y_{ni})$$

Next, proceed with data **Y** (with covariance matrix $\tilde{\mathbf{S}}$) as the input data, and traditional ADF estimation. This is the method is referred to here as YRADF.

For multivariate normality conditions, a confirmatory factor model ($\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon}$, where **x** is a vector of observed variables, $\boldsymbol{\Lambda}$ is a factor loading matrix, $\boldsymbol{\xi}$ is a vector of normally distributed common latent factors, and $\boldsymbol{\varepsilon}$ is a vector of unique unobserved factors) was employed

consisting of three common latent factors, each with five manifest variable indicators, and 15 unique errors independent from factors and one another. The population covariance matrix is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'+\boldsymbol{\Psi},$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are the covariance matrices of ξ and $\boldsymbol{\varepsilon}$, respectively. This model has been used in other Monte Carlo simulation studies (e.g. Yuan & Chan, 2016), and is used here to validate the simulation and as part of a recommended strategy to compare results to existing findings in literature (see Boomsma, 2013). Note the model factor loading matrix was slightly different than that of Study 1 in order to draw comparisons to recent results from Yuan and Chan (2016). Loadings are marginally different reflecting more variability across the loadings,

$\boldsymbol{\Lambda}'=$

$$\begin{bmatrix} 0.70 & 0.75 & 0.80 & 0.85 & 0.90 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 & 0.75 & 0.80 & 0.85 & 0.90 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.70 & 0.75 & 0.80 & 0.85 & 0.90 \end{bmatrix}.$$

Although loadings are generally about the same magnitude to those in Study 1, it is important to keep this difference in mind when making direct comparisons between studies. Variances of factors were set to one, with covariances between factors equal to 0.30, 0.40, and 0.50, respectively. Variances of unique factors were set to values such that under normality variances of observed variables are one.

Data generation was accomplished with R software (Version 3.2.3; R Core Team, 2015) using the simulation function in *simsem* package (Pornprasertmanit, Miller, & Schoemann, 2016) based on the population model described above with latent factor and errors exhibiting the following characteristics for three different cases: For Case A, no skewness or kurtosis; Case B, nonnormality of 3 factors with true factor kurtoses of -1, 2, 5 and 15 errors with true kurtoses of -1.0, 0.5, 2.5, 4.5, 6.5, -1.0, 1.0, 3.0, 5.0, 7.0, -0.5, 1.5, 3.5, 5.5 ,7.5, chosen to mimic

Condition 3 in Hu et al. (1992); and lastly, Case C, nonnormality with the same true kurtoses values as Case B, but also a fixed covariance matrix for variances and covariances of factors such that asymptotic robustness is violated. In other words, normal theory methods that might be robust in Case B should no longer, at least in theory, be robust in Case C (Satorra & Bentler, 1990; Satorra, 1992). The true kurtoses values are scaled such that zero reflects the absence of kurtosis. In all cases, factors and unique variates are independently distributed. These particular nonnormality conditions are of interest for this study because while distribution free methods should perform well across these cases, research reviewed in the introduction reveals this is not the case, except at very large sample sizes.

Data generation.

Multivariate nonnormal data were simulated using the Vale and Maurelli (1983) method. The nonnormal data propagation included latent variable and the indicator errors rather than simply rescaling the indicators directly, which can remove interdependence between latent variables and indicators (Kock, 2016) conflating simulation results. The method used here (often referred to as "sequential" referring to a chained equation approach; see *simsem* documentation for further detail) generates data for latent errors and variables directly, then for indicators, better preserving relationships. For these simulations, both latent factor scores and residual scores used to generate the data were saved and examined to validate process. The resulting indicator variables were also examined.

In general, the Vale and Maurelli method is one of the most common methods applied to simulate nonnormal data (Tadikammalla, 1980). This is a multivariate extension of the univariate Fleishman (1978) method. Commonly referred to as the power method, the Fleishman approach transforms a normally distributed variable, x, to a nonnormal by a the polynomial transformation

$$y = a + bx + cx^2 + dx^3,$$

where coefficients *a, b, c,* and *d* are constants determined so that (univariate) *y* has a particular desired skewness and kurtosis. Fleishman determined the simultaneous equations to solve for all coefficients, and Vale and Maurelli extended this method to allow for multiple variables and intercorrelations between them.

The general idea of the Vale and Maurelli method is as follows: After all constants are found for each variable as if they are univariate (via the Fleishman Polynomial), an intermediate matrix is calculated. This is the intermediate population matrix from which intermediate data are generated. It is important to remember this intermediate matrix will not be equivalent to the final desired matrix. This matrix is a pairwise matrix with intermediate entries found by calculating each correlation using Vale and Maurelli's third degree polynomial equation, using the constants calculated in the first step, and the pre-specified final correlation matrix of interest. This step is necessary because the correlations for nonnormality will differ in value than those under normality. In the final step the univariate equation is applied to each data point found from simulating the multivariate normal data from the intermediate correlational matrix. The final data set will have the structure of the desired pre-specified population matrix.

This simulation focused on sample sizes greater than 120, due to the fact p^* is the lower bound for inversion of the weight matrix for ADF and YRADF. Sample sizes were comparable to several conditions of Study 1, as well as conditions used by Yuan and Chan (*N*s = 150, 200, 300, 500, 1000, 2000, 3000, & 5000). Sample moments were collected to compare with population moments, including information with respect to matrix eigenvalues.

Each of 1000 independent samples of each sample size was analyzed with the *lavaan* package in R by specifying the correct model with 87 *df*, for normality Cases A and B, and 93 degrees of freedom for Case C, employing ML, GLS, and ADF estimation methods.

The method described previously in Study 1 was once again employed again for RGLS. For YRADF, the ADF method was employed as usual, given the data were the subject of modification rather than the weight matrix. A user-supplied weight matrix was supplanted for RGLSI and AGLSI methods. This matrix was the inverse of the modified matrix, $\hat{\mathbf{\Gamma}}_{\mathbf{I}}$**.** Each test statistic for AGLSI was evaluated based on the appropriate degrees of freedom calculated based on Equation 5 in Yuan and Chan (2016) for each individual replication. For multivariate normal data, optimal penalty values that were empirically determined by simulation in Yuan & Chan were applied in order to closely mimic simulation results. Sample size conditions above 3,000 were not tested in the original paper, so a very small tuning parameter is used in line with the decreasing trend across sample size for normal theory conditions (see Table 6). Due to the fact nonnormal conditions differ from Yuan and Chan, the exact penalty values would not necessarily be the exact optimal values, however interestingly values of the ridge penalty under nonnormality did not vary substantially across nonnormality conditions and sample size in Yuan & Chan (ranging .80 -.90). Therefore, the average value of all conditions was applied in this study for nonnormal conditions, a = .85. This value may not generalize perfectly because the conditions of the data, (a feature that may influence the parameter), are not exactly the same. However, findings demonstrated a lack of variability in the original simulation across normality conditions.

Table 6.
Tuning Parameters for Ridge GLS

N	Distribution	
	Normal	Nonnormal
150	.55	.85
200	.50	.85
300	.40	.85
500	.35	.85
1000	.25	.85
2000	.15	.85
3000	.10	.85
5000	.05	.85
10000	.01	.85

A pilot simulation was conducted to ensure methodology for ridge GLS were correctly implemented and coded, including normal-theory results for ML and RGLSI for *N*s ranging 60-3000, and with replications equal to 500, the same number as Yuan & Chan's simulations. Rejection rates are available in the Appendix. Results reflect similar outcomes, and differences between RGLSI are similar to differences between ML, with the exception of the smallest of sample size (which was not used in this study, $N = 60$).

For purposes of this study, the maximum iterations for convergence was set to 1000. This "liberal" number of iterations was used for investigation, given the novel application of the regularization method. If a replication did not converge within 100 iterations, the replication was flagged and recorded as an extreme number of iterations but allowed to continue up to 1000 iterations. *R* script was written so that number of iterations, convergence, and warnings regarding negative variances were recorded for each method.

Poor solutions were examined including non-convergences, the number of flagged large number of iterations, and number of results with at least one negative variance. Each test statistic was averaged across the number of converged replications and compared to expected values (the

degrees of freedom for the model assuming a χ^2 reference distribution, here 87 for Cases A and B, and 93 for Case C). Standard deviations of test statistics for each method were compared to the square root of twice the degrees of freedom (variance = 2*df*, SD = $\sqrt{2df} \approx 13.19$). Empirical rejection rates were calculated by summing number of replications for which the model was rejected based on the nominal alpha of .05. Since the fitted model is the correct model, this rate should approximate .05.

4.3 Case A: Normal Results

4.3.1 Condition of Matrices

Results are compiled in Table 7. The average condition numbers are displayed in the left half of the table, and standard deviations are displayed in the right half. As expected, sample condition numbers are larger than the population condition number (27.599). Furthermore, condition numbers of **S** are larger at smaller sample sizes and more variable. The average condition numbers of $\hat{\mathbf{\Gamma}}$ are quite huge, especially at small *N*s. These condition numbers are larger than those reported Huang & Bentler (2015), though of similar magnitude at smaller *N*s. Average condition numbers and their associated variability decreases as sample increases for the unmodified matrices. When regularizing the sample covariance matrix, results reflect $\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$ is somewhat improved over $\hat{\mathbf{\Gamma}}$. For example, at the smallest of sample sizes, condition numbers of $\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$, and the variability of $\kappa(\hat{\mathbf{\Gamma}}_{\mathbf{YR}})$ are about half the size of the unmodified matrix. However, the condition numbers of $\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$ are still quite large. Although condition numbers of $\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$ are improved over $\hat{\mathbf{\Gamma}}$ across all sample sizes, this difference decreases as *N* increases. An interesting pattern emerges when examining $\hat{\mathbf{\Gamma}}_{\mathbf{I}}$. The average condition numbers are quite small at small sample sizes, and equivalent to the original matrix at the largest sample size.

Table 7.
Condition Numbers of Matrices and SDs by Sample Size (Normal Case A)

Samp Size	$\mathbf{S}$ Cond	$\hat{\mathbf{\Gamma}}$ Cond	$\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$ Cond	$\hat{\mathbf{\Gamma}}_{\mathbf{I}}$ Cond	$\mathbf{S}$ Cond SD	$\hat{\mathbf{\Gamma}}$ Cond SD	$\hat{\mathbf{\Gamma}}_{\mathbf{YR}}$ Cond SD	$\hat{\mathbf{\Gamma}}_{\mathbf{I}}$ Cond SD
150	40.72	38746.70	19001.46	36.41	6.24	15159.18	7226.74	10.75
200	37.34	8684.65	4766.96	43.45	4.71	2541.97	1409.90	11.07
300	34.48	3409.81	2188.62	62.60	3.57	812.44	517.27	13.12
500	32.07	1878.56	1388.10	74.44	2.56	350.37	254.67	12.29
1000	30.27	1261.02	1076.44	112.54	1.73	175.20	145.23	13.30
2000	29.29	1036.32	949.08	189.93	1.22	106.34	95.95	16.51
3000	28.86	967.17	911.51	266.86	0.98	79.01	73.81	18.84
5000	28.53	912.20	881.13	744.09	0.74	58.99	52.96	45.89
10000	28.18	867.86	854.12	867.86	0.53	40.60	39.27	40.60

Note. Samp = Sample, Cond = Condition, SD = Standard deviation, Population condition number is approximately 27.60.

4.3.2 Solution Propriety

All replications (1000) converged regardless of sample size or methodology. At the smallest sample size (N = 150) the ADF methods revealed a small number of solutions with at least one negative variance, (ADF = 16 & YRADF = 6). Results for other Ns, and for all other methods across sample sizes reflected no boundary condition issues.

4.3.3 Performance of Test Statistics

Test statistic results for all methods are compiled in Table 8. AGLSI is included for completion in all tables however this method cannot be directly evaluated in summary statistics or compared to other methods except in regards to empirical rejection rates because the degrees of freedom change across replications.

Table 8.

Simulation Results of Test Statistics for Eight Methods by Sample Size (Normal Case A)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	92.16	92.77	84.46	88.71	229.53	196.92	45.41	99.09
200	89.97	90.45	84.48	87.66	167.65	147.62	49.11	96.56
300	89.54	89.83	86.03	88.18	132.35	120.34	56.58	95.68
500	88.87	89.02	86.66	87.96	110.67	104.47	60.43	92.66
1000	87.74	87.74	86.59	87.25	97.46	94.58	68.16	90.75
2000	87.81	87.81	87.27	87.60	92.67	91.23	75.45	89.87
3000	88.09	88.13	87.64	87.87	90.98	90.15	82.24	89.89
5000	87.02	87.03	86.82	86.95	88.84	88.38	88.50	88.68
10000	87.08	87.09	86.96	87.02	88.03	87.75	88.15	88.03

Note . Model is oblique 3 factor model in Yuan & Chan (2016), Replications = 1000, RGLS = Regularized GLS, YRADF = Y Regularized ADF, RGLSI = Ridge GLS Identity Matrix- Rescaled AGLSI = Ridge GLS Identity Matrix- Adjusted, Model *df* = 87, AGLSI *df* differ per replication.

When *N* = 10,000, very little bias is reflected for all methods. On average, test statistics are close to the expected values. Additionally, results parallel the normal condition in Hu, Bentler, and Kano (1992) for comparable sample sizes and methods (ML, GLS, SB and ADF), further validating the simulation. In general ML, GLS, SB, and RGLS perform well and similarly at larger sample sizes (> 500). At smaller *N*s, test statistics for ML and SB are slightly positively biased, GLS test statistics are slightly negatively biased, and RGLS reflects very little bias at all. These findings are also comparable to previous results from Study 1.

Traditional ADF results parallel results from the normal theory conditions in the literature (Hu, Bentler, & Kano, 1992; Huang & Bentler, 2015). Bias is quite high at smaller sample sizes (ranging up to 164%), consistent with research. Regarding regularization methods, the test statistic for YRADF is positively biased. However, YRADF performs better than traditional ADF in terms of the magnitude of bias, especially at small *N*s, though this difference decreases as *N* increases. Additionally, test statistics for RGLSI are also positively biased at smaller *N*s though bias improves as *N* increases.

Considering methods for which bias of test statistics is lower than 10% across all *N*s, only ML, SB, and GLS methods meet this criteria. Regarding ADF methods, bias is only smaller than 10% for RGLSI, YRADF, and ADF when *N*s are fairly large: at least 500, 1000, and 2000, respectively. Overall, results reflect under multivariate normality RGLSI performs the best of all of the ADF methods, while RGLS performs the best across all methods.

Variability of Test Statistics

Normal theory methods displayed in Table 9 are fairly similar with respect the variability to each other, and to expected standard deviations (13.19). Test statistics for GLS are less variable than ML and SB, while variability for RGLS is greater than GLS but smaller than ML. This is the same pattern as resulted from Study 1.

Table 9.
SDs of Test Statistics for Eight Methods by Sample Size (Normal Case A)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	RGLSI	AGLSI
150	13.97	14.01	12.73	13.27	51.92	46.48	18.86	8.57
200	13.53	13.65	12.76	13.10	34.05	30.68	18.03	9.16
300	13.46	13.53	13.14	13.40	24.62	22.97	17.37	10.27
500	13.09	13.13	12.86	13.05	19.01	17.60	16.69	10.87
1000	13.52	13.55	13.23	13.32	15.72	15.52	15.69	11.77
2000	12.93	12.94	12.96	13.00	14.22	13.91	14.32	12.13
3000	13.48	13.48	13.36	13.40	14.11	13.97	14.49	13.26
5000	12.89	12.88	12.90	12.92	13.20	13.28	13.15	13.12
10000	13.29	13.30	13.25	13.26	13.46	13.42	13.46	13.48

Note . Model is oblique 3 factor model in Yuan & Chan (2016), Replications = 1000, RGLS = Regularized GLS, YRADF = Y Regularized ADF, RGLSI = Ridge GLS Identity Matrix- Rescaled AGLSI = Ridge GLS Identity Matrix- Adjusted, Model df = 87, AGLSI df differ per replication.

In general, ADF and YRADF methods exhibit more variability than other methods. While test statistics for YRADF were less variable than ADF, this difference decreases as *N* increases. RGLSI are less variable than traditional ADF and YRADF. This is especially true at small *N*s.

4.3.4 Rejection Rates

At the largest sample size, methods perform similarly and close to the expected number of rejections (50, see Table 10).

Table 10.
Rejection Rates for Eight Methods by Sample Size (Normal Case 2A)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	108	115	34	74	1000	987	120	263
200	85	96	35	61	977	909	109	224
300	75	81	40	62	827	645	111	185
500	65	68	48	57	496	370	93	159
1000	58	57	46	51	209	154	87	121
2000	47	48	48	51	117	96	67	80
3000	55	57	48	48	101	92	78	90
5000	47	48	44	46	60	55	55	55
10000	52	52	50	51	64	64	52	64

Note . Model is oblique 3 factor model in Yuan & Chan (2016), Replications = 1000, RGLS = Regularized GLS, YRADF = Y Regularized ADF, RGLSI = Ridge GLS Identity Matrix- Rescaled AGLSI = Ridge GLS Identity Matrix- Adjusted, Model df = 87, AGLSI df differ per replication.

At smaller *N*s, ML and SB rejection rates were higher than nominal rates, while GLS rejected models too infrequently. Rejection rates decreased (increased) as *N* increased for ML(GLS). At larger *N*s, equal to or greater than 500, rejection rates were very similar for these two methods and close to the expected number of rejections. RGLS rejection rates are closest to nominal rates for most conditions, and performed optimally at small *N*s. Other methods reflect unacceptable rejection rates especially at small *N*s. When considering ADF and YRADF, almost all true models are rejected for *N*s below 300. However, rejection rates for YRADF reflects improvement that is more accelerated than ADF. Rejection rates for both RGLSI and AGLSI were improved over ADF and YRADF, especially at small *N*s. Results for these methods parallel findings in Yuan and Chan (2016). In general, AGLSI reflects empirical rejection rates closer to

expectation, performing better than RGLSI across sample sizes, however the regularization method with the MAP estimator, RGLS, performs better still.

4.4 Case B: Nonnormal Results

4.4.1 Data Distributional Characteristics

To assess the adequacy of the nonnormal data generation process, sample characteristics over replications were examined. Since data were generated to a specified amount of skewnesses and kurtoses, sample values for latent factors and errors were averaged across replications and were compared to desired values. Table 11 presents these descriptive summaries to examine the adequacy of the method by sample size.

Table 11.
Average Kurtosis by Sample Size Across Replications

N	Factor Kurtosis			Error Kurtosis
	F1	F2	F3	E15
150	-0.99	1.47	3.63	5.27
200	-0.98	1.71	3.85	5.32
300	-0.99	1.80	4.29	6.01
500	-1.00	1.88	4.64	6.54
1000	-1.00	1.92	4.86	7.00
2000	-1.00	1.96	4.76	7.08
3000	-1.00	1.98	4.82	7.22
5000	-1.00	1.97	4.81	7.31
10000	-1.00	2.00	4.92	7.53

Note. True values: F1 = -1.0, F2 = 2.0, F3 = 5.0, E15 = 7.5, E15 was selected as a representative for all errors.

The multivariate normal distribution skewness is zero and kurtosis is three but commonly the constant of three is subtracted so that both skew and kurtosis are zero. Such is the case here. These descriptive statistics suggest under nonnormality, kurtoses are under estimated at small sample sizes, which conforms to a well-known finding that true values are dampened given sample data (Kock, 2016; Olvera Astivia & Zumbo, 2014). It should be noted that other methods

(e.g. Headrick, 2002; Mair, Satorra, & Bentler, 2012) could allow for a wider scope of nonnormality conditions, and may reveal differing results had they been employed. At this stage of comparison of many methods, using VM allows for comparison to other studies since Vale and Maurelli technique has been the 'status quo'.

4.4.2 Condition of Matrices

Average condition numbers are displayed in Table 12. The average MAP estimator for sample covariance matrices ($\tilde{\mathbf{S}}$) are included for comparison, and since nonnormality was not considered in Study 1. The condition numbers of $\hat{\boldsymbol{\Gamma}}$ are quite huge, and also worsened with nonnormality. For example, at the smallest *N*, the average condition number is double the average condition number for the same sample size in Case A. Moreover, the variability of condition number of $\hat{\boldsymbol{\Gamma}}$ is about half the size of the average condition number under sample size conditions ranging up to about 2,000.

In general, $\hat{\boldsymbol{\Gamma}}_{\mathrm{YR}}$ improves upon $\hat{\boldsymbol{\Gamma}}$ in two ways: Improving condition of the matrices on average and reducing variability. In contrast to $\hat{\boldsymbol{\Gamma}}$, variability of condition numbers of $\hat{\boldsymbol{\Gamma}}_{\mathbf{YR}}$ are relatively smaller than the average condition number. However, $\hat{\boldsymbol{\Gamma}}_{\mathbf{YR}}$ condition numbers are still quite large, reflecting ill-conditioning even at the largest sample size condition.

Table 12.

Condition Numbers of Matrices and SDs by Sample Size (Nonnormal Case B)

Samp Size	$\mathbf{S}$ Cond	$\tilde{\mathbf{S}}$ Cond	$\hat{\Gamma}$ Cond	$\hat{\Gamma}_{YR}$ Cond	$\hat{\Gamma}_I$ Cond	$\mathbf{S}$ Cond SD	$\tilde{\mathbf{S}}$ Cond SD	$\hat{\Gamma}$ Cond SD	$\hat{\Gamma}_{YR}$ Cond SD	$\hat{\Gamma}_I$ Cond SD
150	40.05	25.79	79632.70	21506.96	59.33	5.20	3.41	41767.86	7076.98	34.93
200	37.22	26.01	18249.19	5537.31	70.85	4.37	3.08	9056.44	1810.37	31.10
300	34.53	26.41	7542.35	2537.90	111.83	3.29	2.59	3388.23	611.62	47.86
500	32.16	26.91	4154.60	1577.99	136.29	2.38	2.02	1749.41	289.05	53.48
1000	30.33	27.50	2673.06	1183.26	207.82	1.60	1.46	867.85	168.40	63.53
2000	29.22	27.76	2060.57	1036.89	345.29	1.03	0.99	424.04	139.09	68.13
3000	28.86	27.87	1888.79	993.59	488.97	0.85	0.83	310.69	110.74	76.48
5000	28.54	27.88	1744.26	960.36	773.41	0.67	0.63	238.79	68.67	99.88
10000	28.23	29.01	1635.35	944.99	1343.36	0.50	0.48	155.07	47.73	122.93

Note. Samp = Sample, Cond = Condition, SD = Standard Deviation, Population condition number of the data matrix is approximately 27.60.

4.4.3 Solution Propriety

At N = 150, the ADF condition had five replications that did not converge, six cases of high number of iterations and many cases of improper solution for ADF (53), though far fewer for YRADF (15). At N = 200, ADF had three replications of non-convergence, and 11 improper results.

4.4.4 Performance of Test Statistics

Patterns of results in Table 13 parallel Case A results for the normal theory methods. As expected, asymptotic properties of methods were retained under Case B. RGLSI performs similarly to SB, and RGLS follows suit, enjoying the least amount of bias. Traditional ADF results parallel results from Case A and nonnormal conditions from previous studies such that test statistics are still generally positively biased at small sample sizes. Bias is quite high at smallest sample sizes (ranging up to150%). Results reflect the ADF bias of test statistics

improves quickly as sample size increases. On average, in terms of test statistics, YRADF performs better than ADF, yet still not to expectation until *N*s of at least 2,000.

Table 13.

Simulation Results for Test Statistics for Eight Methods across Sample Size (Nonnormal Case B)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	89.10	89.94	81.36	86.05	217.46	198.40	31.34	91.86
200	90.82	91.45	85.03	88.70	164.83	152.68	33.29	90.85
300	89.33	89.63	85.49	88.06	129.99	121.57	35.87	90.48
500	88.78	89.03	86.73	88.28	110.23	104.58	36.10	89.28
1000	87.70	87.84	86.68	87.47	97.51	94.15	39.59	87.66
2000	87.15	87.21	86.48	86.88	91.70	88.80	40.42	86.64
3000	86.62	86.87	86.23	86.49	89.45	86.81	40.70	86.26
5000	87.05	87.07	86.85	87.01	88.85	86.19	41.37	86.83
10000	87.01	87.40	87.29	87.99	88.04	85.83	85.83	88.07

Note . Model is oblique 3 factor model in Yuan & Chan (2016), Replications = 1000, RGLS = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF, Model *df* = 87, Replications for *Ns* = 150 and 200 for ADF are 995 and 997, respectively.

Variability of Test Statistics

Results in Table 14 suggest normal theory methods are fairly similar with respect the variability to each other, though at small *N*s, test statistics for GLS are less variable than ML, while variability for RGLS is greater than GLS but smaller than ML, like SB. On the other hand, at small sample sizes ADF and YRADF exhibit more variability than all other methods, followed by RGLSI.

The variability of RGLSI is fairly constant across *N*s and about half the size of ADF and YRADF at small sample sizes. However, this pattern does not hold at larger *N*s given variability decreases for ADF and YRADF but in general, not for RGLSI.

Table 14.

SD of Test Statistics for Eight Methods by Sample Sizes (Nonnormal Case B)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	14.27	14.60	12.98	13.74	48.94	45.63	6.84	19.56
200	13.78	13.92	13.39	13.70	31.38	31.80	7.05	18.77
300	12.98	12.78	12.57	12.91	22.56	22.37	7.44	18.71
500	12.85	12.89	12.58	12.81	17.73	17.36	8.36	19.65
1000	12.64	12.55	12.67	12.79	14.58	14.76	8.51	18.71
2000	13.35	13.32	13.26	13.27	14.20	14.04	8.96	19.17
3000	13.23	13.24	13.20	13.24	13.83	13.74	8.65	18.31
5000	13.16	13.16	13.16	13.16	13.67	13.29	8.91	18.64
10000	12.08	12.06	12.13	12.16	12.21	12.01	10.16	12.16

Note . Model is oblique 3 factor model in Yuan & Chan (2016), 1000 Replications, RGLS = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF, Replications for *Ns* = 150 and 200 for ADF are 995 and 997, respectively.

4.4.5 Rejection Rates

Results in Table 15 reflects that normal theory methods (ML & GLS) as well as SB perform well at most sample sizes. RGLS superior to these methods at the smallest of sample sizes, with rejection rates closest to the nominal level at *N* =150.

Table 15.

Rejection Rates for Eight Methods by Sample Sizes (Nonnormal Case B)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	86	86	31	66	995	985	44	181
200	97	102	49	85	975	932	45	162
300	64	58	35	48	810	704	46	155
500	55	56	39	49	483	371	63	145
1000	49	44	43	51	201	139	57	121
2000	59	60	50	57	111	77	48	111
3000	51	49	52	53	74	60	41	113
5000	50	49	47	49	62	45	47	112
10000	40	40	40	40	44	44	42	112

Note . Model is the oblique 3 factor model in Yuan & Chan (2016), 1000 replications, RGLS = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF, Replications for *N* s = 150 and 200 for ADF are 995 and 997, respectively.

Although results reflect unacceptable rejection rates at small *N*s for ADF and YRADF. In fact, these are still worse than normal theory methods. Rejection rates greatly improve across sample sizes, converging toward 5%. Improvement of rejection rates for YRADF across sample sizes is accelerated over ADF and reaches rejection rates closer to nominal rates quicker than ADF, though these sample sizes are both quite large (N = 2000 for YRADF, N = 3000 for ADF).

Regarding other regularization methods, rejection rates of AGLSI are often improved over other methods. Although results reveal AGLSI tends to consistently under-reject while RGLSI over rejects the true model, (consistent with Yuan and Chan's 2016 findings), AGLSI doesn't under-reject to the extent of their simulation findings and actually performs well, especially at smaller *N*s similar in performance to the regularization method RGLS.

4.5 Case C: Nonnormal Results

4.5.1 Condition of Matrices

Condition numbers in Table 16 reveal that results are essentially the same as to those in Case B, as expected.

Table 16.

Condition Numbers of Matrices and SDs by Sample Size (Nonormal Case C)

Samp Size	$\mathbf{S}$ Cond	$\tilde{\mathbf{S}}$ Cond	$\hat{\Gamma}$ Cond	$\hat{\Gamma}_{YR}$ Cond	$\hat{\Gamma}_I$ Cond	$\mathbf{S}$ Cond SD	$\tilde{\mathbf{S}}$ Cond SD	$\hat{\Gamma}$ Cond SD	$\hat{\Gamma}_{YR}$ Cond SD	$\hat{\Gamma}_I$ Cond SD
150	40.39	25.82	81070.29	22403.57	13.67	5.32	3.46	46637.33	8059.52	6.96
200	37.41	26.02	18450.99	5626.91	13.68	4.38	3.09	9338.32	1782.62	5.89
300	34.53	26.41	7542.35	2537.90	14.21	8.25	2.04	3388.23	611.21	5.71
500	32.16	26.91	4154.59	1579.99	14.25	2.38	2.02	1749.41	289.05	5.23
1000	30.33	27.50	2673.06	1183.55	14.18	1.46	1.60	867.85	186.39	4.02
2000	29.23	27.76	2060.57	1036.89	13.80	1.04	0.99	424.04	139.09	2.54
3000	28.86	27.87	1888.80	993.59	13.83	0.85	0.83	310.69	110.74	2.01
5000	28.55	27.94	1739.52	962.85	13.75	0.67	0.66	233.53	66.23	1.52
10000	28.20	27.90	1635.38	942.32	13.85	0.47	0.46	154.38	55.10	1.17

Note. Samp = Sample, Cond = Condition, SD = Standard Deviation, Population condition number is approximately 27.60.

4.5.2 Solution Propriety

All methods reached convergence for each replication. Problem solutions occurred at N = 150 for methods concerning ADF. Negative variances occurred for 66 replications for traditional ADF, 19 for YRADF, and in one case for RGLSI/AGLSI. At N = 200, the number was greatly reduced to two cases for ADF, and one for YRADF. Additionally, there were no other issues for other methods or at other sample sizes.

4.5.3 Performance of Test Statistics

Test statistic results for all methods are compiled in Table 17. Recall for Case C, the degrees of freedom of the model is 93. At smaller sample sizes, biases are much more noticeable for normal theory methods under Case C then for other conditions. Additionally, RGLS results reflect upward biased of test statistics at smaller sizes. YRADF also does not improve over ADF except at the very smallest *N*. Additionally, results for RGLSI are inconsistent across sample sizes as results reflect positive bias at smallest sample size and negative bias at largest sample sizes.

Table 17.

Simulation Results for Test Statistics for Eight Methods by Sample Size (Nonormal Case C)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	97.82	97.89	91.31	99.39	285.60	276.69	24.85	101.61
200	96.90	96.78	92.32	98.40	196.14	194.25	24.96	97.63
300	95.44	95.02	92.43	96.71	146.71	146.65	25.48	95.05
500	94.87	94.26	93.28	95.93	121.19	121.25	26.04	93.41
1000	93.85	93.14	93.10	94.46	105.68	106.45	26.49	91.77
2000	93.18	92.39	92.60	93.28	98.34	98.15	26.32	89.51
3000	92.75	91.95	92.42	92.88	95.82	95.52	26.51	89.62
5000	92.96	92.12	92.82	93.09	94.82	94.14	26.57	89.34
10000	92.87	92.03	92.77	92.92	93.43	92.49	26.87	90.03

Note . Model is oblique 3 factor model in Yuan & Chan (2016), Replications = 1000, RGLS = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF, Model df = 93 (df for AGLSI differs across replications).

Variability of Test Statistics

Similar to Cases A and B, results in Table 18 suggest normal theory methods are fairly similar with respect the variability to each other, though at small *N*s, variability for RGLS is greater than both the normal theory and SB methods. On the other hand, a similar pattern occurs under Case C to that of Case B. Small sample sizes ADF and YRADF exhibit more variability than all other methods, followed by RGLSI. Here again, variability of RGLSI is fairly constant across *N*s and about half the size of ADF and YRADF at small sample sizes. However, this pattern does not hold at larger *N*s given variability decreases for ADF and YRADF but in general, not for RGLSI. Unlike Case B, under Case C variability of ADF and YRADF are quite similar to one another.

Table 18.
SD of Test Statistics for Eight Methods by Sample Size (Nonormal Case C)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	14.73	14.77	14.31	15.75	70.73	69.77	7.16	29.06
200	14.44	14.42	14.23	15.15	39.74	41.44	6.63	25.52
300	13.57	13.41	13.61	14.25	25.82	27.83	6.49	24.24
500	13.28	13.25	13.25	13.66	19.34	19.44	7.03	24.97
1000	13.14	12.95	13.08	13.51	15.55	16.42	7.18	24.49
2000	13.87	13.75	13.86	13.96	15.03	15.13	7.01	23.79
3000	13.69	13.62	13.70	13.77	14.45	14.59	6.99	23.58
5000	12.95	12.82	12.93	12.97	13.58	13.35	6.69	22.41
10000	12.94	12.82	12.96	12.98	13.16	13.32	6.68	22.93

Note . Model is oblique 3 factor model in Yuan & Chan (2016), 1000 Replications, RGLS = = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF.

4.5.4 Rejection Rates

With regards to rejection rates, Table 19 reflects RGLS may be sensitive to Case 3 misspecification at the smallest of sample sizes, performing in a similar manner to normal theory methods, with too many rejections. This is not the case for GLS and AGLSI. These methods both

perform well across the sample size conditions. YRADF does not enjoy much improvement over ADF. Under Case C, both have about an equal number of rejection rates. Although RGLSI improves on ADF, it generally over-rejects across sample sizes.

Table 19.
Empirical Rejection Rates for Eight Methods by Sample Size (Nonormal Case C)

Samp Size	ML	SB	GLS	RGLS	ADF	YRADF	AGLSI	RGLSI
150	99	100	53	141	1000	1000	79	246
200	96	96	61	121	993	985	59	213
300	66	60	44	83	878	868	42	165
500	59	50	40	76	563	568	47	156
1000	47	40	52	66	237	258	40	166
2000	63	51	54	60	128	109	25	122
3000	51	47	50	53	82	78	39	135
5000	46	36	46	50	60	56	40	102
10000	46	42	50	51	52	54	34	108

Note . Model is the oblique 3 factor model in Yuan & Chan (2016), 1000 replications, RGLS = Regularized GLS, RGLSI = Ridge GLS Identity Matrix, YRADF = Y Regularized ADF.

4.6 Simulation 2: Overall Results

Findings for YRADF at N =10,000 (as close to asymptotic N as considered in this simulation) indicate the procedure was correctly implemented as results are very close to theoretical values. Though YRADF improved ADF quantitatively, the two methods are qualitatively similar to one another in performance. The reason for this may be explained by examining condition numbers. Only a very modest improvement is found when examining condition numbers of $\hat{\mathbf{\Gamma}}_{YR}$, and largely the condition values still reflect ill-conditioning. Also, while YRADF improves on ADF under normal conditions, this finding does is not as pronounced in Case B. Under Case C, little difference is found between the traditional method and the regularized version. Again, condition numbers gives us insight in regards to this finding, as the matrices are very ill-conditioned even at larger sample sizes, therefore any improvement in

the regularized **S** is not reflected further in the sample fourth order covariance matrix. Some improvement was evident in solution propriety, given fewer instances of negative variances.

Under the normal distribution condition RGLS improved upon normal theory methods, even with slightly more variable factor loadings. Additionally, results revealed RGLS outperforms SB even at smallest of *N*s. RGLS also seems to be robust under Case B, however under Case C may not offer an advantage at the smallest *N*s.

This study also offers some new insights on RGLSI and AGLSI methods, which were examined under new nonnormal conditions and a larger range of sample sizes and number of replications than initial simulations. Caution should be applied to these results given tuning parameters were adopted from the original study for simulation efficiency. Findings overall suggest that under Case A, AGLSI performs similarly to SB, and perhaps even better at smaller sample sizes under nonnormal cases. Results also suggest AGLSI is generally preferable to RGLSI across conditions.

CHAPTER 5

Simulation Study 3

5.1 Asymptotic Distribution Free & Correction methods

This simulation was carried out to illustrate the potential for broader application of MAP estimator beyond the sample covariance matrix. Here, regularization is applied to the asymptotic covariance matrix of sample fourth-order moments, useful for calculations of the weight matrix for the Asymptotic Distribution Free (ADF) method, as well as calculations of corrections to Maximum Likelihood test statistics. These proposed approaches will be detailed next and were evaluated in a large-scale simulation relative to their traditional counterparts under a variety of sample sizes.

It is a strength of the ADF method that it does not require any distributional assumptions however as reviewed earlier, ADF methods do not perform well until sample sizes are quite large (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Huang & Bentler, 2015). Therefore, it is important to examine whether regularization can offer insight on why this method doesn't achieve its desired asymptotic properties, and whether regularization can improve ADF performance. Thus, nonnormal conditions for the simulation study are considered as well.

This study builds on the previous two studies. Like Study 2, the MAP estimator approach is once again applied to ADF methodology, however instead of indirectly affecting the weight matrix via the sample covariance matrix, it aims to directly improve the weight matrix like the approach in Simulation Study 1. In this case, the covariance matrix of interest is the asymptotic covariance matrix, a matrix of much larger dimension ($p^* x\ p^*$) than the $p\ x\ p$ sample covariance matrix. Therefore, an important first step is to demonstrate the differences in the condition of this special kind of covariance matrix before and after regularization. Secondly, like previous

simulations, this study also aims to demonstrate improvements in solution propriety. High iterations, negative variances, and convergence problems are typical for ADF under small sample size conditions, and these were expected to improve given the improvement of the weight matrix.

Third, the regularized matrix of fourth order sample moments is extended to other methods and studied. This includes another ADF method referred to here a Residual Test Statistic (Browne, 1984; RES), and correction methods including Satorra-Bentler (1994; SB) scaled Chi-square test statistic, as well as the mean scaled and variance adjusted statistic (SBMV). Detailed descriptions and equations follow below. Overall, it is expected that results will reveal better behaved test statistics and improved empirical rejection rates, particularly at small sample sizes and these results should not worsen under nonnormal conditions due to the same rationales presented for the similar simulation Study 2. Similar results are expected given the method is the same, but the application of the estimator is more direct so could be improved.

5.2 Details of Method and Simulation

The application of the MAP estimator is quite straightforward in this case. For the sample counterpart of $\boldsymbol{\Gamma}$, ($\hat{\boldsymbol{\Gamma}}$) determine the corresponding MAP estimator $\hat{\boldsymbol{\Gamma}}_{\mathrm{R}}$ for appropriate calculations related to test statistics of interest for the following methodologies: ADF, RES, SB, and SBMV. Recall the general quadratic form of the GLS function

$$F_{ADF} = [\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\hat{\boldsymbol{\Gamma}}^{-1}[\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})],$$

where $\hat{\boldsymbol{\Gamma}}$ is a $p^* \times p^*$ matrix defined by the asymptotic distribution of the residual

$$\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{\Gamma})$$

where $n = N - 1$. The typical element of $\hat{\boldsymbol{\Gamma}}$ consists of estimates of the second- and fourth-order product moments around the mean and is calculated as

$$[\gamma]_{ij,kl} = s_{ijkl} - s_{ij}s_{kl}, i \geq j, k \geq l \qquad (5.1)$$

, where

$$s_{ijkl} = N^{-1}\sum_{n=1}^{N}(y_{ni} - \bar{y}_i)(y_{nj} - \bar{y}_j)(y_{nk} - \bar{y})(y_{nl} - \bar{y}_l)$$

Browne (1982, 1984) developed and introduced another asymptotic distribution free test statistic, however this alternative was not studied empirically until much later (Satorra & Bentler, 1991; Yuan & Bentler, 1998; Bentler & Yuan, 1999). This method is applied to the normal theory maximum likelihood estimator to obtain a test statistic that is asymptotically chi-square distributed, even under distributional violations given correct model specification. This is a residual-based test statistic defined theoretically by Browne (1984; see Proposition 4) that includes a sandwich triple product in its calculations. Let vech (·) be an operator that transforms a symmetric matrix into a vector by selecting the non-duplicated elements in the matrix such that $\mathbf{s} = \text{vech}(\mathbf{S})$ and $\boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}}) = \text{vech}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}))$. Designate the $p^* \times q$ Jacobian matrix, (where q is the estimated parameters) corresponding to $\boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$ as $\dot{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\theta}})$. When the $p^* \times p^*$ matrix $\widehat{\boldsymbol{\Gamma}}^{-1}$ is non-singular,

$$\text{T}_\text{B}(\widehat{\boldsymbol{\theta}}) = n\hat{\mathbf{e}}'[\widehat{\boldsymbol{\Gamma}}^{-1} - \widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\theta}})\{\dot{\boldsymbol{\sigma}}'(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\theta}})\}^{-1}\dot{\boldsymbol{\sigma}}'(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Gamma}}^{-1}]\hat{\mathbf{e}}, \qquad (5.2)$$

where $\hat{\mathbf{e}} = \mathbf{s} - \boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$, and the discrepancy between the data and the model or residual, is estimated by any consistent estimator such as ML. In practice, the elements of $\widehat{\boldsymbol{\Gamma}}$ are defined in Eq. 5.1. This test statistic will be referred to as Browne's residual-based ADF test statistic, abbreviated here as RES, which evaluates the model based on the chosen estimates.

Other correction methods are considered next. In the case of ML, the scaled test statistic is defined as

$$\overline{\text{T}}_\text{M} = \frac{\text{T}_\text{ML}}{c} \qquad (5.3)$$

The scaling constant *c* is estimated by

$$\hat{c} = \frac{\text{tr}(\widehat{\mathbf{U}}\widehat{\mathbf{\Gamma}})}{df}, \tag{5.4}$$

, whose $\widehat{\mathbf{\Gamma}}$ elements are defined in Eq. 5.1 and $\widehat{\mathbf{U}}$ is a consistent estimator for the residual weight matrix under the model when there are no constraints

$$\mathbf{U} = \mathbf{W} - \mathbf{W}\,\dot{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\theta}})\{\dot{\boldsymbol{\sigma}}'(\widehat{\boldsymbol{\theta}})\mathbf{W}\,\dot{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\theta}})\}^{-1}\dot{\boldsymbol{\sigma}}'(\widehat{\boldsymbol{\theta}})\mathbf{W},$$

where $\dot{\boldsymbol{\sigma}}$ is the Jacobian matrix evaluated at $\widehat{\boldsymbol{\theta}}$ and depends on the model, tr(·) is the trace of the matrix product or the sum of the non-null eigenvalues, and **W** is a weight matrix. A consistent estimator for **W** is

$$\mathbf{W} = .5\mathbf{K}_{\mathbf{P}}'(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{K}_{\mathbf{P}}$$

, where $\mathbf{K}_{\text{p}}$ is a p^2 x p^* transition matrix and $\widehat{\boldsymbol{\Sigma}}$ is a matrix that converges in probability to $\boldsymbol{\Sigma}$, such as **S** in the case of GLS or $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ in the case of RWLS which is asymptotically equivalent to ML and the estimator used here.

In order to better approximate the Chi-square distribution, Satorra and Bentler (1994) also consider a variance adjustment, in addition to mean scaling. This mean scaled and variance adjusted test statistic is an extension of $\overline{\text{T}}_M$ often denoted as $\overline{\overline{\text{T}}}_{\text{MV}}$, includes a Satterthwaite second moment adjustment to the degrees of freedom. The test statistic is defined as

$$\overline{\overline{\text{T}}}_{\text{MV}} = \frac{v}{tr(\widehat{\mathbf{U}}\widehat{\mathbf{\Gamma}})}\text{T}, \tag{5.5}$$

,where here, T is T_{ML} and

$$v = \left[tr\left(\widehat{\mathbf{U}}\widehat{\mathbf{\Gamma}}\right)\right]^2 / tr[\left(\widehat{\mathbf{U}}\widehat{\mathbf{\Gamma}}\right)^2], \tag{5.6}$$

which represents the estimated degrees of freedom. The mean and variance of $\overline{\overline{\text{T}}}_{\text{MV}}$, match a Chi-square distribution with the Satterthwaite estimated degrees of freedom as an approximation under the null hypothesis.

This MAP estimator $\hat{\boldsymbol{\Gamma}}_R$ is applied and studied in the following ways:

1) Use data **X** with covariance matrix **S** as input data, and calculate $\hat{\boldsymbol{\Gamma}}$. Apply shrinkage to $\hat{\boldsymbol{\Gamma}}$ resulting in MAP estimator $\hat{\boldsymbol{\Gamma}}_R$, and proceed with classical ADF estimation. Since $\hat{\boldsymbol{\Gamma}}^{-1}$ is the matrix of optimal weights, invert the matrix $\hat{\boldsymbol{\Gamma}}_R$, and supplant it as the weight matrix. This method is referred to as RADF.
2) Use data **X** with covariance matrix **S** as the input data, and proceed with ML estimation. The residual-based ADF test statistic (T_B) is considered in Eq. 5.2, now based on the MAP estimator $\hat{\boldsymbol{\Gamma}}_R$ instead of $\hat{\boldsymbol{\Gamma}}$.This methods is referred to as RRES.
3) Use data **X** with covariance matrix **S** as the input data, and proceed with ML estimation. The Satorra Bentler Mean Scaled Chi-square test statistic ($\overline{T}_{SB}$) is computed with the MAP estimator $\hat{\boldsymbol{\Gamma}}_R$ in place of $\hat{\boldsymbol{\Gamma}}$ in Eq. 5.4. This method is referred to as RSB.
4) Additionally, the Satorra Bentler mean scaled and variance adjusted test statistic ($\overline{\overline{T}}_{SB}$) is computed based on Eqs. 5.5 and 5.6 using the MAP estimator $\hat{\boldsymbol{\Gamma}}_R$ in place of $\hat{\boldsymbol{\Gamma}}$. This method is referred to as RSBMV.

A Monte Carlo simulation study was carried out to illustrate and compare the performance of these eight test statistics associated to traditional and regularized ADF and correction methods (ADF, RADF, RES, RRES, SB, RSB, SBMV, RSBMV). Like Study 1, the choice of $\hat{\alpha}_r$ could just as well have considered over $\hat{\alpha}_R$ instead, in turn leading to $\hat{\boldsymbol{\Gamma}}_r$ instead of $\hat{\boldsymbol{\Gamma}}_R$ (i.e., rADF, rRES, etc.). Considering the good performance of alpha ($\hat{\alpha}_R$) in Study 1, we begin with this approach, though will return to this idea in the results section.

These methods are also compared to those in Study 1 (GLS and ML) under conditions of multivariate normality, and non-normality. For multivariate normality conditions, a confirmatory

factor model ($\mathbf{x} = \mathbf{\Lambda}\xi + \boldsymbol{\varepsilon}$, where $\mathbf{x}$ is a vector of observed variables, $\mathbf{\Lambda}$ is a factor loading matrix, ξ is a vector of normally distributed common latent factors, and $\boldsymbol{\varepsilon}$ is a vector of unique unobserved factors) was employed consisting of three common latent factors, each with five manifest variable indicators, and 15 unique errors independent from factors and one another. The population covariance matrix is given by

$$\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda\Phi\Lambda}'+\Psi,$$

where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are the covariance matrices of ξ and $\boldsymbol{\varepsilon}$, respectively. Note the model factor loading matrix was the same of that of Study One,

$\mathbf{\Lambda}'=$

$$\begin{bmatrix} 0.70 & 0.70 & 0.75 & 0.80 & 0.80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.70 & 0.70 & 0.75 & 0.80 & 0.80 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.70 & 0.70 & 0.75 & 0.80 & 0.80 \end{bmatrix}$$

. Variances of factors were set to one, with covariances between factors equal to 0.30, 0.40, and 0.50, respectively. Variances of unique factors were set to values such that under normality, variances of observed variables are one. This model has been used in other Monte Carlo simulation studies (e.g. Hu, et al., 1992), and is used here to validate the simulation and as part of a recommended strategy to compare results to existing findings in literature (see Boomsma, 2013).

Data generation was accomplished with *R* software (Version 3.2.3; R Core Team, 2015) using the simulation function in *simsem* package (Pornprasertmanit, Miller, & Schoemann, 2016) based on the population model described above with latent factor and errors exhibiting the following characteristics for three different cases: For Case A, no skewness or kurtosis; Case B, nonnormality of 3 factors with true factor kurtoses of -1, 2, 5 and 15 errors with true kurtoses of -1.0, 0.5, 2.5, 4.5, 6.5, -1.0, 1.0, 3.0, 5.0, 7.0, -0.5, 1.5, 3.5, 5.5 ,7.5, chosen to mimic Condition 3 in Hu et al. (1992); and lastly, Case C, nonnormality with the same true kurtoses

values as Case B, but also a fixed covariance matrix for variances and covariances of factors such that asymptotic robustness is violated. In other words, normal theory methods that might be robust in Case B should no longer, at least in theory, be robust in Case C (Satorra & Bentler, 1990; Satorra, 1992). The true kurtoses values are scaled such that zero reflects the absence of kurtosis. In all cases, factors and unique variates are independently distributed. These particular nonnormality conditions are of interest for this study because while distribution free methods should perform well across these cases, research reviewed in the introduction reveals this is not the case, except at very large sample sizes. The multivariate nonnormal data were simulated using the Vale and Maurelli (1983) method described previously in Study 2.

This simulation focused on sample sizes greater than 120, due to the fact p^* is the lower bound for inversion of the weight matrix for ADF. Sample sizes were comparable to several conditions of Study 1, as well as sample size conditions in Hu et al. (*N*s = 150, 200, 500, 1000, 2000, 2500 3000, & 5000). Sample moments were collected to compare with population moments, including information with respect to matrix eigenvalues.

Each of 1000 independent samples of each sample size was analyzed with the *lavaan* package in R by specifying the correct model with 87 *df*, employing ML, GLS, and ADF estimation methods. Each test statistic for the SBMV were evaluated based on the appropriate degrees of freedom calculated based on Eq. 5.6 for each individual replication. Therefore in the results section only empirical rejection rates for SBMV will be discussed and compared to other methods. For the nonnormal Case C, the *df* is 93.

For purposes of this study, the maximum iterations for convergence was set to 1000. This "liberal" number of iterations was used for investigation, given the novel application of the regularization method. If a replication did not converge within 100 iterations, the replication was

flagged and recorded as an extreme number of iterations but allowed to continue up to 1000 iterations. R script was written so that number of iterations, convergence, and warnings regarding negative variances were recorded for each method.

Poor solutions were examined including non-convergences, the number of flagged large number of iterations, and number of results with at least one negative variance. Each test statistic was averaged across the number of converged replications and compared to expected values (the degrees of freedom for the model assuming a χ^2 reference distribution). Standard deviations of test statistics for each method were compared to the square root of twice the degrees of freedom (variance = 2*df*, SD = $\sqrt{2df} \approx 13.19$). Empirical rejection rates were calculated by summing number of replications for which the model was rejected based on the nominal alpha of .05. Since the fitted model is the correct model, this rate should approximate .05.

5.3 Case A: Normal Results

A small pilot study was conducted to decide on the specification for the mixture parameter for the MAP estimator. In Study 1, alpha ($\hat{\alpha}_R$) was found to be favorable over alpha ($\hat{\alpha}_r$), therefore may also be favored across these studies. On the other hand, the matrix considered in this study is quite a bit larger than the sample covariance matrix in Studies 1 and 2. If the number of non-zero eigenvalues are equal to the number of elements on the diagonal of a square matrix, there are close to 10 times the number of eigenvalues for $\hat{\mathbf{\Gamma}}$ than **S** in this case (120 & 15, respectively).

To empirically examine potential for performance, the two mixture parameters were auditioned. A condition was selected that was small enough to be low cost in terms of computing time and to reflect differences in outcomes of interest, while at the same time would not hindered by problems with solution propriety (N = 250). Both mixture parameters were tested and

compared in terms of condition numbers of the MAP estimators. Additionally, the MAP estimator was inverted and employed as the weight matrix for ADF, (the regularized ADF method described above, rADF & RADF) and results across 1,000 replications were compared (see Table 20) with the goal of selecting the mixture parameter with the best potential.

Table 20.
Comparison of Alpha Mixture Parameter for Study 3

Alpha	Condition Number	SD	ADF	Emp Rej
-	2087.79	532.43	143.78	911
$\hat{\alpha}_R$	442.41	83.52	140.19	903
$\hat{\alpha}_r$	269.48	61.39	93.68	147

Note. Replications = 1000, N = 250, Case A.

Results suggest that the condition number of the unmodified matrix is quite large and >> than 1, indicating ill-conditioning. Both MAP estimators' condition numbers are reduced, though more so for $\hat{\alpha}_r$ than $\hat{\alpha}_R$. Moreover, in terms of variability all methods are relatively similar. In terms of test statistics, both methods improve upon the traditional ADF method. When computing the MAP estimator using $\hat{\alpha}_r$ the ADF rejection rate is closer to the expected 50, while the MAP estimator computed using $\hat{\alpha}_R$ resulted in only a small improvement, with a rejection rate well above the nominal value. These results are somewhat similar to results in Study 1 suggesting $\hat{\alpha}_r$ resulted in lower conditions numbers and smaller test statistics on average than $\hat{\alpha}_R$ However, this phenomena is not as severe in this case. This could be due to several factors, since both the method and the size of the matrix differ across these studies. Nevertheless, based on these results, the $\hat{\mathbf{\Gamma}}_r$ matrix calculated using $\hat{\alpha}_r$ is employed next for the normal condition (Case A) to keep simulations manageable.

5.3.1 Condition of Matrices

Results are compiled in Table 21. Condition numbers are displayed on the left half, while standard deviations, and coefficients of variation are displayed in the right half of the tabled results. The coefficient of variation (c_v) is a relative standard deviation calculated as the ratio of the standard deviation to the mean. It is useful for comparing the degree of variation from one data series to another, especially when means are quite variable. Such is the case here, as the average condition numbers are quite huge at small *N*, especially relative to those at larger sample sizes making comparisons of variability using the typical SD challenging.

An empirical condition number based on 100,000 replications for $\hat{\mathbf{\Gamma}}$ was calculated for comparison benchmark and was found to be large, 254.94. Condition numbers of $\hat{\mathbf{\Gamma}}$ are quite huge and variable at small *N*s, though these decease as sample increases for the unmodified matrices. The $\hat{\mathbf{\Gamma}}_r$ is quite improved over $\hat{\mathbf{\Gamma}}$, especially when at the smallest of sample sizes, though the difference decreases as *N* increases. The coefficients of variation reflect condition numbers are less dispersed at larger sample sizes than smaller ones, but they also indicate regularized matrices have smaller dispersion of condition numbers relative to the unmodified matrices, though differences are not very pronounced except at the smallest of sample sizes.

Table 21.

Average Condition Numbers, SDs, and Coefficient of Variations of $\hat{\Gamma}$ *and* $\hat{\Gamma}_R$ *by Sample Size (Normal Case A)*

Samp Size	$\hat{\Gamma}$ Cond	$\hat{\Gamma}_R$ Cond	$\hat{\Gamma}$ SD Cond	$\hat{\Gamma}_R$ SD Cond	$\hat{\Gamma}$ c_v	$\hat{\Gamma}_R$ c_v
150	18459.87	181.84	7160.42	53.79	0.39	0.30
250	2087.80	269.48	532.43	61.39	0.26	0.23
500	742.46	317.04	132.13	51.20	0.18	0.16
1000	455.46	307.77	59.74	38.12	0.13	0.12
2000	352.19	290.06	33.89	27.03	0.10	0.09
2500	332.16	284.42	28.78	23.80	0.09	0.08
3000	318.75	279.98	24.96	21.53	0.08	0.08
5000	294.96	272.82	18.27	16.71	0.06	0.06
10000	275.68	265.03	12.75	12.20	0.05	0.05

Note. Samp = Sample, Cond = Condition, SD = Standard Deviation, c_v = Coefficient of Variation, Emprical Condition Number = 254.94.

5.3.2 Solution Propriety

At the smallest sample size, two replications for the ADF method did not converge. All other replications had converged results regardless of sample size or methodology. Additionally, there were a small number of solutions with at least one negative variance, (ADF = 22) and three replications with high number of iterations. Results for other *N*s, and for all other methods across sample sizes did not reflect any problem solutions.

5.3.3 Performance of Test Statistics

Test statistic results for all methods are compiled in Table 22. When N = 10,000, very little bias is reflected. On average, test statistics are close to the expected values at that large *N*. Results parallel the normal condition in Study 2 and previous literature (e.g. Hu, Bentler, and Kano, 1992) for comparable sample sizes and methods including ML, GLS, SB and ADF validating the simulation. In general, ML and GLS perform similarly to previous results in Study 1 and will not be reviewed again here except in relation to new methods.

Table 22.

Simulation Results for Test Statistics for Ten Methods by Sample Size (Normal Case A)

Samp Size	ML	GLS	ADF	rADF	SB	rSB	SBMV	rSBMV	TRES	rRES
150	92.08	84.43	229.50[a]	90.42	92.70	69.36	49.57	48.62	241.31	89.78
250	89.24	84.78	143.38	93.68	89.52	74.11	58.68	55.47	143.55	93.22
500	88.87	86.64	110.78	94.06	89.07	80.49	70.50	66.56	110.51	93.78
1000	87.80	86.76	97.73	91.42	87.92	83.48	77.70	74.72	97.57	91.27
2000	87.73	87.17	92.56	89.82	87.76	85.49	82.33	80.47	92.47	89.74
2500	88.03	87.61	91.74	89.62	88.05	86.22	83.64	82.08	91.68	89.55
3000	88.24	87.77	91.20	89.47	88.29	86.75	84.57	83.23	91.14	89.42
5000	87.11	86.88	88.94	87.95	87.13	86.23	84.90	84.05	88.90	87.91
10000	86.67	86.59	87.62	87.14	86.68	86.22	85.55	85.11	87.60	87.12

Note . Model is oblique 3 factor model in Hu et al. (1992), Replications = 1000, [a] Replications = 998, *df* = 87, *df* s for SBMV and rSBMV differ across replications.

ADF results are consistent with the literature reflecting large and positive bias at smaller sample sizes, ranging up to 164%. Positive bias of test statistics is large across small and moderate sample sizes until at least sample sizes of 2000. Although test statistics for rADF are positively biased at smaller *N*s, rADF is improved over traditional ADF. For example, at *N* = 150, bias is only 4% and ranges up to 8.1% across all sample sizes. The other distribution free test statistic for RES performs comparably to ADF across the range of sample sizes and is biased upward, ranging up to 177% at the smallest of sample size. In parallel to ADF results, the regularized RES test statistic performs like rADF, with only a small amount of positive bias (ranging up to 7.79%). As expected at small sample sizes, test statistics for SB were inflated and results for rSB reveal the test statistic is negatively biased. Compared to normal theory methods, rADF is performing similarly and even better than the normal theory and SB methods under a few conditions, especially at *N* = 150.

Variability of Test Statistics

In general, ADF and RES test statistics exhibit more variability than all other methods (see Table 23). Test statistics of rADF and rRES are less variable than their traditional counterparts. This is especially true at small *N*s for which variability for rADF and rRES are about a third of the size of ADF and RES. By *N*s of 2000 variability of all methods are fairly comparable to each other and to expected standard deviations (13.19).

Table 23.

Standard Deviations of Test Statistics for Ten Methods by Sample Size (Normal Case A)

Samp Size	ML	GLS	ADF	rADF	SB	rSB	SBMV	rSBMV	RES	rRES
150	14.23	12.75	50.71[a]	16.47	14.23	11.51	7.22	7.43	60.02	16.51
250	12.83	12.26	27.16	15.86	12.83	10.98	8.08	7.83	27.60	15.84
500	12.98	12.81	18.58	15.33	13.05	11.89	10.04	9.58	18.58	15.31
1000	13.71	13.47	16.00	14.89	13.72	13.06	11.91	11.50	15.98	14.87
2000	12.89	12.86	14.13	13.67	12.90	12.56	12.00	11.73	14.12	13.66
2500	13.49	13.41	14.18	13.90	13.51	13.28	12.86	12.66	14.17	13.90
3000	13.88	13.86	14.76	14.38	13.87	13.57	13.07	12.83	14.75	14.37
5000	12.75	12.74	13.09	12.93	12.75	12.62	12.37	12.25	13.09	12.93
10000	13.01	12.98	13.22	13.15	13.01	12.94	12.82	12.75	13.21	13.14
100000	12.81	12.82	12.83	12.82	12.82	12.80	12.79	12.79	12.83	12.82

Note . Model is oblique 3 factor model in Hu et al. (1992), SD = Standard Deviation, Replications = 1000,

[a] Replications = 998.

5.3.4 Rejection Rates

Empirical rejection rates for all ten methods are displayed in Table 24. Even at *N* = 10,000 there is some variability in performance across methods. Both rADF and rRES methods produce exactly 50 rejections. ADF and RES both have the highest number of rejections (both 58), while regularized SB methods have the fewest (both 38). At smaller *N*s, ML rejection rates were higher than nominal rates, and GLS rejected models too infrequently, while ADF and RES reflect unacceptable rejection rates. Almost all true models are rejected at the smallest of sample size conditions for these methods. However, rejection rates for rADF reflects a huge improvement over ADF, as does rRES over RES. Importantly, these methods reflect

improvement over ML and SB at the smallest sample sizes, (but not SBMV which performs next to perfect).

On the other hand, the regularized versions of SB methods did not perform well. These methods over accept at small *N*s not reaching acceptable reject rates until at least 2,500. Rejection rates of SB and SBMV were the best performing overall methods: SBMV outperformed SB at the smallest sample size condition.

Table 24.

Rejection Rates for Ten Methods by Sample Size (Normal Case A)

Samp Size	ML	GLS	ADF	rADF	SB	rSB	SBMV	rSBMV	RES	rRES
150	112	37	998[a]	109	126	0	46	0	1000	109
250	68	29	911	147	68	2	33	0	911	139
500	60	48	504	152	61	12	45	8	491	147
1000	61	52	207	119	63	29	48	27	204	119
2000	50	44	125	79	49	28	48	26	125	79
2500	63	58	125	87	64	47	58	38	124	86
3000	64	56	101	78	63	49	61	48	100	77
5000	47	47	68	58	48	38	47	37	68	58
10000	40	43	58	50	41	38	39	38	58	50

Note . Model is oblique 3 factor model in Hu et al. (1992). Replications = 1000, [a] Replications = 998.

5.4 Case B: Nonnormal Condition Results

A small pilot study was conducted to decide on the specification for the mixture parameter for the MAP estimator for nonnormal conditions (Case B and C).

To empirically examine potential for performance, the two mixture parameters were auditioned. A condition was selected that was small enough to be low cost in terms of computing time and to reflect differences in outcomes of interest, while at the same time would not hindered by problems with solution propriety ($N = 250$). Both mixture parameters were tested and compared in terms of condition numbers of the MAP estimators. Additionally, the MAP estimator was inverted and employed as the weight matrix for ADF, (the regularized ADF

method described above, rADF & RADF) and results across 1,000 replications were compared (see Table 25) with the goal of selecting the mixture parameter with the best potential.

Table 25.
Comparison of Alpha Mixture Parameter for Study 3

Alpha	Condition Number	SD	ADF	Emp Rej
-	4443.25	945.25	140.4	907
$\hat{\alpha}_R$	569.52	136.83	137.01	896
$\hat{\alpha}_r$	174.21	43.41	69.23	1

Note. Replications = 1000, N = 250, Case B.

Results suggest that the condition number of the unmodified matrix is quite large and quite a bit larger than 1, indicating ill-conditioning (see Table 25). Both MAP estimators' condition numbers are reduced, though more so for $\hat{\alpha}_r$ than $\hat{\alpha}_R$. Moreover, in terms of variability all methods are relatively similar (C_v =.211-.249). In terms of test statistics, both methods improve upon the traditional ADF method. When computing the MAP estimator using $\hat{\alpha}_r$ the ADF rejection rate over-accepts compared to the expected 50, while the MAP estimator computed using $\hat{\alpha}_R$ resulted in only a small improvement, with a rejection rate well above the nominal value. These results are somewhat similar to results in Study 1 suggesting $\hat{\alpha}_r$ resulted in lower conditions numbers and severely smaller test statistic compared to $\hat{\alpha}_R$. This could be due to several factors, since both the method and the size of the matrix differ across these studies. Additionally the normality changes between Case A and B, and we also see differences in results. The pattern of results is similar but here we see a more severe penalty creates an over acceptance. Nevertheless, based on these results, the $\hat{\mathbf{\Gamma}}_R$ matrix was calculated using $\hat{\alpha}_R$. Although results reflect neither method is ideal, choosing the more conservative result might help uncover patterns

to further understand relationships between condition of matrices and solutions rather than choosing $\hat{\alpha}_r$ resulting in continued over-acceptance.

5.4.1 Condition of Matrices

Results are compiled in Table 26. Condition numbers are displayed on the left half, while standard deviations, and coefficient of variations are displayed in the right half of the tabled results. The empirical condition number based on 100,000 replications for $\hat{\boldsymbol{\Gamma}}$ was 516.43, larger than the condition number under normality. Condition numbers of $\hat{\boldsymbol{\Gamma}}$ are quite huge and variable at small *N*s, and larger than condition numbers for normal condition (ranging up to about 265,000). The coefficient of variation (CV) reveals variability of condition numbers of $\hat{\boldsymbol{\Gamma}}$ exceeds the average condition number for $\hat{\boldsymbol{\Gamma}}$ at the smallest sample size. Moreover, the variability is about equal to the average for $N = 200$. Condition number and variability decreases as sample sizes increase for the unmodified matrices. CVs reflect the decrease of variability is more accelerated than the average condition number.

Similar to the normal condition, the $\hat{\boldsymbol{\Gamma}}_R$ is well improved over $\hat{\boldsymbol{\Gamma}}$ especially the smallest of sample sizes, though the difference decreases as *N* increases. Interestingly the condition numbers of $\hat{\boldsymbol{\Gamma}}_R$ do not increase monotonically across sample size conditions. Rather, both the average condition number and variability of condition numbers increase until *N*s of about 2000 and 2500 then decrease, though CVs reflect the variability does decrease relative to the signal across *N*.

Table 26.

Average Condition Numbers, SDs, and Coefficient of Variations of $\hat{\Gamma}$ *and* $\hat{\Gamma}_R$ *by Sample Size (Nonnormal Case B)*

Samp Size	$\hat{\Gamma}$ Cond	$\hat{\Gamma}_R$ Cond	$\hat{\Gamma}$ SD Cond	$\hat{\Gamma}_R$ SD Cond	$\hat{\Gamma}$ c_v	$\hat{\Gamma}_R$ c_v
150	37892.69	1343.28	21554.27	412.52	0.57	0.31
250	4444.25	569.52	1935.42	136.83	0.44	0.24
500	1601.97	235.42	642.65	56.47	0.40	0.24
1000	953.67	624.80	296.47	309.41	0.31	0.50
2000	697.74	632.13	139.03	159.57	0.20	0.25
2500	649.49	589.05	113.58	113.33	0.17	0.19
3000	624.30	572.26	99.68	99.38	0.16	0.17
5000	562.75	531.30	73.27	72.90	0.13	0.14
10000	514.47	499.63	50.08	49.79	0.10	0.10

Note. Samp = Sample, Cond = Condition Number, SD = Standard Deviation, Empirical condition number is approximately 476.03 (N = 100,000).

5.4.2 Solution Propriety

Under nonnormality, no convergence issues emerged for any of the methods except ADF. Five replications had analyses that did not converge for the traditional ADF method. At the smallest sample size the ADF method also had 21 high number of iterations and results for 47 replications revealed improper solutions. At N = 250, only one replication for ADF resulted in at least one negative variance.

5.4.3 Performance of Test Statistics

The normal theory and ADF methods performed as expected based on past research and Studies 1 and 2 (see Table 27). As in previous conditions, test statistics for the RES condition continued to perform very similarly to ADF. Regarding regularization methods, both RADF and RRES improve upon traditional methods at the smallest *N*s however both methods reflect huge positive bias until *N*s of about 3,000. SB is positively biased at the smallest of *N*s, while the regularized version actually worsens the positive bias.

Table 27.

Simulation Results for Test Statistics for Ten Methods by Sample Size (Nonnormal Case B)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	91.24	83.50	219.91	200.93	91.90	108.73	43.31	63.86	231.45	200.95
250	88.97	84.72	140.40	137.00	89.38	98.81	52.70	64.56	140.56	136.30
500	88.81	86.76	110.18	110.20	88.81	94.45	65.24	74.67	109.95	109.84
1000	87.69	86.69	97.46	97.93	87.84	88.44	74.01	74.89	97.32	97.78
2000	87.17	86.48	91.69	91.76	87.23	87.30	79.67	79.77	91.61	91.40
2500	87.84	87.43	91.59	91.65	87.90	87.96	81.71	81.78	91.53	91.58
3000	86.77	86.37	89.59	89.64	86.81	86.85	81.63	81.69	89.53	89.58
5000	86.52	86.29	88.36	88.38	86.54	86.56	83.35	83.38	88.32	88.35
10000	87.12	87.02	88.06	88.06	87.14	87.14	85.51	85.50	88.05	88.06

Note . Model is oblique 3 factor model in Hu et al. (1992), Replications = 1000, ADF had 995 replications at $N = 150$, $df = 87$, *dfs* for SBMV and RSBMV differ across replications.

Variability of Test Statistic

When examining the variability in Table 28, ADF and RES methods are similar in regards to the spread of test statistics and at the smallest of sample size they are the most variable of all methods. On the other hand, the regularized versions of these methods reflect variability is relatively smaller especially at smaller sample sizes. Additionally, RADF and RRES are less variable than SB and across all sample sizes.

Table 28.

Standard Deviations of Test Statistics for Ten Methods by Sample Size (Nonnormal Case B)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	14.05	12.79	46.82	39.19	14.00	16.84	7.23	8.94	54.98	40.01
250	13.53	12.98	25.24	23.48	13.50	15.03	8.24	9.50	25.70	23.49
500	12.79	12.52	17.58	17.01	12.79	13.65	9.65	10.67	17.56	16.96
1000	12.52	12.58	14.49	14.56	12.44	12.53	10.41	10.53	14.47	14.53
2000	13.40	13.30	14.23	14.24	13.37	13.38	12.09	12.11	14.22	14.23
2500	12.72	12.71	13.61	13.61	12.75	12.76	11.80	11.81	13.60	13.61
3000	13.15	13.11	13.72	13.73	13.15	13.16	12.32	12.33	13.71	13.72
5000	13.06	13.02	13.41	13.42	13.03	13.04	12.51	12.52	13.41	13.41
10000	14.01	13.97	14.17	14.17	14.01	14.01	13.71	13.71	14.16	14.16

Note . Model is oblique 3 factor model in Hu et al. (1992), Replications = 1000, ADF had 995 replications at $N = 150$.

5.4.4 Rejection Rates

According to Table 29, ADF and RES methods perform similarly in that reject almost every true model under smaller sample sizes. As sample size increases they perform closer to normal theory methods in terms of rejections, however not at acceptable levels until *N*s of at least 2500. The regularized counterparts are not improved. Neither RADF nor RRES reach acceptable rejection rates until *N*s of at least 2,500 like their traditional counterparts. Moreover, results for the regularized methods RSB reflect performance is worse than SB. On the other hand, RSBMV over "improves" upon SBMV. While SBMV over accepts models at the smallest of sample sizes, RSBMV over rejects, and does not reach nominal rejection rates until *N*s of 1,000.

Table 29.

Rejection Rates for Ten Methods by Sample Size (Nonnormal Case B)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	103	27	995	1000	100	455	30	261	1000	1000
250	72	37	907	896	71	230	28	122	906	891
500	57	38	481	482	55	126	33	95	471	473
1000	48	45	203	212	47	51	35	43	200	209
2000	64	56	117	117	62	65	53	54	114	115
2500	47	45	98	100	50	51	47	47	98	98
3000	53	52	73	73	54	54	50	50	73	73
5000	40	42	57	57	42	42	38	39	57	57
10000	53	52	65	65	53	53	52	52	65	65

Note . Model is oblique 3 factor model in Hu et al. (1992), Replications = 1000, ADF had 995 replications at N = 150,

5.5 Case C: Nonnormal Results

5.5.1 Condition of Matrices

Condition numbers in Table 30 reveal that results are essentially the same as to those in Case B, as expected. Refer to Section 5.4.1.

Table 30.

Average Condition Numbers, SDs, and Coefficient of Variations of $\hat{\Gamma}$ *and* $\hat{\Gamma}_R$ *by Sample Size (Case C)*

Samp Size	$\hat{\Gamma}$ Cond	$\hat{\Gamma}_R$ Cond	$\hat{\Gamma}$ SD Cond	$\hat{\Gamma}_R$ SD Cond	$\hat{\Gamma}$ c_v	$\hat{\Gamma}_R$ c_v
150	37892.69	1343.28	21554.21	412.52	0.57	0.31
250	4443.25	569.52	1935.42	136.83	0.44	0.24
500	1601.97	235.42	642.61	56.47	0.40	0.24
1000	953.68	624.79	296.37	309.41	0.31	0.50
2000	697.73	625.72	139.02	139.11	0.20	0.22
2500	649.49	589.05	113.58	113.33	0.17	0.19
3000	624.31	572.62	99.68	99.38	0.16	0.17
5000	560.73	529.02	73.01	72.29	0.13	0.14
10000	515.59	500.68	47.95	47.63	0.09	0.10

Note. Samp = Sample, Cond = Condition, SD = Standard Deviation, c_v = coefficient of variation.

5.5.2 Solution Propriety

All replications converged for all methods. Additionally, no replications exceeded 100 iterations but at $N = 150$ results included improper solutions (53 for ADF, 3 for RADF). Furthermore, results for two other methods reflected problems at $N = 150$: GLS had one replication that reflected non-convergence, one with high iterations, and a replication with an improper solution. Although at $N = 250$ all replications reflected convergence for ADF, 19 solutions had at least one negative variance.

5.5.3 Performance of Test Statistics

Results in Table 31 reflect typical findings under severe nonnormality for ML, GLS, ADF and SB methods (Model $df = 93$). As reflected in other conditions, RES performed closely to ADF. Regularization methods are all positively biased at most sample sizes except at the largest sample sizes. This positive bias worsened under these severe nonnormality conditions.

Table 31.
Simulation Results for Test Statistics for Ten Methods by Sample Size (Nonnormal Case C)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	97.79	91.08	282.60	242.90	97.92	119.58	44.55	62.99	288.64	241.84
250	95.16	91.77	161.04	158.11	94.97	108.37	54.35	69.53	160.77	157.23
500	94.87	93.22	120.91	124.69	94.29	103.64	67.38	80.58	120.63	124.27
1000	93.80	93.03	105.47	106.49	93.14	94.65	77.05	94.65	105.32	106.32
2000	93.13	92.53	98.23	98.40	92.39	92.66	83.33	83.77	89.89	90.01
2500	93.92	93.56	98.05	98.21	93.16	93.40	85.59	85.97	97.99	98.14
3000	92.83	92.49	95.87	96.01	92.07	92.28	85.64	85.96	98.15	98.32
5000	92.98	92.83	94.79	94.88	92.19	92.32	87.98	88.19	94.76	94.85
10000	92.87	92.76	93.40	93.44	92.08	92.15	89.62	89.73	93.39	93.43

Note . Model is oblique 3 factor model in Hu et al. (1992), Replications = 1000, $df = 93$, df for SBMV and RSBMV differ across replications,.

Variability of Test Statistics

When examining the variability in Table 32, findings reflect that ADF and RES methods

are similar in regards to the spread of test statistics and at the smallest of sample size. Both of these methods are the most variable of all methods. On the other hand, the regularized versions of these methods reflect variability that is relatively smaller, especially at smaller sample sizes. Additionally, RADF and RRES are less variable than even RSB and traditional SB across all sample sizes.

Table 32.

SDs of Test Statistics for Ten Methods by Sample Size (Nonnormal Case C)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	14.84	14.21	70.38	52.76	14.86	14.86	7.54	9.44	75.76	53.41
250	14.02	14.08	29.86	27.70	14.06	16.05	8.61	9.95	30.13	27.66
500	13.20	13.16	19.09	19.28	13.12	14.42	9.98	11.12	19.09	19.25
1000	13.17	13.18	15.40	15.57	12.83	13.84	10.77	11.06	15.38	15.55
2000	13.92	13.89	15.07	15.09	13.80	13.84	12.39	12.44	15.06	15.08
2500	13.07	13.06	14.09	14.11	12.97	13.01	11.97	11.93	14.08	14.10
3000	13.58	13.57	14.07	14.09	13.50	13.53	12.57	12.61	14.06	14.08
5000	12.92	12.88	13.48	13.49	12.80	12.82	12.20	12.23	13.46	13.49
10000	12.93	12.95	13.16	13.17	12.84	12.85	12.48	12.50	13.16	13.17

Note . Model is oblique 3 factor model in Hu et al. (1992).

5.5.4 Rejection Rates

As Table 33 reflects, the RES method once again parallels the ADF results. Empirical rejection rates for regularization methods are similar to results under nonnormality Case B. That is, at the smallest sample sizes these methods do not reject any true models. Although rejection rates improved as *N* increased, they do not reach nominal rejections until the largest of sample sizes.

On the other hand, regularization applied to correction methods (SB/SBMV) actually increase the rate of rejections at the smallest sample size. The goal is to decrease rejections at small *N* for SB, not increase them. On the other hand, although inflating SBMV is desired, the inflation is too severe. By *N*s of 1000 both regularized methods return reasonable rejection rates.

Table 33.

Rejection Rates for Ten Methods by Sample Sizes (Nonnormal Case C)

Samp Size	ML	GLS	ADF	RADF	SB	RSB	SBMV	RSBMV	RES	RRES
150	110	47	1000	1000	113	543	28	339	1000	1000
250	67	49	957	955	67	299	23	183	957	954
500	52	39	557	624	49	191	22	141	555	636
1000	49	50	236	257	42	59	40	26	232	254
2000	64	53	131	136	53	56	44	46	128	133
2500	58	55	95	97	54	56	38	39	95	96
3000	51	51	78	80	48	48	40	42	78	79
5000	46	44	58	58	40	42	36	38	58	58
10000	40	40	48	48	34	34	34	34	48	48

Note . Model is oblique 3 factor model in Hu et al. (1992).

5.6 Simulation 3: Overall Results

The average condition number at a sample size of 100,000 (254.94) worsened at under nonnormality (516.43), though this number was still under 1,000 (one heuristic for ill-conditioning) for all three conditions. For a summary of condition numbers of unmodified matrices across conditions see Table 34.

Table 34.

Condition Numbers, SDs, and Coefficients of Variation of $\hat{\Gamma}$ *by Sample Size and Normality*

Sample Size	Normal			Nonnormal		
	$\hat{\Gamma}$ Cond	$\hat{\Gamma}$ SD Cond	c_v	$\hat{\Gamma}$ Cond	$\hat{\Gamma}$ SD Cond	c_v
150	18459.87	7160.42	0.39	37892.69	21554.21	0.57
250	2087.80	532.43	0.26	4443.25	1935.42	0.44
500	742.46	132.13	0.18	1601.97	642.61	0.40
1000	455.46	59.74	0.13	953.68	296.37	0.31
2000	352.19	33.89	0.10	697.73	139.02	0.20
2500	332.16	28.78	0.09	649.49	113.58	0.17
3000	318.75	24.96	0.08	624.31	99.68	0.16
5000	294.96	18.27	0.06	560.73	73.01	0.13
10000	275.68	12.75	0.05	515.59	47.95	0.09

Note. Cond = Condition Number, SD = Standard Deviation, c_v = Coefficient of Variation.

In general, the average condition number and SDs for unmodified matrices increased as *N* decreased. Additionally, average condition numbers were quite large at small *N*s for the normal condition, and even greater still under nonnormality. In general they were about double the size of normal condition numbers, and discrepancy was largest at smaller sample sizes. Results also reveal high variability of condition numbers at small sample sizes especially under nonnormal conditions.

Regarding solution propriety in Table 35, poor solutions worsened under nonnormality at small sample sizes for ADF. These issues greatly improved under regularization, and were essentially zero accept a few negative variances (4 for Case B & 3 for Case 3 at $N = 150$) especially for both normal and nonnormal conditions.

Table 35.
Improper Solutions Results for each Normality Condition at Ns of 150 and 250

N	Nonnormality	ADF		
		Non-Conv	High Iter	Neg Var
150	Normal	2	3	22
	Nonnormal B	5	21	47
	Nonnormal C	0	0	53

Note . Model is oblique 3 factor model in Hu et al. (1992), 1000 replications attempted; ADF = 1 Neg Var for N =250, Case C condition, GLS = 1 Non-Conv, High Iter, & Neg Var at N =150.

Test statistics under regularization in normal conditions reflect rADF and rRES are greatly improved, with rRES having a slight advantage over rADF. Both methods exhibit small positive bias at small *N*s, with somewhat inflated variability. These inflations are reflected in rejection rates that are too high at the smallest *N*s. On the other hand, under nonnormality the pilot study revealed at $N = 250$ the test statistic was severely negatively biased, reflecting too few

rejections (1 of 1000 replications). Yet, RRADF and RRES did not improve on ADF and RES under nonnormal conditions, performing on par with the traditional methods.

Findings for regularization applied to SB reflect similar results with SBMV more severely impacted- while rSB recovers at larger sample sizes rSBMV does not, consistently over accepting the true model. As expected, under normality the typical scaled SB method performs similarly to ML at small *N* conditions but improves as *N* increases, while the typical SBMV performs quite well even at small *N* conditions and in general outperforms other methods. Under nonnormality SBMV loses the advantage and over-accepts the true model across a wide range of sample sizes, while SB performs well under moderate and severe nonnormality at most sample sizes except the smallest. Under nonnormality, SB and SBMV worsened under regularization and did not recover until *N*s of 1000.

CHAPTER 6

DISCUSSION

6.1 Summary of Findings and Conclusions

Major objectives of these studies included the application of regularization to SEM in order to improve results of existing estimators that are known to break down under finite conditions, and to conduct systematic simulation studies to elucidate potential mechanisms for the poor performance. In particular, ill-conditioned matrices were examined as culprits, and a MAP approach was proposed and applied. The idea to employ a shrinkage estimator that modifies the eigenvalues of offending matrices, improving the "data", but not modifying the method itself, allowed for great flexibility in application. The form of the target matrix to regularize differed across the simulation studies, however the overall goal remained the same—to stably estimate a covariance matrix in order to improve SEM results that depend upon these matrices (or their inverse).

Major findings include the following- poorly conditioned matrices play a role in problems in SEM that arise under non-ideal circumstances. Condition numbers reflect ill-conditioning, especially when sample sizes are small, as well as under conditions of nonnormality. Since there are no cutoffs or guidelines for what constitutes an "worse" ill-conditioned matrix, the magnitude can be assessed in a qualitative manner, such that it is conservative to say condition numbers are large under small sample size, larger for nonnormal than normal data, and quite huge for higher dimensional matrices. Additionally, these factors (tend to) interact creating conditions where the magnitude of condition numbers were even higher than these reported in recent literature. While most findings were very similar to Huang & Bentler (2015), condition numbers decreased at a much slower decent across sample size under the nonnormal conditions in Study 2, and still were quite large at sizes typical for SEM use (*N*s

of 200 and 300). Basically, these condition numbers were more pronounced. One difference between the two studies is that the true factor loadings were more variable, (and therefore so were the associated error variances), which may be another basis for ill-conditioning of matrices.

The proposed regularization methods always improved condition numbers, while RGLS also improved small sample performance of test statistics. In fact, the RGLS method was found to be quite a well performing method for such finite samples. RGLS outperformed traditional estimators at these small *N* conditions. As expected by previous literature reviewed in the introduction (e.g. Bentler & Yuan, 1999) conventional test statistics of ML(GLS) break down under less than asymptotic conditions, over (under) estimating the test statistics associated with too many (few) rejections. Therefore, the superior performance of RGLS is valuable given the more common practice of intensive data collection. Researchers now have access to data sets that do not fit into the classical statistical framework used in SEM, that of asymptotic theory requiring large samples on a relatively small number of variables.

Additionally, under normal conditions ADF methods were improved when regularization was applied to both the sample covariance matrix and the asymptotic covariance matrix (though much less for the former than the latter). Other methods outperformed RADF and RRES however at the smallest of sample size there was a notably improved difference. More importantly, beyond performance results, these findings provide empirical evidence to support the idea that estimating a large amount of information (e.g. elements in sample covariance matrices) based on little information (such as small *N*), leads to instability and degenerate cases (when the number of elements is quite large in the matrix) and are at least partially to blame for problems. Additional evidence of this included results concerning solution propriety. When matrices were regularized and applied to SEM convergence was improved, number of iterations

for reaching convergence was reduced, and fewer cases of improper solutions emerged. The findings add to research examining similar phenomena (e.g. Yuan & Bentler, 2017).

Although findings support that under some conditions regularization techniques greatly improve upon traditional ones, unexpectedly, the choice of tuning and mixture parameters added complexity to the methodology. When aiming to fix problems, other issues can arise and sometimes a better solution will come with trade-offs. For example, while there are many approaches for cross validation for the penalty parameter, it is not intuitive as to what data would best be supplied for the folds for the asymptotic matrix. This process was made possible here by using the n x p^* "data" matrix, which is a type of sum of cross products matrix, yet this may not be the only, or best approach. Other methods reviewed in the introduction like that of Kamada (2011) may be useful for these types of matrices because using information criteria has lower cost. This possible solution should be studied comparatively.

Additionally, finding the penalty and relative weight independent of one another may not be ideal. Fixing one parameter while finding the other is an ad-hoc method. When multiple parameters need to be estimated this common practice eases complexity. Yet, this can come at a cost. For the RGLS case, the suggested estimator of the mixture parameter (α) was not ideal which motivated another proposition put forth in Study 1 that worked well. On the other hand, this suggestion may not generalize. Given results from Simulation 3, a parameter that depends on features of the data may be more fruitful in terms of results, especially given the application to a much larger, complex matrix. Chi and Lange (2014) suggest a choice of α to match the scale of the data. If large sample eigenvalues are inflated, the mean may also be too large, leading to an α that is too small, which was the case for Study 1. It is important to investigate this idea further.

Of course, one could run a simulation study to identify the ideal value to use, but this is not very efficient, or practical for users.

A next step towards better penalty parameters for the MAP estimator will be devoted to finding a rationale for selection or even simpler, finding good tools to select the optimal parameters, empirically. One possibility might include extending cross validation for finding the mixture-parameter α though this adds potential complexity. A way to ameliorate this might be to adjust the number of folds for cross validation.

Furthermore the Elastic Net (Zou & Hastie, 2005) regularization method for regression might be a good source for a possible solution. Elastic Net is a hybrid between ridge regression and LASSO, and uses a convex linear combination of the $l1$ and $l2$ norms. A typical scheme for finding the necessary parameters with Elastic Net is to employ a cross validation scheme using a grid of candidate α values within [0,1] representing the proportion of $l1$ versus $l2$ penalty, a parallel to our problem. A naïve approach that is sometimes taken is to fix the alpha to some value and use cross validation for lambda. However, this ad hoc approach is simple but not always ideal for Elastic Net. Zou and Hastie (2005) recommend to cross validate on a two dimensional surface which may point to a solution for our similar issue. There is a literature around such methods (e.g. Hastie, Tibshirani, & Friedman, 2009; Lorbert & Ramadge, 2010). While not extrinsically obvious, parameters assign relative weights to different characteristics of the data. Therefore it is reasonable that performance not only relies on choice of regularization to some degree but also choice of these types of parameters.

The present study adds to existing regularization literature reviewed in Chapter 2 particularly for SEM. Just in the past few years there have been several publications in SEM journals surrounding regularization techniques. This work adds to that contemporary research but

also extends the literature given this is a novel approach for other matrices beside the sample covariance matrix, and applications of estimators beyond ML (though also see Yuan & Chan, 2016). Moreover, these findings as a whole contributes to the growing body of literature about the condition of matrices. This has implications for both methodologists and practitioners.

6.2 Significance and Implications

6.2.1 Implications for Methodologists: Future Research

The identification of some ill-conditioned matrices used in SEM, and their impact on results, points toward the consideration of why matrices are poorly conditioned and how this ill conditioning is impactful to different methodologies. For researchers who use Monte Carlo simulation as a tool, these results might encourage a practice of reporting condition numbers of important matrices alongside typical reporting of methods and/or results, even if regularization is not the focus of the study. This will not only add to the body of understanding of condition numbers over a wider range of conditions than those considered here, but also identifies potential problems before a simulation gets underway. Additionally, findings here open up new avenues of research.

Simulation conditions considered in these studies were limited. This is typical for a first application of any technique like MAP to SEM. Constricting conditions allowed for greater control (i.e. internal validity), but this of course has the opposite effect on generalizability. It is a strength of these studies that a model that has been tested in several papers was utilized allowing for verification of the simulation. Extensions of conditions such as inclusions of different nonnormality conditions and possibly others not considered here (e.g. misspecification or missing data) could be reasonable next steps. Additionally, many extensions of the approach can

be considered, including drawing from the rich bank of regularization techniques outlined in Chapter 2.

Finally, findings could also inspire developers of programs for SEM analyses to either inform the user about the condition of the matrices used in the analyses. This could be in the form of descriptive information commonly reported at the beginning of output, or even some sort of warning message.

6.2.2 Implications for Practitioners: Applications

Study results also indicate a clear need to educate SEM users about the benefits of regularization as applied to SEM. Given the fact as reviewed in introductory chapters, the actual usage of regularization in the social and behavioral sciences is fairly limited relative to the huge amount of literature of applications that exists in other areas. Perhaps little is known about the usefulness of such techniques to address problems. Key findings here suggest poor matrices could be culprits of problems of convergence, high number of iterations and improper solutions, encountered by SEM users. For one such example, consider Mark and Belyea (2009) who do not report SEM results, stating a lack of convergence of the model moving from a univariate to bivariate model, doubling the number of manifest variables. Although other problems could be the culprit besides high dimensionality, it points to the disadvantage of not being able to employ a model that might be more in line with theory and settling for another that might be less-ideal, or even biased. A suggestion based on findings from these studies is to evaluate matrices as part of ones data screening and model building process. Although poor solution propriety can also be indicators of a variety of other issues such as a poorly hypothesized models, or an identification problem, evaluating relevant matrices can point the analyst toward the problem (or at least rule one out).

Of course, any limitation can hinder wide-spread adoption of practice or use of a novel procedure, even a beneficial one. However, calculations of condition numbers are fairly simple, and condition number estimators can be found in freeware. For example in R, an open source software, one such package exists called *kappa*. This package uses the $l2$ norm in calculations, equivalent to the intuitive definition of the ratio of largest to smallest eigenvalue. For data that are "ultra high-dimensional", such as fMRI data or large scale data often associated with educational research for a small subset of participants, these calculations may not be cheap in terms of time or computing resources, though advances of computer memory may mitigate these concerns. Further, cheaper approximations are available, even in the *kappa* package, if necessary.

Reporting condition numbers will have a two-fold impact (1) it is a part of best practice for understanding your data and points researchers toward the selection of an appropriate method to analyze data and (2) it grows the body of knowledge for what might be considered "typical" in terms of conditions of matrices, perhaps even quicker than large number of simulations with far-reaching conditions. Moreover, this practice could provide more authentic information about characteristics of real-world data. Furthermore, reporting this information might identify issues with particular matrices more common to specific lines of research and in a didactic pattern, inspire new methods, informing methodologists about areas fruitful for research.

6.3 Concluding Remarks

First, a brief overview of well-known problems in SEM was given. These problems have been longstanding, and while they are frequently acknowledged, little has been done to fix issues such as ill-conditioned matrices that might have a role in creating such problems. Additionally, a review of regularization methods was provided, including contemporary work in the SEM, as

well as of other techniques lesser known to social and behavioral scientists. Details were delineated for one such method, the MAP estimator, as well as plans for its applications to SEM.

The idea of fixing the covariance matrix is not necessarily a brand new idea. One such case is reviewed in a paper by Tanaka (1987) thirty years ago. Borrowing from economics, the Minimum Entropy Estimator (ME2; Vinrod, 1982) matrix was proposed in order to estimate a sample covariance matrix whose diagonal was adjusted for measurement error for small sample problems. This method proved not to be a consistent estimate of the population matrix, and therefore, was not useful. However the MAP estimator considered in this present investigation is importantly a consistent estimator and ideal for such a purpose.

In sum, the proposed MAP method was subjected to Monte Carlo simulation study where true parameters and conditions could be controlled so that performance could be evaluated in three large-scale studies. Importantly, studies included both normal and nonnormal conditions and a wide ranging number of sample sizes to create contexts for which matrices would be ill-conditioned, and moreover, model fit breaks down. In addition, the MAP approach was compared to other methods originally designed to ameliorate but not directly address these long-standing problems, including such methods as asymptotic distribution free methods, correction methods, and another contemporary regularization method using a ridge approach.

In the following chapter, the well-performing RGLS estimator is employed with real-world data. Broadly speaking, one goal of regularization is to change a system so it obeys laws – regularization may be one key to fixing problems in structural equation modeling under non-ideal finite circumstances, such that they perform like they are obeying asymptotic properties.

Chapter 7

AN APPLICATION OF REGULARIZED GLS

7.1 Background and Related Literature

This illustrative example is presented in order to study the application of regularization methods using real-world data. This example is particularly instructive given the small sample size relative to the number of items on a measure of parental motivational practices for encouragement of children's academic motivation.

In addition to school based factors, home based factors including attitudes, beliefs, and behaviors of parents play an important role in children's educational success (Fan & Chen, 2001). For example, mothers' encouragement of task endogeny has been shown to be positively related to children's intrinsic academic motivation (A. E. Gottfried, Fleming, & Gottfried, 1994). Furthermore, research suggests rewarding children contingent upon performance can discourage autonomy, and has been shown to be negatively related to future academic achievement.

Though research abounds in respect to mothers' educational involvement, less is known about paternal involvement. This illustration focuses specifically on fathers' task intrinsic and task extrinsic motivational practices. Task intrinsic practices comprise parental encouragement of children's pleasure in the learning process, curiosity, persistence, and task involvement. On the other hand, parental task extrinsic practices comprise external control using external rewards and consequences contingent on children's performance. In general, research of motivational practices surround mothers' practices and relatively less is known about fathers' practices. This is not unusual in educational and developmental research. In fact, as a whole in comparison to mothers, the importance on the role of the father and his influence on the development and growth of his child is often neglected or assumed equivalent to mothers, (though there are exceptions, see Flouri, 2005 and other research by Flouri & Buchanan, 2003, 2004 for example).

Although the subject of studies in educational or developmental research is parents, often mothers are the typical respondent, not fathers. Relatively fewer parent-child studies specifically investigate fathers' roles. Additionally, even when responses surround paternal involvement, responses are often reported by children's mother in regards to fathers rather than collecting data directly from fathers. Therefore, these analyses begin with examining a measurement model, an important first step toward examining parent motivational practices with father-reported data.

7.2 Methods

7.2.1 Participants

The present study employed data from the Fullerton Longitudinal Study (FLS; e.g., A.W. Gottfried, Gottfried, & Guerin, 2006), a contemporary ongoing long-term longitudinal investigation in which 130 children and their families were followed from infancy (age 1-year) through adulthood. The infants were selected from notifications of all births from hospitals surrounding the university. Families were invited to participate just prior to the infants' 1-year birthday. Infants free of neurological and visual problems and of normal birth weight were eligible to enter the study. Socioeconomic status (SES) of families was determined by the Hollingshead Four-Factor Index of Social Status (see Hollingshead, 1975). This index is based on both mothers' and fathers' level of education and occupational ranking. SES varied ranging from semi-skilled workers with no high school degree through professionals.

Over the course of study, participants' retention was high with at least 80% returning for any assessment and with no evidence of attrition bias (A. W. Gottfried, Gottfried, Bathurst, Guerin, & Paramore, 2003). For further details concerning sample characteristics and study design, see A. W. Gottfried and Gottfried (1984), as well as A.W. Gottfried, Gottfried, Bathurst, and Guerin (1994). In the course of investigation, participants were administered a battery of

standardized tests in the university laboratory. Additionally, parents responded to standardized home environmental inventories. These analyses concern 72 fathers of FLS children at the 16 and 17 year waves.

7.2.2 Measures

Parental motivational practices

Parental motivational practices for encouragement of their children's academic motivation were assessed with the Parental Motivational Practices Survey (PMPS; A.E. Gottfried et al., 1994). Psychometric properties can be found in A.E. Gotffried et al. (1994) and A. E. Gottfried, Marcoulides, Gottfried, and Oliver (2009). The instrument consists of 10 item subscales, named Task Intrinsic and Task Extrinsic. The items were responded to using a 6 point Likert-type scale, where 6 = *not at all true* to 1 = *very true*. Examples of items for tasks fostering intrinsic motivation include "*I encourage my child to be persistent in school work*." and "*I try to expose my child to new experiences*". Examples of task-extrinsic PMPS items include "*Reward him/her with money*" and "*Provide child with learning materials*". For the purposes of demonstration these items are treated as continuous. With six response categories the items could be represented as ordered categories, though descriptive results presented later reflect evidence of multivariate normality, important for methods employed here. Items that were not appropriate to adolescence were removed (e.g., items having to do with toys or teachers), such that 11 total items were retained, with seven Task-Intrinsic items and four Task-Extrinsic items.

7.2.3 Hypothesized Model

The proposed model (see Figure 6.1) was fit to the data using *R* (R Core Team, 2017). A two factor solution for each wave is proposed based on both theoretical rationale highlighted briefly in the introduction and previous research outlined already using maternal report. It should

be noted that response choice is not forced between extrinsic and intrinsic items. Although previous research reflected factors were distinguishable and correlation between factors was nonsignificant (r = .10), given the longitudinal nature of the data and the fact the earlier finding was based on maternal report the traditional inclusion of covariances between factors was followed. Errors of the same items across wave are also allowed to covary. The model degrees of freedom were 203 (22*23/2 =253 - 61 = 192). The model was fit with Maximum Likelihood (ML), Generalized Least Squares (GLS), rGLS as well as RGLS methods.

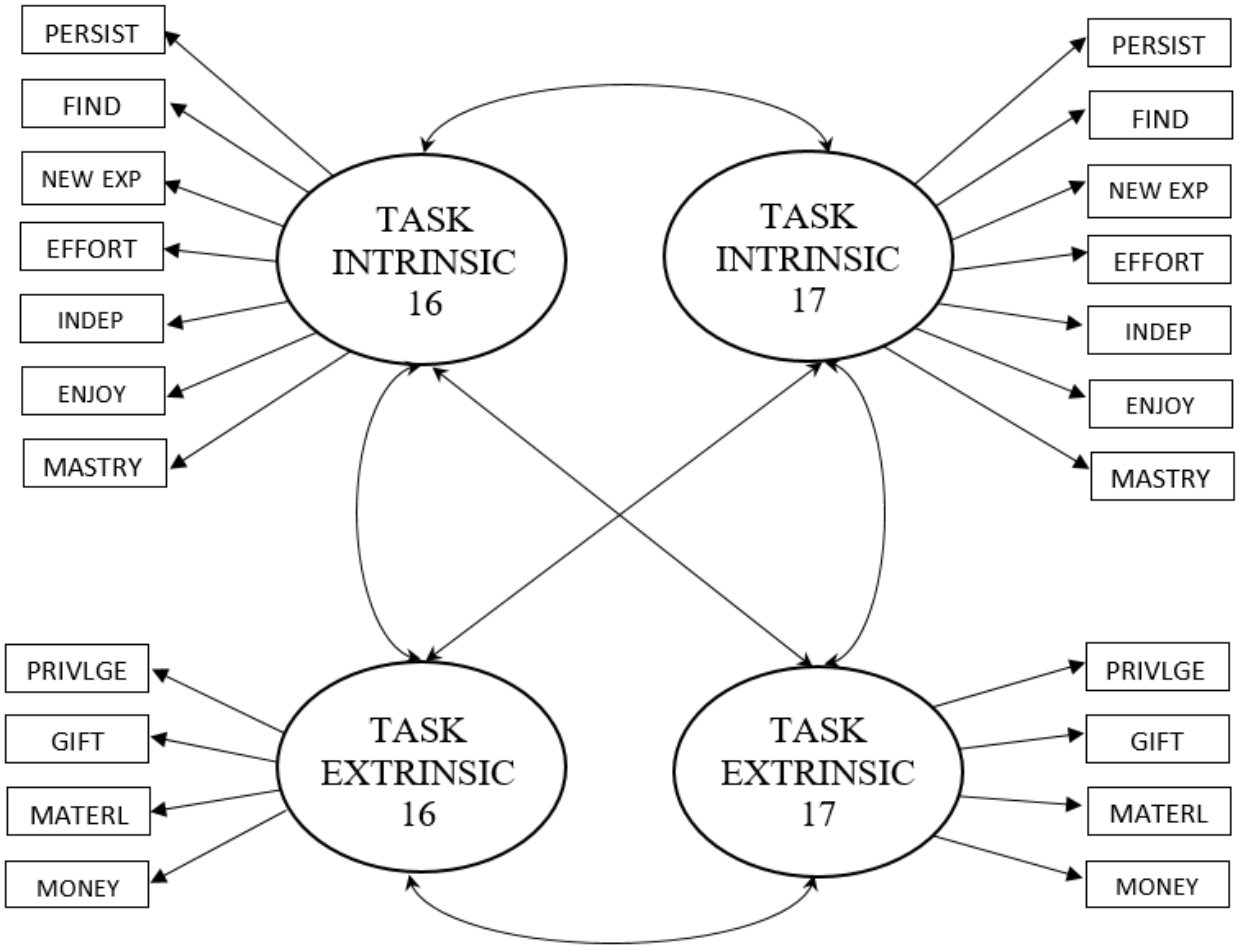

Figure 6.1. CFA model for parental motivational practices.

7.3 Results

Descriptive statistics including means and standard deviations of items are presented in Table 37. On average, means for Intrinsic items were slightly higher than average scores for Extrinsic items. Results suggest a similar amount of variability across items. Though some items are slightly negatively skewed and fewer items exhibited mild positive kurtosis (i.e., Items 1 and 13), Mardia's normalized multivariate kurtosis coefficient (1970, 1974) evidenced multivariate normality (2.16) given the estimate is smaller than a benchmark of 3 (Bentler, 2006). When examining zero-order correlations, items of the Task Intrinsic practices subscale reflect moderate to strong relationships, ranging up to .74. Correlations for Task Extrinsic practices reflected similar inter-relationships, ranging up to .71. As expected the same items across waves enjoyed moderate to strong relationships, though more so for indicators of the Extrinsic (up to $r = .71$) than the Intrinsic ($r = .57$) factor. Generally, most items between the two subscales are not related, though a very few number items reflect somewhat small to moderate relationships, with correlations ranging up to .31. The condition number of the sample covariance matrix was 100.79. The condition numbers for MAP estimators $\widehat{\mathbf{\Sigma}}_{\mathbf{r}}$ and $\widehat{\mathbf{\Sigma}}_{\mathbf{R}}$ were 16.66 and 20.33, respectively, reflecting improvement from the original condition number.

All models converged. The methods had similar numbers of iterations (ML = 42, Regularization methods = 41 each), while GLS had the highest number (46). The test statistics for the four methods and associated *p*-values are displayed in Table 36. ML has the largest test statistic, followed by RGLS then GLS and finally rGLS. When examining *p*-values, ML rejects the model, while RGLS, GLS and rGLS retain the model, given the traditional benchmark of .05.

Table 36.

Model Results for Test Statistics by Method

Method	*T*	*df*	*T*/*df* Ratio	*p*
ML	253.66	192	1.32115	<.001
GLS	201.79	192	1.05099	.299
rGLS	142.18	192	0.74052	.997
RGLS	217.11	192	1.13078	.104

Note. N = 72

Parameter estimates are displayed in Table 37. Naïve standardized estimates based on the product of the estimate and the square root of the variance of the associated latent factor the item loaded on was calculated in order to allow for comparisons across method. In general GLS estimates are smaller than ML with RGLS/rGLS estimates falling between the two (see bolded entries in Table to demonstrate how often this occurred). RGLS and rGLS estimates are quite similar, or in some cases even identical.

Table 37.

Descriptive Statistics and Loadings for Items from the Parental Motivational Practices Survey

Item	Description	Mean	SD	Factor	RGLS		ML		GLS		rGLS	
					Est	Std.lv	Est	Std.lv	Est	Std.lv	Est	Std.lv
1	Persist	4.03	1.21	Int 16	1	-	1	-	1	-	1	-
2	Find Out	4.49	0.99		.55	**.52**	.75	.72	.45	.44	.55	.53
3	New Exp	4.44	1.22		.85	**.81**	.98	.94	.69	.68	.84	.81
4	Effort	4.51	1.31		.98	**.93**	.99	.95	.91	.90	.97	.93
5	Indep	4.31	1.16		.92	**.88**	1.01	.97	.53	.52	.89	.86
6	Enjoy	4.73	1.15		.99	.94	.95	.91	.94	.93	.98	.94
7	Mastery	4.51	1.21		.91	**.87**	.96	.92	.67	.66	.89	.86
8	Privilege	3.79	1.64	Ext 16	1		1		1		1	-
9	Gift	3.33	1.76		.88	**.81**	.97	1.04	.90	.68	.88	.82
10	Material	2.89	1.43		.55	**.51**	.59	.63	.43	.33	.56	.52
11	Money	2.61	1.55		.80	**.74**	.92	.98	.72	.55	.82	.76
12	Persist	4.59	1.29	Int 17	1	-	1	-	1	-	1	-
13	Find Out	4.16	1.01		.75	1.04	.69	.92	.59	.76	.74	1.02
14	New Exp	4.38	1.25		.85	1.18	.85	1.13	.61	.78	.84	1.16
15	Effort	4.35	1.14		.96	**1.33**	.89	1.18	1.14	1.47	.95	1.31
16	Indep	4.46	1.24		.80	**1.11**	.86	1.14	.54	.69	.78	1.08
17	Enjoy	4.29	1.15		.79	1.09	.81	1.08	.65	.84	.79	1.09
18	Mastery	4.82	1.11		.60	**.83**	.65	.86	.51	.66	.59	.81
19	Privilege	3.49	1.62	Ext 17	1	-	1	-	1	-	1	-
20	Gift	3.07	1.57		1.29	1.38	1.28	1.39	1.23	1.25	1.31	1.38
21	Material	2.51	1.51		.80	**.86**	.86	.94	.59	.60	.81	.85
22	Money	2.61	1.59		1.19	**1.27**	1.28	1.39	.95	.97	1.21	1.27

Note. Est = unstandardized estimate, Std.lv = Est standardized by latent factor variance, Bolded estimates reflect RGLS Std.lv estimates that are > GLS and < ML.

7.4 Conclusions and Implications

Regularized methods improve upon GLS in regards to solution propriety. Additionally, with regards to model evaluation both regularized solutions are superior to the traditional method ML, with respect to smaller test statistics and ratios, though rGLS should not be trusted given simulation results in Study 1. ML is known to in general over-reject true models while GLS is known to under-reject models at small sample sizes such as this (e.g., Bentler & Yuan, 1999), therefore those test statistics and model evaluation results should not be trusted. In fact, the pattern of results closely also reflect the same patterns of results demonstrated in Study 1.

If a researcher used the default ML as the estimation method, entering into an exploratory mode via model modification would most likely be a next step and could very well improve the outcome for ML. For example, a typical next step might be to investigate the potential for cross loadings, or residual correlations. This was not the point of this illustration therefore will not be further explored, but this points to another potential benefit of regularization—fewer model modifications may improve generalizability of the model to other samples rather than overfitting a model to sample characteristics as may occur when trying to get a poor model to fit. As of now, because the method is novel and coding does not take advantage of a built package, model modifications are difficult because the model matrices have to be reparametrized manually for each change. While this is not practical for the typical user of a black box program, certainly another next step would be to modify the process to be more accessible, efficient, and user-friendly.

With the ability to fit the model to a smaller sample, follow-up analyses that may have been impossible before may be possible. For example, additional analyses might examine ecological validity of the scale via inclusion of an outcome of interest, such as children's

intrinsic and extrinsic motivation or academic achievement. Given the usual practice of parceling or using a composite to maintain manageable number of *p*s relative to small *N*s, the new regularization methodology may allow for the retention of and inclusion of additional latent factors with indicators that fully account for measurement error and furthermore, better reflect the construct of interest.

More broadly, as researchers are interested in studying questions surrounding non-traditional groups that may be much smaller than majority groups (e.g., non-biological fathers, or fatherhood among gay men) and the number of data points collected are increased given more efficient avenues of data collection, regularization techniques like these are especially relevant given the small *N*/large *p* problems. Moreover, self-identifying categories allow for inclusion and acknowledgement of diversity, but these same categories create problems for analyses given such interest in smaller sub-samples. Rather than collapsing categories or wasting data collected, these new techniques may allow for more flexible analyses and allow researchers to retain the benefits that SEM affords.

APPENDIX

Table A.

Replication of Yuan & Chan 2016 results: Rejection rates for Larger N Conditions of Three Test Staistics

Method	N = 150			N = 200			N = 300			N = 500			N = 1000			N = 2000			N = 3000		
	Y & C	Sim	Dif	Y & C	Sim	Dif	Y & C	Sim	Dif	Y & C	Sim	Dif	Y & C	Sim	Dif	Y & C	Sim	Dif	Y & C	Sim	Dif
T_{ML}	30	55	25	35	43	8	36	42	6	30	36	6	21	32	11	21	21	0	24	28	4
T_{RML}	33	58	25	38	46	8	34	44	10	31	38	7	21	32	11	21	22	1	26	28	2
T_{RGLSI}	104	128	24	103	110	7	95	96	1	65	84	19	47	62	15	38	35	-3	40	41	1

Note. Three factor oblique model, 500 Replications, a = Tuning Parameter, Y & C = Yuan & Chan (2016), Sim = Simulation, Dif = Difference.

REFERENCES

Akaike, H. (1971). Autoregressive model fitting for control. *Ann. Inst. Statist. Math., 23*, 163-180.

Anderson, T. W. & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions, *The Annals of Statistics, 16,* 759–771.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155–173.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, Statistical Section, 3,* 77-85.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual differences, 42*, 815-824.

Banerjee, S., Monni, S., & Wells, M. T. (2016). A regularized profile likelihood approach to covariance matrix estimation. *Journal of Statistical Planning and Inference, 179*, 36-59.

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

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47,* 563-592.

Bentler, P. M., & Yuan, K. -H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34,* 181-197.

Bickel, P. J., & Li, B. (2006). Regularization in statistics. *TEST: Sociedad de Estadística e Investigación Operativa, 15*, 271-344.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis

models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction*: *Part 1* (pp. 149-173). Amsterdam, The Netherlands: North-Holland.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 52*, 345–370.

Boomsma, A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 518-540.

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

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics, 18*, 5 - 46.

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

Chou, C. -P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*, (pp. 37-55). Thousand Oaks, CA: Sage.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Donoho, D. L. (2000). High dimensional data analysis: the curses and blessings of

dimensionality [PDF document]. In Math Challenges of 21st Century American Mathematical Society. Retrieved from http://www-stat.stanford.edu/donoho/Lectures/AMS2000/.

Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika, 81,* 425-455.

El Karouri, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory, *Annals of Statistics, 36,* 2757-2790.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13,* 1-22.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43,* 521-532.

Flouri, E. (2005). *Fathering and child outcomes.* New York, NY, US: John Wiley & Sons Ltd.

Flouri, E. & Buchanan, A. (2003). What predicts fathers' involvement with their children? A prospective study of intact families. *British Journal of Developmental Psychology,* 21, 81-97.

Flouri, E. & Buchanan, A. (2004) Early father's and mother's involvement and child's later educational outcomes. *British Journal of Educational Psychology, 74,* 141-153

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal, 7,* 356-410.

Gottfried, A. E., Flemming, J. S., & Gottfried, A.W. (1994). Role of parental motivational practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology, 86,* 104-113.

Gottfried, A. W., & Gottfried, A. E. (1984). Home environment and cognitive development in young children of middle-socioeconomic status families. In A. W. Gottfried (Ed.), *Home*

environment and early parent motivation practices, 737 *Cognitive Development: Longitudinal Research* (pp. 57–115). New York: Academic Press.

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. (2009). A latent curve model of parental motivational practices and developmental decline in math and science academic intrinsic motivation. *Journal of Educational Psychology,101,* 729-739.

Gottfried, A.W., Gottfried, A. E., & Guerin, D.W. (2006). The Fullerton longitudinal study: A long-term investigation of intellectual and motivational giftedness. *Journal for the Education of the Gifted,* 29, 430-450.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning.* Springer Series in Statistics. New York, NY: Springer New York Inc.

Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and sample sizes. *British Journal of Mathematical and Statistical Psychology, 57*, 327-351.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three–Testing the theory in structural equation models! *Personality and Individual Differences, 42*, 841-850.

Headrick, T. C. (2002). Fast fifth order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis, 40,* 685-711.

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

Hoerl, A. E., Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics. 12,* 55–67.

Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscript.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26,* 329-367.

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

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

Jackson, D. L., Voth, J., & Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural Equation Modeling: A Multidisciplinary Journal, 20,* 86-97.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23,* 1-12.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443-482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183-202.

Kamada, A. (2011). *Selecting a shrinkage parameter in structural equation modeling with a near singular covariance matrix by the GIC minimization method.* (TR 11-04). Hiroshima, Japan: Statistical Research Group, Hiroshima University.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44,* 486-507.

Kock, N. (2016). Non-normality propogation amoung latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods, 15*, 299-315.

Lai, K., & Kelley, K. (2011). Accuracy in parameter estimation for targeted effects in structural equation modeling: Sample size planning for narrow confidence intervals. *Psychological Methods, 16,* 127-148.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88,* 365-411.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics, 40,* 1024–1060.

Ledoit, O., & Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis, 139*, 360-384.

Liang, L. (2016). *Closeness of factor analysis and principal component analysis in semi-high-dimensional condition*. (Unpublished doctoral dissertation). University of Hawai'i, Mānoa, Hawaii.

Lorbert, A. & Ramadge, P. J. (2010). Descent methods for tuning parameter refinement. *Journal of Machine Learning Research, 9,* 469-476.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

Marčenko, V. A., & Pastur, L. A., (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb., 1*, 507–536.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519-30.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics-Series B, 36,* 115-128.

Mark, B. A., & Belyea, M. (2009). Nurse staffing and medication errors: Cross-sectional or Longitudinal Relationships? *Research in Nursing & Health, 32*, 18-30.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Mooijaart, A. (1985). A note on the computational efficiency in asymptotically distribution-free correlational models. *British Journal of Mathematical and Statistical Psychology, 38,* 112-115.

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, 19*, 86-98.

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171-189.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8,* 287-312.

Pornprasertmanit, S., Miller, P. & Schoemann, A. (2016). simsem: SIMulated Structural Equation Modeling. R package version 0.5-13. https://CRAN.R project.org/package= simsem.

Pourahmadi, M. (2013). *High-dimensional covariance estimation.* Hoboken, NJ: John Wiley & Sons, Inc.

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

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology, 22,* 249-278.

Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis, 10,* 235-249.

Satorra, A. & Bentler P. M. (1991). Goodness-of-fit Test under IV Estimation: Asymptotic robustness of a NT Test Statistic. In Gutiérrez, R. & M. J. Valderrama (Eds.), *Proceedings of the Fifth International Symposium on Applied Stochastic Models and Data Analysis,* (pp. 555-567), London: World Scientific.

Savalei, V. (2008). Is the ML chi-square ever robust to nonnormality? A cautionary note with missing data. *Structural Equation Modeling, 15,* 1-22.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics, 6,* 461-464.

Stein, C. (1956). *Some problems in multivariate analysis: Part 1* (Report No. 6). Department of Statistics, Stanford University.

Swain, A. J. (1975). *Analysis of parametric structures for variance matrices* (Unpublished doctoral dissertation). University of Adelaide, Australia.

Taasoobshirazi, G. & Wang, S. (2016). The performance of the SRMR, RMSEA, CFI, and TLI:

An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods, 11*, 31-39.

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika, 45*, 273-279.

Tanaka J. S. (1984). *Some results on the estimation of covariance structure models***. (**Doctoral Dissertation) Retrieved from ProQuest, UMI: 8413682.

Tanaka, J. S. (1987). How big is big enough?: Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,* 134-146.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B., 58,* 267-288.

Tikhonov, A. N. (1943). Об устойчивости обратных задач [On the stability of inverse problems]. *Doklady Akademii Nauk SSSR. 39,* 195–198.

Tong, T., Wang, C. & Wang, Y. (2014). Estimation of variances and covariances for highdimensional data: A selective review. *WIREs Computational Statistics, 6,* 255-264.

Touloumis, A. (2015). Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings. *Computational Statistics and Data Analysis, 83,* 251-261.

Tsukuma, H. (2016). Estimation of a high-dimensional covariance matrix with the Stein loss. *Journal of Multivariate Analysis, 148,* 1-17.

Vale, C. D. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48,* 465-471.

Vinrod, H. D. (1982). Maximum entropy measurement error estimates of singular covariance matrices in undersized samples. *Journal of Econometrics, 20,* 163-174.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonormal variables. Problems and remedies. In R. H. Hoyle (Ed.). *Structural Equation Modeling. Concepts, issues, and applications*, (pp. 56–75). Thousand Oaks, CA: Sage.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73,* 913-934.

Won, J. -H., Lim, J., Kim, S. -J., & Rajaratnam, B. (2013). Condition-number-regularized covariance estimation. *Journal of the Royal Statistical Society B, 75,* 427-450.

Xu, K., & He, D. (2015). Further results on estimation of covariance matrix. *Statistics and Probability Letters, 101,* 11-20.

Yuan, K. -H., & Bentler, P. M. (2017). Improving the convergence rate and speed of Fisher-scoring algorithm: Ridge and anti-ridge methods in structural equation modeling. *Annals of the Institute of Statistical Mathematics, 69,* 1-27.

Yuan, K. -H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics and Data Analysis, 52,* 4842-4858.

Yuan, K. -H., & Chan, W. (2016). Structural equation modeling with unknown population distributions: Ridge Generalized Least Squares. *Structural Equation Modeling: A Multidisciplinary Journal, 23,* 163-179.

Yuan, K. -H., Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika, 80,* 379-405.

Yuan, K. -H., Wu, R., Bentler, P. M. (2011). Ridge structural equation modelling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology, 64,* 107–133.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodology, 67,* 301-320.