June 2010

Factor Analysis via Components Analysis

Peter M. Bentler and Jan de Leeuw

University of California, Los Angeles

Research supported by grants 5K05DA000017-32 and 5P01DA001070-35 from the National Institute on Drug Abuse to P. M. Bentler and grant 4R44CA137841-03 from the National Cancer Institute to P. Mair. Bentler acknowledges a financial interest in EQS and its distributor, Multivariate Software.

Correspondence concerning this article should be addressed to Peter M. Bentler, Departments of Psychology and Statistics UCLA, Box 951563, Los Angeles, CA 90095-1563. E-mail: bentler@ucla.edu

Factor Analysis via Components Analysis

Abstract

Under the null hypothesis, component loadings are linear combinations of factor loadings, and vice versa. This interrelation permits defining new optimization criteria and estimation methods for exploratory factor analysis. Although this note is primarily conceptual in nature, an illustrative example shows the methodology to be promising.

Factor Analysis Via Components

Although limiting conditions have been developed under which components, or principal components as an important special case, and latent factors of factor analysis coincide (e.g., Guttman, 1956; Bentler & Kano, 1990), in most treatments components analysis and factor analysis are considered to be alternative but basically unrelated methods for determining sources of variance in variables (e.g., Mulaik, 2009). The goal of this paper is to estimate the parameters of the common factor model via components. To do this, we require a stronger linking between the models than has been previously described. To begin, we provide some matrix background on the existence of a factor analysis model.

Suppose C is a symmetric positive semidefinite matrix of order p and rank r with $C = U\Delta^2 U'$, where U is a $p \times r$ orthonormal matrix of eigenvectors with $U'U = I$, and Δ^2 is an $r \times r$ diagonal matrix of eigenvalues. Then we have

<u>Result 1</u>. A $p \times q$ matrix X satisfies $C = XX'$ if and only if $X = U\Delta V'$, where V is $q \times r$ and satisfies $V'V = I$. Thus q cannot be smaller than r, and the rank of X is equal to r.

Postmultiplying $X = U\Delta V'$ by its transpose gives C, which is of rank r, establishing that $q \geq r$ and the rank of X. Assuming to the contrary that $V'V \neq I$ violates the assumption that $C = U\Delta^2 U'$, thus leading to a contradiction. Note that $X = U\Delta V'$ gives its singular value decomposition.

Now let us take $C = \Sigma$, a population covariance matrix with eigenvector decomposition $\Sigma = U\Delta^2 U'$. The orthogonal factor analysis model states that the covariance matrix has decomposition $\Sigma = \Lambda\Lambda'$ where $\Lambda = [\lambda \,|\, \psi]$ is a $p \times q (= k + p)$ partitioned matrix of factor loadings that contains a $p \times k$ common factor loading matrix λ and the $p \times p$ diagonal unique loading matrix ψ.[1] We apply Result 1 to the factor model by letting $X = \Lambda$, obtaining

[1] Note that our lower case λ is the loading matrix for the common factors that is more typically given as cap Λ.

Result 2. The orthogonal factor analysis model $\Sigma = \Lambda\Lambda'$ is true if and only if there exists $\Lambda = [\lambda \mid \psi]$ with λ, diagonal ψ, and orthonormal matrix V such that $\Lambda = U\Delta V'$.

It follows that $\Lambda V_{\perp} = 0$, where $V_{\perp}$ is a $q \times (q-p)$ orthogonal complement such that $V_{\perp}{}'V = 0$. In the following we assume that the factor analysis model holds in the population.

Interrelations between Models

Under a component model, a random p-variate vector of observed variables x may be expressed as a population model based on a linear combination of underlying components ζ with coefficients L

$$x = L\zeta \,. \tag{1}$$

L is typically called a component loading matrix, and is sometimes misleadingly called a factor loading matrix. We take L to be square and full rank, so that this is a complete components representation. It is not necessarily unique, since (1) allows a rotation. One way to make it unique is to specify that $L'L$ is diagonal with elements ordered from large to small, i.e., a principal components representation. In practice, when the decomposition (1) is applied with sample data, it is used as an approximation, and the number of columns of L is taken to be substantially below p.

Under a common factor analysis model, the same p-variate vector of observed variables x is given a different population decomposition. In particular, using the factor notation above,

$$x = \Lambda\xi = [\lambda \mid \psi]\xi \tag{2}$$

where the q random variables ξ are factors consisting of k common factors and p unique factors. Now, accepting Result 2, we assume that both models (1) and (2) are true in the population. In this case, we may write $L\zeta = \Lambda\xi$, and, since in the complete components loading matrix L is invertible, we have

$$\zeta = L^{-1}\Lambda\xi . \tag{3}$$

That is, we have the obvious result

Result 3. Components ζ are linear combinations of factors ξ.

In particular, components are combinations of common and unique factors. Since unique factors contain specificity plus random error, components also contain these sources of variance.

Next we consider the covariance structures of these two models. Under the usual assumptions that $E(x)=0, E(\zeta)=0, E(\xi)=0$, the components ζ are mutually uncorrelated, and, in exploratory factor analysis, the factors ξ are mutually uncorrelated, when both models are true we have

$$\Sigma = LL' = \Lambda\Lambda' . \tag{4}$$

Now we perform the singular value decomposition given in Result 2, namely, $\Lambda = U\Delta V'$, where U is $p \times p$ with $U'U = I$, Δ is a $p \times p$ diagonal matrix of singular values, and V' is $p \times q$ with $V'V = I$. Furthermore, since $\Lambda VV' = \Lambda$, we may write $LL' = \Lambda VV'\Lambda'$. Since we have specified no special structure for L, and since ΛV is of dimension $p \times p$, this allows us to take

$$L = \Lambda V . \tag{5}$$

Thus we have

Result 4. The component loading matrix L is a rank-reducing linear combination of elements of the factor loading matrix Λ.

It follows trivially from the singular value decomposition of Λ and (5) that

$$\Lambda = LV' . \tag{6}$$

Hence we have

Result 5. The factor loading matrix Λ is a rank-increasing linear combination of elements of the component loading matrix L.

Suppose D_p is $p \times p$ a nonsingular diagonal matrix, and that the observed variables are scaled by this matrix. The effect on the component loadings is to yield $L_D = D_p L$, and on the factor loadings is to yield $\Lambda_D = D_p \Lambda$. Clearly, these rescaled loading matrices maintain the relations given by (5) and (6), and hence we may without loss of generality perform all analyses on the correlation matrix. Thus we have

<u>Result 6.</u> The relations between component and factor loadings given by (5) and (6) are invariant to rescaling of the observed variables.

Earlier we stated that the choice of component representation in the above relations is arbitrary. It may be worthwhile to be explicit why this is so. Consider transforming the component loading matrix by an orthonormal matrix. If T is a matrix such that $T'T = TT' = I$, and the left and right hand sides of (5) are postmultiplied by T, the relation (5) is maintained for the new loading matrix $L_T = LT$ and the new $V_T = VT$, where V_T possesses the same orthogonality properties as the original V. Hence

<u>Result 7.</u> The choice of components, such as principal components, is arbitrary. Any convenient component representation maintains the key relations between components and factors.

It may be useful to give the explicit interrelations between components and factors at the level of population covariances. We postmultiply (3) by ξ' and take expectations of both sides, yielding

$$\Phi_{\zeta\xi} = E(\zeta\xi') = L^{-1}\Lambda . \tag{7}$$

If the starting point of our analysis had allowed the factors to be correlated with covariance matrix Φ, this covariance matrix would show up on the right side of (7). In any case, the interpretation of elements of $\Phi_{\zeta\xi}$ will hinge critically on the chosen scaling of variables, the identification conditions utilized and possible rotations imposed on the components as well as the factors. Thus we have

Result 8. The covariance matrix relating components and factors, or their correlation matrix when standardized, is given by $L^{-1}\Lambda$.

The above equations are population relations that hold if the hypothesized factor model is true. Otherwise, they will only be approximations, and the quality of the approximation will depend on the correctness of (2), as well as a correct choice of the number of factors. For example, if (2) holds with k common factors, a factor model with $k-1$ factors would not be consistent with (5)-(6) in either the population or in a sample. Equations (5)-(6) would be approximations rather than equalities. In addition, when applied to real data, where a sample covariance matrix S will replace its corresponding population Σ to yield a sample component loading matrix $\hat{L}$, (5) and (6) will no longer hold exactly. This implies the need for a methodology to optimize the approximations.

Practical Approaches to Factor Analysis via Components

The interrelations developed above can be used in several ways to obtain new factor analytic estimation methods. The most obvious approach is to define optimization functions based on the population relations (5)-(6), and then to apply them to parameter estimation with sample data. In the population, minimized values of these functions provide sample-size independent definitions of noncentrality parameters when the chosen factor model is not true, while when implemented with sample components, they define discrepancy functions to be minimized in an estimation methodology. Because of the scale invariance noted in Result 6, we can work with the correlation matrix without loss of generality, i.e., we may consider the above relations based on standardized observed variables. That is, we consider $P = LL'$, where P is the population correlation matrix.

Some interesting population functions can be defined on the discrepancies between the left and right sides of (5) and (6). In this paper, we consider only relatively simple nonnegative functions that take on the value of zero under the null hypothesis (2). First we consider a kind of generalized least squares function based on (5) that, in contrast to previous approaches to factor analysis that fit the factor model to the correlation matrix, fits the factor model to the component loading matrix

$$\delta_1 = \frac{1}{2} tr(L - \Lambda V)' W (L - \Lambda V) \text{ with } \Lambda V_\perp = 0 . \tag{8}$$

W is a weight matrix that can be chosen in various ways. When $W = P^{-1}$, (8) is a scale-invariant way to measure how close $\Lambda V L^{-1}$ is to an identity matrix based on $\delta_1 = .5 tr(I - \Lambda V L^{-1})'(I - \Lambda V L^{-1})$ and $\Lambda V_\perp = 0$. Although here we use a trace function to quantify this, it also is possible to use $\left|\Lambda V L^{-1}\right|$ alone or in combination with (8). When $W = I$ this is a least squares (LS) discrepancy function. When a sample $\hat{L}$ is fitted, the formulation allows such options as $W = R^{-1}$, based on the sample correlation matrix R, and $W = \hat{P}^{-1}$ based on the estimated model $\hat{P} = \hat{\Lambda}\hat{\Lambda}'$; these may be called generalized least squares (GLS) and reweighted least squares (RLS), respectively.

In parallel to the above, we also may consider a function based on (6) that fits the factor loading matrix to a transformed components loading matrix, specifically

$$\delta_2 = \frac{1}{2} tr(\Lambda - L V')' W (\Lambda - L V') . \tag{9}$$

In this case, when $W = P^{-1}$, the function simplifies to the discrepancy $\delta_2 = p - tr(\Lambda V L^{-1})$, and as with (8), alternative specifications of W lead to LS-, GLS-, and RLS-type methods. Of course, additional methods are possible as noted in the Discussion.

Although there are a lot of options for estimation, a careful study of alternatives is beyond the scope of this paper. We may call this class of procedures EFAC, exploratory factor analysis via components. We simply illustrate the proposed approach with one method and one data set, taking $L = \hat{L}$ from $R = \hat{L}\hat{L}'$ and treating it as fixed.

Computational Approach

We illustrate our results by minimizing (9) with $W = I$, i.e., using the least squares special case. What is different in our approach is that V is also an unknown parameter matrix that needs to be estimated along with Λ, although it will typically not be of special interest. Here we consider a simple alternating least squares approach to obtain the unknown optimum parameter estimates $\hat{\Lambda}$ and $\hat{V}$. We use two steps

1. Given a current Λ, say Λ_i, find a V, say V_i;

2. Given a current V_i, find a new Λ, say Λ_{i+1},

repeated in sequence until convergence. Abstractly, Step 1 is implemented by obtaining the gradient $\partial\delta_i / \partial V$ and computing up a type of gradient projection step (Jennrich, 2002) involving a least squares orthonormalization similar to that of orthogonal Procrustes rotation (Schönemann, 1966). This makes use of a Lagrangian constraint to assure that V is an orthonormal matrix at each iteration, and assures that $\Lambda V_\perp = 0$. Step 2 is implemented by a similar gradient project step that assures that the estimated $\hat{\Lambda}$ has the required form (2). To obtain a unique $\hat{\Lambda}$ that does not allow rotation, at each step we take it to have an upper-right triangle of zeros. After convergence, we will rotate it into a more interesting form, in the example, to maximize the varimax criterion.

Specifically, in Step 1, using the current estimates we compute the steepest-descent update matrix on the left of the equality

$$V - \alpha(VL'L - \Lambda'L) = PDQ' \quad , \tag{10}$$

and obtain its singular value decomposition as given on the right of (10). Here, α is a step-size number chosen to guarantee a decrease in function (9); Jennrich proved that such a number exists. The usual properties assure that $P'P = I, Q'Q = I$, and D is the diagonal matrix of singular values. Then the new estimate

$$V_i = PQ' \tag{11}$$

is obtained. This is the least squares orthogonal normal approximation to the matrix in (10). This new estimate V_i is used in Step 2, which proceeds by computing $\partial\delta_{2LS} / \partial\Lambda = 0 \Rightarrow \Lambda_{i+1} = [LV_i']_f$ where $[.]_f$ extracts the elements corresponding to free parameters in Λ and ignores the rest. Then we return to Step 1 and continue cycling until a minimum of (9) is obtained, yielding the final $\hat{\Lambda}$ and $\hat{V}$.

Example

To illustrate this methodology, we analyze the well known Holzinger-Harman 24 psychological tests and compare the results to maximum likelihood (ML) and least squares (LS) solutions. Convergence to the minimum of (9) was straightforward. Since the unrotated solutions of ML and EFAC are not comparable due to different conditions on the initial solutions, both results were rotated by varimax. Table 1 gives the two varimax solutions. The EFAC solution on the left side of Table 1 is remarkably similar to the ML solution given in the right side of Table 2: the root mean square difference across all elements of the two matrices is .012. A similar comparison of the EFAC solution to the standard iterative principal axes solution (not shown) obtained a root mean square difference of .009.

Discussion

As far as can be determined, the interrelations between components and latent factor scores as well as between components and factor loading matrices that were developed in this manuscript have not previously been recognized. The results provide the basis for a class of new estimation methods, dubbed EFAC, to the century old factor analysis model. In this approach, the components loading matrix rather than the correlation matrix is used to define a discrepancy function. The illustrative results verify that an EFAC solution can be surprisingly similar to the classical maximum likelihood and least squares solutions on the classical 24 psychological variables, suggesting that further research into its properties may be of interest in the future.

Being primarily conceptual in nature, this paper has not addressed the wide variety of estimation methods that can be developed based on this conceptualization. The functions (8) and (9) are just illustrative, and many variants are possible. For example, δ_2 is based on a $p \times p$ weight matrix W, but the variant $\delta_2^* = \frac{1}{2} tr(\Lambda - LV')W(\Lambda - LV')'$ would be based on a $q \times q$ weight matrix. Or a discrepancy function could be based on the differences between the singular values σ_i of L and singular values Δ_i of Λ, such as $\sum(\sigma_i - \Delta_i)^2$. This is an adaptation of a criterion introduced by de Leeuw (2004), the sum of squared differences between singular values of a raw score data matrix and those of Λ. Especially interesting would be the development of discrepancy functions that provide the type of statistics of typical interest in covariance structure analysis, such as goodness-of-fit χ^2 tests and standard error estimates for the parameters. These might include such obvious variants as asymptotically distribution free methods, elliptical, normal theory and robust methods based on distribution of $\hat{L}$ and the data.

Although we showed that the proposed EFAC methodology can yield results that are equivalent to those from standard methods, an important question to consider is whether any variants of this methodology actually can yield improvements over existing methods. If not, our results will be of interest mainly in providing a new theoretical perspective on the relations between components and factor analysis.

In this manuscript, for simplicity we considered only the exploratory factor analysis model. It is possible to extend our approach so as to relate components to other latent variable models as well.

References

Bentler, P. M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*, 67-74.

de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In K. van Montfort , J. Oud, & A. Satorra, (Eds.), R*ecent developments on structural equation models: Theory and applications* (Ch 7). Dordrecht, Netherlands: Kluwer Academic. See also http://preprints.stat.ucla.edu, Preprint #360.

Guttman, L. (1956). Best possible systematic estimates of communalities. *Psychometrika*, *21*, 273-285.

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, *67*, 7-20.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2^{nd} Ed.). Boca Raton, FL: Chapman & Hall/CRC.

Schönemann, P. H. (1966). A general solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1-10.

Table 1

EFAC (left) and ML (right) Varimax Rotated Solutions for 24 Psychological Variables

1	0.153	0.704	0.169	0.130	1	0.160	0.689	0.187	0.161
2	0.119	0.435	0.099	0.081		0.117	0.436	0.083	0.096
3	0.141	0.534	-0.010	0.159		0.137	0.571	-0.020	0.109
4	0.227	0.553	0.088	0.054		0.233	0.528	0.099	0.079
5	0.741	0.185	0.216	0.148		0.739	0.185	0.213	0.150
6	0.764	0.209	0.067	0.232		0.767	0.205	0.066	0.233
7	0.809	0.201	0.152	0.070		0.806	0.197	0.153	0.075
8	0.569	0.342	0.233	0.139		0.569	0.339	0.242	0.131
9	0.810	0.209	0.043	0.215		0.806	0.201	0.040	0.227
10	0.168	-0.098	0.824	0.157		0.168	-0.118	0.831	0.167
11	0.176	0.111	0.530	0.383		0.179	0.119	0.511	0.378
12	0.019	0.209	0.719	0.087		0.019	0.210	0.716	0.089
13	0.181	0.428	0.535	0.084		0.187	0.437	0.525	0.083
14	0.202	0.045	0.081	0.574		0.197	0.050	0.081	0.554
15	0.120	0.125	0.077	0.514		0.121	0.116	0.075	0.522
16	0.068	0.424	0.053	0.517		0.069	0.408	0.062	0.525
17	0.138	0.069	0.221	0.583		0.142	0.062	0.219	0.573
18	0.022	0.308	0.339	0.447		0.026	0.294	0.336	0.455
19	0.145	0.244	0.167	0.364		0.148	0.240	0.161	0.365
20	0.378	0.421	0.107	0.288		0.378	0.402	0.118	0.300
21	0.175	0.407	0.433	0.200		0.175	0.381	0.438	0.223
22	0.365	0.414	0.125	0.283		0.366	0.399	0.122	0.301
23	0.370	0.517	0.234	0.224		0.369	0.501	0.244	0.238
24	0.367	0.179	0.492	0.294		0.370	0.158	0.496	0.303
Col SSQ	3.647	2.975	2.657	2.257		3.647	2.875	2.654	2.292