LINEAR COMPRESSIONS USING THE IMAGE FACTOR ANALYSIS CRITERION

Michael D. Pore, University of New Orleans

Image analysis, as developed by Guttman [1953], is the partitioning of an observable random vector, X, into an image vector, WX, and an anti-image vector, (I-W)X.

(1) $$X = WX + (I-W)X$$

where WX is the minimum squared error linear estimate of X such that

$$\text{diag } W = \phi.$$

Diag W (implies W is a square matrix) is a diagonal matrix with elements on the main diagonal identical to the main diagonal of W.

Let R be the sample correlation matrix associated with a random sample of $p \geq N$ observations (each $N \times 1$) on X.

Theorem 1: The coefficient regression matrix, W (defined above), can be shown to by any solution to the system

(2) $$\text{Diag } W = \phi$$

(3) $$WR = R - D$$

for some diagonal matrix D.

A literature search has not revealed a general solution to the system (2) and (3), but only a solution in the case that R is non-singular. This solution follows by post-multiplying (3) by R^{-1} to yield

(4) $$W = I - DR^{-1},$$

and then impose the restriction (2) to show that

(5) $$D = (\text{diag } R^{-1})^{-1}.$$

Current computer programs in image covariance factor analysis, for example in the Statistical Programs in the Social Sciences (SPSS), use (4) and (5) but are not operable for singular R. The property that each element of (5) is non-negative, and several other properties shown by Guttman [1956] have made (5) a nearly universal estimate of the covariance of the "unique" variables in factor analysis. In order that image analysis might be more generally used and reasonable estimates for factor analysis developed, a general solution to Theorem 1 is now developed. The following preliminary results are used to develop Theorem 2. R^+ represents the pseudo-inverse of R, as defined in Boullion and Odell [1971].

Lemma 1: There are no zeros on the main diagonal of R^+.

Notation: Let $A = \text{diag}(I - R^+R) = (I - \text{diag } R^+R)$.

Theorem 2: A solution to the system of equations

$$W R = R - D$$

$$D \text{ is diagonal}$$

$$\text{diag } W = \phi$$

is

(6) $$W_0 = I - (I - AA^+)(\text{diag } R^+)^{-1} R^+ - A^+(I - R R^+)$$

and

(7) $$D_0 = (I - AA^+)(\text{diag } R^+)^{-1}.$$

In general, all solutions to Theorem 2 are of the form

$$W = I - H_1 (I - A A^+) R^+ + H_2 (I - R R^+)$$

for some H_1 and H_2 where H_1 is diagonal. Theorem 2 presents the solution when H_2 is also diagonal.

Factor Analysis.

Factor analysis hypothesizes the existence of an integer M, $0 < M < N$, and random vectors Y and U such that

(8) $$\underset{N\times 1}{X} = \underset{N\times M}{\Lambda}\ \underset{N\times M}{Y} + \underset{N\times 1}{U}$$

where Λ is an $N \times M$ matrix of constant coefficients and

(9) $$E(Y) = \phi,\ E(U) = \phi,$$

(10) $$\text{Cov}(Y,Y) = I$$

(11) $$\text{Cov}(Y,U) = \phi$$

(12) $$\text{Cov}(U,U) \text{ is diagonal.}$$

This model implies the following partitioning of the correlation matrix:

(13) $$R = \Lambda\Lambda^T + \text{Cov}(U,U)$$

Two general methods exist for fitting data to this model. The first (1) is to initially estimate $\Lambda\Lambda^T$ with $R-D_0$, then use one of three criteria to reduce the rank of $R-D_0$. The three criteria are principal factor analysis (PFA), cannonical factor analysis (CFA), and alpha factor analysis (AFA). These three criteria are discussed later. The second method (2) is to initially estimate $\Lambda\Lambda^T$ with the image covariance matrix, $G = \text{Cov}(W_0X, W_0X) = W_0R\,W_0^T = R + D_0R^+D_0 - 2D_0$, then use a "rank reducing" method on G.

Neither method has been possible for singular R, although iterative approximation schemes have been developed for the first (1) method. Tests of significance for "goodness of fit" exist for only one method, CFA, and then only under assumptions of normality. Hence, the present model-fitting methods are unsatisfactory, often giving rough approximations with gross errors.

Y is interpreted as the common factors. This means that conceptually

ΛY and WX are identical, yet Y is not a linear combination of X. (A proof is in Pore [1973]). Also, image analysis makes a strong argument against the factor analysis model including equation (12): since the anti-image covariance matrix Cov(I-W)X, (I-W)X, is not diagonal.

These criticisms, coupled with the experience of researchers declaring that equation (12) was not a significant assumption in their analyses, has led the author to drop this assumption and continue the analysis. This "relaxed" factor analysis model can not only be precisely fit with any set of data (for all M, $0 < M < N$) but many such solutions exist. Hence, to restrict it for more meaningful interpretations, the common factors, Y, are restricted to being a linear compression of X. This is

$$(14) \qquad Y = BX$$

where B is a full rank $N \times M$ matrix of unknown constant coefficients. The "modified" factor analysis model (MFA) is given by

$$X = \Lambda Y + U,$$

where

$$E(Y) = \phi, \quad E(U) = \phi,$$
$$\text{Cov}(Y,Y) = I,$$
$$\text{Cov}(Y,U) = \phi$$

and

$$(15) \qquad Y = BX.$$

It can easily be shown that

$$(16) \qquad \Lambda = RB^T$$

and

$$(17) \qquad U = (I - RB^TB)X,$$

with the only restriction on B being that it is $M \times N$ and satisfies

$$(18) \qquad BRB^T = I.$$

The researcher need no longer estimate common factor scores for particular individuals. They are specified in the model (15) as a linear compression of the observation. This also makes interpretation of Y a precise linear compression of X.

In being so general the MFA model allows data to be fitted to the model in ways that are meaningless to the researcher. That is, thare exist matrices, B, such that the MFA model is satisfied, yet y is uninterpretable. The problem is that

$$BRB^T = I$$

does not sufficiently restrict B to meaningfull solutions. Classical partitioning procedures may be imposed to optimize interpretability. One application of this model, MPC, is to partition B as

$$B = PA,$$

so that AX are the principal components of X. P then reduces the rank to the M largest principal components and scales B so that (18) is satisfied. This is one meaningful way to select B, but it is not in keeping with the concepts of factor analysis: that is, it does nothing toward separating error and underlying factors, etc. The following method is designed to do just that.

Application: B is partitioned so that

$$(19) \qquad B = PW$$

where WX is the image of X and P, an $M \times N$ matrix, reduces the rank of B and scales it so the (18) is satisfied. (See (1) and Theorem 2 for an explicit definition of W).

The superior properties of the image vector in factor analysis have been discussed. It is, conceptually, the type of common factor "filter" for which the researcher uses factor analysis. Y, then, represents a linear compression of the image of X. The precise type of linear compression will depend on the researcher's criteria for optimization. Classical criteria include principal components, canonical correlation, and generalizability. Methods using each of these criterion are now presented.

MFA Applied to PFA Criterion (MPF).

PFA is the method where Cov(U,U) is estimated with D_0 of Theorem 2, and then the largest M principal components of $R-D_0$ are extracted. MFA follows the same principle. The main difference is that the classical factor analysis model is not being approximated, hence estimates of Cov(U,U) are no longer diagonal. Image theory justifies using the anti-image covariance matrix as an initial extimate of D. The procedure is, then, to extract the largest M principal components from the image covariance matrix.

The image covariance matrix can be shown to be

$$(20) \qquad G = R + D_0R^+D_0 - 2D_0.$$

Application: Let E be the $N \times N$ p.s.d. diagonal matrix with the eigenvalues of G on the main diagonal in descending order and H the respective $N \times N$ eigenvector matrix such that

$$(21) \quad G = HEH^T = [H_1 \quad H_2] \begin{bmatrix} E_1 & \phi \\ \phi & E_2 \end{bmatrix} \begin{bmatrix} H_1^T \\ H_2^T \end{bmatrix} = H_1E_1H_1^T + H_2E_2H_2^T$$

for partitionings of E and H such that E_1 is $M \times M$ p.d., and H_1 is $N \times M$.

Now P in (19) is

(22) $$P = E_1^{-\frac{1}{2}} H_1.$$

Notice that
$$BRB^T = E_1^{-\frac{1}{2}} H_1 W R W^T H^T H_1^T E_1^{-\frac{1}{2}}$$
$$= E_1^{-\frac{1}{2}} H_1 G H_1^T E_1^{-\frac{1}{2}}$$
$$= E_1^{-\frac{1}{2}} E_1 E_1^{-\frac{1}{2}}$$
(23) $$= I.$$

BX is the scaled linear compression of the image variables that retains the M largest principal components of the image variables. Those researchers partial to PFA will find the same favorable principles in MPF, but applied to the image variables, rather than a hypothetical set of variables with covariance R-D.

MFA Applied to CFA Criteria (MCF).

CFA is the method of constructing factors, Y, that have maximum canonical correlation with the observations X. There are at least two possible ways to apply this principle to MFA. These are: (1) maximizing the correlation of Y and X, as above, with the restriction Y = PWX (2) maximizing the canonical correlation of Y and WX, with the restriction Y = PWX. The first method is what will be used to develop MCF. The second method remains undeveloped.

Proceeding in the fashion developed for CFA:

(24) $$\mathrm{Cov}(Y,Y) = D$$

(25) $$\mathrm{Cov}(X,X) = R$$

(26) $$\mathrm{Cov}(X,Y) = RW^T P^T = \Lambda.$$

Hence, $\begin{matrix} Y \\ X \end{matrix}$ has the super covariance matrix

$$\Sigma = \begin{bmatrix} I & \Lambda^T \\ \Lambda & R \end{bmatrix}$$

Following the same procedure as Anderson [1957] (Chapter 12),

$$|\Lambda\Lambda^T - \lambda R| = 0$$

but since our estimate of $\Lambda\Lambda^T$ is G, the image covariance matrix, then

(28) $$|G - \lambda R| = 0.$$

Now since R can be written

$$R = Q[e_i]Q^T$$

where $[e_i]$ is an $r \times r$ p.d. diagonal matrix, r is the rank or R, and Q is an $N \times r$ matrix of eigenvectors of R respective to $[e_i]$ then (28) implies

$$[e_i]^{-\frac{1}{2}} Q^T G Q [e_i]^{-\frac{1}{2}} - \lambda I = 0.$$

where the λ is the canonical correlation involved.

Application: Let $[u_i]$ be the p.d. diagonal matrix of eigenvalues of $[e_i]^{-\frac{1}{2}} Q^T G Q [e_i]^{-\frac{1}{2}}$ in descending order, and S the respective matrix of eigenvectors.

It can be shown that

(29) $$P = [u_i]_1^{-\frac{1}{2}} S_1^T [e_i]^{-\frac{1}{2}} Q^T.$$

Notice that (18) is assured by
$$WRW^T = G$$
and it follows that
$$PGP^T = I.$$

The researcher preferring CFA will find MCF more theoretically defensible, yet based on the same optimization procedure. MCF also incoropates the properties of image analysis, hence optimization is bi-dimensional, rather that simply in the one aspect, correlation.

MFA Applied to AFA Criteria (MAF).

The AFA criterion is to define the common factors in such a way that their reliability coefficient is maximized. MAF will use the same reliability coefficient, α , that AFA used. (Tryon [1957] has shown that practically all reliability coefficients are equivalent to α.) The α coefficient is

$$\alpha = \frac{N}{N-1}\left[1 - \frac{\omega^T(\mathrm{Trace}\ \Lambda\Lambda^T)\omega}{\omega^T \ \Lambda\Lambda^T \ \omega}\right]$$

where ω is the coefficient vector in defining the common factors

(30) $$Y = \omega^T W X.$$

By substituting the initial estimate of $\Lambda\Lambda^T$ with G (defined by (20)) yields

(31) $$\alpha = \frac{N}{N-1}\left[1 - \frac{\omega^T(I - D)\omega}{\omega^T G \omega}\right]$$

which, in turn, yields an identical eigenvalue equation as in AFA (see Kaiser and Caffrey [1965]) except R - D is replaced by G. Hence,

(32) $$(H^{-1}GH^{-1} - U^2 I)e = \phi$$

where $H^2 = I - D$.

Application: Let $[e_i]$ be the $N \times N$ diagonal matrix of eigenvalues of $H^{-1}GH^{-1}$, in descending order. Also let Q be the respective eigenvector matrix. It can be shown that

$$P = [e_i]_1^{-\frac{1}{2}} Q_1^T H^{-1}.$$

Notice that

$$PGP^T = I,$$

hence the restrictions for MFA are satisfied. Since P is the objective of the procedure above, it differs slightly from AFA; but the principle, general method of analysis, and interpretability are all identical to AFA. Although AFA is not as widely used as the other forms of factor analysis, researchers may find an increasing need for psychometric sampling (or Q-analysis, as it is sometimes called: [Rummel, 1970]). As researchers do, they may find AFA claims the type of analysis they are looking for; but MAF will do likewise, and it will fit a model precisely, as opposed to the approximation technique of classical methods.

REFERENCES

Anderson, T.W. An Introduction To Multivariate Statistical Analysis. New York: John Wiley & Sons, Inc,, 1957.

Boullion, Thomas L., and Odell, Patrick L. Generalized Inverse Matrices. New York: John Wiley & Sons, Inc., 1971.

Guttman, L. Image theory for the structure of quantitative variates. Psychometrika, 1953, 18, 277-296.

__________. Some necessary conditions for common factor analysis. Psychometrika, 1954, 19, 149-161.

__________. "Best Possible" systematic estimates of communalities. Psychometrika, 1956, 21, 273-285.

Kaiser, H.R., and Caffrey, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Pore, Michael D. Factor analysis as a data compression technique. Unpublished Ph.D. Dissertation, Texas Tech University, 1973.

Rummel, Rudolph J. Applied Factor Analysis. Evanston, Illinois: Northwestern University Press, 1970.

Tryon, R.C. Reliability and behavior domain validity: reformulation and historical critique. Psychological Bulletin, 1957, 54, 229-249.