ABSTRACT

UNFRIED, ALANA JANE. Variable Selection in Factor Analysis. (Under the direction of Dr. Dennis Boos and Dr. Leonard Stefanski.)

Factor analysis is commonly used to describe the covariance structure for a group of variables through a set of underlying latent factors. Variables that are associated with the same factor are correlated with one another. It is often desirable to remove variables from the model that do not contribute to any factors, which can be accomplished by identifying and eliminating variables that are uncorrelated with any others. Traditionally, variable selection in factor analysis is done by removing variables based on the magnitude of their factor loadings after the model has been estimated. However, this approach relies on somewhat arbitrary cutoffs, with ambiguous cases at borders.

We propose a pre-screening method that identifies uncorrelated variables prior to estimating the factor model. These uncorrelated variables are then removed from the dataset before factor analysis is conducted. This two-step pre-screening procedure first orders variables according to some measure of their correlation with other variables, then uses likelihood-based stopping criteria to determine which variables in the ordering should be estimated as uncorrelated. Ordering methods considered include measures of absolute and maximum correlation, the squared multiple correlation, a likelihood ratio, and a new technique that forces measurement error into the likelihood in order to identify unimportant variables. Various forms of a correlation matrix are also considered in order to best separate correlated and uncorrelated variables. Finally, stopping criteria include information criteria approaches as well as likelihood ratios, with one possible adaptation focused on controlling the false selection rate. Simulation studies show that some of these new pre-screening methods are competitive, though not uniformly superior to, the traditional approaches to variable selection in factor analysis.

Variable Selection in Factor Analysis

by
Alana Jane Unfried

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2016

APPROVED BY:

Dr. David Dickey	Dr. Adam Meade
Dr. Dennis Boos Co-chair of Advisory Committee	Dr. Leonard Stefanski Co-chair of Advisory Committee

DEDICATION

To Kyra. Pursue your dreams.

BIOGRAPHY

Alana Jane Unfried was born in 1987 in Castro Valley, California, to Jim and Corby Nichol, joining siblings Brenton and Trevor. She spent her childhood in California, Colorado, and Wisconsin, prior to graduating from Northridge High School in Greeley, Colorado, in 2005. She received her B.S. in Mathematics from Point Loma Nazarene University in San Diego, California, in 2009, where she also met her now-husband, Jordan Unfried. Their daughter, Kyra Jane Unfried, was born in 2014. In the fall of 2009, Alana began graduate school in the Department of Statistics at North Carolina State University in Raleigh, North Carolina. During her time as a graduate student, Alana spent two years as a teaching assistant or instructor, and five years as a research assistant for Dr. Eric Wiebe at The Friday Institute for Educational Innovation. She received her Master's degree in Statistics in 2011. She will complete her Ph.D. in Statistics in 2016 under the direction of Drs. Dennis Boos and Leonard Stefanski. In the fall of 2016, she will join the Department of Mathematics and Statistics at California State University - Monterey Bay as an Assistant Professor of Statistics.

ACKNOWLEDGEMENTS

The completion of my Ph.D. in Statistics would not be possible without the support of many wonderful people. First and foremost, I would like to thank my advisors, Drs. Dennis Boos and Leonard Stefanski. Thank you for staying with me for almost four years of working on this dissertation. Dr. Boos, thank you for initially brainstorming projects we might work on that aligned with my research interests. Thank you for continually checking in on me when I slowed down a bit, for gracefully answering my endless questions, and for encouraging me to finish and pursue my own career and family goals. Dr. Stefanski, you have made me a better statistician. Thank you for asking me tough questions and pushing me to provide more rigorous answers than I sometimes want to. I would not be nearly as prepared for my career without you as my "bad cop." To both of you, thank you for supporting me as I have chosen to pursue both a Ph.D. and a family at the same time. You will never know how much your support has meant to me. I could not have chosen a better pair of advisors.

Thank you also to my committee members, Drs. David Dickey and Adam Meade, for your time spent reading my work and giving valuable feedback. I thoroughly enjoyed taking classes from both of you, and hope to inspire and challenge my own students the way that you do yours.

I would also like to thank Dr. Eric Wiebe, my supervisor-turned-colleague at The Friday Institute, for hiring me and keeping me employed for five years. The MISO project provided such a rich learning experience for designing a study, collecting responses, analyzing messy, complicated data, and communicating findings to others outside of my field. I would have a huge hole in my education as a statistician were it not for these applied experiences. Further, you taught me the "soft skills" of academia, how to collaborate with others, and how to strike a good balance between statistical rigor and practicality. Thank you for helping me to believe in myself as a statistician, and always being available for questions and advice during my job search. Your confidence in my abilities translated into my own confidence in my interviews. I look forward to continually working with you on future projects.

My desire as a statistician is to help the world understand and see the importance of statistics, and to make it interesting, through a career as an educator. A wonderful mentor in this area has been Dr. Herle McGowan, who supervised me during my year in NC State's Preparing the Professoriate program. Dr. McGowan, thank you for the incredible amount of time you gave toward helping me grow as a statistics educator. I learned so much about

teaching both from watching you in the classroom and discussing the details of learning. You have also been a wonderful example of a woman who has chosen both family and career, which made entering that same role a little less terrifying for me. Don't be surprised if I keep asking you questions even as I start my own career.

Graduate school would not have been survivable without my "grad school besties" Kristin Linn, Gina-Maria Pomann, and Kasturi Talapatra. Since we view the past through rose-colored glasses, I will remember our late nights spent on homework and studying fondly, rather than with the terror we sometimes felt at the time. I look forward to many reunions at conferences, and can you imagine, maybe just for fun! Thank you also to Alison McCoy for your relentless belief that we could make it, and for always giving me someone to talk to when the days got a little stressful. I am aslo grateful for Pam Arroway, who, in my very first year as a graduate student, was already tuned in to my research interests and found me my position at The Friday Institute.

Moving back in time, it would be remiss of me to not acknowledge my undergraduate advisors at Point Loma Nazarene University, Drs. Greg Crow and Maria Zack. You provided not only an excellent undergraduate education, but also unparalleled mentorship, and my first experience with research. Without your encouragement, I would definitely not have applied to a Ph.D. program, and the trajectory of my life might look very different today. Dr. Zack, thank you as well for the numerous hours you spent enlightening me about the job search process. I am thankful to now consider you both friends, and hope that I can be even half as good in the classroom as you are. Thank you for giving me something to aspire to.

To my parents. Thanking you both has to start with the very beginning of my life, because you have shaped me since before I was born. You raised me to believe in my abilities and to always ask questions. I never felt odd for being a woman who loved math. You have always encouraged me to pursue my dreams at the expense of comfort or proximity to family. I am so grateful for your endless support and encouragement through my many teary-eyed phone calls and stressful days. I really could not have asked for better parents. I am truly so excited to live closer to you so that we can enjoy life together.

To my darling daughter Kyra, my biggest joy. Sometimes I have wanted to quit graduate school so that I could just play with you and watch you grow up. But, I want you to grow up knowing that you are capable, that you can pursue your dreams. And so I hope that I have been an example of that. You have provided me with the greatest motivation to finish, so that I can be a more present parent, and so that someday during challenging times, you

can look back and say “Mom did it, so I can do it too.” And so this is for you.

Lastly, to Jordan. You are last not because you are least cared about, but because you are my foundation. You have been beside me for every step of the past seven years. I struggle even thinking of words to describe how thankful I am that I have not journeyed alone. You have been my constant encouragement, and your endless dedication and care does not go unnoticed. You have graciously given me permission to focus on school at the expense of your own bell-being. I promise that once I graduate, I will help with the dishes! I am so looking forward to this next phase of life with you.

“To God be the glory forever and ever! Amen.” -Galatians 1:5

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 Motivation

Exploratory factor analysis is commonly used to describe the covariance structure for a group of variables through a set of underlying latent factors. If variables in a certain subset are highly correlated with one another, but lowly correlated with variables outside the subset, then the subset of variables might represent a single latent factor that explains the correlations (Johnson and Wichern, 2007, p. 481). Therefore, variables that are highly correlated within a group, but are uncorrelated with variables outside of the group, are said to "load" on the same factor.

Factor analysis is often used in psychology research, specifically for developing questionnaires or tests (Conway and Huffcutt, 2003; Costello and Osborne, 2005; Tabachnick and Fidell, 2001, p. 582). For example, a survey might aim to measure student opinions about various school subjects. Survey items that ask about mathematics might correspond

to one factor, and items asking about physical education might correspond to another factor.

In order to ensure that these underlying factors accurately measure what they intend to measure, it is often desirable to remove variables from the factor model that are uncorrelated with any others, as they do not contribute to measuring any particular factor and will reduce the accuracy in estimating the covariance structure. Variables are traditionally removed after the factor model has been generated, based on how high the correlations are between each variable and the factors. However, these approaches often lack objectivity since they are based on rules of thumb. Another task in factor analysis is choosing the appropriate number of factors, which must be specified prior to estimating factor loadings. Many traditional methods have been shown to be ineffective, although others work well but are not widely used.

More recently, statisticians and psychometricians have made advances in new ways to identify uncorrelated variables and select the number of factors. The most promising algorithms need to be more widely tested.

We address the issue of variable selection in factor analysis in a two-phase pre-screening process. First, we order variables according to various measures of their correlation with other variables. These methods include applications of the likelihood ratio, squared multiple correlation, and measurement error methods introduced in Stefanski et al. (2014). Second, we develop stopping criteria to determine the ideal cutoff in the ordering for grouping variables as correlated and uncorrelated. Therefore, this pre-screening process identifies and eliminates marginally correlated variables prior to estimating the factor model. After uncorrelated variables are removed from the model, we recommend the use of parallel analysis for determining the number of factors to estimate (Horn, 1965). Then, maximum likelihood factor analysis is utilized to estimate factor loadings (Lawley and Maxwell, 1971, pp. 25-34). We compare this pre-screening process to more traditionally-used variable selection techniques in factor analysis, as well as to a penalty method for selecting the number of factors and appropriate subset of variables (Hirose and Konishi, 2012).

Chapter 1 describes the factor model, methods of estimation, and current options for selecting variables and the number of factors. Chapter 2 introduces new pre-screening methods for identifying uncorrelated variables, including the measurement error approach,

as well as initial simulation results. Chapter 3 pairs these pre-screening methods with parallel analysis for determining the number of factors. A simulation study is presented comparing these new approaches with other approaches that have been developed for identifying uncorrelated variables in factor analysis. Examples can be found in Chapter 4, with a conclusion in Chapter 5.

1.2 The Common Factor Model

1.2.1 Model Definition

Consider the unstandardized common factor model

$$\underset{p\times 1}{\mathbf{X}} = \underset{p\times 1}{\boldsymbol{\mu}} + \underset{p\times k}{\mathbf{L}}\,\underset{k\times 1}{\mathbf{F}} + \underset{p\times 1}{\boldsymbol{\epsilon}}, \tag{1.2.1}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ is a random vector with p elements with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathbf{L}$ is a matrix of factor loadings, k is the number of factors, $\mathbf{F} = (F_1, F_2, \ldots, F_k)^T$ is a vector of common factors or factor scores, and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_p)^T$ is a vector of errors, also called specific factors. The random variables in $\mathbf{F}$ and $\boldsymbol{\epsilon}$ are all unobservable, which distinguishes the factor model from the regression model. $\mathbf{L}$ is sometimes expressed in terms of its rows and columns as

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{1,\cdot} \\ \vdots \\ \mathbf{L}_{p,\cdot} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{\cdot,1}, & \ldots & \mathbf{L}_{\cdot,k} \end{pmatrix},$$

with individual elements (L_{ij}) where $i = 1, \ldots, p$ and $j = 1, \ldots, k$. Additional assumptions are necessary to make the model identifiable:

1. $E(\mathbf{F}) = \mathbf{0}$, $\text{Cov}(\mathbf{F}) = \mathbf{I}$,
2. $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{D}_{\boldsymbol{\Psi}} = \text{diag}(\boldsymbol{\Psi})$ where $\boldsymbol{\Psi} = (\Psi_1, \Psi_2, ..., \Psi_p)^T$,
3. $\mathbf{F}$ and $\boldsymbol{\epsilon}$ are independent, so $\text{Cov}(\mathbf{F}, \boldsymbol{\epsilon}) = \mathbf{0}$.

1.2.2 Covariance Structure

Because of these assumptions, the factor analysis model covariance is

$$\text{Cov}(\mathbf{X}) = \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}}. \tag{1.2.2}$$

The variance of each X_i can be viewed as a combination of communality and specific variance; that is,

$$\text{Var}(X_i) = \sigma^2_{X_i} = h_i^2 + \Psi_i \tag{1.2.3}$$

where communality of X_i is defined as $h_i^2 = L_{i1}^2 + \cdots + L_{ik}^2$, the sum of squares of the loadings of X_i on the k factors. Communality can be interpreted as the portion of variance in X_i explained by the k factors. Further, $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$, which means that each factor loading L_{ij} can be interpreted as the covariance between the variable X_i and the factor F_j.

The population correlation matrix $\boldsymbol{\rho}$ can also be factored. Define $\mathbf{D}_{\mathbf{\Sigma}}$ as the $p \times p$ diagonal matrix of population variances and $\mathbf{D}_{\mathbf{\Sigma}}^{1/2}$ as the $p \times p$ diagonal matrix of population standard deviations. Then standardized variables are defined as $\mathbf{Z} = \mathbf{D}_{\mathbf{\Sigma}}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, and

$$\begin{aligned} \boldsymbol{\rho} &= \mathbf{D}_{\mathbf{\Sigma}}^{-1/2}\mathbf{\Sigma}\mathbf{D}_{\mathbf{\Sigma}}^{-1/2} && (1.2.4)\\ &= (\mathbf{D}_{\mathbf{\Sigma}}^{-1/2}\mathbf{L})(\mathbf{D}_{\mathbf{\Sigma}}^{-1/2}\mathbf{L})^T + \mathbf{D}_{\mathbf{\Sigma}}^{-1/2}\mathbf{\Psi}\mathbf{D}_{\mathbf{\Sigma}}^{-1/2} && (1.2.5)\\ &= \mathbf{L}_{\mathbf{Z}}\mathbf{L}_{\mathbf{Z}}^T + \mathbf{\Psi}_{\mathbf{Z}}. && (1.2.6) \end{aligned}$$

In practice, it is most common to work with the standardized factor model (Johnson and Wichern, 2007, p. 497). Standardization prevents any particular variable from unduly influencing the estimated factor structure due to a large variance.

1.2.3 Rotational Indeterminacy and Simple Structure

The matrix of factor loadings $\mathbf{L}$ is not unique. Consider an orthogonal matrix $\mathbf{T}$, meaning $\mathbf{T}\mathbf{T}^T = \mathbf{I}$. Define a new loadings matrix $\mathbf{L}^* = \mathbf{L}\mathbf{T}$. Then

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}} = \mathbf{L}(\mathbf{T}\mathbf{T}^T)\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}} = (\mathbf{L}^*)(\mathbf{L}^*)^T + \mathbf{D}_{\mathbf{\Psi}} \tag{1.2.7}$$

Therefore $\mathbf{L}$ and $\mathbf{L}^*$ both lead to the same $\mathbf{\Sigma}$, even though the factor loadings in each

matrix differ. Communalities and uniquenesses are also unaffected by the orthogonal transformation of **L**. This non-uniqueness of **L** is the rationale for factor rotation, or multiplying the loadings matrix by an orthogonal matrix.

In estimating the factor model, conditions are imposed in order to find unique estimates for **L** and **Ψ**. After a unique solution is found, factor rotations may be employed in order to find a more interpretable form of the loading matrix. "Interpretable" typically implies what is called "simple structure." Ideal simple structure shows each variable having a high absolute loading on only one factor, with all other loadings close to zero (Thurstone, 1947; Johnson and Wichern, 2007, p. 504). Further, interpretability indicates that each factor should have an interpretable meaning to the researcher, which is a somewhat qualitative characteristic.

1.2.3.1 Varimax Rotation

One factor rotation method is called Varimax rotation (Kaiser, 1958). In a study of the use of factor analysis in psychology research, Conway and Huffcutt (2003) found that Varimax rotation was the most common rotation method in practice, used in almost 38% of published articles that were reviewed. Varimax rotation is viewed as an analytical interpretation of simple structure (Johnson and Wichern, 2007, p. 507). This rotation aims to maximize the variance among the p factor loadings for each factor. This would allow the researcher to obtain a set of variables with high loadings on the given factor, and another set of variables with low loadings on the factor. The normalized varimax criterion is defined as

$$v = \sum_{j=1}^{k} \frac{p \sum_{i=1}^{p} \left(L_{ij}^2/h_i^2\right)^2 - \left(\sum_{i=1}^{p} L_{ij}^2/h_i^2\right)^2}{p^2}. \tag{1.2.8}$$

We let $\mathbf{T}_v$ be the orthogonal transformation matrix that maximizes the criterion. Due to the scaling of the unrotated squared loadings by their corresponding communality, the matrix that is rotated is $\mathbf{H}^{-1}\mathbf{L}$, where $\mathbf{H}^2$ is a $p \times p$ diagonal matrix of the communalities. Then the actual varimax loadings matrix is $\mathbf{L}^* = \mathbf{H}^{-1}\mathbf{L}\mathbf{T}_v$ (Mulaik, 2010, p. 311). The loadings are scaled by the communalities such that variables with smaller communalities have a greater influence on the solution (Johnson and Wichern, 2007, p. 507). The original, or

"raw," varimax criterion (Kaiser, 1958) did not include this scaling by communality, and has consistently been found to not produce as clear of a simple structure as the normalized varimax criterion (Mulaik, 2010, p. 311).

1.3 Estimation Methods

For an observed dataset $\mathbf{X}$ ($n \times p$), with $p \times 1$ independent observations $\mathbf{X}_1 \ldots \mathbf{X}_n$, the $p \times p$ matrix of sample covariances is defined as

$$\mathbf{S}_{n-1} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T, \tag{1.3.1}$$

where $\bar{\mathbf{X}}$ is a $p \times 1$ vector of sample means. The diagonal elements of $\mathbf{S}_{n-1}$ are the sample variances $s^2_{n-1,i} = [1/(n-1)] \sum_{j=1}^{n} (X_{ji} - \bar{X}_i)^2$, and off-diagonal elements are the sample covariances $s_{n-1,ik} = [1/(n-1)] \sum_{j=1}^{n} (X_{ji} - \bar{X}_i)(X_{jk} - \bar{X}_k)$. Then the diagonal matrix of sample variances can be defined as

$$\mathbf{D}_{\mathbf{S}_{n-1}} = \begin{pmatrix} s^2_{n-1,1} & 0 & \ldots & 0 \\ 0 & s^2_{n-1,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s^2_{n-1,p} \end{pmatrix}, \tag{1.3.2}$$

The sample correlation matrix is defined as

$$\mathbf{R} = \mathbf{D}^{-1/2}_{\mathbf{S}_{n-1}} \mathbf{S}_{n-1} \mathbf{D}^{-1/2}_{\mathbf{S}_{n-1}}, \tag{1.3.3}$$

with elements $r_{ik} = s_{n-1,ik} / [(s_{n-1,i})(s_{n-1,k})]$. Conversely, $\mathbf{S}_{n-1}$ can be defined in terms of $\mathbf{R}$ as $\mathbf{S}_{n-1} = \mathbf{D}^{1/2}_{\mathbf{S}_{n-1}} \mathbf{R} \mathbf{D}^{-1/2}_{\mathbf{S}_{n-1}}$.

In factor analysis, data are often standardized so that variables with large variance do not overly influence the estimates of factor loadings. In this case, the sample covariance is equal to the sample correlation $\mathbf{R}$ and is used to estimate $\boldsymbol{\rho}$. In general, the factor model can be estimated using either $\mathbf{S}_{n-1}$ or $\mathbf{R}$, where use of $\mathbf{R}$ treats the data as standardized (Johnson and Wichern, 2007, pp. 489, 496).

Given these data, we would like to know if the factor model is a good representation. The goal of factor analysis is to estimate the factor loadings L_{ij} and specific variances Ψ_i with $k < p$ factors that adequately represent the data. There are several different approaches to estimation. The most common approaches are principal components factor analysis, principal axis factoring, and maximum likelihood factor analysis (Johnson and Wichern, 2007, pp. 488-497).

1.3.1 Principal Components Factor Analysis

The principal components factor analysis solution uses spectral decomposition as a way to factor the covariance matrix. Spectral decomposition of $\mathbf{S}_{n-1}$ leads to p eigenvalue-eigenvector pairs $(\hat{\phi}_i, \hat{\mathbf{e}}_i)$ where eigenvalues are ordered $\hat{\phi}_1 \geq \hat{\phi}_2 \geq \cdots \geq \hat{\phi}_p$ and eigenvectors have dimension $(p \times 1)$. Each column of estimated factor loadings is composed of the first k eigenvalue-eigenvector pairs; that is,

$$\hat{\mathbf{L}} = \left(\sqrt{\hat{\phi}_1}\hat{\mathbf{e}}_1, \sqrt{\hat{\phi}_2}\hat{\mathbf{e}}_2, \ldots, \sqrt{\hat{\phi}_k}\hat{\mathbf{e}}_k\right). \tag{1.3.4}$$

Uniquenesses are estimated as $\hat{\psi}_i = s^2_{n-1,i} - \sum_{j=1}^{k} \hat{L}^2_{ij}$. Principal components factor analysis can also be implemented on $\mathbf{R}$ in place of $\mathbf{S}_{n-1}$, with $\hat{\psi}_i = 1 - \sum_{j=1}^{k} \hat{L}^2_{ij}$. In the factor analysis procedure in SAS software, principal components is the default method (SAS Institute Inc., 2013).

Note that this procedure stems from principal components analysis, which summarizes the covariance structure of the data through principal components. Principal components are linear combinations of the variables in $\mathbf{X}$, with coefficients defined by eigenvectors. That is, the ith sample principal component is defined as $\hat{y}_i = \hat{\mathbf{e}}_i^T \mathbf{X} = \hat{e}_{i1}X_1 + \hat{e}_{i2}X_2 + \cdots + \hat{e}_{ip}X_p$ (Johnson and Wichern, 2007, p. 442). Principal components factor analysis utilizes principal components to define factor loadings. In particular, in principal axis factoring, "factor loadings are the scaled coefficients of the first few sample principal components" (Johnson and Wichern, 2007, p. 490). The coefficients of the sample principal components are defined by the eigenvectors, which in principal components factor analysis are scaled by the square roots of the eigenvalues in order to define factor loadings, as in Equation (1.3.4).

1.3.2 Principal Axis Factoring

Principal axis factoring, also called principal factor analysis, is an adaptation of the principal components method (Conway and Huffcutt, 2003; Johnson and Wichern, 2007, p. 494). The procedure is typically conducted using $\mathbf{R}$ rather than $\mathbf{S}_{n-1}$, with one modification. The diagonal elements in $\mathbf{R}$ are replaced with estimates of communalities, giving a reduced sample correlation matrix

$$\mathbf{R}_r = \begin{pmatrix} \hat{h}_1^2 & r_{12} & \dots & r_{1p} \\ r_{12} & \hat{h}_2^2 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \dots & \hat{h}_p^2 \end{pmatrix}. \tag{1.3.5}$$

In the population correlation $\boldsymbol{\rho}$ as defined by the factor model, diagonal elements are defined as $\rho_{ii} = 1 = h_i^2 + \psi_i$, a combination of uniqueness and communality. The off-diagonal elements are defined solely by the factor loadings. That is, $\rho_{ij} = (\mathbf{L_z})_{i,.}(\mathbf{L_z})_{j,.}^T$, $i \neq j$. Therefore the diagonal elements of $\mathbf{R}_r$ are $\hat{h}_i^2 = 1 - \hat{\psi}_i$, where variance unexplained by the k factors has been removed. Then $\mathbf{R}_r$ should be explained solely by the common factors. That is, the estimated loadings are defined as the eigenvalue-eigenvector pairs found from $\mathbf{R}_r$, $\hat{\mathbf{L}}_r = \left\{\sqrt{\hat{\phi}_{1r}}\hat{\mathbf{e}}_{1r}, \sqrt{\hat{\phi}_{2r}}\hat{\mathbf{e}}_{2r}, \dots, \sqrt{\hat{\phi}_{kr}}\hat{\mathbf{e}}_{kr}\right\}$ and $\mathbf{R}_r = \hat{\mathbf{L}}_r\hat{\mathbf{L}}_r^T$ (Johnson and Wichern, 2007, p. 494).

In order to perform this estimation, initial estimates of the communalities h_i^2 must be given, after which a final solution is found iteratively. Most commonly, initial estimates are given as the squared multiple correlation, or R^2, between each each variable X_i and the remaining $p-1$ variables. These values can easily be calculated from the inverse sample correlation matrix $\mathbf{R}^{-1}$ as

$$\hat{h}_i^2 = 1 - 1/(\mathbf{R}^{-1})_{ii} \tag{1.3.6}$$

(Johnson and Wichern, 2007, p. 495). See Appendix A.1, for a detailed explanation of how R^2 is calculated from a correlation matrix.

Principal axis factoring can be implemented in SAS software using principal components factor analysis combined with the statement `priors=smc` (SAS Institute Inc., 2013). Traditional principal components factor analysis, as defined in Section 1.3.1, can be thought of as another case of principal axis factoring, with initial communality estimates of 1.

1.3.3 Maximum Likelihood Factor Analysis

Maximum likelihood estimates of factor loadings and specific variances, and therefore of $\mathbf{\Sigma}$, can be found if the common factors $\mathbf{F}$ and specific factors $\boldsymbol{\epsilon}$ are assumed to follow normal distributions. In particular, assume that $\mathbf{F}_j \sim N(0, \mathbf{I})$ and $\boldsymbol{\epsilon}_j \sim N(0, \mathbf{D_\Psi})$ for $j = 1, \ldots, n$, where n is the sample size. Given the assumptions, this implies that $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ for $j = 1, \ldots, n$. The likelihood is therefore

$$\mathrm{L}(\boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-np/2} |\mathbf{\Sigma}|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} (\mathbf{X}_j - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu}) \right\}, \tag{1.3.7}$$

where $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D_\Psi}$. Through a series of algebraic manipulations (see Appendix A.2), this simplifies to

$$\mathrm{L}(\boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-np/2} |\mathbf{\Sigma}|^{-n/2} \exp\left\{ -\frac{n}{2} \mathrm{tr}[\mathbf{\Sigma}^{-1}\mathbf{S}] - \frac{n}{2} (\bar{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \right\}. \tag{1.3.8}$$

where $\mathbf{S} = (1/n) \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T$. Following Johnson and Wichern (2007, p. 527), we use $\mathbf{S}$ in place of $\mathbf{S}_{n-1}$, as $\mathbf{S}$ is known to be the unstructured maximum likelihood estimator of the covariance matrix $\mathbf{\Sigma}$. Although still employed by some analysts, the use of $\mathbf{S}_{n-1}$ is appropriate for the likelihood obtained from the Wishart distribution, rather than the Normal distribution. Then the log likelihood is

$$l(\boldsymbol{\mu}, \mathbf{\Sigma}) = -\frac{n}{2} \left[p \log(2\pi) + \log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) + (\bar{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \right]. \tag{1.3.9}$$

This likelihood will be maximized when the last additive term is equal to zero; that is, when $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$. Therefore, the profile log likelihood is

$$l(\mathbf{\Sigma}) = -\frac{n}{2} \left[p \log(2\pi) + \log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) \right]. \tag{1.3.10}$$

Several other functions of the log likelihood are also used in practice. For instance, Lawley and Maxwell (1971, p. 26) note that it is more convenient to minimize

$$H(\mathbf{\Sigma}) = \log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \log|\mathbf{S}| - p. \tag{1.3.11}$$

This function is bounded below by 0, which occurs when $\boldsymbol{\Sigma} = \mathbf{S}$. Any discrepancies between $\hat{\boldsymbol{\Sigma}}$ and $\mathbf{S}$ lead to a positive value of H.

1.3.3.1 Finding a Unique Solution

The profile likelihood in Equation (1.3.10) is still not well-defined because there are many possible choices for $\mathbf{L}$ due to rotational indeterminacy. $\mathbf{L}$ becomes well-defined by imposing the arbitrary, but convenient, maximum likelihood uniqueness condition

$$\mathbf{L}^T \mathbf{D}_{\boldsymbol{\Psi}}{}^{-1} \mathbf{L} = \boldsymbol{\Delta} \quad \text{a diagonal matrix} \tag{1.3.12}$$

(Lawley and Maxwell, 1971, p. 27). This condition allows for maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ to be found through numerical optimization (Lawley and Maxwell, 1971, pp. 24-34; Johnson and Wichern, 2007, pp. 528-529). After a solution is found, factor rotations can be employed to obtain a more interpretable solution.

1.3.3.2 Maximum Likelihood Estimation of $\boldsymbol{\rho}$

Common practice is to perform maximum likelihood factor analysis (MLFA) using $\mathbf{R}$ instead of $\mathbf{S}$ to estimate the population correlation matrix $\boldsymbol{\rho}$, as defined in Equation (1.2.6). Although the use of $\mathbf{S}$ is appropriate in Equation (1.3.10), $\mathbf{S}$ can be replaced with $\mathbf{R}$. Due to the invariance property of the MLE, this substitution will result in the standardized estimates $\hat{\mathbf{L}}_{\mathbf{z}} = \mathbf{D}_{\mathbf{S}}^{-1/2} \hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}_{\mathbf{z}} = \mathbf{D}_{\mathbf{S}}^{-1/2} \hat{\boldsymbol{\Psi}} \mathbf{D}_{\mathbf{S}}^{-1/2}$, where $\mathbf{D}_{\mathbf{S}}^{-1/2}$ is a diagonal matrix of the reciprocals of the sample standard deviations as defined by $\mathbf{S}$ (Johnson and Wichern, 2007, pp. 496-497). Similarly, if $\mathbf{S}_{n-1}$ were originally used rather than $\mathbf{S}$ then the same estimates $\hat{\mathbf{L}}_{\mathbf{z}}$ and $\hat{\boldsymbol{\Psi}}_{\mathbf{z}}$ can be found by appropriately multiplying the unstandardized estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$, given from the use of $\mathbf{S}_{n-1}$, by $\mathbf{D}_{\mathbf{S}_{n-1}}^{-1/2}$.

1.3.4 Discussion

Each of the three estimation methods discussed remain popular today (Conway and Huffcutt, 2003; Johnson and Wichern, 2007, p. 488). In their review of 371 organizational research studies that employed exploratory factor analysis, Conway and Huffcutt (2003) found that almost 40% of studies utilized principal components factor analysis, around 22% used

principal axis factoring, and just under 4% employed maximum likelihood factor analysis. Of course these findings are not indicative of use across all fields of study, but this does give useful insight into the use of these estimation methods.

The maximum likelihood approach is increasing in popularity, mostly because it uses a more formal statistical approach that allows for inference (Fabrigar et al., 1999). Although Fabrigar et al. (1999) generally prefer maximum likelihood factor analysis due to the usefulness of available fit indices, the authors do recommend either transforming data or avoiding the use of maximum likelihood factor analysis in the presence of severe non-normality.

Principal components factor analysis and principal axis factoring have been found to produce similar results, particularly when the number of variables is large and the number of factors is small (Johnson and Wichern, 2007, p. 495). However, Conway and Huffcutt (2003) note that the motivation of the two methods differ. The goal of principal axis factoring (or any other method considered a "common factor model," such as maximum likelihood factor analysis) is to understand the latent structure of the data. On the contrary, methods such as principal components factor analysis focus solely on data reduction and do not assume a factor model, without concern for the interpretation of the components. Some authors have argued that principal components factor analysis gives almost identical results to common factor analysis methods (Velicer and Jackson, 1990). Others have shown that differences are evident in certain cases; for example, when the common factor model assumptions hold for the data, principal components factor analysis has been found to inflate the estimates of loadings, whereas common factor models are less biased (Widaman, 1993; Gorsuch, 1990).

More recently, de Winter and Dodou (2012) specifically compared principal axis factoring and maximum likelihood factor analysis. The authors found that principal axis factoring outperforms maximum likelihood factor analysis in cases of low population loadings and cases with equal loadings within factors, but that maximum likelihood factor analysis is the most flexible approach that can more reasonably handle model misspecification and unequal loadings within factors. Their final recommendation is to employ multiple estimation methods and compare the findings before selecting a final approach.

1.4 Determining the Number of Factors

Each estimation method discussed in Section 1.3 requires prespecifying the number of factors k to estimate. The unstructured covariance contains $p(p+1)/2$ parameters to be estimated. The factor model assumes that these parameters can be estimated using only the pk factor loadings and p specific variances. If $k = p$ then $\mathbf{LL}^T$ exactly reproduces $\mathbf{\Sigma}$, where $\mathbf{L}$ must be estimated through spectral decomposition of $\mathbf{\Sigma}$ (Johnson and Wichern, 2007, p. 485). However, factor analysis is not useful in that case. The goal is to reduce the number of parameters that must be estimated and to summarize multiple variables with one factor. This occurs when k is less than p. However, the problem lies in deciding what exact value k should take.

If a method other than pure spectral decomposition is used, such as maximum likelihood factor analysis, then the maximum number of factors allowed can be defined based on the number of parameters that must be estimated. The number of unknown parameters to be estimated in the factor model is $pk + p - k(k-1)/2$ due to the $p \times k$ loadings matrix, the p uniquenesses, and the k(k-1)/2 constraints on the loadings matrix in order to give a unique solution (Mulaik, 2010, pp. 173-174). This total number of parameters must be less than the number of unique elements in the covariance matrix, $p(p+1)/2$, in order to make the model identifiable. Solving for k, we find that

$$k \leq \frac{(2p+1) - \sqrt{(8p+1)}}{2}, \tag{1.4.1}$$

which is known as the Lederman inequality (Lederman, 1937; Mulaik, 2010, p. 174).

Although this may be the maximum number of factors allowed, this is likely not the ideal number of factors to estimate. Overextraction, or retaining too many factors, can be problematic (Wood et al., 1996; Fabrigar et al., 1999). Comrey (1978) found that overextraction, when coupled with varimax rotation, leads to inflated loadings in the extraneous factors, and deflated loadings in the appropriate factors, causing a deterioration in the estimation of loadings. Velicer and Jackson (1990) also found that when too many factors were retained, the pattern of significant loadings in the population $\mathbf{L}$ matrix was unable to be recovered by the estimated loadings matrix, even after varimax rotation.

Underextraction, or retaining too few factors, is often considered a more severe problem than overextraction, and has been shown to lead to error in estimation of factor loadings (Fava and Velicer, 1992; Fabrigar et al., 1999). Comrey (1978) also noted that underextraction can force multiple factors to become one composite factor, which gives a misleading pattern of loadings. Further, estimates of loadings for variables that should load highly on a certain variable can be distorted (Wood et al., 1996).

Due to the problems that come with under- or overestimating the number of factors, properly determining k is essential to accurate estimation of $\mathbf{L}$. The value of k is traditionally chosen by considering solutions from a variety of methods, since many popular methods often do not agree. Interpretability criteria are also considered, meaning that the most desirable value of k will induce simple structure with factors that are easily identified and interpreted according to the variables that compose each factor. Following is a summary of some of the most common approaches to selecting the number of factors to retain, although other methods also exist.

1.4.1 Kaiser Criterion

Perhaps the most commonly used method for choosing k is the Kaiser criterion (Kaiser, 1960; Zwick and Velicer, 1986). Also called the eigenvalues-greater-than-one criterion, this rule stems from principal components analysis, using the p eigenvalues $\hat{\phi}_i$ found from $\mathbf{R}$ (Section 1.3.1). The method sets k equal to the number of factors that have an eigenvalue greater than 1.0 (Crawford et al., 2010). However, this method tends to severely overestimate the true number of factors and is seen to be arbitrary in cases where an eigenvalue of 1.01 leads to a factor, but 0.99 does not (Conway and Huffcutt, 2003; Zwick and Velicer, 1986; Fabrigar et al., 1999). In fact, in their review of the use of factor analysis in psychological research, Fabrigar et al. (1999) found no known studies that indicate the Kaiser criterion is accurate in defining k. Despite the Kaiser criterion's obvious shortcomings, it remains the default in many statistical packages, contributing to its popularity (Zwick and Velicer, 1986; Mulaik, 2010, p. 186).

1.4.2 Proportion of Variance Explained

Another eigenvalue approach is to choose k to explain a "suitable proportion" of the total sample variance, which is defined by the researcher (Johnson and Wichern, 2007, p. 491). Since the sum of the p eigenvalues derived from $\mathbf{R}$ is p, the proportion of sample variance explained by the ith factor is $\hat{\phi}_i/p$. The number of factors under consideration is increased until the sum of the $\hat{\phi}_j/p$ values reaches a suitable value. This approach lacks clear, objective criteria, as the researcher determines the desired proportion, and therefore k.

1.4.3 Scree Plot

Yet another common eigenvalue method is based on Cattell's scree plot, in which a plot of the ordered eigenvalues against their order number is created from the spectral decomposition in principal components factor analysis (Cattell, 1966; Johnson and Wichern, 2007, p. 445). The researcher determines where an elbow in the graph appears and sets k equal to the number of eigenvalues above the elbow. This approach depends on a decision made by a researcher viewing a graph that can be difficult to interpret, so reliability is a concern (Gorsuch, 1983; Zwick and Velicer, 1986).

1.4.4 Likelihood Ratio Test

When a normal population density is assumed, one recommendation is to use a likelihood ratio test to determine k (Johnson and Wichern, 2007, pp. 501-503). Under the null hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}}$, where $\mathbf{L}$ is determined by k factors, $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \mathbf{D}_{\hat{\mathbf{\Psi}}}$ where $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$ are maximum likelihood estimates as defined in Section 1.3.3. Using a Bartlett correction, the likelihood ratio statistic is defined as

$$LR_k = \left(n - 1 - \frac{2p + 4k + 5}{6}\right) \log\left\{\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \mathbf{D}_{\hat{\mathbf{\Psi}}}|}{|\mathbf{S}|}\right\} \tag{1.4.2}$$

(Bartlett, 1954; Johnson and Wichern, 2007, p. 502). If LR_k is larger than the critical value of a χ^2 distribution with degrees of freedom $((p-k)^2 - p - k)/2$ at some significance level α^*, then H_0 is rejected. In practice, the researcher should first perform the test with 0 factors

and continue to increase the number of factors allowed until the likelihood ratio test fails to reject H_0.

Some researchers warn, however, that personal judgment is required. If n is large, the method will tend to select too many factors, since trivial differences between $\hat{\mathbf{\Sigma}}$ and $\mathbf{S}$ will lead to a rejection of the null hypothesis (Maccallum, 1990; Johnson and Wichern, 2007, p. 503). Fabrigar et al. (1999) note that small sample sizes can also be problematic, because even large discrepancies may not be detected, leading to a failure to reject the null hypothesis and therefore too few factors retained (Humphreys and Montanelli Jr., 1975). Further, many argue that likelihood ratio tests are unrealistic because the null hypothesis is that the factor model is a perfect fit to the data. Rather, factor analysis aims to create a parsimonious approximation to reality, which typically involves fewer factors than the likelihood ratio test would tend to include (Browne and Cudeck, 1992; Maccallum, 1990).

1.4.5 Parallel Analysis

One method for determining the number of factors to retain that appears to have the fewest shortcomings is parallel analysis (Horn, 1965). Parallel analysis first performs principal components analysis on the observed correlation matrix $\mathbf{R}$, calculating the eigenvalues $(\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p)$, as in Section 1.3.1. Then, multiple sample correlation matrices $\mathbf{R}_{PA,j}$ (e.g., $j = 1, \ldots, N$, where $N = 100$ replications) are generated for p uncorrelated variables of the same sample size n as the observed data. Data is typically generated $i.i.d.$ from the standard normal distribution. Eigenvalues $(\phi^*_{1j}, \phi^*_{2j}, \ldots, \phi^*_{pj})$ are calculated for each randomly generated sample correlation matrix $\mathbf{R}_{PA,j}$. Each set of eigenvalues is ordered largest to smallest. A vector of p average ordered eigenvalues $(\bar{\phi}^*_1, \bar{\phi}^*_2, \ldots, \bar{\phi}^*_p)$ is calculated from the generated vectors of eigenvalues. The number of factors to retain is set to the number of eigenvalues from the observed data that are larger than the respective average eigenvalues from the randomly generated data (Crawford et al., 2010). In other words, k is set to the value that gives $\hat{\phi}_k > \bar{\phi}^*_k$ but $\hat{\phi}_{k+1} < \bar{\phi}^*_{k+1}$.

The motivation for this approach is that at the population level, the eigenvalues for a correlation matrix of all uncorrelated variables would all be 1. Therefore, the Kaiser criterion is akin to comparing the eigenvalues from $\mathbf{R}$ to the eigenvalues from a population correlation matrix of p uncorrelated variables. However, at the sample level, the first eigenvalues will

be larger than 1, and the last eigenvalues will be smaller than 1 (Horn, 1965). Therefore, parallel analysis is considered an adaptation of the Kaiser criterion. The Kaiser criterion is a population-based approach, whereas parallel analysis is sample-based. Therefore the parallel analysis approach offers a more realistic set of eigenvalues to compare to the eigenvalues of the observed data.

Simulation studies have found parallel analysis to be accurate across a broad range of scenarios (Humphreys and Montanelli Jr., 1975). Zwick and Velicer (1986) found parallel analysis to be the most accurate of the factor selection methods they tested through simulation studies (compared to the Kaiser criterion and scree test, among others). The accuracy of parallel analysis is only minimally influenced by sample size. Accuracy tends to increase with larger loadings, and with more variables associated with each factor. In general, they found that the consistency of parallel analysis did not vary widely according to simulation factors. Although not the most commonly used method, parallel analysis is one of the only methods for choosing k that is consistently recommended (Fabrigar et al., 1999; Zwick and Velicer, 1986).

Two important adaptations have been suggested for parallel analysis. First, researchers have found that parallel analysis tends to overestimate the number of factors when it is incorrect, rather than underestimate (Glorfeld, 1995; Harshman and Reddon, 1983). Zwick and Velicer (1986) also noted that, although parallel analysis was the best method for determining k that they found, it still did tend to slightly overestimate k. Therefore, Glorfeld (1995) suggested using the 95th percentile of each of set of N eigenvalues, rather than the average, as a comparison to the eigenvalues from the observed data. Harshman and Reddon (1983) noted that the use of average eigenvalues is like setting the Type 1 Error rate to $\alpha^* = 0.5$ in hypothesis testing, whereas using the 95th percentile more appropriately sets the Type 1 error rate to $\alpha^* = 0.05$. The 95th percentile approach is more conservative, leads to retaining fewer factors, and is commonly applied in parallel analysis today (Lorenzo-Seva et al., 2011).

The second adaptation is to apply parallel analysis to the reduced sample correlation matrix $\mathbf{R}_r$, defined in Equation (1.3.5), in place of $\mathbf{R}$. This adaptation performs principal axis factoring in order to calculate eigenvalues, as in Section 1.3.2. This approach has been both recommended and questioned in the past (Humphreys and Montanelli Jr., 1975; Crawford

and Koopman, 1973). More recently, Crawford et al. (2010) studied parallel analysis combinations of principal components versus principal axis factoring, and mean eigenvalues versus the 95th percentile eigenvalues. Mixed results, depending on the number of factors and number of variables per factor, did not allow for any general recommendations. Green et al. (2012) also studied these adaptations of parallel analysis and found that using both principal axis factoring and the 95th percentile rule adaptations provides more accurate estimation of k than traditional parallel analysis. Fabrigar et al. (1999) recommend the joint use of a scree plot and parallel analysis, with $\mathbf{R}_r$, for determining the number of factors.

1.4.6 Other Approaches

Less common in practice, but considered more recently, is the use of information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for selecting the number of factors (Akaike, 1973; Schwarz, 1978). Similar to the likelihood ratio test in Section 1.4.4, series of models are fit using the likelihood approach, with k increasing by 1 in each iteration. The value of k that leads to the minimum AIC or BIC is selected. To our knowledge, only one study has assessed the performance of AIC and BIC in terms of a general success rate for correctly determining k. The study found that BIC (76% success rate) was an improvement over AIC (44% success rate), but BIC performed poorly when sample size was large. (Lorenzo-Seva et al., 2011). In comparison, parallel analysis had an 81% success rate in their study.

There are many other approaches developed to determine k than described here (e.g. Velicer's Minimum Average Partial Test; Velicer, 1976), although, for better or for worse, those described in this section appear to be the most well-known. Researchers continue to develop new methods (e.g., the Hull method from Lorenzo-Seva et al., 2011; a Bayesian method from Lopes and West, 2004; and the GBIC method from Hirose and Konishi, 2012).

1.5 Variable Selection in Factor Analysis

The goal of factor analysis is to identify groups of correlated variables, where each grouping corresponds to one factor. A secondary task aside from determining the number of factors k, then, is to also identify uncorrelated variables that do not contribute to any factor, leaving

a total of m correlated variables in the model. In psychological applications, it is important to identify questionnaire items that do not correlate with other variables so that they can be removed from the survey, or revised (Mavridis and Ntzoufras, 2014). As Tabachnick and Fidell (2001, p. 590) state it,

> "A variable with a low squared multiple correlation with all other variables and low correlations with all important factors is an outlier among the variables. The variable is usually ignored in the current factor analysis and either deleted or given friends in future research."

The same sentiment is expressed by Gorsuch (1988). Recall that squared multiple correlation is often used as an initial estimate of the communality h_i^2 of a variable. Also, a variable loading is defined as the correlation between a variable X_i and a given factor, when $\boldsymbol{\rho}$ is factored in place of $\boldsymbol{\Sigma}$. Therefore, a variable with low correlations with all factors should have loadings that are all close to zero and low communality (Section 1.2.2). Further, if a variable X_i has loadings that are all close to zero such that $\mathbf{L}_{i,\cdot} \approx \mathbf{0}$, then its correlations with the other $(p-1)$ variables are also close to zero. This can be seen when defining an element of $\boldsymbol{\Sigma}$ in terms of two rows of the $\mathbf{L}$ matrix:

$$\begin{aligned} \Sigma_{ij} &= \mathbf{L}_{i,\cdot}\mathbf{L}_{j,\cdot}^T && (1.5.1) \\ &= L_{i1}L_{j1} + L_{i2}L_{j2} + \cdots + L_{ik}L_{jk}. && (1.5.2) \end{aligned}$$

Therefore, if $\mathbf{L}_{i,\cdot} \approx \mathbf{0}$ then $\Sigma_{ij} \approx 0$ for all j. Identifying these lowly correlated variables is crucial for proper estimation of the factor model, since distortion of the results can occur when variables with low communalities are retained in the model (Fabrigar et al., 1999; MacCallum et al., 1999; Hogarty et al., 2004).

One general recommendation for factor analysis is that there should be a minimum of three to five variables associated with each factor (MacCallum et al., 1999; Velicer and Fava, 1998). In cases where a factor is estimated to have less than the recommended number of variables, the recommendations of Tabachnick and Fidell (2001) come into play to either remove these variables or add more relating to the same factor

The concept of variable selection in factor analysis has generally been less studied than other issues in factor analysis, such as estimation methods, determining the number

of factors, and factor rotation (Hogarty et al., 2004). In practice, the process of selecting variables is often informal, with researchers relying on intuition or what has been done in the past (Little et al., 1999). Traditional approaches to variable selection, discussed in the next section, are often ad hoc and subjective (Mavridis and Ntzoufras, 2014). More recently, researchers have been searching for more objective approaches to variable selection in factor analysis through the use of methods such as stepwise selection and penalized likelihood. However, a more formal and objective guideline that is also widely accepted and recommended does not currently exist.

1.5.1 Traditional Approaches

The traditional approaches to variable selection align with the comments of Tabachnick and Fidell (2001) and Gorsuch (1988). In practice, two approaches are a magnitude-of-loadings rule and a magnitude-of-communalities rule. Typically these rules are defined assuming loadings are standardized and we are estimating the population correlation matrix $\boldsymbol{\rho}$.

1.5.1.1 Magnitude of Factor Loadings

The "magnitude rule," as we will call it, states that a variable X_i should be removed if it does not have at least one factor loading of great enough magnitude; that is, for a predetermined cutoff c, if $\text{abs}(L_{ij}) < c,\ j = 1,\ldots,k$, for all j, then the variable should be removed. The actual value of c varies based on who is making the recommendation. One widely applied rule today is to remove a variable if it has no absolute loading greater than $c = 0.32$, as this corresponds to a correlation of about $r_{ij} = 0.1$ (Tabachnick and Fidell, 2001, p. 625). However, other cutoffs are also used in practice, such as $c = 0.3$ (Hair et al., 2006; Hogarty et al., 2004) or $c = 0.4$ (Stevens, 1992). There is no consensus on what cutoff value should be used. This guideline also does not account for aspects such as sample size, number of factors, and rotation method. Further, there is a great deal of ambiguity at borderline cases, such as a variable X_i with a maximum absolute loading $\text{abs}(L_{ij}) = 0.31$. Based on one criterion this variable would be retained in the model, and based on another criterion it would be removed.

As an aside, note that any variable that has more than one factor loading greater than c is said to *cross-load*. Cross-loading is generally undesirable, as factor analysis aims to sum-

marize each group of correlated variables with just one factor. Having a variable associated with multiple factors does not lend itself to a clean interpretation, and is misaligned with the goal of simple structure.

1.5.1.2 Communalities of Variables

A second traditional rule is the "communality rule." This method states that a variable X_i should be excluded from the model if its communality h_i^2 is not above a certain threshold c (Kano and Harada, 2000; Mavridis and Ntzoufras, 2014). Child (2006) recommends using a threshold of $c = 0.2$ when selecting variables based on their communalities. This approach faces problems similar to the traditional magnitude rule in terms of ambiguity for borderline cases and the subjectivity of setting c. Further, since communality of X_i is defined as $h_i^2 = L_{i1}^2 + \cdots + L_{ik}^2$ it is possible for a variable to have low loadings on several factors that total to a communality higher than c, even though no particular loading is high enough to indicate a substantial correlation between variables. Kano and Harada (2000) found that the communality rule did a poor job of identifying uncorrelated variables.

1.5.2 Modern Approaches

In light of the difficulties found in the traditional methods for variable selection in factor analysis, a few alternatives have been proposed.

1.5.2.1 Stepwise Selection using a Likelihood Ratio Test

One alternative method for selecting variables is that of Kano and Harada (2000). Working with a correlation matrix and a prespecified number of factors (methods for selecting k are not discussed), backward elimination and forward selection are used to select appropriate variables to improve model fit. Model fit is assessed by comparing the sample correlation matrix and model-implied correlation matrix through a Lagrange Multiplier Test, which is a substitute for the likelihood ratio test in order to reduce computation time (Kano and Ihara, 1994). The form of the likelihood ratio is identical to that defined for determining the number of factors in Section 1.4.4, including the use of the Bartlett correction. However, in this case, the likelihood ratio test is not used to determine the number of factors, since k is

assumed to be predetermined. Rather, if the likelihood ratio test is significant, then variables are removed from the model until a subset of variables is found that is consistent with the k factor model. However, the hypothesis can be defined equivalently as $H_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}}$. The likelihood ratio statistic is similarly defined as

$$\text{LR}_{\text{HK}} = \Big(n - \frac{2p+4k+5}{6}\Big)(\log|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \mathbf{D}_{\hat{\mathbf{\Psi}}}| - \log|\mathbf{S}|), \tag{1.5.3}$$

The likelihood ratio statistic is compared to a χ^2 distribution with degrees of freedom $((p-k)^2 - p - k)/2$.

If the value of LR_{HK} is greater than the corresponding critical value, then H_0 is rejected and a new model is tested. The new model removes one variable and calculates the likelihood ratio test using $\mathbf{S}_{(p-1)}$, the sample covariance matrix for the $(p-1)$ remaining variables, and $\hat{\mathbf{\Sigma}}_{(p-1)}$, the estimated covariance structure under the factor model for the $(p-1)$ variables. This process is repeated for each of the p variables. Next, one deletes the variable whose removal causes the greatest improvement in model fit. If the reduced likelihood ratio test is still significant, continue the process iteratively to remove variables until the model is not rejected; this implies a subset of variables has been found that is consistent with a k-factor model (Hogarty et al., 2004).

Generally speaking, the stepwise selection method starts with the full set of p variables in the model and tests how dropping each variable would affect the model fit, given k factors. Once variables have been removed, forward selection also becomes an option. Kano and Harada (2000) do not perform simulation studies, but they compare their procedure to the traditional communalities rule for a few sample datasets. They find that their stepwise method is superior, and the communalities approach results in a different set of variables that lead to a poor model fit.

Through simulation studies, Hogarty et al. (2004) compare the stepwise procedure to the traditional magnitude rule with $c = 0.3$ (Section 1.5.1.1). One feature of the stepwise procedure that they note is that the stepwise procedure aims to include as many variables as possible under the constraint that the model correlation does not differ significantly from the sample correlation. Therefore, excluding a variable is based on obtaining statistical significance, so low power may lead to including anomalous variables. This is a reversal of how the stepwise algorithm is used in cases like regression, where low power leads to the

exclusion of important variables.

Hogarty et al. (2004) give a theoretical argument against the stepwise procedure. They note that the stepwise method only identifies certain types of anomalous variables. The algorithm searches for variables consistent with a k-factor model. However, some variables, such as those uncorrelated with the m variables that should be included, would be difficult to identify as anomalous since the sample and model correlations would still be similar. Consider a covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{p-1} & \mathbf{0} \\ \mathbf{0} & \sigma_p^2 \end{pmatrix}. \tag{1.5.4}$$

where $\mathbf{\Sigma}_{p-1}$ is the population covariance matrix for $(p-1)$ variables and σ_p^2 is the population variance for X_p. In this case, X_p is uncorrelated with all other variables and should be identified for removal. However, this covariance structure can still be adequately represented by a factor model, with loadings matrix

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{p-1} \\ \mathbf{0} \end{pmatrix}, \tag{1.5.5}$$

where $\mathbf{L}_{p-1}$ is a $(p-1) \times k$ matrix of factor loadings. Therefore the stepwise method of Kano and Harada (2000) would likely fail to identify X_p for removal.

The simulation study by Hogarty et al. (2004) confirms these ideas about the stepwise method. Namely, the stepwise method fails in specificity, meaning it tends to include anomalous variables in the factor model that should be excluded. The traditional magnitude rule performs much better in this sense. The stepwise and traditional methods are generally comparable regarding sensitivity (correctly identifying the m variables that should be retained in the model).

Hirose and Konishi (2012) found in their own simulation study that the stepwise method often retained all p variables rather than eliminating uncorrelated variables.

1.5.2.2 Group Lasso Factor Analysis

More recently, researchers have tested penalized likelihoods for selecting variables in the factor model. The lasso is a popular regularization method developed by Tibshirani (1996) for variable selection and estimation. Hirose and Konishi (2012) apply the group lasso

(Yuan and Lin, 2006), which was designed for selecting groups of variables, and gives this method the name GLAFA (Group Lasso Factor Analysis). Hirose and Konishi (2012) define elements of one row of the loading matrix, $\mathbf{L}_{i,\cdot} = (L_{i1}, L_{i2}, \ldots, L_{ik})$ as a group, so that these elements will be penalized together. Combined with a weighted penalty, this approach will set entire rows of the loading matrix to zero, instead of setting individual elements L_{ij} to zero. A variable with all loadings set to zero can be removed from the model. The penalized likelihood to maximize is

$$l_\theta(\mathbf{\Sigma}) = l(\mathbf{\Sigma}) - n\theta P_w(\mathbf{L}), \tag{1.5.6}$$

where $l(\mathbf{\Sigma})$ is the profile log likelihood defined in Equation (1.3.10), n is the sample size, and θ is the regularization parameter. $P_w(\mathbf{L})$ is the weighted penalty function defined as

$$P_w(\mathbf{L}) = \sum_{i=1}^{p} \hat{w}_i ||\mathbf{L}_{i,\cdot}|| \quad \text{where} \quad ||\mathbf{L}_{i,\cdot}|| = \sqrt{\mathbf{L}_{i,\cdot}\mathbf{L}_{i,\cdot}^T} = h_i, \tag{1.5.7}$$

$$\hat{w}_i = c * \frac{1}{||\hat{\mathbf{L}}_{i,\cdot}^{ML}||}, \qquad c = \frac{p}{\sum_{h=1}^{p} 1/||\hat{\mathbf{L}}_{h,\cdot}^{ML}||},$$

where h_i is the square root of the communality and $\hat{\mathbf{L}}_{i,\cdot}^{ML}$ is the traditional maximum likelihood estimate of $\mathbf{L}_{i,\cdot}$ from Section 1.3.3. The weights $\hat{w}_i$ are based on the adaptive lasso technique and aim to select variables with high communality, and therefore low uniqueness (Zou, 2006). Note that $\mathbf{L}_{i,\cdot}\mathbf{L}_{i,\cdot}^T = L_{i1}^2 + \cdots + L_{ik}^2$ is the definition of communality for X_i. Therefore if a variable has a high estimated communality, the weight $\hat{w}_i$ will be small and the corresponding loadings will not be as highly penalized. Hirose and Konishi (2012) also considered an unweighted penalized likelihood, but found it inferior to the weighted approach.

The EM algorithm is used to maximize the penalized likelihood, and therefore estimate $\mathbf{L}$ and $\mathbf{D}_{\mathbf{\Psi}}$. The model is estimated for a grid of regularization parameters θ and number of factors k. For each model, the generalized BIC (called GBIC) is calculated. The GBIC is an adaptation of the BIC specifically for models estimated via penalized likelihoods (Konishi et al., 2004). The final values of θ and k are chosen to minimize the GBIC. After estimation is complete, they apply the Varimax rotation to obtain an interpretable solution.

Through a simulation study, Hirose and Konishi (2012) compared this approach to the

stepwise selection approach from Section 1.5.2.1. Since the stepwise selection method does not specify a method for selecting k, they used maximum likelihood estimation and the BIC approach to determine k (Section 1.4.6) prior to running the stepwise algorithm. GLAFA was found to select the true model more often than the stepwise approach, as noted in Section 1.5.2.1. The stepwise method often selected all variables rather than properly identifying uncorrelated variables.

When investigating the stability of the estimated factor loadings in GLAFA, Hirose and Konishi (2012) found that the mean squared error between $\hat{\mathbf{L}}$ and $\mathbf{L}$ was smaller for traditional maximum likelihood estimation than for GLAFA. They believe this is due to the sometimes large value of θ, which led to attenuated factor loading estimates. However, they argue that determining the correct factor model is more important than precisely estimating the loadings themselves.

Since GLAFA is fairly new, additional researchers have not yet compared it to other factor analysis approaches, so it has not been externally validated.

1.5.2.3 Related Methods

Mavridis and Ntzoufras (2014) have implemented a Markov chain Monte Carlo algorithm for selecting variables in factor analysis, and also extend their algorithm to determine the number of factors. In simulation studies and real datasets, the authors find that their approach can successfully detect the pattern of factor loadings. However, it does not appear that they compare their methods to any previously established methods for selecting variables in factor analysis.

While not directly removing variables from the factor model, other recent approaches have focused on estimating a sparse $\mathbf{L}$ matrix via penalized likelihoods as an alternative to rotation methods such as Varimax. If all estimated loadings associated with a variable X_i are small enough, or even zero, these approaches might also be considered variable selection methods. Choi et al. (2010) use the l_1 regularization to create sparsity via the penalized likelihood

$$l_\theta(\mathbf{\Sigma}) = l(\mathbf{\Sigma}) - \frac{1}{2}\theta P_\theta(\mathbf{L}), \tag{1.5.8}$$

$$P_\theta(\mathbf{L}) = \sum_{j=1}^{k}\sum_{i=1}^{p}|L_{ij}|. \tag{1.5.9}$$

Choi et al. (2010) also discuss using the adaptive lasso instead of the regular lasso. They recommend a two-step process, where first the ordinary lasso estimator is computed, and then those estimates are used to form weights for the new penalty. Hirose and Konishi (2012) also used an adaptive-lasso-like weight, but used the maximum likelihood estimates of $\mathbf{L}$ to form the weights. Consider the ordinary lasso-estimated weights $\hat{L}_{ij}^{OL}$. To define the new adaptive-lasso estimator, the authors state that if $\hat{L}_{ij}^{OL} = 0$ then $\hat{w}_{ij} = \infty$, otherwise $\hat{w}_{ij} = 1/|\hat{L}_{ij}^{OL}|$. Then the adaptive-lasso penalty is

$$P_\lambda(\mathbf{L}) = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{p}\hat{w}_{ij}|L_{ij}|. \tag{1.5.10}$$

A generalized EM algorithm is used to compute the estimates. The number of factors k is treated as a meta-parameter of the MLE. The k which minimizes the Kullback-Leibler loss, calculated using a validation dataset, is selected. The authors conclude that if the data are generated from a truly sparse model, the penalized model not only discovers the loadings that should be set to zero, but also is significantly more accurate than the regular MLE model. Ning and Georgiou (2011) use the same penalty but a different algorithm, and conclude that the penalized approach leads to a sparser estimation of $\mathbf{L}$ than Varimax rotation. However, this is not necessarily an indication that the penalized approach leads to a more obvious simple structure, as defined in Section 1.2.3.

Hirose and Yamamoto (2015) also developed an EM algorithm using coordinate descent to create a sparser solution than traditional rotation methods. Their method also works for cases where $p > n$, and can be used with a wide range of penalties such as the lasso, SCAD, and MC+. While this method is promising, the authors note that it would be beneficial to find a faster algorithm.

1.5.3 The Need for an Alternative

While several approaches to variable selection in factor analysis currently exist, each has its unique shortcomings. The traditional rules are somewhat subjective and lead to borderline cases where the correct decision is often unclear. The stepwise selection method fails to identify what should be obviously uncorrelated variables. The GLAFA method, while appearing promising, has not been widely studied due to its recent development, and

also has sub-par performance in stability of factor loading estimates when compared to the traditional magnitude approach. Further, the model must be estimated for multiple values of k in order to find an optimal solution before selecting the final subset of variables, leading to a computationally intensive procedure. There is clearly room for improvement on these variable selection techniques.

The main variable selection methods discussed thus far also all require that the factor model be estimated before variables can be identified for removal. In Chapter 2, we explore an alternative approach to variable selection in factor analysis that pre-screens variables, without assuming a factor model, in order to identify uncorrelated variables.

CHAPTER

2

PRE-SCREENING VARIABLES

2.1 Introduction

The main variable selection methods discussed thus far (traditional approaches, stepwise selection, and group lasso factor analysis) all require that the factor model be estimated before variables can be removed. Further, in some cases the model must be estimated for multiple values of k in order to find an optimal solution before selecting the final subset of variables. Alternatively, we consider an approach that allows variables to be screened for removal prior to estimating the factor model or determining k. Our approach is related to the recommendations in Section 1.5 that state a variable should be removed from the factor model if it has only low correlations with other variables. Our method identifies variables that have low correlations with all other variables, without first needing to choose k or estimate the factor model.

This process of pre-screening variables is done in two phases, without any assumption

of a factor model. First, we order the set of p variables according to the magnitude of correlation with other variables. The magnitude of correlation can be measured in several ways, which are introduced in Section 2.2. Next, Section 2.3 defines several stopping criteria that can be used to determine where a cutoff should be placed in the sequence of ordered variables in order to classify correlated and uncorrelated variables. Recommendations are given in Section 2.4

After this classification is complete, factor analysis can be completed in the usual way, with uncorrelated variables already removed from the model.

2.1.1 Structure of $\boldsymbol{\Sigma}$

Consider a dataset of n iid p-vectors $\mathbf{X}_1, \ldots \mathbf{X}_n$ where $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T$. The corresponding population covariance matrix can be partitioned as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \underset{m\times m}{\boldsymbol{\Sigma}_{11}} & \underset{m\times u}{\boldsymbol{\Sigma}_{12}} \\ \underset{u\times m}{\boldsymbol{\Sigma}_{21}} & \underset{u\times u}{\boldsymbol{\Sigma}_{22}} \end{pmatrix}, \tag{2.1.1}$$

where $m + u = p$. Similarly, the sample covariance matrix can be partitioned as

$$\mathbf{S} = \begin{pmatrix} \underset{m\times m}{\mathbf{S}_{11}} & \underset{m\times u}{\mathbf{S}_{12}} \\ \underset{u\times m}{\mathbf{S}_{21}} & \underset{u\times u}{\mathbf{S}_{22}} \end{pmatrix}. \tag{2.1.2}$$

Our goal is to estimate the covariance matrix $\boldsymbol{\Sigma}$ in such a way that we clearly identify variables that are uncorrelated with the rest. Without loss of generality, we place the uncorrelated variables in the last u positions in the vector of p variables. Recall from Section 1.2.2 that $\mathbf{D}_{\boldsymbol{\Sigma}}$ is the $(p \times p)$ diagonal matrix of population variances. Similarly, $\mathbf{D}_{\boldsymbol{\Sigma}_{22}}$ is the $(u \times u)$ diagonal matrix of population variances for the u uncorrelated variables. Let the notation $\mathbf{A}_{-u}$ define any covariance or correlation matrix with the last u variables set as uncorrelated. That is, if we know that u variables are uncorrelated in the population, then $\boldsymbol{\Sigma}$ can be defined as

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{-u} = \begin{pmatrix} \underset{m\times m}{\boldsymbol{\Sigma}_{11}} & \underset{m\times u}{\mathbf{0}} \\ \underset{u\times m}{\mathbf{0}} & \underset{u\times u}{\mathbf{D}_{\boldsymbol{\Sigma}_{22}}} \end{pmatrix}, \tag{2.1.3}$$

In other words, $\mathbf{\Sigma}_{-u}$ is the covariance matrix of $(X_1,\ldots,X_m,X_{m+1},\ldots X_p)$ where $(X_1,\ldots,X_m)$ have covariance matrix $\mathbf{\Sigma}_{11}$, and $(X_{m+1},\ldots,X_p)$ are uncorrelated with $(X_1,\ldots,X_m)$ and with each other, with covariance matrix

$$\mathbf{D}_{\mathbf{\Sigma}_{22}} = \begin{pmatrix} \sigma^2_{X_{m+1}} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma^2_{X_p} \end{pmatrix}, \tag{2.1.4}$$

For simplicity, define $\mathbf{A}_{bb}$ as any correlation or covariance matrix for a subset of b variables. Then $\mathbf{\Sigma}_{-u}$ can also be denoted

$$\mathbf{\Sigma}_{-u} = \begin{pmatrix} \mathbf{\Sigma}_{mm} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{\Sigma}_{uu}} \end{pmatrix}. \tag{2.1.5}$$

2.1.2 Block-Diagonal Covariance as MLE

For iid normal random vectors, it is known that the unstructured maximum likelihood estimator of $\mathbf{\Sigma}$ is $\mathbf{S}$. If the constraint is added that the last u variables in the population are uncorrelated, then $\mathbf{\Sigma}$ can be estimated with $\hat{\mathbf{\Sigma}}_{-u}$, defined as the sample covariance $\mathbf{S}$ with covariances of the u uncorrelated variables set to 0:

$$\hat{\mathbf{\Sigma}}_{-u} = \begin{pmatrix} \mathbf{S}_{mm} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}} \end{pmatrix}. \tag{2.1.6}$$

$\mathbf{S}_{mm}$ is the sample covariance matrix of the m correlated variables. Since $\mathbf{D}_{\mathbf{S}}$ is defined in Section 1.3 as the $p \times p$ diagonal matrix of sample variances, then $\mathbf{D}_{\mathbf{S}_{uu}}$ is defined as the $u \times u$ diagonal matrix of the sample variances of the u uncorrelated variables.

The derivation of $\hat{\mathbf{\Sigma}}_{-u}$ is as follows. If $\mathbf{\Sigma}$ is constrained such that u variables are uncorrelated, then $\mathbf{\Sigma}$ is partitioned as defined as in Equation (2.1.5), and the profile log likelihood from Equation (1.3.10) is denoted $l(\mathbf{\Sigma}_{-u}) = (-n/2)\big[p\log(2\pi) + \log|\mathbf{\Sigma}_{-u}| + \text{tr}(\mathbf{\Sigma}_{-u}^{-1}\mathbf{S})\big]$. $\mathbf{S}$ is then partitioned as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{mm} & \mathbf{S}_{mu} \\ \mathbf{S}_{um} & \mathbf{S}_{uu} \end{pmatrix}. \tag{2.1.7}$$

One property of determinants is that

$$\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D}| \tag{2.1.8}$$

for any $n \times n$ matrix $\mathbf{A}$ and $m \times m$ matrix $\mathbf{D}$ (Silvester, 2000). Then $|\mathbf{\Sigma}_{-u}| = |\mathbf{\Sigma}_{mm}||\mathbf{D}_{\mathbf{\Sigma}_{uu}}|$. Now consider $\text{tr}(\mathbf{\Sigma}_{-u}^{-1}\mathbf{S})$.

$$\begin{aligned} \text{tr}(\mathbf{\Sigma}_{-u}^{-1}\mathbf{S}) &= \text{tr}\left[\begin{pmatrix} \mathbf{\Sigma}_{mm}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{\Sigma}_{uu}} \end{pmatrix}\begin{pmatrix} \mathbf{S}_{mm} & \mathbf{S}_{mu} \\ \mathbf{S}_{um} & \mathbf{S}_{uu} \end{pmatrix}\right] && (2.1.9) \\ &= \text{tr}\left[\begin{pmatrix} \mathbf{\Sigma}_{mm}^{-1}\mathbf{S}_{mm} & \mathbf{\Sigma}_{mm}^{-1}\mathbf{S}_{mu} \\ \mathbf{D}_{\mathbf{\Sigma}_{uu}}\mathbf{S}_{um} & \mathbf{D}_{\mathbf{\Sigma}_{uu}}\mathbf{S}_{uu} \end{pmatrix}\right] && (2.1.10) \\ &= \text{tr}(\mathbf{\Sigma}_{mm}^{-1}\mathbf{S}_{mm}) + \text{tr}(\mathbf{D}_{\mathbf{\Sigma}_{uu}}^{-1}\mathbf{S}_{uu}). && (2.1.11) \end{aligned}$$

Then the profile log likelihood is

$$\begin{aligned} l(\mathbf{\Sigma}) &= (-n/2)\left[p\log(2\pi) + \log|\mathbf{\Sigma}| + \text{tr}(\mathbf{\Sigma}^{-1}\mathbf{S})\right] && (2.1.12) \\ &= (-n/2)\left[p\log(2\pi) + \log(|\mathbf{\Sigma}_{mm}||\mathbf{D}_{\mathbf{\Sigma}_{uu}}|) + \text{tr}(\mathbf{\Sigma}_{mm}^{-1}\mathbf{S}_{mm}) + \text{tr}(\mathbf{D}_{\mathbf{\Sigma}_{uu}}^{-1}\mathbf{S}_{uu})\right] && (2.1.13) \\ &= (-n/2)\left[p\log(2\pi) + \log|\mathbf{\Sigma}_{mm}| + \text{tr}(\mathbf{\Sigma}_{mm}^{-1}\mathbf{S}_{mm})\right] \\ &\quad + (-n/2)\left[p\log(2\pi) + \log|\mathbf{D}_{\mathbf{\Sigma}_{uu}}| + \text{tr}(\mathbf{D}_{\mathbf{\Sigma}_{uu}}^{-1}\mathbf{S}_{uu})\right]. && (2.1.14) \end{aligned}$$

The profile log likelihood is factored into two pieces; one for m correlated variables, and one for u uncorrelated variables. Since we are not assuming a factor model, the unstructured maximum likelihood estimator of $\mathbf{\Sigma}_{mm}$ is $\hat{\mathbf{\Sigma}}_{mm} = \mathbf{S}_{mm}$ (Johnson and Wichern, 2007, pp. 168-172). The second piece of the profile log likelihood could be further factored to the form of u univariate profile log likelihood equations, each giving the maximum likelihood estimator of the population variance as $\hat{\sigma}^2_{X_i} = s_i^2$. Then the maximum likelihood estimator of $\mathbf{D}_{\mathbf{\Sigma}_{uu}}$ is $\mathbf{D}_{\hat{\mathbf{\Sigma}}_{uu}} = \mathbf{D}_{\mathbf{S}_{uu}}$. Then $\hat{\mathbf{\Sigma}}_{-u}$ is as defined in Equation (2.1.6).

2.1.3 Simplified Likelihood due to Trace

When estimating $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}_{-u}$, the trace portion of the maximized likelihood simplifies to $\text{tr}(\hat{\boldsymbol{\Sigma}}_{-u}^{-1}\mathbf{S}) = p$. This can be seen through partitioning the matrices.

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{-u}^{-1}\mathbf{S} &= \begin{pmatrix} \mathbf{S}_{mm}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}}^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{S}_{mm} & \mathbf{S}_{mu} \\ \mathbf{S}_{um} & \mathbf{S}_{uu} \end{pmatrix} && (2.1.15)\\
&= \begin{pmatrix} \mathbf{S}_{mm}^{-1}\mathbf{S}_{mm} & \mathbf{S}_{mm}^{-1}\mathbf{S}_{mu} \\ \mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{um} & \mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu} \end{pmatrix} && (2.1.16)\\
&= \begin{pmatrix} \mathbf{I}_{mm} & \mathbf{S}_{mm}^{-1}\mathbf{S}_{mu} \\ \mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{um} & \mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu} \end{pmatrix} && (2.1.17)
\end{aligned}$$

Then $\text{tr}(\hat{\boldsymbol{\Sigma}}_{-u}^{-1}\mathbf{S}) = \text{tr}(\mathbf{I}_{mm}) + \text{tr}(\mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu}) = m + \text{tr}(\mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu})$. Now the diagonal elements of $\mathbf{S}_{uu}$ are equal to the diagonal elements of $\mathbf{D}_{\mathbf{S}_{uu}}$. Then the diagonal elements of $\mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu}$ are equal to one. Therefore, $\text{tr}(\mathbf{D}_{\mathbf{S}_{uu}}^{-1}\mathbf{S}_{uu}) = u$, and $\text{tr}(\hat{\boldsymbol{\Sigma}}_{-u}^{-1}\mathbf{S}) = m + u = p$.

When estimating $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}_{-u}$, then, the maximized log likelihood can be calculated as

$$l(\hat{\boldsymbol{\Sigma}}) = -\frac{n}{2}\left[p\log(2\pi) + \log|\hat{\boldsymbol{\Sigma}}_{-u}| + p\right]. \tag{2.1.18}$$

The objective function in Equation (1.3.11) can also be re-expressed as

$$\begin{aligned}
H(\hat{\boldsymbol{\Sigma}}) &= \log|\hat{\boldsymbol{\Sigma}}_{-u}| + \text{tr}(\mathbf{S}\hat{\boldsymbol{\Sigma}}_{-u}^{-1}) - \log|\mathbf{S}| - p && (2.1.19)\\
&= \log|\hat{\boldsymbol{\Sigma}}_{-u}| + p - \log|\mathbf{S}| - p && (2.1.20)\\
&= \log|\hat{\boldsymbol{\Sigma}}_{-u}| - \log|\mathbf{S}|. && (2.1.21)
\end{aligned}$$

2.1.4 Likelihood in terms of Correlation

The estimated value of the likelihood can also be expressed in terms of the sample correlation matrix $\mathbf{R}$, since $\mathbf{R} = \mathbf{D}_{\mathbf{S}}^{-1/2}\mathbf{S}\mathbf{D}_{\mathbf{S}}^{-1/2}$; see Equation (1.3.3) and Section 1.3.3.2. Similarly,

$\mathbf{S}_{mm} = \mathbf{D}_{\mathbf{S}_{mm}}^{1/2} \mathbf{R}_{mm} \mathbf{D}_{\mathbf{S}_{mm}}^{1/2}$ and $\mathbf{D}_{\mathbf{S}_{uu}} = \mathbf{D}_{\mathbf{S}_{uu}}^{1/2} \mathbf{I}_{uu} \mathbf{D}_{\mathbf{S}_{uu}}^{1/2}$. Then $\hat{\boldsymbol{\Sigma}}_{-u}$ can be factored as

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{-u} &= \begin{pmatrix} \mathbf{S}_{mm} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}} \end{pmatrix} && (2.1.22)\\
&= \begin{pmatrix} \mathbf{D}_{\mathbf{S}_{mm}}^{1/2} \mathbf{R}_{mm} \mathbf{D}_{\mathbf{S}_{mm}}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}}^{1/2} \mathbf{I}_{uu} \mathbf{D}_{\mathbf{S}_{uu}}^{1/2} \end{pmatrix} && (2.1.23)\\
&= \begin{pmatrix} \mathbf{D}_{\mathbf{S}_{mm}}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}}^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{R}_{mm} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{uu} \end{pmatrix} \begin{pmatrix} \mathbf{D}_{\mathbf{S}_{mm}}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{S}_{uu}}^{1/2} \end{pmatrix} && (2.1.24)\\
&= \mathbf{D}_{\mathbf{S}}^{1/2} \hat{\boldsymbol{\rho}}_{-u} \mathbf{D}_{\mathbf{S}}^{1/2}. && (2.1.25)
\end{aligned}$$

where $\hat{\boldsymbol{\rho}}_{-u}$ is the maximum likelihood estimate of the population correlation matrix under the constraint that u variables are uncorrelated. Then

$$\begin{aligned}
\log|\hat{\boldsymbol{\Sigma}}_{-u}| &= \log|\mathbf{S}_{mm}| + \log|\mathbf{D}_{\mathbf{S}_{uu}}| && (2.1.26)\\
&= \log|\mathbf{D}_{\mathbf{S}_{mm}}| + \log|\mathbf{R}_{mm}| + \log|\mathbf{D}_{\mathbf{S}_{uu}}| && (2.1.27)\\
&= \log|\mathbf{D}_{\mathbf{S}}| + \log|\mathbf{R}_{mm}| && (2.1.28)\\
&= \log|\mathbf{D}_{\mathbf{S}}| + \log|\mathbf{R}_{mm}| + \log|\mathbf{I}_{uu}| && (2.1.29)\\
&= \log|\mathbf{D}_{\mathbf{S}}| + \log|\hat{\boldsymbol{\rho}}_{-u}| && (2.1.30)
\end{aligned}$$

since $|\mathbf{I}| = 1$ so $\log|\mathbf{I}_{uu}| = 0$. Then $l(\hat{\boldsymbol{\Sigma}}_{-u}) = (-n/2)\big[p\log(2\pi) + \log|\mathbf{D}_{\mathbf{S}}| + \log|\hat{\boldsymbol{\rho}}_{-u}| + p\big]$. Alternatively,

$$\begin{aligned}
H(\hat{\boldsymbol{\Sigma}}_{-u}) &= \log|\hat{\boldsymbol{\Sigma}}_{-u}| - \log|\mathbf{S}| && (2.1.31)\\
&= \log|\mathbf{D}_{\mathbf{S}}| + \log|\mathbf{R}_{mm}| - \log|\mathbf{D}_{\mathbf{S}}| - \log|\mathbf{R}| && (2.1.32)\\
&= \log|\mathbf{R}_{mm}| - \log|\mathbf{R}| && (2.1.33)\\
&= \log|\mathbf{R}_{mm}| + \log|\mathbf{I}_{uu}| - \log|\mathbf{R}| && (2.1.34)\\
&= \log|\hat{\boldsymbol{\rho}}_{-u}| - \log|\mathbf{R}|. && (2.1.35)
\end{aligned}$$

The task that remains is to determine the appropriate value of u, the number of variables to estimate as uncorrelated, and which of the p variables will comprise this set of u variables estimated as uncorrelated. Variables will first be ordered according to a measure of their

correlation with other variables. Later, stopping criteria will be developed for determining how many variables in the ordered sequence will be estimated as uncorrelated.

2.2 Ordering Variables by Correlation

Several techniques can be used to order variables according to the magnitude of their correlations. The five methods that we consider are discussed in Sections 2.2.1 through 2.2.5. The methods are followed by alternative definitions of sample correlation matrices to consider in Section 2.2.6, and a simulation study to compare the ordering methods in Section 2.2.7. Let $\mathbf{Q}$ be defined as the $(p \times 1)$ sequence of variables ordered by correlation. For each method described here, Q_1 represents the variable with greatest magnitude of correlation, and Q_p represents the variable with smallest magnitude of correlation. After the proper values of m and u are determined by stopping criteria in Section 2.3, the variables $(Q_1, \ldots, Q_m)$ will correspond to the correlated variables in $\mathbf{\Sigma}_{mm}$, and the remaining variables $(Q_{m+1}, \ldots, Q_p)$ will correspond to uncorrelated variables in $\mathbf{D}_{\mathbf{\Sigma}_{uu}}$.

2.2.1 Average Absolute Correlation

One simple approach to ranking variables according to their correlations is to utilize the matrix of absolute values of sample correlations, $\mathbf{A} = \text{abs}(\mathbf{R})$ with elements A_{ij}. The average absolute correlation (AAC) for each variable X_i is calculated as

$$\text{AAC}_i = \frac{1}{p-1} \sum_{j=1}^{p} A_{ij}, \qquad j \neq i. \tag{2.2.1}$$

Variables are ordered according to their AAC value, so that the variable with the largest AAC value is defined as Q_1.

2.2.2 Maximum Absolute Correlation

Similar to the AAC, maximum absolute correlation (MAC) orders variables based on which has the largest maximum absolute correlation. That is, the MAC for each variable X_i is

calculated as

$$\text{MAC}_i = \max(\mathbf{A}_{i,\cdot}), \text{ for } j \neq i, \tag{2.2.2}$$

where $\mathbf{A}_{i,\cdot}$ is row i of the matrix of absolute correlations. Variables are ordered according to their MAC value, so that the variable with the largest MAC value is first in the sequence $\mathbf{Q}$.

2.2.3 Squared Multiple Correlation (R^2)

The squared multiple correlation (SMC) method calculates R^2, also called the coefficient of determination, for each of the p variables regressed on all other $(p-1)$ variables, as defined in Equation (1.3.6). The SMC for each X_i can be interpreted as the proportion of variance in X_i that is explained by the linear relationship with the remaining $(p-1)$ variables (Johnson and Wichern, 2007, p. 367). The R^2 value for X_i regressed on the remaining $(p-1)$ variables can be calculated as

$$R_i^2 = 1 - 1/(\mathbf{R}^{-1})_{ii}. \tag{2.2.3}$$

The variable with the smallest R_i^2 is removed from the dataset and designated as Q_p in the SMC ordering. R^2 is calculated again for each remaining X_j regressed on the remaining $(p-2)$ variables, using the $(p-1)\times(p-1)$ correlation matrix $\mathbf{R}_{(p-1)(p-1)}$. The variable with the smallest R^2 value is removed from the dataset and placed as Q_{p-1} in the ordering.

The process is repeated iteratively until just two variables remain. At this step, since both variables will have the same value of R^2, they are simply placed as Q_1 and Q_2 based on which variable appeared first in the dataset.

We also considered a non-iterative SMC procedure, where variables were ordered solely on the first step outlined above. That is, the R^2 values were calculated as $R_i^2 = 1 - 1/(\mathbf{R}^{-1})_{ii}$ for $i = 1, \ldots, p$, and variables were ordered based on these values alone. This method was found to be inferior to the iterative procedure, meaning it fails to place truly uncorrelated variables toward the end of the $\mathbf{Q}$ vector as often as the iterative procedure does. It was not considered further.

2.2.4 Likelihood Ratio

A likelihood ratio (LR) can be used to order variables as well. A likelihood ratio can be thought of a distance measure between two covariance matrices.

To start, consider the hypothesis that $u = 1$ variable is uncorrelated in the population. That is,

$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_{-1}$

$H_a : \mathbf{\Sigma} \neq \mathbf{\Sigma}_{-1}$.

The likelihood ratio for this hypothesis is defined as

$$\begin{aligned} \text{LR}_{-1} &= -2\log\left(\frac{\text{L}(\hat{\mathbf{\Sigma}}_{-1})}{\text{L}(\mathbf{S})}\right) && (2.2.4)\\ &= 2l(\mathbf{S}) - 2l(\hat{\mathbf{\Sigma}}_{-u}) && (2.2.5)\\ &= -n\left[p\log(2\pi) + \log|\mathbf{S}| + p\right] + n\left[p\log(2\pi) + \log|\hat{\mathbf{\Sigma}}_{-u}| + p\right] && (2.2.6)\\ &= n\log|\hat{\mathbf{\Sigma}}_{-u}| - n\log|\mathbf{S}| && (2.2.7)\\ &= n\log|\hat{\boldsymbol{\rho}}_{-1}| - n\log|\mathbf{R}|, && (2.2.8) \end{aligned}$$

which is a measure of the difference between the unstructured sample correlation and the sample correlation with correlations for one X_i set to zero. This likelihood ratio can be calculated p times, once with each X_i set as the uncorrelated variable. The variable X_i with the smallest likelihood ratio is set as Q_p. This indicates that the measure of difference between $\mathbf{S}$ and $\hat{\mathbf{\Sigma}}_{-1}$ was smallest by setting correlations for X_i to zero.

After Q_p is determined, the process continues iteratively. The variable X_i corresponding to Q_p is set as uncorrelated, and now we consider setting each of the remaining $(p-1)$ variables as uncorrelated as well, for a total of two variables set as uncorrelated. This likelihood ratio test can be adapted for setting any t variables as uncorrelated, and comparing the corresponding covariance structure to one step prior where only $(t-1)$ variables are set as uncorrelated. Thus, at any step for $t = 1 \ldots p$, we test the hypothesis

$H_0 : \mathbf{\Sigma}_{-(t-1)} = \mathbf{\Sigma}_{-t}$

$H_a : \mathbf{\Sigma}_{-(t-1)} \neq \mathbf{\Sigma}_{-t}$,

that t variables are uncorrelated, compared to the previous step where $(t-1)$ variables are uncorrelated. The likelihood ratio is calculated as

$$\begin{aligned} \text{LR}_{-t} &= -2\log\left(\frac{\text{L}(\hat{\mathbf{\Sigma}}_{-t})}{\text{L}(\hat{\mathbf{\Sigma}}_{-(t-1)})}\right) && (2.2.9)\\ &= n\log|\hat{\boldsymbol{\rho}}_{-t}| - n\log|\hat{\boldsymbol{\rho}}_{-(t-1)}|. && (2.2.10) \end{aligned}$$

Q_{p-t+1} is set as the variable X_i that resulted in the smallest likelihood ratio at step t. Also note that the likelihood ratio can be calculated solely from the sample correlation **R** rather than requiring the sample covariance matrix **S**.

2.2.5 Measurement Error Model Selection

Measurement error model selection (MEMSEL) methods, adapted from Stefanski et al. (2014), can be used to identify and eliminate lowly correlated variables from the data. Once again, we do not assume a factor model structure for $\mathbf{\Sigma}$.

Through forcing measurement error into the normal likelihood model, the MEMSEL method results in an estimated reliability ratio (RR) for each of the p variables. The reliability ratio is interpreted as the proportion of variance in the data free from measurement error. The estimated reliability ratios are used to order the p variables in terms of most correlated with other X_j (large RR), to least correlated with other X_j (small RR).

For this ordering method, we start by assuming that data are standardized, meaning centered at zero and scaled to a variance of one. This is done so that adding measurement error to one variable versus another does not unduly have a greater impact due to differing variances among the variables. Further, it ensures that measurement error variances, introduced in Section 2.2.5.1, are all on the same scale (Stefanski et al., 2014). The standardized data will still be referred to as **X** throughout this procedure. Because of standardization, though, the sample covariance matrix is equal to the sample correlation matrix, $\mathbf{S} = \mathbf{R}$. As a reminder that data are standardized, we use **R** in the likelihood equation and throughout the description of this procedure.

2.2.5.1 Four-Step Procedure

The general approach is to force measurement error into the model. An explanation for why this works will be given after the procedure is described.

Step 1

Start with a normal likelihood $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which is assumed to be true. We will work with the profile log likelihood from Equation (1.3.10); that is,

$$l(\boldsymbol{\Sigma}) = -\frac{n}{2}\left[p\log(2\pi) + \log|\boldsymbol{\Sigma}| + \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{R})\right]. \tag{2.2.11}$$

where $\boldsymbol{\Sigma}$ is the unknown covariance to be estimated, and $\mathbf{R}$ is the sample correlation matrix, equal to the standardized sample covariance matrix.

Step 2

Second, "falsely" assume that measurement error exists in the data. That is, assume that instead of observing $\mathbf{X}$, we observe $\mathbf{W}$.

$$\underset{p\times 1}{\mathbf{W}} = \underset{p\times 1}{\mathbf{X}} + \mathbf{D}_{\boldsymbol{\sigma}_U}\underset{p\times 1}{\mathbf{Z}} \tag{2.2.12}$$

where $\mathbf{Z} \sim \text{N}_\mathbf{p}(\mathbf{0}, \mathbf{I})$, and $\mathbf{D}_{\boldsymbol{\sigma}_U^2}$ is a $p \times p$ diagonal matrix of additional variances due to measurement error for each variable with $\boldsymbol{\sigma}_U^2 = (\sigma_{U1}^2, \sigma_{U2}^2, \ldots, \sigma_{Up}^2)$. Therefore $\text{Cov}(\mathbf{W}) = \boldsymbol{\Sigma} + \mathbf{D}_{\boldsymbol{\sigma}_U^2}$, where $\boldsymbol{\Sigma}$ is the true correlation of $\mathbf{X}$, free from measurement error. Note that due to standardization, $\text{Var}(W_i) = \sigma_{Xi}^2 + \sigma_{Ui}^2 = 1 + \sigma_{Ui}^2$. We can further define the population correlation matrix of $\mathbf{W}$ as

$$\boldsymbol{\rho}_W = \mathbf{D}^{-1}_{\sqrt{1+\sigma_U^2}}(\boldsymbol{\Sigma} + \mathbf{D}_{\boldsymbol{\sigma}_U^2})\mathbf{D}^{-1}_{\sqrt{1+\sigma_U^2}}, \tag{2.2.13}$$

where $\mathbf{D}_{\sqrt{1+\sigma_U^2}}$ is a $(p \times p)$ diagonal matrix of the standard deviations of $\mathbf{W}$, with diagonal elements $\sqrt{1+\sigma_{Ui}^2}$. Construct the measurement error likelihood under this assumption of measurement error in the data, giving $l_{MEM}(\boldsymbol{\rho}_W) = -\frac{n}{2}\left[p\log(2\pi) + \log|\boldsymbol{\rho}_W| + \text{tr}[\boldsymbol{\rho}_W^{-1}\mathbf{R}_W]\right]$, where $\mathbf{R}_W$ is the sample correlation of $\mathbf{W}$. However, evaluate the likelihood using the observed, error-free values $\mathbf{X}$, giving

$$l_{MEM}(\boldsymbol{\Sigma}, \boldsymbol{\sigma}_U^2) = -\frac{n}{2}\left[p\log(2\pi) + \log|\boldsymbol{\rho}_W| + \text{tr}[\boldsymbol{\rho}_W^{-1}\mathbf{R}]\right]. \tag{2.2.14}$$

We want to maximize l_{MEM} (or minimize the corresponding objective function) when estimating both the correlation matrix $\boldsymbol{\Sigma}$ and the measurement error variances $\boldsymbol{\sigma}_U^2$.

Step 3

In the third step, we replace $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}} = \mathbf{R}$ (because the data are standardized), giving $\hat{\boldsymbol{\rho}}_W = \mathbf{D}^{-1}_{\sqrt{1+\sigma^2_U}}(\mathbf{R}+\mathbf{D}_{\boldsymbol{\sigma}^2_U})\mathbf{D}^{-1}_{\sqrt{1+\sigma^2_U}}$. This leads to a pseudo-profile log likelihood $\hat{l}_{MEM}(\mathbf{R},\boldsymbol{\sigma}^2_U)$ where only the measurement error variances must be estimated:

$$\hat{l}_{MEM}(\mathbf{R},\boldsymbol{\sigma}^2_U) = -\frac{n}{2}\big[p\log(2\pi)+\log|\hat{\boldsymbol{\rho}}_W|+\text{tr}[\hat{\boldsymbol{\rho}}^{-1}_W\mathbf{R}]\big]. \tag{2.2.15}$$

Step 4

Following the framework of Stefanski et al. (2014), the fourth step is to re-express the pseudo-profile log likelihood in terms of precision rather than variance, with $\lambda_j = 1/\sigma^2_{Uj}$. Then $\hat{\boldsymbol{\rho}}_W$ can be re-expressed as $\hat{\boldsymbol{\rho}}_W(\boldsymbol{\lambda})$ with

$$\begin{aligned}
\hat{\boldsymbol{\rho}}_W(\boldsymbol{\lambda}) &= \mathbf{D}^{-1}_{\sqrt{1+\sigma^2_U}}(\mathbf{R}+\mathbf{D}_{\boldsymbol{\sigma}^2_U})\mathbf{D}^{-1}_{\sqrt{1+\sigma^2_U}} && (2.2.16)\\
&= \mathbf{D}^{-1}_{\sqrt{1+1/\boldsymbol{\lambda}}}(\mathbf{R}+\mathbf{D}_{1/\boldsymbol{\lambda}})\mathbf{D}^{-1}_{\sqrt{1+1/\boldsymbol{\lambda}}} && (2.2.17)\\
&= \mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}}(\mathbf{R}+\mathbf{D}_{1/\boldsymbol{\lambda}})\mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}} && (2.2.18)\\
&= \mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}}(\mathbf{R})\mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}} + \mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}}\mathbf{D}_{1/\boldsymbol{\lambda}}\mathbf{D}_{\frac{1}{\sqrt{1+1/\boldsymbol{\lambda}}}} && (2.2.19)\\
&= \mathbf{D}_{\sqrt{\frac{\boldsymbol{\lambda}}{1+\boldsymbol{\lambda}}}}(\mathbf{R})\mathbf{D}_{\sqrt{\frac{\boldsymbol{\lambda}}{1+\boldsymbol{\lambda}}}} + \mathbf{D}_{\frac{1}{1+\boldsymbol{\lambda}}}, && (2.2.20)
\end{aligned}$$

where $\mathbf{D}_{\sqrt{\boldsymbol{\lambda}/(1+\boldsymbol{\lambda})}}$ and $\mathbf{D}_{1/(1+\boldsymbol{\lambda})}$ are diagonal matrices with vectors $\sqrt{\boldsymbol{\lambda}/(1+\boldsymbol{\lambda})}$ and $1/(1+\boldsymbol{\lambda})$ defining the diagonal elements, respectively. This results in the measurement error selection log likelihood

$$\hat{l}_{SEL}(\boldsymbol{\lambda}) = -\frac{n}{2}\big[p\log(2\pi)+\log|\hat{\boldsymbol{\rho}}_W(\boldsymbol{\lambda})|+\text{tr}[\hat{\boldsymbol{\rho}}^{-1}_W(\boldsymbol{\lambda})\mathbf{R}]\big]. \tag{2.2.21}$$

This log likelihood is maximized subject to the constraints $\lambda_i \geq 0$ and $\sum_j \lambda_j = \tau$, where $0 \leq \tau < \infty$ is the tuning parameter. We set $\tau = p$ for reasons discussed in Section 2.2.5.5.

2.2.5.2 Estimation Details

Similar to Equation (1.3.11), we define the corresponding objective function to be minimized,

$$\hat{H}_{SEL}(\boldsymbol{\lambda}) = \log|\hat{\boldsymbol{\rho}}_W(\boldsymbol{\lambda})|+\text{tr}[\hat{\boldsymbol{\rho}}^{-1}_W(\boldsymbol{\lambda})\mathbf{R}]-\log|\mathbf{R}|-p. \tag{2.2.22}$$

Estimation is done through numerical optimization, which is straightforward since only p parameters $(\lambda_1, \ldots, \lambda_p)$ must be estimated. The resulting estimate $\hat{\boldsymbol{\lambda}}$ can be transformed into the selection likelihood estimate of the correlation matrix (or standardized covariance matrix), with elements

$$\hat{\boldsymbol{\rho}}_{W,ij} = \frac{r_{ij}}{\sqrt{1+1/\hat{\lambda}_i}\sqrt{1+1/\hat{\lambda}_j}} = \frac{r_{ij}}{\sqrt{1+\hat{\sigma}^2_{Ui}}\sqrt{1+\hat{\sigma}^2_{Uj}}} \tag{2.2.23}$$

where r_{ij} is the sample correlation (i.e., standardized sample covariance) for X_i and X_j.

2.2.5.3 Why MEMSEL Works

The general idea of of the MEMSEL method is to separate important and unimportant variables, whether in the context of correlation as it is used here, or in other situations such as regression (Stefanski et al., 2014). If a variable X_i is unimportant, then if noise is added to X_i, the ability to estimate the model does not change. However, if X_i is an important variable in the model, then adding noise to X_i will cause a reduction in predictive power.

More specifically, measurement error is forced into the model because $\tau < \infty$. When τ is set to a finite value, each estimate λ_i must also be finite. Therefore $1/\sigma^2_{Ui}$ is also finite, so $\sigma^2_{Ui} > 0$.

Measurement error variance is assigned to the p variables in such a way that the likelihood is least diminished. Therefore, the measurement error variance is assigned to variables that already have low sample correlations. Recall that a large measurement error variance σ^2_{Ui} corresponds to a value of λ_i that is close to zero.

$$(\hat{\lambda}_i = 0) \Rightarrow (\hat{\sigma}^2_{Ui} = \infty) \Rightarrow \frac{r_{ij}}{\sqrt{1+\hat{\sigma}^2_{Ui}}\sqrt{1+\hat{\sigma}^2_{Uj}}} = 0, \quad \forall j. \tag{2.2.24}$$

Therefore, a $\hat{\lambda}_i$ estimated to be close to zero implies that $\hat{\boldsymbol{\rho}}_{W,ij}$ is close to zero for all j, so X_i is estimated as uncorrelated with all other variables.

2.2.5.4 Estimation of Reliability Ratios

It is more convenient to consider reliability ratios RR instead of λ values. Reliability ratios are unit-free, scaled functions of λ that are constrained to be between zero and one. If λ values are ordered from smallest to largest, reliability ratios will retain the same order. The reliability ratio can be viewed as the proportion of variance in the data free from measurement error:

$$RR_j = \frac{\sigma^2_{Xj}}{\sigma^2_{Xj} + \sigma^2_{Uj}} = \frac{1}{1+\sigma^2_{Uj}} = \frac{1}{1+1/\lambda_j} = \frac{\lambda_j}{1+\lambda_j}. \tag{2.2.25}$$

A high reliability ratio indicates that a variable X_i was assigned a small amount of measurement error relative to other variables, and therefore X_i generally has higher correlations with other variables. The estimated reliability ratios can be calculated from the estimates $\hat{\boldsymbol{\lambda}}$, found from minimizing equation (2.2.22), as

$$\hat{RR}_j = \frac{\hat{\lambda}_j}{1+\hat{\lambda}_j}. \tag{2.2.26}$$

The reliability ratios are ordered largest to smallest such that $\hat{RR}_{(1)} \geq \hat{RR}_{(2)} \geq \cdots \geq \hat{RR}_{(p)}$. Each ordered reliability ratio corresponds to a variable. The variable X_i that corresponds to $\hat{RR}_{(1)}$ is set as Q_1, and so on. Then the MEMSEL method has given an ordering of the variables based on magnitude of correlations.

2.2.5.5 Choice of Tuning Parameter

The magnitude of τ controls the magnitude of the estimates $\hat{\boldsymbol{\lambda}}$, due to the constraint $\sum_j \lambda_j = \tau$, and therefore determines how much measurement error is added to the model. Rather than considering a grid of possible values for τ, we simply use $\tau = p$ due to its straightforward interpretation. Consider the case where all p variables are uncorrelated with one another. In this case, if $\tau = p$, we would expect the measurement error to be distributed evenly among the p variables, and for each X_i, $\hat{\lambda}_i = 1$ since the sum of the λ values is constrained to be τ. A value $\hat{\lambda}_i = 1$ corresponds to a reliability ratio $\hat{RR}_i = 1/2$. If some X_i values are correlated, then some reliability ratios will be estimated as greater

than 1/2 and some will be estimated as smaller. Further, as was discussed in Section 2.2.5.4, the actual values of $\hat{\boldsymbol{\lambda}}$ are not as important as the ordering of those values, making the magnitude of τ less important.

2.2.6 Types of Sample Correlation

Conveniently, each of the five ordering methods discussed utilize only the sample correlation matrix **R**, rather than **S**, in determining the vector of ordered variables, **Q**. However, we also considered variations on **R** to determine if it was possible to create a better distinction between correlated and uncorrelated variables. Each ordering method can be paired with each of four sample correlation matrices.

First, let $\mathbf{A} \circ \mathbf{B}$ define the elementwise multiplication of two matrices with elements $(\mathbf{A} \circ \mathbf{B})_{ij} = (A_{ij})(B_{ij})$ for $i, j = 1, \ldots, p$. This elementwise multiplication is sometimes referred to as the Hadamard, or Schur, product (Davis, 1962). Define the first alternative correlation matrix as $\mathbf{R}_2 = \mathbf{R} \circ \mathbf{R}$. Then define $\mathbf{R}_4$ as $\mathbf{R}_2^2$ after it is normalized back to a correlation matrix, resulting in $\mathbf{R}_4 = \mathbf{D}_{\mathbf{R}_2^2}^{-1/2} \mathbf{R}_2^2 \mathbf{D}_{\mathbf{R}_2^2}^{-1/2}$, where $\mathbf{D}_{\mathbf{R}_2^2}$ is defined as the diagonal matrix of variances from $\mathbf{R}_2^2$. Lastly, define $\mathbf{R}_8 = \mathbf{R}_4 \circ \mathbf{R}_4$. The idea is that these three alternative correlation matrices might provide better separation among variable correlations.

Each of these adaptations of **R** retains the property of being a correlation matrix. A correlation matrix is defined as any symmetric positive semidefinite matrix with each diagonal entry equal to one (Styan, 1973). Schur (1911) proves that the Hadamard product of any two positive semidefinite matrices is also positive semidefinite. Therefore, since **R** is positive semidefinite, $\mathbf{R}_2$ is also positive semidefinite. Since $\mathbf{R}_2$ also has diagonal elements of 1 and is symmetric, $\mathbf{R}_2$ is a correlation matrix.

Before proceeding, note that for any symmetric positive semidefinite matrix **A**, $\mathbf{AA} = \mathbf{A}^2$ is also positive semidefinite. The proof is as follows. Consider a matrix **A** that is symmetric and positive semidefinite. Then the symmetric square root of **A**, denoted $\mathbf{A}^{1/2}$, can be defined through spectral decomposition of **A** such that $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$ (Johnson and Wichern, 2007, p. 66). Then $\mathbf{A}^2 = \mathbf{A}^{1/2}\mathbf{A}\mathbf{A}^{1/2}$. Then for any vector $\mathbf{v}$ $(p \times 1)$, $\mathbf{v}^T(\mathbf{A}^{1/2}\mathbf{A}\mathbf{A}^{1/2})\mathbf{v} = (\mathbf{v}^T\mathbf{A}^{1/2})\mathbf{A}(\mathbf{A}^{1/2}\mathbf{v}) \geq 0$ since **A** is positive semidefinite. Therefore $\mathbf{A}^2$ is positive semidefinite.

Now $\mathbf{R}_4$ is the normalized version of $\mathbf{R}_2^2$. Any symmetric positive semidefinite matrix is a covariance matrix. We have already shown that $\mathbf{R}_2$ is positive semidefinite. Then $\mathbf{R}_2^2$ is also

positive semidefinite and is a covariance matrix. If $\mathbf{D}_{\mathbf{R}_2^2}$ defined as the diagonal matrix of variances from $\mathbf{R}_2^2$, then $\mathbf{R}_4 = \mathbf{D}_{\mathbf{R}_2^2}^{-1/2}\mathbf{R}_2^2\mathbf{D}_{\mathbf{R}_2^2}^{-1/2}$, a correlation matrix.

The proof that $\mathbf{R}_8$ is a correlation matrix follows from the proof that $\mathbf{R}_2$ is a correlation matrix. Since $\mathbf{R}_4$ is positive semidefinite, $\mathbf{R}_8$ is also positive semidefinite. Since $\mathbf{R}_8$ also has diagonal elements of 1 and is symmetric, $\mathbf{R}_8$ is a correlation matrix.

2.2.7 Simulation Study

Thus far we have presented five methods for ordering variables based on their correlations, as well as four different forms of a correlation matrix that can be combined with each ordering method.

A simulation study was conducted to compare the performance of these methods in terms of appropriately ordering variables based on some measure of correlation. Based on the known correct set of uncorrelated variables in the population, ROC curves were used to determine to what extent methods are able to appropriately order variables.

2.2.7.1 ROC Curves

The ROC curve is a plot of the True Positive Rate (TPR) on the vertical axis against the False Positive Rate (FPR) on the horizontal axis, at various thresholds (Fawcett, 2006). TPR is also called Sensitivity, and FPR is 1 - Specificity. For our work, True Positive is defined as appropriately estimating an uncorrelated variable as uncorrelated, and False Positive is inappropriately estimating a correlated variable as uncorrelated. For convenience, identifying a variable as uncorrelated will sometimes be referred to as "removing" the variable, since variables estimated as uncorrelated will be removed from the dataset before conducting factor analysis. At a given threshold, the TPR is therefore defined as

$$TPR = \frac{\#\text{ uncorrelated variables estimated as uncorrelated}}{\text{total }\#\text{ uncorrelated variables}}. \tag{2.2.27}$$

Similarly, FPR is defined as

$$FPR = \frac{\#\text{ correlated variables estimated as uncorrelated}}{\text{total }\#\text{ correlated variables}}. \tag{2.2.28}$$

The thresholds used are each step in the reverse ordering $\mathbf{Q}_{\text{rev}}$ of the p total variables, sorting

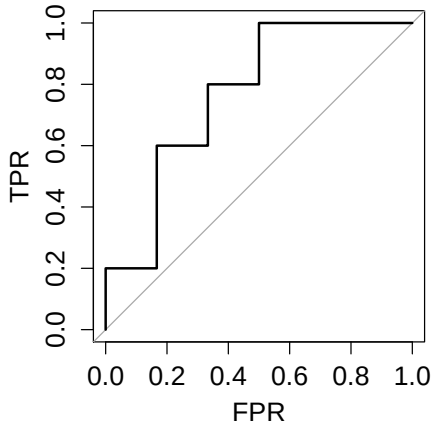


Figure 2.1 Example ROC curve for $m = 6$ and $u = 5$ using the MEMSEL ordering with **R**. AUC is 0.77. Diagonal line is given as reference.

from least correlated variable to most correlated, based on which ordering method is used. At the first step, each ordering method identifies a "lowly correlated" variable that should be removed. A binary classifier indicates whether or not this variable is truly correlated or uncorrelated in the population, and TPR and FPR are then calculated for the first step. The steps proceed through the orderings for each method, and the ROC curve is built. FPR and TPR therefore both begin at zero when no variables have been selected, and grow as the threshold increases, eventually both reaching one when all p variables have been identified for removal.

Note that the ROC curve in no way measures where the cutoff should be placed to determine u and m; rather, the ROC curve is an indicator of whether or not each ordering method appropriately groups uncorrelated variables at one end of **Q**, and correlated variables at the other end. An example ROC curve, created using the `pROC` package in `R`, can be seen in Figure 2.1 (Robin et al., 2011).

The Area Under the Curve (AUC) can be calculated from the ROC curve exactly as one would expect. An AUC of 1 indicates the ordering perfectly discriminates between correlated and uncorrelated variables, identifying all uncorrelated variables for removal from the model prior to identifying any correlated variables. In this case the ROC curve would move vertically to a TPR of 1 before moving horizontally at all. The AUC is used in our

simulation as the measure of how well each method orders variables based on correlation.

2.2.7.2 Simulation Design

A simulation study was used to compare the AUC for each ordering method in various circumstances. In an ideal ordering, all uncorrelated variables would be at the beginning of $\mathbf{Q}_{\text{rev}}$, followed by all correlated variables. The factors considered for the study were:

1. Number of correlated variables ($m = 6$, 12, or 40)
2. Number of uncorrelated variables ($u = 1$, 5, or 10)
3. Sample size (n=100 or 400)
4. Magnitude of factor loadings (controlled by $\alpha = .25$, .5, .75, or 1)
5. Structure of Correlation ($\mathbf{R}$, $\mathbf{R}_2$, $\mathbf{R}_4$ or $\mathbf{R}_8$)
6. Ordering Method (AAC, MAC, SMC, LR or MEMSEL)

All data was generated from the factor model. The three types of population $\mathbf{L}$ matrices with 6, 12, or 40 correlated variables each contained 2 factors. In the initial $\mathbf{L}$ matrices with 6 or 40 correlated variables, the non-zero factor loadings were all 0.9. The initial $\mathbf{L}$ matrix with 12 correlated variables was taken from Hirose and Konishi (2012) as an example with varying magnitudes of loadings within each factor, as well as lack of simple structure. Note that the design was not full-factorial; $m = 6$ correlated variables was paired with $u = 1$ or 5 uncorrelated variables, $m = 12$ was only paired with $u = 5$, and $m = 40$ was paired with $u = 1$ or 10 uncorrelated variables.

Noise was added to each factor model by multiplying the $\mathbf{L}$ matrix element-wise by a value of α =.25, .5, .75, or 1 to attenuate the loadings. Additional rows of zeros were added to include uncorrelated variables. The population covariance $\mathbf{\Sigma}$ was assumed to be standardized so that $\mathbf{\Psi}$ was completely determined by $\mathbf{L}$, since $\sigma^2_{X_i} = 1 = L^2_{i1} + \cdots + L^2_{ik} + \Psi_i$ when standardization is assumed. The population loadings matrices are denoted $\mathbf{L}_\alpha$ to indicate the magnitude of factor loadings. For instance, the rows of the loadings matrices defined by $m = 6$ correlated variables are given as:

$$\mathbf{L}_{.25} = \begin{pmatrix} .23 & 0 \\ .23 & 0 \\ .23 & 0 \\ 0 & .23 \\ 0 & .23 \\ 0 & .23 \end{pmatrix}, \quad \mathbf{L}_{.50} = \begin{pmatrix} .45 & 0 \\ .45 & 0 \\ .45 & 0 \\ 0 & .45 \\ 0 & .45 \\ 0 & .45 \end{pmatrix}, \quad \mathbf{L}_{.75} = \begin{pmatrix} .68 & 0 \\ .68 & 0 \\ .68 & 0 \\ 0 & .68 \\ 0 & .68 \\ 0 & .68 \end{pmatrix}, \quad \mathbf{L}_{1} = \begin{pmatrix} .90 & 0 \\ .90 & 0 \\ .90 & 0 \\ 0 & .90 \\ 0 & .90 \\ 0 & .90 \end{pmatrix}$$

Similarly, the $\mathbf{L}$ matrices with $m = 40$ correlated variables were defined by two factors, each with 20 non-zero loadings with magnitude determined by α.

The four types of population $\mathbf{L}$ matrices with $m = 12$ correlated variables were modeled after the simulation study in Hirose and Konishi (2012). They are given as:

$$\mathbf{L}_{.25} = \begin{pmatrix} .24 & 0 & 0 \\ 0 & .24 & 0 \\ 0 & 0 & .24 \\ .19 & 0 & 0 \\ 0 & .19 & 0 \\ 0 & 0 & .19 \\ .18 & .13 & .13 \\ .13 & .18 & .13 \\ .13 & .13 & .18 \\ .23 & .08 & .08 \\ .08 & .23 & .08 \\ .08 & .08 & .23 \end{pmatrix}, \quad \mathbf{L}_{.50} = \begin{pmatrix} .48 & 0 & 0 \\ 0 & .48 & 0 \\ 0 & 0 & .48 \\ .38 & 0 & 0 \\ 0 & .38 & 0 \\ 0 & 0 & .38 \\ .35 & .25 & .25 \\ .25 & .35 & .25 \\ .25 & .25 & .35 \\ .45 & .15 & .15 \\ .15 & .45 & .15 \\ .15 & .15 & .45 \end{pmatrix}$$

$$\mathbf{L}_{.75} = \begin{pmatrix} .71 & 0 & 0 \\ 0 & .71 & 0 \\ 0 & 0 & .71 \\ .56 & 0 & 0 \\ 0 & .56 & 0 \\ 0 & 0 & .56 \\ .53 & .38 & .38 \\ .38 & .53 & .38 \\ .38 & .38 & .53 \\ .68 & .23 & .23 \\ .23 & .68 & .23 \\ .23 & .23 & .68 \end{pmatrix}, \quad \mathbf{L}_{1} = \begin{pmatrix} .95 & 0 & 0 \\ 0 & .95 & 0 \\ 0 & 0 & .95 \\ .75 & 0 & 0 \\ 0 & .75 & 0 \\ 0 & 0 & .75 \\ .70 & .50 & .50 \\ .50 & .70 & .50 \\ .50 & .50 & .70 \\ .90 & .30 & .30 \\ .30 & .90 & .30 \\ .30 & .30 & .90 \end{pmatrix}$$

As an aside, recall that in factor analysis, a variable is traditionally considered to be uncorrelated if it has all factor loadings below a certain cutoff, such as .3 or .4. Therefore, for a couple of these population loadings matrices, the magnitude of factor loadings would generally be considered very low.

Uncorrelated variables (either $u = 1$, 5, or 10) were created by adding extra rows to the $\mathbf{L}$ matrices with all loadings equal to zero, giving a total of $p = m + u$ variables ranging from $p = 7$ to $p = 50$. The population covariance matrix was defined by the $(p \times k)$ loadings matrix as $\mathbf{L}\mathbf{L}^T + \mathbf{D}_{\boldsymbol{\Psi}}$, which due to the standardization of $\mathbf{L}$ and $\boldsymbol{\Psi}$, is also equal to the population correlation matrix $\boldsymbol{\rho}$. Without loss of generality, population means were assumed to be zero. For each population $\boldsymbol{\Sigma}$ matrix and sample size, five hundred datasets (denoted by $N = 500$) were generated as $\mathbf{X}_j \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ for $j = 1, \ldots, n$. The same seed was used for each combination of m, u, n, and α, although of course samples differed due to different $\boldsymbol{\Sigma}$ matrices. This has a blocking effect on the data, allowing simulation results to showcase differences due to $\boldsymbol{\Sigma}$, rather than seeing random differences in results dependent on the seed used.

Each ordering method was used in conjunction with each of the four correlation matrices from each replication. An AUC value was calculated in each case.

2.2.7.3 Results

The following tables show the average AUC for each ordering method in each combination of factors. In each combination of factors, the ordering method with highest AUC is given in bold. However, this highest AUC may not be significantly different from the next highest AUC. To assist in comparisons, an average standard error is given at the bottom of each column. Standard error is calculated as

$$SE(AUC) = \frac{s}{\sqrt{N}}, \tag{2.2.29}$$

where s is the standard deviation of the AUC value for $N = 500$ replications for each ordering method in each combination of factors. Average standard error is therefore the average among the 20 standard errors given for all methods at each combination of factors.

These tables also include results from paired t-tests that compared each of the four correlation methods within an ordering method. Each p-value compares the ordering method average AUC for one type of correlation matrix to the correlation matrix of one step lower complexity. For example, a p-value of .03 for ordering method MEMSEL on correlation $\mathbf{R}_2$ indicates that on average, the $\mathbf{R}_2$ matrix AUC for MEMSEL was significantly higher than the $\mathbf{R}$ matrix AUC for MEMSEL. Similarly, a p-value of .10 for MEMSEL AUC using $\mathbf{R}_4$ would indicate that on average, $\mathbf{R}_4$ did not result in a significantly higher MEMSEL AUC when compared to $\mathbf{R}_2$. Again, these p-values are not comparing ordering methods to each other, but rather are comparing types of correlation matrices to one another in a stepwise fashion. In the tables, rather than reporting the actual p-value, a * is used if the corresponding p-value was less than .05.

2.2.8 Conclusions

From these simulation results, we see that for high enough loadings in the population $\mathbf{L}$ matrix (such as when $\alpha = 1$), each method correctly orders all uncorrelated variables first in $\mathbf{Q}_{\text{rev}}$. Loadings that are slightly lower ($\alpha = .75$) can still produce correct orderings if the sample size is larger ($n = 400$).

A summary was created to focus more closely on situations where ordering methods and alternative correlations showed more differences. Table 2.6 displays a summary of average

Table 2.1 Average AUC for $m = 6$, $u = 1$ variables for $N = 500$ replications. * indicates p-value $< .05$. Bold values are highest AUC in the given category. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation. See page 45 for **L** definitions.

Corr	Ordering	$n = 100$				$n = 400$			
	Method	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
R	MEMSEL	0.788	0.882	1.000	1	0.802	**0.998**	1	1
	LR	0.785	0.883	0.999	1	0.802	0.998	1	1
	SMC	0.820	0.865	0.999	1	0.799	0.997	1	1
	MAC	**0.824**	0.878	**1.000**	1	**0.809**	0.998	1	1
	AAC	0.786	**0.892**	0.999	1	0.806	**0.998**	1	1
$\mathbf{R}_2$	MEMSEL	0.790	0.886	1.000	1	0.793	0.999	1	1
	LR	0.789	0.886	1.000 *	1	0.796	0.999 *	1	1
	SMC	0.820	0.885 *	1.000 *	1	0.801	0.999 *	1	1
	MAC	**0.824**	0.878	1.000	1	**0.809**	0.998	1	1
	AAC	0.785	**0.896**	1.000	1	0.799	**0.999**	1	1
$\mathbf{R}_4$	MEMSEL	0.789	0.885	1.000	1	0.792	0.999	1	1
	LR	0.789	0.887	1.000	1	0.796	0.999	1	1
	SMC	**0.821**	0.885	1.000	1	0.801	0.999	1	1
	MAC	0.815	0.885	1.000	1	**0.805**	0.998	1	1
	AAC	0.785	**0.895**	1.000	1	0.799	**0.999**	1	1
$\mathbf{R}_8$	MEMSEL	0.791	0.874	1.000	1	0.791	0.994	1	1
	LR	0.790	0.884	1.000	1	0.789	0.998	1	1
	SMC	**0.821**	**0.887**	1.000	1	**0.805**	0.998	1	1
	MAC	0.815	0.885	1.000	1	**0.805**	0.998	1	1
	AAC	0.789	0.885	1.000	1	0.796	**0.999**	1	1
Av SE		0.008	0.007	0.000	0	0.008	0.001	0	0

Table 2.2 Average AUC for $m = 6$, $u = 5$ variables for $N = 500$ replications. * indicates p-value $< .05$. Bold values are highest AUC in the given category. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation. See page 45 for **L** definitions.

Corr	Ordering	$n = 100$				$n = 400$			
	Method	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
R	MEMSEL	0.637	0.824	0.999	1	**0.683**	0.997	1	1
	LR	0.639	0.823	0.999	1	0.683	0.997	1	1
	SMC	**0.646**	0.827	1.000	1	**0.683**	**0.997**	1	1
	MAC	0.644	0.824	**1.000**	1	0.676	**0.997**	1	1
	AAC	0.640	0.812	0.996	1	0.674	0.987	1	1
$\mathbf{R}_2$	MEMSEL	**0.648** *	0.849 *	1.000 *	1	0.690 *	**0.999** *	1	1
	LR	**0.648** *	0.849 *	1.000 *	1	**0.690** *	0.998 *	1	1
	SMC	0.647	0.844 *	1.000	1	0.682	**0.999** *	1	1
	MAC	0.644	0.824	1.000	1	0.676	0.997	1	1
	AAC	0.642	0.847 *	1.000 *	1	0.683 *	0.998 *	1	1
$\mathbf{R}_4$	MEMSEL	**0.648**	**0.850**	1.000	1	0.690	**0.999**	1	1
	LR	**0.648**	0.849	1.000	1	**0.690**	0.998	1	1
	SMC	0.647	0.845	1.000	1	0.682	**0.999**	1	1
	MAC	0.644	0.829	1.000	1	0.678 *	0.998	1	1
	AAC	0.642	0.848	1.000	1	0.684	0.998	1	1
$\mathbf{R}_8$	MEMSEL	0.646	0.844	1.000	1	0.685	0.997	1	1
	LR	0.647	0.843	1.000	1	0.684	0.998	1	1
	SMC	0.645	0.839	1.000	1	0.682	0.998	1	1
	MAC	0.644	0.829	1.000	1	0.678	0.998	1	1
	AAC	**0.648**	**0.853** *	1.000	1	**0.690** *	**0.998**	1	1
Av SE		0.005	0.006	0.000	0	0.006	0.001	0	0

Table 2.3 Average AUC for $m = 12$, $u = 5$ variables for $N = 500$ replications. * indicates p-value $< .05$. Bold values are highest AUC in the given category. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation. See page 45 for **L** definitions.

Corr	Ordering	$n = 100$				$n = 400$			
	Method	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
R	MEMSEL	0.614	0.848	0.997	1	0.692 *	0.995	1	1
	LR	0.614	0.845	0.995	1	0.693 *	0.995	1	1
	SMC	**0.622**	0.854	0.997	1	0.694 *	0.997	1	1
	MAC	0.620	0.865	**0.998**	1	0.677	0.996	1	1
	AAC	0.614	**0.901**	0.996	1	**0.707** *	0.996	1	1
$\mathbf{R}_2$	MEMSEL	0.619	0.898 *	**0.999** *	1	0.708	**0.999** *	1	1
	LR	0.619	0.898 *	0.999 *	1	0.707	0.998 *	1	1
	SMC	**0.626**	0.896 *	0.999 *	1	0.703	0.999 *	1	1
	MAC	0.620	0.865	0.998	1	0.677	0.996	1	1
	AAC	0.622 *	**0.911** *	0.998 *	1	**0.717**	0.998 *	1	1
$\mathbf{R}_4$	MEMSEL	0.619	0.898 *	**0.999**	1	0.708	**0.999**	1	1
	LR	0.619	0.898	0.999	1	0.707	0.998	1	1
	SMC	**0.626**	0.897 *	0.999	1	0.703	0.999	1	1
	MAC	0.621	0.872 *	0.998 *	1	0.681 *	0.996 *	1	1
	AAC	0.623	**0.913** *	0.998	1	**0.718** *	0.999	1	1
$\mathbf{R}_8$	MEMSEL	0.621	0.892	**0.999**	1	0.699	0.998	1	1
	LR	0.621	0.891	0.999	1	0.698	0.998	1	1
	SMC	**0.625**	0.889	0.999	1	0.696	0.998	1	1
	MAC	0.621	0.872	0.998	1	0.681	0.996	1	1
	AAC	0.621	**0.907**	0.999 *	1	**0.710**	**0.999**	1	1
Av SE		0.005	0.004	0.000	0	0.006	0.000	0	0

Table 2.4 Average AUC for $m = 40$, $u = 1$ variables for $N = 500$ replications. * indicates p-value $< .05$. Bold values are highest AUC in the given category. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation. See page 45 for **L** definitions.

Corr	Ordering	$n = 100$				$n = 400$			
	Method	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
R	MEMSEL	0.749	0.890	1.000 *	1	0.821	0.999	1	1
	LR	0.759	0.869	0.997 *	1	0.823	1.000	1	1
	SMC	0.761	0.878	0.997 *	1	0.817	0.999	1	1
	MAC	**0.766**	0.951	**1.000**	1	0.830	1.000	1	1
	AAC	0.749	**0.988**	**1.000**	1	**0.880**	**1.000**	1	1
$\mathbf{R}_2$	MEMSEL	0.768 *	0.981 *	1.000	1	0.872 *	1.000 *	1	1
	LR	0.767	0.981 *	1.000	1	0.872 *	1.000 *	1	1
	SMC	**0.770**	0.979 *	1.000	1	0.870 *	1.000 *	1	1
	MAC	0.766	0.951	1.000	1	0.830	1.000	1	1
	AAC	0.757 *	**0.990** *	1.000	1	**0.888** *	1.000	1	1
$\mathbf{R}_4$	MEMSEL	**0.768**	0.981	1.000	1	0.872	1.000	1	1
	LR	0.767	0.982 *	1.000	1	0.872	1.000	1	1
	SMC	0.768	0.980 *	1.000	1	0.870	1.000	1	1
	MAC	0.766	0.962 *	1.000	1	0.831	1.000	1	1
	AAC	0.758	**0.989**	1.000	1	**0.888**	1.000	1	1
$\mathbf{R}_8$	MEMSEL	0.767	0.978	1.000	1	0.851	1.000	1	1
	LR	0.767	0.977	1.000	1	0.851	1.000	1	1
	SMC	**0.771**	0.977	1.000	1	0.852	1.000	1	1
	MAC	0.766	0.962	1.000	1	0.831	1.000	1	1
	AAC	0.767 *	**0.988**	1.000	1	**0.876**	1.000	1	1
Av SE		0.007	0.003	0.000	0	0.006	0.000	0	0

Table 2.5 Average AUC for $m = 40$, $u = 10$ variables for $N = 500$ replications. * indicates p-value $< .05$. Bold values are highest AUC in the given category. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation. See page 45 for **L** definitions.

Corr	Ordering	$n = 100$				$n = 400$			
	Method	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
R	MEMSEL	0.592	0.842	0.998	1	0.733	0.999	1	1
	LR	0.587	0.802	0.991	1	0.739	1.000	1	1
	SMC	0.588	0.835	0.996	1	0.744	1.000	1	1
	MAC	0.591	0.944	**1.000**	1	0.738	**1.000**	1	1
	AAC	**0.619**	**0.986**	**1.000**	1	**0.846**	**1.000**	1	1
$\mathbf{R}_2$	MEMSEL	0.616 *	0.980 *	1.000 *	1	0.831 *	1.000 *	1	1
	LR	0.615 *	0.980 *	1.000 *	1	0.831 *	1.000 *	1	1
	SMC	0.608 *	0.979 *	1.000 *	1	0.817 *	1.000 *	1	1
	MAC	0.591	0.944	1.000	1	0.738	1.000	1	1
	AAC	**0.624** *	**0.988** *	1.000	1	**0.858** *	1.000	1	1
$\mathbf{R}_4$	MEMSEL	0.614	0.980 *	1.000	1	0.831	1.000	1	1
	LR	0.616	0.981 *	1.000	1	0.831 *	1.000	1	1
	SMC	0.609	0.981 *	1.000	1	0.818 *	1.000	1	1
	MAC	0.592 *	0.960 *	1.000	1	0.742 *	1.000	1	1
	AAC	**0.624**	**0.988**	1.000	1	**0.859** *	1.000	1	1
$\mathbf{R}_8$	MEMSEL	0.606	0.979	1.000	1	0.789	1.000	1	1
	LR	0.606	0.978	1.000	1	0.789	1.000	1	1
	SMC	0.603	0.977	1.000	1	0.777	1.000	1	1
	MAC	0.592	0.960	1.000	1	0.742	1.000	1	1
	AAC	**0.621**	**0.988**	1.000	1	**0.838**	1.000	1	1
Av SE		0.004	0.002	0.000	0	0.004	0.000	0	0

Table 2.6 Summary of Average AUC, averaged across AUC values for $\mathbf{L}_{.25}$ when $n = 100$, $\mathbf{L}_{.50}$ when $n = 100$, and $\mathbf{L}_{.25}$ when $n = 400$. MEMSEL = Measurement Error Model Selection Likelihood, LR = Likelihood Ratio, SMC = Squared Multiple Correlation, MAC = Maximum Absolute Correlation, AAC = Average Absolute Correlation

Ordering Method	Corr	$m = 6$ $u = 1$	$m = 6$ $u = 5$	$m = 12$ $u = 5$	$m = 40$ $u = 1$	$m = 40$ $u = 10$
AAC	$\mathbf{R}$	.83	.71	.74	.87	.82
	$\mathbf{R}_2$	.83	.72	.75	.88	.82
	$\mathbf{R}_4$	.83	.72	.75	.88	.82
LR	$\mathbf{R}$	.82	.71	.72	.82	.71
	$\mathbf{R}_2$	.82	.73	.74	.87	.81
	$\mathbf{R}_4$	.82	.73	.74	.87	.81
MAC	$\mathbf{R}$	.84	.71	.72	.85	.76
	$\mathbf{R}_2$	.84	.71	.72	.85	.76
	$\mathbf{R}_4$	.83	.72	.72	.85	.76
MEMSEL	$\mathbf{R}$	.82	.71	.72	.82	.72
	$\mathbf{R}_2$	.82	.73	.74	.87	.81
	$\mathbf{R}_4$	.82	.73	.74	.87	.81
SMC	$\mathbf{R}$	.83	.72	.72	.82	.72
	$\mathbf{R}_2$	.84	.72	.74	.87	.80
	$\mathbf{R}_4$	.84	.72	.74	.87	.80

AUC values for each combination of population correlation matrix, ordering method and number of uncorrelated/correlated variables. Each value in Table 2.6 is an average of 3 AUC values from Table 2.1 through Table 2.5; $\mathbf{L}_{.25}$ when $n = 100$, $\mathbf{L}_{.50}$ when $n = 100$, and $\mathbf{L}_{.25}$ when $n = 400$. These particular values are included in the summary since they are where the greatest differences are evident in AUC across the combinations of factors; the other combinations had average AUC values very close to, or equal to, 1.

As expected, in general the AUC is lower when there are more uncorrelated variables present. More specific conclusions regarding alternative correlation matrices are as follows:

1. Viewing the four correlation types sequentially, $\mathbf{R}_2$ appears to provide the most significant improvement over the simpler correlation structure $\mathbf{R}$. $\mathbf{R}_4$ is a significant improvement of $\mathbf{R}_2$ for $\mathbf{L}_{.50}$ when $n = 100$, and in some other sporadic situations. However, these statistically significant results may not be practically significant.

2. $\mathbf{R}_8$ does not appear to offer any improvement over $\mathbf{R}_4$. Therefore, it will not be considered further.

3. If desired, then, $\mathbf{R}_2$ or $\mathbf{R}_4$ seem to be the best alternatives to $\mathbf{R}$. $\mathbf{R}_2$ is more easily interpretable which might be reason enough to prefer it over $\mathbf{R}_4$.

4. Table 2.6 confirms the overall trend, across ordering method, that $\mathbf{R}_2$ offers a small improvement over $\mathbf{R}$. However, these differences may not be large enough to be practically significant.

Specific conclusions regarding ordering methods are as follows:

1. No ordering method stands out as clearly superior to all others. Methods that give the largest average AUC in at least one combination of factors are MAC, AAC, SMC, and MEMSEL. LR ties for first place in a couple cases. However, many of these differences are not statistically significant.

2. AAC gives the highest (or tied for highest) AUC in 44 of the cases given in the tables. SMC gives the highest average AUC in 14 cases. MAC gives the highest AUC in 13 cases. MEMSEL also gives the highest AUC in 13 cases. LR method is the highest in 4 cases. Again, note that these cases are most definitely not all statistically significant differences between ordering methods.

3. Table 2.6 confirms that various ordering methods produce the highest average AUC in various cases. For example, MAC and SMC are the most accurate ordering methods for $m = 6$ correlated and $u = 1$ uncorrelated variables. However, AAC appears to be the most accurate ordering method for $m = 40$ variables, regardless of u.

4. LR and MEMSEL results are almost identical across different values of m and u in Table 2.6. SMC is better than MEMSEL in some cases but not in others. MAC performs poorly compared to other ordering methods when there is a higher number of variables.

5. On average, as seen in Table 2.6, AAC ordering appears to consistently provide slightly higher AUC values than other ordering methods. In fact, Table 2.5 shows that AAC

was the only ordering method to give the highest average AUC for any combination of factors, except when tied at AUC=1 with other methods.

In general, these simulation results show that each ordering method has its merits, with no particular method standing out as clearly superior at ordering variables based on correlation. However, the Average Absolute Correlation does appear to be slightly more consistent across various combinations of factors. Although $\mathbf{R}_2$ and $\mathbf{R}_4$ may offer marginal improvements compared to the use of $\mathbf{R}$ for ordering, the improvements may not be large enough to prove useful in a practical sense. At the very least, $\mathbf{R}_2$ is preferred over $\mathbf{R}_4$ for the ease of interpretation.

2.3 Stopping Criteria for Classifying Variables

Thus far, several ordering methods for identifying uncorrelated variables have been compared, including MEMSEL, SMC, LR, AAC, and MAC. After creating these full orderings of all p variables, we must use a stopping criterion to determine from the ordering where to place a cutoff to break the variables into groups that will be estimated as correlated (those listed first in $\mathbf{Q}$) versus uncorrelated. Several methods are proposed; BIC, AIC, and three forms of a likelihood ratio test (LRT). A simulation study is used to compare these stopping criteria combined with each ordering method, for several combinations of factors.

Returning to our original purposes, the u variables that are estimated as uncorrelated will be screened out, or removed from the dataset, prior to continuing on with factor analysis. The determination of appropriate values m and u can be completed with a backward-elimination approach to variable selection. The approach is considered backward elimination, rather than forward selection, because each stopping criterion method begins by assuming all p variables are included in the model and proceeds with setting the correlations to zero for one X_i at a time. This process creates a sequence of nested models that can be compared to find the model that best balances the number of parameters to estimate with the model fit, assessed by BIC, AIC, or several versions of a LRT.

Each stopping criterion begins with the unconstrained sample correlation matrix $\mathbf{R}$. The rows and columns in $\mathbf{R}$ are ordered according the vector $\mathbf{Q}$ for a given ordering method. This ordered sample correlation matrix can be referred to as $\hat{\rho}_{-0}$, since no variables are

estimated as uncorrelated and $u = 0$. The first row and column in $\hat{\boldsymbol{\rho}}_{-0}$ represent correlations for variable Q_1, designated as the most highly correlated variable. Then row and column p in $\hat{\boldsymbol{\rho}}_{-0}$ give correlations for Q_p, the least correlated variable according to one of the five ordering methods.

2.3.1 AIC and BIC

BIC and AIC are generally defined as

$$\text{BIC}=-2l(\hat{\boldsymbol{\Theta}})+q\log(n) \quad \text{and} \tag{2.3.1}$$

$$\text{AIC}=-2l(\hat{\boldsymbol{\Theta}})+2q, \tag{2.3.2}$$

where $l(\hat{\boldsymbol{\Theta}})$ is the maximized log likelihood for a set of parameters $\boldsymbol{\Theta}$, n is the sample size and q is the number of parameters being estimated (Akaike, 1973; Schwarz, 1978). For our scenario, when u variables are constrained to be uncorrelated, the maximized log likelihood is defined in Equation (2.1.18) as $l(\hat{\boldsymbol{\Sigma}}_{-u}) = (-n/2)\big[p\log(2\pi)+\log|\hat{\boldsymbol{\Sigma}}_{-u}|+p\big]$. The number of parameters estimated in $\hat{\boldsymbol{\Sigma}}_{-u}$ is $q = m(m-1)/2+p$ for the p variances and $m(m-1)/2$ unique covariances. Then BIC is equivalent to

$$\begin{aligned}
\text{BIC} &= -2l(\hat{\boldsymbol{\Sigma}}_{-u})+\left[\frac{m(m-1)}{2}+p\right]\log(n) && (2.3.3)\\
&= n\big[p\log(2\pi)+\log|\hat{\boldsymbol{\Sigma}}_{-u}|+p\big]+\left[\frac{m(m-1)}{2}+p\right]\log(n) && (2.3.4)\\
&= n\big[p\log(2\pi)+\log|\mathbf{D_S}|+\log|\hat{\boldsymbol{\rho}}_{-u}|+p\big]+\left[\frac{m(m-1)}{2}+p\right]\log(n) && (2.3.5)\\
&= (n)\log|\hat{\boldsymbol{\rho}}_{-u}|+\left[\frac{m(m-1)}{2}\right]\log(n)+C_{\text{BIC}}, && (2.3.6)
\end{aligned}$$

where $C_{\text{BIC}} = n\big[p\log(2\pi)+\log|\mathbf{D_S}|+p\big]+(p)\log(n)$. As BIC is calculated for nested models where additional variables have correlations constrained to zero, the constant C_{BIC} does not change and is therefore excluded from the calculation. Then BIC can be calculated solely from the correlation matrix $\mathbf{R}$, since $\hat{\boldsymbol{\rho}}_{-u}$ was defined in Section 2.1.4 as the sample correlation matrix $\mathbf{R}$ with correlations in the last u rows and columns set to zero, corresponding to the u variables that are estimated as uncorrelated based on some ordering.

Similarly, AIC is calculated as

$$\begin{aligned} \text{AIC} &= -2l(\hat{\boldsymbol{\Sigma}}_{-u}) + 2\left[\frac{m(m-1)}{2} + p\right] && (2.3.7) \\ &= (n)\log|\hat{\boldsymbol{\rho}}_{-u}| + m(m-1) + C_{\text{AIC}}, && (2.3.8) \end{aligned}$$

where $C_{\text{AIC}} = n\left[p\log(2\pi) + \log|\mathbf{D_S}| + p\right] + 2p$. Once more, in calculations the constant C_{AIC} is ignored and therefore AIC can be calculated using only the sample correlation matrix $\mathbf{R}$.

2.3.1.1 Algorithm

As methods for stopping criteria, AIC and BIC are calculated for each possible value of u, defined by $t = 0, \ldots, p$. At $t = 0$, all variables are allowed to freely correlate and, for example, $\text{BIC} = (n)\log|\hat{\boldsymbol{\rho}}_{-0}| + [p(p-1)/2]\log(n)$ where $m = p$. Sequences of BIC and AIC values are calculated as additional variables are set as uncorrelated, starting with Q_p. For $\mathbf{Q}$ from a given ordering method, the algorithm for BIC is as follows.

1. Use one of the ordering methods to determine $\mathbf{Q}$.
2. Set $t = 0$. Reorder the sample correlation matrix according to $\mathbf{Q}$, giving $\hat{\boldsymbol{\rho}}_{-0}$.
3. Define a $(p+1) \times 1$ vector $\mathbf{V}_{\text{BIC}}$ to store BIC for value for each $t = 0, \ldots, p$.
4. Calculate $\text{BIC}_0 = (n)\log|\hat{\boldsymbol{\rho}}_{-0}| + [p(p-1)/2]\log(n)$ and store as $V_{\text{BIC},1}$.
5. Set $t = 1$. Then $m = p-1$. Constrain the correlations in the pth row and column of $\hat{\boldsymbol{\rho}}_{-0}$ to be zero, giving $\hat{\boldsymbol{\rho}}_{-1}$.
6. Calculate $\text{BIC}_1 = (n)\log|\hat{\boldsymbol{\rho}}_{-1}| + [(p-1)(p-2)/2]\log(n)$ and store as $V_{\text{BIC},2}$.
7. Set $t = 2$. Then $m = p-2$. Constrain the correlations in the last two rows and columns of $\hat{\boldsymbol{\rho}}_{-0}$ to be zero, giving $\hat{\boldsymbol{\rho}}_{-2}$.
8. Calculate $\text{BIC}_2 = (n)\log|\hat{\boldsymbol{\rho}}_{-2}| + [(p-2)(p-3)/2]\log(n)$ and store as $V_{\text{BIC},3}$.
9. Repeat the process iteratively for $t = 3, \ldots, p$, calculating $\hat{\boldsymbol{\rho}}_{-t}$ and the corresponding BIC_t, and storing the result in $V_{\text{BIC},(t+1)}$

AIC is calculated similarly, with 2 in place of $\log(n)$ in calculations. Note that the last two values in the BIC or AIC sequence will be identical, because $\hat{\boldsymbol{\rho}}_{-p} = \hat{\boldsymbol{\rho}}_{-(p-1)} = \mathbf{I}_p$, where $\mathbf{I}_p$ is a $(p \times p)$ identity matrix. It is not possible to have just one correlated variable; a minimum of two variables must be unconstrained for there to be non-zero correlations.

When the full sequence of $(p+1)$ BIC or AIC values has been calculated, identify the minimum BIC or AIC. Say the minimum BIC is found to be $V_{\text{BIC},i}$. Then $u = i-1$ is determined to be the appropriate number of variables to set as uncorrelated, given the ordering method used. In other words, the AIC or BIC method has then identified a set of m variables that are correlated, and has identified u variables that can be removed from the data prior to running factor analysis.

Based on how BIC and AIC are calculated, BIC adds a larger penalty when more parameters are estimated. This means that when a higher number of variables are allowed to correlate, the BIC calculation will have a larger penalty $q\log(n)$ than the AIC penalty, $2q$. Therefore, the BIC tends to have a minimum when more variables are constrained to be uncorrelated, in comparison to the AIC. This will be seen in the simulation study in Section 2.3.3 as well.

2.3.2 Likelihood Ratio Test with Bartlett Correction

The likelihood ratio was first introduced as an ordering method in Section 2.2.4. However, the likelihood ratio test (LRT) can be formalized and further used as a stopping criterion rather than just an ordering method. The LRT as a stopping criterion can be used in conjunction with any of the ordering methods discussed in Section 2.2. After employing a Bartlett correction, the likelihood ratio statistic is compared to a critical value from a χ^2 distribution to determine if the constrained correlation matrix is significantly different than the less constrained version.

We define three possible likelihood ratio tests. The Full LRT always compares $\hat{\boldsymbol{\Sigma}}_{-u}$ to the full ordered sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{-0}$. The Sequential LRT compares $\hat{\boldsymbol{\Sigma}}_{-u}$ to the previous step with one less set of constrained correlations, $\hat{\boldsymbol{\Sigma}}_{-(u-1)}$. The False Selection Rate (FSR) LRT is an adaptation of the Sequential LRT that attempts to control the rate of falsely including uncorrelated variables. Like BIC and AIC, each of these likelihood ratio tests can be simplified to only require the use of $\mathbf{R}$ instead of $\mathbf{S}$.

2.3.2.1 Sequential LRT

As in Section 2.2.4, consider the hypothesis that t variables are uncorrelated with all other variables, compared to the previous step where only $(t-1)$ variables were uncorrelated. Then

$H_0 : \boldsymbol{\Sigma}_{-(t-1)} = \boldsymbol{\Sigma}_{-t}$

$H_a : \boldsymbol{\Sigma}_{-(t-1)} \neq \boldsymbol{\Sigma}_{-t}$.

Note that $m = p - t$. Then the likelihood ratio statistic is defined as

$$\begin{aligned} \text{LR}_{\text{SEQ},-t} &= -2\log\left(\frac{\text{L}(\hat{\boldsymbol{\Sigma}}_{-t})}{\text{L}(\hat{\boldsymbol{\Sigma}}_{-(t-1)})}\right) && (2.3.9)\\ &= n\left[\log|\hat{\boldsymbol{\rho}}_{-t}| - \log|\hat{\boldsymbol{\rho}}_{-(t-1)}|\right] && (2.3.10)\\ &= n\left[\log|\mathbf{R}_{mm}| + \log|\mathbf{I}_{tt}| - \log|\mathbf{R}_{(m+1)(m+1)}| - \log|\mathbf{I}_{(t-1)(t-1)}|\right] && (2.3.11)\\ &= n\left[\log|\mathbf{R}_{mm}| - \log|\mathbf{R}_{(m+1)(m+1)}|\right], && (2.3.12) \end{aligned}$$

since the determinant of an identity matrix is 1, and $\log(1) = 0$. Therefore the sequential likelihood ratio test at t only requires the correlation matrix of the $m+1$ most highly correlated variables, rather than all p variables. That is, for any t, the Sequential LRT tests

$$H_0 : \boldsymbol{\rho}_{(m+1)(m+1)} = \begin{pmatrix} \boldsymbol{\rho}_{mm} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}. \quad (2.3.13)$$

Bartlett Correction

Keeping the likelihood ratio in the form of Equation (2.3.12) is appropriate for ordering variables. However, for a likelihood ratio test as stopping criteria, we must go one step further and compare this statistic to the critical value from the χ^2 distribution with degrees of freedom $(m+1)(m)/2-(m)(m-1)/2 = m = p-t$, representing the m unique correlations that are constricted to be zero. The form of the likelihood in Equation (2.3.12), however, is not well-approximated by the χ^2 distribution. In fact, Appendix A.3 shows that the uncorrected likelihood ratio test leads to greatly inflated Type 1 error rates, which should be around .05 if the significance level is .05.

Therefore, a Bartlett correction is employed in order to improve the χ^2 approximation

to the sampling distribution of $\mathrm{LR}_{\mathrm{SEQ},-t}$ (Johnson and Wichern, 2007, p. 502).

Rencher and Christensen (2012, pp. 271-272) present the Bartlett-corrected likelihood ratio test for testing that a covariance matrix is block diagonal with b blocks and p_j variables in block j. The corresponding likelihood ratio statistic is

$$\mathrm{LR}_{\mathrm{B}} = \left(\frac{n-1-2a_3+3a_2}{6a_2} \right) \left(\sum_{j=1}^{b} \log|\mathbf{R}_j| - \log|\mathbf{R}| \right), \tag{2.3.14}$$

where $a_2 = p^2 - \sum_{j=1}^{b} p_j^2$, $a_3 = p^3 - \sum_{j=1}^{b} p_j^3$, and $\mathbf{R}_j$ is the correlation matrix of the jth block (StataCorp, 2013).

At any given step, the Sequential LRT tests that $\boldsymbol{\rho}_{(m+1)(m+1)}$ is a block diagonal matrix, with one (1×1) block for the variable corresponding to Q_{m+1}, and one $(m \times m)$ block with correlations of the m correlated variables. Therefore we can apply the block-diagonal likelihood ratio statistic formula, with $b = 2$ blocks with $p = m+1$, $p_1 = m$ and $p_2 = 1$. Then the Bartlett-corrected sequential likelihood ratio test statistic for testing t uncorrelated variables is

$$\mathrm{LR}_{\mathrm{SEQ,-t}} = \left(\frac{n-1-2a_3+3a_2}{6a_2} \right) \left(\log|\mathbf{R}_{mm}| - \log|\mathbf{R}_{(m+1)(m+1)}| \right), \tag{2.3.15}$$

where $a_2 = (m+1)^2 - m^2 - 1 = 2m$ and $a_3 = (m+1)^3 - m^3 - 1 = 3m^2 + 3m$. The null hypothesis can now appropriately be rejected if $\mathrm{LR}_{\mathrm{SEQ,-t}} > \chi^2_{\alpha^*,m}$ for a given significance level α^* and degrees of freedom m.

Algorithm

Using $\alpha^* = .05$, the algorithm for using the Sequential likelihood ratio test as a stopping criterion is as follows:

1. Use one of the ordering methods to determine $\mathbf{Q}$.
2. Set $t = 0$. Reorder the sample correlation matrix according to $\mathbf{Q}$, giving $\hat{\boldsymbol{\rho}}_{-0}$.
3. Define a $(p+1) \times 1$ vector $\mathbf{V}_{\mathrm{SEQ}}$ to store the p-value corresponding to $\mathrm{LRT}_{\mathrm{SEQ}}$ for each $t = 0, \ldots, p$.
4. For $t = 0$, set $V_{\mathrm{SEQ},1}$ as a p-value of 1 since this compares $\mathbf{R}$ to $\mathbf{R}$.

5. Set $t = 1$. Then $m = p - 1$. Calculate $\text{LRT}_{\text{SEQ},-1}$, which tests $H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-1}$ and $H_a : \mathbf{\Sigma}_{-0} \neq \mathbf{\Sigma}_{-1}$. Store the corresponding p-value as $V_{\text{SEQ},2}$.

6. Set $t = 2$. Then $m = p - 2$. Calculate $\text{LRT}_{\text{SEQ},-2}$, which tests $H_0 : \mathbf{\Sigma}_{-1} = \mathbf{\Sigma}_{-2}$ and $H_a : \mathbf{\Sigma}_{-1} \neq \mathbf{\Sigma}_{-2}$. Store the corresponding p-value as $V_{\text{SEQ},3}$.

7. Repeat the process iteratively for $t = 3, \ldots, p$, calculating $\text{LRT}_{\text{SEQ},-t}$, which tests $H_0 : \mathbf{\Sigma}_{-(t-1)} = \mathbf{\Sigma}_{-t}$ and $H_a : \mathbf{\Sigma}_{-(t-1)} \neq \mathbf{\Sigma}_{-t}$. Store the corresponding p-value as $V_{\text{SEQ},(t+1)}$.

Note that the last p-value in the Sequential LRT sequence will be equal to one, for the same reason that the last two BIC or AIC values are equivalent. Once only one variable is remaining in the ordering to have correlations constrained, $\mathbf{R}_{(m+1)(m+1)}$ is an identity matrix. Constraining additional correlations has no impact, so the p-value will be one.

When the full sequence of $(p+1)$ p-values from the Sequential LRT has been calculated, begin at $\mathbf{V}_{\text{SEQ},1}$ and search for the first time a $p-\text{value} < .05$ occurs. Say that this occurs for $\mathbf{V}_{\text{SEQ},i}$. Then the hypothesis that $H_0 : \mathbf{\Sigma}_{-(i-1)} = \mathbf{\Sigma}_{-i}$ has been rejected, but the previous hypothesis were not rejected. Then $u = i - 1$ is determined to be the appropriate number of variables to set as uncorrelated, given the ordering method used. In other words, the Sequential LRT stopping criterion has identified a set of m variables that are correlated, and has identified u variables that can be removed from the data prior to running factor analysis.

2.3.2.2 Full LRT

The Sequential LRT considers the hypothesis that t variables are uncorrelated with all other variables, compared to the previous step where only $(t-1)$ variables were uncorrelated. In comparison, the Full LRT compares the hypothesized structure to the ordered, unconstrained covariance. That is, the hypothesis test for the Full LRT is defined as

$H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-t}$
$H_a : \mathbf{\Sigma}_{-0} \neq \mathbf{\Sigma}_{-t}$.

Without the Bartlett correction, this test statistic would be calculated as.

$$\begin{aligned} \text{LR}_{\text{FULL},-t} &= -2\log\left(\frac{\text{L}(\hat{\boldsymbol{\Sigma}}_{-t})}{\text{L}(\hat{\boldsymbol{\Sigma}}_{-0})}\right) && (2.3.16)\\ &= n\left[\log|\hat{\boldsymbol{\rho}}_{-t}| - \log|\hat{\boldsymbol{\rho}}_{-0}|\right] && (2.3.17)\\ &= n\left[\log|\mathbf{R}_{mm}| + \log|\mathbf{I}_{tt}| - \log|\mathbf{R}|\right] && (2.3.18)\\ &= n\left[\log|\mathbf{R}_{mm}| - \log|\mathbf{R}|\right], && (2.3.19) \end{aligned}$$

The Full LRT statistic also benefits from the Bartlett correction described by Rencher and Christensen (2012). For this design, we must continue to consider the correlation matrix for all p variables, rather than reducing the problem to only the correlation matrix of $(m+1)$ variables as was done for the Sequential LRT. At any given step, the Full LRT tests that $\boldsymbol{\rho}_{-0}$ is a block diagonal matrix, with one $(m \times m)$ block with correlations of the m correlated variables and u blocks of dimension (1×1) for the variables corresponding to $Q_{m+1}, \ldots, Q_p$. Therefore, we can again apply the block-diagonal likelihood ratio statistic formula, with $b = t+1$ blocks with $m = p-t$, $p_1 = m$ and $p_2, \ldots, p_{t+1} = 1$. Then the Bartlett-corrected sequential likelihood ratio test statistic for testing t uncorrelated variables is

$$\text{LR}_{\text{FULL,-t}} = \left(\frac{n-1-2a_3+3a_2}{6a_2}\right)\left(\log|\mathbf{R}_{mm}| - \log|\mathbf{R}|\right), \qquad (2.3.20)$$

where $a_2 = p^2 - m^2 - t$ and $a_3 = p^3 - m^3 - t$. The degrees of freedom is given as $f = p(p-1)/2 - m(m-1)/2$. The null hypothesis for the Full LRT can appropriately be rejected if $\text{LR}_{\text{FULL,-t}} > \chi^2_{\alpha^*,f}$ for a given significance level α^*.

Algorithm

Using $\alpha^* = .05$, the algorithm for the Full LRT as a stopping criterion is as follows:

1. Use one of the ordering methods to determine $\mathbf{Q}$.
2. Set $t = 0$. Reorder the sample correlation matrix according to $\mathbf{Q}$, giving $\hat{\boldsymbol{\rho}}_{-0}$.
3. Define a $(p+1)\times 1$ vector $\mathbf{V}_{\text{FULL}}$ to store the p-value corresponding to LRT_{FULL} for each $t = 0, \ldots, p$.

4. For $t = 0$, set $V_{\text{FULL},1}$ as a p-value of 1 since this compares **R** to **R**.

5. Set $t = 1$. Then $m = p - 1$. Calculate $\text{LRT}_{\text{FULL},-1}$, which tests $H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-1}$ and $H_a : \mathbf{\Sigma}_{-0} \neq \mathbf{\Sigma}_{-1}$. Store the corresponding p-value as $V_{\text{FULL},2}$. This p-value will match the corresponding Sequential LRT p-value.

6. Set $t = 2$. Then $m = p - 2$. Calculate $\text{LRT}_{\text{FULL},-2}$, which tests $H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-2}$ and $H_a : \mathbf{\Sigma}_{-0} \neq \mathbf{\Sigma}_{-2}$. Store the corresponding p-value as $V_{\text{FULL},3}$.

7. Repeat the process iteratively for $t = 3, \ldots, p$, calculating $\text{LRT}_{\text{FULL},-t}$, which tests $H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-t}$ and $H_a : \mathbf{\Sigma}_{-0} \neq \mathbf{\Sigma}_{-t}$. Store the corresponding p-value as $V_{\text{FULL},(t+1)}$.

Note that the last two p-values in the Full LRT sequence will be equal to one another, for the same reason that the last two BIC or AIC values are equivalent.

When the full sequence of $(p+1)$ p-values from the Full LRT has been calculated, begin at $V_{\text{FULL},1}$ and search for the first time a p-value $< .05$ occurs. Say that this occurs for $V_{\text{FULL},i}$. Then the hypothesis that $H_0 : \mathbf{\Sigma}_{-0} = \mathbf{\Sigma}_{-i}$ has been rejected, but the previous null hypotheses were not rejected. Then $u = i - 1$ is determined to be the appropriate number of variables to set as uncorrelated, given the ordering method used. In other words, the Full LRT stopping criterion has identified a set of m variables that are correlated, and has identified u variables that can be removed from the data prior to running factor analysis.

2.3.2.3 Sequential LRT Adapted for FSR

One more likelihood ratio test that can be used as a stopping criterion is the False Selection Rate (FSR) LRT, which is an adaptation of the Sequential LRT. First explored by Luo et al. (2006), the false selection rate is the proportion of variables included in a model that are actually uninformative. Ideally, a good variable selection technique will include a high proportion of informative variables and a low proportion of uninformative variables. However, the false selection rate should be larger than zero, because it is advantageous to retain some chance of including weak, but informative, variables. For our purposes, a "falsely selected variable" is one that has been estimated as correlated, but is uncorrelated in the population.

Controlling the false selection rate γ requires tuning the significance level α^* used as criteria in forward selection or backward elimination. Wu et al. (2007) suggest controlling

the FSR through the use of pseudovariables. These phony variables are added to a model and considered alongside the p original variables in a forward selection procedure. The rate at which phony variables are added to the model is monitored, and α^* is adjusted in order to keep γ at a certain level, such as $\gamma = .05$. This approach is called Variable Addition Model Selection (VAMS). However, Boos et al. (2009) give a new version called Fast VAMS that uses straightforward calculation rather than requiring the generation of pseudovariables.

The Sequential LRT in Section 2.3.2.1 is a backward elimination method, which begins by allowing all variables in the model to freely correlate and continues to set correlations to zero as long as the likelihood ratio test is not significant. A forward-selection technique, beginning with all variables constrained to have correlations of zero, would cause problems at certain steps because the p-values are not monotone. This is due to the expected block structure of the covariance matrix when a factor model with simple structure is assumed. Suppose that at a certain step in forward selection, i variables have been allowed to correlate from k factors, based on some ordering method. If the next variable X_{i+1} to have correlations estimated as unconstrained is from an additional factor $k+1$, X_{i+1} will be uncorrelated with the i variables currently in the model, resulting in a large p-value. Then the forward selection procedure will stop, even though if other variables from factor k were now added, correlations would be nonzero between X_{i+1} and the additional variables, so corresponding p-values would be small. Therefore, a Sequential LRT forward selection approach would tend to estimate too many variables as uncorrelated.

However, the Fast VAMS method assumes forward selection, so we must reverse the Sequential LRT method and start with constraining all variables as uncorrelated, one variable to freely correlate at a time until all p variables are unconstrained. In order to avoid the problems due to the block structure of the factor model, the p-values in the sequence $\mathbf{V}_{\text{FSR}}$ are defined as the monotonized p-values from $\mathbf{V}_{\text{SEQ}}$. Then the VAMS procedure can be used to determine an appropriate adjusted value of α^*, say α^*_γ, no larger than .05.

Once $\mathbf{V}_{\text{FSR}}$ is monotonized and α^*_γ is estimated via Fast VAMS, the rest of the FSR LRT procedure is similar to the Sequential LRT procedure. Identify the element $V_{\text{FSR},i} > \alpha^*_\gamma$ such that $V_{\text{FSR},(i+1)} < \alpha^*_\gamma$. Then $u = i-1$ is set as the number of variables to estimate as uncorrelated, given the ordering method used. In other words, variables with a p-value in $\mathbf{V}_{\text{FSR}}$ larger than α^*_γ are estimated as uncorrelated. Once again, these variables are removed

from the dataset prior to running factor analysis.

2.3.2.4 Difference from Stepwise Selection Method

Kano and Harada (2000) also recommended the use of a likelihood ratio test, with Barlett correction, for selecting variables in factor analysis. However, a key distinction compared to our stopping criteria methods is that Kano and Harada (2000) tested the null hypothesis that the covariance structure was that of a factor model, $H_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}}$. The corresponding likelihood ratio statistic is given in Equation (1.5.3). As Hogarty et al. (2004) pointed out, this causes the stepwise selection method to fail to identify certain uncorrelated variables. Specifically, the stepwise selection method does not remove a variable X_i that is uncorrelated with the remaining $p-1$ variables, because X_i still fits the factor model with $\mathbf{L}_{i,\cdot} = \mathbf{0}$. However, this is precisely the type of variable that is identified as uncorrelated in the pre-screening methods presented in this chapter.

2.3.3 Simulation Study

Each of the five stopping criteria defined in Section 2.3, paired with a method for ordering variables from Section 2.2, produces an estimate of which m variables are correlated and which u variables are uncorrelated. Each combination of ordering method and stopping criteria can be considered one pre-screening method. A simulation study was conducted to determine if these pre-screening methods are able to accurately classify variables into the correct subsets of u and m variables.

2.3.3.1 Assessing Stopping Criteria

Rand Index

The Rand index is used to track how closely the two clusters of variables (uncorrelated and correlated) agree with the population clusters. The Rand index is a measure of agreement between two sets of categorizations of the same data, in our case for categorizing uncorrelated and correlated variables in the population and the sample (Rand, 1971). The Rand index ranges between zero, for no agreement between the sample and population

classifications, and one, indicating perfect agreement between the two classifications (Chi et al., 2016).

It is generally recommended to use the adjusted Rand Index (ARI), which accounts for the chance that pairs of points match randomly (Hubert and Arabie, 1985). The adjusted Rand Index still has an upper bound at one, indicating perfect similarity, but a value of 0 is the expected value of the Rand Index, corresponding to random assignment to groups. Negative values are also possible. In `R`, we use the `adjustedRandIndex` function from the `mclust` package to perform the calculation (Fraley et al., 2012).

Average Number of Correlated Variables

The Rand index does not indicate if a stopping criterion is biased, meaning the stopping criterion tends to either over- or under-estimate the true number of correlated variables. Therefore, averages of the number of variables m that are classified as correlated by each stopping criterion were also calculated.

Perfect Ordering

Since this simulation focuses on stopping criteria rather than ordering methods, the only ordering method from Section 2.2 that is used in this study is the AAC ordering paired with the correlation matrix $\mathbf{R}_2$. This ordering method combination performed consistently well across cases in the ordering method simulation study. The full set of ordering method/stopping criterion combinations will be considered in the simulation study in Chapter 3.

However, in an attempt to separate out the influence of ordering methods and stopping criteria, "perfect" orderings were also randomly generated for this study. If in the population, m variables are correlated and u variables are uncorrelated, then the $\mathbf{Q}$ matrix of one perfect ordering is created as follows. Create a random permutation of the numbers 1 through m, representing an ordering of correlated variables, and store this permutation as the first m elements of $\mathbf{Q}$. Then permute the numbers 1 through u, add m to each value, and store this permutation as the last u elements of $\mathbf{Q}$.

Note, however, that these perfect orderings do not take the sample data into account. Even though the groups will be perfectly classified as correlated or uncorrelated, the order-

ing within each group of m and u variables might actually produce a less accurate ordering than an ordering method that account for variations in each particular dataset, such as AAC. The corresponding value of the maximized likelihood for a given step in the criterion might therefore be uncharacteristic of what would be seen in practice.

2.3.3.2 Simulation Design

A simulation study was used to determine **Q** from either a random perfect ordering or from the AAC ordering using $\mathbf{R}_2$, and to classify correlated and uncorrelated variables using each of the five stopping criteria. In an ideal situation, the AAC ordering method would give all correlated variables at the beginning of **Q**, followed by all uncorrelated variables. This situation is exactly what the perfect ordering method engineers. Then, ideally, the stopping criterion would perfectly partition the uncorrelated and correlated variables. This ability to correctly group the correlated and uncorrelated variables is numerically represented by the Rand Index. Bias in estimating m (and therefore $u = p - m$) is assessed by calculating the average number of variables set as correlated in each data set.

The data generation process was mostly identical to that of Section 2.2.7.2. However the case, $\alpha = 0$ was added as a baseline measure to see how many variables each stopping criterion sets as uncorrelated when data is generated from a covariance matrix equal to the identity. The factors considered for the study were then:

1. Number of correlated variables ($m = 6$, 12, or 40)
2. Number of uncorrelated variables ($u = 1$, 5, or 10)
3. Sample size ($n = 100$ or $n = 400$)
4. Magnitude of factor loadings (controlled by α =0, .25, .50, .75 or 1)
5. Ordering Method (AAC with $\mathbf{R}_2$ or Perfect Ordering)
6. Stopping Criterion (BIC, AIC, Full LRT, Seq LRT, or FSR LRT)

The population **L** matrices were defined as in Section 2.2.7.2, with the addition of the **L** matrix composed of all zeros corresponding to $\alpha = 0$. One hundred sets of data were

generated for each population **L** matrix and sample size. All data was generated from the factor model.

Due to the way the seed was set in the simulation, the 100 perfect orderings that were generated were identical across $L_{.25}$ up to L_1 for a given m and u, although of course the sample correlations differed.

2.3.3.3 Results

The following tables show the average adjusted Rand index for each ordering method and stopping criterion combination in each combination of factors. Similarly, the average number of correlated variables selected is given to determine if the stopping criterion is over- or under-estimating on average. To assist in comparisons, an average standard error is given at after the last row for each sample size. Standard error is calculated as

$$SE = \frac{s}{\sqrt{N}}, \tag{2.3.21}$$

where s is the standard deviation of the $N = 100$ replications in each combination of factors. Average standard error is therefore the average among the 5 standard errors given for each stopping criterion at each combination of factors.

Results using perfect orderings are displayed in Table 2.7 through Table 2.11. Results using AAC ordering with $\mathbf{R}_2$ are displayed in Table 2.12 through Table 2.16.

2.3.4 Conclusions

Here we give a summary of the simulation results comparing the various types of stopping criteria and ordering methods. For convenience, estimating a variable as correlated is often referred to as "selecting" the variable, because such a variable will be selected, or retained, to be included in the factor analysis model once pre-screening is complete.

1. In general, using perfect orderings does not seem to lead to any different conclusions than using AAC ordering with $\mathbf{R}_2$. The general patterns in findings remain the same, although the use of perfect orderings seems to lead to attenuated average ARI values in some cases. For example, Table 2.11 for perfect orderings and $m = 40$, $u - 10$, shows that the Full LRT gives average adjusted Rand indices between .83 and .86 for

Table 2.7 Average adjusted Rand index and estimated number of correlated variables for $m = 6$ correlated, $u = 1$ uncorrelated variables, perfect orderings. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	-0.01	-0.04	0.92	1	0.04	0.10	0.77	5.82	6
	AIC	0	-0.01	0.52	0.98	0.98	0.59	0.62	4.49	6.02	6.02
	Full LRT	0.01	0.04	0.42	0.97	0.98	0.55	0.47	3.90	6.00	6.02
	Seq LRT	0	-0.06	0.49	0.98	0.98	1.78	1.78	4.58	6.02	6.02
	FSR LRT	0	0.02	0.42	0.98	0.98	0.25	0.33	3.32	6.02	6.02
	Av SE	0.01	0.02	0.04	0.02	0.01	0.12	0.12	0.20	0.02	0.01
400	BIC	0	-0.01	0.58	1	1	0.02	0.12	4.83	6	6
	AIC	0	0.03	0.92	0.93	0.93	0.81	1.61	6.06	6.07	6.07
	Full LRT	0	0.08	0.93	0.94	0.94	0.47	1.36	6.04	6.06	6.06
	Seq LRT	-0.08	0.03	0.93	0.94	0.94	1.93	2.62	6.05	6.06	6.06
	FSR LRT	0	0.04	0.93	0.94	0.94	0.29	0.66	6.05	6.06	6.06
	Av SE	0.01	0.02	0.03	0.02	0.02	0.12	0.17	0.05	0.02	0.02

Table 2.8 Average adjusted Rand index and estimated number of correlated variables for $m = 6$ correlated, $u = 5$ uncorrelated variables, perfect orderings. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	0	0.01	0.93	1	0.02	0.04	0.44	5.75	6
	AIC	0	0.06	0.61	0.97	0.97	0.67	0.81	4.45	6.09	6.09
	Full LRT	0	0.05	0.33	0.88	0.92	1.20	1.12	3.40	6.20	6.32
	Seq LRT	0	0.14	0.57	0.84	0.84	3.64	3.25	5.54	6.59	6.59
	FSR LRT	0	0.06	0.41	0.91	0.91	0.55	0.50	3.14	6.34	6.34
	Av SE	0	0.02	0.03	0.02	0.02	0.21	0.20	0.23	0.09	0.07
400	BIC	0	0	0.59	1	1	0.02	0.08	4.39	6	6
	AIC	0.01	0.13	0.97	0.97	0.97	0.76	1.40	6.08	6.08	6.08
	Full LRT	0	0.08	0.80	0.89	0.89	1.28	1.81	6.22	6.46	6.46
	Seq LRT	0.01	0.28	0.83	0.83	0.83	3.18	4.29	6.68	6.68	6.68
	FSR LRT	0	0.10	0.89	0.89	0.89	0.27	1.03	6.41	6.42	6.42
	Av SE	0	0.02	0.03	0.02	0.02	0.19	0.23	0.12	0.08	0.08

Table 2.9 Average adjusted Rand index and estimated number of correlated variables for $m = 12$ correlated, $u = 5$ uncorrelated variables, perfect orderings. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	0	-0.03	0.80	1	0	0.04	0.96	11.03	12
	AIC	0	-0.02	0.61	0.99	0.99	0.85	1.08	9.54	12.01	12.03
	Full LRT	0	0.05	0.50	0.89	0.93	1.31	2.15	9.87	12.14	12.32
	Seq LRT	0.01	0.24	0.75	0.88	0.89	6.78	7.12	11.86	12.50	12.52
	FSR LRT	0	0.04	0.65	0.90	0.91	0.37	0.80	10.09	12.40	12.42
	Av SE	0	0.02	0.03	0.02	0.02	0.26	0.30	0.27	0.10	0.07
400	BIC	0	-0.01	0.34	1	1	0.02	0.10	7.81	11.99	12
	AIC	0	-0.02	0.98	0.99	0.99	0.96	2.39	12.00	12.04	12.04
	Full LRT	0	0.17	0.87	0.93	0.93	2.53	5.22	12.11	12.35	12.35
	Seq LRT	0.02	0.46	0.85	0.85	0.85	6.27	10.02	12.65	12.68	12.68
	FSR LRT	0	0.17	0.86	0.87	0.87	0.21	3.55	12.56	12.61	12.61
	Av SE	0	0.02	0.03	0.02	0.02	0.29	0.36	0.15	0.08	0.08

Table 2.10 Average adjusted Rand index and estimated number of correlated variables for $m = 40$ correlated, $u = 1$ uncorrelated variables, perfect orderings. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	0	-0.01	-0.02	0.97	0.02	0.04	0.71	14.37	39.93
	AIC	-0.01	-0.01	0.01	0.98	0.99	0.69	0.91	17.75	39.97	40.01
	Full LRT	-0.01	0.20	0.65	0.92	0.95	11.75	19.76	38.85	39.97	40.05
	Seq LRT	-0.01	0.29	0.72	0.95	0.95	26.82	32.12	39.16	40.04	40.05
	FSR LRT	0	0.02	0.59	0.95	0.95	1.54	2.34	32.90	40.04	40.05
	Av SE	0	0.02	0.02	0.02	0.02	0.75	0.73	0.51	0.08	0.02
400	BIC	0	0	-0.03	0.94	1	0	0.10	9.21	39.82	40
	AIC	0.02	-0.02	1	1	1	0.69	2.61	39.99	40	40
	Full LRT	0	0.40	0.97	0.97	0.97	7.65	35.90	40.03	40.03	40.03
	Seq LRT	0	0.51	0.97	0.97	0.97	25.03	37.32	40.03	40.03	40.03
	FSR LRT	0	0.25	0.97	0.97	0.97	0.98	16.94	40.03	40.03	40.03
	Av SE	0.01	0.02	0.01	0.01	0.01	0.69	0.60	0.08	0.02	0.01

Table 2.11 Average adjusted Rand index and estimated number of correlated variables for $m = 40$ correlated, $u = 10$ uncorrelated variables, perfect orderings. See page 45 for **L** definitions.

n	Stopping Criteria	Adjusted Rand Index						Average m				
		$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$		$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	0	-0.02	-0.08	0.99		0.10	0.17	1.03	13.75	39.92
	AIC	0	-0.02	0.01	0.99	1	0.83	0.88		16.80	40	40.04
	Full LRT	0	0.16	0.55	0.70	0.75		19.00	22.94	39.97	42.06	42.61
	Seq LRT	0	0.44	0.63	0.66	0.66		37.23	39.17	42.99	43.44	43.45
	FSR LRT	0	0.03	0.57	0.69	0.69		1.18	2.43	36.04	43.19	43.20
	Av SE	0	0.02	0.02	0.02	0.02		0.89	0.85	0.69	0.30	0.24
400	BIC	0	0	-0.10	0.98	1		0.06	0.09	8.56	39.76	40
	AIC	0.01	-0.03	1	1	1		0.70	1.73	39.98	40	40
	Full LRT	0	0.42	0.83	0.86	0.86		12.67	34.25	41.05	41.39	41.39
	Seq LRT	-0.01	0.61	0.77	0.77	0.77		31.85	40.22	42.38	42.38	42.38
	FSR LRT	0	0.29	0.81	0.81	0.81		1.19	17.57	41.99	41.99	41.99
	Av SE	0	0.02	0.02	0.02	0.02		0.88	0.74	0.28	0.21	0.20

Table 2.12 Average adjusted Rand index and estimated number of correlated variables for $m = 6$ correlated, $u = 1$ uncorrelated variables, AAC orderings with $\mathbf{R}_2$. See page 45 for **L** definitions.

n	Stopping Criteria	Adjusted Rand Index					Average m				
		$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0.01	-0.02	-0.09	0.66	1	0.88	0.9	2.01	5.09	6
	AIC	0	-0.03	0.16	0.89	0.89	2.23	2.48	4.15	6.11	6.11
	Full LRT	0.01	0.01	0.04	0.89	0.91	0.11	0.25	2.95	6.05	6.09
	Seq LRT	0	0	0.09	0.90	0.91	1.66	1.82	3.79	6.08	6.09
	FSR LRT	0	-0.01	0.06	0.90	0.91	0.38	0.45	2.71	6.08	6.09
	Av SE	0.02	0.02	0.03	0.03	0.02	0.09	0.10	0.16	0.05	0.02
400	BIC	0.01	-0.02	0.12	1	1	0.42	0.67	3.59	6	6
	AIC	0.01	-0.04	0.90	0.91	0.91	2.01	2.74	6.08	6.09	6.09
	Full LRT	0	-0.03	0.86	0.92	0.92	0.15	0.75	5.97	6.08	6.08
	Seq LRT	0.01	-0.09	0.91	0.92	0.92	1.69	2.18	6.06	6.08	6.08
	FSR LRT	0.01	-0.01	0.91	0.92	0.92	0.32	0.64	6.06	6.08	6.08
	Av SE	0.02	0.01	0.03	0.02	0.02	0.09	0.12	0.05	0.02	0.02

Table 2.13 Average adjusted Rand index and estimated number of correlated variables for $m = 6$ correlated, $u = 5$ uncorrelated variables, AAC orderings with $\mathbf{R}_2$. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	0	0.06	0.81	1	0.97	1.12	1.94	5.38	6
	AIC	-0.01	0.01	0.22	0.86	0.88	3.24	3.32	4.51	6.40	6.35
	Full LRT	0	0	0.08	0.82	0.96	0.16	0.28	2.22	5.76	6.13
	Seq LRT	0	0	0.21	0.83	0.81	2.93	2.99	4.46	6.51	6.6
	FSR LRT	0	0	0.10	0.89	0.87	0.53	0.57	2.46	6.31	6.43
	Av SE	0.01	0.01	0.02	0.03	0.02	0.12	0.12	0.17	0.08	0.06
400	BIC	0	0	0.31	1	1	0.38	0.54	3.57	6	6
	AIC	0	0.02	0.88	0.90	0.92	2.87	3.16	6.27	6.28	6.23
	Full LRT	0	0	0.77	0.97	0.97	0.15	0.48	5.54	6.10	6.10
	Seq LRT	-0.01	0.02	0.88	0.86	0.86	2.60	3.26	6.29	6.41	6.44
	FSR LRT	0	0	0.90	0.92	0.91	0.29	0.6	6.09	6.25	6.31
	Av SE	0.01	0.01	0.02	0.02	0.02	0.11	0.12	0.08	0.05	0.05

Table 2.14 Average adjusted Rand index and estimated number of correlated variables for $m = 12$ correlated, $u = 5$ uncorrelated variables, AAC orderings with $\mathbf{R}_2$. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	-0.01	-0.09	0.36	1	0.99	0.91	3.55	8.67	12
	AIC	0	-0.02	0.26	0.91	0.99	4.09	4.31	8.23	11.87	12.06
	Full LRT	0	0	0.15	0.80	0.97	0.31	0.51	7.09	11.34	12.16
	Seq LRT	0.02	-0.01	0.35	0.85	0.85	4.74	5.04	9.63	12.44	12.67
	FSR LRT	0	-0.01	0.16	0.84	0.87	0.70	0.73	7.20	12.25	12.60
	Av SE	0.01	0.01	0.02	0.02	0.01	0.17	0.17	0.21	0.10	0.07
400	BIC	0	-0.03	0.12	1	1	0.35	0.82	6.86	11.99	12
	AIC	0	-0.03	0.91	0.97	0.98	3.61	5.26	11.85	12.11	12.08
	Full LRT	0	-0.02	0.77	0.98	0.97	0.09	1.84	11.18	12.08	12.14
	Seq LRT	0.01	0	0.89	0.92	0.89	4.52	6.41	12.20	12.36	12.49
	FSR LRT	0	-0.03	0.88	0.92	0.91	0.65	2.04	12.07	12.33	12.41
	Av SE	0.01	0.01	0.01	0.01	0.01	0.15	0.19	0.08	0.05	0.06

Table 2.15 Average adjusted Rand index and estimated number of correlated variables for $m = 40$ correlated, $u = 1$ uncorrelated variables, AAC orderings with $\mathbf{R}_2$. See page 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0.01	-0.01	-0.04	-0.02	0.74	0.80	1.08	5.05	15.27	38.94
	AIC	0.01	-0.02	0	0.77	0.98	5.20	6.76	19.73	39.22	40.02
	Full LRT	0	0	0.21	0.75	0.93	0.70	4.08	32.83	39.53	40.07
	Seq LRT	0	0	0.23	0.87	0.93	14.78	19.41	34.11	39.89	40.07
	FSR LRT	0.01	0	0.20	0.87	0.93	2.64	5.23	31.86	39.89	40.07
	Av SE	0.01	0.01	0.01	0.02	0.02	0.39	0.5	0.44	0.10	0.05
400	BIC	0	-0.01	-0.03	0.54	1	0.31	0.92	12.5	38.05	40
	AIC	-0.01	-0.03	0.86	1	1	4.62	11.08	39.54	40	40
	Full LRT	0	0.01	0.94	0.96	0.96	0.34	20.58	39.99	40.04	40.04
	Seq LRT	-0.01	0.03	0.96	0.96	0.96	13.94	24.13	40.03	40.04	40.04
	FSR LRT	-0.01	0	0.96	0.96	0.96	2.96	16.67	40.03	40.04	40.04
	Av SE	0	0	0.02	0.02	0.01	0.28	0.44	0.08	0.04	0.01

Table 2.16 Average adjusted Rand index and estimated number of correlated variables for $m = 40$ correlated, $u = 10$ uncorrelated variables, AAC orderings with $\mathbf{R}_2$. Seepage 45 for **L** definitions.

n	Stopping	Adjusted Rand Index					Average m				
	Criteria	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$	$\mathbf{L}_0$	$\mathbf{L}_{.25}$	$\mathbf{L}_{.50}$	$\mathbf{L}_{.75}$	$\mathbf{L}_1$
100	BIC	0	-0.02	-0.09	-0.07	0.90	0.66	1.31	5.29	15.28	38.84
	AIC	0.01	-0.05	0.01	0.92	0.99	5.10	5.77	19.38	39.19	40.11
	Full LRT	0	-0.02	0.44	0.81	0.82	0.97	4.96	32.22	39.58	41.77
	Seq LRT	0	0.01	0.61	0.76	0.69	20.82	24.48	39.06	42.28	43.26
	FSR LRT	0.01	-0.01	0.51	0.77	0.70	3.36	5.45	36.37	42.15	43.16
	Av SE	0.01	0.01	0.02	0.02	0.02	0.48	0.58	0.50	0.24	0.24
400	BIC	0	-0.02	-0.10	0.80	1	0.29	0.94	12.4	37.6	40
	AIC	0.01	-0.1	0.95	1	1	4.58	10.64	39.45	40	40
	Full LRT	0	0.03	0.91	0.98	0.92	0.39	20.04	39.30	40.22	40.83
	Seq LRT	0.01	0.17	0.90	0.86	0.81	18.24	28.53	41.04	41.49	42.03
	FSR LRT	0	0	0.90	0.86	0.81	3.66	15.65	41.04	41.49	42.03
	Av SE	0	0.01	0.01	0.01	0.02	0.33	0.51	0.14	0.13	0.17

moderate to high factor loadings. However, Table 2.16 gives average adjusted Rand indices between .91 and .98 for the same cases. On a related note, in these cases the AAC ordering results in less bias when determining m compared to perfect orderings, which tend to allow allow too many variables to freely correlate. Since AAC orderings are also what would be used in practice rather than perfect orderings, we will only consider the results from AAC orderings further (Tables 2.12 to 2.16).

2. As expected, the average adjusted Rand index for $\mathbf{L}_0$ is approximately zero, since the classifications truly are random. However, this same adjusted Rand index might correspond to many different values of average m, as indicated in any one of the tables.

3. FSR LRT appears to be a clear improvement over the Sequential LRT. Although both methods perform equally well when loadings are high, the FSR approach offers a great improvement for the baseline case of an identity matrix, where zero variables should be selected as correlated. As it should, FSR LRT better controls the false selection rate, which in this case means that it prevents uncorrelated variables from being estimated as correlated much better than the Sequential LRT. Therefore, Sequential LRT will not be considered further, since FSR LRT is a clear improvement.

4. When loadings are high ($\alpha = 1$), and the number of correlated variables is smaller ($m =$ 6 or 12), BIC is able to perfectly detect the correct correlated variables (Tables 2.12, 2.13, and 2.14). AIC and LRT methods tend to over-select correlated variables in these cases. However, when there is a higher number of correlated variables (40), BIC tends to slightly under-select correlated variables, whereas AIC appears the least biased and LRT methods over-select correlated variables (Tables 2.15 and 2.16).

5. Since the methods displayed in these simulations are pre-screening methods for factor analysis, estimating too many variables as uncorrelated is more problematic than estimating too few variables as uncorrelated. Variables estimated as uncorrelated are removed from the dataset prior to running factor analysis, so when m is set too low, valuable information might be lost. Unfortunately, BIC tends to severely under-select correlated variables as soon as population loadings are below 0.9. For example, Table 2.15 shows that for a population with 40 correlated variables and moderate

loadings (corresponding to $\alpha = .75$), BIC only identifies 15 of the 40 variables as correlated on average. At $\alpha = .5$, BIC only identifies five of the 40 variables as correlated on average. These large discrepancies of BIC lead us to reject BIC as a stopping criterion method.

6. AIC shows the most problems in Tables 2.15 and 2.16 where the number of variables is large. For $L_{.50}$ with small sample size, AIC only correctly selects about half of the correlated variables on average. The LRT methods find upwards of 75% of the appropriate number variables. This finding is problematic for AIC. However, when sample size is large, AIC retains much closer to the appropriate number of $m = 40$ correlated variables.

7. When comparing Full LRT and FSR LRT, results in Table 2.16 for $m = 40$ $u = 10$ are interesting. At $\alpha = .5$ and $n = 100$, FSR LRT appears less biased on average for determining m. Full LRT has an average m of around 32, and an average ARI of .81, whereas FSR LRT has an average m of around 36. However, the corresponding average ARI for FSR LRT is only .77. This indicates that even though FSR LRT is generally better at estimating the correct *number* of correlated variables, this does not necessarily mean the classification of correlated and uncorrelated variables is also better.

8. For high loadings, m=40, and u=10 (Table 2.16), the FSR LRT method appears to retain too many variables are uncorrelated when compared to the Full LRT, which exhibits less bias.

9. The case of $\alpha = .25$ is somewhat ambiguous. This corresponds to population loadings around .23, which in traditional factor analysis interpretation would be considered too low to retain a variable. The stopping criteria methods tend to retain around anywhere from zero to one half of the correlated variables in this case, depending on m, u, and n. It is difficult to interpret this results as either a positive or negative finding.

2.4 Proposed Procedures

This chapter has explored multiple methods for pre-screening variables in order to remove uncorrelated variables prior to estimating the factor model. As a reminder, it is considered more problematic to remove too many variables than retain too many variables since this is a pre-screening method.

Based on the findings presented, all five ordering methods (AAC, MAC, SMC, LR, and MEMSEL) are contenders for ordering variables based on correlation. These ordering methods could be used with either $\mathbf{R}$ or $\mathbf{R}_2$ to create the vector or ordered variables $\mathbf{Q}$.

Out of the five stopping criteria considered (BIC, AIC, Full LRT, Sequential LRT, and FSR LRT), three were found to have merit for appropriately determining m and u (AIC, Full LRT, and FSR LRT).

Therefore, moving forward we suggest considering the five-times-two ordering method combinations each paired with the three stopping criteria for pre-screening variables. Chapter 3 will describe a simulation study pairing these pre-screening methods with parallel analysis for determining k, and then conducting maximum likelihood factor analysis on the reduced set of correlated variables.

These pre-screening methods show promising, objective alternatives to traditional approaches commonly used in factor analysis for selecting both variables and factors.

CHAPTER

3

SIMULATION STUDY

3.1 Introduction

In Chapter 2, we introduced pre-screening methods for identifying uncorrelated variables. This chapter presents a simulation study that compares these pre-screening methods with other factor analysis variable selection techniques discussed in Chapter 1. Whereas Chapter 2 focused only on ordering and classifying variables according to correlation, this chapter discusses the entire factor analysis process, including determining the number of factors k, estimating the factor model, and rotating the resulting loadings matrix.

3.2 Methods to Compare

Each approach to variable selection in factor analysis follows recommendations from Chapters 1 and 2. We compare our new pre-screening methods, the traditional magnitude

and communality rules, and the more recently developed Group Lasso Factor Analysis approach.

3.2.1 Pre-Screening Methods

Our five ordering methods (MAC, AAC, SMC, LR, and MEMSEL), two types of correlation matrices for ordering ($\mathbf{R}$ and $\mathbf{R}_2$), and three stopping criteria (AIC, Full LRT, and FSR LRT) were used to pre-screen variables, for a total of 30 combinations of pre-screening methods.

Once each of these pre-screening methods identified a subset of uncorrelated variables, the u uncorrelated variables were removed from the dataset. Using only the remaining m correlated variables, parallel analysis was conducted to determine the appropriate number of factors to retain. We employed the recommendations in Section 1.4.5 and use the 95th percentile, rather than the mean, to determine the comparisons between sample and random eigenvalues (Glorfeld, 1995). Further, we applied principal axis factoring rather than principal components to determine the appropriate eigenvalues in parallel analysis (Green et al., 2012). Parallel analysis was implemented using the `paran` package in `R` (Dinno, 2009).

Next, maximum likelihood factor analysis was conducted using the m correlated variables and the number of factors k determined by parallel analysis. MLFA was implemented using the `factanal` function in `R`. The resulting estimate of $\mathbf{L}$ was rotated using the normalized varimax criterion from Section 1.2.3.1 (`varimax` function in `R`).

3.2.2 Traditional Methods

Both the magnitude rule from Section 1.5.1.1 and the communality rule from Section 1.5.1.2 were used as traditional approaches to variable selection in factor analysis.

Parallel analysis was conducted on all p variables using the 95th percentile and principal axis factoring. Using k factors as determined from parallel analysis, the factor model was estimated with MLFA for all p variables. The resulting estimate of $\mathbf{L}$ was rotated using the normalized varimax criterion. The traditional communality rule identified variables in $\hat{\mathbf{L}}$ with communalities less than 0.2 for removal from the model (Child, 2006). The traditional magnitude rule identified any variables for removal that had no absolute loadings greater

than $c = 0.32$ (Tabachnick and Fidell, 2001).

With either of these traditional rules, the u identified variables were then removed from the dataset and parallel analysis and MLFA were run once more on the reduced set of m variables. The resulting $\hat{\mathbf{L}}$ was rotated with the normalized varimax criterion as well.

3.2.3 Group Lasso Factor Analysis

The Group Lasso Factor Analysis (GLAFA), introduced in Section 1.5.2.2, was implemented using an R package obtained from the authors (Hirose and Konishi, 2012). Given the number of factors k and a regularization parameter θ, the package will estimate the loadings matrix via the EM algorithm and return the values of $\hat{\mathbf{L}}$ and the GBIC. Recall that the model must be estimated over a grid of $\theta > 0$ and k, from which final values are chosen to minimize the GBIC. Therefore, we specified the maximum value of k based on Equation (1.4.1) as $k_{\max} = \text{floor}(\left[(2p+1) - \sqrt{(8p+1)}\right]/2)$. The range of θ was chosen in an ad hoc manner so that the minimum GBIC was never found at the boundary of the range of θ values, resulting in a maximum θ of 0.2. We chose to consider a total of 20 regularization parameters in this range, giving the vector of possible values $\boldsymbol{\theta} = (\theta_1 = .01, \theta_2 = .02, \ldots, \theta_20 = .2)^T$. This resulted in a $20 \times k_{\max}$ grid over which to estimate the GLAFA model. No more than 20 regularization parameters were considered due the computational intensity of the procedure.

Due to the iterative nature of the EM algorithm, the GLAFA package also requires setting a maximum number of iterations and a tolerance level for convergence. The maximum iterations allowed was set to 20,000, and the tolerance was set to 1^{-6}. Although the default in the package was a tolerance of 1^{-8}, we found that such a small tolerance led to many errors regarding lack of convergence. Rather than increasing the maximum number of iterations, which greatly increased the already-lengthy computational time, we relaxed the tolerance requirements. Even with this change, the occasional warning of non-convergence still occurred.

Once the model was estimated for each point in the grid over θ and k, the solution was chosen that minimized the GBIC. The resulting $\hat{\mathbf{L}}$ had rows that are entirely equal to zero if they were identified for removal from the model. These rows were removed, and the resulting $\mathbf{L}$ matrix was rotated using the normalized varimax criterion. However, Section 3.4.2 will discuss an adaptation that we made to GLAFA based on these solutions.

3.3 Simulation Design

A simulation was designed to compare these 33 approaches to variable selection in factor analysis (30 pre-screening, 2 traditional, and 1 GLAFA). The goal is to properly identify the u uncorrelated and m uncorrelated variables, determine the appropriate number of factors k, and return $\hat{\mathbf{L}}$ so that it closely resembles $\mathbf{L}$. The factors considered in this study were:

1. Number of factors ($k = 2$ or 5)
2. Number of correlated variables per factor ($mpk = 3$ or 9)
3. Number of uncorrelated variables ($u = 1$ or 10)
4. Magnitude of factor loadings (high, low, or mixed)
5. Sample size ($n = 100$ or $n = 400$)
6. Variable selection method (30 pre-screening, 2 traditional, 1 GLAFA)

Based on this simulation design, the same number of variables was associated with each factor. High magnitude of factor loadings indicates that all nonzero factor loadings were 0.8. Low magnitude of factor loadings similarly indicates that all nonzero factor loadings were 0.4. Mixed factor loadings indicates that 1/3 of the nonzero factor loadings were 0.4, 1/3 were 0.6, and 1/3 were 0.8. One additional row of zeros was added to the population $\mathbf{L}$ matrix for each uncorrelated variable. Loadings were assigned to preserve simple structure, meaning each variable only had a nonzero loading on one factor. For example, the population loadings matrix for $k = 2$ factors, $mpk = 9$ correlated variables per factor, and $u = 1$ uncorrelated variable is defined as:

$$\mathbf{L} = \begin{pmatrix} .4 & .4 & .4 & .6 & .6 & .6 & .8 & .8 & .8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .4 & .4 & .4 & .6 & .6 & .6 & .8 & .8 & .8 & 0 \end{pmatrix}^T$$

Note that the design was full-factorial except that $k = 2$ factors was paired with $mpk = 9$ but not $mpk = 3$. This led to 36 combinations of simulation factors on which each of the 33 variable selection methods was tested. Based on the combinations of number of factors,

correlated variables per factor, and number of uncorrelated variables, this design leads to a range of the number of correlated variables from $m = 15$ to $m = 45$, and the total number of variables from $p = 16$ to $p = 55$.

Simulation factors 1 through 4 were used to create the population $\mathbf{L}$ matrices, from which the population covariance matrices were defined. As in the simulations from Chapter 2, $\mathbf{\Sigma}$ was assumed to be standardized so that $\mathbf{\Psi}$ was completely determined by $\mathbf{L}$, since $\sigma^2_{X_i} = 1 = L^2_{i1} + \cdots + L^2_{ik} + \Psi_i$ when standardization is assumed. Then $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{D}_{\mathbf{\Psi}}$, with diagonal elements of one.

3.3.1 Data Generation

Suppose that $\mathbf{Z}_j = (Z_1, Z_2, \ldots, Z_p)^T$ is an observable random vector with p elements, and $\mathbf{Z}_j \sim N_p(\mathbf{0}, \mathbf{I})$ for $j = 1, \ldots, n$ where n is the sample size and p is the total number of variables. Then define the $(n \times p)$ data matrix $\mathbf{X} = \mathbf{Z}\mathbf{\Sigma}^{1/2}$ where $\mathbf{\Sigma}^{1/2}$ is the positive definite symmetric square root (found through spectral decomposition) of a $p \times p$ covariance matrix as defined by a factor model from the simulation design. Then $E(\mathbf{X}) = E(\mathbf{Z})\mathbf{\Sigma}^{1/2} = \mathbf{0}$, and $\text{Cov}(\mathbf{X}) = \mathbf{\Sigma}^{1/2}\text{Cov}(\mathbf{Z})\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$, so $\mathbf{X}_j \sim N_p(\mathbf{0}, \mathbf{\Sigma})$.

$N = 100$ sets of data were generated for each population $\mathbf{\Sigma}$ matrix and sample size. The same 100 $(n \times p)$ $\mathbf{Z}$ datasets were used with each combination of sample size and population covariance matrix of the same dimension. Specifically, for each combination of k, mpk, and u, there are three population covariance matrices; one for each magnitude of factor loadings. These three covariance matrices were each multiplied by the same $\mathbf{Z}$ matrix to create the final set of data used in the simulation. Just as in Chapter 2, this has a blocking effect on the data, allowing simulation results to showcase differences due to $\mathbf{\Sigma}$, rather than seeing random differences in results dependent on the seed used.

Aside from generating normally distributed data from the factor model, we also considered generating non-normal data, and data from a non-factor model using a convex combination of correlation matrices.

For non-normal data, consider the factor model $\mathbf{X} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}$ where without loss of generality, we assume $\boldsymbol{\mu} = 0$. It is known that $E(\mathbf{F}) = \mathbf{0}$, $\text{Cov}(\mathbf{F}) = \mathbf{I}$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, and $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{D}_{\mathbf{\Psi}}$. Then consider generating $\mathbf{F}_j \sim N_k(\mathbf{0}, \mathbf{I})$ and each $\epsilon_i \sim \exp(\lambda = [1/\sqrt{\psi_i}]) - \sqrt{\psi_i}$, for $i = 1 \ldots p$. Then $E(\epsilon_i) = \lambda^{-1} - \sqrt{\psi_i} = \sqrt{\psi_i} - \sqrt{\psi_i} = 0$, and $\text{Var}(\epsilon_i) = \lambda^{-2} = \psi_i$. Then data can be generated

from the factor model using $\mathbf{L}$ along with $\mathbf{F}$ and ϵ.

Conversely, consider generating data from the normal distribution but with a convex combination of correlation matrices such that data is not perfectly generated from the factor model. Define $\mathbf{\Sigma}_f$ as the $(p \times p)$ population correlation matrix defined from the factor model, and $\mathbf{C}$ as a $(p \times p)$ correlation matrix with equicorrelations such that $C_{ij} = .2$, i$\neq j$. Then define $\mathbf{\Sigma} = (.8)\mathbf{\Sigma}_f + (.2)\mathbf{C}$ and again generate data as $\mathbf{X} = \mathbf{Z}\mathbf{\Sigma}^{1/2}$.

Although generating either non-normal data nor non-factor-model data generally led to less ability to correctly identify uncorrelated variables, the relative pattern of performance between the 33 selection methods remained unchanged. Since these alternative ways of generating data did not contribute to distinguishing between the variable selection methods, they were not considered further.

3.3.2 Summary Statistics

As in Chapter 2, the adjusted Rand index (ARI) was used as a measure of how well each selection method was able to classify variables as uncorrelated and correlated. For each replication, the estimated values of u, and k were also recorded. Note that determining u also determines m, since $p = m + u$. For these statistics, in addition to summarizing these statistics with an average for the $N = 100$ replications, proportions were also calculated that indicate the proportion of times each variable selection method was able to achieve the desired result, indicated by an adjusted Rand index of one, or by giving estimates of u and k equal to the population values. Full tables of these statistics can be seen in Appendix B.1.2.

3.3.2.1 Procrustes Transformation of $\hat{\mathbf{L}}$

One more statistic of interest in the simulation was a comparison between $\mathbf{L}$ and $\hat{\mathbf{L}}$ to determine how well the pattern of factor loadings was recovered in the estimate.

Due to the rotational indeterminacy of $\mathbf{L}$, it is desirable to rotate $\hat{\mathbf{L}}$ so that is looks as similar as possible to $\mathbf{L}$ before calculating a statistic such as a mean squared error. This rotation can be accomplished using the Procrustes transformation (Hurley and Cattell, 1962; Schönemann, 1966). Consider two matrices $\mathbf{A}$ and $\mathbf{B}$, both of dimension $p \times k$. Alternatively, if, for example, the number of columns in $\mathbf{A}$ is smaller than in $\mathbf{B}$, simply add columns of zeros to $\mathbf{A}$ so the dimensions of $\mathbf{A}$ and $\mathbf{B}$ are the same (Johnson and Wichern, 2007, p. 733).

Then the Procrustes transformation can be used to rotate **A** so that it most closely aligns with **B**. The orthogonal rotation matrix **R** is defined as

$$\underset{k\times k}{\mathbf{R}} = \arg\min_{\mathbf{\Omega}} ||\mathbf{A\Omega} - \mathbf{B}||_F \text{ subject to } \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}, \tag{3.3.1}$$

where F is the Frobenius norm, defined as

$$||\mathbf{X}||_F = \sum_{i,j}(X_{ij})^2 \tag{3.3.2}$$

for the individual elements X_{ij} of **X**. The solution, given in Schönemann (1966), and shown in an alternate form in (Johnson and Wichern, 2007, pp. 733-735), is given by the singular value decomposition of

$$\underset{k\times k}{\mathbf{M}} = \mathbf{A}^T\mathbf{B} \tag{3.3.3}$$

$$= \mathbf{U\Lambda V}^T. \tag{3.3.4}$$

Then $\mathbf{R} = \mathbf{UV}^T$ and **A** can be rotated to $\mathbf{A}_p = \mathbf{AR}$ to most closely match **B**, as measured by the Frobenius norm.

In rotating $\hat{\mathbf{L}}$ to most closely match **L**, define $\mathbf{A} = \hat{\mathbf{L}}$ and $\mathbf{B} = \mathbf{L}$. Then $\mathbf{M} = \hat{\mathbf{L}}^T\mathbf{L}$, and singular value decomposition of **M** gives **R**. Then the rotation of the estimated loading matrix that most closely maps to **L** is $\hat{\mathbf{L}}_p = \hat{\mathbf{L}}\mathbf{R}$, and the minimized distance between **L** and $\hat{\mathbf{L}}_p$ is given as

$$F = ||\hat{\mathbf{L}}_p - \mathbf{L}||_F. \tag{3.3.5}$$

Since **R** is orthogonal, we retain the property that $\hat{\mathbf{L}}_p\hat{\mathbf{L}}_p^T = \hat{\mathbf{L}}\mathbf{R}\mathbf{R}^T\hat{\mathbf{L}}_p^T = \hat{\mathbf{L}}\mathbf{I}\hat{\mathbf{L}}_p^T = \hat{\mathbf{L}}\hat{\mathbf{L}}^T$.

The value of the Frobenius norm is divided by the number of elements (pk) and then used as the measure of similarity between $\hat{\mathbf{L}}_p$ and **L**. This is referred to in the results as the mean squared error (MSE) between $\hat{\mathbf{L}}_p$ and **L**, where MSE $= F/(pk)$.

3.4 Results

Summaries of the four key statistics of interest (estimated ARI, u, k, and MSE) for each of the 36 simulation cases are given in Appendix B.1. In particular, Appendix B.1.1 gives figures that easily display differences between selection methods, and Appendix B.1.2 gives tables of the specific values corresponding to each figure, along with the proportion of correct cases.

To assist in comparisons, an average standard error is given at after the last row in each table, just as in the simulations in Chapter 2. Standard error is calculated as $SE = s/\sqrt{N}$, where s is the standard deviation of a particular statistic (such as ARI) for the $N = 100$ replications in each combination of factors. Average standard error is therefore the average among the 33 standard errors given for each variable selection method, for each statistic of interest, at each combination of factors.

3.4.1 Top Methods

From the figures and tables in Appendix B.1, it was clear that certain variable selection methods were not competitive with other approaches. In general, the likelihood ratio (LR) and squared multiple correlation (SMC) ordering methods did not perform as well as other pre-screening methods. They tended to classify too many variables as uncorrelated, regardless of the use of $\mathbf{R}$ or $\mathbf{R}_2$. They also showed higher variation in estimating u in relation to the stopping criterion used, which led to errors in the number of factors selected as well.

However, the LR and SMC ordering methods still performed better overall than the traditional communality rule and GLAFA. The communality rule tended to greatly overestimate the number of uncorrelated variables, leading to low ARI values and underestimation of the number of factors. GLAFA was prone to large MSE values, overestimation of the number of uncorrelated variables, and inaccurate estimates of the number of factors even in cases where other methods had no issues (see Section 3.4.2 for more). Therefore, these cases were excluded from further analysis.

Summarized results for the remaining top-performing methods are shown in Figures 3.1 and 3.2. The horizontal axis displays the six remaining ordering method combinations, as well as the traditional magnitude rule. Each figure is a summary of 18 simulation cases

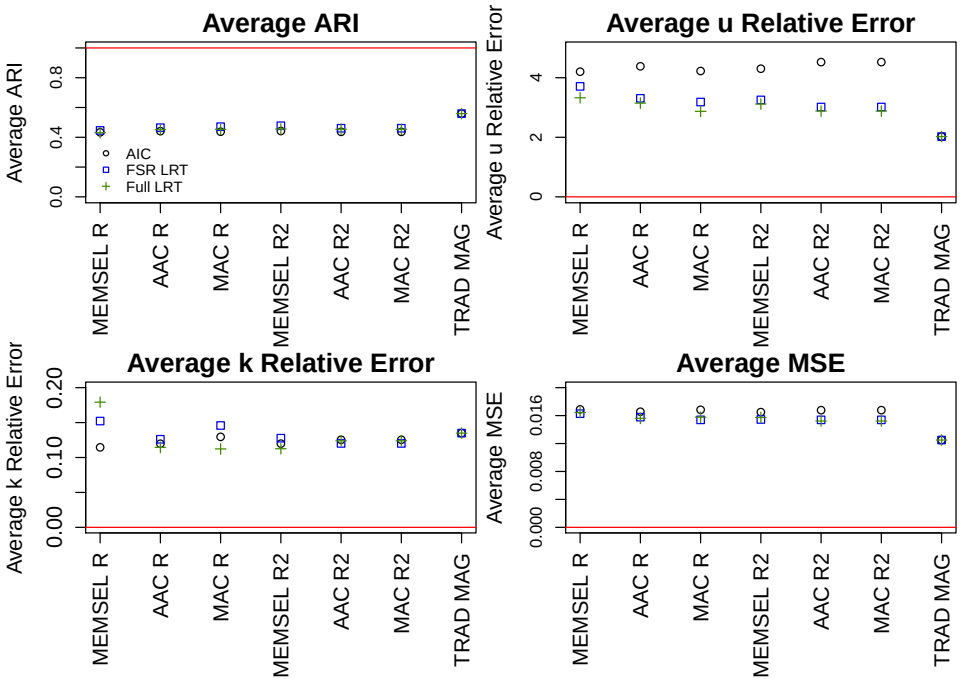


Figure 3.1 Simulation results for variable selection methods, averaged across all cases with $n = 100$. ARI= adjusted Rand index, u = number of uncorrelated variables, k = number of factors, MSE is between $\mathbf{L}$ and $\hat{\mathbf{L}}$. Red lines indicate desired values. Each point is an average of 100 replications $*$ 18 simulation cases.

for the same sample size. For each set of plots, ARI and MSE are averaged across the 18 simulation cases in each sample size. Because each of the 18 cases did not utilize the same population values of u and k, the corresponding plots display average relative error across the 18 cases, with an ideal value of zero. For each case, the relative error of u is calculated as the absolute value of the difference between the population value of u and the average estimated value of u across the $N = 100$ replications, divided by the population value of u. That is, $\text{RE}_u = \text{abs}(\hat{u} - u)/u$ where $\hat{u}$ is the average from the $N = 100$ replications. Then the 18 relative errors for each variable selection method at a given sample size are averaged and displayed. Relative error for k is calculated similarly.

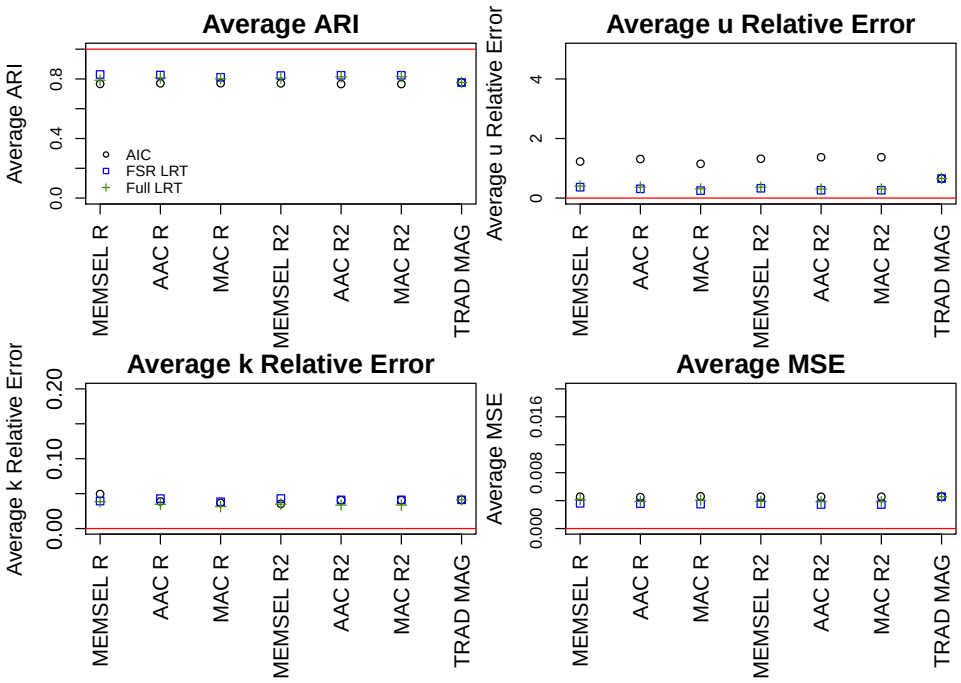


Figure 3.2 Simulation results for variable selection methods, averaged across all cases with $n = 400$. ARI= adjusted Rand index, u = number of uncorrelated variables, k = number of factors, MSE is between $\mathbf{L}$ and $\hat{\mathbf{L}}$. Red lines indicate desired values. Each point is an average of 100 replications $*$ 18 simulation cases.

3.4.2 Adaptations to GLAFA

The GLAFA method unexpectedly showed poor performance for determining the correct number of factors in 23 of the 36 simulation cases. GLAFA tended to greatly overestimate k; see Appendix B.2 for figures of original GLAFA results. For example, Figure B.38 shows that for $k = 2$, $u = 10$, and 9 correlated variables per factor, GLAFA overestimated k on average in every simulation case regardless of sample size or magnitude of factor loadings. The average selected k was between 5 and 17, rather than 2. Even after creating a wider and more dense grid of θ values, the same results were found. Recall that k is determined by finding the minimum GBIC value, so this appears to be an issue with how the GBIC is calculated; see Hirose and Konishi (2012) for lengthy calculation details. This type of finding was not discussed in the original results from Hirose and Konishi (2012).

Further, when viewing the loadings matrix of one of these cases with an extreme number of factors, it is evident that many of the extra columns include loadings significantly different from zero (in both the rotated and non-rotated versions). Therefore, there is not a simple fix such as removing any columns with all elements close to zero.

Notably, the GLAFA estimates of u and ARI remained competitive with other variable selection methods, so it appears that the GBIC calculation prioritizes these estimates over k. We also found that GLAFA tended to have larger average MSE values that other selection methods, although this finding was confirmed in the original simulation study by Hirose and Konishi (2012). They also found that GLAFA showed higher values of MSE than traditional maximum likelihood factor analysis, likely due to high regularization parameters causing factor loading estimates to be attenuated. However, they surmise that even attenuated loadings can still be interpreted, and perhaps it is more important to select the correct model in terms of k and m, rather than to obtain a smaller MSE.

Due to the large amount of distortion in estimates of k, we propose an adaptation of GLAFA that preserves the classifications of u and m, but re-estimates the number of factors k based on the BIC. AIC was also considered but caused no significant improvements over BIC; see Appendix B.2. The algorithm is as follows:

1. Consider the varimax-rotated $\hat{\mathbf{L}}$ matrix from GLAFA with $\hat{m}$ rows and $\hat{k}_0$ columns.
2. Sort $\hat{\mathbf{L}}$ by column with maximum absolute value. Call this matrix $\hat{\mathbf{L}}_0$.

3. Calculate $\text{BIC}_0 = -2l(\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}_0\hat{\mathbf{L}}_0^T + \mathbf{D}_{\hat{\boldsymbol{\Psi}}_0}) + [\hat{m}\hat{k}_0 - \hat{k}_0(\hat{k}_0 - 1)/2 + \hat{m}]\log(n)$, where $\hat{\boldsymbol{\Psi}}_0$ is determined by $\hat{\mathbf{L}}_0$ since standardization is assumed.

4. Set $\hat{\mathbf{L}}_1$ as $\hat{\mathbf{L}}_0$ with the last column removed. Define $\hat{k}_1 = \hat{k}_0 - 1$.

5. Calculate $\text{BIC}_1 = -2l(\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}_1\hat{\mathbf{L}}_1^T + \mathbf{D}_{\hat{\boldsymbol{\Psi}}_1}) + [\hat{m}\hat{k}_1 - \hat{k}_1(\hat{k}_1 - 1)/2 + \hat{m}]\log(n)$.

6. Continue iteratively until there only one column remains.

7. Choose value of $\hat{k}_i$ that gives the minimum BIC.

This way, the estimated numbers of correlated and uncorrelated variables, and Rand index, will not change, but a better estimate of the number of factors should be achieved. MSE is also affected by the change.

All GLAFA simulation results were post-processed using this method. Appendix B.2 gives figures of original versus BIC and AIC revised estimates of the number of factors. These tables show that although the post-processing approach provides a great improvement in the estimate of k when the original value was large, it can also incorrectly reduce the number of factors for some simulation cases where the original GLAFA method was accurate. See Figure B.39 for a few examples of this. Further, the post-processing procedure led to increases in the average MSE, which is not ideal. Therefore, even this post-processing BIC procedure does not solve all of GLAFA's problems.

The results for GLAFA in the full simulation results figures and tables in Appendix B.1 are the post-processed BIC results rather than the original results, since they are at least preferred over the original GLAFA results.

3.5 Discussion

Simulation results have shown that the SMC and LR ordering methods, as well as the traditional communality rule and GLAFA, do not perform as well as other approaches for determining the correct subset of variables in factor analysis. However, many of the other methods tested are promising.

Figures 3.1 and 3.2 summarize the better-performing variable selection methods in factor analysis: the traditional magnitude rule, and all pre-screening methods with either

MEMSEL, AAC, or MAC ordering. In general, all methods perform better at the higher sample size, as expected. For smaller sample sizes, the traditional magnitude rule for selecting variables generally performs well in comparison to the pre-screening methods. The magnitude rule tends to give a more accurate estimate of u and better ARI and MSE than pre-screening methods. On average, the pre-screening methods tend to overestimate the number of uncorrelated variables as compared to the traditional magnitude rule. However, when sample size is large, pre-screening methods that utilize an LRT-type stopping criterion give a more accurate measure of u than the traditional magnitude rule, and also a slightly higher ARI on average.

Figures 3.1 and 3.2 show that regardless of ordering method, on average the AIC stopping criterion estimates too many variables as uncorrelated. Other than AIC results, all other pre-screening methods appear to perform similarly at the larger sample size on average.

For sample size $n = 100$, ignoring AIC results, we see that the use of $\mathbf{R}$ rather than $\mathbf{R}_2$ leads to variation in k according to the stopping criterion used. These variations also tend to be further from the true value of k. On average, the use of $\mathbf{R}_2$ also appears to give slightly more accurate estimates of u and higher ARI than the use of $\mathbf{R}$. Therefore, we recommend the use of $\mathbf{R}_2$ when conducting ordering of variables.

When analyzing the figures of 36 cases and focusing on MEMSEL, AAC, and MAC ordering with $\mathbf{R}_2$, we see that when Full LRT and FSR LRT differ, FSR LRT generally has higher average ARI values, and tends to estimate a smaller value of u than Full LRT; see Figure B.35 for an example of both. Recall that underestimating u is generally preferred to overestimating u since this is a pre-screening method and it is undesirable to remove too many variables prior to conducting factor analysis. However, when differences between Full LRT and FSR LRT are present, FSR LRT tends to have larger average MSE values than Full LRT (See Figure B.22).

It can also be seen in the figures from the 36 cases that, on average, AAC and MAC have values of ARI, u, and k closer to optimal values more often than MEMSEL ordering does. Any differences between AAC and MAC ordering methods are small, and either might be used with a similar level of accuracy. Regarding MSE statistics, there are few noticeable differences between methods

In most cases, the traditional magnitude rule appears to work well. However, pre-

screening methods are more accurate than the traditional magnitude rule in most cases where low population factor loadings are combined with a large sample size; see Figure B.5 for an example. In other cases with low loadings and small sample size, though, pre-screening methods tend to classify too many variables as uncorrelated, whereas the magnitude approach is much more accurate (also indicated by a higher adjusted Rand index). See Figure B.8 for another example.

Based on these summaries, we recommend the use of AAC or MAC orderings, paired with $\mathbf{R}_2$ and either Full LRT or FSR LRT as stopping criterion, if a pre-screening method is desired. However, the traditional magnitude rule performs well across most cases and should also be considered. The new pre-screening methods are competitive with the traditional magnitude rule, but not uniformly better.

CHAPTER

4

EXAMPLES

Introduction

In Chapter 3, a simulation study was presented that compares various variable selection methods for factor analysis. This chapter presents examples of a few of the techniques applied to real data, along with discussion of issues that may be encountered in practice.

4.1 Job Applicants Data

Hirose and Konishi (2012) used job application data found in Kendall (1980) as an example of their GLAFA method compared to maximum likelihood factor analysis. For each of 48 applicants, the dataset contains 15 scores on measures such as honesty, drive, and potential. Traditional maximum likelihood factor analysis, with BIC as the criterion for determining k, finds a a four-factor solution. The varimax-rotated estimated loadings can be seen in

Table 4.1 $\hat{\mathbf{L}}$ matrix for Applicants data, estimated from maximum likelihood factor analysis and rotated by varimax criterion. BIC and parallel analysis with 95th %ile and principal axis factoring for eigenvalues both determined the number of factors to be 4, leading to the same solution.

	Factor1	Factor2	Factor3	Factor4
Form of Letter Application	0.13	0.72	0.11	-0.12
Appearance	0.45	0.14	0.25	0.16
Academic Ability	0.07	0.13	0.00	0.68
Likeability	0.23	0.24	0.83	-0.05
Self Confidence	0.92	-0.10	0.15	-0.09
Lucidity	0.84	0.11	0.29	0.06
Honesty	0.25	-0.22	0.74	-0.02
Salesmanship	0.89	0.24	0.08	-0.07
Experience	0.09	0.78	-0.05	0.17
Drive	0.76	0.39	0.18	-0.06
Ambition	0.90	0.19	0.11	-0.06
Grasp	0.78	0.28	0.36	0.15
Potential	0.73	0.35	0.44	0.25
Keenness to Join	0.42	0.39	0.56	-0.59
Suitability	0.36	0.77	0.05	0.14

Table 4.1. After varimax rotation, all variables have at least one absolute loading greater than $c = 0.32$, so all variables are retained based on the traditional magnitude rule. However, five of the variables cross-load in this solution (Drive, Grasp, Potential, Keenness to Join, and Suitability).

Using the GLAFA method, Hirose and Konishi (2012) found a three-factor solution for this dataset, and the Academic Ability variable was estimated to have loadings all equal to zero, indicating it should be removed from the model. We confirmed this finding with our own use of the GLAFA package, which found a minimum GBIC at $k = 3$ and $\theta = 0.056$ and gave almost identical loadings as found in Hirose and Konishi (2012). These findings are given in Table 4.2. Hirose and Konishi (2012) claim that the three-factor solution is more interpretable than the four-factor solution, with factors of *Career and Adequacy, Motivation and Ability,* and *Character*.

We applied a pre-screening method on the same dataset. The average absolute correlation was paired with $\mathbf{R}_2$ as an ordering method, and FSR LRT was used as a stopping

Table 4.2 $\hat{\mathbf{L}}$ matrix for Applicants data, estimated from GLAFA with $\theta = 0.056$ and rotated by varimax criterion.

	Factor1	Factor2	Factor3
Form of Letter Application	0.54	0.09	0.12
Appearance	0.13	0.31	0.19
Academic Ability	0	0	0
Likeability	0.17	0.10	0.87
Self Confidence	-0.14	0.78	0.10
Lucidity	0.07	0.68	0.25
Honesty	-0.19	0.18	0.58
Salesmanship	0.18	0.73	0.08
Experience	0.68	0.05	-0.05
Drive	0.30	0.62	0.12
Ambition	0.10	0.74	0.09
Grasp	0.21	0.64	0.24
Potential	0.29	0.57	0.34
Keenness to Join	0.20	0.32	0.48
Suitability	0.73	0.26	0.05

criterion. We found that no variables were estimated as uncorrelated, and therefore retained all 15 variables when running factor analysis. Parallel analysis with the 95th percentile and eigenvaules from principal axis factoring was conducted, which resulted in $k = 4$ factors being retained. Then maximum likelihood estimation with varimax rotation was used to estimate $\mathbf{L}$, leading to the same solution found in Table 4.1.

The four-factor solution shows that Factor 4 has only two loadings larger than $c = 0.32$. Therefore, per the guidelines in Section 1.5, these two variables should either be removed, or more variables should be added to give a minimum of three to five variables per factor. Therefore, even though traditional methods and pre-screening methods do not indicate that any variables are uncorrelated with all others, the estimated factor model still leaves room for improvement.

4.2 T-STEM Data

The T-STEM Science questionnaire asks K-12 teachers about their perceived self-efficacy in teaching science, as well as their belief that educators affect student learning (Unfried et al., 2016). This unpublished instrument is an adaptation of the Science Teaching Efficacy Belief Instrument (STEBI; Riggs and Enochs, 1990). The pilot T-STEM Science instrument used the STEBI survey items directly. Due to the dated nature of the original STEBI instrument, when the pilot T-STEM Science instrument was administered in 2011, it was expected that some survey items might not load on factors as expected. The pilot T-STEM Science survey results are used here as another example of variable selection techniques in factor analysis. The full set of pilot T-STEM Science items can be seen in Appendix A.4. The pilot instrument was later revised and can be requested online at `http://miso.ncsu.edu/articles/t-stem-survey-2`.

After data cleaning, the pilot administration resulted in 133 usable responses on 24 items such as "I know the steps necessary to teach science concepts effectively," and "If students are underachieving in science, it is most likely due to ineffective science teaching." Negatively-worded items were reverse-coded prior to analysis.

When using the traditional magnitude of factor loadings approach to analyze these data, parallel analysis with principal axis factoring and the 95th percentile rule retained four factors. Based on the estimated loadings matrix, variables 12 and 20 were removed since they did not have any absolute loading greater than $c = 0.32$ on any factor. Then, parallel analysis was conducted again, resulting in three factors. The final maximum likelihood factor analysis solution based on the traditional magnitude rule is given in Table 4.3. In Table 4.3, we see that V5 and V9 now also do not have any absolute factor loadings above $c = 0.32$. The traditional magnitude rule could be applied again if desired, although this was not done in the simulation in Chapter 3. Variables V7, V13, V14, and V23 cross-load.

Pre-screening methods were used in the same manner as the Applicants data, with AAC ordering paired with $\mathbf{R}_2$, and FSR LRT stopping criterion. Pre-screening led to the deletion of two variables, V9 and V19. Note that these are two different variables than those identified by the traditional magnitude rule. After the two variables were removed, parallel analysis was conducted and suggested five factors. The resulting matrix of factor loadings

Table 4.3 $\hat{\mathbf{L}}$ matrix for pilot T-STEM Science data, after the traditional magnitude rule was applied, removing V12 and V20. Results were estimated from maximum likelihood factor analysis and rotated by varimax criterion. Number of factors was chosen by parallel analysis with 95th %ile and principal axis factoring for eigenvalues.

	Factor1	Factor2	Factor3
V1	-0.07	0.57	0.03
V2	0.25	0.42	0.15
V3	-0.05	0.58	0.13
V4	0.67	0.21	-0.03
V5	0.26	-0.16	0.30
V6	-0.21	0.24	0.37
V7	0.42	-0.16	0.38
V8	0.15	0.21	0.35
V9	-0.02	0.08	0.23
V10	0.16	0.71	0.10
V11	0.81	0.16	-0.08
V13	0.00	0.36	0.47
V14	0.10	0.53	0.33
V15	0.19	0.59	0.04
V16	0.71	-0.01	0.17
V17	0.66	0.16	-0.06
V18	0.73	-0.04	0.00
V19	0.07	0.06	0.47
V21	0.70	-0.04	0.13
V22	0.73	0.16	-0.03
V23	0.51	0.11	0.35
V24	0.47	0.05	0.27

is given in Table 4.4.

Although the traditional magnitude rule results in three factors and the pre-screening method results in five factors, each method results in three factors that can be similarly interpreted as 1) Confidence in teaching science, 2) Positive results of teacher effort and performance, and 3) Lack of responsibility for underachieving students. However, the five-factor solution additionally has constructs 4) Lack of confidence in teaching science, broken out from the first factor, and 5) Resistance to observation. This fifth factor is primarily defined by variable V20, which was removed from the dataset in the traditional magnitude solution.

This example shows that the particular subset of variables that is used to conduct parallel analysis can have a significant impact on the number of factors retained. Tables 4.3 and 4.4 differ by only two variables, yet come to different forms of $\hat{\mathbf{L}}$. Further, the difference between the use of eigenvalues from principal components and eigenvalues from principal axis factoring is notable. If the principal components eigenvalues are used in parallel analysis after pre-screening methods are applied, then only two factors are retained instead of five (results not shown). Principal components eigenvalues are perhaps more widely used than principal axis factoring eigenvalues, but the principal axis factoring approach is growing more popular, as discussed in Section 1.4.5.

4.3 S-STEM Data

The pre-screening methods studied in Chapters 2 and 3 have focused on moderate sample sizes from $n = 100$ to $n = 400$. In their review of psychological literature, Conway and Huffcutt (2003) found that almost 69% of the studies reviewed conducted factor analysis with samples of less than $n = 400$, with 46% of all studies using less than $n = 200$. However, although we have not studied it explicitly in this dissertation, the performance of variable selection methods for larger sample sizes is of interest as well.

The S-STEM middle/high (MH) questionnaire surveys 6-12th grade students on their attitudes toward science, technology, engineering, and mathematics (STEM), as well as 21st century learning skills (Unfried et al., 2015). The survey contains 43 items such as "I know I can do well in science" and "I believe I can be successful in a career in engineering."

Table 4.4 $\hat{\mathbf{L}}$ matrix for pilot T-STEM Science data, after applying the AAC, $\mathbf{R_2}$, and FSR LRT pre-screening method, removing V9 and V19. Results were estimated from maximum likelihood factor analysis and rotated by varimax criterion. Number of factors was chosen by parallel analysis with 95th %ile and principal axis factoring for eigenvalues.

	Factor1	Factor2	Factor3	Factor4	Factor5
V1	-0.07	0.56	0.03	-0.01	0.02
V2	0.25	0.40	0.17	0.07	-0.04
V3	-0.03	0.53	0.16	-0.08	-0.07
V4	0.68	0.14	0.22	-0.06	0.13
V5	0.12	0.00	-0.05	0.72	0.05
V6	-0.19	0.20	0.30	0.02	-0.30
V7	0.30	-0.07	0.18	0.50	0.05
V8	0.12	0.18	0.42	0.08	-0.15
V10	0.15	0.77	0.16	-0.08	0.17
V11	0.82	0.11	0.00	0.08	0.07
V12	0.01	0.21	0.45	-0.06	0.16
V13	-0.03	0.40	0.28	0.20	-0.25
V14	0.15	0.51	0.24	0.06	-0.47
V15	0.20	0.66	-0.06	0.02	0.00
V16	0.65	0.00	0.16	0.30	0.06
V17	0.71	0.10	0.00	0.03	-0.13
V18	0.67	0.01	-0.03	0.27	0.22
V20	0.11	0.07	0.10	0.13	0.57
V21	0.61	0.01	0.15	0.27	0.27
V22	0.77	0.09	0.05	0.03	-0.10
V23	0.42	0.14	0.37	0.25	0.17
V24	0.43	-0.04	0.58	0.06	0.01

Unfried et al. (2015) conducted exploratory factor analysis on a sample of 8,659 middle and high school students across North Carolina. The authors found a four-factor solution with one variable that did not have any absolute factor loadings greater than $c = 0.32$, along with a few other variables with loadings that were borderline.

When the AAC, $\mathbf{R}_2$, FSR LRT pre-screening method is applied, zero variables are classified as uncorrelated in the model. The GLAFA approach also fails to identify any variables to be removed from the model.

Another version of the S-STEM survey was developed for upper elementary (UE) school students, grades 4 and 5 (Unfried et al., 2015). Factor analysis was conducted with a sample of 768 upper elementary school students from across North Carolina. Unfried et al. (2015) again found a four-factor solution, but in this instance, four variables did not have any absolute loadings above $c = 0.32$. In contrast, both GLAFA and pre-screening methods again fail to identify any variables that should be removed from the factor model.

These two S-STEM examples indicate that the use of pre-screening methods on large sample sizes can cause stopping criteria to be overly sensitive to changes in the estimated correlation matrix. The large sample sizes lead to extremely small p-values, causing the methods to reject any form of a constrained correlation matrix as presented in Section 2.3. Adapting the pre-screening methods to work appropriately with larger sample sizes is an area of future work.

CHAPTER

5

CONCLUSION

This dissertation has presented a new approach to variable selection in factor analysis. The goal of factor analysis is to summarize the covariance structure of a set of variables by a set of factors, where each factor represents multiple variables. Crucial decisions in this process are the estimation method, determining the number of factors to retain, and rotating the estimated loadings matrix to an interpretable solution. Based on the literature, we generally recommend parallel analysis with the 95th percentile approach for determining the number of factors, maximum likelihood factor analysis for model estimation, and varimax rotation for finding an interpretable solution.

Another key issue in factor analysis, as well as the focus of this dissertation, is the idea of variable selection. Since factor analysis is focused on representing each group of correlated variables with one factor, variables that are lowly correlated with all other variables should be identified and either removed from the model or developed into a new factor with additional variables. The most common traditional approach to variable selection in factor

analysis is based on the magnitude of factor loadings, although it can be a subjective process in practice. Recently, researchers have studied new techniques such as a stepwise approach (Kano and Harada, 2000) and GLAFA, a weighted group lasso approach (Hirose and Konishi, 2012).

We have proposed a new approach to variable selection in factor analysis that pre-screens variables according to their correlation, prior to estimating the factor model. In this two-step procedure, variables are first ordered by some measure of correlation. Given an ordering, stopping criteria are used to determine the appropriate number of variables to estimate as uncorrelated. These uncorrelated variables are removed from the dataset prior to conducting factor analysis. Although several pre-screening approaches generally perform similarly, we recommend either the Average Absolute Correlation (AAC) or Maximum Absolute Correlation (MAC) methods for ordering variables. Each of these ordering methods can be implemented with either the sample correlation $\mathbf{R}$ or the Hadamard product of $\mathbf{R}$, defined as $\mathbf{R}_2$. For stopping criteria, we recommend either the Full likelihood ratio test (Full LRT), or the likelihood ratio test that controls the false selection rate (FSR LRT). Our simulations show that these recommended pre-screening approaches outperform GLAFA and are competitive with the traditional magnitude of factor loadings rule, thought not uniformly better.

Future work on variable selection in factor analysis might be conducted in a couple of areas. The proposed pre-screening methods might be overly sensitive for large sample sizes, so alternative stopping criteria should be developed. Further, parallel analysis only appears to be studied up to a sample size of around $n = 500$ (Crawford et al., 2010; Green et al., 2012). Another consideration is how exactly to define a lowly correlated variable. In our simulations, we only sought to remove variables that had correlations of zero with all other variables at the population level. However, in factor analysis, a simpler solution is sought even at the cost of removing "lowly" correlated variables. It is desirable for each variable to be strongly associated with its corresponding factor, and therefore with other variables loading on the same factor. Pre-screening methods could search for lowly correlated variables rather than just uncorrelated variables, although the definition of "low" is debatable. This might include the use of stopping criteria that are not dependent on p-values for solutions. Pre-screening methods should also be further tested with other forms of parallel analysis,

such as the use of eigenvalues from the principal components solution rather than from the principal axis factoring solution.

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, eds. Petrox, B. N. and Caski, F., Budapest, Hungary: Akademiai Kiado, p. 267.

Bartlett, M. S. (1954), "A Note on the Multiplying Factors for Various Chi-Square Approximations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 16, 296–298.

Boos, D. D., Stefanski, L. A., and Wu, Y. (2009), "Fast FSR variable selection with applications to clinical trials," *Biometrics*, 65, 692–700.

Browne, M. and Cudeck, R. (1992), "Alternative ways of assessing model fit," *Sociological Methods & Research*, 21, 230–258.

Cattell, R. B. (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, 1, 245–276.

Chi, J. T., Chi, E. C., and Baraniuk, R. G. (2016), "k-POD: A Method for k-Means Clustering of Missing Data," *The American Statistician*, 70, 91–99.

Child, D. (2006), *The Essentials of Factor Analysis*, New York, NY: Continuum International Publishing Group, 3rd ed.

Choi, J., Zou, H., and Oehlert, G. (2010), "A Penalized Maximum Likelihood Approach to Sparse Factor Analysis," *Statistics and Its Interface*, 0, 1–8.

Comrey, A. L. (1978), "Common methodological problems in factor analytic studies," *Journal of Consulting and Clinical Psychology*, 46, 648–659.

Conway, J. M. and Huffcutt, A. I. (2003), "A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research," *Organizational Research Methods*, 6, 147–168.

Costello, A. B. and Osborne, J. W. (2005), "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis," *Practical Assessment, Research & Evaluation*, 10, 1–9.

Crawford, A. V., Green, S. B., Levy, R., Lo, W. J., Scott, L., Svetina, D., and Thompson, M. S. (2010), "Evaluation of Parallel Analysis Methods for Determining the Number of Factors," *Educational and Psychological Measurement*, 70, 885–901.

Crawford, C. B. and Koopman, P. (1973), "A Note on Horn's Test for the Number of Factors in Factor Analysis," *Multivariate Behavioral Research*, 8, 117–125.

Davis, C. (1962), "The Norm of the Schur Product Operation," *Numerische Mathematik*, 4, 343–344.

de Winter, J. C. F. and Dodou, D. (2012), "Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size," *Journal of Applied Statistics*, 39, 695–710.

Dinno, A. (2009), "Exploring the Sensitivity of Horn's Parallel Analysis to the Distributional Form of Random Data," *Multivariate Behavioral Research*, 44, 362–388.

Fabrigar, L. R., Wegener, D. T., Maccallum, R. C., and Strahan, E. J. (1999), "Evaluating the Use of Exploratory Factor Analysis in Psychological Research," *Psychological Methods*, 4, 272–299.

Fava, J. L. and Velicer, W. F. (1992), "The Effects of Overextraction on Factor and Component Analysis," *Multivariate Behavioral Research*, 27, 387–415.

Fawcett, T. (2006), "An Introduction to ROC Analysis," *Pattern Recognition Letters*, 27, 861–874.

Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012), "mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation," Tech. rep., University of Washington, Seattle, WA.

Glorfeld, L. W. (1995), "An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain," *Educational and Psychological Measurement*, 55, 377–393.

Gorsuch, R. L. (1983), *Factor Analysis*, Hillsdale, NJ: Lawrence Erlbaum Associates, 2nd ed.

— (1988), "Exploratory Factor Analysis," in *Handbook of multivariate experimental psychology*, eds. Nesselroade, J. R. and Cattell, R. B., New York, NY: Plenum Press, 2nd ed., pp. 231–258.

— (1990), "Common Factor Analysis versus Component Analysis: Some Well and Little Known Facts," *Multivariate Behavioral Research*, 25.

Green, S. B., Levy, R., Thompson, M. S., Lu, M., and Lo, W.-J. (2012), "A Proposed Solution to the Problem With Using Completely Random Data to Assess the Number of Factors With Parallel Analysis," *Educational and Psychological Measurement*, 72, 357–374.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (2006), *Multivariate Data Analysis*, Upper Saddle River, NJ: Prentice Hall, 6th ed.

Harshman, R. A. and Reddon, J. R. (1983), "Determining the Number of Factors by Comparing Real with Random Data: A Serious Flaw and some Possible Corrections," in *Proceedings of the Classification Society of North America*, Philadelphia, pp. 14–15.

Hirose, K. and Konishi, S. (2012), "Variable selection via the weighted group lasso for factor analysis models," *Canadian Journal of Statistics*, 40, 345–361.

Hirose, K. and Yamamoto, M. (2015), "Sparse estimation via nonconcave penalized likelihood in factor analysis model," *Statistics and Computing*, 25, 863–875.

Hogarty, K., Kromrey, J., Ferron, J., and Hines, C. (2004), "Selection of variables in exploratory factor analysis: An empirical comparison of a stepwise and traditional approach," *Psychometrika*, 69, 593–611.

Horn, J. L. (1965), "A rationale and test for the number of factors in factor analysis," *Psychometrica*, 30, 179–185.

Hubert, L. and Arabie, P. (1985), "Comparing partitions," *Journal of Classification*, 2, 193–218.

Humphreys, L. G. and Montanelli Jr., R. G. (1975), "An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors," *Multivariate Behavioral Research*, 10, 193–205.

Hurley, J. R. and Cattell, R. B. (1962), "The Procrustes Program: Producing Direct Rotation to Test a Hypothesized Factor Structure," *Computers in Behavioral Science*, 7, 258–262.

Johnson, R. A. and Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis (6th Edition)*, Pearson.

Kaiser, H. F. (1958), "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, 23, 187–200.

— (1960), "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, 20, 141–151.

Kano, Y. and Harada, A. (2000), "Stepwise variable selection in factor analysis," *Psychometrika*, 65, 7–22.

Kano, Y. and Ihara, M. (1994), "Identification of Inconsistent Variates in Factor Analysis," *Psychometrika*, 59, 5–20.

Kendall, M. G. (1980), *Multivariate Analysis*, London: Charles Griffin, 2nd ed.

Konishi, S., Ando, T., and Imoto, S. (2004), "Bayesian information criteria and smoothing parameter selection in radial basis function networks," *Biometrika*, 91, 27–43.

Lawley, D. N. and Maxwell, A. E. A. E. (1971), *Factor analysis as a statistical method*, New York: American Elsevier Pub. Co.„ 2nd ed.

Lederman, W. (1937), "On the Rank of the Reduced Correlation Matrix in Multiple-Factor Analysis," *Psychometrika*, 2, 85–93.

Little, T. D., Lindenberger, U., and Nesselroade, J. R. (1999), "On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good," *Psychological Methods*, 4, 192–211.

Lopes, H. F. and West, M. (2004), "Bayesian model assessment in factor analysis," *Statistica Sinica*, 14, 41–67.

Lorenzo-Seva, U., Timmerman, M. E., and Kiers, H. A. L. (2011), "The Hull Method for Selecting the Number of Common Factors," *Multivariate Behavioral Research*, 46, 340–364.

Luo, X., Stefanski, L. a., and Boos, D. D. (2006), "Tuning Variable Selection Procedures by Adding Noise," *Technometrics*, 48, 165–175.

Maccallum, R. C. (1990), "The Need for Alternative Measures of Fit in Covariance Structure Modeling," *Multivariate Behavioral Research*, 25, 157–162.

MacCallum, R. C., Widaman, K. F., Zhang, S., and Hong, S. (1999), "Sample size in factor analysis." *Psychological Methods*, 4, 84–99.

Mavridis, D. and Ntzoufras, I. (2014), "Stochastic search item selection for factor analytic models," *British Journal of Mathematical and Statistical Psychology*, 67, 284–303.

Mulaik, S. A. (2010), *Foundations of Factor Analysis*, Boca Raton: Chapman & Hall, 2nd ed.

Ning, L. and Georgiou, T. T. (2011), "Sparse factor analysis via likelihood and l1-regularization," *IEEE Conference on Decision and Control and European Control Conference*, 5188–5192.

Rand, W. M. (1971), "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, 66, 846–850.

Rencher, A. C. and Christensen, W. F. (2012), *Methods of Multivariate Analysis*, Hoboken NJ: Wiley, 3rd ed.

Riggs, I. M. and Enochs, L. G. (1990), "Toward the Development of an Elementary Teaher's Science Teaching Efficacy Belief Instrument," *Science Education*, 74, 625–637.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011), "pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves." *BMC Bioinformatics*, 12, 1–8.

SAS Institute Inc. (2013), "The FACTOR Procedure," in *SAS/STAT® 13.1 User's Guide*, Cary, NC: SAS Institute Inc., pp. 2371–2382.

Schönemann, P. H. (1966), "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, 31, 1–10.

Schur, J. (1911), "Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen," *Journal für die reine und angewandte Mathematik*, 140, 1–28.

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

Silvester, J. R. . (2000), "Determinants of Block Matrices," *The Mathematical Gazette*, 84, 460–467.

StataCorp (2013), "Multivariate Tests of Covariances," in *Stata 13 Base Reference Manual*, College Station, TX: Stata Press, p. 6.

Stefanski, L. A., Wu, Y., and White, K. (2014), "Variable Selection in Nonparametric Classification via Measurement Error Model Selection Likelihoods," *Journal of the American Statistical Association*, 37–41.

Stevens, J. (1992), *Applied Multivariate Statistics for the Social Sciences*, Hillsdale, NJ: Erlbaum, 2nd ed.

Styan, G. P. H. (1973), "Hadamard Products and Multivariate Statistical Analysis," *Linear Algebra and its Applications*, 6, 217–240.

Tabachnick, B. G. and Fidell, L. S. (2001), *Using multivariate statistics*, Allyn and Bacon, 4th ed.

Thurstone, L. L. (1947), *Multiple-Factor Analysis*, Chicago: University of Chicago Press.

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

Tucker, L. R. and Maccallum, R. C. (1997), "Factor Extraction by Matrix Factoring Techniques," in *Exploratory Factor Analysis*, Unpublished, chap. 8, pp. 180–251.

Unfried, A., Faber, M., Stanhope, D. S., and Wiebe, E. (2015), "The Development and Validation of a Measure of Student Attitudes Toward Science, Technology, Engineering, and Math (S-STEM)," *Journal of Psychoeducational Assessment*, 33, 622–639.

Unfried, A., Faber, M., and Wiebe, E. (2016), "Teacher Efficacy and Attitudes toward STEM (T-STEM) Survey: Development and Psychometric Properties," Tech. rep., North Carolina State University, Raleigh, NC.

Velicer, W. F. (1976), "The Relation Between Factor Score Estimates, Image Scores, and Principal Component Scores," *Educational and Psychological Measurement*, 36, 149–159.

Velicer, W. F. and Fava, J. L. (1998), "Effects of Variable and Subject Sampling on Factor Pattern Recovery Pattern Recovery," *Psychological Methods*, 3, 231–251.

Velicer, W. F. and Jackson, D. (1990), "Component Analysis Versus Common Factor-Analysis: Some Issues in Selecting an Appropriate Procedure," *Multivariate Behavioral Research*, 25, 1–28.

Widaman, K. F. (1993), "Common Factor Analysis Versus Principal Component Analysis: Differential Bias in Representing Model Parameters?" *Multivariate Behavioral Research*, 28, 263–311.

Wood, J. M., Tataryn, D. J., and Gorsuch, R. L. (1996), "Effects of under- and overextraction on principal axis factor analysis with varimax rotation." *Psychological Methods*, 1, 354–365.

Wu, Y., Boos, D. D., and Stefanski, L. a. (2007), "Controlling Variable Selection by the Addition of Pseudovariables," *Journal of the American Statistical Association*, 102, 235–243.

Yuan, M. and Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, 68, 49–67.

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.

Zwick, W. R. and Velicer, W. F. (1986), "Comparison of five rules for determining the number of components to retain." *Psychological Bulletin*, 99, 432–442.

APPENDICES

APPENDIX

A

ADDITIONAL FINDINGS AND PROOFS

A.1 R^2 from Correlation Matrix

R^2 can be calculated from the correlation matrix as follows. Consider the partition $\mathbf{W} = (Y, \mathbf{X})^T$ where $Y = W_1$ and $\mathbf{X} = (W_2, \ldots, W_p)^T$. If we assume that $\mathbf{W}$ follows a multivariate normal distribution, then

$$\begin{pmatrix} Y \\ \mathbf{X} \end{pmatrix} \sim \text{MVN}\left\{ \begin{pmatrix} \mu_Y \\ \boldsymbol{\mu}_X \end{pmatrix}, \begin{pmatrix} \Sigma_Y & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_X \end{pmatrix} \right\}.$$

Further, the conditional distribution of Y given $\mathbf{X}$ is defined as

$$Y|\mathbf{X} \sim \text{MVN}(\mu_{Y|\mathbf{X}}, \Sigma_{Y|\mathbf{X}}) \tag{A.1.1}$$

where

$$\mu_{Y|\mathbf{X}} = \mu_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}(\mathbf{X} - \boldsymbol{\mu}_X) \tag{A.1.2}$$

and

$$\Sigma_{Y|\mathbf{X}} = \Sigma_Y - \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}. \tag{A.1.3}$$

We know that R^2 can be defined as

$$\begin{aligned} R^2 &= 1 - \frac{SS_{resid}}{SS_{tot}} && \text{(A.1.4)} \\ &= 1 - \frac{E\{Y - E(Y|\mathbf{X})\}^2}{E(Y^2)} && \text{(A.1.5)} \end{aligned}$$

By definition, the numerator $E\{Y - E(Y|\mathbf{X})\}^2 = \text{Cov}(Y|\mathbf{X}) = \Sigma_{Y|\mathbf{X}}$. Further, for our work we always assume $\mu_{W_i} = 0$ and $\Sigma_{W_i} = 1 \; \forall i$. Therefore the denominator, $E(Y^2) = E(Y - \mu_Y)^2. = \Sigma_Y = 1$. Then we have

$$\begin{aligned} R^2 &= 1 - \frac{E\{Y - E(Y|\mathbf{X})\}^2}{E(Y^2)} && \text{(A.1.6)} \\ &= 1 - \frac{(1 - \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY})}{1} && \text{(A.1.7)} \\ &= \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY} && \text{(A.1.8)} \end{aligned}$$

Therefore R^2 can be calculated entirely using only Σ_W. Also note that if desired, Adjusted R^2 can be defined as

$$Adj.R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{A.1.9}$$

Now this proof only shows how R^2 can be found for one W_1. The derivation could be repeated iteratively for all p variables, but a simpler calculation to find the set of p R^2 values is derived in Tucker and Maccallum (1997). The authors give an alternative notation for Equation (A.1.8) on p. 194, and further show how this formulation leads to Equation (1.3.6).

A.2 Likelihood Function Derivation

Using observed data $\mathbf{X}_1, \ldots, \mathbf{x}_n$ on p variables, the goal of analysis is to estimate the factor loadings L_{ij} and the specific variances ψ_i assuming the model $\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\epsilon}$. One method to do so is the maximum likelihood method. This method assumes that the common factors $\mathbf{F}_j$ and the specific factors $\boldsymbol{\epsilon}_j$ are normally distributed, meaning $\mathbf{F}_j \sim N(0, \mathbf{I})$ and $\boldsymbol{\epsilon}_j \sim N(0, \mathbf{D_v})$. Given the assumptions, this implies that $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Therefore, the density for one $\mathbf{X}_j$ vector is

$$f_{\mathbf{X}_j}(X) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu}) \right\}, \tag{A.2.1}$$

and the likelihood is therefore

$$\mathrm{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} (\mathbf{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu}) \right\}. \tag{A.2.2}$$

This matches Equation (1.3.7). It is known that for any $k \times k$ matrix $\mathbf{A}$ and $k \times 1$ vector $\mathbf{X}$,

$$\mathbf{X}^T \mathbf{A} \mathbf{X} = \mathrm{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \mathrm{tr}(\mathbf{A} \mathbf{X} \mathbf{X}^T). \tag{A.2.3}$$

Using this result, part of the likelihood becomes

$$\begin{aligned} \sum_{j=1}^{n} (\mathbf{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu}) &= \sum_{j=1}^{n} \mathrm{tr}[(\mathbf{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu})] && \text{(A.2.4)} \\ &= \sum_{j=1}^{n} \mathrm{tr}[\boldsymbol{\Sigma}^{-1} (\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T] && \text{(A.2.5)} \\ &= \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^{n} (\mathbf{X}_j - \boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})^T \right) \right] && \text{(A.2.6)} \end{aligned}$$

Now, if we add and subtract $\bar{\mathbf{X}}$ in each term $(\mathbf{X}_j - \boldsymbol{\mu})$, then we have

$$\sum_{j=1}^{n}(\mathbf{X}_j-\boldsymbol{\mu})(\mathbf{X}_j-\boldsymbol{\mu})^T = \sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}}+\bar{\mathbf{X}}-\boldsymbol{\mu})(\mathbf{X}_j-\bar{\mathbf{X}}+\bar{\mathbf{X}}-\boldsymbol{\mu})^T \tag{A.2.7}$$

$$= \sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T+\sum_{j=1}^{n}(\bar{\mathbf{X}}-\boldsymbol{\mu})(\bar{\mathbf{X}}-\boldsymbol{\mu})^T \tag{A.2.8}$$

$$= \sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T+n(\bar{\mathbf{X}}-\boldsymbol{\mu})(\bar{\mathbf{X}}-\boldsymbol{\mu})^T \tag{A.2.9}$$

since the cross-product in (A.2.7) goes to zero. At this time, the quantities can be substituted into the original likelihood equation to result in the form

$$\mathrm{L}(\boldsymbol{\mu},\boldsymbol{\Sigma})=(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T+n(\bar{\mathbf{X}}-\boldsymbol{\mu})(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\right)\right]\right\}. \tag{A.2.10}$$

Now focus briefly on the term included in the "trace" function. Using Equation (A.2.3) again, we have

$$\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T+n(\bar{\mathbf{X}}-\boldsymbol{\mu})(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\right)\right]$$

$$= \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T\right)\right]+n\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu})(\bar{\mathbf{X}}-\boldsymbol{\mu})^T] \tag{A.2.11}$$

$$= \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T\right)\right]+n(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu}) \tag{A.2.12}$$

This can then be plugged back into the likelihood equation, giving

$$\mathrm{L}(\boldsymbol{\mu},\boldsymbol{\Sigma})=(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T\right)\right]-\frac{n}{2}(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu})\right\}. \tag{A.2.13}$$

Note that if $\mathbf{S}=\frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T$, then the trace function in the likelihood can be

rewritten as $\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})^T\right)\right] = \text{tr}[\boldsymbol{\Sigma}^{-1}n\mathbf{S}]$. Therefore, the likelihood can be written as

$$\text{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{n}{2}\text{tr}[\boldsymbol{\Sigma}^{-1}\mathbf{S}]-\frac{n}{2}(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu})\right\}. \tag{A.2.14}$$

This is the form of the likelihood given in Equation (1.3.8).

A.3 Type 1 Error Rates for LRT

Section 2.3.2 discusses the need for a Bartlett correction when using the likelihood ratio test as a stopping criterion. The Type 1 error rate corresponds to the designated significance level for a test. If a significance level is set as $\alpha^* = .05$, for example, then 5% of the time, the null hypothesis will be incorrectly rejected. However, when the Bartlett correction is not employed for the likelihood ratio tests, simulations show that Type 1 error rates are sometimes close to 100% for certain cases.

A.3.1 Simulation

A simulation was conducted using the same design as Section 2.2.7.2. Only the first step in the sequence of likelihood ratio tests was considered, corresponding to the hypothesis $H_0 : \boldsymbol{\Sigma}_{-0} = \boldsymbol{\Sigma}_{-1}$ and $H_a : \boldsymbol{\Sigma}_{-0} \neq \boldsymbol{\Sigma}_{-1}$.

Note that at the first step, $\text{LR}_{\text{FULL}} = \text{LR}_{\text{SEQ}}$. Therefore, the uncorrected likelihood ratio was calculated as

$$\text{LR}_{-t} = n\left[\log|\hat{\boldsymbol{\rho}}_{-1}| - \log|\hat{\boldsymbol{\rho}}_{-0}|\right]. \tag{A.3.1}$$

Similarly, the corrected likelihood ratio statistic is calculated as

$$\text{LR}_{-t} = \left(\frac{n-1-2a_3+3a_2}{6a_2}\right)\left(\log|\hat{\boldsymbol{\rho}}_{-1}| - \log|\hat{\boldsymbol{\rho}}_{-0}|\right), \tag{A.3.2}$$

where a_2 and a_3 are as defined in Section 2.3.2.1. If either form of the likelihood ratio test is approximated by the χ^2 distribution, then H_0 should be rejected 5% of the time.

As in Section 2.2.7.2, data were generated from a 2-factor model. Three main types of loadings matrices were used for generating data, with 6, 12, or 40 correlated variables.

Uncorrelated variables (either 1, 5, or 10) were created by adding extra rows to the **L** matrix with all loadings equal to zero. The magnitude of factor loadings was determined by $\alpha = .75$, although the magnitude can be shown to be inconsequential. Samples sizes considered were n =50, 100, 400, and 1000.

A.3.2 Results

For 10,000 replicated datasets at each combination of factors, the percent of p-values that are smaller than .05 for the corresponding LR test statistic was found. In each case, the p^{th} variable was set as uncorrelated, as it was set to be uncorrelated in the population. Therefore this percent shows the probability of a Type 1 error since we ideally would fail to reject H_0 at the first step. Table A.1 gives the proportion of p-values below .05 for each combination of factors.

Table A.1 Non-Bartlett-Corrected LRT: Proportion of p-values $< .05$ out of 10,000 replications. *Sample size was 51 instead of 50, since there were 50 variables total.

m	u	$n = 50$	$n = 100$	$n = 400$	$n = 1000$
6	1	.080	.066	.054	.056
6	5	.108	.079	.052	.052
12	5	.192	.098	.061	.054
40	1	.951	.365	.087	.062
40	10	1*	.554	.105	.068

These results show that the Type 1 error rate is larger than expected, most notably for small n and large p. However, using the Bartlett correction almost entirely eradicates this problem, as seen in Table A.2. The only cases where the Type 1 error rate strays from .05 is when the ratio of p to n is very large, although even these differences are not nearly as drastic as the non-corrected results.

Table A.2 Bartlett-Corrected LRT: Proportion of p-values $< .05$ out of 10,000 replications. *Sample size was 51 instead of 50, since there were 50 variables total.

m	u	$n = 50$	$n = 100$	$n = 400$	$n = 1000$
6	1	.051	.053	.050	.055
6	5	.052	.053	.048	.050
12	5	.053	.051	.051	.050
40	1	.295	.069	.049	.050
40	10	.934*	.083	.052	.050

A.4 Pilot T-STEM Survey

Subjects rated these survey items from 1 (strongly disagree) to 5 (strongly agree).

1. When a student does better than usual in science, it is often because the teacher exerted a little extra effort.
2. I am continually finding better ways to teach science.
3. When the science grades of students improve, it is most often due to their teacher having found a more effective teaching approach.
4. I know the steps necessary to teach science concepts effectively.
5. I am not confident that I can monitor science experiments well.
6. If students are underachieving in science, it is most likely due to ineffective science teaching.
7. I am not confident that I can teach science effectively.
8. The inadequacy of a student's science background can be overcome by good teaching.
9. The low science achievement of students cannot generally be blamed on their teachers.
10. When a low achieving child progresses in science, it is usually due to extra attention given by the teacher.

11. I understand science concepts well enough to be effective in teaching science.
12. Increased effort in science teaching produces little change in students' science achievement.
13. The teacher is generally responsible for the achievement of students in science.
14. Students' achievement in science is directly related to their teacher's effectiveness in science teaching.
15. If parents comment that their child is showing more interest in science at school, it is probably due to the performance of the child's teacher.
16. I am not confident that I can explain to students why science experiments work.
17. I am confident that I can answer students' science questions.
18. I wonder if I have the necessary skills to teach science.
19. Effective science teaching has little influence on the achievement of students with low motivation.
20. Given a choice, I would not invite the principal to evaluate my science teaching.
21. When a student has difficulty understanding a science concept, I am not confident that I know to how to help the student understand it better.
22. When teaching science, I am confident enough to welcome student questions.
23. I don't know what to do to turn students on to science.
24. Even teachers with good science teaching abilities cannot help students learn science.

APPENDIX

B

ADDITIONAL SIMULATION RESULTS

This Appendix contains additional simulation results from Chapter 3, both for the full simulation and for comparisons of different revisions to GLAFA.

B.1 Full Simulation Results

Figures and tables of the full simulation results from Chapter 3 are given below.

B.1.1 Figures

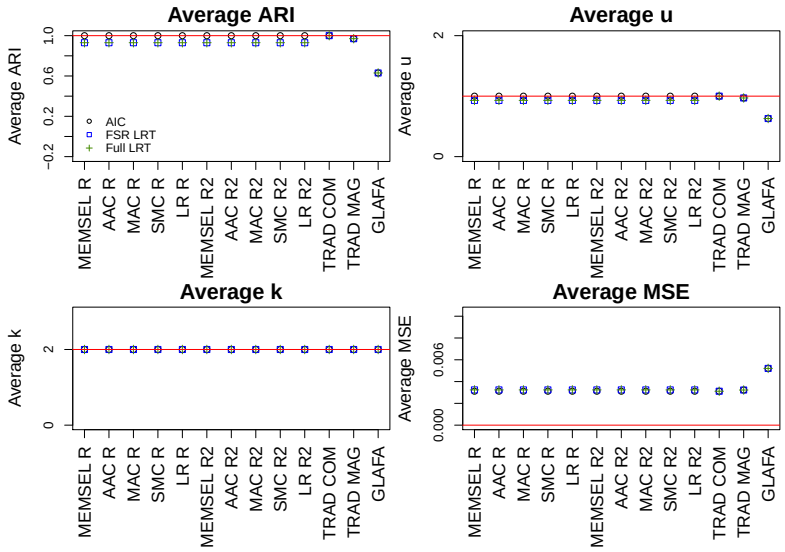


Figure B.1 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

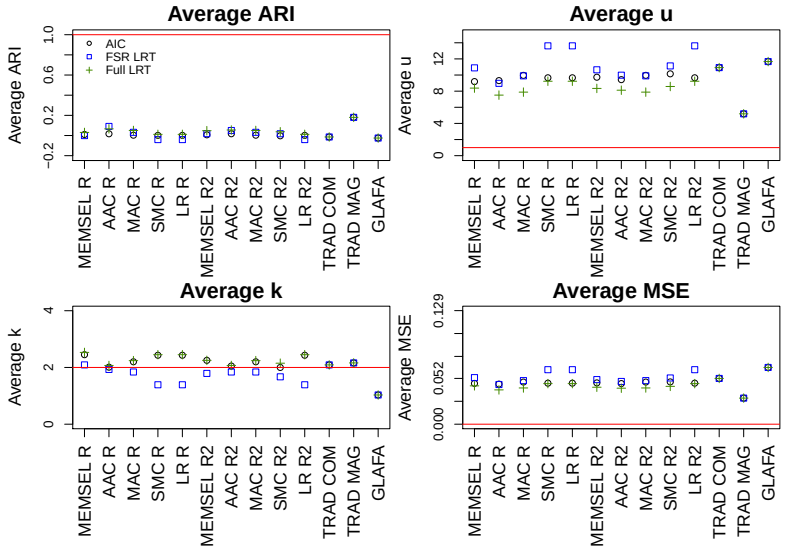


Figure B.2 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

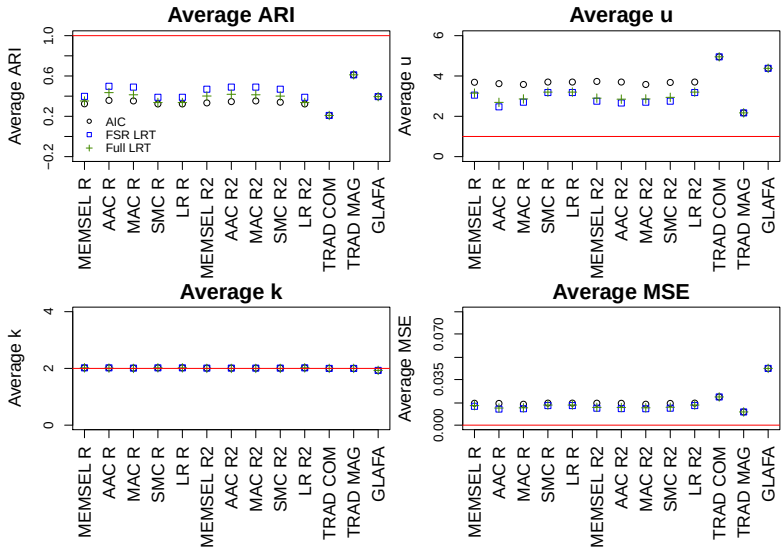


Figure B.3 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

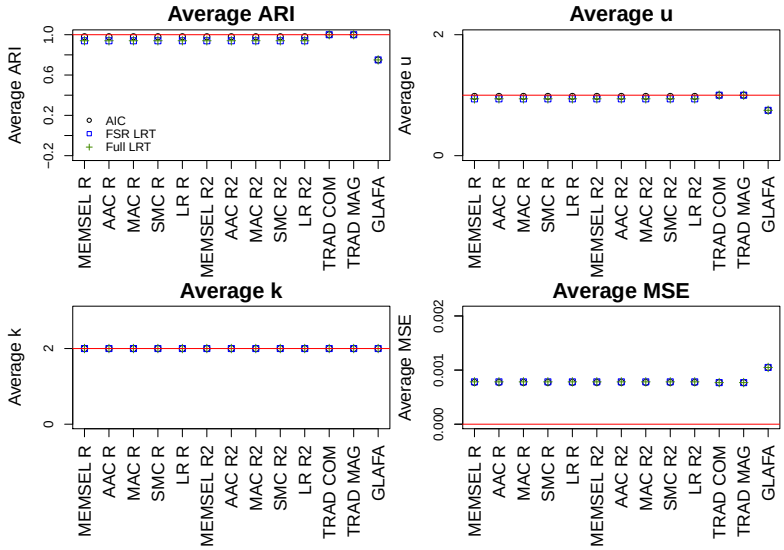


Figure B.4 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

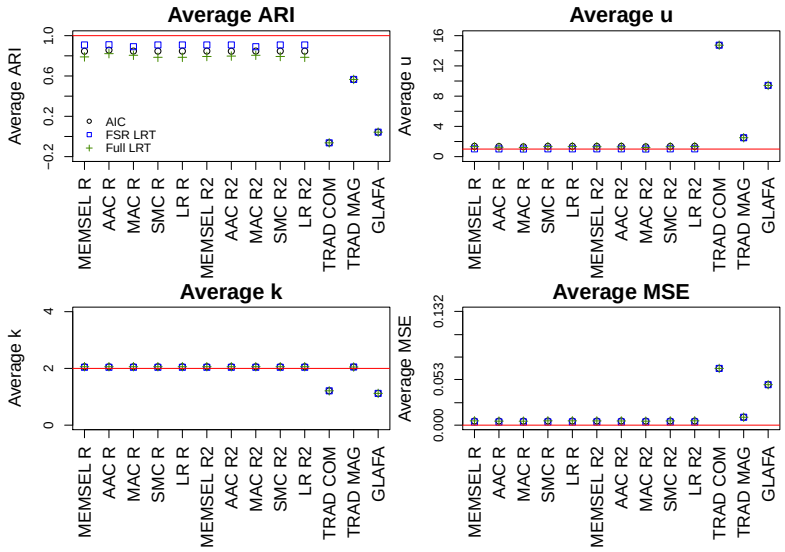


Figure B.5 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

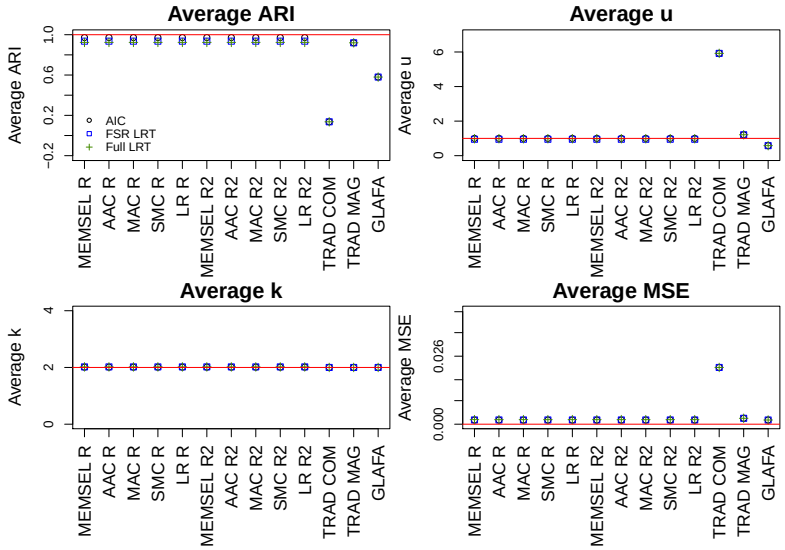


Figure B.6 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

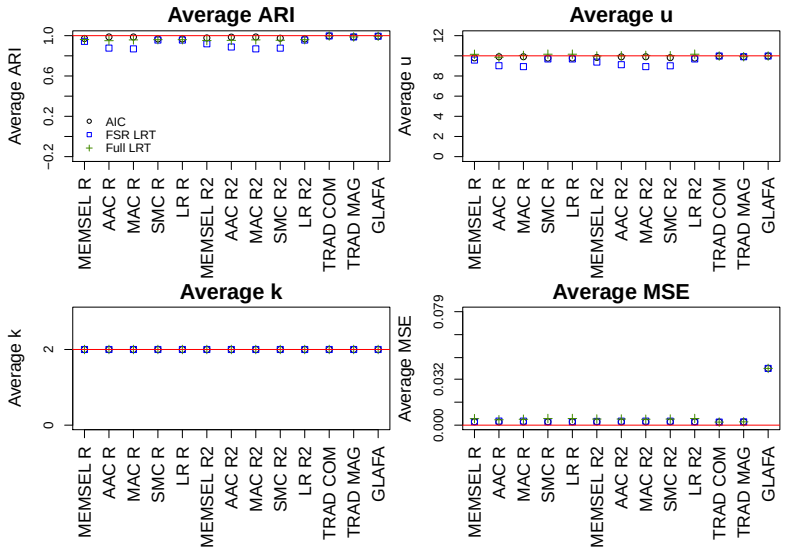


Figure B.7 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

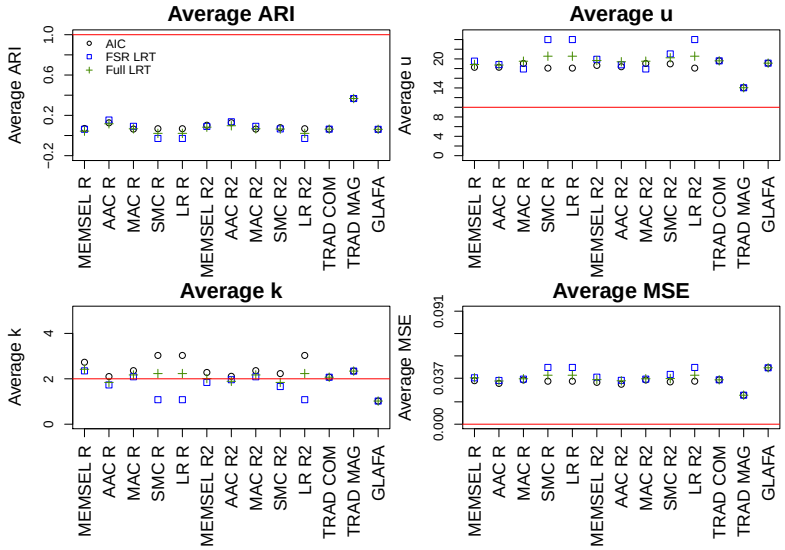


Figure B.8 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

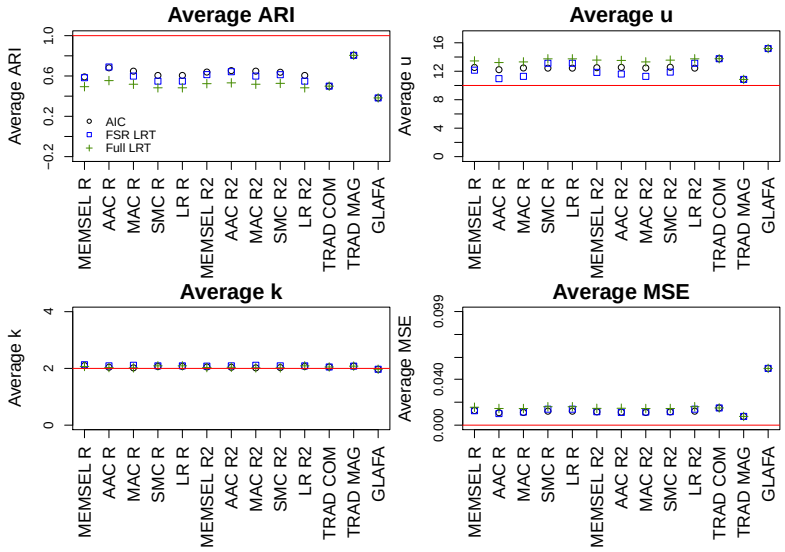


Figure B.9 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

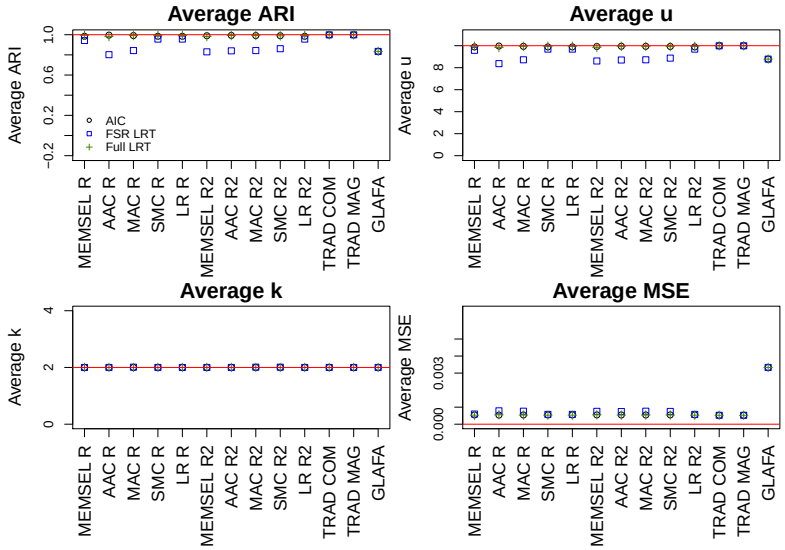


Figure B.10 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

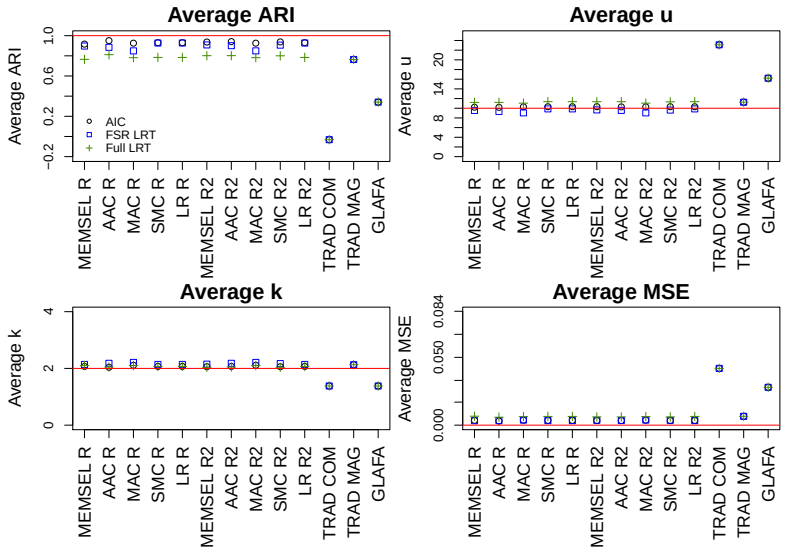


Figure B.11 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

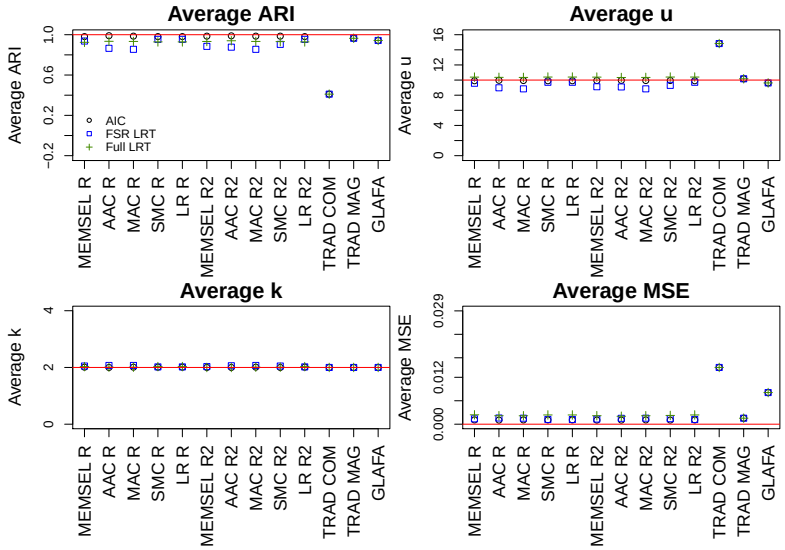


Figure B.12 Simulation results for population with $k = 2$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

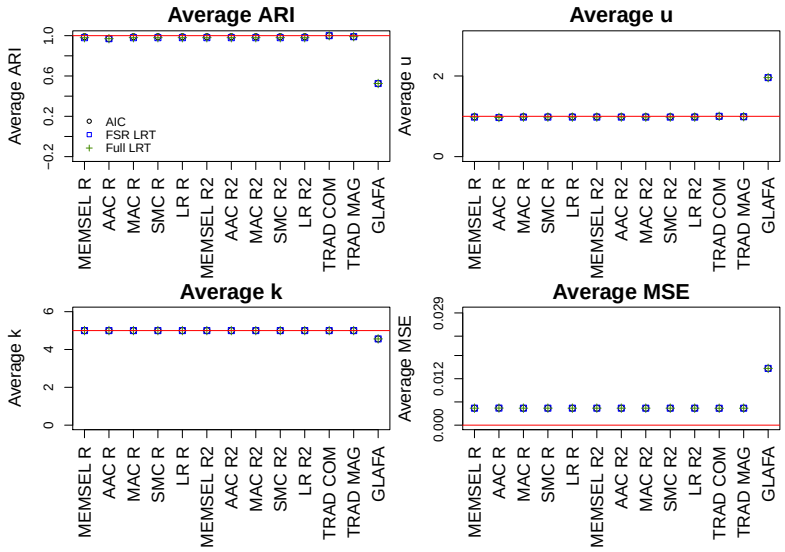


Figure B.13 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

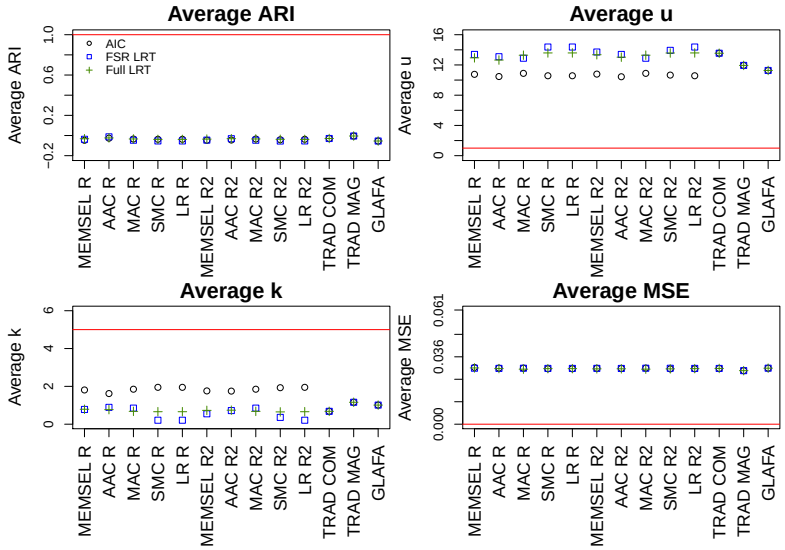


Figure B.14 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

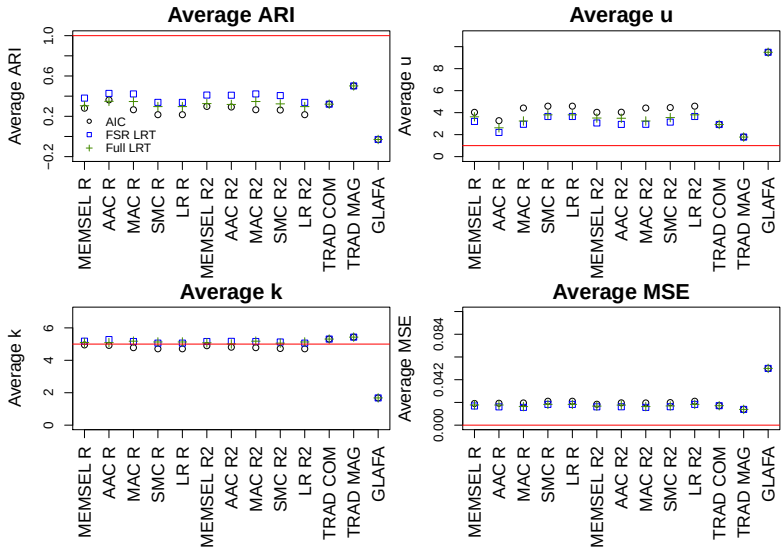


Figure B.15 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

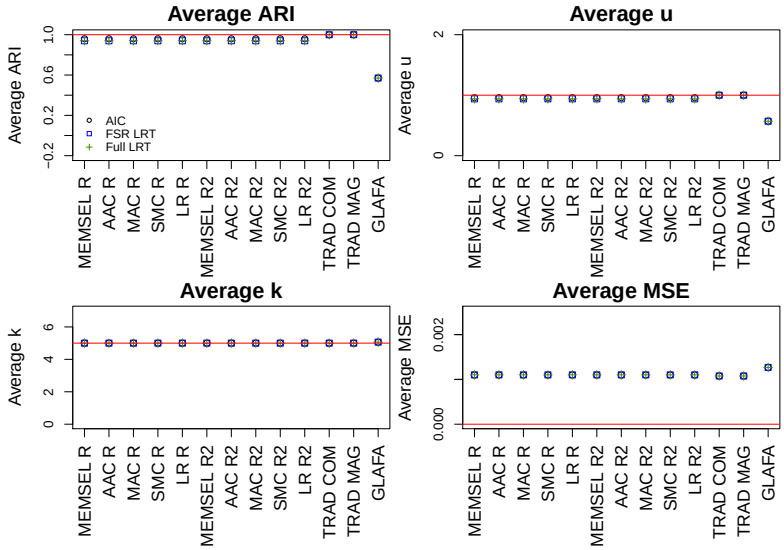


Figure B.16 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

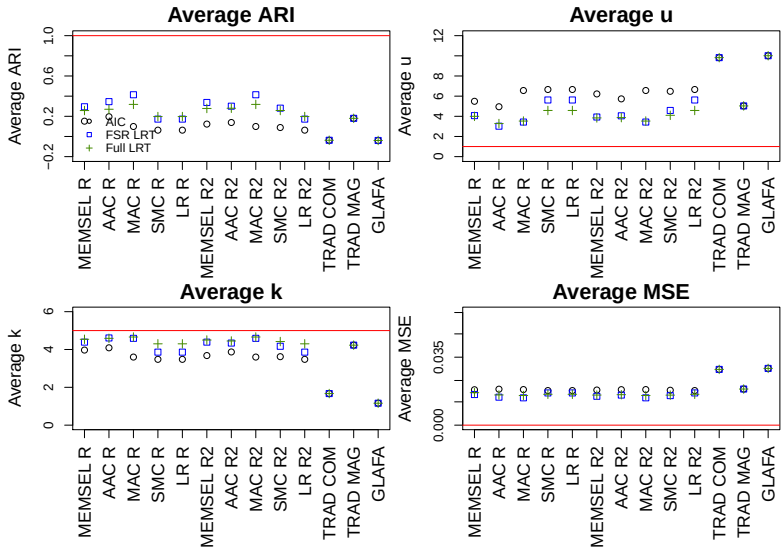


Figure B.17 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

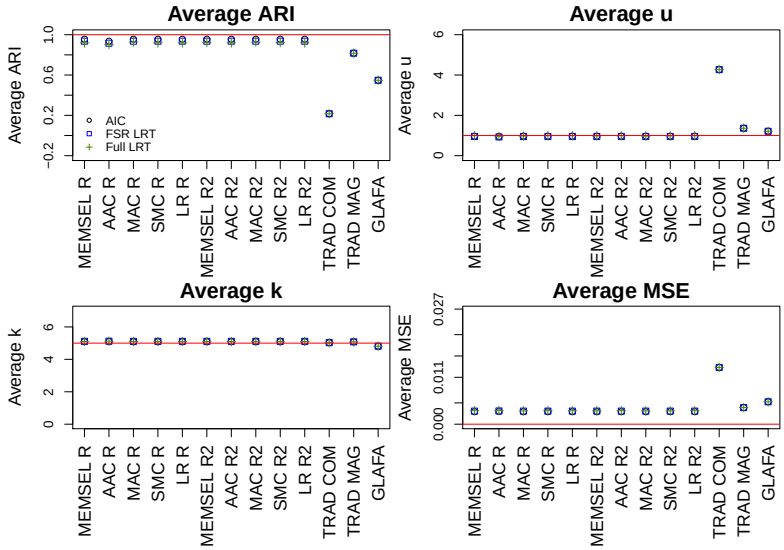


Figure B.18 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 1$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

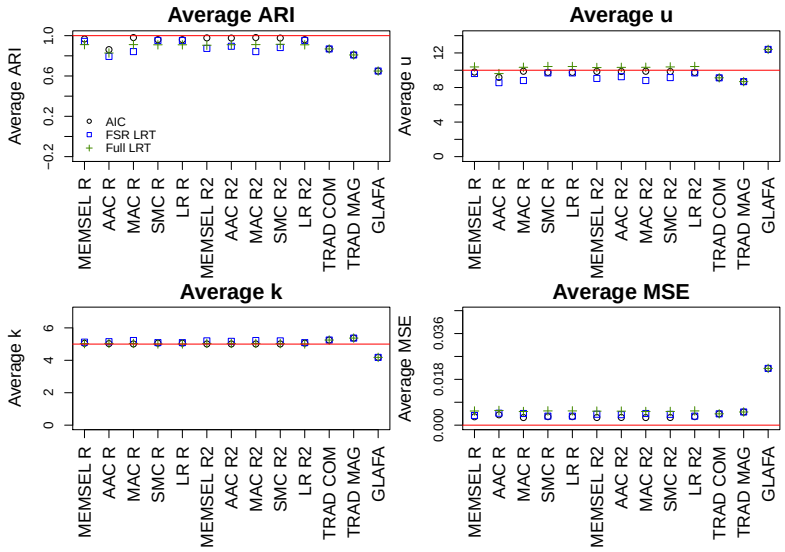


Figure B.19 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

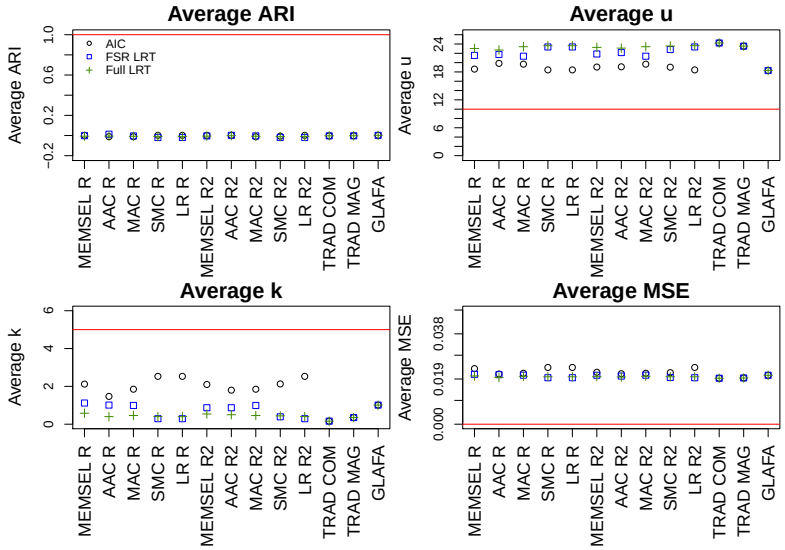


Figure B.20 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

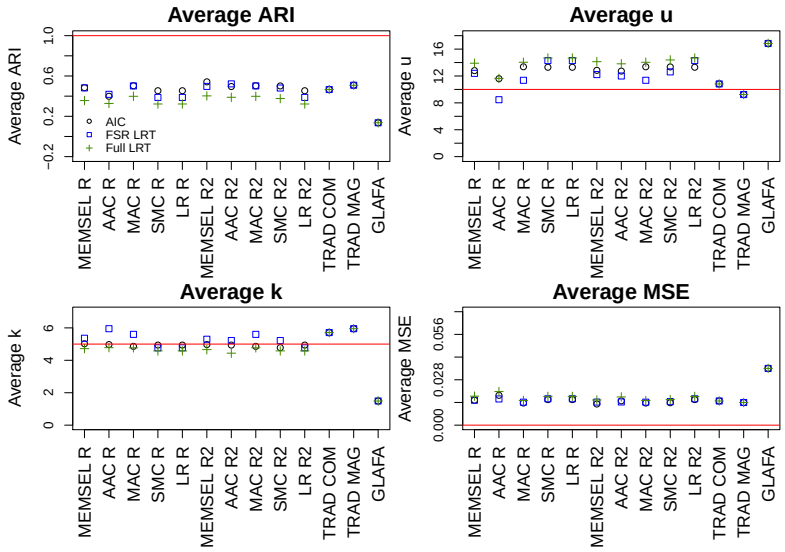


Figure B.21 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

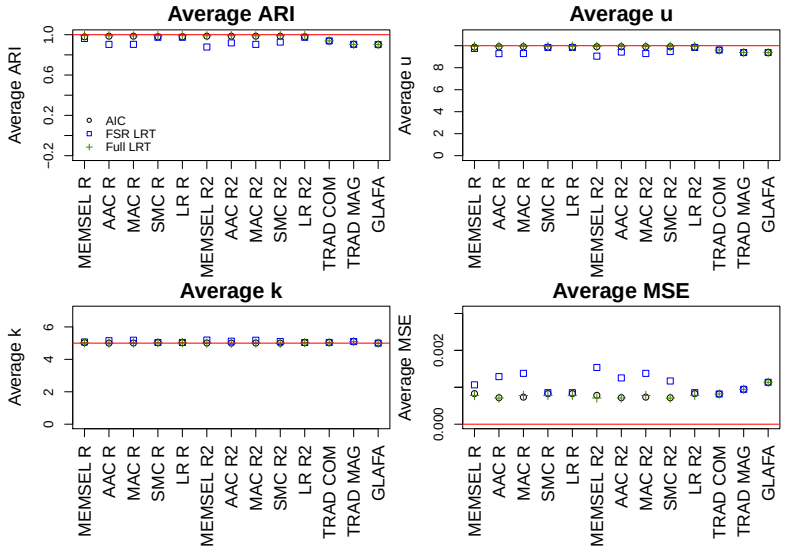


Figure B.22 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

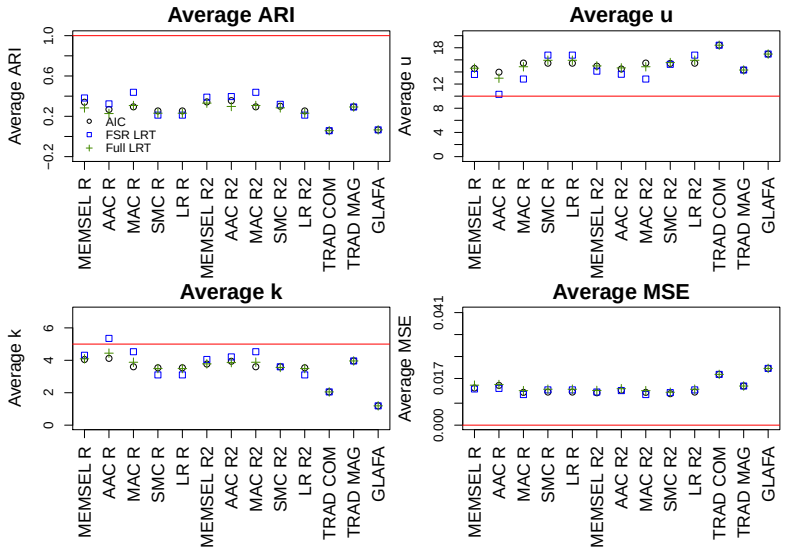


Figure B.23 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

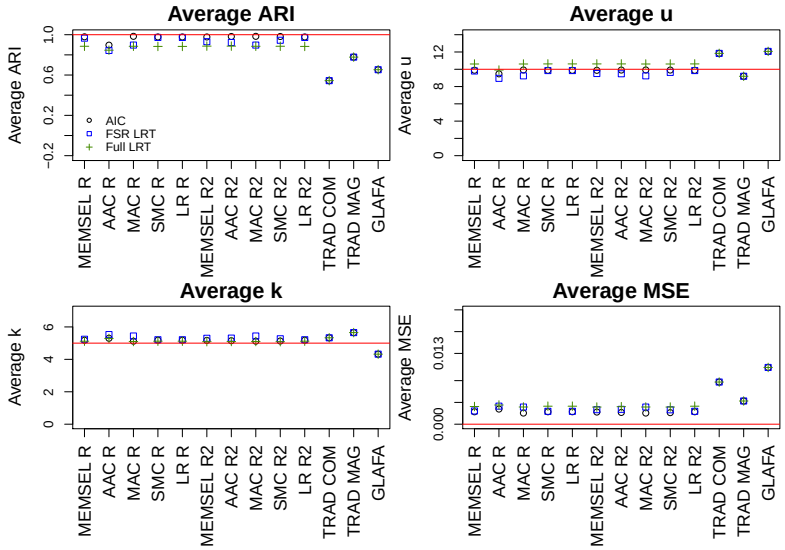


Figure B.24 Simulation results for population with $k = 5$ factors, 3 correlated variables per factor, $u = 10$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

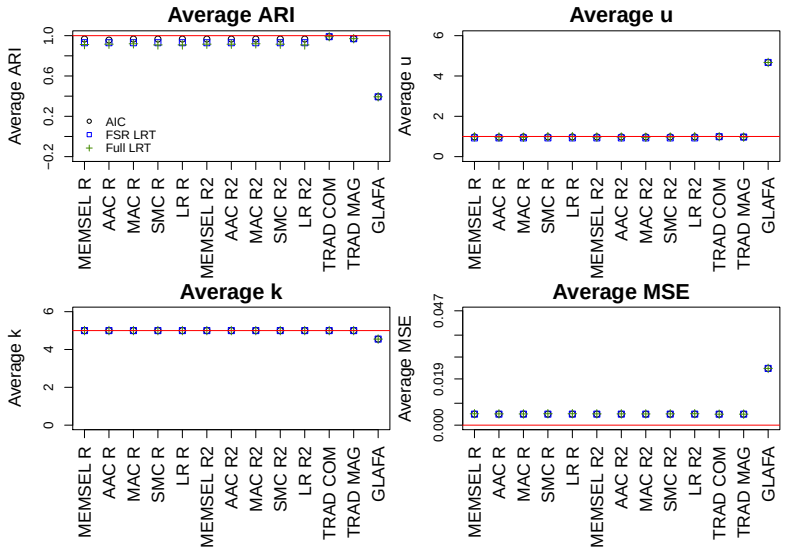


Figure B.25 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

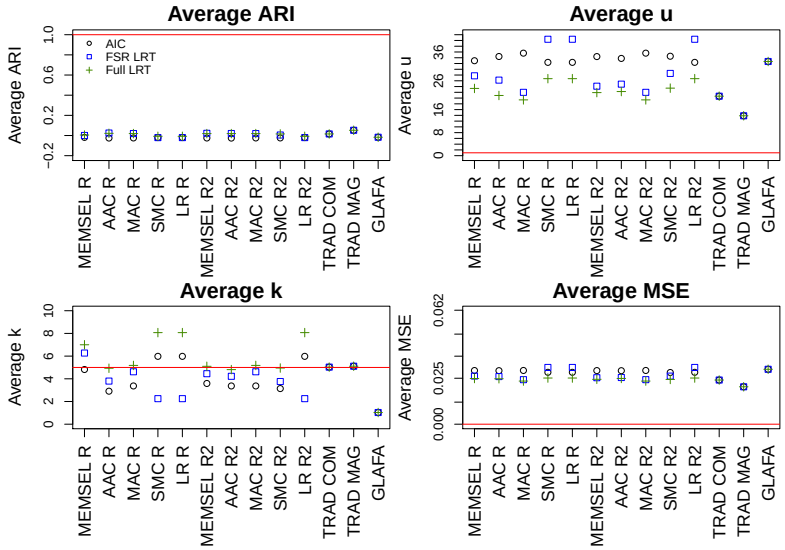


Figure B.26 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

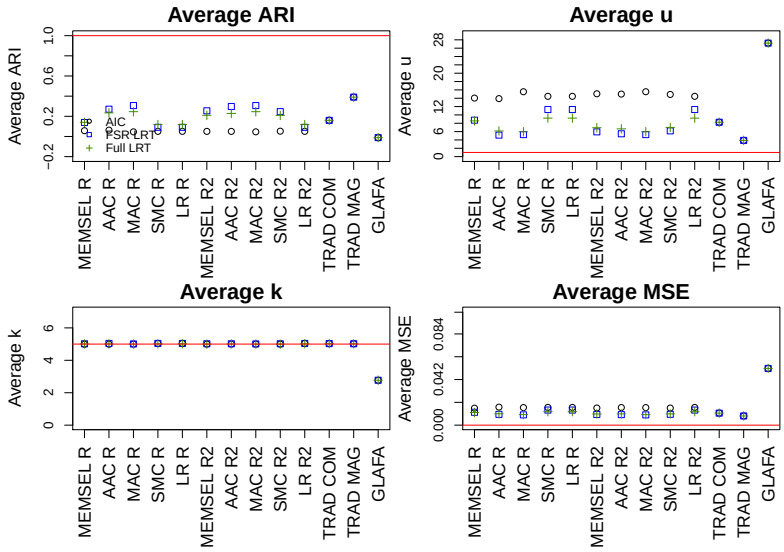


Figure B.27 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

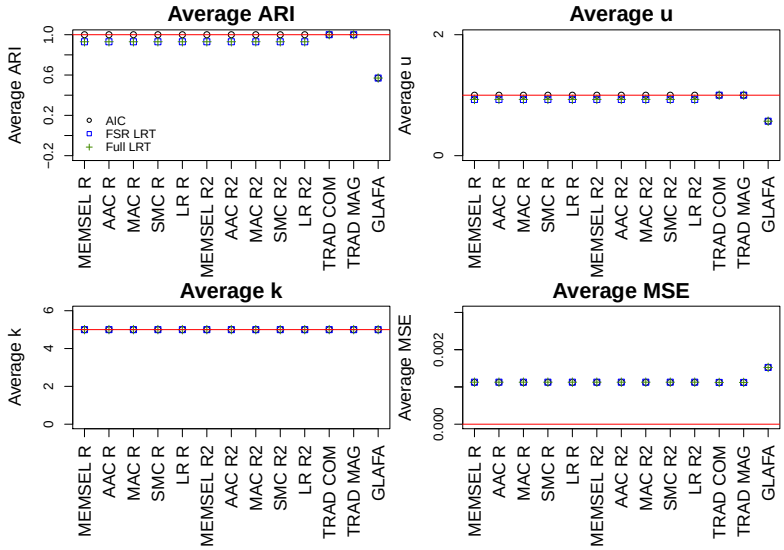


Figure B.28 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

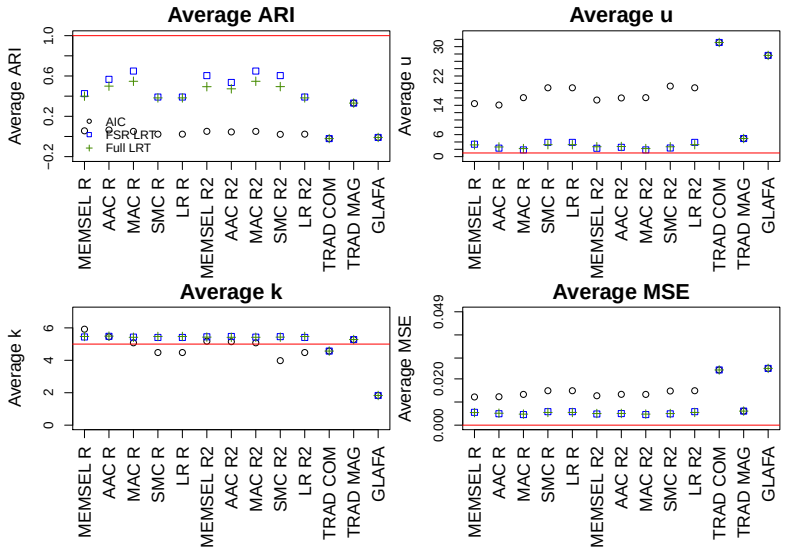


Figure B.29 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

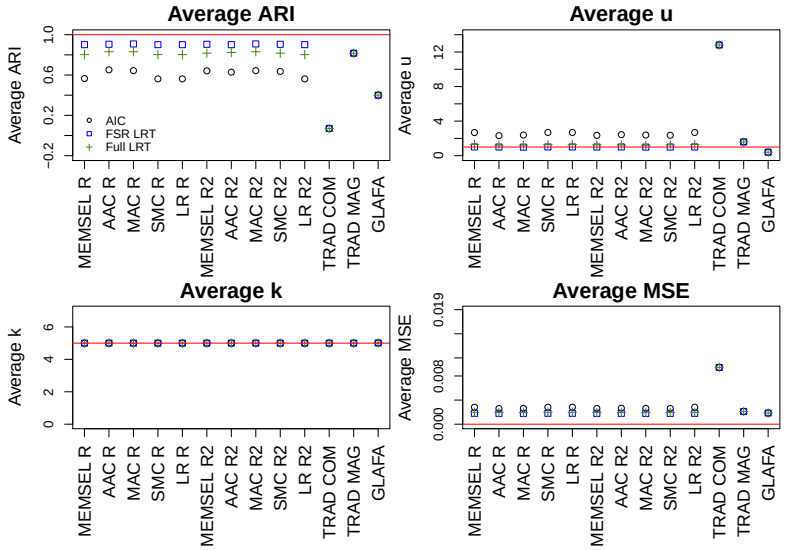


Figure B.30 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 1$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

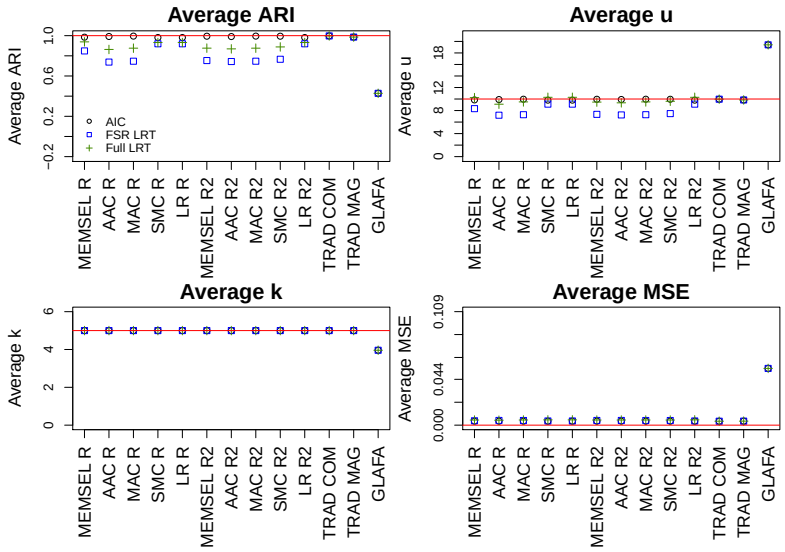


Figure B.31 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

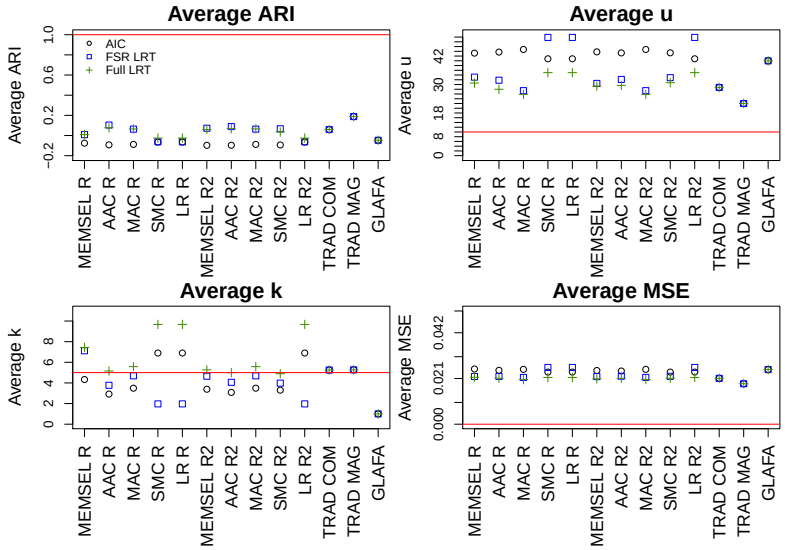


Figure B.32 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

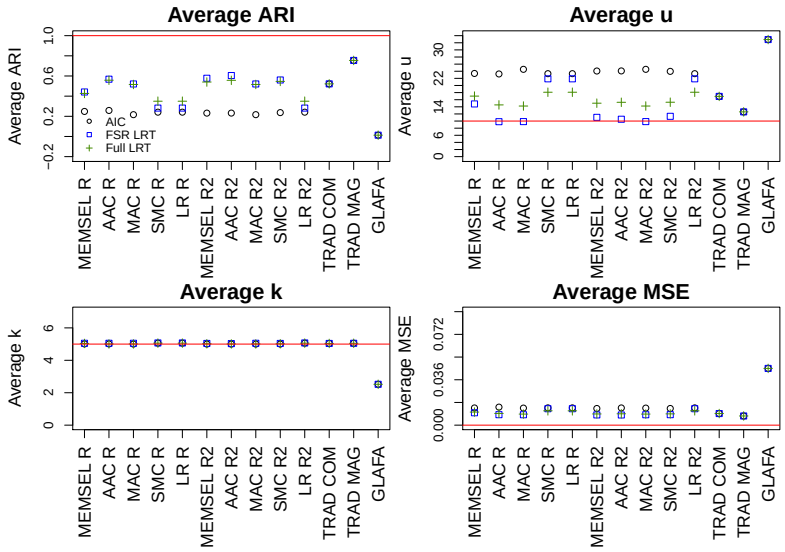


Figure B.33 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 100$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

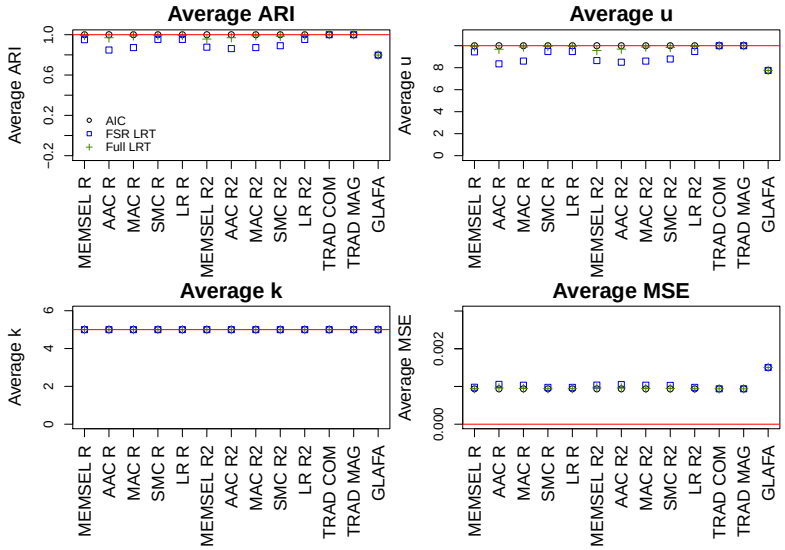


Figure B.34 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and high magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

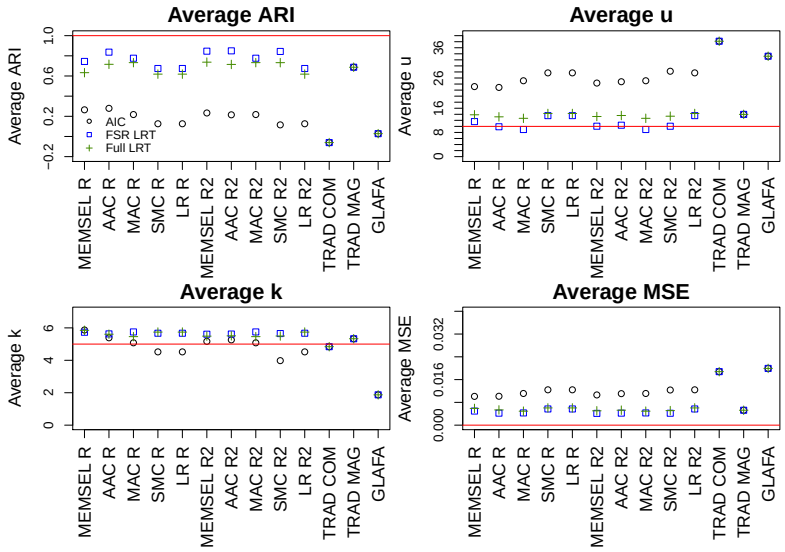


Figure B.35 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and low magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

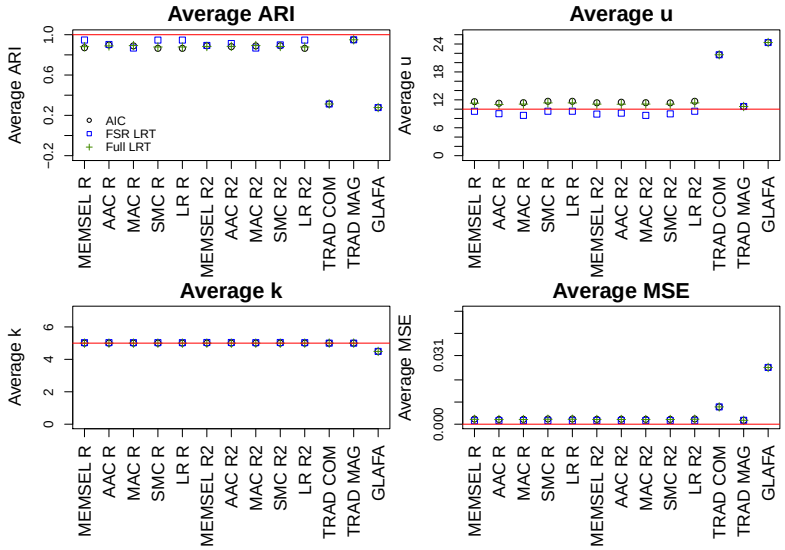


Figure B.36 Simulation results for population with $k = 5$ factors, 9 correlated variables per factor, $u = 10$, $n = 400$, and mixed magnitude of loadings. Red lines indicate desired values. Each point is an average of 100 replications.

B.1.2 Tables

Table B.1 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
	$\mathbf{R}_2$	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
AAC	**R**	AIC	1.00	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
	$\mathbf{R}_2$	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
MAC	**R**	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
	$\mathbf{R}_2$	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
SMC	**R**	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
	$\mathbf{R}_2$	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
LR	**R**	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
	$\mathbf{R}_2$	AIC	1	1	1	1	2	1	0.003
		FSR LRT	0.93	0.93	0.93	0.93	2	1	0.003
		Full LRT	0.93	0.93	0.93	0.93	2	1	0.003
Traditional Communality			1	1	1	1	2	1	0.003
Traditional Magnitude			0.97	0.97	0.97	0.97	2	1	0.003
GLAFA			0.63	0.63	0.63	0.63	2	1	0.005
Average Standard Error			0.02		0.02		0		0.000

Table B.2 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.01	0.00	9.19	0.00	2.45	0.51	0.046
		FSR LRT	0.00	0.00	10.89	0.00	2.09	0.43	0.053
		Full LRT	0.03	0.00	8.38	0.00	2.53	0.53	0.044
	$\mathbf{R}_2$	AIC	0.01	0.00	9.73	0.00	2.25	0.55	0.047
		FSR LRT	0.02	0.00	10.65	0.00	1.79	0.42	0.051
		Full LRT	0.05	0.00	8.34	0.00	2.25	0.60	0.042
AAC	**R**	AIC	0.02	0.00	9.31	0.00	2.01	0.58	0.046
		FSR LRT	0.09	0.02	8.97	0.02	1.93	0.55	0.045
		Full LRT	0.07	0.00	7.51	0.01	2.07	0.70	0.039
	$\mathbf{R}_2$	AIC	0.02	0.00	9.43	0.00	2.05	0.61	0.046
		FSR LRT	0.05	0.02	9.98	0.03	1.84	0.45	0.049
		Full LRT	0.06	0.00	8.12	0.00	2.07	0.70	0.041
MAC	**R**	AIC	0.00	0.00	9.96	0.00	2.20	0.55	0.048
		FSR LRT	0.03	0.00	9.90	0.01	1.84	0.47	0.050
		Full LRT	0.05	0.00	7.88	0.00	2.24	0.66	0.041
	$\mathbf{R}_2$	AIC	0.00	0.00	9.96	0.00	2.20	0.55	0.048
		FSR LRT	0.03	0.00	9.90	0.01	1.84	0.47	0.050
		Full LRT	0.05	0.00	7.88	0.00	2.24	0.66	0.041
SMC	**R**	AIC	0.00	0.00	9.66	0.00	2.43	0.43	0.047
		FSR LRT	-0.04	0.00	13.62	0.00	1.39	0.28	0.062
		Full LRT	0.01	0.00	9.23	0.00	2.45	0.48	0.046
	$\mathbf{R}_2$	AIC	0.00	0.00	10.15	0.00	2.00	0.49	0.048
		FSR LRT	0.02	0.00	11.13	0.00	1.67	0.31	0.053
		Full LRT	0.04	0.00	8.58	0.00	2.15	0.56	0.043
LR	**R**	AIC	0.00	0.00	9.66	0.00	2.43	0.43	0.047
		FSR LRT	-0.04	0.00	13.62	0.00	1.39	0.28	0.062
		Full LRT	0.01	0.00	9.23	0.00	2.45	0.48	0.046
	$\mathbf{R}_2$	AIC	0.00	0.00	9.66	0.00	2.43	0.43	0.047
		FSR LRT	-0.04	0.00	13.62	0.00	1.39	0.28	0.062
		Full LRT	0.01	0.00	9.23	0.00	2.45	0.48	0.046
Traditional Communality			-0.01	0.00	10.92	0.00	2.09	0.72	0.052
Traditional Magnitude			0.18	0.00	5.19	0.01	2.16	0.80	0.030
GLAFA			-0.03	0.00	11.67	0.00	1.03	0.03	0.064
Average Standard Error			0.01		0.31		0.09		0.001

Table B.3 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.32	0.00	3.69	0.00	2.02	0.98	0.017
		FSR LRT	0.40	0.07	3.05	0.09	2.02	0.98	0.014
		Full LRT	0.35	0.01	3.15	0.04	2.02	0.98	0.015
	$\mathbf{R}_2$	AIC	0.33	0.04	3.73	0.04	2.01	0.99	0.017
		FSR LRT	0.47	0.17	2.75	0.18	2.01	0.99	0.013
		Full LRT	0.40	0.06	2.90	0.07	2	1	0.014
AAC	**R**	AIC	0.36	0.06	3.62	0.06	2.02	0.98	0.017
		FSR LRT	0.50	0.20	2.47	0.23	2.02	0.98	0.012
		Full LRT	0.44	0.09	2.69	0.10	2.01	0.99	0.013
	$\mathbf{R}_2$	AIC	0.35	0.06	3.70	0.06	2.01	0.99	0.017
		FSR LRT	0.49	0.19	2.66	0.20	2.01	0.99	0.013
		Full LRT	0.42	0.08	2.85	0.08	2	1	0.014
MAC	**R**	AIC	0.35	0.04	3.58	0.04	2.01	0.99	0.016
		FSR LRT	0.49	0.20	2.70	0.21	2.01	0.99	0.013
		Full LRT	0.41	0.07	2.86	0.08	2	1	0.013
	$\mathbf{R}_2$	AIC	0.35	0.04	3.58	0.04	2.01	0.99	0.016
		FSR LRT	0.49	0.20	2.70	0.21	2.01	0.99	0.013
		Full LRT	0.41	0.07	2.86	0.08	2	1	0.013
SMC	**R**	AIC	0.32	0.00	3.70	0.00	2.02	0.98	0.017
		FSR LRT	0.39	0.06	3.18	0.06	2.02	0.98	0.015
		Full LRT	0.34	0.01	3.20	0.04	2.02	0.98	0.015
	$\mathbf{R}_2$	AIC	0.34	0.04	3.68	0.04	2.01	0.99	0.017
		FSR LRT	0.47	0.17	2.75	0.18	2.01	0.99	0.013
		Full LRT	0.40	0.05	2.94	0.05	2	1	0.014
LR	**R**	AIC	0.32	0.00	3.70	0.00	2.02	0.98	0.017
		FSR LRT	0.39	0.06	3.18	0.06	2.02	0.98	0.015
		Full LRT	0.34	0.01	3.20	0.04	2.02	0.98	0.015
	$\mathbf{R}_2$	AIC	0.32	0.00	3.70	0.00	2.02	0.98	0.017
		FSR LRT	0.39	0.06	3.18	0.06	2.02	0.98	0.015
		Full LRT	0.34	0.01	3.20	0.04	2.02	0.98	0.015
Traditional Communality			0.21	0.00	4.95	0.00	2	1	0.022
Traditional Magnitude			0.61	0.25	2.17	0.26	2	1	0.010
GLAFA			0.40	0.37	4.38	0.37	1.93	0.93	0.044
Average Standard Error			0.02		0.14		0.01		0.001

Table B.4 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
AAC	**R**	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
MAC	**R**	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
SMC	**R**	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
LR	**R**	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.98	0.98	0.98	2	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	2	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	2	1	0.001
Traditional Communality			1	1	1	1	2	1	0.001
Traditional Magnitude			1	1	1	1	2	1	0.001
GLAFA			0.75	0.75	0.75	0.75	2	1	0.001
Average Standard Error			0.02		0.02		0		0.000

Table B.5 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.85	0.68	1.36	0.69	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.57	1.32	0.57	2.06	0.95	0.005
	$\mathbf{R}_2$	AIC	0.85	0.68	1.35	0.68	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.58	1.31	0.58	2.06	0.95	0.005
AAC	**R**	AIC	0.86	0.71	1.34	0.71	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.02	0.87	2.04	0.96	0.004
		Full LRT	0.82	0.65	1.24	0.65	2.06	0.95	0.005
	$\mathbf{R}_2$	AIC	0.85	0.68	1.36	0.68	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.80	0.59	1.30	0.59	2.06	0.95	0.005
MAC	**R**	AIC	0.85	0.72	1.30	0.73	2.06	0.95	0.005
		FSR LRT	0.89	0.86	1.01	0.86	2.04	0.96	0.004
		Full LRT	0.81	0.64	1.21	0.64	2.06	0.95	0.005
	$\mathbf{R}_2$	AIC	0.85	0.72	1.30	0.73	2.06	0.95	0.005
		FSR LRT	0.89	0.86	1.01	0.86	2.04	0.96	0.004
		Full LRT	0.81	0.64	1.21	0.64	2.06	0.95	0.005
SMC	**R**	AIC	0.85	0.68	1.36	0.69	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.56	1.33	0.56	2.06	0.95	0.005
	$\mathbf{R}_2$	AIC	0.85	0.68	1.35	0.68	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.58	1.31	0.58	2.06	0.95	0.005
LR	**R**	AIC	0.85	0.68	1.36	0.69	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.56	1.33	0.56	2.06	0.95	0.005
	$\mathbf{R}_2$	AIC	0.847	0.68	1.36	0.69	2.06	0.95	0.005
		FSR LRT	0.91	0.87	1.03	0.87	2.04	0.96	0.004
		Full LRT	0.79	0.56	1.33	0.56	2.06	0.95	0.005
Traditional Communality			-0.06	0.00	14.74	0.00	1.21	0.28	0.066
Traditional Magnitude			0.57	0.24	2.50	0.24	2.05	0.95	0.009
GLAFA			0.04	0.06	9.42	0.06	1.12	0.12	0.047
Average Standard Error			0.03		0.07		0.03		0.000

Table B.6 Results for $k = 2$, 9 correlated variables per factor, $u = 1$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
	$\mathbf{R}_2$	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
AAC	**R**	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
	$\mathbf{R}_2$	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
MAC	**R**	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
	$\mathbf{R}_2$	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
SMC	**R**	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
	$\mathbf{R}_2$	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
LR	**R**	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
	$\mathbf{R}_2$	AIC	0.97	0.96	1.00	0.96	2.01	0.99	0.002
		FSR LRT	0.94	0.93	0.95	0.93	2.02	0.98	0.002
		Full LRT	0.92	0.90	0.98	0.90	2.02	0.98	0.002
Traditional Communality			0.14	0.00	5.92	0.00	2.00	1.00	0.022
Traditional Magnitude			0.92	0.81	1.21	0.81	2.00	1.00	0.002
GLAFA			0.58	0.58	0.58	0.58	2.00	1.00	0.002
Average Standard Error			0.02		0.03		0.01		0.000

Table B.7 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.97	0.84	9.80	0.84	2	1	0.002
		FSR LRT	0.94	0.78	9.58	0.78	2	1	0.002
		Full LRT	0.96	0.75	10.15	0.75	2	1	0.005
	$\mathbf{R}_2$	AIC	0.98	0.88	9.84	0.88	2	1	0.002
		FSR LRT	0.92	0.76	9.38	0.76	2	1	0.003
		Full LRT	0.95	0.77	9.98	0.77	2	1	0.004
AAC	**R**	AIC	0.99	0.93	9.91	0.93	2	1	0.002
		FSR LRT	0.88	0.70	9.02	0.70	2	1	0.003
		Full LRT	0.95	0.80	9.86	0.80	2	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.91	9.90	0.91	2	1	0.002
		FSR LRT	0.89	0.72	9.12	0.72	2	1	0.003
		Full LRT	0.95	0.77	10.00	0.77	2	1	0.004
MAC	**R**	AIC	0.99	0.91	9.91	0.91	2	1	0.002
		FSR LRT	0.87	0.71	8.93	0.71	2	1	0.003
		Full LRT	0.96	0.77	10.04	0.77	2	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.91	9.91	0.91	2	1	0.002
		FSR LRT	0.87	0.71	8.93	0.71	2	1	0.003
		Full LRT	0.96	0.77	10.04	0.77	2	1	0.004
SMC	**R**	AIC	0.97	0.80	9.77	0.80	2	1	0.002
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.002
		Full LRT	0.96	0.74	10.16	0.74	2	1	0.005
	$\mathbf{R}_2$	AIC	0.98	0.86	9.82	0.86	2	1	0.002
		FSR LRT	0.88	0.71	9.01	0.71	2	1	0.003
		Full LRT	0.96	0.77	10.01	0.77	2	1	0.004
LR	**R**	AIC	0.97	0.80	9.77	0.80	2	1	0.002
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.002
		Full LRT	0.96	0.74	10.16	0.74	2	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.80	9.77	0.80	2	1	0.002
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.002
		Full LRT	0.96	0.74	10.16	0.74	2	1	0.005
Traditional Communality			1.00	0.99	9.98	0.99	2	1	0.002
Traditional Magnitude			0.99	0.93	9.91	0.93	2	1	0.002
GLAFA			0.99	0.96	9.98	0.96	2	1	0.040
Average Standard Error			0.01		0.08		0		0.000

Table B.8 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.07	0.00	18.27	0.00	2.73	0.29	0.035
		FSR LRT	0.06	0.00	19.54	0.01	2.35	0.27	0.037
		Full LRT	0.04	0.00	18.84	0.00	2.42	0.26	0.037
	$\mathbf{R}_2$	AIC	0.10	0.00	18.66	0.00	2.28	0.54	0.034
		FSR LRT	0.09	0.00	19.92	0.04	1.84	0.27	0.038
		Full LRT	0.08	0.00	19.67	0.00	2.01	0.49	0.036
AAC	**R**	AIC	0.13	0.00	18.28	0.00	2.10	0.53	0.033
		FSR LRT	0.15	0.00	18.83	0.05	1.73	0.41	0.035
		Full LRT	0.12	0.00	18.73	0.01	1.84	0.53	0.035
	$\mathbf{R}_2$	AIC	0.13	0.00	18.38	0.00	2.11	0.52	0.032
		FSR LRT	0.14	0.00	18.78	0.03	1.96	0.40	0.035
		Full LRT	0.10	0.00	19.40	0.00	1.89	0.45	0.035
MAC	**R**	AIC	0.06	0.00	19.05	0.01	2.36	0.40	0.036
		FSR LRT	0.09	0.00	17.92	0.04	2.09	0.32	0.036
		Full LRT	0.07	0.00	19.52	0.02	2.17	0.34	0.037
	$\mathbf{R}_2$	AIC	0.06	0.00	19.05	0.01	2.36	0.40	0.036
		FSR LRT	0.09	0.00	17.92	0.04	2.09	0.32	0.036
		Full LRT	0.07	0.00	19.52	0.02	2.17	0.34	0.037
SMC	**R**	AIC	0.07	0.00	18.10	0.00	3.03	0.27	0.035
		FSR LRT	-0.03	0.00	24.00	0.00	1.08	0.15	0.046
		Full LRT	0.02	0.00	20.56	0.00	2.23	0.26	0.039
	$\mathbf{R}_2$	AIC	0.08	0.00	18.98	0.00	2.23	0.43	0.034
		FSR LRT	0.07	0.00	21.03	0.04	1.66	0.19	0.040
		Full LRT	0.06	0.00	20.23	0.00	1.84	0.35	0.037
LR	**R**	AIC	0.07	0.00	18.10	0.00	3.03	0.27	0.035
		FSR LRT	-0.03	0.00	24.00	0.00	1.08	0.15	0.046
		Full LRT	0.02	0.00	20.56	0.00	2.23	0.26	0.039
	$\mathbf{R}_2$	AIC	0.07	0.00	18.10	0.00	3.03	0.27	0.035
		FSR LRT	-0.03	0.00	24.00	0.00	1.08	0.15	0.046
		Full LRT	0.02	0.00	20.56	0.00	2.23	0.26	0.039
Traditional Communality			0.06	0.00	19.60	0.01	2.07	0.67	0.036
Traditional Magnitude			0.37	0.00	14.04	0.04	2.34	0.68	0.023
GLAFA			0.06	0.00	19.12	0.00	1.02	0.02	0.045
Average Standard Error			0.01		0.38		0.11		0.001

Table B.9 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.59	0.00	12.53	0.02	2.10	0.91	0.013
		FSR LRT	0.58	0.01	12.15	0.05	2.13	0.88	0.013
		Full LRT	0.49	0.01	13.45	0.03	2.07	0.93	0.016
	$\mathbf{R}_2$	AIC	0.64	0.02	12.50	0.02	2.05	0.95	0.012
		FSR LRT	0.61	0.06	11.84	0.06	2.08	0.92	0.012
		Full LRT	0.52	0.01	13.57	0.02	2.03	0.97	0.015
AAC	**R**	AIC	0.68	0.07	12.22	0.08	2.04	0.96	0.011
		FSR LRT	0.69	0.11	10.96	0.17	2.09	0.91	0.010
		Full LRT	0.55	0.02	13.22	0.04	2.03	0.97	0.015
	$\mathbf{R}_2$	AIC	0.65	0.05	12.54	0.05	2.04	0.96	0.011
		FSR LRT	0.64	0.06	11.61	0.10	2.09	0.91	0.011
		Full LRT	0.53	0.01	13.52	0.02	2.04	0.96	0.015
MAC	**R**	AIC	0.65	0.01	12.46	0.03	2.02	0.98	0.011
		FSR LRT	0.60	0.04	11.28	0.06	2.11	0.92	0.011
		Full LRT	0.52	0.00	13.31	0.00	2.02	0.98	0.014
	$\mathbf{R}_2$	AIC	0.65	0.01	12.46	0.03	2.02	0.98	0.011
		FSR LRT	0.60	0.04	11.28	0.06	2.11	0.92	0.011
		Full LRT	0.52	0.00	13.31	0.00	2.02	0.98	0.014
SMC	**R**	AIC	0.61	0.00	12.44	0.01	2.07	0.94	0.012
		FSR LRT	0.55	0.00	13.12	0.01	2.09	0.92	0.014
		Full LRT	0.48	0.00	13.74	0.03	2.09	0.92	0.016
	$\mathbf{R}_2$	AIC	0.64	0.03	12.56	0.04	2.03	0.97	0.012
		FSR LRT	0.61	0.06	11.88	0.08	2.09	0.92	0.012
		Full LRT	0.53	0.01	13.56	0.02	2.02	0.98	0.015
LR	**R**	AIC	0.61	0.00	12.44	0.01	2.07	0.94	0.012
		FSR LRT	0.55	0.00	13.12	0.01	2.09	0.92	0.014
		Full LRT	0.48	0.00	13.74	0.03	2.09	0.92	0.016
	$\mathbf{R}_2$	AIC	0.61	0.00	12.44	0.01	2.07	0.94	0.012
		FSR LRT	0.55	0.00	13.12	0.01	2.09	0.92	0.014
		Full LRT	0.48	0.00	13.74	0.03	2.09	0.92	0.016
Traditional Communality			0.50	0.00	13.76	0.00	2.05	0.96	0.015
Traditional Magnitude			0.81	0.23	10.85	0.28	2.08	0.94	0.008
GLAFA			0.38	0.00	15.19	0.00	1.97	0.97	0.049
Average Standard Error			0.02		0.18		0.03		0.000

Table B.10 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.98	0.90	9.89	0.90	2	1	0.001
		FSR LRT	0.94	0.76	9.58	0.76	2	1	0.001
		Full LRT	0.99	0.96	9.96	0.96	2	1	0.001
	$\mathbf{R}_2$	AIC	0.99	0.95	9.93	0.95	2	1	0.001
		FSR LRT	0.83	0.66	8.61	0.66	2	1	0.001
		Full LRT	0.98	0.95	9.85	0.95	2	1	0.001
AAC	**R**	AIC	1.00	0.98	9.97	0.98	2	1	0.001
		FSR LRT	0.80	0.63	8.37	0.63	2	1	0.001
		Full LRT	0.98	0.95	9.84	0.95	2	1	0.001
	$\mathbf{R}_2$	AIC	0.99	0.97	9.95	0.97	2	1	0.001
		FSR LRT	0.84	0.69	8.70	0.69	2	1	0.001
		Full LRT	0.99	0.96	9.94	0.96	2	1	0.001
MAC	**R**	AIC	0.99	0.96	9.94	0.96	2	1	0.001
		FSR LRT	0.84	0.68	8.72	0.68	2.01	0.99	0.001
		Full LRT	0.99	0.96	9.94	0.96	2	1	0.001
	$\mathbf{R}_2$	AIC	0.99	0.96	9.94	0.96	2	1	0.001
		FSR LRT	0.84	0.68	8.72	0.68	2.01	0.99	0.001
		Full LRT	0.99	0.96	9.94	0.96	2	1	0.001
SMC	**R**	AIC	0.98	0.90	9.89	0.90	2	1	0.001
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.001
		Full LRT	0.99	0.96	9.95	0.96	2	1	0.001
	$\mathbf{R}_2$	AIC	0.99	0.95	9.93	0.95	2	1	0.001
		FSR LRT	0.86	0.72	8.87	0.72	2.01	0.99	0.001
		Full LRT	0.99	0.96	9.92	0.96	2	1	0.001
LR	**R**	AIC	0.98	0.90	9.89	0.90	2	1	0.001
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.001
		Full LRT	0.99	0.96	9.95	0.96	2	1	0.001
	$\mathbf{R}_2$	AIC	0.98	0.90	9.89	0.90	2	1	0.001
		FSR LRT	0.96	0.78	9.69	0.78	2	1	0.001
		Full LRT	0.99	0.96	9.95	0.96	2	1	0.001
Traditional Communality			1.00	0.99	9.99	0.99	2	1	0.001
Traditional Magnitude			1.00	0.99	9.99	0.99	2	1	0.001
GLAFA			0.83	0.37	8.78	0.37	2	1	0.003
Average Standard Error			0.01	0.00	0.08	0.00	0.00	0.00	0.000

Table B.11 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.92	0.56	10.22	0.62	2.07	0.94	0.004
		FSR LRT	0.90	0.55	9.55	0.57	2.14	0.89	0.003
		Full LRT	0.77	0.08	11.21	0.17	2.10	0.91	0.007
	$\mathbf{R}_2$	AIC	0.94	0.64	10.33	0.64	2.06	0.95	0.003
		FSR LRT	0.91	0.65	9.66	0.66	2.15	0.88	0.004
		Full LRT	0.80	0.13	11.35	0.14	2.04	0.96	0.006
AAC	**R**	AIC	0.95	0.72	10.19	0.73	2.04	0.96	0.003
		FSR LRT	0.88	0.62	9.33	0.63	2.18	0.87	0.003
		Full LRT	0.81	0.16	11.21	0.18	2.05	0.95	0.006
	$\mathbf{R}_2$	AIC	0.94	0.67	10.28	0.68	2.07	0.94	0.003
		FSR LRT	0.90	0.64	9.56	0.65	2.18	0.86	0.003
		Full LRT	0.80	0.15	11.35	0.15	2.05	0.95	0.006
MAC	**R**	AIC	0.93	0.63	10.28	0.64	2.11	0.91	0.004
		FSR LRT	0.85	0.55	9.05	0.55	2.21	0.83	0.004
		Full LRT	0.78	0.13	11.05	0.17	2.10	0.92	0.006
	$\mathbf{R}_2$	AIC	0.93	0.63	10.28	0.64	2.11	0.91	0.004
		FSR LRT	0.85	0.55	9.05	0.55	2.21	0.83	0.004
		Full LRT	0.78	0.13	11.05	0.17	2.10	0.92	0.006
SMC	**R**	AIC	0.93	0.63	10.32	0.66	2.07	0.94	0.004
		FSR LRT	0.93	0.65	9.87	0.66	2.14	0.90	0.003
		Full LRT	0.78	0.10	11.36	0.15	2.07	0.93	0.006
	$\mathbf{R}_2$	AIC	0.94	0.64	10.32	0.64	2.06	0.95	0.003
		FSR LRT	0.91	0.63	9.62	0.63	2.17	0.87	0.004
		Full LRT	0.80	0.13	11.32	0.14	2.04	0.96	0.006
LR	**R**	AIC	0.93	0.63	10.32	0.66	2.07	0.94	0.004
		FSR LRT	0.93	0.65	9.87	0.66	2.14	0.90	0.003
		Full LRT	0.78	0.10	11.36	0.15	2.07	0.93	0.006
	$\mathbf{R}_2$	AIC	0.93	0.63	10.32	0.66	2.07	0.94	0.004
		FSR LRT	0.93	0.65	9.87	0.66	2.14	0.90	0.003
		Full LRT	0.78	0.10	11.36	0.15	2.07	0.93	0.006
Traditional Communality			-0.03	0.00	23.11	0.00	1.38	0.33	0.042
Traditional Magnitude			0.76	0.15	11.25	0.22	2.13	0.90	0.007
GLAFA			0.34	0.17	16.21	0.17	1.38	0.36	0.028
Average Standard Error			0.01		0.12		0.04		0.000

Table B.12 Results for $k = 2$, 9 correlated variables per factor, $u = 10$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.98	0.88	9.91	0.88	2.01	0.99	0.001
		FSR LRT	0.94	0.74	9.58	0.74	2.05	0.95	0.001
		Full LRT	0.93	0.51	10.39	0.53	2.03	0.97	0.002
	$\mathbf{R}_2$	AIC	0.99	0.92	9.94	0.92	2	1	0.001
		FSR LRT	0.88	0.68	9.12	0.68	2.03	0.97	0.001
		Full LRT	0.93	0.54	10.39	0.54	2	1	0.002
AAC	**R**	AIC	0.99	0.95	9.97	0.95	2	1	0.001
		FSR LRT	0.87	0.68	8.98	0.68	2.07	0.94	0.002
		Full LRT	0.93	0.56	10.36	0.56	2.01	0.99	0.002
	$\mathbf{R}_2$	AIC	0.99	0.94	9.96	0.94	2	1	0.001
		FSR LRT	0.88	0.68	9.09	0.68	2.06	0.95	0.002
		Full LRT	0.94	0.58	10.34	0.58	2	1	0.002
MAC	**R**	AIC	0.99	0.92	9.94	0.92	2.01	0.99	0.001
		FSR LRT	0.86	0.67	8.84	0.67	2.07	0.93	0.002
		Full LRT	0.93	0.56	10.35	0.56	2.01	0.99	0.002
	$\mathbf{R}_2$	AIC	0.99	0.92	9.94	0.92	2.01	0.99	0.001
		FSR LRT	0.86	0.67	8.84	0.67	2.07	0.93	0.002
		Full LRT	0.93	0.56	10.35	0.56	2.01	0.99	0.002
SMC	**R**	AIC	0.98	0.88	9.91	0.88	2.01	0.99	0.001
		FSR LRT	0.95	0.76	9.71	0.76	2.02	0.98	0.001
		Full LRT	0.93	0.52	10.40	0.52	2.03	0.97	0.002
	$\mathbf{R}_2$	AIC	0.99	0.92	9.94	0.92	2	1	0.001
		FSR LRT	0.90	0.71	9.30	0.71	2.05	0.95	0.001
		Full LRT	0.93	0.53	10.40	0.53	2.01	0.99	0.002
LR	**R**	AIC	0.98	0.88	9.91	0.88	2.01	0.99	0.001
		FSR LRT	0.95	0.76	9.71	0.76	2.02	0.98	0.001
		Full LRT	0.93	0.52	10.40	0.52	2.03	0.97	0.002
	$\mathbf{R}_2$	AIC	0.98	0.88	9.91	0.88	2.01	0.99	0.001
		FSR LRT	0.95	0.76	9.71	0.76	2.02	0.98	0.001
		Full LRT	0.93	0.52	10.40	0.52	2.03	0.97	0.002
Traditional Communality			0.41	0.00	14.83	0.00	2	1	0.015
Traditional Magnitude			0.97	0.78	10.17	0.79	2	1	0.002
GLAFA			0.94	0.63	9.62	0.63	2	1	0.008
Average Standard Error			0.01		0.08		0.01		0.000

Table B.13 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
AAC	**R**	AIC	0.97	0.97	0.97	0.97	5	1	0.004
		FSR LRT	0.97	0.97	0.97	0.97	5	1	0.004
		Full LRT	0.97	0.97	0.97	0.97	5	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
MAC	**R**	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
SMC	**R**	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
LR	**R**	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
	$\mathbf{R}_2$	AIC	0.99	0.99	0.99	0.99	5	1	0.004
		FSR LRT	0.98	0.98	0.98	0.98	5	1	0.004
		Full LRT	0.98	0.98	0.98	0.98	5	1	0.004
Traditional Communality			1	1	1	1	5	1	0.004
Traditional Magnitude			0.99	0.99	0.99	0.99	5	1	0.004
GLAFA			0.53	0.45	1.96	0.46	4.56	0.59	0.015
Average Standard Error			0.01		0.02		0.00		0.000

Table B.14 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	-0.05	0.00	10.76	0.00	1.81	0.01	0.030
		FSR LRT	-0.04	0.00	13.39	0.00	0.78	0.00	0.030
		Full LRT	-0.03	0.00	12.94	0.00	0.79	0.00	0.030
	$\mathbf{R}_2$	AIC	-0.05	0.00	10.79	0.00	1.76	0.00	0.030
		FSR LRT	-0.05	0.00	13.71	0.00	0.55	0.00	0.030
		Full LRT	-0.04	0.00	13.35	0.00	0.73	0.00	0.030
AAC	**R**	AIC	-0.03	0.00	10.47	0.00	1.62	0.02	0.030
		FSR LRT	-0.01	0.00	13.09	0.00	0.89	0.01	0.030
		Full LRT	-0.02	0.00	12.65	0.00	0.77	0.01	0.030
	$\mathbf{R}_2$	AIC	-0.05	0.00	10.44	0.00	1.75	0.01	0.030
		FSR LRT	-0.03	0.00	13.39	0.00	0.72	0.01	0.030
		Full LRT	-0.03	0.00	13.03	0.00	0.74	0.00	0.030
MAC	**R**	AIC	-0.04	0.00	10.89	0.00	1.85	0.02	0.030
		FSR LRT	-0.05	0.00	12.89	0.00	0.85	0.03	0.030
		Full LRT	-0.03	0.00	13.29	0.00	0.68	0.01	0.029
	$\mathbf{R}_2$	AIC	-0.04	0.00	10.89	0.00	1.85	0.02	0.030
		FSR LRT	-0.05	0.00	12.89	0.00	0.85	0.03	0.030
		Full LRT	-0.03	0.00	13.29	0.00	0.68	0.01	0.029
SMC	**R**	AIC	-0.04	0.00	10.57	0.00	1.95	0.01	0.030
		FSR LRT	-0.05	0.00	14.37	0.00	0.21	0.00	0.030
		Full LRT	-0.04	0.00	13.59	0.00	0.66	0.00	0.030
	$\mathbf{R}_2$	AIC	-0.04	0.00	10.66	0.00	1.93	0.00	0.030
		FSR LRT	-0.06	0.00	13.93	0.00	0.36	0.00	0.030
		Full LRT	-0.04	0.00	13.58	0.00	0.65	0.00	0.030
LR	**R**	AIC	-0.04	0.00	10.57	0.00	1.95	0.01	0.030
		FSR LRT	-0.05	0.00	14.37	0.00	0.21	0.00	0.030
		Full LRT	-0.04	0.00	13.59	0.00	0.66	0.00	0.030
	$\mathbf{R}_2$	AIC	-0.04	0.00	10.57	0.00	1.95	0.01	0.030
		FSR LRT	-0.05	0.00	14.37	0.00	0.21	0.00	0.030
		Full LRT	-0.04	0.00	13.59	0.00	0.66	0.00	0.030
Traditional Communality			-0.03	0.00	13.55	0.00	0.68	0.03	0.030
Traditional Magnitude			-0.01	0.00	11.94	0.00	1.16	0.02	0.029
GLAFA			-0.06	0.00	11.27	0.00	1.01	0.00	0.030
Average Standard Error			0.01		0.24		0.10		0.000

Table B.15 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.28	0.02	4.03	0.02	4.95	0.53	0.020
		FSR LRT	0.38	0.05	3.20	0.05	5.18	0.57	0.018
		Full LRT	0.31	0.01	3.62	0.02	5.10	0.57	0.019
	$\mathbf{R}_2$	AIC	0.30	0.03	4.03	0.03	4.90	0.53	0.019
		FSR LRT	0.41	0.09	3.06	0.10	5.16	0.62	0.017
		Full LRT	0.33	0.04	3.51	0.06	5.08	0.65	0.018
AAC	**R**	AIC	0.36	0.08	3.27	0.11	4.92	0.59	0.020
		FSR LRT	0.43	0.13	2.20	0.21	5.28	0.57	0.017
		Full LRT	0.35	0.07	2.62	0.14	5.09	0.63	0.019
	$\mathbf{R}_2$	AIC	0.29	0.03	4.04	0.04	4.81	0.49	0.021
		FSR LRT	0.41	0.06	2.93	0.08	5.17	0.61	0.017
		Full LRT	0.32	0.02	3.49	0.05	5.01	0.58	0.019
MAC	**R**	AIC	0.27	0.03	4.42	0.03	4.78	0.45	0.021
		FSR LRT	0.42	0.12	2.94	0.14	5.16	0.65	0.016
		Full LRT	0.35	0.07	3.24	0.10	5.18	0.63	0.017
	$\mathbf{R}_2$	AIC	0.27	0.03	4.42	0.03	4.78	0.45	0.021
		FSR LRT	0.42	0.12	2.94	0.14	5.16	0.65	0.016
		Full LRT	0.35	0.07	3.24	0.10	5.18	0.63	0.017
SMC	**R**	AIC	0.22	0.00	4.59	0.00	4.71	0.44	0.022
		FSR LRT	0.34	0.05	3.65	0.05	5.06	0.55	0.019
		Full LRT	0.30	0.02	3.80	0.02	5.13	0.51	0.019
	$\mathbf{R}_2$	AIC	0.26	0.03	4.46	0.03	4.73	0.44	0.021
		FSR LRT	0.41	0.09	3.14	0.10	5.13	0.62	0.017
		Full LRT	0.32	0.04	3.55	0.06	5.06	0.63	0.018
LR	**R**	AIC	0.22	0.00	4.59	0.00	4.71	0.44	0.022
		FSR LRT	0.34	0.05	3.65	0.05	5.06	0.55	0.019
		Full LRT	0.30	0.02	3.80	0.02	5.13	0.51	0.019
	$\mathbf{R}_2$	AIC	0.22	0.00	4.59	0.00	4.71	0.44	0.022
		FSR LRT	0.34	0.05	3.65	0.05	5.06	0.55	0.019
		Full LRT	0.30	0.02	3.80	0.02	5.13	0.51	0.019
Traditional Communality			0.32	0.05	2.92	0.12	5.31	0.53	0.018
Traditional Magnitude			0.50	0.23	1.78	0.31	5.43	0.64	0.015
GLAFA			-0.03	0.00	9.49	0.00	1.68	0.00	0.052
Average Standard Error			0.02		0.16		0.08		0.001

Table B.16 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
AAC	**R**	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
MAC	**R**	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
SMC	**R**	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
LR	**R**	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	0.96	0.96	0.96	0.96	5	1	0.001
		FSR LRT	0.94	0.94	0.94	0.94	5	1	0.001
		Full LRT	0.94	0.94	0.94	0.94	5	1	0.001
Traditional Communality			1	1	1	1	5	1	0.001
Traditional Magnitude			1	1	1	1	5	1	0.001
GLAFA			0.57	0.57	0.57	0.57	5.06	0.94	0.001
Average Standard Error			0.02		0.02		0.00		0.000

Table B.17 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.15	0.02	5.49	0.02	3.97	0.20	0.018
		FSR LRT	0.30	0.08	4.07	0.08	4.39	0.40	0.016
		Full LRT	0.26	0.02	4.00	0.04	4.54	0.46	0.017
	$\mathbf{R}_2$	AIC	0.12	0.03	6.22	0.03	3.68	0.12	0.018
		FSR LRT	0.34	0.10	3.93	0.10	4.39	0.40	0.015
		Full LRT	0.28	0.03	3.84	0.03	4.50	0.43	0.016
AAC	**R**	AIC	0.20	0.04	4.95	0.04	4.09	0.31	0.019
		FSR LRT	0.35	0.11	3.02	0.15	4.61	0.47	0.014
		Full LRT	0.27	0.05	3.28	0.10	4.59	0.44	0.016
	$\mathbf{R}_2$	AIC	0.14	0.02	5.73	0.02	3.87	0.19	0.018
		FSR LRT	0.30	0.06	4.05	0.07	4.34	0.43	0.015
		Full LRT	0.28	0.04	3.87	0.06	4.45	0.49	0.016
MAC	**R**	AIC	0.10	0.02	6.56	0.02	3.60	0.09	0.018
		FSR LRT	0.41	0.19	3.43	0.19	4.59	0.47	0.014
		Full LRT	0.32	0.07	3.53	0.08	4.66	0.49	0.015
	$\mathbf{R}_2$	AIC	0.10	0.02	6.56	0.02	3.60	0.09	0.018
		FSR LRT	0.41	0.19	3.43	0.19	4.59	0.47	0.014
		Full LRT	0.32	0.07	3.53	0.08	4.66	0.49	0.015
SMC	**R**	AIC	0.06	0.00	6.66	0.00	3.48	0.10	0.018
		FSR LRT	0.17	0.05	5.62	0.05	3.86	0.20	0.017
		Full LRT	0.20	0.01	4.58	0.03	4.30	0.36	0.016
	$\mathbf{R}_2$	AIC	0.09	0.01	6.48	0.01	3.62	0.12	0.018
		FSR LRT	0.28	0.10	4.58	0.10	4.17	0.35	0.015
		Full LRT	0.26	0.03	4.10	0.03	4.42	0.41	0.015
LR	**R**	AIC	0.06	0.00	6.66	0.00	3.48	0.10	0.018
		FSR LRT	0.17	0.05	5.62	0.05	3.86	0.20	0.017
		Full LRT	0.20	0.01	4.58	0.03	4.30	0.36	0.016
	$\mathbf{R}_2$	AIC	0.06	0.00	6.66	0.00	3.48	0.10	0.018
		FSR LRT	0.17	0.05	5.62	0.05	3.86	0.20	0.017
		Full LRT	0.20	0.01	4.58	0.03	4.30	0.36	0.016
Traditional Communality			-0.04	0.00	9.82	0.00	1.67	0.00	0.029
Traditional Magnitude			0.18	0.02	5.03	0.02	4.23	0.44	0.019
GLAFA			-0.04	0.00	10.01	0.00	1.16	0.00	0.029
Average Standard Error			0.02		0.21		0.09		0.000

Table B.18 Results for $k = 5$, 3 correlated variables per factor, $u = 1$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.89	0.99	0.89	5.10	0.90	0.003
	$\mathbf{R}_2$	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.90	0.98	0.90	5.10	0.90	0.003
AAC	**R**	AIC	0.94	0.93	0.95	0.93	5.10	0.90	0.003
		FSR LRT	0.92	0.91	0.93	0.91	5.12	0.88	0.003
		Full LRT	0.90	0.88	0.96	0.88	5.12	0.88	0.003
	$\mathbf{R}_2$	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.89	0.99	0.89	5.10	0.90	0.003
MAC	**R**	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.93	0.91	0.97	0.91	5.10	0.90	0.003
	$\mathbf{R}_2$	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.93	0.91	0.97	0.91	5.10	0.90	0.003
SMC	**R**	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.89	0.99	0.89	5.10	0.90	0.003
	$\mathbf{R}_2$	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.90	0.98	0.90	5.10	0.90	0.003
LR	**R**	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.89	0.99	0.89	5.10	0.90	0.003
	$\mathbf{R}_2$	AIC	0.96	0.95	0.97	0.95	5.09	0.91	0.003
		FSR LRT	0.94	0.93	0.95	0.93	5.11	0.89	0.003
		Full LRT	0.92	0.89	0.99	0.89	5.10	0.90	0.003
Traditional Communality			0.22	0.00	4.27	0.00	5.03	0.93	0.013
Traditional Magnitude			0.82	0.66	1.36	0.67	5.08	0.92	0.004
GLAFA			0.55	0.51	1.21	0.51	4.81	0.82	0.005
Average Standard Error			0.02		0.03		0.03		0.000

Table B.19 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.97	0.83	9.79	0.83	5.05	0.95	0.003
		FSR LRT	0.94	0.79	9.62	0.79	5.12	0.90	0.004
		Full LRT	0.91	0.48	10.38	0.48	5.03	0.95	0.006
	$\mathbf{R}_2$	AIC	0.98	0.89	9.86	0.89	5.01	0.99	0.003
		FSR LRT	0.87	0.73	9.05	0.73	5.20	0.85	0.004
		Full LRT	0.91	0.50	10.31	0.50	5.04	0.95	0.005
AAC	**R**	AIC	0.86	0.45	9.17	0.45	5.03	0.95	0.004
		FSR LRT	0.79	0.37	8.56	0.37	5.15	0.88	0.005
		Full LRT	0.83	0.31	9.61	0.38	5.04	0.94	0.006
	$\mathbf{R}_2$	AIC	0.98	0.88	9.85	0.88	5.01	0.99	0.003
		FSR LRT	0.90	0.74	9.25	0.74	5.17	0.88	0.004
		Full LRT	0.91	0.50	10.33	0.50	5.04	0.95	0.006
MAC	**R**	AIC	0.98	0.91	9.88	0.91	5.01	0.99	0.003
		FSR LRT	0.84	0.69	8.81	0.69	5.22	0.84	0.005
		Full LRT	0.91	0.50	10.34	0.50	5.04	0.96	0.005
	$\mathbf{R}_2$	AIC	0.98	0.91	9.88	0.91	5.01	0.99	0.003
		FSR LRT	0.84	0.69	8.81	0.69	5.22	0.84	0.005
		Full LRT	0.91	0.50	10.34	0.50	5.04	0.96	0.005
SMC	**R**	AIC	0.96	0.80	9.75	0.80	5.06	0.94	0.003
		FSR LRT	0.95	0.79	9.69	0.79	5.09	0.92	0.004
		Full LRT	0.91	0.47	10.43	0.47	5.01	0.97	0.006
	$\mathbf{R}_2$	AIC	0.98	0.88	9.85	0.88	5.01	0.99	0.003
		FSR LRT	0.88	0.73	9.15	0.73	5.20	0.85	0.004
		Full LRT	0.91	0.51	10.38	0.51	5.03	0.95	0.005
LR	**R**	AIC	0.96	0.80	9.75	0.80	5.06	0.94	0.003
		FSR LRT	0.95	0.79	9.69	0.79	5.09	0.92	0.004
		Full LRT	0.91	0.47	10.43	0.47	5.01	0.97	0.006
	$\mathbf{R}_2$	AIC	0.96	0.80	9.75	0.80	5.06	0.94	0.003
		FSR LRT	0.95	0.79	9.69	0.79	5.09	0.92	0.004
		Full LRT	0.91	0.47	10.43	0.47	5.01	0.97	0.006
Traditional Communality			0.87	0.54	9.12	0.54	5.25	0.79	0.004
Traditional Magnitude			0.81	0.43	8.68	0.43	5.37	0.69	0.005
GLAFA			0.65	0.24	12.41	0.24	4.18	0.25	0.022
Average Standard Error			0.01		0.10		0.03		0.000

Table B.20 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.00	0.00	18.61	0.00	2.12	0.02	0.023
		FSR LRT	0.00	0.00	21.55	0.00	1.12	0.02	0.021
		Full LRT	-0.01	0.00	23.04	0.00	0.58	0.01	0.020
	$\mathbf{R}_2$	AIC	0.00	0.00	19.05	0.00	2.10	0.04	0.022
		FSR LRT	0.00	0.00	21.86	0.01	0.87	0.03	0.021
		Full LRT	-0.01	0.00	23.28	0.00	0.54	0.01	0.020
AAC	**R**	AIC	-0.01	0.00	19.84	0.00	1.47	0.02	0.021
		FSR LRT	0.01	0.00	21.75	0.01	1.01	0.02	0.021
		Full LRT	-0.01	0.00	22.71	0.01	0.40	0.00	0.020
	$\mathbf{R}_2$	AIC	0.00	0.00	19.10	0.00	1.80	0.01	0.021
		FSR LRT	0.00	0.00	22.16	0.01	0.87	0.01	0.020
		Full LRT	0.00	0.00	23.12	0.00	0.50	0.00	0.020
MAC	**R**	AIC	-0.01	0.00	19.67	0.00	1.85	0.00	0.021
		FSR LRT	0.00	0.00	21.38	0.02	0.99	0.03	0.021
		Full LRT	-0.01	0.00	23.42	0.00	0.46	0.01	0.020
	$\mathbf{R}_2$	AIC	-0.01	0.00	19.67	0.00	1.85	0.00	0.021
		FSR LRT	0.00	0.00	21.38	0.02	0.99	0.03	0.021
		Full LRT	-0.01	0.00	23.42	0.00	0.46	0.01	0.020
SMC	**R**	AIC	0.00	0.00	18.44	0.00	2.53	0.04	0.024
		FSR LRT	-0.02	0.00	23.37	0.00	0.29	0.00	0.020
		Full LRT	-0.01	0.00	23.64	0.00	0.40	0.01	0.020
	$\mathbf{R}_2$	AIC	-0.01	0.00	19.02	0.00	2.13	0.01	0.022
		FSR LRT	-0.02	0.00	22.82	0.00	0.39	0.01	0.020
		Full LRT	-0.01	0.00	23.57	0.00	0.44	0.01	0.020
LR	**R**	AIC	0.00	0.00	18.44	0.00	2.53	0.04	0.024
		FSR LRT	-0.02	0.00	23.37	0.00	0.29	0.00	0.020
		Full LRT	-0.01	0.00	23.64	0.00	0.40	0.01	0.020
	$\mathbf{R}_2$	AIC	0.00	0.00	18.44	0.00	2.53	0.04	0.024
		FSR LRT	-0.02	0.00	23.37	0.00	0.29	0.00	0.020
		Full LRT	-0.01	0.00	23.64	0.00	0.40	0.01	0.020
Traditional Communality			0.00	0.00	24.22	0.00	0.16	0.00	0.019
Traditional Magnitude			0.00	0.00	23.54	0.00	0.35	0.00	0.019
GLAFA			0.00	0.00	18.30	0.00	1.01	0.00	0.021
Average Standard Error			0.00		0.26		0.11		0.000

Table B.21 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.48	0.01	12.80	0.06	5.02	0.39	0.015
		FSR LRT	0.48	0.02	12.39	0.10	5.36	0.38	0.015
		Full LRT	0.36	0.00	13.90	0.02	4.72	0.31	0.018
	$\mathbf{R}_2$	AIC	0.54	0.02	12.84	0.08	4.97	0.52	0.013
		FSR LRT	0.50	0.04	12.21	0.08	5.30	0.42	0.014
		Full LRT	0.40	0.00	14.13	0.02	4.66	0.44	0.016
AAC	**R**	AIC	0.40	0.00	11.60	0.13	4.97	0.36	0.018
		FSR LRT	0.42	0.00	8.47	0.10	5.95	0.33	0.016
		Full LRT	0.33	0.00	11.64	0.15	4.78	0.25	0.021
	$\mathbf{R}_2$	AIC	0.50	0.02	12.73	0.12	4.94	0.38	0.015
		FSR LRT	0.52	0.02	12.00	0.15	5.22	0.39	0.014
		Full LRT	0.39	0.00	13.81	0.04	4.44	0.32	0.017
MAC	**R**	AIC	0.50	0.02	13.37	0.07	4.86	0.42	0.014
		FSR LRT	0.50	0.04	11.36	0.07	5.60	0.45	0.014
		Full LRT	0.40	0.00	14.04	0.01	4.77	0.49	0.015
	$\mathbf{R}_2$	AIC	0.50	0.02	13.37	0.07	4.86	0.42	0.014
		FSR LRT	0.50	0.04	11.36	0.07	5.60	0.45	0.014
		Full LRT	0.40	0.00	14.04	0.01	4.77	0.49	0.015
SMC	**R**	AIC	0.45	0.00	13.30	0.04	4.94	0.40	0.016
		FSR LRT	0.39	0.00	14.29	0.01	4.77	0.38	0.016
		Full LRT	0.32	0.00	14.67	0.02	4.58	0.35	0.018
	$\mathbf{R}_2$	AIC	0.50	0.02	13.35	0.08	4.78	0.42	0.014
		FSR LRT	0.48	0.04	12.61	0.09	5.22	0.41	0.015
		Full LRT	0.38	0.00	14.39	0.02	4.58	0.38	0.016
LR	**R**	AIC	0.45	0.00	13.30	0.04	4.94	0.40	0.016
		FSR LRT	0.39	0.00	14.29	0.01	4.77	0.38	0.016
		Full LRT	0.32	0.00	14.67	0.02	4.58	0.35	0.018
	$\mathbf{R}_2$	AIC	0.45	0.00	13.30	0.04	4.94	0.40	0.016
		FSR LRT	0.39	0.00	14.29	0.01	4.77	0.38	0.016
		Full LRT	0.32	0.00	14.67	0.02	4.58	0.35	0.018
Traditional Communality			0.47	0.01	10.82	0.13	5.71	0.37	0.015
Traditional Magnitude			0.51	0.04	9.24	0.14	5.95	0.34	0.014
GLAFA			0.14	0.00	16.84	0.00	1.49	0.00	0.035
Average Standard Error			0.02		0.25		0.12		0.001

Table B.22 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.98	0.87	9.87	0.87	5.03	0.97	0.001
		FSR LRT	0.96	0.84	9.75	0.84	5.07	0.93	0.001
		Full LRT	0.99	0.96	9.94	0.96	5.03	0.97	0.001
	$\mathbf{R}_2$	AIC	0.99	0.91	9.91	0.91	5.01	0.99	0.001
		FSR LRT	0.88	0.75	9.05	0.75	5.19	0.85	0.002
		Full LRT	0.99	0.96	9.93	0.96	5	1	0.001
AAC	**R**	AIC	0.99	0.91	9.91	0.91	5	1	0.001
		FSR LRT	0.90	0.80	9.28	0.80	5.15	0.88	0.001
		Full LRT	0.99	0.96	9.94	0.96	5	1	0.001
	$\mathbf{R}_2$	AIC	0.99	0.91	9.91	0.91	5	1	0.001
		FSR LRT	0.92	0.80	9.42	0.80	5.12	0.90	0.001
		Full LRT	0.99	0.96	9.94	0.96	5	1	0.001
MAC	**R**	AIC	0.99	0.91	9.91	0.91	5.01	0.99	0.001
		FSR LRT	0.90	0.77	9.29	0.77	5.18	0.86	0.001
		Full LRT	0.99	0.96	9.92	0.96	5.01	0.99	0.001
	$\mathbf{R}_2$	AIC	0.99	0.91	9.91	0.91	5.01	0.99	0.001
		FSR LRT	0.90	0.77	9.29	0.77	5.18	0.86	0.001
		Full LRT	0.99	0.96	9.92	0.96	5.01	0.99	0.001
SMC	**R**	AIC	0.98	0.87	9.87	0.87	5.03	0.97	0.001
		FSR LRT	0.98	0.85	9.84	0.85	5.04	0.96	0.001
		Full LRT	0.99	0.96	9.94	0.96	5.03	0.97	0.001
	$\mathbf{R}_2$	AIC	0.99	0.91	9.91	0.91	5	1	0.001
		FSR LRT	0.93	0.80	9.48	0.80	5.10	0.91	0.001
		Full LRT	0.99	0.96	9.94	0.96	5	1	0.001
LR	**R**	AIC	0.98	0.87	9.87	0.87	5.03	0.97	0.001
		FSR LRT	0.98	0.85	9.84	0.85	5.04	0.96	0.001
		Full LRT	0.99	0.96	9.94	0.96	5.03	0.97	0.001
	$\mathbf{R}_2$	AIC	0.98	0.87	9.87	0.87	5.03	0.97	0.001
		FSR LRT	0.98	0.85	9.84	0.85	5.04	0.96	0.001
		Full LRT	0.99	0.96	9.94	0.96	5.03	0.97	0.001
Traditional Communality			0.94	0.69	9.61	0.69	5.04	0.96	0.001
Traditional Magnitude			0.90	0.58	9.38	0.58	5.10	0.92	0.001
GLAFA			0.90	0.47	9.38	0.47	5	1	0.001
Average Standard Error			0.01		0.06		0.02		0.000

Table B.23 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.34	0.03	14.57	0.04	4.05	0.30	0.014
		FSR LRT	0.38	0.03	13.59	0.09	4.31	0.33	0.013
		Full LRT	0.28	0.00	14.62	0.04	4.09	0.24	0.015
	$\mathbf{R}_2$	AIC	0.34	0.03	14.96	0.03	3.77	0.14	0.012
		FSR LRT	0.39	0.04	14.13	0.07	4.05	0.24	0.012
		Full LRT	0.33	0.00	15.00	0.01	3.78	0.18	0.013
AAC	**R**	AIC	0.27	0.01	13.96	0.07	4.12	0.25	0.014
		FSR LRT	0.32	0.01	10.30	0.15	5.35	0.24	0.013
		Full LRT	0.23	0.00	12.97	0.09	4.44	0.23	0.015
	$\mathbf{R}_2$	AIC	0.36	0.03	14.49	0.04	3.94	0.26	0.013
		FSR LRT	0.40	0.03	13.62	0.07	4.21	0.24	0.013
		Full LRT	0.30	0.00	14.67	0.04	3.86	0.21	0.013
MAC	**R**	AIC	0.29	0.02	15.47	0.03	3.60	0.09	0.012
		FSR LRT	0.44	0.05	12.83	0.09	4.53	0.29	0.011
		Full LRT	0.31	0.00	14.86	0.01	3.88	0.20	0.013
	$\mathbf{R}_2$	AIC	0.29	0.02	15.47	0.03	3.60	0.09	0.012
		FSR LRT	0.44	0.05	12.83	0.09	4.53	0.29	0.011
		Full LRT	0.31	0.00	14.86	0.01	3.88	0.20	0.013
SMC	**R**	AIC	0.25	0.00	15.45	0.00	3.54	0.10	0.012
		FSR LRT	0.21	0.01	16.78	0.02	3.10	0.11	0.013
		Full LRT	0.23	0.00	15.93	0.01	3.49	0.08	0.013
	$\mathbf{R}_2$	AIC	0.30	0.01	15.41	0.01	3.60	0.11	0.011
		FSR LRT	0.32	0.04	15.26	0.05	3.60	0.15	0.012
		Full LRT	0.28	0.00	15.55	0.01	3.56	0.08	0.012
LR	**R**	AIC	0.25	0.00	15.45	0.00	3.54	0.10	0.012
		FSR LRT	0.21	0.01	16.78	0.02	3.10	0.11	0.013
		Full LRT	0.23	0.00	15.93	0.01	3.49	0.08	0.013
	$\mathbf{R}_2$	AIC	0.25	0.00	15.45	0.00	3.54	0.10	0.012
		FSR LRT	0.21	0.01	16.78	0.02	3.10	0.11	0.013
		Full LRT	0.23	0.00	15.93	0.01	3.49	0.08	0.013
Traditional Communality			0.06	0.00	18.42	0.00	2.05	0.06	0.018
Traditional Magnitude			0.29	0.01	14.32	0.06	3.96	0.23	0.014
GLAFA			0.07	0.00	16.96	0.00	1.19	0.00	0.021
Average Standard Error			0.02		0.27		0.12		0.000

Table B.24 Results for $k = 5$, 3 correlated variables per factor, $u = 10$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.98	0.87	9.89	0.87	5.18	0.84	0.002
		FSR LRT	0.97	0.84	9.80	0.84	5.24	0.81	0.002
		Full LRT	0.89	0.41	10.61	0.41	5.12	0.91	0.003
	$\mathbf{R}_2$	AIC	0.98	0.87	9.88	0.87	5.17	0.85	0.002
		FSR LRT	0.93	0.81	9.51	0.81	5.30	0.79	0.003
		Full LRT	0.89	0.41	10.62	0.41	5.09	0.94	0.003
AAC	**R**	AIC	0.90	0.55	9.47	0.57	5.30	0.75	0.003
		FSR LRT	0.84	0.50	8.95	0.50	5.53	0.66	0.003
		Full LRT	0.85	0.30	9.91	0.37	5.29	0.77	0.004
	$\mathbf{R}_2$	AIC	0.98	0.89	9.91	0.89	5.15	0.87	0.002
		FSR LRT	0.92	0.79	9.48	0.79	5.31	0.77	0.003
		Full LRT	0.89	0.41	10.62	0.41	5.10	0.92	0.003
MAC	**R**	AIC	0.98	0.90	9.92	0.90	5.12	0.88	0.002
		FSR LRT	0.90	0.77	9.24	0.77	5.44	0.77	0.003
		Full LRT	0.89	0.41	10.61	0.41	5.10	0.93	0.003
	$\mathbf{R}_2$	AIC	0.98	0.90	9.92	0.90	5.12	0.88	0.002
		FSR LRT	0.90	0.77	9.24	0.77	5.44	0.77	0.003
		Full LRT	0.89	0.41	10.61	0.41	5.10	0.93	0.003
SMC	**R**	AIC	0.98	0.86	9.88	0.86	5.18	0.84	0.002
		FSR LRT	0.97	0.84	9.85	0.84	5.21	0.82	0.002
		Full LRT	0.88	0.41	10.62	0.41	5.12	0.91	0.003
	$\mathbf{R}_2$	AIC	0.98	0.89	9.91	0.89	5.13	0.87	0.002
		FSR LRT	0.95	0.82	9.64	0.82	5.27	0.80	0.003
		Full LRT	0.89	0.41	10.61	0.41	5.09	0.94	0.003
LR	**R**	AIC	0.98	0.86	9.88	0.86	5.18	0.84	0.002
		FSR LRT	0.97	0.84	9.85	0.84	5.21	0.82	0.002
		Full LRT	0.88	0.41	10.62	0.41	5.12	0.91	0.003
	$\mathbf{R}_2$	AIC	0.98	0.86	9.88	0.86	5.18	0.84	0.002
		FSR LRT	0.97	0.84	9.85	0.84	5.21	0.82	0.002
		Full LRT	0.88	0.41	10.62	0.41	5.12	0.91	0.003
Traditional Communality			0.54	0.01	11.84	0.20	5.33	0.74	0.008
Traditional Magnitude			0.78	0.24	9.19	0.33	5.65	0.57	0.004
GLAFA			0.65	0.26	12.07	0.26	4.32	0.47	0.010
Average Standard Error			0.01		0.09		0.05		0.000

Table B.25 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.91	0.88	0.98	0.88	5	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.89	0.97	0.89	5	1	0.005
AAC	**R**	AIC	0.96	0.95	0.97	0.95	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.89	0.97	0.89	5	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.89	0.97	0.89	5	1	0.005
MAC	**R**	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.90	0.96	0.90	5	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.90	0.96	0.90	5	1	0.005
SMC	**R**	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.91	0.87	0.99	0.87	5	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.92	0.89	0.97	0.89	5	1	0.005
LR	**R**	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.91	0.87	0.99	0.87	5	1	0.005
	$\mathbf{R}_2$	AIC	0.97	0.96	0.98	0.96	5	1	0.005
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.005
		Full LRT	0.91	0.87	0.99	0.87	5	1	0.005
Traditional Communality			0.99	0.99	0.99	0.99	5	1	0.005
Traditional Magnitude			0.97	0.97	0.97	0.97	5	1	0.005
GLAFA			0.39	0.39	4.67	0.39	4.55	0.77	0.023
Average Standard Error			0.02		0.05		0.00		0.000

Table B.26 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	-0.02	0.00	32.97	0.00	4.82	0.18	0.029
		FSR LRT	0.00	0.00	27.73	0.00	6.27	0.12	0.026
		Full LRT	0.01	0.00	23.35	0.00	7.00	0.19	0.025
	$\mathbf{R}_2$	AIC	-0.03	0.00	34.41	0.00	3.59	0.17	0.029
		FSR LRT	0.02	0.00	24.12	0.00	4.45	0.25	0.025
		Full LRT	0.02	0.01	21.90	0.01	5.10	0.32	0.024
AAC	**R**	AIC	-0.03	0.00	34.44	0.00	2.91	0.07	0.029
		FSR LRT	0.03	0.00	26.23	0.01	3.80	0.19	0.026
		Full LRT	0.02	0.00	20.90	0.00	4.93	0.29	0.025
	$\mathbf{R}_2$	AIC	-0.03	0.00	33.77	0.00	3.37	0.14	0.029
		FSR LRT	0.02	0.00	24.83	0.00	4.22	0.23	0.025
		Full LRT	0.02	0.01	22.27	0.01	4.81	0.29	0.025
MAC	**R**	AIC	-0.02	0.00	35.60	0.00	3.37	0.10	0.029
		FSR LRT	0.02	0.00	21.99	0.02	4.62	0.28	0.024
		Full LRT	0.02	0.00	19.37	0.02	5.18	0.32	0.023
	$\mathbf{R}_2$	AIC	-0.02	0.00	35.60	0.00	3.37	0.10	0.029
		FSR LRT	0.02	0.00	21.99	0.02	4.62	0.28	0.024
		Full LRT	0.02	0.00	19.37	0.02	5.18	0.32	0.023
SMC	**R**	AIC	-0.02	0.00	32.40	0.00	5.98	0.14	0.028
		FSR LRT	-0.02	0.00	40.42	0.00	2.25	0.04	0.031
		Full LRT	-0.01	0.00	26.74	0.00	8.07	0.07	0.025
	$\mathbf{R}_2$	AIC	-0.02	0.00	34.53	0.00	3.14	0.07	0.028
		FSR LRT	0.01	0.00	28.56	0.00	3.76	0.13	0.026
		Full LRT	0.02	0.01	23.45	0.01	4.95	0.26	0.024
LR	**R**	AIC	-0.02	0.00	32.40	0.00	5.98	0.14	0.028
		FSR LRT	-0.02	0.00	40.42	0.00	2.25	0.04	0.031
		Full LRT	-0.01	0.00	26.74	0.00	8.07	0.07	0.025
	$\mathbf{R}_2$	AIC	-0.02	0.00	32.40	0.00	5.98	0.14	0.028
		FSR LRT	-0.02	0.00	40.42	0.00	2.25	0.04	0.031
		Full LRT	-0.01	0.00	26.74	0.00	8.07	0.07	0.025
Traditional Communality			0.02	0.00	20.65	0.00	5.03	0.52	0.024
Traditional Magnitude			0.05	0.00	13.85	0.00	5.11	0.54	0.020
GLAFA			-0.02	0.00	32.70	0.00	1.03	0.00	0.030
Average Standard Error			0.01		0.66		0.20		0.000

Table B.27 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.06	0.00	14.04	0.00	4.99	0.99	0.016
		FSR LRT	0.14	0.00	8.74	0.00	5.02	0.98	0.012
		Full LRT	0.14	0.00	8.57	0.00	5.01	0.99	0.012
	$\mathbf{R}_2$	AIC	0.05	0.00	15.07	0.00	5	1	0.016
		FSR LRT	0.26	0.03	5.94	0.06	5.02	0.98	0.010
		Full LRT	0.21	0.03	6.93	0.05	5	1	0.010
AAC	**R**	AIC	0.06	0.00	13.91	0.00	4.99	0.99	0.017
		FSR LRT	0.27	0.04	5.16	0.08	5.04	0.96	0.010
		Full LRT	0.24	0.03	6.15	0.04	5.01	0.99	0.010
	$\mathbf{R}_2$	AIC	0.05	0.00	15.00	0.00	5	1	0.016
		FSR LRT	0.30	0.06	5.48	0.09	5.02	0.98	0.010
		Full LRT	0.23	0.05	6.65	0.07	5.01	0.99	0.010
MAC	**R**	AIC	0.05	0.00	15.56	0.00	5	1	0.016
		FSR LRT	0.31	0.04	5.28	0.05	5.01	0.99	0.009
		Full LRT	0.25	0.03	5.96	0.05	5	1	0.010
	$\mathbf{R}_2$	AIC	0.05	0.00	15.56	0.00	5	1	0.016
		FSR LRT	0.31	0.04	5.28	0.05	5.01	0.99	0.009
		Full LRT	0.25	0.03	5.96	0.05	5	1	0.010
SMC	**R**	AIC	0.05	0.00	14.46	0.00	5.04	0.96	0.016
		FSR LRT	0.09	0.00	11.29	0.00	5.04	0.96	0.014
		Full LRT	0.12	0.00	9.23	0.00	5.02	0.98	0.012
	$\mathbf{R}_2$	AIC	0.05	0.00	14.91	0.00	5	1	0.016
		FSR LRT	0.25	0.03	6.18	0.05	5.02	0.98	0.010
		Full LRT	0.21	0.03	7.00	0.05	5	1	0.010
LR	**R**	AIC	0.05	0.00	14.46	0.00	5.04	0.96	0.016
		FSR LRT	0.09	0.00	11.29	0.00	5.04	0.96	0.014
		Full LRT	0.12	0.00	9.23	0.00	5.02	0.98	0.012
	$\mathbf{R}_2$	AIC	0.05	0.00	14.46	0.00	5.04	0.96	0.016
		FSR LRT	0.09	0.00	11.29	0.00	5.04	0.96	0.014
		Full LRT	0.12	0.00	9.23	0.00	5.02	0.98	0.012
Traditional Communality			0.16	0.00	8.24	0.00	5.03	0.97	0.011
Traditional Magnitude			0.39	0.03	3.89	0.03	5.02	0.98	0.008
GLAFA			-0.01	0.00	27.21	0.00	2.77	0.00	0.052
Average Standard Error			0.01		0.29		0.01		0.000

Table B.28 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
AAC	**R**	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
MAC	**R**	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
SMC	**R**	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
LR	**R**	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	1	1	5	1	0.001
		FSR LRT	0.93	0.93	0.93	0.93	5	1	0.001
		Full LRT	0.93	0.93	0.93	0.93	5	1	0.001
Traditional Communality			1	1	1	1	5	1	0.001
Traditional Magnitude			1	1	1	1	5	1	0.001
GLAFA			0.57	0.57	0.57	0.57	5	1	0.002
Average Standard Error			0.02		0.02		0		0.000

Table B.29 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.06	0.00	14.47	0.00	5.92	0.40	0.012
		FSR LRT	0.43	0.06	3.39	0.09	5.44	0.66	0.006
		Full LRT	0.40	0.01	3.20	0.06	5.47	0.63	0.005
	$\mathbf{R}_2$	AIC	0.05	0.00	15.44	0.00	5.18	0.53	0.013
		FSR LRT	0.60	0.31	2.28	0.32	5.45	0.63	0.005
		Full LRT	0.49	0.07	2.70	0.07	5.43	0.66	0.005
AAC	**R**	AIC	0.06	0.00	14.08	0.00	5.45	0.53	0.012
		FSR LRT	0.57	0.21	2.32	0.22	5.48	0.61	0.005
		Full LRT	0.50	0.11	2.64	0.12	5.49	0.62	0.005
	$\mathbf{R}_2$	AIC	0.05	0.00	16.01	0.00	5.15	0.56	0.013
		FSR LRT	0.54	0.16	2.54	0.16	5.47	0.63	0.005
		Full LRT	0.47	0.05	2.76	0.05	5.42	0.67	0.005
MAC	**R**	AIC	0.05	0.00	16.09	0.00	5.08	0.57	0.013
		FSR LRT	0.65	0.36	1.84	0.36	5.43	0.66	0.005
		Full LRT	0.55	0.15	2.20	0.15	5.42	0.65	0.005
	$\mathbf{R}_2$	AIC	0.05	0.00	16.09	0.00	5.08	0.57	0.013
		FSR LRT	0.65	0.36	1.84	0.36	5.43	0.66	0.005
		Full LRT	0.55	0.15	2.20	0.15	5.42	0.65	0.005
SMC	**R**	AIC	0.02	0.00	18.78	0.00	4.48	0.23	0.015
		FSR LRT	0.39	0.05	3.89	0.08	5.41	0.61	0.006
		Full LRT	0.38	0.01	3.26	0.06	5.47	0.65	0.005
	$\mathbf{R}_2$	AIC	0.02	0.00	19.24	0.00	3.98	0.19	0.015
		FSR LRT	0.60	0.30	2.36	0.31	5.46	0.63	0.005
		Full LRT	0.50	0.07	2.69	0.07	5.43	0.66	0.005
LR	**R**	AIC	0.02	0.00	18.78	0.00	4.48	0.23	0.015
		FSR LRT	0.39	0.05	3.89	0.08	5.41	0.61	0.006
		Full LRT	0.38	0.01	3.26	0.06	5.47	0.65	0.005
	$\mathbf{R}_2$	AIC	0.02	0.00	18.78	0.00	4.48	0.23	0.015
		FSR LRT	0.39	0.05	3.89	0.08	5.41	0.61	0.006
		Full LRT	0.38	0.01	3.26	0.06	5.47	0.65	0.005
Traditional Communality			-0.02	0.00	31.14	0.00	4.57	0.44	0.024
Traditional Magnitude			0.33	0.03	4.94	0.03	5.28	0.77	0.006
GLAFA			-0.01	0.00	27.61	0.00	1.83	0.00	0.025
Average Standard Error			0.02		0.24		0.08		0.000

Table B.30 Results for $k = 5$, 9 correlated variables per factor, $u = 1$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.57	0.14	2.68	0.14	5	1	0.003
		FSR LRT	0.90	0.85	1.01	0.85	5	1	0.002
		Full LRT	0.80	0.58	1.29	0.58	5	1	0.002
	$\mathbf{R}_2$	AIC	0.64	0.28	2.36	0.28	5	1	0.003
		FSR LRT	0.91	0.86	1.00	0.86	5	1	0.002
		Full LRT	0.82	0.62	1.25	0.62	5	1	0.002
AAC	**R**	AIC	0.65	0.28	2.32	0.28	5	1	0.003
		FSR LRT	0.91	0.86	1.00	0.86	5	1	0.002
		Full LRT	0.83	0.66	1.21	0.66	5	1	0.002
	$\mathbf{R}_2$	AIC	0.63	0.26	2.44	0.26	5	1	0.003
		FSR LRT	0.90	0.85	1.01	0.85	5	1	0.002
		Full LRT	0.82	0.64	1.23	0.64	5	1	0.002
MAC	**R**	AIC	0.64	0.29	2.39	0.29	5	1	0.003
		FSR LRT	0.91	0.87	0.99	0.87	5	1	0.002
		Full LRT	0.83	0.66	1.21	0.66	5	1	0.002
	$\mathbf{R}_2$	AIC	0.64	0.29	2.39	0.29	5	1	0.003
		FSR LRT	0.91	0.87	0.99	0.87	5	1	0.002
		Full LRT	0.83	0.66	1.21	0.66	5	1	0.002
SMC	**R**	AIC	0.56	0.13	2.69	0.13	5	1	0.003
		FSR LRT	0.90	0.85	1.01	0.85	5	1	0.002
		Full LRT	0.80	0.58	1.29	0.58	5	1	0.002
	$\mathbf{R}_2$	AIC	0.64	0.26	2.36	0.26	5	1	0.003
		FSR LRT	0.91	0.86	1.00	0.86	5	1	0.002
		Full LRT	0.82	0.62	1.25	0.62	5	1	0.002
LR	**R**	AIC	0.56	0.13	2.69	0.13	5	1	0.003
		FSR LRT	0.90	0.85	1.01	0.85	5	1	0.002
		Full LRT	0.80	0.58	1.29	0.58	5	1	0.002
	$\mathbf{R}_2$	AIC	0.56	0.13	2.69	0.13	5	1	0.003
		FSR LRT	0.90	0.85	1.01	0.85	5	1	0.002
		Full LRT	0.80	0.58	1.29	0.58	5	1	0.002
Traditional Communality			0.07	0.00	12.81	0.00	5	1	0.009
Traditional Magnitude			0.82	0.56	1.59	0.56	5	1	0.002
GLAFA			0.40	0.40	0.40	0.40	5.01	0.99	0.002
Average Standard Error			0.03		0.07		0.00		0.000

Table B.31 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 100$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.99	0.88	9.85	0.88	5	1	0.004
		FSR LRT	0.85	0.38	8.33	0.38	5	1	0.004
		Full LRT	0.94	0.39	10.24	0.39	5	1	0.005
	$\mathbf{R}_2$	AIC	1.00	0.94	9.96	0.94	5	1	0.004
		FSR LRT	0.75	0.40	7.34	0.40	5	1	0.004
		Full LRT	0.88	0.40	9.47	0.40	5	1	0.005
AAC	**R**	AIC	0.99	0.94	9.91	0.94	5	1	0.004
		FSR LRT	0.74	0.42	7.19	0.42	5	1	0.005
		Full LRT	0.86	0.45	9.10	0.45	5	1	0.005
	$\mathbf{R}_2$	AIC	0.99	0.93	9.92	0.93	5	1	0.004
		FSR LRT	0.74	0.39	7.24	0.39	5	1	0.005
		Full LRT	0.87	0.39	9.34	0.39	5	1	0.005
MAC	**R**	AIC	1.00	0.95	9.97	0.95	5	1	0.004
		FSR LRT	0.75	0.40	7.28	0.40	5	1	0.004
		Full LRT	0.88	0.37	9.50	0.37	5	1	0.005
	$\mathbf{R}_2$	AIC	1.00	0.95	9.97	0.95	5	1	0.004
		FSR LRT	0.75	0.40	7.28	0.40	5	1	0.004
		Full LRT	0.88	0.37	9.50	0.37	5	1	0.005
SMC	**R**	AIC	0.98	0.81	9.81	0.81	5	1	0.004
		FSR LRT	0.92	0.40	9.11	0.41	5	1	0.004
		Full LRT	0.93	0.36	10.30	0.38	5	1	0.005
	$\mathbf{R}_2$	AIC	1.00	0.94	9.96	0.94	5	1	0.004
		FSR LRT	0.77	0.39	7.47	0.39	5	1	0.004
		Full LRT	0.89	0.40	9.61	0.40	5	1	0.005
LR	**R**	AIC	0.98	0.81	9.81	0.81	5	1	0.004
		FSR LRT	0.92	0.40	9.11	0.41	5	1	0.004
		Full LRT	0.93	0.36	10.30	0.38	5	1	0.005
	$\mathbf{R}_2$	AIC	0.98	0.81	9.81	0.81	5	1	0.004
		FSR LRT	0.92	0.40	9.11	0.41	5	1	0.004
		Full LRT	0.93	0.36	10.30	0.38	5	1	0.005
Traditional Communality			1.00	0.97	9.97	0.97	5	1	0.004
Traditional Magnitude			0.99	0.85	9.84	0.85	5	1	0.004
GLAFA			0.43	0.07	19.41	0.07	3.96	0.08	0.054
Average Standard Error			0.01		0.15		0.00		0.000

Table B.32 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 100$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	-0.08	0.00	43.24	0.00	4.33	0.16	0.025
		FSR LRT	0.01	0.00	33.11	0.01	7.12	0.12	0.022
		Full LRT	0.01	0.00	30.62	0.00	7.44	0.13	0.022
	$\mathbf{R}_2$	AIC	-0.10	0.00	43.77	0.00	3.39	0.16	0.025
		FSR LRT	0.07	0.00	30.47	0.01	4.65	0.23	0.022
		Full LRT	0.06	0.00	29.31	0.01	5.27	0.26	0.021
AAC	**R**	AIC	-0.09	0.00	43.67	0.00	2.92	0.09	0.025
		FSR LRT	0.10	0.00	31.82	0.01	3.77	0.19	0.022
		Full LRT	0.08	0.00	28.00	0.00	5.16	0.24	0.021
	$\mathbf{R}_2$	AIC	-0.10	0.00	43.32	0.00	3.08	0.16	0.024
		FSR LRT	0.09	0.00	32.15	0.01	4.06	0.27	0.022
		Full LRT	0.07	0.00	29.64	0.00	5.00	0.32	0.021
MAC	**R**	AIC	-0.09	0.00	44.78	0.00	3.49	0.12	0.025
		FSR LRT	0.06	0.00	27.41	0.00	4.69	0.29	0.021
		Full LRT	0.07	0.00	25.84	0.01	5.58	0.26	0.020
	$\mathbf{R}_2$	AIC	-0.09	0.00	44.78	0.00	3.49	0.12	0.025
		FSR LRT	0.06	0.00	27.41	0.00	4.69	0.29	0.021
		Full LRT	0.07	0.00	25.84	0.01	5.58	0.26	0.020
SMC	**R**	AIC	-0.07	0.00	40.85	0.00	6.90	0.10	0.024
		FSR LRT	-0.06	0.00	49.86	0.00	1.97	0.03	0.026
		Full LRT	-0.03	0.00	35.01	0.00	9.67	0.11	0.021
	$\mathbf{R}_2$	AIC	-0.09	0.00	43.37	0.00	3.30	0.07	0.024
		FSR LRT	0.07	0.00	32.89	0.03	3.98	0.21	0.022
		Full LRT	0.04	0.00	30.77	0.01	4.89	0.29	0.021
LR	**R**	AIC	-0.07	0.00	40.85	0.00	6.90	0.10	0.024
		FSR LRT	-0.06	0.00	49.86	0.00	1.97	0.03	0.026
		Full LRT	-0.03	0.00	35.01	0.00	9.67	0.11	0.021
	$\mathbf{R}_2$	AIC	-0.07	0.00	40.85	0.00	6.90	0.10	0.024
		FSR LRT	-0.06	0.00	49.86	0.00	1.97	0.03	0.026
		Full LRT	-0.03	0.00	35.01	0.00	9.67	0.11	0.021
Traditional Communality			0.06	0.00	28.77	0.00	5.25	0.41	0.021
Traditional Magnitude			0.19	0.00	21.99	0.00	5.27	0.48	0.019
GLAFA			-0.05	0.00	39.98	0.00	1.00	0.00	0.025
Average Standard Error			0.01		0.78		0.23		0.000

Table B.33 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 100$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.25	0.00	23.38	0.00	5.01	0.97	0.014
		FSR LRT	0.44	0.00	14.81	0.03	5.05	0.95	0.010
		Full LRT	0.43	0.00	17.01	0.02	5.04	0.96	0.010
	$\mathbf{R}_2$	AIC	0.23	0.00	24.07	0.00	5	1	0.013
		FSR LRT	0.58	0.02	11.04	0.10	5.04	0.96	0.008
		Full LRT	0.54	0.00	15.02	0.05	5.02	0.98	0.009
AAC	**R**	AIC	0.26	0.00	23.24	0.00	5.01	0.99	0.014
		FSR LRT	0.57	0.00	9.82	0.09	5.05	0.95	0.008
		Full LRT	0.56	0.00	14.54	0.08	5.01	0.99	0.009
	$\mathbf{R}_2$	AIC	0.23	0.00	24.09	0.00	5	1	0.014
		FSR LRT	0.60	0.01	10.51	0.09	5.03	0.97	0.008
		Full LRT	0.56	0.00	15.27	0.04	5.01	0.99	0.009
MAC	**R**	AIC	0.22	0.00	24.54	0.00	5	1	0.014
		FSR LRT	0.52	0.01	9.85	0.04	5.05	0.95	0.008
		Full LRT	0.52	0.00	14.22	0.05	5.01	0.99	0.009
	$\mathbf{R}_2$	AIC	0.22	0.00	24.54	0.00	5	1	0.014
		FSR LRT	0.52	0.01	9.85	0.04	5.05	0.95	0.008
		Full LRT	0.52	0.00	14.22	0.05	5.01	0.99	0.009
SMC	**R**	AIC	0.24	0.00	23.32	0.00	5.06	0.94	0.014
		FSR LRT	0.28	0.00	21.87	0.00	5.07	0.93	0.013
		Full LRT	0.35	0.00	18.10	0.00	5.06	0.94	0.011
	$\mathbf{R}_2$	AIC	0.24	0.00	23.97	0.00	5	1	0.013
		FSR LRT	0.56	0.02	11.33	0.09	5.04	0.96	0.008
		Full LRT	0.55	0.00	15.29	0.05	5.02	0.98	0.009
LR	**R**	AIC	0.24	0.00	23.32	0.00	5.06	0.94	0.014
		FSR LRT	0.28	0.00	21.87	0.00	5.07	0.93	0.013
		Full LRT	0.35	0.00	18.10	0.00	5.06	0.94	0.011
	$\mathbf{R}_2$	AIC	0.24	0.00	23.32	0.00	5.06	0.94	0.014
		FSR LRT	0.28	0.00	21.87	0.00	5.07	0.93	0.013
		Full LRT	0.35	0.00	18.10	0.00	5.06	0.94	0.011
Traditional Communality			0.52	0.00	16.93	0.01	5.04	0.96	0.009
Traditional Magnitude			0.75	0.03	12.58	0.07	5.05	0.95	0.007
GLAFA			0.01	0.00	32.91	0.00	2.52	0.00	0.045
Average Standard Error			0.01		0.39		0.02		0.000

Table B.34 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 400$, high magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.95	0.56	9.44	0.56	5	1	0.001
		Full LRT	0.99	0.96	9.92	0.96	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.88	0.55	8.65	0.55	5	1	0.001
		Full LRT	0.96	0.93	9.56	0.93	5	1	0.001
AAC	**R**	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.85	0.57	8.35	0.57	5	1	0.001
		Full LRT	0.97	0.94	9.68	0.94	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.86	0.56	8.50	0.56	5	1	0.001
		Full LRT	0.97	0.94	9.69	0.94	5	1	0.001
MAC	**R**	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.87	0.58	8.60	0.58	5	1	0.001
		Full LRT	0.99	0.96	9.88	0.96	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.87	0.58	8.60	0.58	5	1	0.001
		Full LRT	0.99	0.96	9.88	0.96	5	1	0.001
SMC	**R**	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.95	0.56	9.47	0.56	5	1	0.001
		Full LRT	0.99	0.96	9.93	0.96	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.89	0.55	8.79	0.55	5	1	0.001
		Full LRT	0.99	0.96	9.84	0.96	5	1	0.001
LR	**R**	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.95	0.56	9.47	0.56	5	1	0.001
		Full LRT	0.99	0.96	9.93	0.96	5	1	0.001
	$\mathbf{R}_2$	AIC	1	1	10	1	5	1	0.001
		FSR LRT	0.95	0.56	9.47	0.56	5	1	0.001
		Full LRT	0.99	0.96	9.93	0.96	5	1	0.001
Traditional Communality			1	1	10	1	5	1	0.001
Traditional Magnitude			1	1	10	1	5	1	0.001
GLAFA			0.80	0.15	7.75	0.15	5	1	0.002
Average Standard Error			0.01		0.08		0		0.000

Table B.35 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 400$, low magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

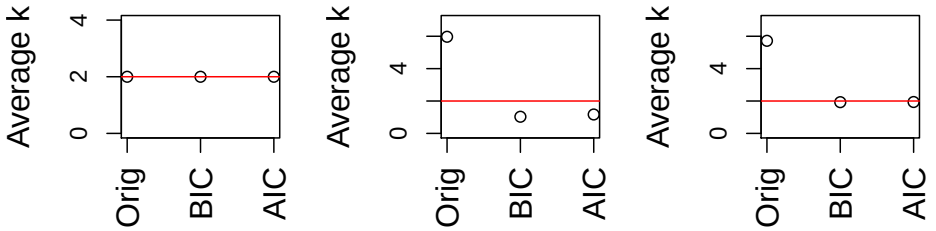
Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.26	0.00	23.16	0.00	5.87	0.42	0.010
		FSR LRT	0.74	0.05	11.59	0.20	5.73	0.49	0.005
		Full LRT	0.63	0.00	13.84	0.03	5.85	0.48	0.006
	$\mathbf{R}_2$	AIC	0.23	0.00	24.33	0.00	5.17	0.55	0.011
		FSR LRT	0.85	0.19	10.06	0.23	5.61	0.56	0.004
		Full LRT	0.74	0.02	13.27	0.03	5.49	0.64	0.005
AAC	**R**	AIC	0.28	0.00	22.89	0.00	5.39	0.57	0.010
		FSR LRT	0.84	0.13	9.88	0.23	5.63	0.52	0.004
		Full LRT	0.72	0.01	13.16	0.03	5.61	0.55	0.005
	$\mathbf{R}_2$	AIC	0.22	0.00	24.77	0.00	5.26	0.54	0.011
		FSR LRT	0.85	0.15	10.39	0.20	5.62	0.55	0.004
		Full LRT	0.72	0.02	13.60	0.03	5.51	0.63	0.005
MAC	**R**	AIC	0.22	0.00	25.09	0.00	5.08	0.57	0.011
		FSR LRT	0.78	0.14	8.97	0.16	5.75	0.48	0.004
		Full LRT	0.73	0.02	12.67	0.05	5.46	0.62	0.005
	$\mathbf{R}_2$	AIC	0.22	0.00	25.09	0.00	5.08	0.57	0.011
		FSR LRT	0.78	0.14	8.97	0.16	5.75	0.48	0.004
		Full LRT	0.73	0.02	12.67	0.05	5.46	0.62	0.005
SMC	**R**	AIC	0.13	0.00	27.70	0.00	4.52	0.22	0.012
		FSR LRT	0.67	0.03	13.56	0.12	5.67	0.45	0.006
		Full LRT	0.62	0.00	14.27	0.02	5.74	0.48	0.006
	$\mathbf{R}_2$	AIC	0.12	0.00	28.24	0.00	3.98	0.19	0.012
		FSR LRT	0.84	0.18	10.05	0.20	5.65	0.55	0.004
		Full LRT	0.73	0.02	13.37	0.03	5.50	0.61	0.005
LR	**R**	AIC	0.13	0.00	27.70	0.00	4.52	0.22	0.012
		FSR LRT	0.67	0.03	13.56	0.12	5.67	0.45	0.006
		Full LRT	0.62	0.00	14.27	0.02	5.74	0.48	0.006
	$\mathbf{R}_2$	AIC	0.13	0.00	27.70	0.00	4.52	0.22	0.012
		FSR LRT	0.67	0.03	13.56	0.12	5.67	0.45	0.006
		Full LRT	0.62	0.00	14.27	0.02	5.74	0.48	0.006
Traditional Communality			-0.06	0.00	38.16	0.00	4.84	0.49	0.019
Traditional Magnitude			0.69	0.01	13.97	0.01	5.33	0.77	0.005
GLAFA			0.03	0.00	33.20	0.00	1.87	0.00	0.020
Average Standard Error			0.01		0.30		0.09		0.000

Table B.36 Results for $k = 5$, 9 correlated variables per factor, $u = 10$, $n = 400$, mixed magnitude of loadings. Av = average of $N = 100$ replications, PC = proportion finding correct value.

Selection Method			ARI		u		k		MSE
Ordering	**R**	Stop Crit	Av	PC	Av	PC	Av	PC	Av
ME	**R**	AIC	0.87	0.16	11.61	0.16	5	1	0.002
		FSR LRT	0.95	0.50	9.54	0.51	5.03	0.97	0.002
		Full LRT	0.88	0.12	11.29	0.12	5	1	0.002
	$\mathbf{R}_2$	AIC	0.89	0.27	11.37	0.27	5	1	0.002
		FSR LRT	0.89	0.49	8.94	0.49	5.04	0.96	0.002
		Full LRT	0.88	0.18	11.08	0.18	5	1	0.002
AAC	**R**	AIC	0.90	0.29	11.27	0.29	5	1	0.002
		FSR LRT	0.90	0.51	9.02	0.51	5.03	0.97	0.002
		Full LRT	0.89	0.17	10.99	0.17	5	1	0.002
	$\mathbf{R}_2$	AIC	0.88	0.25	11.49	0.25	5	1	0.002
		FSR LRT	0.91	0.50	9.14	0.50	5.04	0.96	0.002
		Full LRT	0.88	0.17	11.07	0.17	5	1	0.002
MAC	**R**	AIC	0.89	0.29	11.39	0.29	5	1	0.002
		FSR LRT	0.87	0.50	8.66	0.50	5.03	0.97	0.002
		Full LRT	0.89	0.19	11.09	0.19	5	1	0.002
	$\mathbf{R}_2$	AIC	0.89	0.29	11.39	0.29	5	1	0.002
		FSR LRT	0.87	0.50	8.66	0.50	5.03	0.97	0.002
		Full LRT	0.89	0.19	11.09	0.19	5	1	0.002
SMC	**R**	AIC	0.87	0.13	11.69	0.13	5	1	0.002
		FSR LRT	0.95	0.50	9.55	0.52	5.03	0.97	0.002
		Full LRT	0.88	0.10	11.35	0.10	5	1	0.002
	$\mathbf{R}_2$	AIC	0.89	0.26	11.36	0.26	5	1	0.002
		FSR LRT	0.90	0.48	8.99	0.48	5.04	0.96	0.002
		Full LRT	0.88	0.18	11.08	0.18	5	1	0.002
LR	**R**	AIC	0.87	0.13	11.69	0.13	5	1	0.002
		FSR LRT	0.95	0.50	9.55	0.52	5.03	0.97	0.002
		Full LRT	0.88	0.10	11.35	0.10	5	1	0.002
	$\mathbf{R}_2$	AIC	0.87	0.13	11.69	0.13	5	1	0.002
		FSR LRT	0.95	0.50	9.55	0.52	5.03	0.97	0.002
		Full LRT	0.88	0.10	11.35	0.10	5	1	0.002
Traditional Communality			0.31	0.00	21.69	0.00	5	1	0.008
Traditional Magnitude			0.95	0.55	10.57	0.57	5	1	0.002
GLAFA			0.28	0.09	24.32	0.09	4.49	0.54	0.026
Average Standard Error			0.01		0.14		0.01		0.000

B.2 GLAFA Original Results

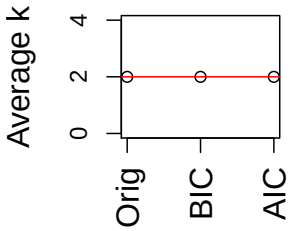
This section contains plots showing the different average estimates of k for the original GLAFA method, as compared to the BIC- and AIC-revised GLAFA methods.



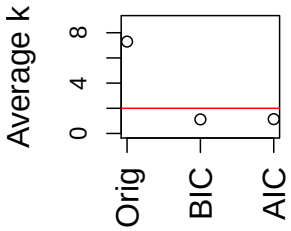
(a) $n = 100$, high **L**

(b) $n = 100$, low **L**

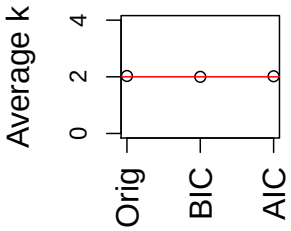
(c) $n = 100$, mixed **L**



(d) $n = 400$, high **L**

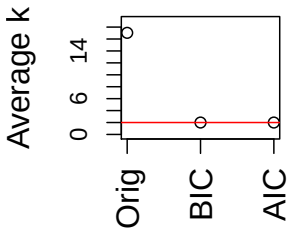


(e) $n = 400$, low **L**

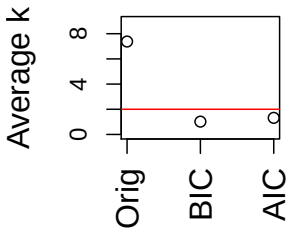


(f) $n = 400$, mixed **L**

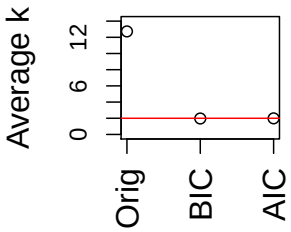
Figure B.37 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 2$ factors, 9 variables per factor, $u = 1$, $N = 100$ replications.



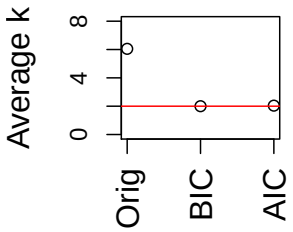
(a) $n = 100$, high **L**



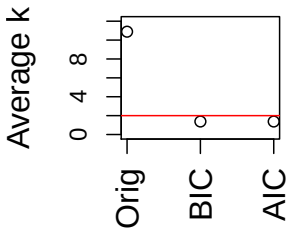
(b) $n = 100$, low **L**



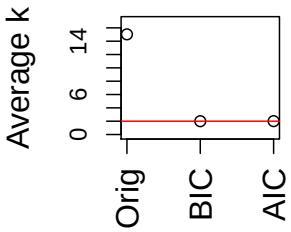
(c) $n = 100$, mixed **L**



(d) $n = 400$, high **L**

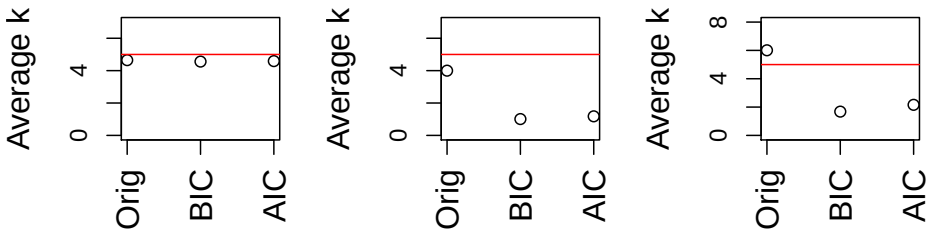


(e) $n = 400$, low **L**

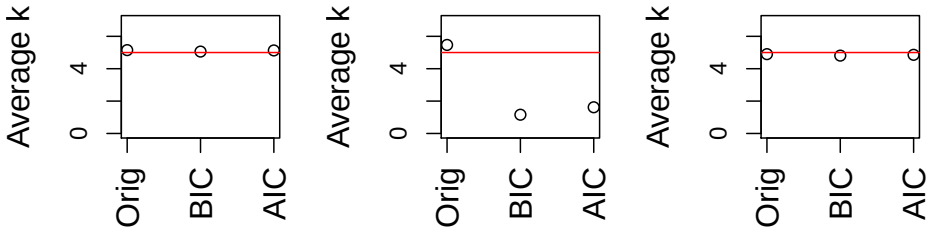


(f) $n = 400$, mixed **L**

Figure B.38 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 2$ factors, 9 variables per factor, $u = 10$, $N = 100$ replications.

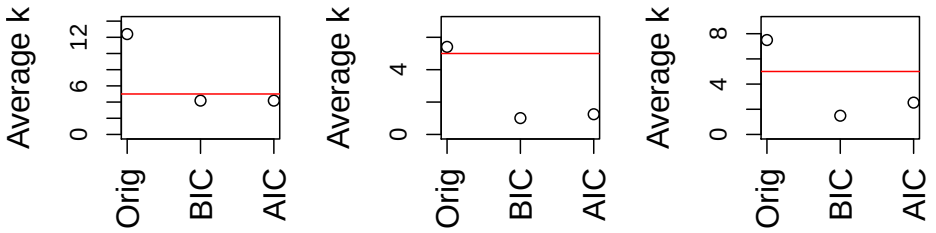


(a) $n = 100$, high **L** **(b)** $n = 100$, low **L** **(c)** $n = 100$, mixed **L**

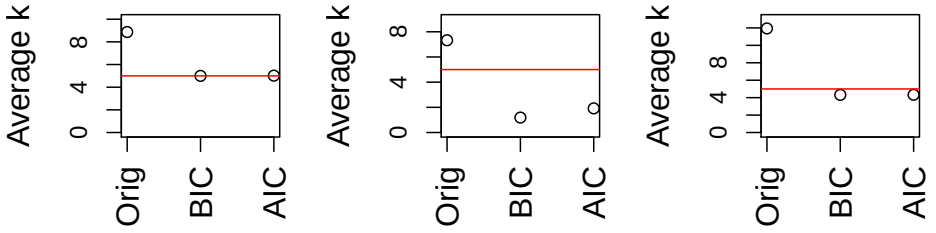


(d) $n = 400$, high **L** **(e)** $n = 400$, low **L** **(f)** $n = 400$, mixed **L**

Figure B.39 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 5$ factors, 3 variables per factor, $u = 1$, $N = 100$ replications.

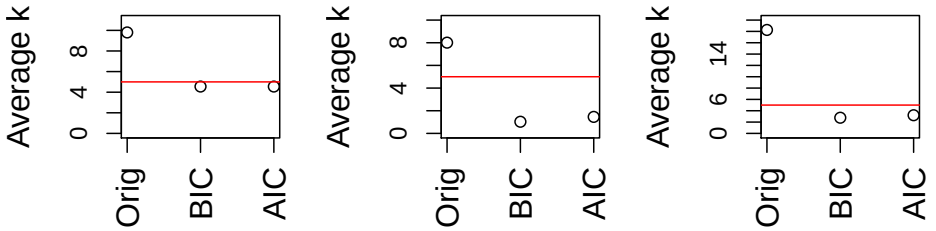


(a) $n = 100$, high **L** **(b)** $n = 100$, low **L** **(c)** $n = 100$, mixed **L**

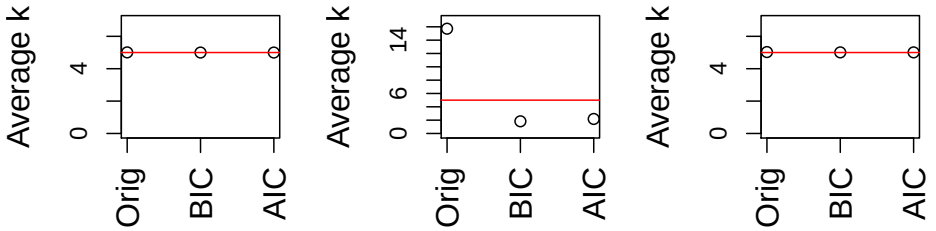


(d) $n = 400$, high **L** **(e)** $n = 400$, low **L** **(f)** $n = 400$, mixed **L**

Figure B.40 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 5$ factors, 3 variables per factor, $u = 10$, $N = 100$ replications.

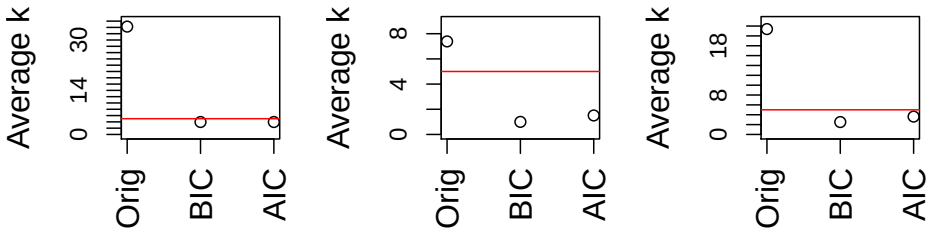


(a) $n = 100$, high **L** **(b)** $n = 100$, low **L** **(c)** $n = 100$, mixed **L**

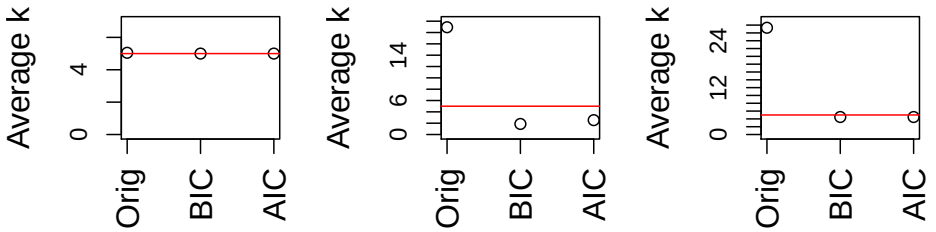


(d) $n = 400$, high **L** **(e)** $n = 400$, low **L** **(f)** $n = 400$, mixed **L**

Figure B.41 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 5$ factors, 9 variables per factor, $u = 1$, $N = 100$ replications.



(a) $n = 100$, high **L** **(b)** $n = 100$, low **L** **(c)** $n = 100$, mixed **L**



(d) $n = 400$, high **L** **(e)** $n = 400$, low **L** **(f)** $n = 400$, mixed **L**

Figure B.42 Average number of factors chosen for original GLAFA, BIC-revised GLAFA, and AIC-revised GLAFA. $k = 5$ factors, 9 variables per factor, $u = 10$, $N = 100$ replications.