SEFAiR SO FAR

BEN GOODRICH (BGOKGM@GMAIL.COM)

For more information on FAiR (including installation instructions) go to

`http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:fair`

1. Introduction

This paper introduces a new estimation technique called semi-exploratory factor analysis (SEFA). SEFA can more-or-less be seen as a generalization of confirmatory factor analysis (CFA) in the direction of exploratory factor analysis (EFA). In CFA, the analyst uses prior theory to specify which coefficients are fixed — usually to zero — and estimates the remaining coefficients. In SEFA, the analyst specifies how many coefficients are zero for each factor, but the *locations* of these exact zeros are estimated along with the values of the corresponding non-zero coefficients using an optimization algorithm that has not been applied to factor analysis problems before.

This ongoing project synthesizes three Big Ideas that are most certainly not new, but some of them may be new to many practitioners of factor analysis. The first Big Idea is a philosophy of factor analysis, which is inspired by Yates (1987), although it differs in almost all of the operational details. For a largely favorable book review, see MacCallum (1989), but in brief, Yates (1987) argues that the traditional work-flow of EFA — extract factors, test whether the number of factors is sufficient, transform using quartimin or varimax — renders EFA merely another form of cluster analysis, a descriptive tool for describing how in-sample outcomes relate to each other. Yates (1987) aims to restore factor analysis's original purpose as a scientific tool for inferring the relationships between latent factors and outcomes in the population and proposes new algorithms for extracting and transforming factors to promote the *invariance* of the solution to moderate changes in the battery. I argue these goals are more attainable through SEFA.

The second Big Idea is a free, libre, open source software (FLOSS) philosophy. All of the estimation techniques referenced here are implemented in a package for the R language (see R Core Development Team 2007) called FAiR that can be downloaded and used for free. R has been quite popular among statisticians for a few years but is only slowly gaining influence in psychometrics. For example, de Leeuw and Mair (2007) guest edited a special edition of the *Journal of Statistical Software* dedicated to "Psychometrics in R", *Structural Equation Modeling* published Fox (2006) on using R to estimate structural equation models (including CFA), and Doug Bates, who is part of the R Core

I would like to thank Jas Sekhon for his advice and for extensively modifying RGENOUD to make this paper possible, Mike "Combo" Kellerman for solving some combinatorics problems and commenting on an application of these methods, and the participants at the Applied Statistics Workshop at Harvard University and the Psychology Methods Group at the University of Rhode Island for valuable comments.

Development Team is scheduled to present a workshop on the use of R to estimate mixed-effects models at the 2008 International Meeting of the Psychometric Society.

FAiR, like R itself, is licensed under the General Public License (GPL), which gives everyone the right to use FAiR for any purpose, to view all the source code, and redistribute modifications under the same license.[1] This license is designed to foster collaboration and hopefully a sizeable group of people will eventually contribute to FAiR's development. As far as I know, no other major software package for factor analysis is currently covered by a GPL-compatible license, aside from the scattered tools for factor analysis that have already made it into R or R packages.[2]

The third Big Idea is a philosophy of optimization and is related to the second because the main estimation techniques that are (or will be) implemented in FAiR essentially require the use of other GPL software, such as a genetic optimization algorithm called RGENOUD (see Mebane and Sekhon 2007) or eventually the headers in a Markov Chain Monte Carlo library called MCMCpack (see Martin and Quinn 2007). Difficult factor analysis problems can be solved with FAiR because of RGENOUD's unique ability to find the global optimum of a nonlinear and / or discontinuous function with many local optima and plateaus, possibly subject to inequality constraints on the parameters.

There are three main sections in this paper, written in sharply increasing order of difficulty to understand. The next section discusses the common factor analysis model in the population and contrasts SEFA with the two most popular ways to estimate it, CFA and EFA. Anyone who is familiar with EFA and CFA will be able to understand the simple examples that are given. This paper, however, does not cover *how* to use FAiR to estimate the SEFA model.

The following section explains how RGENOUD works in general and specifically for SEFA. RGENOUD represents a fundamental departure from traditional optimization algorithms that are used for factor analysis. It is unnecessary to have a deep understanding of RGENOUD in order to use FAiR, but it is important to understand how constraints are operationalized. I do not discuss the possibility of using Markov Chain Monte Carlo (MCMC) to estimate the SEFA model, but doing so is reasonably straightforward in some cases.

The final major section discusses some of the "advanced" features of SEFA. In particular, the recurring topics in Yates' (1987) discussion of EFA — like Thurstone's (1935) definition of simple structure, the positive manifold, suppressor variables, the distinguishability of factors, the invariance of the solution, etc. — are reexamined in light of SEFA. In short, I believe FAiR can accomplish Yates' (1987) goals much better than Yates' (1987) EFA algorithms could. However, as with any new estimation technique, a lot more simulations, as well as improvements to the code and the theory, need to be done in order to determine SEFA's full potential.

[1] See http://www.fsf.org/licensing/licenses/gpl.html for the license terms.

[2] For descriptions of existing tools, see http://cran.r-project.org/src/contrib/Views/Psychometrics.html or Mair and Hatzinger (2007) for a static version of the same document. In particular, Bernaards and Jennrich (2005) is the basis for the GPArotation package; Revelle (2007) is a package providing tools with an emphasis on personality research; Husson, Josse, Le, and Mazet (2007) is a package that focuses on components but also has tools for factor analysis; and Falissard (2007) is a package that emphasizes tools for measurement. Bill Rozeboom recently agreed to license HYBALL under the GPL so that parts of it might eventually be included in FAiR.

2. The Common Factor Analysis Model in the Population and A New Way To Estimate It

This section briefly introduces some notation for the factor analysis model. I assume the reader has seen the equations for factor analysis before but restate a couple here. For a full treatment, see Harman (1976) or another textbook on factor analysis. EFA, CFA, and SEFA only differ in how they estimate the same population model.

For $j = 1, 2, \ldots, n$ outcome variables (also called manifest variables or tests), $p = 1, 2, \ldots, r \ll n$ common factors, and $i = 1, 2, \ldots, N$ units of observation, the common factor analysis model in the population can be written as a linear regression model

$$Y_{ij} = \sum_{p=1}^{r} \beta_{jp} X_{ip} + \Theta_j E_{ij},$$

where Y_{ij} is an observable score on outcome j, X_{ip} is a score on an unobserved common factor with effect β_{jp}, E_{ij} is a score on an unobserved factor that is unique to the jth outcome with effect Θ_j, and all variables are expressed as deviations from their expectations. If the unique factors are independent of each other and the common factors, this model implies that

$$\underset{n\times n}{\boldsymbol{\Sigma}} = \underset{n\times r}{\boldsymbol{\beta}} \underset{r\times r}{\boldsymbol{\Phi}} \underset{r\times n}{\boldsymbol{\beta}'} + \underset{n\times n}{\boldsymbol{\Theta}^2},$$

where $\boldsymbol{\Sigma}$ is a covariance matrix among outcomes, $\boldsymbol{\beta}$ is the primary pattern matrix of coefficients, $\boldsymbol{\Phi}$ is the correlation matrix among common factors, and $\boldsymbol{\Theta}^2$ is a diagonal matrix with unique variances along the diagonal (in this paper).[3]

Let $\mathbf{S}$ be an estimate of $\boldsymbol{\Sigma}$ that is derived from a sample of data with N observations. If the common and specific factors are normally distributed, $\mathbf{S}$ is distributed Wishart with $\mathbb{E}[\mathbf{S}] = \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}' + \boldsymbol{\Theta}^2$ and $N - 1$ degrees of freedom. Let tr $(\bullet)$ be the trace function. The Wishart log-likelihood of can be written as

$$\ell\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Phi}}, \widetilde{\boldsymbol{\Theta}^2}\right) \propto -\frac{N-1}{2}\left(\ln\left|\widetilde{\mathbf{S}}\right| + \text{tr}\left(\mathbf{S}\widetilde{\mathbf{C}}^{-1}\right)\right),$$

where $\widetilde{\mathbf{C}} = \widetilde{\boldsymbol{\beta}}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\beta}}' + \widetilde{\boldsymbol{\Theta}^2}$ is a proposal for the optimum. Most software actually optimizes a monotonic function of ℓ instead of ℓ itself, but FA*i*R primarily uses this log-likelihood function. I do not discuss other discrepancy functions in this paper, but others are (or will be) available in FA*i*R. Some of the limitatations of maximum likelihood estimation of factor analysis models can be mitigated with SEFA, as may become clear in the last section. One feature of maximum likelihood is that it is scale invariant, and $\mathbf{S}$ can thus be taken as a correlation matrix for most purposes.

If at least r^2 restrictions are placed on the factor analysis model, then the remaining parameters can be estimated without rotational indeterminacy. EFA and CFA differ in the nature of the restrictions they impose. There are several kinds of restrictions. "Identification conditions" have the effect of choosing one solution over all others with the same

[3] At present, FA*i*R does not allow $\boldsymbol{\Theta}^2$ to be non-diagonal, but adding this feature will be straightforward.

value of $\widehat{\ell}$. “Exclusion restrictions” specify cells of $\boldsymbol{\beta}$ are exactly zero. I do not discuss restrictions where cells of $\boldsymbol{\beta}$ are assigned non-zero value, but such restrictions are supported in FAiR. I also do not discuss restrictions where multiple cells of $\boldsymbol{\beta}$ are constrained to be equal, and such restrictions are not currently supported in FAiR. “Inequality restrictions”, which are discussed in the last section and are fundamental in FAiR, are conditions where some parameter or function of parameters is greater than or less than another parameter, a function of parameters, or a threshold that the analyst specifies. Inequality restrictions are generally not “binding” in the sense that if some set of parameters satisfies an inequality restriction, then a small perturbation of those parameters will also satisfy that inequality restriction.

When estimating a CFA model, the analyst imposes at least r^2 restrictions on $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ that reflect substantive beliefs about the population and then tests whether the restrictions were reasonable, perhaps relative to another CFA model. A useful theorem on rotational indeterminacy is proven in / paraphrased from Howe (1955, p.87):

Theorem 1. *A set of sufficient conditions for eliminating rotational indeterminacy is that, after eliminating rows of $\boldsymbol{\beta}$ that contain all zeros,*

(i) $\boldsymbol{\Phi}$ is a full-rank correlation matrix, imposing r restrictions that the common factor scores have unit variance.

(ii) At least $r-1$ cells in each of the r columns of $\boldsymbol{\beta}$ are zero, which imposes at least $r\left(r-1\right)$ exclusion restrictions.

(iii) $\overset{p}{\boldsymbol{\beta}}$ is of rank $r-1$, where $\overset{p}{\boldsymbol{\beta}}$ is the submatrix of $\boldsymbol{\beta}$ with exact zeros in the pth column.

Theorem 1, which can also be extended to estimates of $\boldsymbol{\beta}$, is said to be “minimally satisfied if the identification condition given in (i) holds and *exactly* $r-1$ cells in each of the r columns of $\boldsymbol{\beta}$ are fixed at zero such that condition (iii) holds. There are slightly fewer than $\frac{1}{r!}\prod_{p=0}^{r-1}\left\{\binom{n}{r-1}-p\right\}$ unique ways to minimally satisfy theorem 1, all of which yield the same $\widehat{\mathbf{C}}$, and hence the same $\widehat{\ell}$ at the optimal estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Theta}^2$ (denoted by $\widehat{\ }$). Thus, there is no statistical way to decide among different CFA models that minimally satisfy theorem 1 in different ways.

The CFA tradition presumes that the the exclusion restrictions are chosen *a priori* based on theory. In SEFA, the number of exclusion restrictions is specified *a priori* but the locations of the zeros in $\widehat{\boldsymbol{\beta}}$ chosen *a posteriori* by the algorithm. In other words, SEFA asserts that theorem 1 holds in a *non-minimal* fashion and “specifies” that a given number of exclusion restrictions are located wherever they, along with the corresponding free parameters, maximize ℓ (possibly subject to other restrictions), which can be restated as a corollary of theorem 1:

Corollary 1. *A set of sufficient conditions for eliminating the rotational indeterminacy for $\widehat{\boldsymbol{\beta}}$ is that, after eliminating rows of $\widehat{\boldsymbol{\beta}}$ that contain all zeros,*

(i) $\widehat{\boldsymbol{\Phi}}$ is a full-rank correlation matrix,

(ii) $r-1$ cells in each of the r columns of $\widehat{\boldsymbol{\beta}}$ are zero,

(iii) $\overset{p}{\widehat{\boldsymbol{\beta}}}$ is of rank $r-1$, where $\overset{p}{\widehat{\boldsymbol{\beta}}}$ is the submatrix of $\widehat{\boldsymbol{\beta}}$ with exact zeros in the pth column,

(iv) At least one additional binding restriction is imposed.

TABLE 1. The 16 "Unique" Configurations of the Primary Pattern Matrix that Satisfy Corollary 1 when There Are Five Outcomes, Two Factors, and Two Required Zeros per Factor

0	X	0	X	0	X	0	X	0	X	0	X	X	0	X	X
0	X	0	X	0	X	X	0	X	0	X	X	0	X	0	X
X	0	X	0	X	X	0	X	0	X	0	X	0	X	0	X
X	0	X	X	X	0	X	0	X	X	X	0	X	X	X	0
X	X	X	0	X	0	X	X	X	0	X	0	X	0	X	0

0	X	0	X	0	X	0	X	0	X	0	X	X	0	X	0
X	0	X	0	X	X	X	0	X	0	X	X	0	X	0	X
X	0	X	X	X	0	X	0	X	X	X	0	X	X	X	X
0	X	0	X	0	X	X	X	X	0	X	0	0	X	X	0
X	X	X	0	X	0	0	X	0	X	0	X	X	0	0	X

Note: X signifies a non-zero coefficient. It is possible to switch the columns and /or reflect factors to obtain additional configurations, but these additional configurations are not substantively important.

This corollary is really just a special case of theorem 1 applied to an estimate of $\boldsymbol{\beta}$, but it is important to note that schemes that minimally satisfy theorem 1 do *not* satisfy corollary 1 due to condition (iv). Typically, condition (iv) is satisfied with at least one more exclusion restriction, but there are some other possibilities as well.

If corollary 1 is satisfied, $\widehat{\ell}$ is not guaranteed to be unique, but it typically seems to be in practice in well-constructed research designs (up to sign changes and reorderings of the factors). Nevertheless, Millsap (2001), Bollen (1989), and Bollen and Jöreskog (1985) have generated CFA examples where rotational indeterminacy is eliminated but some parameters remain unidentified. Following most of the literature, I will ignore this inconvenient truth and proceed "as if" corollary 1 were sufficient to ensure $\widehat{\ell}$ is unique.[4] It is always possible to verify whether there is at least *local* identification in a SEFA model by simply checking whether the information matrix is invertible at the optimum.

Let the number of zeros per factor required for a SEFA model be $b \geq r-1$ to define an r-dimensional hyperplane in common factor space. In geometric terms, SEFA estimates the locations the hyperplane edges as part of the optimization algorithm. The most straightforward way to satisfy conditions (ii) and (iv) of corollary 1 is to set $b = r$, which corresponds to the second of Thurstone's (1935, 1947) rules of thumb for assessing the probable uniqueness of $\widehat{\boldsymbol{\beta}}$. When corollary 1 is satisfied such that $b = r$, I will refer to the model as a "vanilla" SEFA model. Table 1 enumerates the 16 vanilla configurations of $\widetilde{\boldsymbol{\beta}}$ that satisfy corollary 1 when there are $n = 5$ outcomes and $r = 2$ factors.

It would be straightforward to obtain CFA estimates of the free parameters for each of these 16 vanilla configurations and then analyze the results for the model where $\widehat{\ell}$ was the largest among the 16. One can *conceptualize* SEFA estimates as the CFA estimates with the largest $\widehat{\ell}$ among all CFA models that could have been estimated with b exclusion restrictions per factor. However, the SEFA algorithm does not literally estimate all such CFA models by

[4] The problematic examples in Millsap (2001) occur when theorem 1 is minimally satisfied, which is insufficient for SEFA. Thus, SEFA models appear to handle the examples in Millsap (2001) well, although I believe there is some rounding error in the covariance matrices presented in tables 1 and 3 of Millsap (2001). The problematic example in Bollen and Jöreskog (1985) does satisfy corollary 1. However, I am not aware of an example where rotational indeterminacy is eliminated, some parameters are not identified, there are at least as many zeros per factor as there are factors, and there are at least three non-zeros per factor.

TABLE 2. Three Primary Pattern Configurations for Harman's Data on Eight Physical Variables

Solution	Transformed EFA		Pretty Good CFA		SEFA	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
height	0.871	0.079	0.912		0.850	0.126
arm span	0.970	-0.055	1.005	-0.123	0.945	
forearm	0.935	-0.049	0.970	-0.115	0.913	
lower leg	0.874	0.042	0.895		0.853	0.090
weight	-0.005	0.957	-0.024	0.967		0.954
bitro	-0.005	0.800	-0.023	0.811		0.798
girth	-0.065	0.793		0.758	-0.060	0.788
width	0.130	0.609		0.676	0.130	0.615
Factor 1	1.000	0.476	1.000	0.542	1.000	0.427
Factor 2	0.476	1.000	0.542	1.000	0.427	1.000
$\widehat{\ell}$	-199.464		-204.285		-199.474	

"brute force" because doing so quickly becomes impractical when n or r is of even moderate size. Instead, a genetic algorithm — which will be discussed in the next section — is used for SEFA that maximizes the likelihood jointly over the locations of the exact zeros and the corresponding free parameters.

To further fix ideas about the vanilla SEFA model, consider Harman's (1976) data on $n = 8$ physical variables where $r = 2$, and four variables that pertain to "lankiness" are followed by four variables that pertain to "stockiness". Excluding reorderings and reflections of the factors, there are 210 ways to place exactly two zeros into each column of $\boldsymbol{\beta}$ that satisfy corollary 1 for these data. Table 2 shows three estimated primary pattern matrices. The first configuration has no *exact* zeros and hence does not satisfy corollary 1 because it is a EFA model that is similar to the result obtained by Harman (1976) using a graphical transformation.[5] The second solution is a CFA with zeros on the first factor for the girth and width measurements and on the second factor for the height and lower leg measurements. The third is the vanilla SEFA model where the locations of the zeros were unspecified but happen to maximize ℓ when located on the first factor for the weight and bitro measurement and on the second factor for the arm span and forearm measurement. While both of the last two configurations satisfy corollary 1, are consistent with the lankiness-stockiness division, and produce a good fit at their respective optima, the $\widehat{\ell}$ for the SEFA model is slightly higher than that of the CFA model and is only a hundredth worse than that of the EFA model. We will return to the issue of model comparison often, but clearly the SEFA model is as adequate as the EFA and CFA models in this case.

EFA is often characterized as an "unrestricted" model but is properly speaking a minimally restricted model that imposes exactly r^2 restrictions on the parameters. *One* way to estimate a EFA model is to impose the preliminary restrictions that the factors are orthogonal and that the upper triangle of the primary pattern matrix is filled with zeros. Doing so imposes r^2 restrictions on the model, which can be estimated using CFA algorithms as if the these arbitrary

[5] Thurstone's (1935, 1947) criterion was used to transform the factors, which is possible for the first time thanks to RGENOUD. See Goodrich (2008) for more details.

TABLE 3. Major Differences between EFA, CFA, and SEFA

Consideration	EFA	CFA	SEFA
Number of restrictions	r^2	$\geq r^2$	$> r^2$
Transformation required / allowed	Yes	No	No
Specific cells of $\boldsymbol{\beta}$ constrained to be zero in advance	No	Yes	Maybe
Simultaneous estimation of second order factor analysis model	No	Maybe	Maybe
Computationally simple these days	Yes	Yes	No

restrictions represented the theory to be confirmed. This way of estimating the EFA model is not too prevalent in software, but for the purposes of discussion, it does not matter what algorithm is used to obtain EFA estimates of the preliminary primary pattern matrix, here denoted $\widehat{\boldsymbol{\Lambda}}$.

EFA introduces a second equation that $\boldsymbol{\Phi} = \mathbf{T}'\mathbf{T}$, where $\mathbf{T}$ is a $r \times r$ matrix with unit-length columns. By specifying these r^2 cells of $\mathbf{T}$, the analyst thereby substitutes r^2 restrictions for the r^2 preliminary restrictions used to obtain $\widehat{\boldsymbol{\Lambda}}$. It can be shown that $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Lambda}} \left[\mathbf{T}'\right]^{-1}$ and that $\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}' = \widehat{\mathbf{C}} - \widehat{\boldsymbol{\Theta}^2} = \widehat{\boldsymbol{\beta}}\mathbf{T}'\mathbf{T}\widehat{\boldsymbol{\beta}}'$. The process of choosing a $\mathbf{T}$ is known as "choosing a transformation" of the factors by optimizing some objective function, generically denoted $f(\mathbf{T})$. The most popular algorithm for selecting $\mathbf{T}$ in applied research is called "varimax" (see Kaiser 1958).

A much less popular, but much more effective, $f(\mathbf{T})$ is the "simplimax" criterion (see Kiers 1994), which is also noteworthy for sharing with SEFA the general attitude of "I know there are certain number of zeros in $\boldsymbol{\beta}$ but I will tell you where they are later." However, when using the simplimax criterion with EFA, the analyst specifies that some gross number of cells in $\widehat{\boldsymbol{\beta}}$ are (near) zero but does not require a particular allocation of zeros across the r factors. In SEFA, condition (ii) of corollary 1 requires $r-1$ exact zeros per column of $\widehat{\boldsymbol{\beta}}$.

Table 3 summarize the major differences between EFA, CFA, and SEFA. Besides the computational difficulty of SEFA to be discussed in the next section, the major differences between the three methods pertain to the restrictions. The conceptual difference between EFA and SEFA is this: An EFA analyst chooses the $\mathbf{T}$ that is closest to the optimum of a given $f(\mathbf{T})$, which implies an estimate of $\boldsymbol{\beta}$ that is optimal among all combinations of estimates that yield $\widehat{\ell}_{EFA}$. A SEFA model produces estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ that are optimal with respect to ℓ among all combinations of estimates that satisfy the restrictions the analyst imposes. In other words, EFA yields an estimate of $\boldsymbol{\beta}$ that satisfies some identification condition among all estimates with the same fit, and SEFA yields an estimate of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ with the best fit among all estimates that satisfy the restrictions imposed.

This distinction is subtle but important. Exploratory factor analysts have long conceptualized choosing $\mathbf{T}$ as following Ockham's Razor by choosing the "simplest" among equally good explanations for $\mathbf{S}$, although defining "simplest" has proven to be complicated. The most common definition for the "simplest" $\widehat{\boldsymbol{\beta}}_{EFA}$ is the one with the most zeros, which would better be termed "sparsest", since this conception of simplicity and / or its utility is sharply contested by

Butler (1969), Cureton and D'Agostino (1983), Yates (1987), and others. SEFA tries to avoid this issue by placing restrictions on the model so that there (hopefully) is a unique, best-fitting solution, obviating the need to define "simplest" in order to choose among equally good explanations for $\mathbf{S}$. In other words, the primary reason to prefer SEFA to EFA is that it avoids the necessity to choose $\mathbf{T}$. SEFA is closer to CFA in that deciding among models is, in part, a matter of choosing the best tradeoff between model fit and the number of estimated parameters.

How much violence do the SEFA restrictions to the model fit? Note that $\widehat{\ell}_{EFA}$ is an upper bound for ℓ_{SEFA}, so one should always verify whether $\widehat{\ell}_{SEFA}$ is within plausible sampling variability of $\widehat{\ell}_{EFA}$ for an EFA model with r factors (as was the case with the example from Harman (1976) above). If $\widehat{\ell}_{SEFA} \ll \widehat{\ell}_{EFA}$, then one should be skeptical of the restrictions placed on the SEFA model, unless the restrictions have a strong theoretical motivation, as is discussed in the last section. Unfortunately, $\widehat{\ell}_{SEFA} \approx \widehat{\ell}_{EFA}$ does not necessarily imply that $\widehat{\boldsymbol{\beta}}_{SEFA}$ and $\widehat{\boldsymbol{\Phi}}_{SEFA}$ are good estimates, and it is still necessary to justify the theoretical appropriateness of the restrictions and verify that the estimates appear reasonable. But $2\left(\widehat{\ell}_{EFA} - \widehat{\ell}_{SEFA}\right) \approx 0$ does imply that even if the True $\mathbf{T}'\mathbf{T} = \boldsymbol{\Phi}$ were found by optimizing $f\left(\mathbf{T}\right)$, the differences between $\widehat{\boldsymbol{\beta}}_{EFA}$ and $\widehat{\boldsymbol{\beta}}_{SEFA}$ could plausibly be attributed to sampling variability.

The conceptual difference between SEFA and CFA is this: SEFA estimates the locations of a given number of zeros in $\widehat{\boldsymbol{\beta}}$; CFA specifies the locations of the zeros *a priori*. However, it is also possible to estimate a "mixed" SEFA model, which specifies some cells in $\widehat{\boldsymbol{\beta}}$ *a priori* and estimates the locations of the remaining zeros, such that the restrictions collectively satisfy corollary 1. Thus, a "pure" SEFA model that requires $b \geq r$ zeros per factor but leaves all their locations unspecified nests all CFA models that have at least b zeros per factor. The exception to this generalization is the rare case of a CFA model that minimally satisfies theorem 1, in which case a SEFA model is *more* restrictive from a statistical perspective. However, even in that case the SEFA model would be considered *less* restrictive in everyday language because it does not specify the locations of the b exact zeros for each factor.

I have so far glossed over *how* a SEFA model is estimated, which is a nontrivial but not insurmountable task to be discussed in the next section. The goal of this section was merely to situate SEFA between EFA and CFA and emphasize the virtues of SEFA. To oversimplify, but only slightly, SEFA is a fully statistical model that is similar to EFA but without the necessity of the transformation step, and CFA is a special case of SEFA. Another way to conceptualize SEFA is that it maximizes the likelihood of $\mathbf{S}$ over a huge number of CFA models that impose at least b exclusion restrictions per factor. In other words, once SEFA estimates are obtained, $\widehat{\boldsymbol{\beta}}_{SEFA}$ has b zeros per factor, implying that the SEFA estimates are "as if" a CFA model had imposed *those* exclusion restrictions from the outset, and any theorems that are applicable to CFA estimates are also applicable to SEFA estimates.

Lisa Harlow remarked that one virtue of SEFA, which will become more apparent in the last section, is that it makes transparent the process of establishing restrictions on the factor analysis model. I think that is an important point, and I hope it also makes the plausibility of the restrictions easier to judge. SEFA could largely replace both

EFA and CFA because it genuinely has the ability to span almost the entire range of the continuum between EFA and CFA models, depending on the number and nature of the restrictions that the analyst chooses to impose. For example, a vanilla SEFA is no less "exploratory" than a EFA, which usually implies that $\widehat{\mathbf{\Theta}^2}_{SEFA} \approx \widehat{\mathbf{\Theta}^2}_{EFA}$ and any differences between $\widehat{\boldsymbol{\beta}}_{SEFA}$ and $\widehat{\boldsymbol{\beta}}_{EFA}$ will usually be primarily attributable to the choice of $\mathbf{T}$ for the EFA model. Alternatively, SEFA can be "confirmatory" and orient its restrictions toward testing a specific theory, including fixing some coefficients to particular values (zero or otherwise), placing bounds on the parameters, etc. My general preference is to estimate SEFA models with $b = r$ and impose some theoretically motivated inequality restrictions on the parameters, which are discussed in the last section.

3. Estimating the SEFA Model via a Genetic Algorithm

The computational difficulty of SEFA comes from a variety of sources, but it is obvious that the number of ways to satisfy corollary 1 increases at an increasing rate in both n and r. Thus, SEFA requires either a genetic algorithm to simultaneously find the optimal configuration of zeros and the optimal values of the corresponding free parameters or MCMC to sample from the joint posterior distribution (which is not discussed in this paper). It appears as if the phrase "genetic algorithm" has appeared in only five *Psychometrika* articles, and by any measure, this approach to optimization seems to be largely unknown in psychology and other fields that use factor analysis.[6]

The particular genetic algorithm used by FAiR is called RGENOUD (see Mebane and Sekhon 2008 and the references to proofs therein). The main virtues of genetic algorithms are that they are robust to merely local optima and are flexible enough to work in all kinds of non-standard problems. Table 2 in Mebane and Sekhon (2008) compares RGENOUD favorably against a genetic algorithm described in Yao, Liu, and Lin (1999) in their ability to minimize twenty-two functions, many of which are much too difficult for a traditional optimization algorithm, such as

$$\begin{aligned}
f(\boldsymbol{\theta}) &= \sum_{i=1}^{30} u(\theta_i, 10, 100, 4) + 10\sin^2(\pi\tau_1) + \sum_{i=1}^{29}\left\{(\tau_i - 1)^2\left[1 + 10\sin^2(\pi\tau_{i+1})\right] + \tau_{30} - 1\right\}, \text{ where} \\
\tau_i &= 1 + \frac{\theta_i + 1}{4}, \text{ and} \\
u(\theta_i, k, a, r) &= \begin{cases} k(\theta_i - a)^r & \text{if}\, \theta_i > a, \\ 0 & \text{if } -a \leq \theta_i \leq a, \\ k(-\theta_i - a)^r & \text{otherwise,} \end{cases}
\end{aligned}$$

which is a piecewise sinusoidal function in thirty dimensions. SEFA problems can be even *more* difficult than this, but RGENOUD can find the global optimum nonetheless.

[6] The Wikipedia entry (http://en.wikipedia.org/wiki/Genetic_algorithm) appears to be a fair introduction to the topic, as of the time of this writing.

A genetic algorithm uses principles of Evolution to find the optimum of an objective function. Consider a population of parameter vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ where K is "large" (like 1,000 or more) and $\boldsymbol{\theta}_k$ — which is called the kth "individual" — is a vector that contains all the free or potentially free parameters to be estimated. For example, each $\boldsymbol{\theta}_k$ would be a vector of length $0.5r\left(r-1\right)+nr+n$ if we are estimating the off-diagonals of $\boldsymbol{\Phi}$, all cells of $\boldsymbol{\beta}$, and the diagonal elements of $\boldsymbol{\Theta}^2$, even though at least $r\left(r-1\right)$ of the coefficients will be pegged to zero. The parameters can be thought of as constituting the DNA of an individual.

We need an objective function, $f\left(\boldsymbol{\theta}\right)$, which can be just about any real function, such as a log-likelihood function or other discrepancy function. One point that must always be kept in mind is that the *ranks* of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ with respect to $f\left(\boldsymbol{\theta}\right)$ are more relevant for RGENOUD than are the numeric *values* of $f\left(\boldsymbol{\theta}_1\right), f\left(\boldsymbol{\theta}_2\right), \ldots, f\left(\boldsymbol{\theta}_K\right)$ or the numeric values of the gradients $\frac{\partial f(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}, \frac{\partial f(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}, \ldots, \frac{\partial f(\boldsymbol{\theta}_K)}{\partial \boldsymbol{\theta}_K}$, which stands in sharp contrast to most of the optimization algorithms that have been used in factor analysis previously.

A genetic algorithm is Evolutionary rather than iterative, and we next need rules that govern how $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ progress over the generations, which are indexed by the superscripts $t = 0, 1, 2, \ldots, \infty$. RGENOUD uses nine "heuristics" or "operators" but I will only mention three of the most intuitive here. The first is "cloning", which implies that $\boldsymbol{\theta}_k^{[t+1]} = \boldsymbol{\theta}_k^{[t]}$ and that the kth individual is the same in generation $t+1$ as in generation t. The second is "uniform mutation", which is like noisy cloning and implies that $\boldsymbol{\theta}_k^{[t+1]}$ is the same as $\boldsymbol{\theta}_k^{[t]}$, except in one element, which is randomly changed. The third is "simple crossover", which first selects two "parents", such that $\boldsymbol{\theta}_k^{[t]} \neq \boldsymbol{\theta}_{k'}^{[t]}$, and a random integer a. Then two children in the $t+1$ generation are created Frankenstein-style, such that $\boldsymbol{\theta}_k^{[t+1]}$ consists of the first a elements of $\boldsymbol{\theta}_k^{[t]}$ and the remaining elements of $\boldsymbol{\theta}_{k'}^{[t]}$ while $\boldsymbol{\theta}_{k'}^{[t+1]}$ consists of the first a elements of $\boldsymbol{\theta}_{k'}^{[t]}$ and the remaining elements of $\boldsymbol{\theta}_k^{[t]}$. The nine operators are probabilistically applied to the population at the end of each generation. See table 1 in Mebane and Sekhon (2008) for more details on operators.

Evolution implies that if individuals with higher fitness are more likely to be cloned, to be parents, etc. than are less fit individuals, then the average fitness of the population improves in expectation each generation. And if the average fitness of the population improves in expectation, then as $t, K \rightarrow \infty$, some $\boldsymbol{\theta}_k^{[t]}$ must eventually approach the global optimum of $f\left(\boldsymbol{\theta}\right)$. More formally, the population of individuals can be seen as a suitable Markov chain because $\boldsymbol{\theta}_k^{[t]}$ depends only on the parameter values in the $t-1$ generation, suitable Markov chains asymptotically converge to their stationary distributions, and the stationary distribution of this Markov chain is where $\boldsymbol{\theta}_k = \boldsymbol{\theta}\, \forall k$ at the global optimum of $f\left(\boldsymbol{\theta}\right)$. In practice, we have to impose stopping conditions based on the gradient and / or level of $f\left(\boldsymbol{\theta}\right)$ to obtain an answer in finite time with finite K, but the result is typically correct to the specified precision if care is taken.

Finally, we need "mapping rules" from $\boldsymbol{\theta}$ to $\widetilde{\boldsymbol{\Phi}}$, $\widetilde{\boldsymbol{\beta}}$, and $\widetilde{\boldsymbol{\Theta}^2}$. Since RGENOUD only requires that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ be *ranked* with respect to $f\left(\boldsymbol{\theta}\right)$, there is no need for $f\left(\boldsymbol{\theta}\right)$ to be continuous and differentiable everywhere, which permits discontinuous or otherwise complicated mapping rules from $\boldsymbol{\theta}$ to $\widetilde{\boldsymbol{\Phi}}$, $\widetilde{\boldsymbol{\beta}}$, and $\widetilde{\boldsymbol{\Theta}^2}$. To understand the essential relationship between mapping rules and corollary 1, let sort $(\{\bullet\})$ be the standard function that takes a numeric set (represented by $\bullet$) and sorts it in ascending order and let sort $(\{\bullet\})_{1:b}$ indicate the first b ascending elements of the sorted set. $\boldsymbol{\theta}$ fills the off-diagonals of $\widetilde{\boldsymbol{\Phi}}$, the diagonal elements of the $\widetilde{\boldsymbol{\Theta}^2}$, and a preliminary coefficient matrix $\check{\boldsymbol{\beta}}$ (with a supra-check). Then, $\widetilde{\boldsymbol{\beta}}$ (with a supre-tilde) could be derived from $\check{\boldsymbol{\beta}}$ according to a mapping rule, for example:

$$\widetilde{\beta}_{jp} \quad = \quad \begin{cases} 0 & abs\left(\check{\beta}_{jp}\right) \in \text{sort}\left(\left\{abs\left(\check{\boldsymbol{\beta}}_p\right)\right\}\right)_{1:b} \\ \check{\beta}_{jp} & \text{otherwise.} \end{cases}$$

In words, $\widetilde{\beta}_{jp}$ is simply $\check{\beta}_{jp}$, unless $\check{\beta}_{jp}$ is among the b smallest absolute values slated for the pth factor, in which case $\widetilde{\beta}_{jp}$ is squashed to exactly zero to eliminate rotational indeterminacy. In other words, the mapping rule ensures that $\widetilde{\boldsymbol{\beta}}$ satisfies condition (ii) of corollary 1. To give a hypothetical example of this mapping rule where $n = 5$ and $r = 2 = b$,

$$\boldsymbol{\theta}_k = \begin{bmatrix} 0.25 & 0.9 & 0.8 & 0.4 & 0.3 & 0.5 & 0.1 & 0.1 & 0.7 & 0.6 & 0.5 & 0.11 & 0.22 & 0.33 & 0.44 & 0.55 \end{bmatrix}$$

$$\widetilde{\boldsymbol{\Phi}}_k = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \check{\boldsymbol{\beta}}_k = \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.1 \\ 0.4 & 0.7 \\ 0.3 & 0.6 \\ 0.5 & 0.5 \end{bmatrix} \Longrightarrow \widetilde{\boldsymbol{\beta}}_k = \begin{bmatrix} 0.9 & 0 \\ 0.8 & 0 \\ 0 & 0.7 \\ 0 & 0.6 \\ 0.5 & 0.5 \end{bmatrix} \widetilde{\boldsymbol{\Theta}^2}_k = \begin{bmatrix} 0.11 & 0 & 0 & 0 & 0 \\ 0 & 0.22 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0.44 & 0 \\ 0 & 0 & 0 & 0 & 0.55 \end{bmatrix},$$

and the log-likelihood for the kth individual in generation t is evaluated at $\widetilde{\boldsymbol{\beta}}_k$ rather than $\check{\boldsymbol{\beta}}_k$.

This mapping rule from $\boldsymbol{\theta}$ to $\check{\boldsymbol{\beta}}$ to $\widetilde{\boldsymbol{\beta}}$ implies that $\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is flat in some (here four) dimensions, which would be problematic for optimization algorithms that rely on gradients but not for RGENOUD because the population Evolves pseudo-randomly and with discrete jumps over the generations. Any deterministic mapping rule that results in condition (ii) of corollary 1 being satisfied can be used, and FAiR provides several mapping rules to choose from, which are discussed in the next section and will generally yield different fits at their respective optima.[7]

Mapping rules are one cornerstone of SEFA; lexical ranking is the other, which is just a formal method for breaking ties that is also used for playoff seeding in many sports.[8] If $\mathbf{z} = f(\boldsymbol{\theta})$ is a *vector* of fit criteria, then lexical ranking is used to rank $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ with respect to $f(\boldsymbol{\theta})$ as follows:

(1) $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K$ are ranked by the first criterion $\left(z^{[1]}\right)$ produced by $f(\boldsymbol{\theta})$.

(2) If any are tied on $z^{[1]}$, those population members are then ranked by the second criterion $\left(z^{[2]}\right)$.

(3) If still tied, the third, . . ., last criterion is utilized until all ties among unique population members are broken.

[7] Also, FAiR does not actually use the primary pattern matrix to implement these mapping rules. Instead, some combination of the reference structure matrix and the factor contribution matrix are used.

[8] See, for example, http://www.nfl.com/standings/tiebreakingprocedures .

Thus, if $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k'}$ are tied on $z^{[1]}$ (sports: overall record), but $\boldsymbol{\theta}_k$ is better than $\boldsymbol{\theta}_{k'}$ on $z^{[2]}$ (sports: head-to-head record), then $\boldsymbol{\theta}_k$ is ranked better than $\boldsymbol{\theta}_{k'}$ with respect to $f(\boldsymbol{\theta})$ (sports: playoff seeding), regardless of whether $\boldsymbol{\theta}_{k'}$ is better than $\boldsymbol{\theta}_k$ on any subsequent criterion (sports: conference record, net margin of victory, etc.).

It should be noted in passing that EFA *is* an example of lexical optimization, just in two separate steps. First, the analyst finds some set of parameter values that maximizes ℓ. But there are an infinite number of sets of parameters that also maximize ℓ that are within a transformation of the MLEs. So the analyst then breaks ties among all sets of parameters that yield $\widehat{\ell}$ by optimizing a second objective function to choose the optimal $\mathbf{T}$. In theory, but not with floating point numbers, one could conduct an EFA via lexical optimization by specifying that $z^{[1]} = \widetilde{\ell}$ and $z^{[2]} = -f\left(\widetilde{\mathbf{T}}\right)$ and would theoretically obtain the same result as conducting the EFA in two conventional steps. But if EFA were explicitly recognized as a lexical optimization problem, then the full power of lexical optimization could be brought to bear when conducting a EFA, as is discussed in Goodrich (2008).

Lexical optimization with respect to $f(\boldsymbol{\theta})$ is equivalent to optimization with respect to the last criterion among population members with the same values on all previous criteria. If the last criterion is fully continuous and all previous criteria are operationalized as "constraints", then lexical optimization is tantamount to constrained optimization, but is *much* more flexible than traditional constrained optimization techniques like Lagrange multipliers and linear programming. By a "constraint", I mean a fit criterion that equals 1.0 if some qualitative condition is sufficiently satisfied by $\boldsymbol{\theta}_k$ and equals some number less than 1.0 otherwise.

The recognition of this point led to the birth of SEFA, where ℓ is the last criterion in $\mathbf{z} = f(\boldsymbol{\theta})$ and the previous criteria are various constraints. For example, to impose the necessary constraints that unique variances are positive and that condition (iii) of corollary 1 holds, one could operationalize the first two constraints in $\mathbf{z}$ as

$$z^{[1]} = \begin{cases} 1 & \text{if } \widetilde{\Theta_j^2} > 0\,\forall j, \\ 0 & \text{otherwise.} \end{cases} \qquad z^{[2]} = \begin{cases} 1 & \text{if condition}\,(iii)\text{ of corollary 1 holds for } \widetilde{\boldsymbol{\beta}} \\ 0 & \text{otherwise.} \end{cases}$$

Based on these two criteria alone, the ranking of candidates in generation t would be, in descending order:

(1) all individuals where $z^{[1]} = 1$ and $z^{[2]} = 1$

(2) all individuals where $z^{[1]} = 1$ and $z^{[2]} = 0$

(3) all individuals where $z^{[1]} = 0$ and $z^{[2]} = 1$

(4) all individuals where $z^{[1]} = 0$ and $z^{[2]} = 0$

and ties among individuals within each of these four groups are broken by subsequent criteria.[9] In the $t+1$ generation, we can expect at least as many individuals to satisfy these two constraints due to "survival of the fittest" principles.

[9] SEFA actually uses more complicated piecewise functions for computational efficiency, but these only come into play for suboptimal candidates.

Three other criteria are standard in vanilla SEFA for more technical reasons:

$$z^{[3]} = \begin{cases} 1 & \text{if } \widetilde{\mathbf{\Phi}} \text{ is positive definite} \\ 0 & \text{otherwise,} \end{cases} \qquad z^{[4]} = \begin{cases} 1 & \text{if } \left(\sum_{j=1}^{n} \widetilde{\beta}_{jp}\right) > 0 \, \forall p, \\ 0 & \text{otherwise.} \end{cases} \qquad z^{[5]} = \begin{cases} 1 & \text{if } \widetilde{\mathbf{C}} \text{ is positive definite} \\ 0 & \text{otherwise,} \end{cases}$$

Since $\widetilde{\mathbf{\Phi}}$ and $\widetilde{\mathbf{C}}$ are estimates of correlation matrices, it is reasonable to require that they be positive definite up to computer tolerance to avoid the prospect of trying to invert a matrix that is computationally singular. The rationale for the fourth criterion is less obvious but follows from the fact that factors can arbitrarily by multiplied by -1 to reverse their polarity without affecting ℓ. However, RGENOUD has no way of knowing that there are many places in $\boldsymbol{\theta}$-space that yield the same $\widehat{\ell}$ where the MLEs are merely reflections of each other. Thus, the fourth criterion has the effect of confining the search to the region of $\boldsymbol{\theta}$-space where the column-sums of $\widetilde{\boldsymbol{\beta}}$ are positive, which excludes some of the duplicative modes in the likelihood function and concentrates the population into a subset of $\boldsymbol{\theta}$-space.

The ultimate criterion, $z^{[6]} = \widetilde{\ell}$, breaks all ties among unique individuals. Thus, vanilla SEFA seeks the $\boldsymbol{\theta}$ that maximizes ℓ, subject to the condition that $z^{[1]} = z^{[2]} = z^{[3]} = z^{[4]} = z^{[5]} = 1$. Among individuals in generation t that satisfy all the constraints, the ones with the highest values of $\widetilde{\ell}$ are ranked highest, which drives the Evolutionary dynamic toward the global optimum of $f\left(\boldsymbol{\theta}\right)$, which is the constrained optimum of ℓ. More constraints can (and should) be added but the essence of lexical optimization remains the same. Creativity is necessary to construct the mapping rules and the constraints, which ideally would exclude all "bad optima" of ℓ. Carefully choosing the constraints and mapping rule in light of the problem at hand is more important than understanding in great depth how RGENOUD works, although Evolution is a convenient and pertinent metaphor for understanding the process. The most important things to understand about genetic algorithms are that they are more likely to be successful in finite time as the size of the population increases, as the number of generations increases, and as the regularity of $f\left(\boldsymbol{\theta}\right)$ increases in the region of $\boldsymbol{\theta}$-space where the constraints hold.

4. Model Comparison and Additional Restrictions on SEFA Models

CFA makes rather *specific* assumptions about $\boldsymbol{\beta}$ that the analyst thinks are appropriate for a particular research project; for example that $\boldsymbol{\beta}_{12} = 0$. EFA makes rather *general* assumptions about $\boldsymbol{\beta}$ and $\mathbf{\Phi}$ when using an algorithm to select $\mathbf{T}$, such as whether $\mathbf{\Phi} = \mathbf{T}'\mathbf{T}$ is an identity matrix. SEFA makes at least one general assumption by specifying b but can also make assumptions about specific parameters. Many papers in the long history of (mostly exploratory) factor analysis have suggested or implied that $\boldsymbol{\beta}$, $\mathbf{\Phi}$, or functions thereof should satisfy some inequality constraints that are fairly general, but it is very difficult to respect nonlinear inequality constraints with a traditional optimization algorithm. The preceeding section illustrates that FAiR can easily impose *any* constraint by formulating it as a piecewise criterion that precedes ℓ in the lexical ranking. It is also possible to place bounds on any elements of $\boldsymbol{\theta}$. This section discusses possible constraints, some of which represent a quite radical departure from conventional thinking in factor analysis models, and explores alternative mapping rules.

While genetic algorithms have excellent asymptotic properties, SEFA problems are quite difficult, and FAiR may fail to find a global optimum in a reasonable period of time. The primary difficulty is that there are a lot of local optima in a $\boldsymbol{\theta}$-space of very high dimensionality and several good optima may be far apart from each other in $\boldsymbol{\theta}$-space. Thus, the chances of failure can be lessened by imposing additional restrictions on the model that narrow the admissible region of $\boldsymbol{\theta}$-space. Of course, poorly chosen restrictions may serve to exclude the best optimum from consideration, so it is important to think carefully about what restrictions to impose. But there are plenty of reasons to believe that vanilla SEFA will not be scientifically fruitful in many situations, and thus it is typically necessary to put some additional restrictions on a SEFA model.

4.1. **Simple Structure.** For starters, Guttman (1992) criticizes Thurstone's concept of simple structure (and EFA generally) as an unfalsifiable hypothesis because it lacks a test statistic that could provide a basis for rejection in a given application. In other words, one can optimize $f(\mathbf{T})$ to produce $\mathbf{T}$, but nothing indicates whether the resulting $\widehat{\boldsymbol{\beta}}_{EFA}$ is a "simple enough" to sustain the simple structure hypothesis. Perhaps Guttman (1992) reads Thurstone (1935, 1947) a bit literally, but the general point that the choice of $\mathbf{T}$ is not a statistical decision is undeniable.

With SEFA, it is easy to test the hypothesis that a particular simple structure configuration is plausible. Thurstone's definition of simple structure implies that *each* row of $\boldsymbol{\beta}$ contains at least one exact zero in one of its r cells so that each outcome is "in" one or more of the r hyperplanes (or dimensionality $r-1$) in the common factor space. Thus, there are r^n ways to achieve simple structure under this strict definition. Consider a mapping rule where $\boldsymbol{\theta}$ fills a preliminary matrix, $\check{\boldsymbol{\beta}}$, and then a simple structure configuration can be enforced by specifying that

$$\widetilde{\beta}_{jp} = \begin{cases} 0 & \text{if } abs\left(\check{\beta}_{jp}\right) = \min\left\{abs\left(\check{\boldsymbol{\beta}}_j\right)\right\}, \\ \check{\beta}_{jp} & \text{otherwise.} \end{cases}$$

In words, $\widetilde{\beta}_{jp}$ is simply the corresponding element of $\check{\boldsymbol{\beta}}$, unless the corresponding element of $\check{\boldsymbol{\beta}}$ is the smallest absolute value slated for the jth row of $\widetilde{\boldsymbol{\beta}}$, in which case it gets squashed to exactly zero by the mapping rule.

Once $\widehat{\ell}$ is found by SEFA with simple structure as the restriction scheme, one can *test* whether the restrictions hold by comparing model fit statistics to those for an EFA model with an equivalent number of factors. Such a test is not a test of the null hypothesis that simple structure holds; it is a test of a the null hypothesis that the restrictions hold for *the most likely configuration* of the parameters that satisfies Thurstone's definition of simple structure and the other constraints. In other words, rejecting this null implies that the data are inconsistent with the simple structure in the most favorable case but does not necessarily imply that the cumulative evidence for simple structure — taking into account all good, but suboptimal, configurations of $\boldsymbol{\beta}$ that satisfy simple structure — is insufficient. I have no idea how to test a simple structure hypothesis over almost r^n possible ways to achieve simple structure, but such a test is probably not what Guttman (1992) had in mind anyway.

Thurstone recognized that simple structure was an unnecessarily strong hypothesis because it imposed many more restrictions than were necessary and that it was quite possible for any particular test to be a function of all r factors. Thus, a literal simple structure hypothesis may be rejected in a given application, but we can weaken the mapping rule considerably. Let the "complexity" of the jth outcome be the number of non-zeros in $\boldsymbol{\beta}_j$. Thurstone (1947) only "demands that the list of traits be long enough so that, after elimination of several traits of complexity r, enough traits of complexity less than r remain to determine uniquely both the trait configuration and the simple structure (335)."

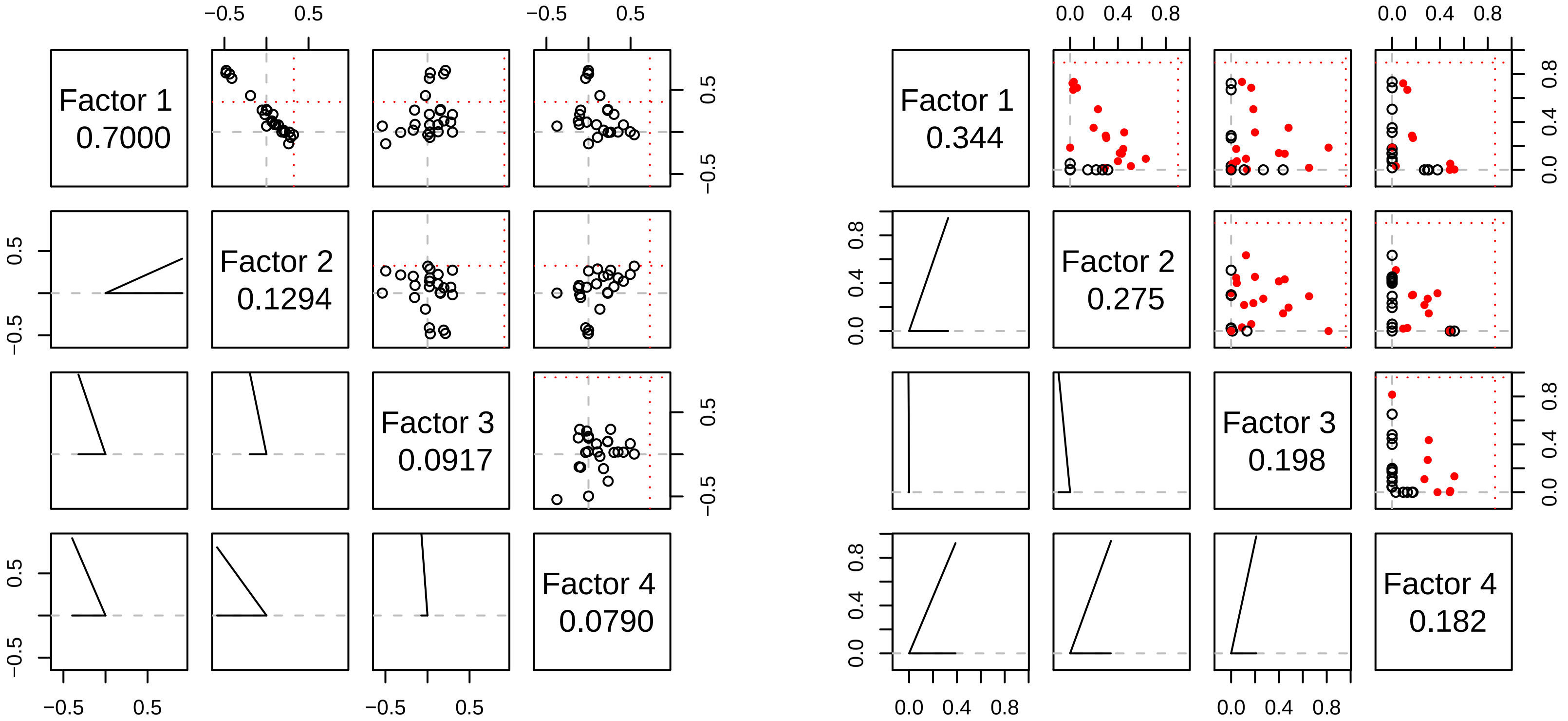
This weaker simple structure hypothesis requires that r tests per factor are of complexity $r - 1$ — although these r^2 tests could have "non-zero" coefficients that are arbitrarily close to zero — and is feasible in the usual case where $r^2 < n$. In other words, up to $n - r^2$ tests are allowed to be of complexity r under the weaker version of the simple structure hypothesis and there are slightly fewer than $\binom{n}{r^2} \times r^{r^2}$ ways to satisfy corollary 1 with this scheme. The only inherent difference between vanilla SEFA and weak simple structure SEFA is that vanilla SEFA allows the r^2 zeros to be allocated such that some tests can be of complexity less than $r - 1$, while weak simple structure SEFA does not.

In this section, I present four-factor solutions for the ubiquitous Holzinger and Swineford (1939) data on twenty-four mental tests to illustrate different SEFA options. In my opinion, one should impose any restrictions that seem appropriate at the outset, examine the model fit statistics and plots, judge which models are plausible, and only then examine attempt to interpret the primary pattern matrix, as if one were evaluating a candidate for the transformed solution in an EFA. To that end, I do not undertake any substantive interpretations in this paper and focus on the process by which a model's plausibility can be judged.

Aside from model fit statistics, the main tool is a figure where the correlations between outcomes and reference factors for each outcome are plotted two factors at a time in the upper right, which is the mechanism Thurstone relied on to judge if the model were plausible. Thurstone (1935, 1947) gives some guidelines for interpretting these plots, but basically the analyst is trying to discern order from the configuration of reference structure correlations, which are discussed in a bit more detail below and to put it simply, are proportional to coefficients in the primary pattern matrix. The angle between each pair of primary factors is indicated by the lines in the lower left of the figure, and the correlation itself can be read off the x-axis. Finally, the diagonal of the figure lists the proportion of communality associated with each factor, which will be defined more rigorously toward the end of this section.

The plot to the left of figure 1 is an example of what the solution should *not* look like. It is an EFA, transformed using Thurstone's (1935) criterion subject to the constraint that the factors must remain distinct, which FA*i*R makes feasible for the first time ever and is discussed more in Goodrich (2008). Although it is possible to obtain better results

FIGURE 1. Simple Structure Solutions for the Holzinger and Swineford (1939) Data



The plot on the left is a EFA transformed to simple structure using Thurstone's (1935) criterion, while the plot on the right is a SEFA that imposes the simple structure restriction that all tests are of complexity $r - 1$. Although technically successful, the first solution should be considered poor, while the second is plausible. The log-likelihoods for the two solutions are -1027.736 and -1035.664 respectively.

using a more complicated scheme for choosing $\mathbf{T}$, this EFA would loosely be considered a "successful" transformation to simple structure in the sense that each outcome has at least one reference structure correlation that is smaller than 0.10 in absolute value. But we have no statistical way to formally test a simple structure with EFA; it is simply the best that could be done with respect to this $f(\mathbf{T})$. Although there was no "factor collapse", the first two factors are unreasonably correlated, and there is virtually no systematic ordereding of the reference structure correlations.

In contrast, the plot to the right of figure figure 1 is almost exactly what Thurstone wanted to see when conducting a factor analysis but is a SEFA with the strong simple structure requirement that there be one exact zero per outcome. Although there appears to be a lot of order to the configuration, one aspect that looks abberrant is that the third factor is nearly orthogonal to the first two factors, while the rest of the inter-factor correlations are moderately positive and is more in accordance with theory on primary mental abilities.

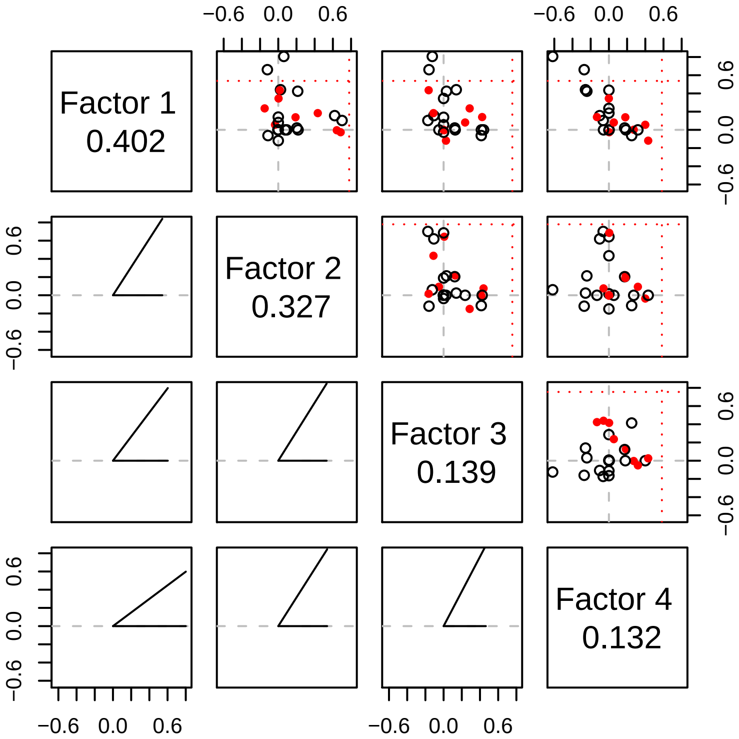
Since the plot appears reasonable, we can turn to some model fit statistics, of which FAiR calculates several. For the sake of brevity, I will focus exclusively on the log-likelihood, although one would not want to do so in actual reasearch. Recall that $\widehat{\ell}_{EFA}$ is an upper bound for the log-likelihood in a SEFA model with an equivalent number of factors. This SEFA with strong simple structure differs in (twice) $\widehat{\ell}$ from the EFA by about (sixteen) eight units but estimates twelve fewer coefficients (due to the exact zeros). Thus, a likelihood-ratio test would fail to reject the SEFA with strong simple structure in favor of the EFA. Hence the combined evidence from prior theory about mental tests, the plots, and the model fit statistic(s) suggest that this simple structure SEFA model is plausible.

Figure 2 contains a SEFA solution with the weak simple structure requirement that 16 of the outcomes have one exact zero and 8 outcomes are allowed to be perfectly complex. The significance of the different symbols is as follows: A closed, red dot indicates that outcome has an exact zero loading on one of the factors *besides* the two being plotted. An open, black circle indicates that the outcome does not fall in *any* other hyperplanes besides (potentially) the two being plotted. Hence for the strong simple structure SEFA, there were no open, black circles in the interior of any plot because every test was restricted to be in one hyperplane but this requirement is relaxed under weak simple structure.

For the weak simple structure SEFA, the first three factors do look too bad when paired with each other but the fourth factor is a disaster. There is little order to the reference structure correlations on the fourth factor, and it is hightly correlated with the other factors, especially the first. Based on this plot alone, one might conclude that the weak simple structure SEFA model should be reestimated with three factors, although there is independent evidence for four factors.

4.2. **Model Comparison and Invariance.** These plots illustrate a counter-intuitive point: The most plausible model is the strong simple structure SEFA, followed by the weak simple structure SEFA, and then the strong simple structure EFA. But the EFA model (necessarily) has the best fit, followed (necessarily) by the weak simple structure SEFA, and then by the strong simple structure SEFA. It is true that the difference in $\widehat{\ell}$ between the EFA and the strong

FIGURE 2. SEFA with Weak Simple Structure for the Holzinger and Swineford (1939) Data



This SEFA requires that r^2 (16) of the n (24) outcomes be of complexity of $r-1$ (3). Although the log-likelihood (-1028.028) is comparable to that the EFA (-1027.736) in figure 1, one should not consider this solution plausible because the reference structure correlations on the fourth factor appear chaotic.

simple structure SEFA would not be considered statistically significant, but it emphasizes that factor analysts (should) typically seek the best estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Theta}^2$ rather than seeking the $\mathbf{C} = \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}' + \boldsymbol{\Theta}^2$ that best fits $\mathbf{S}$. Although the two goals are intertwined, the former does not imply the latter and the latter most certainly does not imply the former. Thus, one cannot decide among models merely by automatic application of some model comparison statistic, especially the simple comparisons of the log-likelihood that are made in this paper.

Moreover, model comparison tools typically are intended to select the model that would produce the best out-of-sample predictions if the same battery were collected for a different random sample of individuals. Yates (1987) argues that the greater concern is not that the model would poorly predict the correlations among the same outcomes in a new sample but that the interpretation would qualitatively change if a different subset of the domain of outcomes were collected on the *same* units of observation. Furthermore, a solution tends to be invariant to pertubations of the battery when it *fails* to take into account minor factors, which necessarily increases the misfit between $\mathbf{S}$ and $\widehat{\mathbf{C}}$. Statistical tools for model comparison tell us nothing about Yates' (1987) concerns, and for all these reasons, common sense and substantive theory should be used as much as any model comparison statistic.

In this case, when the strong simple structure requirement that each outcome be of complexity $r - 1$ is relaxed so that eight outcomes can be perfectly complex, the weak simple structure SEFA allocates the zeros among clusters of similar tests. Yates (1987) only discusses EFA but would have anticipated this phenomenon, which is problem for vanilla or weak simple structure SEFA but is actually a problem with factor analysis generally. While this allocation of zeros is "optimal" in the sense that it maximizes ℓ, it is decidedly suboptimal from a scientific perspective because the resulting factors would not be robust to the removal of a cluster from the battery. The short version of Yates' (1987) argument is that clusters are battery-dependent and are created by the *investigator*, rather than the data-generating process. It is trivially easy to create a cluster by including a sufficiently large number of similar tests from the domain of possible outcomes. But in a homogenous domain, there is a large number (or perhaps even a continuum) of possible outcomes that only differ in degree from each other, depending on differences in how strongly they are related to different latent factors in the population. Thus, the domain is not overtly characterized by clusters, and a model for the domain of interest should generally predict that outcomes differ in degree rather than in kind.

Moreover, Yates (1987) argues that a cluster solution is also difficult to interpret *scientifically* because the one factor that a cluster of outcomes appears to load on may actually be a linear combination of multiple factors. In other words, when $\widehat{\boldsymbol{\beta}}$ exhibits a cluster configuration, it implies that each outcome in the cluster is a *similar* function of one or more latent variables, but it does not necessarily imply that each outcome in the cluster is a function of one scientifically meaningful latent variable. Tucker (1955) notes that if tests fall into r clusters, then it is possible to run one axis through each cluster or to "turn" the axes slightly and make them less acute so that each cluster falls in exactly one edge of the hyperplane. Tucker (1955) prefers the former solution where all outcomes load on exactly one factor over the latter where all outcomes load on all but one factor, but Yates (1987) emphasizes that the presence of *multiple* solutions implies scientifically ambiguity. The only way to resolve this ambiguity is to create a larger and more diverse battery that could disconfirm either or both configurations.

Thus, for Yates (1987), the tragedy of factor analysis is that almost all researchers *seek* a perfect cluster configuration, as if it were a good thing (see Ferguson 1954). This tendency can be seen from the fact that almost all $f(\mathbf{T})$ that are used to choose a transformation in EFA reach their theoretical optimum when every row of $\widehat{\boldsymbol{\beta}}_{EFA}$ is of unit complexity. But this tendency can also been seen in the CFA literature, where it is quite common to see a "congeneric" theory tested where every row of $\widehat{\boldsymbol{\beta}}_{CFA}$ is of unit complexity *a priori*. In both cases, this tendency can be seen in the way batteries are constructed with a view toward collecting data on many similar outcomes so that the resulting coefficient estimates will be of unit complexity. But it is quite reasonable for outcomes to be a function of more than one factor, which certainly seems to be the norm in linear regression models where all variables are observed.

Yates (1987) takes steps in the context of EFA to ensure that the primary axes to run through the edges of the configuration of outcomes in common factor space, rather than picking the lowest hanging fruit by running the axes through clusters of similar tests. In other words, Yates (1987) seeks a $\mathbf{T}$ so that some tests fall in the hyperplanes of dimensionality $r-1$ defined by the r axes, and the rest of the tests fall in the interior of this polyhedral cone in common factor space. If the hyperplanes tightly bind the test configuration in a cone, then the solution will be largely invariant to the addition or removal of interior tests, and if there are an excess of tests that fall in (or near) cone boundaries, then the solution will be largely invariant to the removal of a small number of tests from a hyperplane edge. The rest of this paper considers how best to do so in the context of SEFA, which I think is entirely consistent with the ideas in Yates (1987), although I am more sympathetic to statistical considerations.

To some extent, invariance can often be informally judged by looking at plots: Simply imagine a few conspicuous outcomes being removed from the battery and ask whether the axes would jump to another equilibrium. It is immediate that even if the analyst only *requires* $b = r$ in a SEFA model, invariance requires that the number of exact plus near zeros in each column of $\widehat{\boldsymbol{\beta}}$ exceed b. Otherwise, if one were to take away even one of the b tests in one of the hyperplanes, the algorithm would be *forced* to cut the axes through a denser region of the test configuration.

4.3. **Positive Manifolds and Positive Factor Contributions.** Thurstone (1935, 1937) coined the "positive manifold", although he considered it to be inferior theory to simple structure and not applicable outside the domain of mental abilities. Cureton and Mulaik (1971), Cureton and D'Agostino (1983), Yates (1987), and Comrey and Lee (1992) certainly consider it to be applicable to many domains that can be factor analyzed. Guttman (1992) notes that most practitioners are under the misimpression that a positive manifold refers to the off-diagonal elements of $\mathbf{S}$ being non-negative, whereas a positive manifold is defined as the situation where all cells of $\boldsymbol{\beta}$ are non-negative. Guttman (1992) also emphasizes that the positive manifold hypothesis was a proposed explanation for — not a fancy term for — the non-negativity of the off-diagonals of $\mathbf{S}$ in the selection of mental tests. If $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ are non-negative, then the off-diagonals of $\boldsymbol{\Sigma}$ will be non-negative, but the converse need not hold, particularly in a sample. It should be kept in mind that the jth manifest variable can arbitrarily be reflected by multiplying the observed scores on the jth test by -1, which merely reverses the signs of $\boldsymbol{\beta}_j$ without affecting ℓ in order to move tests into the positive manifold.

If a positive manifold exists, then the primary axes necessarily run through the edges of the test configuration in common factor space, which is why Yates (1987) is supportive of positive manifolds. The SEFA with strong simple structure above is within the positive manifold, even though it was designed to find the best solution with strong simple structure. It also illustrates Thurstone's (1947) preferred strategy of seeking simple structure and hoping to find a positive manifold in doing so, while Yates (1987) prefers to do the reverse. Both Cureton and Mulaik (1971) and Yates (1987) suggest EFA algorithms that make the occurance of a positive manifold somewhat more likely but neither algorithm is entirely satisfactory, much less automated in any software package. As discussed in Goodrich (2008), FAiR makes it possible to obtain a positive manifold via transformation in EFA if so desired.

But a positive manifold can more easily be enforced under SEFA by placing bounds on the estimated cells of $\boldsymbol{\beta}$. Strictly speaking, bounds are a feature of RGENOUD rather than SEFA per se, and thus bounds could be placed on CFA estimates of $\boldsymbol{\beta}$ as well. For example, the analyst could require that $\widetilde{\beta}_{jp} \in [-0.1, 1.2]\ \forall j, p$ to maximize ℓ subject to a "positive" manifold condition (with some allowance for sampling error around zero). Parameter bounds do not generally satisfy any aspect of corollary 1, but they do speed up the computation considerably and moreover represent a substantive belief about the model that can be tested relative to a EFA with a model comparison statistic.[10]

Yates (1987) justifies the related assumption that $\boldsymbol{\Phi}$ is non-negative by channeling Spearman and theorizing that a positive general factor lies behind the correlations among primary factors:

> The occurrence of positive correlations among primary factors and negative correlations among reference vectors in [Thurstone's] box problem actually stems from the fact that variation in the volume of boxes tends to require proportional adjustments in all three basic dimensions, i.e., there is a second order general factor here related to the fact that boxes must be functional containers rather than objects of random dimensions. That there are many possible boxes that do not occur is simply an indication that box dimensions share a common hypothetical determinant or factor such as the volume requirement that simultaneously characterizes and differentiates containers. The point of all this is to indicate that the existence of a common variance in the form of a second order or higher general factor is to be expected in any domain in which it makes sense to look for common factors at all. We should therefore always be able to reflect our first order common factors so that they are mutually positively correlated, as is the case in [Thurstone's] box problem. Otherwise, there is a possibility that the unpredictable parts of the factors (i.e. the reference vectors) share more variance in common than the primary factors per se — this would clearly suggest that we should not continue to entertain the notion of common factors as hypothetical determinants within the domain in question. (26–27)

However, Yates' (1987) hypothesis needs closer scrutiny, since there is little elaboration on the last sentence. The reference factors are an alternate representation of the factor analysis model that were mentioned above. Let $\boldsymbol{\Psi} = \mathbf{D}\boldsymbol{\Phi}^{-1}\mathbf{D}$ be the correlation matrix among reference factors, where $\mathbf{D}$ is a diagonal matrix with the reciprocal square roots of the diagonal elements of $\boldsymbol{\Phi}^{-1}$ along the diagonal of $\mathbf{D}$. The matrix of reference structure correlations, which appears in the upper right of all the plots in this paper, is defined as $\boldsymbol{\beta}\mathbf{D}$, which illustrates the claim made earlier that reference structure correlations are (column-wise) proportional to primary pattern coefficients.

[10] Bounds can also be placed on the estimated off-diagonals of $\boldsymbol{\Phi}$.

Let the second-order model be given by $\mathbf{\Phi} = \mathbf{\Delta}\mathbf{\Delta}' + \mathbf{\Omega}^2$ where $\mathbf{\Delta}$ is a $1 \times r$ matrix of second-order coefficients with $\Delta_{1p} \in [0, 1]\ \forall p$ and $\mathbf{\Omega}^2$ is a diagonal $r \times r$ matrix of uniquenesses for the primary factors. Yates (1987) implies that if these bounds are not imposed, then it is *possible* that the generalized variance (determinant) of the (correlation matrix among) reference factors could be smaller than the generalized variance (determinant) of the (correlation matrix among) primary factors, which would (somehow) imply that the common factor model is inappropriate for the sample.

However, it appears to be the case that if $\mathbf{\Phi} - \mathbf{\Omega}^2$ is of rank 1, then $|\mathbf{\Phi}| \leq |\mathbf{\Psi}|$ regardless of whether $\mathbf{\Delta}$ is non-negative, and it is possible for $|\mathbf{\Phi}| \leq |\mathbf{\Psi}|$ even when there are no second-order factors. Thus, if $|\mathbf{\Phi}| \leq |\mathbf{\Psi}|$ is "clearly" problematic (I don't see the problem), it would be logical to reinterpret Yates (1987) as recommending that the gth lexical criterion in a SEFA be operationalized as

$$z^{[g]} = \begin{cases} 1 & \text{if } |\mathbf{\Phi}| \leq |\mathbf{\Psi}| \\ 0 & \text{otherwise} \end{cases}$$

and only consider solutions that satisfy this constraint. However, counterexamples can be constructed where the factor analysis model holds but the reference vectors do have more shared variance than the primary factors (for example, the plasmode in Cattell and Dickman 1962). In limited testing, I have not found this criterion to be overwhelmingly effective but further investigation (and exposition of Yates' (1987) claim) will be needed before passing judgment.

The CFA literature has alternative justifications for including a general second order factor (whether non-negative or not), and SEFA provides a natural mechanism to test whether this specification is appropriate when $r \geq 3$. Specification of a second-order equation satisfies condition (iv) of corollary 1 and permits the analyst to specify $b = r - 1$ instead of $b \geq r$. If $r \geq 5$, then there can be two or more correlated second-order factors, but corollary 1 then needs to be applied to the second-order model as well. FAiR supports all of these options.

Yates' (1987) justification for the positive manifold assumption can also be criticized for going too far. In the most fundamental passage of the entire book, Yates (1987) states, with the emphasis in the original:

> Short of accepting the extremely chaotic situation in which manifest behavioral variables are related to their latent determinants in a completely unrestricted fashion, about the weakest simplifying assumption we can make about the action of natural causal factors is that they *may influence manifest variables either independently or in conjunction, but do not tend to act at odds to one another*. This assumption is a more fundamental expression of the principle of scientific parsimony than is the notion of simple structure, although it leads to the latter as a special case.
>
> When dealing with the problem of identifying unobserved or "latent" independent variables that can be regarded as the source of observed covariation among dependent variables we really have little choice but to assume that these independent variables do not in general produce effects that cancel

> one another out. If such cancellation of latent determinant effects were typical, then the implications of any observed set of intercorrelations among manifest variables would be ambiguous indeed. In particular, we could not take the *absence* of manifest association between any two observed behavioral variables to mean that they do not share a common cause — it could be that the lack of manifest association occurs simply because the strong mutual effects of one shared determinant are cancelled out by equally strong but contradictory effects of another shared determinant. (87–88)

This "weakest possible assumption" would nevertheless strike many behavioral researchers as too strong. There are, however, possibilities to weaken this assumption that are mentioned in Yates (1987) but are not explained so well.

One is to allow the coefficients to be slightly negative, which Yates (1987, p. 106) labels as a "diffuse" positive manifold. A second is to apply Yates' (1987) principle of non-cancellation only to the systematic variation that an outcome has with itself and not to the covariation outcomes have with each other. In jargon, one can make the assumption that none of the factors are "suppressor variables" for any of the n outcomes, but this point requires further elaboration. Let $\boldsymbol{\Pi} = \boldsymbol{\beta}\boldsymbol{\Phi}$, be the primary structure matrix, which gives the correlation between outcomes and primary factors in each of its cells when the outcomes and factors are standardized. And let $\boldsymbol{\Gamma} = \boldsymbol{\Pi} \times \boldsymbol{\beta}$, where $\times$ indicates element-by-element multiplication, be the "factor contribution matrix" (see White 1966), which indicates the proportion of variance in the jth outcome that is explained by pth factor. Thus, $h_j^2 = \sum_{p=1}^{r} \Gamma_{jp} = \boldsymbol{\beta}_j \boldsymbol{\Phi} \boldsymbol{\beta}_j'$ is the communality of the jth outcome. Although a communality cannot be negative because $\boldsymbol{\Phi}$ is positive definite, any Γ_{jp} could be negative when $r > 1$, in which case the pth factor would be a "suppressor variable" for the jth outcome. As a side note, one can define the proportional "influence" of the pth factor as $\frac{\sum_{j=1}^{n} \Gamma_{jp}}{\sum_{j=1}^{n} h_j^2}$, which is the measure appearing along the diagonal of the figures in this paper.

But Yates (1987) erroneously claims "By assuming consistency in the manner in which manifest variables relate to their hypothetical determinants, then, the positive manifold assumption is equivalent to the assertion that suppressor variable effects are rare or, at least, not typical in nature (119)." The non-negativity of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ are jointly sufficient but neither is necessary for the non-negativity of $\boldsymbol{\Gamma}$: If $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ are both non-negative, then $\boldsymbol{\Pi} = \boldsymbol{\beta}\boldsymbol{\Phi}$ is non-negative and thus $\boldsymbol{\Gamma} = \boldsymbol{\Pi} \times \boldsymbol{\beta}$ is non-negative. But it is (now) overkill for Yates (1987) to assume so when there is a weaker condition available that is both necessary and sufficient to avoid finding suppressor variables in a sample. When estimating a SEFA model, simply operationalize the gth criterion as

$$\begin{aligned} z^{[g]} &= \frac{1}{nr} \sum_{j=1}^{n} \sum_{p=1}^{r} \mathbb{I}\left\{ \widetilde{\Gamma}_{jp} \geq \underline{\Gamma} \right\}, \text{ where} \\ \mathbb{I}\left\{ \widetilde{\Gamma}_{jp} \geq \underline{\Gamma} \right\} &= \begin{cases} 1 & \text{if } \widetilde{\Gamma}_{jp} \geq \underline{\Gamma} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

and $\underline{\Gamma}$ is some analyst-specified threshold like 0 or, to make some allowance for sampling error around zero, some marginally negative number like -0.01. Thus, the optimal $\boldsymbol{\theta}$ is sought subject to the condition that $z^{[g]} = 1$, which holds when there are no suppressor variables in the sample.

Yates' (1987) principle of non-cancellation is broader. We could define "correlation contributions" as the vector comprised by $\mathbf{\Pi}_j \times \boldsymbol{\beta}_{j'}$, which are "factor contributions" in the special case where $j = j'$. Yates (1987) thus contends that all correlation contributions between j and j' should have the same sign to prevent the complicated scenario where $\sum_{p=1}^{r} \Pi_{jp} \times \beta_{j'p} = 0$ but outcome j and outcome j' have offsetting factors in common. This "weakest possible assumption" may very well be plausible in some domains, but it clearly can be weakened further by requiring it only when $j = j'$.

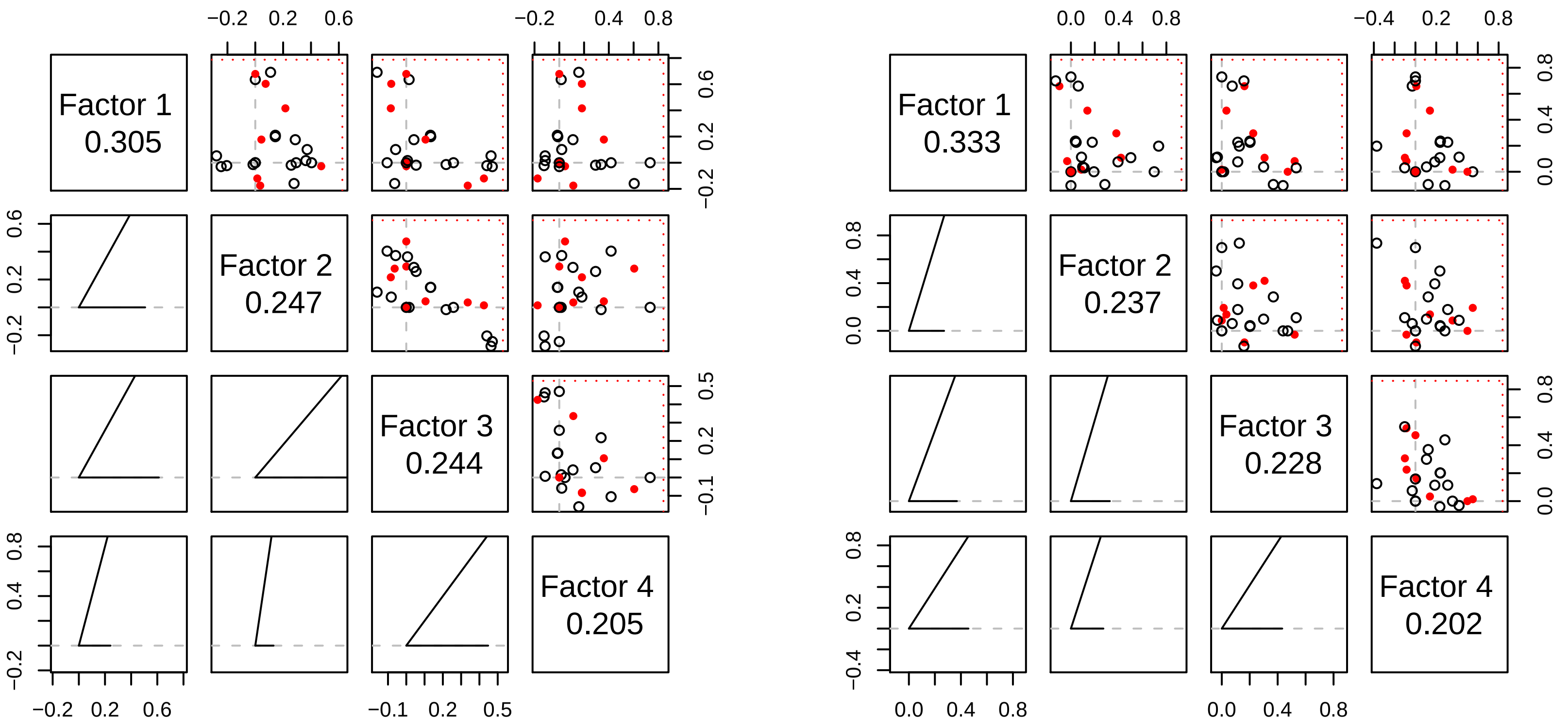
Recall that in figure 1, the solution with the strong simple structure restriction also exhibited a positive manifold. However, one of the correlations between factors was slightly negative, which runs contrary to Yates (1987) rule that the intercorrelations among primary factors should be non-negative. For the Holzinger and Swineford (1939) data, if Yates' (1987) recommendation is faithfully followed by constraining the first-order coefficients to be in the diffuse positive manifold and that the second-order general factor is non-negative, a corner solution (not shown) results where the factors must be orthogonal in order to fit the data reasonably well. Perhaps for this reason, Yates (1987) favored three-factor solutions for mental test batteries.

The left side of figure 3 shows a solution where Yates' (1987) recommendation is relaxed and only requires that $\left|\widehat{\boldsymbol{\Phi}}\right| \leq \left|\widehat{\boldsymbol{\Psi}}\right|$. The plots are not bad, although the configuration does not exhibit as much order as some of the solutions that follow. It appears as if the third factor is inclined too accutely to the other factors by perhaps a few degrees. Although this distortion would probably not be big enough to drastically affect the interpretation of the factors, it is problematic and can usually be attributed to strong pairwise linkages between tests that fall in certain hyperplanes.

The solution on the right side of figure 3 imposes the restriction that prohibits suppressor variables by requiring factor contributions to exceed -0.03. It looks similar to the previous plot, but the inter-factor correlations are somewhat more moderate while still satisfying the (not required) condition that $\left|\widehat{\boldsymbol{\Phi}}\right| \leq \left|\widehat{\boldsymbol{\Psi}}\right|$. There is more order to this configuration, but it looks like the fourth factor is somewhat misaligned and leaves one major negative outlier, presumably due to the cluster of tests that loads highly on factor 1 but not at all on factor 4. As it turns out, this outlier is the test for addition of numbers, which is too fundamental to the study of mental abilities to simply be dismissed. In both solutions in figure 3, $\widehat{\ell}$ differs only trivially from $\widehat{\ell}_{EFA}$.

Many contend that the positive manifold assumption is reasonable for mental test data but too strong for behavioral data. The assumption that none of the factors are suppressor variables for any of the n outcomes is a weaker assumption than the positive manifold / positive intercorrelations assumption, and hopefully it is sufficiently weak to be applicable in the behavioral domains. In other words, it is possible for some, and perhaps several, cells of $\widehat{\boldsymbol{\beta}}$ to be somewhat

FIGURE 3. More SEFA Solutions for the Holzinger and Swineford (1939) Data



The plot on the left is a SEFA that requires $|\mathbf{\Phi}| \leq |\mathbf{\Psi}|$. The plot on the right is a SEFA that prohibits solutions with any factor contributions less than -0.03. The log-likelihoods for the two solutions are -1027.742 and -1027.783 respectively, which are essentially equal to the EFA log-likelihood with four factors. Both solutions seem plausible but unfortunately leave a small number of outliers outside the bulk of the configuration.

negative provided that the off-diagonals of $\widehat{\boldsymbol{\Phi}}$ are not too large. In the extreme case where $\widehat{\boldsymbol{\Phi}} = \mathbf{I}$, then $\widehat{\boldsymbol{\Pi}} = \widehat{\boldsymbol{\beta}}$ and $\widehat{\Gamma}_{jp} = \widehat{\beta}^2_{jp}$, in which case there is no limit on how "bipolar" the factors can be because the factor contributions are all non-negative. Generally speaking, the size of the off-diagonals in $\widehat{\boldsymbol{\Phi}}$ will be largest in a vanilla SEFA, smallest when a positive manifold is also required, and middling when suppressor variables are prohibited.

4.4. **Invariance, Cohyperplanarity, and Collinearity.** Yates (1987) was influenced by Butler (1969), which argues that researchers should choose $\mathbf{T}$ to make the factors invariant to reasonable changes in the battery or the sample. Butler (1969) further argues factors are likely to be invariant when the off-diagonal elements of $\boldsymbol{\Phi}$ are small in magnitude and proposes choosing $\mathbf{T}$ by running the r primary axes directly through the r tests in $\widehat{\boldsymbol{\Lambda}}$ that are "distinguished", i.e. most orthogonal to each other. This simple method of choosing $\mathbf{T}$ thereby creates one test of unit complexity per factor, which is criticized in Cureton and Mulaik (1971) for presuming that r such tests exist in the battery.

One could respond to Cureton and Mulaik's (1971) criticism by estimating a SEFA model with a mapping rule that satisfies condition (ii) of corollary 1 by allocating $r\left(r-1\right)$ zeros among the set of r tests that are least correlated in the common factor space and then evaluating whether these restrictions are plausible at the optimum. Such a mapping rule can be implemented $\forall q \neq p$

$$\widetilde{\beta}_{jp} = \begin{cases} 0 & \text{if}\, abs\left(\check{\beta}_{jq}\right) = \max\left\{abs\left(\check{\boldsymbol{\beta}}_q\right)\right\} \\ \check{\beta}_{jp} & \text{otherwise.} \end{cases}$$

In other words, if $\check{\beta}_{jq}$ is the largest preliminary coefficient for the qth factor, squash all cells in the jth row of $\widetilde{\boldsymbol{\beta}}$ to zero except $\widetilde{\beta}_{jq}$. There is a deeper justification for this mapping rule than it might appear at first sight. Let $\overset{0}{\widetilde{\boldsymbol{\beta}}}$ be the $r \times r$ submatrix of $\widetilde{\boldsymbol{\beta}}$ such that each row of $\overset{0}{\widetilde{\boldsymbol{\beta}}}$ is of complexity 1. The determinant of $\overset{0}{\widetilde{\boldsymbol{\beta}}}$ is simply the product of the non-zero cells, and $\left|\overset{0}{\widetilde{\boldsymbol{\beta}}}\right|$ is thus the largest determinant that could have been formed under this mapping rule. Since $\left|\overset{0}{\widetilde{\boldsymbol{\beta}}}\boldsymbol{\Phi}\overset{0}{\widetilde{\boldsymbol{\beta}}'}\right| = \left|\overset{0}{\widetilde{\boldsymbol{\beta}}}\right|^2 \left|\boldsymbol{\Phi}\right|$, $\left|\overset{0}{\widetilde{\boldsymbol{\beta}}}\boldsymbol{\Phi}\overset{0}{\widetilde{\boldsymbol{\beta}}'}\right|$ is largest than any other reduced correlation matrix among r outcomes in common factor space, which is consistent with the definition Butler (1969) gives for the most distinguished set of r vectors in common factor space.

Yates (1987) is also wary of presuming that unifactorial tests exist in a battery but wholeheartedly adopts Butler's (1969) goal of pursuing invariance. However, Yates (1987) differs from Butler (1969) by arguing that invariance is probable when the tests within a hyperplane are highly distinguished from each other; i.e. widely dispersed in $r-1$ directions throughout the hyperplane (when $r > 2$). In other words, there is cohyperplanarity but not collinearity among tests in a hyperplane. Literally implementing a mapping rule where the zeros are chosen to minimize the correlation among tests in a hyperplane is currently impractical from a computational perspective, but there are a couple of options for cheap approximations.

One is to operationalize the gth criterion as

$$z^{[g]} \quad = \quad \frac{1}{r}\sum_{p=1}^{r} \mathbb{I}\left\{ \left| \overset{p}{\widetilde{\boldsymbol{\beta}}}\widetilde{\boldsymbol{\Phi}}\overset{p}{\widetilde{\boldsymbol{\beta}}}' + \overset{p}{\widetilde{\boldsymbol{\Theta}^2}} \right|^{\frac{1}{b}} \geq \left|\widetilde{\mathbf{C}}\right|^{\frac{1}{n}} \right\},$$

where $\overset{p}{\widetilde{\boldsymbol{\beta}}}$ is again the submatrix of $\widetilde{\boldsymbol{\beta}}$ with exact zeros in the pth column, $\overset{p}{\widetilde{\boldsymbol{\Theta}^2}}$ is the diagonal submatrix of $\widetilde{\boldsymbol{\Theta}^2}$ containing the uniquenesses of the tests with exact zeros in the pth column of $\widetilde{\boldsymbol{\beta}}$. A determinant of a matrix raised to the power of the reciprocal of its order is called the "effective variance" by Peña and Rodriguez (2003), which shows that — in contrast to the generalized variance (determinant) — the effective variance can be meaningfully compared across correlation matrices of different orders. Thus, this criterion essentially requires that, for each hyperplane, the tests in that hyperplane must be "less correlated per variable" than is the battery as a whole in common factor space.

Another approach is to use a mapping rule to encourage cohyperplanarity. Let $\boldsymbol{\theta}$ fill $\widetilde{\boldsymbol{\Phi}}$ and $\check{\boldsymbol{\beta}}$, which is again a preliminary pattern matrix that can be used to construct a preliminary factor contribution matrix, $\check{\boldsymbol{\Gamma}}$. Then, $\forall q \neq p$,

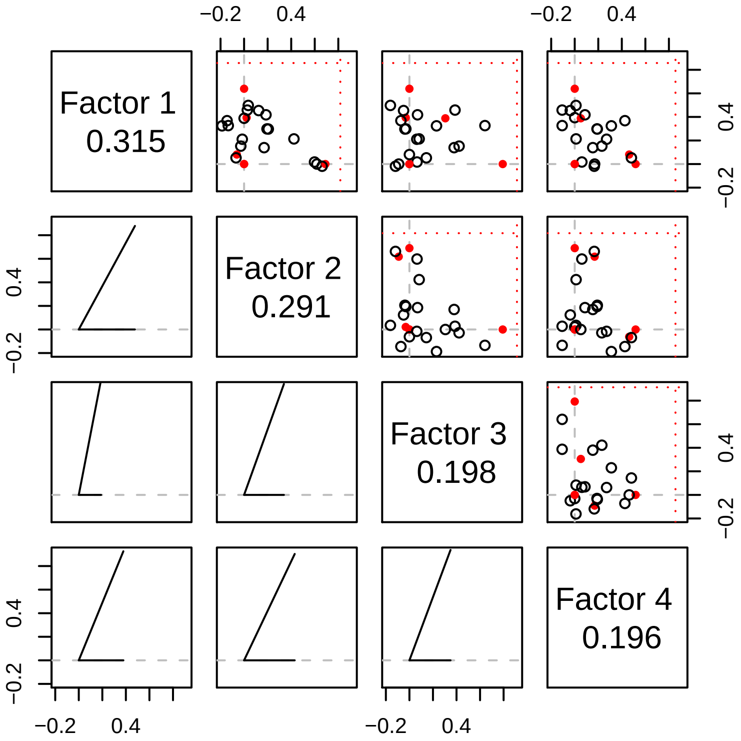
$$\widetilde{\beta}_{jp} \quad = \quad \begin{cases} 0 & \text{if } \check{\Gamma}_{jq} - \check{\Gamma}_{jp} = \max\left\{\check{\boldsymbol{\Gamma}}_q - \check{\boldsymbol{\Gamma}}_p\right\} \\ \check{\beta}_{jp} & \text{otherwise.} \end{cases}$$

In essence, this mapping rule forces $\widetilde{\beta}_{jp}$ to be exactly zero when $\check{\Gamma}_{jq}$ is large, which is to say that the qth factor is quite influential when $\widetilde{\boldsymbol{\beta}}_j$ is in the pth hyperplane. This mapping rule encourages the $r - 1$ tests within the pth hyperplane to be widely dispersed in $r - 1$ directions and tends to encourage tests in hyperplanes to have large communalities.

Figure 4 depicts a solution that uses this mapping rule and imposes the constraint that the tests in each hyperplane must have more effective variance than the battery as a whole. The solution fits well in the sense that $\widehat{\ell}$ is virtually the same as $\widehat{\ell}_{EFA}$. Although a few squares seem imperfect, for each factor, two out of three squares appear quite good, which suggests that it would be difficult to further improve the configuration. In particular, note that every axis runs through a point far from the origin, which is encouraged but not strictly implied by the cohyperplanarity mapping rule.

These high communality tests are denoted by filled, red dots, which indicates that they also fall in other hyperplanes. In fact, although it was *not* required, each factor is represented by one test of unit complexity: Visual Perception (test #1), Sentence Completion (#7), Addition (#10), and Word Recognition (#14). Although this occurence does make interpretation easy, the solution cannot be invariant to the removal of any of these tests (especially #10). On the other hand, the solution is extremely invariant to the removal of any of the other twenty tests or the addition of new tests. If the mapping rule inspired by Butler (1969) is used, the configuration stays the same except that the third axis cuts slightly inside test #10 and through test #12 (Counting Dots).

FIGURE 4. Cohyperplanar SEFA for the Holzinger and Swineford (1939) Data



This SEFA requires that tests within each hyperplane have more effective variance than the configuration as a whole and uses a mapping rule based on differences in factor contributions to encourage cohyperplanarity without collinearity among the tests within a hyperplane. The log-likelihood is -1027.807, which is indistinguishable from that of the EFA with four factors.

The solution also happens to closely resemble the result one would obtain if the quartimin algorithm were used to transform the preliminary EFA solution. From Yates' (1987) anti-quartimin perspective, this similarity would represent a stroke of luck on quartimin's part. Although the coefficients do not fall into a perfect cluster configuration, by optimizing with respect to the quartimin criterion — whose theoretical optimum corresponds to a perfect cluster configuration — and failing to make much headway in that direction, the result is a configuration characterized by cohyperplanarity instead of "progressing" all the way to within-cluster collinearity in common factor space. In contrast, this SEFA model achieved exactly what it set out to do with its mapping rule — namely to place at least four outcomes in each hyperplane edge that are widely dispersed in three directions — a goal that has a deeper, Butler / Yates justification compared to merely seeking an easy-to-interpret, perfect cluster configuration.

Despite the fact that no one has stepped forward to defend quartimin and perfect cluster configurations against the criticisms made by Butler (1969) and Yates (1987) decades ago, FAiR includes a mapping rule that estimates a SEFA model under the presumption that $\boldsymbol{\beta}$ has a perfect cluster configuration:

$$\widetilde{\beta}_{jp} = \begin{cases} \check{\beta}_{jp} & \text{if } abs\left(\check{\beta}_{jp}\right) = \max\left\{abs\left(\check{\boldsymbol{\beta}}_j\right)\right\}, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, all coefficients are squashed to zero unless they are the largest (in magnitude) among those slated for the jth row of $\widetilde{\boldsymbol{\beta}}$. The question is whether the likelihood at the constrained optimum is sufficiently close to $\widehat{\ell}_{EFA}$. My guess is that it typically is not, but this fact can be ascertained in any given situation by estimating both models and reporting the model fit statistics. FAiR actually generalizes this mapping rule and the strong simple structure mapping rule to allow the user to specify the maximum factor complexity for each outcome or for all outcomes.

Table **??** shows the results of three SEFA models with maximum test complexity of 3, 2, and 1 respectively. The solution with complexity 3 is the strong simple structure solution shown in figure 1, which was thought to be good although perhaps not as good as the solution that encouraged cohyperplanarity. Even if one were not persuaded by Yates' (1987) theoretical arguments that perfect cluster configurations should not be sought in a scientific investigation, we can say that the second and third solutions should be rejected on the basis of a likelihood-ratio test relative to the EFA, whereas we cannot reject the first solution with strong simple structure or any of the previous SEFA models shown in this paper. Perhaps if one were interested in constructing an in-sample scale, then estimating a SEFA model with perfect clustering would be useful, albeit not useful for estimating population parameters.

When the data-generating process actually does involve a $\boldsymbol{\beta}$ with a perfect cluster configuration, it is difficult to find the global optimum with a SEFA model. In part, it may be that FAiR is insufficiently optimized for such situations, which I believe to be rare in actual research. In part, there are a lot of potentially free parameters but very little "information" because r clusters of tests are perfectly collinear in the common factor space. And part of the the problem is that ℓ is highly multimodal with modes that are widely separated in $\boldsymbol{\theta}$-space, as noted earlier in the earlier disussion of the differences in the reactions of Tucker (1955) and Yates (1987) to this fact. In theory, RGENOUD can handle any degree of multimodality as $t, K \rightarrow \infty$, but in practice FAiR seems to consistently handle intense multimodality only when the best admissable modes are within shouting distance of each other in $\boldsymbol{\theta}$-space.

5. Conclusions

This paper has introduced SEFA as a method to estimate a factor analysis model. By utilizing RGENOUD, FAiR gives a reasonable (but not guaranteed) assurance that the maximum of the constrained log-likelihood function will be found in finite time. The question becomes — but really always has been — "What constraints should a good factor analysis model meet?" or "How do you know a good estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Theta}^2$ when you see them?" There are surprisingly few papers and books in the history of multiple factor analysis that have addressed this question directly.

For Thurstone (1935), if $\widehat{\boldsymbol{\beta}}$ exhibits simple structure, then it is probably a good estimate of $\boldsymbol{\beta}$. For Ferguson (1954) and Tucker (1955), if $\widehat{\boldsymbol{\beta}}$ is sparse, then it is probably a good estimate of $\boldsymbol{\beta}$. For Butler (1969), if $\widehat{\boldsymbol{\Phi}} \approx \mathbf{I}$, then $\widehat{\boldsymbol{\beta}}$ is probably a good estimate of $\boldsymbol{\beta}$. For Yates (1987), if $\widehat{\boldsymbol{\beta}}$ is within the positive manifold and $\widehat{\boldsymbol{\Phi}}$ is non-negative, then they are probably good estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$. CFA proponents answer this question somewhat tautologically by saying that $\widehat{\boldsymbol{\beta}}$ is a good estimate of $\boldsymbol{\beta}$ when some cells of $\widehat{\boldsymbol{\beta}}$ are fixed to the corresponding values in $\boldsymbol{\beta}$ in accordance with substantive theory. But today many EFA practitioners and methodologists duck this challenging but standard statistical question by saying something to the effect of "If the primary pattern is easy to interpret (read sparse), then it is good" without specifically arguing that it is a good *estimate* of $\boldsymbol{\beta}$ in the population, which requires stronger assumptions.

In this paper, I added the contention that if $b = r$, $\widehat{\boldsymbol{\Gamma}}$ is non-negative, and $\widehat{\ell}_{SEFA} \approx \widehat{\ell}_{EFA}$, then $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Phi}}$ may be good estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ in many circumstances. The best result for the Holzinger and Swineford (1939) data was obtained with a mapping rule that encouraged cohyperplanarity through differences in factor contributions, although it would have also been possible to prohibit suppressor variables from this solution as well. Decent resuls were obtained simply by requiring $\left|\widehat{\boldsymbol{\Phi}}\right| \leq \left|\widehat{\boldsymbol{\Psi}}\right|$, which probably could have been improved upon by adding the restriction on $\widehat{\boldsymbol{\Gamma}}$.

The reasons to use, or at least try SEFA, over EFA and CFA are relatively clear. SEFA is like EFA in that neither requires a complete substantive theory to be fruitful. Thus, SEFA is appropriate for a pilot study and so forth. But unlike EFA, SEFA does not require the analyst to choose $\mathbf{T}$, which is its chief virtue. Because the final results are driven entirely by ℓ and the restrictions, SEFA is on firm statistical ground, and many statisticians have previously objected to EFA because the choice of $\mathbf{T}$ is astatistical. Yates (1987) goes farther by arguing that, in practice, the choice of $\mathbf{T}$ is ascientific and designed to provide in-sample descriptions of clusters rather than to make inferences about the effects of latent variables in the population. With appropriate restrictions, a SEFA can accomplish Yates' (1987) ultimate goal of locating hyperplanes at or near the edges of the test configuration in common factor space.

In a sense, SEFA replaces the transformation problem in EFA with the problem of choosing a solution that is within some tolerance of $\widehat{\ell}_{EFA}$. It may be possible to reject some choices of restrictions by comparing log-likelihoods at the optimum, as was the case using the Holzinger and Swineford (1939) data for the last two solutions in table **??**. But lots of schemes for placing weaker restrictions on a SEFA model would have produced a $\widehat{\ell}_{SEFA}$ that is statistically insignificantly different from (or otherwise close to) $\widehat{\ell}_{EFA}$. Thus, as Yates (1987) stated in no uncertain terms, an additional principle — beyond simple structure and certainly beyond perfect clustering — is typically needed to obtain scientifically meaningful results in non-trivial, real world problems. For Yates (1987), this additional principle is non-cancellation but there are some other possibilities like $|\boldsymbol{\Phi}| \leq |\boldsymbol{\Psi}|$ and the idea of distinguishability in Butler (1969) that need further investigation. FA*i*R could quickly operationalize other principles once proposed.

I like the SEFA's reformulation of the EFA transformation problem because the restrictions imposed on the SEFA model must be made explicit at the outset. And everyone has at least an intuitive understanding of (constrained) discrepancy functions measuring the misfit between $\mathbf{S}$ (data) and $\widehat{\mathbf{C}}$ (model), which ultimately drives the results. In contrast, the identification condition used to select $\mathbf{T}$ and thus $\widehat{\boldsymbol{\beta}}_{EFA}$ is more opaque because $f\left(\mathbf{T}\right)$ drives the results, about which all we know are the conditions under which it reaches its theoretical optimum. At best, for most transformation algorithms $f\left(\mathbf{T}\right)$ tells us something about the misfit between $\widehat{\boldsymbol{\beta}}_{EFA}$ and a hypothetical primary pattern matrix with perfect clustering, but tell us nothing intrinsically about the misfit between $\widehat{\boldsymbol{\beta}}_{EFA}$ and $\boldsymbol{\beta}$.

SEFA is also appropriate when the researcher has a well-specified substantive theory to test. Many will undoubtedly contend that "CFA is grounded in theory" while "SEFA is atheoretical" but this view misunderstands SEFA. Choosing what restrictions to place on a SEFA requires careful consideration of the problem at hand. For example, Thurstone originally proposed the positive manifold as a theoretically motivated hypothesis to explain variation in mental test scores, and FAiR gives researchers the option of estimating a SEFA model subject to the positive manifold restriction to see if this hypothesis is plausible in light of the data. In short, $\beta_{jp} \geq 0\,\forall j, p$ is no less a theory than $\beta_{12} = 0$, but the former is harder to test with well-known CFA software and is thus liable to be considered less theoretical.

Moreover, the purpose of CFA is to test competing theories, but SEFA can be conceptualized as maximizing the likelihood over all CFA models that require at least as many zeros per factor. Thus, if the researcher has a well-specified substantive theory about specific cells of $\boldsymbol{\beta}$, SEFA provides a *stronger* test of that theory than does CFA because SEFA exposes the theory to an incomprehendable number of competing hypotheses that could be backed out of the likelihood function. CFA only exposes the theory in question to at most a handful of other theories that have been articulated well enough to test with a different CFA model. And it is important to keep in mind the possibility — that is already implemented in FAiR but not discussed much here — that most restrictions that can be imposed on a CFA model can also be imposed on a (mixed) SEFA model, and vice versa with the exception of mapping rules.

The weaknesses of SEFA seem small in number and importance. SEFA sacrifices some model fit relative to EFA and to CFA models that minimally satisfy theorem 1. SEFA standard errors do not reflect the locational uncertainty in the exact zeros. Perhaps the biggest concern is that FAiR may bypass the "True" answer in a finite sample when it is a local optimum that is exceeded by another optimum due to sampling variability. Thus, it is fair to say that the research design requirements are generally stronger for a SEFA model than for a pure CFA model. However, it may very well be the case that if someone were to seek a grant to conduct a preliminary EFA followed by a CFA, the money could be better spent on a single SEFA with a larger sample and / or battery.

There are, however, a lot of practical challenges for SEFA to overcome. Many practitioners are familiar with EFA and / or CFA and will undoubtedly be reluctant to change to a new estimator in a new software environment. SEFA can take an hour or more to finish, depending mostly on n and r, and still is not guaranteed to find the optimum in finite time. There are a lot of tuning parameters for RGENOUD, and we do not yet know what combinations work best with FAiR. FAiR is considerably less mature than other packages for factor analysis that have been in development for a decade or more and currently lacks mechanisms to make the estimates "robust" to departures from the traditional factor analysis assumptions. These are largely shortcomings of the computer code underlying FAiR rather than theoretical problems with the SEFA model that can be probably be overcome with continued effort. Most of all, we do not yet know what combinations of restrictions are best to impose on the factor analysis model, despite Yates' (1987) progress on this front. That is the main theoretical question for future research.

REFERENCES

Bernaards, Coen A. and Robert I. Jennrich. 2005. "Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis." *Educational and Psychological Measurement* 65:676–696.

Bollen, K.A. 1989. *Structural equations with latent variables*. Wiley New York.

Bollen, K.A. and K.G. Jöreskog. 1985. "Uniqueness does not Imply Identification: A Note on Confirmatory Factor Analysis." *Sociological Methods & Research* 14(2):155.

Browne, M.W. 2001. "An Overview of Analytic Rotation in Exploratory Factor Analysis." *Multivariate Behavioral Research* 36(1):111–150.

Butler, J.M. 1969. "Simple Structure Reconsidered: Distinguishability and Invariance in Factor Analysis." *Multivariate Behavioral Research* 4(1):5–28.

Cattell, R.B. and K. Dickman. 1962. "A dynamic model of physical influences demonstrating the necessity of oblique simple structure." *Psychological Bulletin* 59:389–400.

Comrey, A.L. and H.B. Lee. 1992. *A First Course in Factor Analysis*. Lawrence Erlbaum Associates.

Cureton, E.E. and R.B. D'Agostino. 1983. *Factor analysis: an applied approach*. L. Erlbaum Associates.

Cureton, E.E. and S.A. Mulaik. 1971. "On Simple Structure and the Solution to Thurstone's Invariant Box Problem." *Multivariate Behavioral Research* 6:375–387.

de Leeuw, Jan and Patrick Mair. 2007. "An Introduction to the Special Volume on "Psychometrics in R"." *Journal of Statistical Software* 20(1):1–5.
URL: *http://www.jstatsoft.org/v20/i01*

Falissard, Bruno. 2007. *psy: Various procedures used in psychometry*. R package version 0.7.

Ferguson, G.A. 1954. "The concept of parsimony in factor analysis." *Psychometrika* 19(4):281–290.

Fox, John. 2006. "Sturctural Equation Modeling with the sem Package in R." *Sturctural Equation Modeling* 13(3):465–486.

Guttman, L. 1992. "The irrelevance of factor analysis for the study of group differences." *Multivariate Behavioral Research* 27(1):75–204.

Harman, H.H. 1976. *Modern Factor Analysis*. University Of Chicago Press.

Holzinger, K.J. and F. Swineford. 1939. *A study in factor analysis*. University of Chicago Chicago, Ill.

Howe, W.G. 1955. Some Contributions to Factor Analysis. Technical report ORNL-1919, Oak Ridge National Lab., Tenn.

Husson, Francois, Julie Josse, Sebastien Le and Jeremy Mazet. 2007. *FactoMineR: Factor Analysis and Data Mining with R*. R package version 1.07.

URL: *http://factominer.free.fr, http://www.agrocampus-rennes.fr/math/*

Kaiser, H.F. 1958. "The varimax criterion for analytic rotation in factor analysis." *Psychometrika* 23(3):187–200.

Kiers, H.A.L. 1994. "Simplimax: Oblique rotation to an optimal target with simple structure." *Psychometrika* 59(4):567–579.

Lorenzo-Seva, U. 2003. "A factor simplicity index." *Psychometrika* 68(1):49–60.

MacCallum, R.C. 1989. "Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis." *Journal of the American Statistical Association* 84(406).

Mair, Patrick and Reinhold Hatzinger. 2007. "Psychometrics Task View." *R News* 3(3):38–40.

URL: *http://CRAN.R-project.org/doc/Rnews/*

Martin, Andrew D., Kevin M. Quinn and Jong Hee Park. 2007. *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. R package version 0.9-2.

URL: *http://mcmcpack.wustl.edu*

Mebane, Walter R. Jr. and Jasjeet Singh Sekhon. N.d. "Genetic Optimization Using Derivatives: The rgenoud package for R." *Journal of Statistical Software*. Forthcoming.

URL: *http://sekhon.berkeley.edu/papers/rgenoudJSS.pdf*

Millsap, R.E. 2001. "When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis." *Structural Equation Modeling* 8(1):1–17.

Peña, D. and J. Rodriguez. 2003. "Descriptive measures of multivariate scatter and linear dependence." *Journal of Multivariate Analysis* 85(2):361–374.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

URL: *http://www.R-project.org*

Raiche, Gilles. 2007. *nFactors: Non Graphical Solution to the Cattell Scree Test*. R package version 2.1.

Revelle, William. 2007. *psych: Procedures for Personality and Psychological Research*. R package version 1.0-27.

URL: *http://personality-project.org/r, http://personality-project.org/r/psych.manual.pdf*

Thurstone, L.L. 1935. *The vectors of mind: multiple-factor analysis for the isolation of primary traits*. University of Chicago Press.

Thurstone, L.L. 1947. *Multiple Factor Analysis: A Development and Expansion of The Vectors of the Mind*. University of Chicago Press.

Tucker, L.R. 1955. "The objective definition of simple structure in linear factor analysis." *Psychometrika* 20(3):209–225.

White, O. 1966. "Some Properties of Three Factor Contribution Matrices." *Multivariate Behavioral Research* 1(3):373–378.

Yao, X., Y. Liu and G. Lin. 1999. "Evolutionary programming made faster." *Evolutionary Computation, IEEE Transactions on* 3(2):82–102.

Yates, A. 1987. *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis*. State University of New York Press.