Factor Analysis

Jaime Frade

December 5, 2007

Part I
9.4: Factor Rotation

1 Review:

1.1 Covariance Structure for the Orthogonal Factor Model

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL^t} +$

$$\begin{aligned} \text{Var}(X_i) &= l_{i1}^2 + \cdots + l_{im}^2 + \Psi \\ \text{Cov}(X_i) &= l_{i1}l_{k1} + \cdots + l_{im}l_{km} + \Psi \end{aligned} \tag{1}$$

2. $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$
   or

$$\text{Cov}(X_i, F_i) = l_{ij}$$

It has been shown **(9-8) pg 487**, that all factor loadings obtained from the initial loadings by an orthogonal transformation have the same ability to reproduce the covariance (or correlation) matrix. An orthogonal transformation is a rotation or reflection of the coordinate axis.

Factor analysis is a generic term for a family of statistical techniques concerned with the reduction of a set of observable variables in terms of a small number of latent factors. It has been developed primarily for analyzing relationships among a number of measurable entities (such as survey items or test scores). The underlying assumption of factor analysis is that there exists a number of unobserved latent variables (or 'factors') that account for the correlations among observed variables, such that if the latent variables are partialled out or held constant, the partial correlations among observed variables all become zero. In other words, the latent factors determine the values of the observed variables.

The primary purpose of factor analysis is data reduction and summarization

2 Definition:

Factor Rotation: An orthogonal transformation of the factor loadings, as well as the implied orthogonal transformation of the factors.
Let $T_{m\times m}$ be an orthogonal matrix,
$(TT^t = T^tT = I_{m\times m})$.

If $\widehat{L}$ is the $p \times m$ matrix of estimated factor loading obtained by any method (PCP, ML, etc.) then

$$\widehat{L}^* = \widehat{L}T, \quad \text{where } TT^t = T^tT = I \tag{2}$$

is the $p \times m$ matrix of 'rotated' loadings. $\widehat{L}^*$ is an orthogonal transformation of the factor loading matrix. The estimated covariance (or correlation) matrix remains unchanged.

$$\begin{aligned} \widehat{L}\widehat{L}^t + \widehat{\Psi} &= \widehat{L}TT^t\widehat{L}^t + \widehat{\Psi} \\ &= \widehat{L}^*(\widehat{L}^*)^t + \widehat{\Psi} \end{aligned} \tag{3}$$

indicates that the residual matrix remains unchanged.

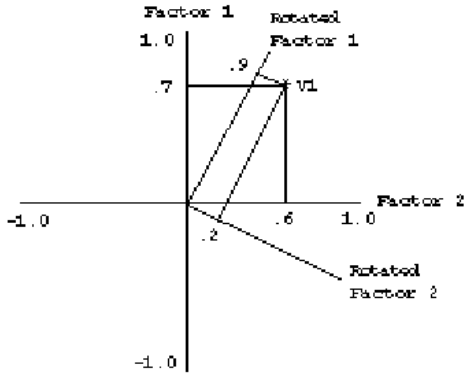
Also, the estimated specific variances and communalities, $\widehat{\Psi}_i$, and hence the $\widehat{h_i^2}$, are unchanged by orthogonal transformations of $\widehat{L}$.

Therefore, can still use either $\widehat{L}$ or $\widehat{L}^*$.

The examples illustrate graphical and analytical methods for determining an orthogonal rotation to a simple structure.

3 Examples

3.1 Example 9.8 (First look at factor rotation.) Oblique transformation.



Presented with the sample correlation matrix of examination scores in $p = 6$ subject areas for $n = 220$ male students. The correlation matrix is

$$R = \begin{bmatrix} 1.0 & 0.439 & 0.410 & 0.288 & 0.329 & 0.248 \\ & 1.0 & 0.351 & 0.354 & 0.320 & 0.329 \\ & & 1.0 & 0.164 & 0.190 & 0.181 \\ & & & 1.0 & 0.595 & 0.470 \\ & & & & 1.0 & 0.464 \\ & & & & & 1.0 \end{bmatrix} \tag{4}$$

Using a Maximum likelihood solution for $m = 2$ factors yields the

RESULTS

`http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_8.htm`

CODE

`http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_8_pg505.sas`

4 Varimax Criterion

Define the 'scaled' loading, $\widetilde{l}^*_{ij}$, of the i-th variable on the j-th factor after rotation. Rotated coffecients scaled by the square root of the communalities.

$$\widetilde{l}^*_{ij} = \frac{\widehat{l}^*_{ij}}{\widehat{h}_i}$$

Then the varimax procedure selects the orthogonal transformation that maximizes

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{j=1}^{p}(\widetilde{l}^*_{ij})^4 - \frac{1}{p}\left(\sum_{i=1}^{p}\left(\widetilde{l}^*_{ij}\right)^2\right)^2\right] \tag{5}$$

Interpretation of (5),

$$V \propto \sum_{j=1}^{m}(\text{Variance of squares of scaled loading for the j-th factor}) \tag{6}$$

Notes: Scaling gives the variables with smaller communalities more influence. Overall, maximizing V (5) corresponds to 'spreading out' the squares of the loadings on each factor as much as possible. Therefore, we hope to find groups of large and negligible coefficients in any *column* of the rotated ladings matrix, $\widetilde{l}^*_{ij}$.

4.1 Example 9.9(2) (Rotated loadings for stock-price data.)

Returning to the data discussed in example 9.3 **pg 493**. The original factor loadings (obtained by PCP), the communalities, and the (varimax) rotated factor loadings are shown in Table below. SAS code given as well.

`http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9-10.htm`

It is clear that the bank stocks (JP Morgan, Citibank, and Wells Fargo) load highly on the first factor, while the oil stocks, (Shell and Mobil) load highly on the second factor. Factor 1 may indicate those unique factors that case bank stocks to move together and factor 2 represents conditions affecting oil stocks.

`http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_10_complete.htm`

`http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_10_pg510.sas`

The factor loadings for the variables are pictured below with respect to the original and (varimax) rotated factor axes in below figure. After rotation each of the p variables should have a high loading on only one factor, however, this is not always possible.

The general market factor was destroyed by rotation. Analyze the relationship to pattern of correlations.

Part II

9.5: Factor Scores

Factor analysis centers the focus on the parameters in the factor model, however the estimated values of the common factors, **factor scores**, may also be of interest. These quantities are often used for diagonstic purposes, as well as inputs to a subsequent analysis.

Factor scores are estimates of values for the unobserved random factor vectors, $\mathbf{F}_j$, $j = 1, 2, ldots, n$.

$$\widehat{\mathbf{f_j}} = \text{estimate of the values } \mathbf{f_j} \text{ attained by } \mathbf{F_j} \quad (j\text{th case})$$

However, estimation is complicated by the fact that unobserved quantities, $\mathbf{f_j}$ and $\boldsymbol{\epsilon_j}$, outnumber the observed $\mathbf{x_j}$. The following two methods, Weighted Least Squares and Regression Methods, describe ways to overcome this complication.

5 Estimation of Factor Scores for both models

Suppose that the mean vector, $\boldsymbol{\mu}$, the loading factors, $\mathbf{L}$, and the specific variance Ψ are known for the factor model,

$$(\mathbf{X_j} - \boldsymbol{\mu})_{(p\times 1)} = \mathbf{L}_{(p\times m}\mathbf{F}_{(m\times 1)} + \epsilon_{(p\times 1} \tag{7}$$

regard the specific factors $\epsilon' = [\epsilon_1, \epsilon_2, \ldots, \epsilon_p]$ as errors. Since the $\text{Var}(\epsilon_i) = \Psi_i$, $i = 1, 2, \ldots, p$ need not be equal, thus

$$\text{Var}(\epsilon_i) = \boldsymbol{\Psi} = \begin{bmatrix} \Psi_1 & 0 & 0 & \ldots & \ldots \\ 0 & \Psi_2 & 0 & \ldots & \ldots \\ 0 & 0 & \Psi_3 & \ldots & \ldots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \Psi_p \end{bmatrix} \tag{8}$$

will use the weighted least squares to estimate the common factor values.

6 Weighted Least Squares Method

WLS estimation of $\widehat{f_j}$ of f_j is

$$\mathbf{f}_j \;=\; \left(\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L}\right)^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) \tag{9}$$

Using (9), take estimates $\widehat{\mathbf{L}}, \widehat{\mathbf{\Psi}}, \&\widehat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ as true values to obtain the factor scores for the j-th case:

$$\widehat{\mathbf{f}}_j \;=\; \left(\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}}\right)^{-1}\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\left(\mathbf{x_j}-\bar{\mathbf{x}}\right) \tag{10}$$

In (10), the estimates $\widehat{\mathbf{L}}, \widehat{\mathbf{\Psi}}$ are obtained by the maximum likelihood method and must satisfy the uniqueness condition, $\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}} = \widehat{\Delta}$, a diagonal matrix.

Note: Ordinary (unweighted) least squares estimation is sometimes used when factor loadings are obtained from the principal component method, since specific variances tend to be more nearly equal. (i.e. $\widehat{\Psi_1} = \cdots = \widehat{\Psi_p}$)

The factor scores are then,

$$\widehat{\mathbf{f}}_j \;=\; \left(\widetilde{\mathbf{L}}'\widetilde{\mathbf{L}}\right)^{-1}\widetilde{\mathbf{L}}'\left(\mathbf{x_j}-\bar{\mathbf{x}}\right) \tag{11}$$

$$\;=\; \left(\widetilde{\mathbf{L_z}}'\widetilde{\mathbf{L_z}}\right)^{-1}\widetilde{\mathbf{L_z}}'\mathbf{z}_j \tag{12}$$

where, (12), is for the standardized data.

Since, (see (9-15) **pg. 490**

$$\widetilde{\mathbf{L}} = \begin{bmatrix} \sqrt{\widehat{\lambda}_1}\widehat{\mathbf{e}}_1 \\ \sqrt{\widehat{\lambda}_2}\widehat{\mathbf{e}}_2 \\ \vdots \\ \sqrt{\widehat{\lambda}_m}\widehat{\mathbf{e}}_m \end{bmatrix} \tag{13}$$

Using (13), into (11) will obtain

$$\widehat{\mathbf{f}}_j = \begin{bmatrix} \dfrac{1}{\sqrt{\widehat{\lambda}_1}}\widehat{\mathbf{e}}_1'\left(\mathbf{x_j}-\bar{\mathbf{x}}\right) \\ \dfrac{1}{\sqrt{\widehat{\lambda}_2}}\widehat{\mathbf{e}}_2'\left(\mathbf{x_j}-\bar{\mathbf{x}}\right) \\ \vdots \\ \dfrac{1}{\sqrt{\widehat{\lambda}_m}}\widehat{\mathbf{e}}_m'\left(\mathbf{x_j}-\bar{\mathbf{x}}\right) \end{bmatrix} \tag{14}$$

Comparing, from princpal component analysis, (see (8-21), **pg 443**), can obtain the ith principal component $\widehat{y}_{ji} = \widehat{\mathbf{e}}_i'(\mathbf{x}-\bar{\mathbf{x}})$ for any observation vector, $\mathbf{x}$.

Comparing (14) with above (8-21), $\widehat{\mathbf{f}}_j$ is the first m (scaled) principal components evaluated at $\mathbf{x}_j$.

7 Regression Method

Starting from original factor model in (7), treat the loadings matrix, $\mathbf{L}$ and $\mathbf{\Psi}$ as known values.

Consider the joint distribution of $(\mathbf{x}_j - \mu)$ and $\mathbf{F}_j$. Assume the multivariate normality as in the maximum likelihood approach to factor analysis.

$$\begin{bmatrix} \mathbf{x}_j - \boldsymbol{\mu} \\ \mathbf{F}_j \end{bmatrix} \sim N_{p+m}\left(\mathbf{0}, \sum{}^{*}\right) \tag{15}$$

7.1 Example 9.12 (Computing Factor Scores by WLS and Regression.)

Using the MLE solution from R for stock data, the estimated rotated loadings and specific variances, the following equations should hold, where the standardized observations, $\mathbf{z}'$

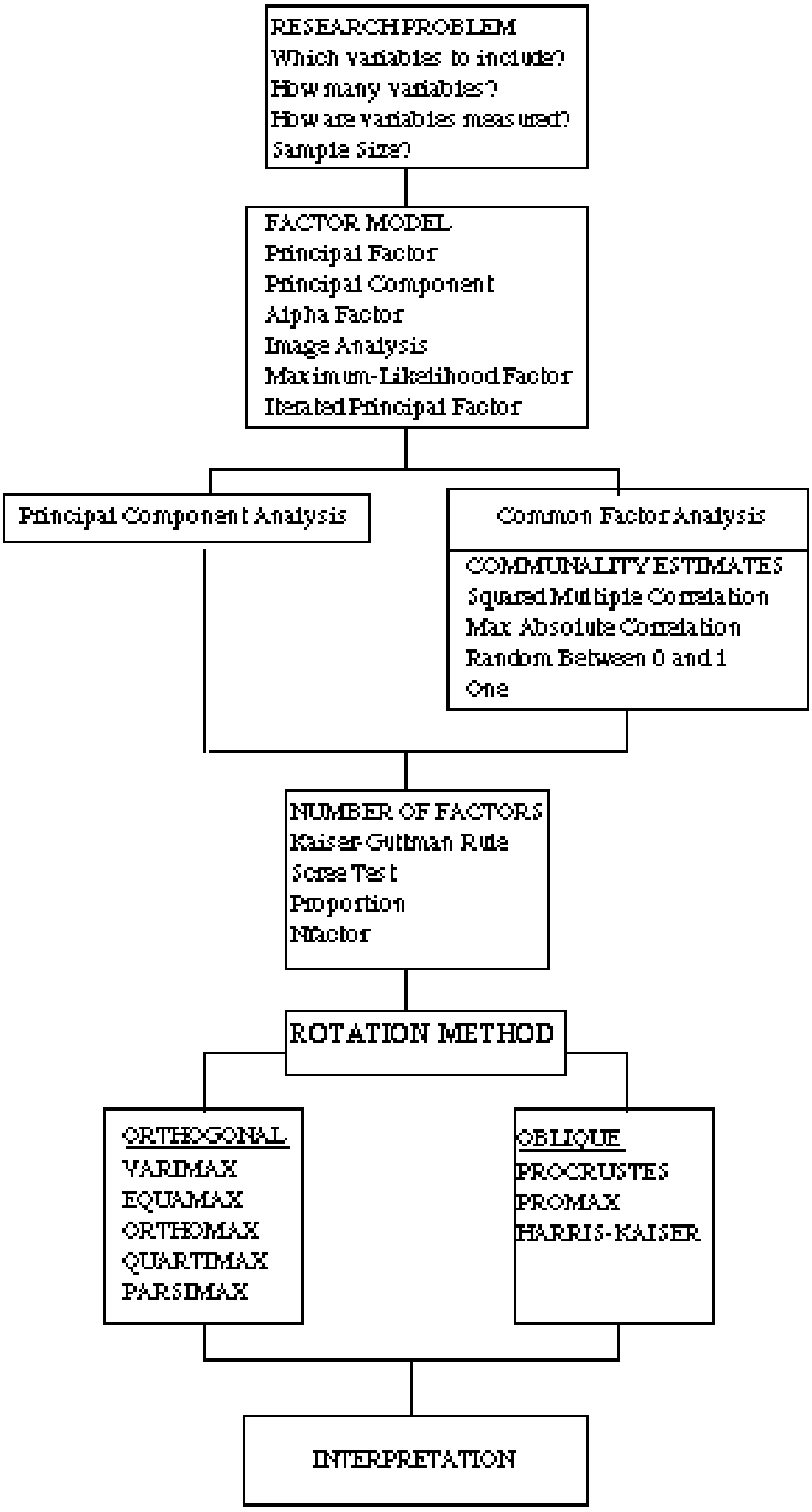
$$\widehat{\mathbf{f}} = \left(\widehat{\mathbf{L}}_{\mathbf{z}}^{*\prime}\widehat{\mathbf{\Psi}}_{\mathbf{z}}^{-\mathbf{1}}\widehat{\mathbf{L}}_{\mathbf{z}}^{*}\right)^{-1}\widehat{\mathbf{L}}_{\mathbf{z}}^{*\prime}\widehat{\mathbf{\Psi}}_{\mathbf{z}}^{-\mathbf{1}}\mathbf{z} \approx \widehat{\mathbf{L}}_{\mathbf{z}}^{*}\mathbf{R}^{-\mathbf{1}}\mathbf{z} \tag{16}$$

Part III

9.6: Perspectives and a Strategy for Factor Analysis

8 Suggestible steps for factor analysis

1. *Perform PCP factor analysis*
   - **a.** Look for observations that are suspicious behavior
   - **b.** Try a varimax rotation
2. *Perform a maximum likelihood factor analysis, including a varimax rotation*
3. *Compare the solutions obtained from the two factor analysis*
   - **a.** Do the loadings group in the same manner?
   - **b.** Plot factor scores using PCP and MLE
4. *Repeat the first three steps for other numbers of common factors* $\boldsymbol{m}$.
5. *For large data sets, split them in half and perform a factor analysis on each part.*


RESEARCH PROBLEM
Which variables to include?
How many variables?
How are variables measured?
Sample Size?
FACTOR MODEL
Principal Factor
Principal Component
Alpha Factor
Image Analysis
Maximum-Likelihood Factor
Iterated Principal Factor
Principal Component Analysis
Common Factor Analysis
COMMUNALITY ESTIMATES
Squared Multiple Correlation
Max Absolute Correlation
Random Between 0 and 1
One
NUMBER OF FACTORS
Kaiser-Guttman Rule
Scree Test
Proportion
Nfactor
ROTATION METHOD
ORTHOGONAL
VARIMAX
EQUAMAX
ORTHOMAX
QUARTIMAX
PARSIMAX
OBLIQUE
PROCRUSTES
PROMAX
HARRIS-KAISER
INTERPRETATION

8.1 Example 9.14 (chicken-bone data)

http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_14.htm

http://www.stat.fsu.edu/~jfrade/PAPERS/Statistics/example9_14.sas