Title
Performance of Collinearity-Resistant Fitting Methods in Exploratory Factor Analysis

Author
Xie, Xiaoxi

Publication Date
2013

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Performance of Collinearity-Resistant Fitting Methods in Exploratory Factor Analysis

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Xiaoxi Xie

2013

Abstract of the Thesis

Performance of Collinearity-Resistant Fitting Methods in Exploratory Factor Analysis

by

Xiaoxi Xie

Master of Science in Statistics

University of California, Los Angeles, 2013

Professor Peter M. Bentler, Chair

The factor analytic model is usually an approximation that may not represent well the latent structure of a set of variables. This paper studies the distortion to common factor analysis due to doublets, pairwise associations in variables over and above those postulated by the factor model. Three methodologies to estimate the parameters of the factor model while minimizing the influence of doublets on the solution, i.e., methodologies that are resistant to the effects of doublets and near doublets, proposed by Yates(1987), Mulaik(2010) and Bentler(2012), are reviewed and compared. Examples and a simulation verify that these methodologies achieve their goal.

The thesis of Xiaoxi Xie is approved.

Hongquan Xu

Frederic R. Paik Schoenberg

Peter M. Bentler, Committee Chair

University of California, Los Angeles

2013

Table of Contents

List of Figures

List of Tables

Acknowledgments

This master's thesis would not have been possible without the support of many people. The author wishes to express her gratitude to her advisor, Prof. Peter M. Bentler who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest gratitude are also due to the members of the committee, Prof. Frederic R. Paik Schoenberg and Prof. Hongquan Xu without whose knowledge and assistance this study would not have been successful.

Special thanks also to all her graduate friends; Xiaochen Lian, Xuening Liu, Bo Chang, Xiaoxiao Tong, Chen Chen, Dong Wang, Yuchen Liu, Jie Fu and Shan He for emotional support, sharing the literature and invaluable assistance. Not forgetting to her best friends who always been there. The author would also like to convey thanks to the staff of the Department of Statistics at the University of California, Los Angeles, especially Glenda Jones and Jason Mesa, for their sincere support during the two years of study. The author wishes to express her love and gratitude to her beloved families; for their understanding and endless love, through the duration of her studies.

CHAPTER 1

Introduction

It has often agreed that strong mathematical models of real-world phenomena, including those of interest in psychometrics, are almost never precisely true (e.g., Bentler & Bonett, 1980 [BB80]; de Leeuw, 1988 [De 88]; Browne & Cudeck, 1993 [BCB93]; MacCallum, 2003 [Mac03]). This viewpoint is also popular in the field of exploratory factor analysis (EFA). MacCallum, Browne, and Cai (2007) [MBC07], for example, state that "...a factor analysis model is not an exact representation of real-world phenomena. Such a model, at least in parsimonious form, is always wrong to some degree, even in the population." This raises the question of why the standard estimation methods in factor analysis, such as least squares (LS) and maximum likelihood (ML), are not modified to deal with what often is, a priori, an inadequate model. Perhaps such default estimation methods are used because no general alternative methodology has been accepted. This paper reviews two methods included in Yates (1987) [Yat87], Mulaik (2010) [Mul10], and an newly proposed exploratory factor analysis methodology by Bentler (2012) [Ben12] to minimize the effects of a substantial part of misspecification that is due to use of an inadequate factor model, that is, a model that allows doublets into the common factor space.

One of the most well-known distortions in the factor model is due to associations between pairs of variables beyond that due to unidentified common factors in the factor model, namely, doublets. A doublet can be thought of as a residual

correlation or covariance or as a factor with nonzero loadings only for a pair of variables. The most thorough attempt to deal with this problem was made by Yates (1987) [Yat87]. Although his book was substantially devoted to the methodology of factor rotation, Yates (pp. 224-231) proposed a method of collinearity-resistant fitting inspired by Tukeys (1977) [Tuk77] approach to exploratory data analysis. He wanted to assure that the factors extracted would not be doublets or minor group factors which he felt would not be invariant to changes in the selection of variables as factor indicators. A different approach to this problem was proposed by Mulaik (2010, pp. 258-262) [Mul10] in his doublet factor analysis method. As far as we know, neither of these methods have been independently evaluated. They will be described once the problem posed by doublets is clearly illustrated and the notation of this paper is introduced. A new approach (Bentler, 2012) to the doublet problem is then developed and subsequently evaluated with simulations.

CHAPTER 2

Doublets and Common Factors

2.1 Exploratory Factor Analysis

In multivariate statistics, exploratory factor analysis (EFA) is a statistical method used to uncover the underlying structure of a relatively large set of variables. EFA is a technique within factor analysis whose overarching goal is to identify the underlying relationships between measured variables (Norris, 2010) [NL10]. It is commonly used by researchers when developing a scale (a scale is a collection of questions used to measure a particular research topic) and serves to identify a set of latent constructs underlying a battery of measured variables. It should be used when the researcher has no a priori hypothesis about factors or patterns of measured variables. Measured variables are any one of several attributes of people that may be observed and measured. Researchers must carefully consider the number of measured variables to include in the analysis. EFA procedures are more accurate when each factor is represented by multiple measured variables in the analysis. There should be at least 3 to 5 measured variables per factor (MacCallum & Robert C, 1990) [Mac90].

EFA is based on the common factor model, which seeks the least number of factors which can account for the common variance (correlation) of a set of variables. Common factors theory proposes that different theoretical and evidence-based approaches to psychotherapy and counseling have common components and that

those components account for outcome more than components that are unique to each approach (Imel & Wampold, 2008) [IW08]. Within the common factor model, measured variables are expressed as a function of common factors, unique factors, and errors of measurement. Common factors influence two or more measured variables, while each unique factor influences only one measured variable and does not explain correlations among measured variables. The Common factor model can be expressed by the following notation:

$$\Sigma = \Lambda\Lambda^{'} + \Psi^2 \tag{2.1}$$

where Σ is a $p \times p$ correlation or covariance matrix, Λ is a $p \times k$ matrix of factor loadings with rows $\lambda_{i.}^{'}$, and Ψ^2 is a $p \times p$ diagonal matrix of unique variances.

2.2 Fitting Procedures

Fitting procedures are used to estimate the factor loadings and unique variances of the model (Factor loadings are the regression coefficients between items and factors and measure the influence of a common factor on a measured variable). There are several factor analysis fitting methods to choose from. Principal axis factoring (PAF) and maximum likelihood (ML) are two extraction methods that are generally recommended. In general, ML or PAF give the best results, depending on whether data are normally-distributed or if the assumption of normality has been violated (Fabrigar, 1999) [FWM99].

The maximum likelihood method is factor extraction method that produces parameter estimates that are most likely to have produced the observed correlation matrix if the sample is from a multivariate normal distribution. The correlations are weighted by the inverse of the uniqueness of the variables, and an iterative algorithm is employed. Let S be a sample correlation or covariance matrix. The

maximum likelihood solutions of Λ and Ψ^2 are obtained by minimizing

$$F = tr[(\Lambda\Lambda^{'} + \Psi^2)^{-1}S] - log|(\Lambda\Lambda^{'} + \Psi^2)^{-1}S| - p \tag{2.2}$$

It has many advantages in that it allows researchers to compute of a wide range of indexes of the goodness of fit of the model, it allows researchers to test the statistical significance of factor loadings, calculate correlations among factors and compute confidence intervals for these parameters. ML is the best choice when data are normally distributed and should not be used if the data are not normally distributed. In this paper, we only study the correlation structure under conditions of multivariate normality.

2.3 Problem of Doublet Factors

Factors are unobserved latent variables that can be inferred from a set of observed variables. Therefore, factors cannot emerge unless there is a sufficient number of observed variables that vary along the latent continuum. You cannot define a factor with a single observed variable. You should have a minimum of three observed variables for each factor expected to emerge. In Thurstone's terminology, the factors defined by only one or two observed variables are called "singlet" or "doublet" factors, which are not desirable.

The doublet factors reflect local dependencies between pairs of individual variables. It is possible for these pairwise linkages between manifest variables to be extremely strong because they are brought about by circumstances prevailing right at the time of data collection; i.e., experimental dependencies produced by recording several responses to the same stimulus, by obtaining alternative indicators of

proficiency at the same task (e.g. latency to respond and accuracy of response) or at extremely similar or overlapping tasks (e.g., adding two digit numbers and adding three digit numbers) (Yates, 1987).

Disturbances in the form of local dependencies among manifest variables are a serious threat to the ultimate goals of common factor analysis because they are highly dependent on test battery composition (i.e., they are not invariant in the least) and tend to reflect superficial methodological or variable sampling idiosyncrasies rather than the effects of latent determinants that are of profound theoretical significance. In particular, they do not reflect the action of latent determinants that are broad and general enough in their action to participate in giving shape to the highly populated polyhedral convex cone of test vectors required for effective application of bounding hyperplane simple structure transformation (Yates, 1987).

The presence of the minor factors in an analysis jeopardizes the overall stability of the solution. A minor factor may relate to only a few manifest variables, but its inclusion in a final solution involves the estimation of just as many parameters (factor pattern coefficients) as does a major factor. The fact that most of these extra parameters are near-zero and hence contribute strongly to Thurstone's original criterion for a best-fitting simple structure means that the inclusion of one or more minor factors in transformation can weaken the precision with which major factors are located (Yates, 1987). Since pairwise linkages among variables do occur in practice, they must be dealt with in some way in order to keep them from contaminating the results of any practical application of factor analytic methods.

Suppose that a 5×5 population correlation matrix with correlations among all variables at .49 is presented for analysis. Asked to extract one factor, all stan-

dard EFA methods will recover the correct result, a loading matrix with all .7 entries. Now suppose that only the 1,2 correlation is increased from .49 to .75, representing the influence of a doublet. Now a 1-factor EFA method yields the incorrect loading matrix, with ML yielding $[.844, .844, .622, .622, .622]'$. The factor has been tilted toward the doublet variables. Of course, one factor could be considered to be an incorrect number of factors, since the 1,2 correlation is not fully accounted for by the factor model. Actually, with a 2-factor solution can perfectly reproduce these correlations, but in the case of ML, with problems including: 1 degrees of freedom (df), arbitrary unique ("error") variances for the 1^{st} two variables, three linear dependencies, and two out of range standard error estimates. It seems that the 1-factor model underfactors, while the 2-factor model overfactors or has too many parameters. Our aim in what follows is to present and evaluate methods for the routine detection and appropriate treatment of minor factors, including doublets, that constitute violations of the invariant common factor model.

CHAPTER 3

Yates' Collinearity-Resistant Fitting Procedure

3.1 Model

Yates' (1987)[Yat87] work is remembered today for his geomin, a widely respected method of oblique factor rotation. He intended this method to help distinguish invariant factors from minor factors including doublets during the rotation phase. As noted, however, he also proposed a method to extract factors so as to minimize the effect of minor factors using a weighted least squares methodology.

Let the exploratory factor analysis model be (2.1). Yates (1987, p. 229) proposed to minimize

$$1^{'}([S - \Lambda\Lambda^{'}]^{(2)} * V)1 \tag{3.1}$$

where 1 is a unit vector, $[.]^{(2)}$ is the matrix of squared elements in [.], "$*$" is the element-wise Hadamard matrix product, and V is a symmetric matrix with elements

$$v_{ij} = 1 - \frac{(\lambda_{i.}^{'}\lambda_{j.})^2}{(\lambda_{i.}^{'}\lambda_{i.})(\lambda_{j.}^{'}\lambda_{j.})} \tag{3.2}$$

The entry within brackets in the numerator of the last term of the foregoing expression is, of course, the k-factor reproduced correlation between variables i and j. When this term is squared and put in ratio to the product of the associated commonalities it yields an index of how collinear the variables in question appear

to be within common factor space, irrespective of the sign of the reproduced correlation and adjusted for variation in the respective lengths of the vectors involved. Variables that are highly collinear in the factor space receive a low weight (3.2) and reduce the influence of these variables during the function minimization. Since the latter index of collinearity ranges from zero to unity, the pairwise distinguishability weight, v_{ij}(3.2), ranges from unity, for a pair of variables between which the common factors account for absolutely no association, to a value of zero, for a pair of variables which relate to the common factors in exactly the same way (i.e., which have proportional loading patterns).

Those entries receiving little weight will correspond to pairs of manifest variables relating in parallel ways to the broad general factors common to the battery as a whole; i.e., to pairs of variables that are apt to show excessive collinearity at the manifest level if they are linked by any minor factor effects. Such variables appear to have been influenced in highly parallel ways by latent determinants that underlie the entire domain sampled, so it is likely that the action of more limited influences would only contribute further to their collinearity at the manifest level. It is in an effort to avoid confounding invariant major common factor effects with these more superficial local disturbances that accurate fitting of the manifest association found between any pair of variables showing latent collinearity is avoided by incorporating pairwise distinguishability weights into criterion (3.1).

Notice that diagonal cells in the residual covariance matrix need not be explicitly skipped in the summations going into (3.1) since each manifest variable is always collinear with itself at the latent level as well as at the manifest level. What this suggests, of course, is that the classical common factor model in which only off-diagonals in the observed correlation matrix are fit is a special case of this collinearity-resistant model, in which it is assumed that excessive collinearity at

the manifest level occurs only when a variable is correlated with itself (i.e. when a specific factor comes into play along with the postulated common factors). Yates have simply generalized the classical model to allow for "correlated specifics" in the form of doublet or minor group factors, the effects of which must not be allowed to distort the solution for factors common to the battery as a whole. The latter must remain invariant with respect to changes in test battery composition, so they cannot account for the action of minor influences specific to any given individual variable or group of highly collinear variables.

It is interesting to consider what Yates' EFA method might do with the doublet example given in the previous chapter. Suppose the model estimated is a 1-factor model and the current estimated factor loadings are arbitrary values $[a, b, c, d, e]'$. A bit of calculation shows that all the weights (3.2) are zero, and hence the function (3.1) attains its minimum of zero at any set of parameter values. Clearly this approach is useless in the 1-factor case.

3.2 Algorithm

In order to encourages tolerance for outliers in the form of strong linkages between isolated pairs or groups of highly similar manifest variables, Yates proposed a simple and practical computing formula inspired by Gauss-Seidel and multidimensional sectioning methods, in which the current approximation to the orthogonal factor matrix ${}_t\Lambda$, after any given move t, is used to obtain the index of distinguishability ${}_tv_{ij}$, with which (3.1) is optimized for an updated ${}_{t+1}\Lambda$ and the procedure is repeated iteratively. In other words, rewrite (3.2) as:

$$ {}_tv_{ij} = 1 - \frac{\left(\sum_{m=1}^{k} ({}_t\lambda_{im})({}_t\lambda_{jm})\right)^2}{\sum_{m=1}^{k} ({}_t\lambda_{im}^2) \sum_{m=1}^{k} ({}_t\lambda_{jm}^2)} \tag{3.3} $$

where ${}_t\lambda_{im}$ is the factor loading of variable i on factor m in current estimate ${}_t\Lambda$. And then keep (3.3) up to date at all times, so the weighted least squares fitting of the common factor model can be accomplished via successive replacement of individual elements of ${}_t\Lambda$ in such a way that residual linkages corresponding to highly collinear (indistinguishable) variables within common factor space are accommodated; i.e. minimize

$$\sum_{i=1}^{p}\sum_{j=1}^{p}({}_tv_{ij})({}_tr_{ij}^2) \tag{3.4}$$

where ${}_tr_{ij}$ is the general element of the residual covariance matrix

$${}_tR = S -_t \Lambda_t\Lambda' \tag{3.5}$$

and

$${}_tr_{ij} = s_{ij} - \sum_{m=1}^{k}({}_t\lambda_{im})({}_t\lambda_{jm}) \tag{3.6}$$

where s_{ij} is the general element of the sample correlation or covariance matrix S. Through several steps of deduction, Yates proposed that an improved estimate of the (i,k)-th element of ${}_t\Lambda$ at any stage of iteration can be expressed as

$${}_{t+1}\lambda_{im} =_t \lambda_{im} + \frac{\sum_{j=1}^{p}({}_tv_{ij})({}_t\lambda_{jm})({}_tr_{ij})}{\sum_{j=1}^{p}({}_tv_{ij})({}_t\lambda_{jm}^2)} \tag{3.7}$$

If the pairwise distinguishability weights for the i-th variable are updated along with the usual updating of residuals within that row and column of ${}_tR$, then an adaptive weighting system is put into effect by (3.7) that systematically relaxes the pressure within weighted least square criterion (3.4) to fit certain entries in the original correlation matrix. Yates suggested to use the result from principal components analysis as the starting point of iteration, and repeat the successive displacement of all k elements in each row of ${}_t\Lambda$ until the change in communality with a given repetition is some fraction (e.g. one tenth) of the largest communality

adjustment produced by the previous major iteration cycle in order to accelerate the iteration– but not to exceed k repetitions in any given row per major cycle (Yates 1987, p.224). But he didn't indicate the termination condition of the algorithm. He also suggested changing all $p \times k$ individual directions of search simply through occasional arbitrary orthogonal transformations of the coordinate system (Yates 1987, p.223).

In practice, we could follow the steps below to achieve collinearity-resistant estimates: Given S, compute ${}_0\Lambda$ as the result of principal components analysis, and then compute ${}_0R$, ${}_0V$ according to (3.5) and (3.3). For the purpose of identification, we could implement orthogonal transformation, and fix the sign of certain element (the element that has the largest absolute value in that column) in each column of Λ. At time t,

- Step 1: For i=1,...,p, m=1,...,k, compute ${}_t\lambda_{im}$, and update elements in ${}_tR$ and ${}_tV$ accordingly in each iteration. And we get ${}_t\Lambda$ after iterations through all $p \times k$ elements.

- Step 2: Orthogonal transformation: a) Let ${}_t\Lambda = UD$, where U is the left-singular vectors and D is the singular values from singular value decomposition; b) Check the sign of the chosen element in each column; if the sign is opposite to what we fixed at the beginning, multiply that column by -1

- Step 3: Compute the absolute changes in communalities from ${}_{t-1}\Lambda$ to ${}_t\Lambda$, i.e. compute $|diag({}_t\Lambda_t\Lambda' - {}_{t-1}\Lambda_{t-1}\Lambda')|$

- Step 4: Update ${}_tR$ and ${}_tV$

Let ${}_th_i^2$ be the communality of variable i at time t. Then if $|{}_th_i^2 - {}_{t-1}h_i^2| > 0.0001$ for any i, i=1,...,p, go to step 1, else quit. The estimate is the last ${}_t\Lambda$. We replicate the covariance residual matrix given in Yates (1987, pp.240-243) by implementing

the above algorithm on the correlation matrix provided by Lord (1956, pp. 36-37)[Lor56].

CHAPTER 4

Mulaik's Model That Includes Doublets Explicitly

4.1 Model

Mulaik (2010)[Mul10] proposed to model doublet factors explicitly using a strong assumption derived from Guttmans (1953)[Gut53] image analysis and the prior work of Butler (1968)[But68]. He proposed the following model equation

$$Y = \Lambda X + K\Delta + \Psi E \tag{4.1}$$

where

Y is a $p \times 1$ vector of observed random variables

Λ is a $p \times k$ common-factor-pattern matrix

X is a $k \times 1$ vector of common factors

K is a $p \times q$ doublet pattern matrix

Δ is a $q \times 1$ vector of doublet factors, $q = p(p-1)/2$

Ψ is a $p \times p$ diagonal matrix of unique-factor-pattern coefficients

E is a $p \times 1$ vector of unique- or singlet-factor variables

He assume the follow relationships between the random variables of the model:

$$\begin{aligned} E(X\Delta^{'}) &= 0 \\ E(\Delta E^{'}) &= 0 \\ E(XE^{'}) &= 0 \\ E(XX^{'}) &= I \end{aligned} \tag{4.2}$$

He further assume that all variables have zero means and unit variances. Then the matrix of covariances or correlations among the observed variables is given by

$$\Sigma = \Lambda\Lambda^{'} + KK^{'} + \Psi^2 \tag{4.3}$$

The matrix K, as a factor-pattern matrix, is unique in that each column all loadings are zero, except two. Its nonzero elements are not identified. Furthermore, each variable in Y is paired with each other variable with nonzero loadings in one of the columns of K. The matrix $KK^{'}$ is the matrix of doublet variances and covariances. Ψ^2 is the diagonal matrix of unique-factor variances. A doublet factor is correlated with only two variables in Y and orthogonal with all the rest. He further stipulate that in each column of λ there are at least three nonzero loadings. So, the common factors are minimally determined if not overdetermined from the observed variables.

He desired that the variance due to the common factors be separated from that due to the doublet and singlet factors. Thus we will seek the matrix $\Sigma - KK^{'} - \Psi^2 = \Lambda\Lambda^{'}$. Factoring the matrix on the left-hand side of the equality will yield a solution on the right (with suitable constraints). Knowing that partial correlation matrix P,

$$P = 2I - [diag(\Sigma^{-1})]^{-1/2}\Sigma^{-1}[diag(\Sigma^{-1})]^{-1/2} \tag{4.4}$$

of which an element is the partial correlation between a pair of variables that obtained when all $(p-2)$ other variables are held constant, contains the (approximate) information about doublet correlations between pairs of variables, Mulaik makes the strong assumption that

$$KK' + \Psi^2 = UPU \tag{4.5}$$

where U is some unknown diagonal matrix. Then the matrix we want to analyzed by finding its eigenvectors and eigenvalues and obtaining an orthogonal solution for the factors is

$$\Sigma - UPU = \Lambda\Lambda' \tag{4.6}$$

Clearly, this specifies that the residuals are allowed to be correlated, specifically, to be proportional to the partial correlations among variables.

Actually, Mulaik seems to base P on the sample covariance or correlation matrix S instead of Σ, according to the iterative methodology he described later (Mulaik 2010, p.261). He computed P at the start of computations and does not update it. A consequence is that model (4.6) is an odd combination of population parameters and sample data.

It is interesting to evaluate model (4.6) for the doublet example given in chapter 2. The partial correlation matrix is

$$\begin{bmatrix} 1.0000 & 0.6070 & 0.1215 & 0.1215 & 0.1215 \\ 0.6070 & 1.0000 & 0.1215 & 0.1215 & 0.1215 \\ 0.1215 & 0.1215 & 1.0000 & 0.2219 & 0.2219 \\ 0.1215 & 0.1215 & 0.2219 & 1.0000 & 0.2219 \\ 0.1215 & 0.1215 & 0.2219 & 0.2219 & 1.0000 \end{bmatrix} \tag{4.7}$$

Clearly, this matrix correctly identifies the doublet – the (1,2) element is quite elevated as compared to all other values. However, since any rescaling of this

matrix by a diagonal U maintains all off-diagonal elements as nonzero, model (4.6) cannot recover the correct model structure whether Λ is taken as a 1-factor or 2-factor model. Mulaik provides an iterative method to estimate Λ and U , but does not specify what objective function is being optimized by it.

4.2 Algorithm

The matrix P is easily obtained in (4.4), and substitute population correlation Σ with sample correlation S. The solution for U may require an iterative algorithm. Mulaik suggested to begin with a good approximation for U in the matrix $B = [diag(\Sigma^{-1})]^{-1/2}$. Then

$$\Sigma - BPB = \Sigma + B^2\Sigma^{-1}B^2 - 2B^2 = G \tag{4.8}$$

This suggests that factoring G will yield eigenvalues that may tell us how many factors to retain.

The iterative solution for the matrix U is as follows. Given S, a sample covariance or correlation matrix, compute S^{-1} and $B^2 = [diag(\Sigma^{-1})]^{-1}$. Find $B = (B^2)^{1/2}$. Then compute $P = -BS^{-1}B + 2I$. Now let $U^2_{(0)} = S^2$. Compute G in (4.8) and determine the number of factors to retain from its eigenvalues. Let this number be r and do not change it afterward during the iterations. Now, the iterations are as follows:

- Step 1: Obtain the eigenvectors A and the eigenvalues D of $C_{(i)} = S - U_{(i)}PU_{(i)}$. Compute $\Lambda_{r(i)} = A_{r(i)}D^{1/2}_{r(i)}$
- Step 2: Compute $W_{(i)} = S - \Lambda_{r(i)}\Lambda^{'}_{r(i)}$
- Step 3: Compute $U^2_{(i+1)} = diag(W_{(i)})$. Compute $U_{(i+1)}$.

By $u(j)^2_{(i)}$ denote the value of the jth diagonal elements of $U^2_{(i)}$ in the ith iteration. Then if $|u(j)^2_{(t+1)} - u(j)^2_{(t)}| > .0001$ for any j, j=1,...,p, go to step 1, else quit. Print the last $\Lambda_{r(i)}$.

In practice, for the purpose of identification, we could fix the sign of certain element (the element that has the largest absolute value in that column) in each column of Λ. And after every iteration, check the sign of the chosen element in each column; if the sign is opposite to what we fixed at the beginning, multiply that column by -1. We replicate the solution for Λ given in Mulaik(2010, p. 261) by implementing the above algorithm on the correlation matrix from Carlson and Mulaik (1993)[CM93].

CHAPTER 5

Bentler's Correlation Structure Methodology

5.1 Partial Correlations in a Universe of Variables

Mulaik used the idea that a partial correlation matrix can provide clues about associations among variables beyond those that depend on invariant common factors. This idea can be made more precise and then utilized in a different way. With the common factor model given by (2.1), it is well known that

$$\Sigma^{-1} = \Psi^{-2} - \Psi^{-2}\Lambda(I + \Lambda'\Psi^{-2}\Lambda)^{-1}\Lambda'\Psi^{-2} \tag{5.1}$$

Re-parameterizing so that $A = \Psi^{-1}\Lambda$, the model itself is $\Sigma = \Psi(AA' + I)\Psi$ and the inverse in (5.1) can be written as

$$\Sigma^{-1} = \Psi^{-2} - \Psi^{-1}A(I + A'A)^{-1}A'\Psi^{-1} \tag{5.2}$$

For identification, we may take $A'A$ to be diagonal. Now let the universe of variables be one where the number of variables gets arbitrarily large but the number of factors does not increase, that is, $p \to \infty$ while $k/p \to 0$. Then the diagonal of $A'A$ contains the sum of squares of an infinite number of (rescaled) loadings, which becomes arbitrarily large and hence, as is well known

$$\Sigma^{-1} \to \Psi^{-2} \tag{5.3}$$

This means that the partial correlation matrix (4.4) becomes $P_\Sigma \to 2I - \Psi\Psi^{-2}\Psi = I$, i.e., the partial correlations vanish.

In practice, one never has a universe of variables, but this ideal case suggests that in real data analysis, a good common factor solution should have enough indicators of each factor so that all partial correlations among variables become very small. When factors do not have enough strong indicators, partial correlations remain nonzero.

5.2 Model

In contrast to Mulaik, Bentler(2012)[Ben12] proposed to accept the standard factor analysis model, but use a new objective function to estimate parameters for this model. Since EFA is virtually always defined for correlations, we take model (2.1) to refer to a correlation structure model with S being the sample correlation matrix. He proposed to fit model (2.1) using a weighted least squares function. With the matrix of sample partial correlations given by

$$P = 2I - DS^{-1}D \tag{5.4}$$

where $D = [diag(S^{-1})]^{-1/2}$, and the i, j^{th} element of P as p_{ij} , we consider a weight matrix W with elements $w_{ii} = 0$, and off-diagonal elements $w_{ij} = 1 - p_{ij}^2$. Then, analogous to Yates(3.1), we may define the weighted least squares function

$$1'([S - \Lambda\Lambda']^{(2)} * W)1 \tag{5.5}$$

The proposed methodology is to consider W as fixed during the iterations, and to minimize (5.5) iteratively with respect to the factor loading matrix Λ, where

the matrix is specified in some arbitrary identified form (e.g., as lower triangular). Note that if a squared partial correlation is large, the i, j^{th} residual $(s_{ij} - \lambda_{i.}^{'}\lambda_{j.}^{'})^2$ in (5.5) is down-weighted during the next optimization step, i.e., fit of the factor model to that particular correlation or covariance becomes relatively less important. If all the partial correlations are about equal, the weighting will have little or no effect and the solution will approach a standard least squares solution. Similarly, when the variables are well sampled from the universe, $r_{ij}^2 \to 0$ and $w_{ij} \to 1$ and the weighting has no effect. Given the estimator $\hat{\Lambda}$, the unique variances are estimated as $\hat{\Psi}^2 = diag(S - \hat{\Lambda}\hat{\Lambda}^{'})$.

5.3 Algorithm

Minimization of (5.5) can be achieved in standard ways with a little modification, since not all the elements in Λ is free. Let Λ_f be a vector of free elements of Λ, e.g., the vector of lower triangular elements of Λ when the upper triangle is fixed at zero for identification. A modified quasi-Newton optimization method can be implemented with the gradient

$$\frac{\partial f}{\partial \Lambda_f} = -4\dot{\Lambda_f} vec\{[(S - \Lambda\Lambda^{'}) * W]\Lambda\} \tag{5.6}$$

where $\dot{\Lambda_f} = \partial\Lambda/\partial\Lambda_f$ is a matrix of 0s and 1s that selects the free elements of Λ.

Again, we could fix the sign of certain element (the element that has the largest absolute value in that column) in each column of Λ for the purpose of identification. And in every iteration, after getting a new $\hat{\Lambda}$, check the sign of the chosen element in each column; if the sign is opposite to what we fixed at the beginning, multiply that column by -1, and then update gradient and approximate of Hessian matrix accordingly.

In fact, we can compute the same residual covariance matrix using standard optimization packages in softwares without imposing any constraints on the model (i.e., fixing the upper triangle elements, and fixing the sign of certain element in each column of Λ). However, the factor loading estimates would be different from what we get from the above algorithm.

CHAPTER 6

Simulations

6.1 Methodology

To analyze the performance of Yates', Mulaik's and Bentler's proposed collinearity-resistant fitting method, compared with maximum likelihood extraction under normality assumption, we firstly generated a dataset of 10 variables and 500 observations, with zero mean and the following covariance matrix (which is the same as its correlation matrix), which follows multivariate normal distribution:

$$\begin{bmatrix} 1 & .75 & .49 & .49 & .49 & 0 & 0 & 0 & 0 & 0 \\ .75 & 1 & .49 & .49 & .49 & 0 & 0 & 0 & 0 & 0 \\ .49 & .49 & 1 & .49 & .49 & 0 & 0 & 0 & 0 & 0 \\ .49 & .49 & .49 & 1 & .49 & 0 & 0 & 0 & 0 & 0 \\ .49 & .49 & .49 & .49 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & .75 & .49 & .49 & .49 \\ 0 & 0 & 0 & 0 & 0 & .75 & 1 & .49 & .49 & .49 \\ 0 & 0 & 0 & 0 & 0 & .49 & .49 & 1 & .49 & .49 \\ 0 & 0 & 0 & 0 & 0 & .49 & .49 & .49 & 1 & .49 \\ 0 & 0 & 0 & 0 & 0 & .49 & .49 & .49 & .49 & 1 \end{bmatrix}_{10\times 10} \tag{6.1}$$

Clearly, the doublets here are produced by variable 1 and 2, and variable 6 and 7. If we could eliminate the effect of doublets completely, the theoretical factor

loading matrix should be

$$\begin{bmatrix} .7 & .7 & .7 & .7 & .7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .7 & .7 & .7 & .7 & .7 \end{bmatrix}' \tag{6.2}$$

In our simulation study, we randomly generated 5000 datasets which follow the multivariate normal distribution stated above, and implemented all the four extraction methods on them. It should be noted that the factor numbers are all set to 2, since we do not want to discuss how to determine the number of factors here. We will not discuss factor rotation methods neither. So in order to compare the results between different random samples, we specified the factor loading matrix in some arbitrary identified form. Specifically, we implemented orthogonal transformation through singular value decomposition (see algorithm in chapter 3) for Yates' method, and fix the sign of the elements [1,1] and [1,2] of the factor loading matrix to be positive. For Mulaik's method, since it is derived through eigendecomposition, we do not need to specify an identified form; but we need to fix the sign of the elements [1,1] and [10,2] of the factor loading matrix to be positive. Lastly, we fixed the upper triangle of the factor loading matrix to be zero for Bentler's method, i.e., the value of element [1,2] is 0 and only 19 elements of the matrix is estimated, and fix the sign of the elements [1,1] and [1,2] to be positive.

We also specified some control parameters for the simulations. For Yates', the termination condition of the iterations is that the changes in communalities are less than 10^{-4}. For Mulaik's, the tolerance is 10^{-6}. For Bentler's, we used BFGS to approximate Hessian matrix in quasi-Newton method, and stop the iterations when the norm of gradient is less than 10^{-6}. It should be noted that for each method in a given sample, when the iteration times is larger than 10^5, we would

take it as a case of non-convergence. In such situation, we stop the iterations, and take the last value as the result, while the replication number is marked for further analysis.

Last but not least, besides the estimated factor loading matrix, we define the doublet covariance residual(DR) as another measurement of the performance of the collinearity-resistant fitting methods. Since the covariance residual matrix is $R = S - \hat{\Lambda}\hat{\Lambda}'$, the doublet covariance residual is

$$dr = \frac{1}{n}\sum_{ij} r_{ij}^2 \tag{6.3}$$

where variable i and j, $i, j = 1, ..., p$, are detected doublets, n is the number of detected doublets, and r_{ij} are the corresponding elements in covariance residual matrix. Here in our datasets, $dr = (r_{12}^2 + r_{67}^2)/2$. The larger the doublet covariance residual is, the higher tolerance of doublet the method has.

6.2 Results

Before showing the result of simulations, let's first check how these methods work on the illustrative example given in chapter 2. As mentioned before, Yates' method is useless in the 1-factor case, so here we only estimate the example using Mulaik's and Bentler's method. We know that the maximum likelihood estimate is $[0.844, 0.844, 0.622, 0.622, 0.622]'$. Mulaik's method yields a loading of $[0.676, 0.676, 0.599, 0.599, 0.599]'$, and Bentler's yields $[0.783, 0.783, 0.67, 0.67, 0.67]'$. The doublet covariance residuals of Mulaik's and Bentler's are 0.0859, 0.0187 respectively, while that of ML is 0.00142. From this small example, we finds that both Mulaik's and Bentler's method are better than ML, and Mulaik's method produce better doublet covariance residual.

The result of 5000 times simulations is as follows. We notice that 9 out of 5000 cases cannot yield converged solutions using Yates' method, while all cases converged using other methods. It is also worth mentioned that this rate of convergence of Yates' method is under the tolerance of 10^{-4}, which was stated in the previous section; if we lower the tolerance (e.g., 10^{-6}) or change the measurement (e.g. use the change in loadings instead of commonalities as the termination condition), this algorithm would become much harder to converge. Therefore, we need to examine the difference of the results between the 4991 converged samples, and the 5000 complete samples.

From Table 6.1, we can see that the difference of mean and standard deviation of the factor loadings between these two groups are very small. In fact, the mean of doublet covariance residuals of the total 5000 samples is 0.080542, while that of converged samples is 0.08046, and the standard deviations are 0.186353 and 0.186447 respectively, which indicates that we can ignore the difference between these two groups. Therefore, we can use the 5000 complete samples' results to compare with that of other fitting methods.

Table 6.2 and Table 6.3 shows the means and standard deviations of factor loadings. Comparing the means cannot provide much information, since we didn't rotate the factors. However, comparing the standard deviation would show the stability of the estimator since they are all specified in some identified form. We can see that among the three resistant estimator, the stability of Bentler's is comparable with ML, while the variances of Yates' and Mulaik's are relatively large.

Doublet covariance residuals are shown in Table 6.4. We can see that Mulaik's has the largest DR mean, and its standard deviation is relatively small. Further-

Mean	N=5000		N=4991	
	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.509515	0.431093	0.509358	0.431545
Variable 2	0.507473	0.429744	0.507125	0.42993
Variable 3	0.444275	0.392055	0.44387	0.39205
Variable 4	0.446907	0.39032	0.446417	0.390405
Variable 5	0.447485	0.392607	0.447508	0.392969
Variable 6	0.001456	-0.00992	0.001368	-0.01016
Variable 7	0.001021	-0.01193	0.000576	-0.01154
Variable 8	-0.00061	-0.01044	-0.00059	-0.01045
Variable 9	0.002542	-0.0114	0.00246	-0.01114
Variable 10	0.001121	-0.01106	0.001061	-0.01085
Std	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.324028	0.27672	0.323854	0.276562
Variable 2	0.324536	0.280712	0.324273	0.280551
Variable 3	0.291425	0.273548	0.289821	0.272312
Variable 4	0.292914	0.26838	0.292208	0.267985
Variable 5	0.290784	0.270755	0.290661	0.270556
Variable 6	0.578148	0.532744	0.578435	0.532681
Variable 7	0.575164	0.5321	0.574299	0.530117
Variable 8	0.515414	0.497398	0.515596	0.497216
Variable 9	0.510276	0.499207	0.510423	0.498936
Variable 10	0.514335	0.499995	0.514663	0.499836

Table 6.1: Factor Loading Means and Standard Deviations of Yates' Method, Complete Samples vs Converged Samples

Mean	Maximum Likelihood		Yates	
	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.8440	0.0002	0.5095	0.4311
Variable 2	0.8436	0.0001	0.5075	0.4297
Variable 3	0.6223	-0.0002	0.4443	0.3921
Variable 4	0.6214	-0.0003	0.4469	0.3903
Variable 5	0.6216	0.0006	0.4475	0.3926
Variable 6	0.0002	0.8438	0.0015	-0.0099
Variable 7	-0.0001	0.8441	0.0010	-0.0119
Variable 8	0.0004	0.6223	-0.0006	-0.0104
Variable 9	0.0005	0.6222	0.0025	-0.0114
Variable 10	0.0004	0.6217	0.0011	-0.0111
	Mulaik		Bentler	
	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.4551	-0.0001	0.7835	0.0000
Variable 2	0.4549	-0.0004	0.7822	-0.0003
Variable 3	0.4035	-0.0006	0.6687	-0.0006
Variable 4	0.4027	-0.0005	0.6679	-0.0007
Variable 5	0.4027	0.0000	0.6681	0.0003
Variable 6	0.0005	0.4230	0.0007	0.7810
Variable 7	0.0004	0.4228	0.0002	0.7814
Variable 8	0.0006	0.3755	0.0007	0.6683
Variable 9	0.0007	0.3751	0.0008	0.6683
Variable 10	0.0008	0.3756	0.0006	0.6678

Table 6.2: Factor Loadings – Means

Standard Deviation	Maximum Likelihood		Yates	
	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.0196	0.0368	0.3240	0.2767
Variable 2	0.0196	0.0367	0.3245	0.2807
Variable 3	0.0331	0.0386	0.2914	0.2735
Variable 4	0.0331	0.0379	0.2929	0.2684
Variable 5	0.0331	0.0381	0.2908	0.2708
Variable 6	0.0366	0.0197	0.5781	0.5327
Variable 7	0.0368	0.0192	0.5752	0.5321
Variable 8	0.0383	0.0335	0.5154	0.4974
Variable 9	0.0384	0.0337	0.5103	0.4992
Variable 10	0.0388	0.0335	0.5143	0.5000
	Mulaik		Bentler	
	Factor 1	Factor 2	Factor 1	Factor 2
Variable 1	0.1869	0.4645	0.0234	0.0000
Variable 2	0.1874	0.4645	0.0234	0.0335
Variable 3	0.1689	0.4116	0.0299	0.0487
Variable 4	0.1683	0.4114	0.0298	0.0475
Variable 5	0.1692	0.4113	0.0298	0.0478
Variable 6	0.4974	0.1781	0.0563	0.0241
Variable 7	0.4977	0.1783	0.0565	0.0233
Variable 8	0.4418	0.1571	0.0546	0.0301
Variable 9	0.4417	0.1581	0.0540	0.0304
Variable 10	0.4415	0.1562	0.0541	0.0300

Table 6.3: Factor Loadings – Standard Deviations

more, one-way ANOVA indicates significant difference between the means, and Figure 6.1 visualize the DR means with confidence intervals, in which the CIs are too small to be plotted except Yates'. Also, two group t-test between ML and other three methods shown in Table 6.5 demonstrates that all these three resistant estimators have significant effect on reducing the influence of doublets in factor analysis.

	ML	Yates	Mulaik	Bentler
Mean	0.0015	0.0805	0.0859	0.0194
Std	0.0006	0.1864	0.0089	0.0050

Table 6.4: Doublet Covariance Residuals

	Yates	Mulaik	Bentler
t value	-30.0047	-667.34	-250.875
DF	4999.116	5049.478	5159.631
p value	2.20E-16	2.20E-16	2.20E-16

Table 6.5: T Test Result between Maximum Likelihood and Others

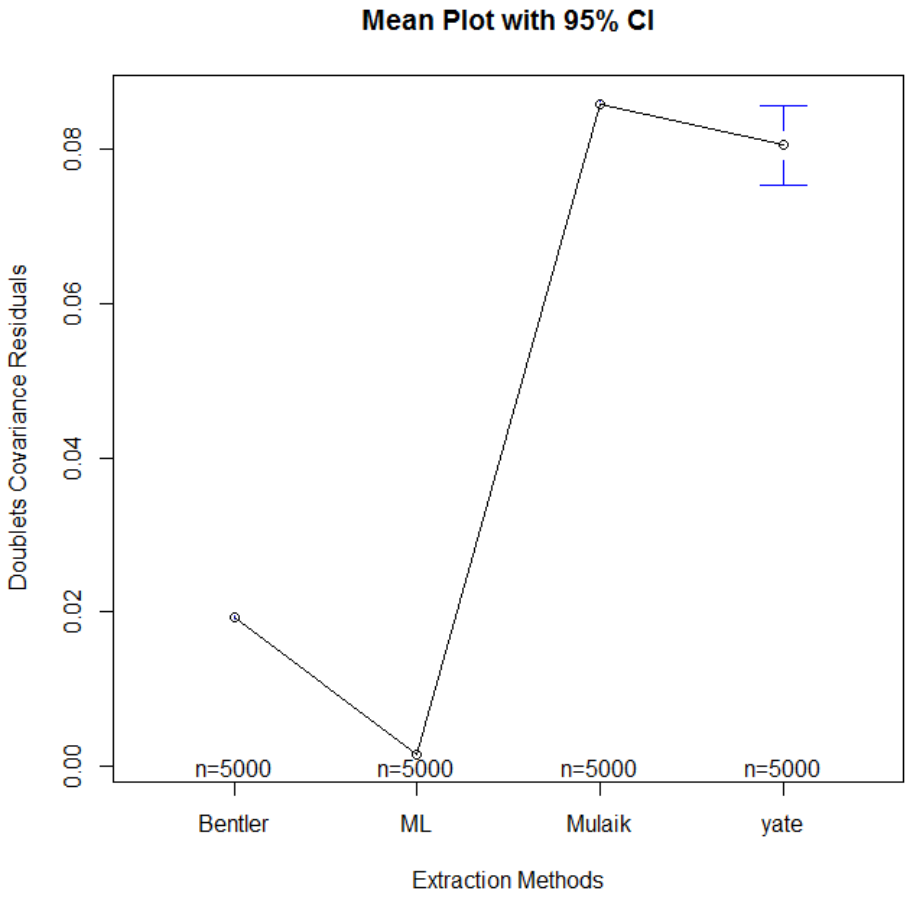

Figure 6.1: Mean Plot with 95% CI

CHAPTER 7

Conclusion and Discussion

In sum, we can conclude that all these three collinearity-resistant fitting methods, Yates', Mulaik's and Bentler's, can help reduce the distortion resulted from doublets. Among the three estimator, Mulaik's estimates produces the largest doublet covariance residual with small variance, and Bentler's estimated loadings have the smallest variance, according to our simulation results.

In Yates(1987), he introduced this collinearity-resistant fitting procedure very thoroughly, including how to deal with "Heywood case", how to accelerate the algorithm with little calculation wasted on displacement of nearly stabilized elements of the factor matrix. The simulations in this paper did not take all the suggestions he gave. Improvement might be able to seen if some modification is implemented according to Yates' suggestions.

It is also interesting to mention that since the weight V is a function of Λ, we can rewrite the (3.2) to

$$V = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}_{p \times p} - \frac{(\Lambda\Lambda')^{(2)}}{(\Lambda^{(2)} 1_{k\times 1})(\Lambda^{(2)} 1_{k\times 1})'} \tag{7.1}$$

and then use a standard optimization method(e.g., quasi-Newton) to minimize the function (3.1). However, we cannot replicate the table(Yates 1987, pp.240-243) in this way. The reason of this problem should be studied in the future.

Bentler(2012) mentioned that the minimization function (5.5) is not the only approach, e.g., we also could minimize $1'([S - \Lambda\Lambda' - \Psi^2]^{(2)} * W)1$ with respect to both Λ and Ψ^2, assuring that now $w_{ii} = 1$. And while this methodology seemed to work well in our simulations, it did not do well when the added doublet association was negative rather than positive. The reason for this is quite simple: a positive doublet association has the consequence that its associated observed variable partial correlation (5.4) is increased over and above the baseline of no doublet. On the other hand, a negative doublet association in the correlation matrix has the effect of reducing the associated partial correlation. Hence such a doublet will not be down-weighted by the proposed procedure. Although doublets most likely will be positive in empirical research due to similarly acting method or extraneous influences, this limitation is an inadequacy of the current approach.

References

[BB80] Peter M Bentler and Douglas G Bonett. "Significance tests and goodness of fit in the analysis of covariance structures." *Psychological bulletin*, **88**(3):588, 1980.

[BCB93] Michael W Browne, Robert Cudeck, Kenneth A Bollen, and J Scott Long. "Alternative ways of assessing model fit." *Sage Focus Editions*, **154**:136–136, 1993.

[Ben12] Peter M. Bentler. "Resistant Exploratory Factor Analysis." 10 2012.

[But68] John M. Butler. "Descriptive factor analysis." *Multivariate Behavioral Research*, **3**(3):355–370, 1968.

[CM93] Marianne Carlson and Stanley A Mulaik. "Trait ratings from descriptions of behavior as mediated by components of meaning." *Multivariate Behavioral Research*, **28**(1):111–159, 1993.

[De 88] Jan De Leeuw. "Model selection in multinomial experiments." *On model uncertainty and its statistical implications*, pp. 118–138, 1988.

[FWM99] Leandre R Fabrigar, Duane T Wegener, Robert C MacCallum, and Erin J Strahan. "Evaluating the use of exploratory factor analysis in psychological research." *Psychological methods*, **4**(3):272, 1999.

[Gut53] Louis Guttman. "Image theory for the structure of quantitative variates." *Psychometrika*, **18**(4):277–296, 1953.

[IW08] Z Imel and B Wampold. "The importance of treatment and the science of common factors in psychotherapy." *Handbook of counseling psychology*, pp. 249–266, 2008.

[Lor56] Frederic M Lord. "A study of speed factors in tests and academic grades." *Psychometrika*, **21**(1):31–50, 1956.

[Mac90] Robert C MacCallum. "The need for alternative measures of fit in covariance structure modeling." *Multivariate Behavioral Research*, **25**(2):157–162, 1990.

[Mac03] Robert C MacCallum. "2001 Presidential Address: Working with Imperfect Models." *Multivariate Behavioral Research*, **38**(1):113–139, 2003.

[MBC07] Robert C MacCallum, Michael W Browne, and Li Cai. "Factor analysis models as approximations." *Factor analysis at*, **100**:153–175, 2007.

[Mul10] Stanley A. Mulaik. *Foundations of Factor Analysis*. Chapman & Hall/CRC, second edition, 2010.

[NL10] Megan Norris and Luc Lecavalier. "Evaluating the use of exploratory factor analysis in developmental disability psychological research." *Journal of autism and developmental disorders*, **40**(1):8–20, 2010.

[Tuk77] John W Tukey. "Exploratory data analysis." *Reading, MA*, **231**, 1977.

[Yat87] Allen Yates. *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. State University of New York Press, Albany, 1987.