

## Chapter 11: Exploratory Factor Analysis

**Prerequisites:** Chapter 9, Sections 3.5 - 3.8

### 11.1 Some Comments on the History of Factor Analysis

In this chapter we are going to cover a set of techniques known as *Exploratory Factor Analysis*. Originally, these techniques were simply known as factor analysis, but when Confirmatory Factor Analysis was invented, the word "Exploratory" was added so as to differentiate the two types of factor analysis. At this point we will be briefly reviewing the basic factor analysis model. The derivation of that model is done with more detail in Chapter 9. The difference between exploratory and confirmatory analyses is partly stylistic. For one thing, in exploratory analysis it is traditional to use a correlation matrix instead of a covariance matrix. In that case, the model specifies that

$$\begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_p \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \dots & \dots & \dots & \dots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pm} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \dots \\ \eta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_m \end{bmatrix}$$

$$\mathbf{z} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}. \quad (11.1)$$

Using Theorem (4.9) we can easily show that

$$\begin{aligned} \hat{\mathbf{R}} &= E[(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon})(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon})'] \\ &= \mathbf{\Lambda} E(\boldsymbol{\eta}\boldsymbol{\eta}') \mathbf{\Lambda}' + \mathbf{\Lambda} E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\eta}') \mathbf{\Lambda}' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'). \end{aligned}$$

Defining  $E(\boldsymbol{\eta}\boldsymbol{\eta}') = V(\boldsymbol{\eta}) = \boldsymbol{\Psi}$ ,  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = V(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}$ , and knowing that the unique factor vector  $\boldsymbol{\varepsilon}$  is independent of the common factors in the vector  $\boldsymbol{\eta}$ , we can conclude that

$$\hat{\mathbf{R}} = \mathbf{\Lambda} \boldsymbol{\Psi} \mathbf{\Lambda}' + \boldsymbol{\Theta}. \quad (11.2)$$

Thus the presence of unmeasured variables can be revealed by a particular structure in the observed correlation matrix. There are a variety of ways of uncovering the structure revealed in Equation (11.2), many of which were invented long before computers. In general, there are two steps involved in doing this. In the first step, the factors are extracted, but in an arbitrary way where the regression weights in  $\mathbf{\Lambda}$  are generally not interpretable. In a second step, the factors are rotated into an orientation that is more interpretable and hopefully in alignment with theoretical expectations. This is all in contrast to the confirmatory approach, where we hypothesize a certain alignment of the loadings from the beginning, and test the proposed model.

One of the earliest ways, and still the most popular method of factor extraction, is called *Principal Factors*. We begin our discussion with that technique.

### 11.2 Principal Factors Factor Extraction

We will begin with the simplifying assumption that the unobserved factors are z-scores and are also uncorrelated. In that case  $\Psi = \mathbf{I}$  and the model of Equation (11.2) simplifies to

$$\hat{\mathbf{R}} = \mathbf{\Lambda}\mathbf{\Lambda}' + \Theta.$$

The part of the correlation matrix due to the common factors, call it  $\mathbf{R}^*$ , is given by

$$\hat{\mathbf{R}}^* = \mathbf{\Lambda}\mathbf{\Lambda}'. \quad (11.3)$$

The off-diagonal elements of  $\hat{\mathbf{R}}$  and  $\hat{\mathbf{R}}^*$  are identical since  $\Theta$  is diagonal. The  $\Theta$  matrix must be diagonal, being the covariance matrix of the unique factors, and "unique" after all, describes a set of independent factors. However,  $\hat{\mathbf{R}}$  and  $\hat{\mathbf{R}}^*$  do differ on the diagonal. Whereas  $\mathbf{R}$  has unities on the diagonal,  $\mathbf{R}^*$  has the proportion of the variance of each variable that it has in common with the other variables. This proportion is known as the *communality* of the variable. A quick look at  $\mathbf{R}^*$  reveals it to appear as below

$$\mathbf{R}^* = \begin{bmatrix} h_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & h_2^2 & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & h_p^2 \end{bmatrix}$$

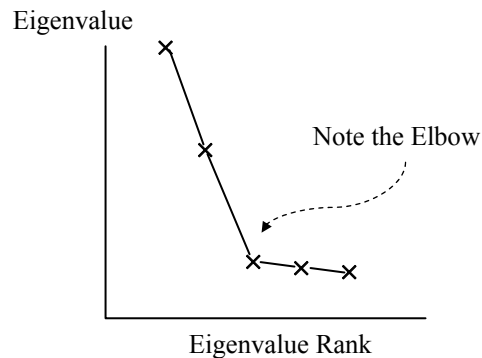
with  $h_i^2$  being the communality of variable  $i$ . The goal of principle factors is to extract factors from  $\mathbf{R}^*$  in such a way as to explain the maximum amount of variance. Extracting the maximum amount of variance is also the goal of eigenstructure, as discussed in Section 3.5. Principle Factors is a technique that uses the eigenstructure of the  $\mathbf{R}^*$  matrix. But before we can proceed, we have to answer two related questions.

1. What are the values of the  $h_i^2$ ?
2. How many factors are there?

If we knew how many factors there were, we could extract that many eigenvalues and eigenvectors from  $\mathbf{R}$ , reproduce  $\hat{\mathbf{R}}$  using the eigenvalues and eigenvectors, then look at the diagonal of this reproduced correlation matrix. Conversely, if we knew what the communalities were, we could deduce the number of factors because while the rank (see Section 3.7) of  $\hat{\mathbf{R}}$  is  $p$ , the rank of  $\hat{\mathbf{R}}^*$  depends on  $m$ , the number of factors as can be seen in Equation (11.3).  $\hat{\mathbf{R}}^*$  is an outer product [Equation (1.21)] with a rank no greater than the number of columns of  $\mathbf{\Lambda}$ . Therefore the number of non-zero eigenvalues of  $\hat{\mathbf{R}}^*$  would tell us exactly how many factors there are. So which comes first: the chicken in the form of the values of the  $h_i^2$ , or the egg in the form of the number of factors?

Even though this is called exploratory factor analysis, we would normally begin with some notion of  $m$ , the number of factors. This notion might come from substantive theory or from an educated guess. Another traditional method is to pick the number of factors based on the number of eigenvalues  $> 1$ . The logic here is that since an eigenvalue represents the variance of the factor, if a factor does not explain even as much as a single observed variable, it is not really pulling its weight.

Another approach is to use a so-called *Scree Chart*.



Given the Scree Chart above, we would pick  $m = 3$  and extract 3 eigenvalues. The third one represents an inflection point, after which there is not much change.

Even if we start with some determined number of factors, it is good to start off with good estimates of the communalities. Here we take advantage of the fact that the lower bound for the communality for a particular variable is the squared multiple correlation,  $R_j^2$ , introduced in Equation (6.21), when that variable is regressed on all the other variables. So we have the relationship

$$R_j^2 \leq h_j^2 \leq 1 \quad (11.4)$$

where  $R_j^2$  is the  $R^2$  value for variable  $j$ , chosen as the dependent variable with all other variables used as independent variables. A very simple computational formula for this is

$$R_j^2 = 1 - \frac{1}{r_{jj}} \quad (11.5)$$

where  $r_{jj}$  is the  $j$ th diagonal element of  $\mathbf{R}^{-1}$ , the inverse of the correlation matrix of all the variables.

We are now ready to discuss the steps of the algorithm known as Principal Factors. We begin with the observed correlation matrix,  $\mathbf{R}$ . According to Equation (11.4), we then can either use the lower bound to the communality, the Squared Multiple Correlation, or use the upper bound, unity. In either case, we find the eigenstructure of  $\mathbf{R}^*$ , and then reproduce that matrix using only the  $m$  largest eigenvalues and their corresponding eigenvectors, i. e.

$$\hat{\mathbf{R}}^* = \mathbf{X}\mathbf{L}\mathbf{X}'$$

Here the columns of the matrix  $\mathbf{X}$  contain the eigenvectors while the diagonal elements of  $\mathbf{L}$  contain the eigenvalues. Now we need only define

$$\mathbf{\Lambda} = \mathbf{X}\mathbf{L}^{1/2}$$

where the square root of a matrix,  $\mathbf{L}^{1/2}$  is uniquely identified since  $\mathbf{L}$  is a diagonal matrix containing the eigenvalues on the diagonal and zeroes elsewhere. Remembering the definition of the Diag function [Equation (2.13)] of a square matrix, by subtraction we can deduce that

$$\Theta = \mathbf{I} - \text{Diag}(\hat{\mathbf{R}}^*).$$

Sometimes Principal Factors is iterated using the following steps.

Step 0. Find the  $m$  largest roots of  $\mathbf{R}$ . Calculate  $\hat{\mathbf{R}} = (\mathbf{X}\mathbf{L}^{1/2})(\mathbf{X}\mathbf{L}^{1/2})' = \mathbf{\Lambda}\mathbf{\Lambda}'$ .

Step 1. Set  $\mathbf{R}^* = \mathbf{R} - [\mathbf{I} - \text{diag}(\hat{\mathbf{R}})]$ .

Step 2. Find the  $m$  largest roots of  $\mathbf{R}^*$ , recalculate  $\hat{\mathbf{R}} = (\mathbf{X}\mathbf{L}^{1/2})(\mathbf{X}\mathbf{L}^{1/2})' = \mathbf{\Lambda}\mathbf{\Lambda}'$ . If  $\hat{\mathbf{R}}$  is not changing from iteration to iteration, stop. Otherwise go back to Step 1.

In Step 0 we can start with unities on the diagonal of  $\mathbf{R}$  and the process will converge down to the  $h_j^2$ , or you start with squared multiple correlations and converge up.

### 11.3 Exploratory Factor Analysis Is a Special Case of Confirmatory

Before the maximum likelihood approach to factor analysis was invented by Lawley (summarized in Lawley and Maxwell 1963), factor analysis existed as a purely descriptive technique. Now we know that exploratory factor analysis is a special case of the confirmatory model discussed in Chapter 9. To implement the special case, we fix the absolute minimum number of parameters necessary to identify the model. The absolute minimum number of parameters that must be fixed to identify an  $m$ -factor model is  $m^2$ . These need to be arranged in the  $\mathbf{\Lambda}$  and  $\mathbf{\Psi}$  matrices in a certain way, however. If we set  $\mathbf{\Psi} = \mathbf{I}$  this fixes  $\frac{m(m+1)}{2}$  parameters leaving

$$\begin{aligned} m^2 - \frac{m(m+1)}{2} &= \frac{2m^2}{2} - \frac{m^2 + m}{2} \\ &= \frac{m(m-1)}{2} \end{aligned} \tag{11.6}$$

restrictions. If you have no hypotheses, other than a hypothesis as to the number of factors,  $m$ , these restrictions may be arbitrarily placed in  $\mathbf{\Lambda}$  with column  $i$  getting  $i - 1$  zeroes at the top. For example, with  $m = 3$  factors we have  $\mathbf{V}(\boldsymbol{\eta}) = \mathbf{\Psi} = \mathbf{I}$  which imposes  $\frac{3(3+1)}{2} = 6$  restrictions. We need  $\frac{3(3-1)}{2} = 3$  more restrictions to make  $m^2 = 9$  all together. In that case we can arbitrarily build  $\mathbf{\Lambda}$  as

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \\ \cdots & \cdots & \cdots \\ \lambda_{p1} & \lambda_{p2} & \lambda_{p3} \end{bmatrix}.$$

The  $\hat{\chi}^2$  tests the null hypothesis that  $\Sigma$  stems from 3 factors vs. the alternative that  $\Sigma$  is arbitrary, or it stems from as many factors as there are variables,  $p$ . Once dimensionality has been statistically determined, rotation may generate hypotheses (for later confirmation) regarding the nature of the dimensions. If there are more than  $m^2$  fixed constants rotation is not possible and meaning that the factor space has been restricted.

#### 11.4 Other Methods of Factor Extraction

In addition to ML factor analysis, we have an approach called MINRES which seeks to minimize the residual or the difference between the predicted correlations in  $\hat{\mathbf{R}}$  and the actual correlations in  $\mathbf{R}$ . The objective function is then

$$f = \sum_{j=2}^p \sum_{k=1}^j \left( r_{jk} - \sum_{l=1}^m \lambda_{jl} \lambda_{kl} \right)^2. \quad (11.7)$$

The reader will note that the component  $\sum_{l=1}^m \lambda_{jl} \lambda_{kl}$  is a scalar version of the inner product of the  $j$ th and  $k$ th rows of  $\Lambda$ , since it would be those two rows used to reproduce element  $r_{jk}$  in Equation (11.3).

*Canonical factoring* maximizes the canonical correlation (see Section 8.18) between the factors and the variables while *Image factoring* and *Alpha factoring* are based on the notion that items measuring any particular factor are sampled from some population of items that might be chosen. These techniques are discussed in Harman (1976).

#### 11.5 Factor Rotation

After the factors have been extracted, whether this be by ML Factor Analysis, Principal Factors, or one of the other techniques, it is possible to rotate the factor axes into a position of possible higher theoretical value. That this is possible can be easily proven by noting that the extraction of factors based on Equation (11.3) is essentially arbitrary. If I define an orthonormal matrix  $\mathbf{C}$  such that  $\mathbf{C}\mathbf{C}' = \mathbf{I}$ , I can always create a new loading matrix, call it  $\tilde{\Lambda}$ , as in

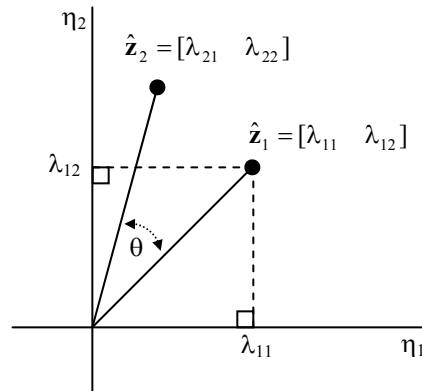
$$\tilde{\Lambda} = \Lambda\mathbf{C}$$

so that

$$\mathbf{R}^* = \tilde{\Lambda}\tilde{\Lambda}' + \Theta$$

$$= \Lambda\mathbf{C}\mathbf{C}'\Lambda + \Theta$$

which of course yields the original Equation (11.3). An orthonormal matrix like  $\mathbf{C}$  imposes a rigid rotation on axes which leaves angles and distances between points unchanged. The geometry of the situation might look like the figure below which shows two variables defined in a space with two factors.



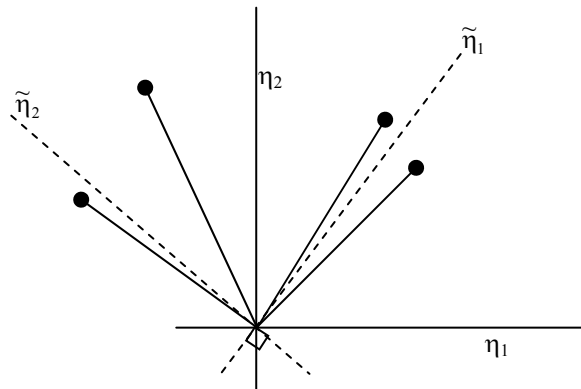
Looking at the Figure, the length of the vector  $\hat{\mathbf{z}}_j$  that corresponds to the (predicted) variable  $j$  is

$$\sqrt{\hat{\mathbf{z}}_j' \hat{\mathbf{z}}_j} = \sqrt{\sum_k \lambda_{jk}^2} = \sqrt{h_j^2}. \text{ Our factors are at right angles, which is to say uncorrelated. At this}$$

point, assuming we have extracted those two factors using Principal Factors, the position of the axes is in the arbitrary orientation of maximal variance. The loadings are now the coordinates of the variables on the axes formed by the factors. The predicted correlation between the two variables is

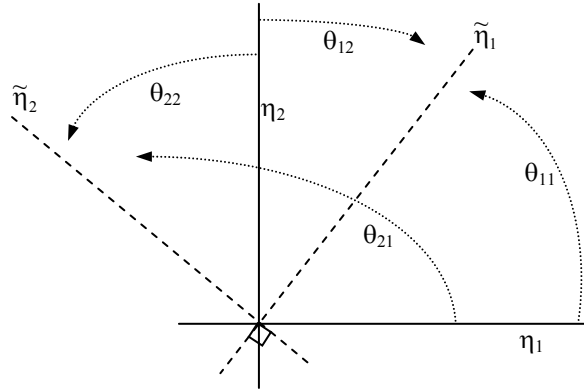
$$\hat{r}_{12} = \sqrt{h_1^2} \sqrt{h_2^2} \cos \theta. \quad (11.8)$$

In the next figure, we complicate things somewhat by having four variables appear.



All variables load on all original axes,  $\eta_1$  and  $\eta_2$ . However, the loadings or coordinates on the new axes,  $\tilde{\eta}_1$  and  $\tilde{\eta}_2$  will be different. Two of the variables will load heavily on  $\tilde{\eta}_1$  while the other

two will load on  $\tilde{\eta}_2$ . The cross loadings will be minimal which creates a much simpler  $\Lambda$  matrix in which the interpretation of the  $\eta$ 's will be facilitated. In order to rotate the original axes into the new positions, we will need a bit of trigonometry. Below we have labeled all of the angles between each original axis and each new one:



We can construct the orthonormal rotation matrix  $\mathbf{C}$  such that

$$\mathbf{C} = \begin{bmatrix} \cos \theta_{11} & \cos \theta_{12} \\ \cos \theta_{21} & \cos \theta_{22} \end{bmatrix}$$

and even though  $\theta_{12}$  is "reversed", since  $\cos \theta = \cos (360 - \theta)$ , it comes out the same. Note that the first subscript refers to the new axis and the second to the old. This concept works in spaces of arbitrary dimensionality. So how do we pick the angle of rotation? What constitutes a good orientation of the axes with the variables? What we are looking for is called *simple structure*. This is an idea due to Thurstone [summarized in Thurstone (1935)] who came up with three principles.

1. Each row of  $\Lambda$  should have at least one zero.
2. Each column of  $\Lambda$  should have at least  $m$  zeroes.
3. For every pair of columns of  $\Lambda$  there should be at least  $m$  variables with zeroes in one column but not in the other.

The most famous implementation of rotation to simple structure is Kaiser's *Varimax procedure* that maximizes the variance of the squared loadings within each column. The original formula, sometimes still called raw varimax, is to pick the rotation that maximizes the variance of the squared loadings in each column  $j$  of  $\Lambda$

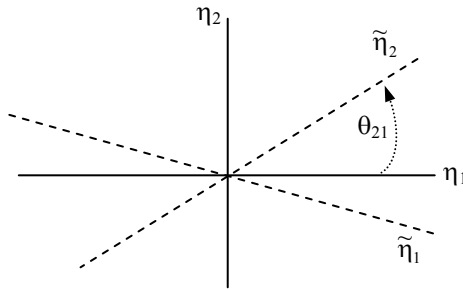
$$\left[ \sum_i^p (\lambda_{ij}^2)^2 - \frac{\left( \sum_i^p \lambda_{ij}^2 \right)^2}{p} \right] \frac{1}{p} \quad (11.9)$$

The formula widely used today (see Harman, 1976, pp. 290-1) weights each factor by the inverse of its total communality, but conceptually it follows the lines of the above equation. Other

approaches maximize the variance within each row (*Quartimax*), or equally between rows and columns (*Equimax*).

### 11.6 Oblique Rotation

Of course nothing guarantees that the factors that we see in marketing will be orthogonal. In order to create a rotation like the one pictured below,



we could use a transformation matrix

$$\mathbf{C} = \begin{bmatrix} \cos \theta_{11} & \cos \theta_{12} & \cdots & \cos \theta_{1m} \\ \cos \theta_{21} & \cos \theta_{22} & \cdots & \cos \theta_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \cos \theta_{m1} & \cos \theta_{m2} & \cdots & \cos \theta_{mm} \end{bmatrix}$$

which would take the old axes into a new set. Note that in this case  $\mathbf{C}\mathbf{C}' \neq \mathbf{I}$ . The elements of  $\mathbf{C}$  are direction cosines, and the sum of cross-products of direction cosines gives the cosine of the angle between the two vectors which according to Equation (11.8) is the same thing as a correlation. Thus we have for the correlations between the new factors:

$$\tilde{\Psi} = \mathbf{C}\mathbf{C}' \quad (11.10)$$

The new loadings, in  $\tilde{\Lambda}$ , can be inferred from the fact that since

$$\hat{\mathbf{R}} = \mathbf{\Lambda}\mathbf{\Lambda}' + \Theta$$

it must also be the case that

$$\mathbf{R} = \tilde{\mathbf{\Lambda}}\mathbf{C}\mathbf{C}'\tilde{\mathbf{\Lambda}}' + \Theta$$

so that obviously  $\mathbf{\Lambda} = \tilde{\mathbf{\Lambda}}\mathbf{C}$ , which then implies further that  $\mathbf{C}^{-1}\mathbf{\Lambda} = \tilde{\mathbf{\Lambda}}$ . When factors are orthogonal, and we have standardized both the variables and the factors to be z-scores, the loadings in the  $\mathbf{\Lambda}$  matrix can also be interpreted as correlations between the variables and the factors. When we have non-orthogonal factors, this is no longer so. We can, however, calculate these correlations, known as the *factor structure*, using



$$\mathbf{S} = \tilde{\mathbf{\Lambda}}\Psi. \quad (11.11)$$

There are a number of analytic techniques available to perform oblique rotation including *Oblimax*, *Quartimin*, *Oblimin* and *Promax*.

#### *References*

Harman, Harry H. (1976) *Modern Factor Analysis. Third Edition*. Chicago: University of Chicago Press.

Jöreskog, Karl Gustav (1967) Some Contributions to Maximum Likelihood Factor Analysis. *Psychometrika*. 32 (December), 443-82.

Lawley, D.N. and A. E. Maxwell (1963) *Factor Analysis as a Statistical Method*. London: Butterworth.

Thurstone, L. L. (1935) *The Vectors of Mind*. Chicago: University of Chicago Press.