Measurement in the Social Sciences: Principal Components and Factor Analysis (2)

Dave Armstrong

University of Oxford
Department of Politics and International Relations
Center for Research Methods in the Social Sciences

t: 01865 285956
e: david.armstrong@politics.ox.ac.uk
w: http://users.ox.ac.uk/∼polf0104

Office: 1101 Manor Road Building (Politics Dept)
Office Hours: TWTh 1400-1500

May 8, 2008

Where we are...

- Last week, we left off with the following model:

$$x_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + U_j$$

- This leaves us with a factor pattern matrix:

$$\begin{array}{c|cccc|c}
 & & & & & \sum_{j=1}^{k} a_{jm}^2 \\
\hline
 & a_{1,1} & a_{1,2} & \cdots & a_{1,m} & h_1^2 \\
 & a_{2,1} & a_{2,2} & \cdots & a_{2,m} & h_2^2 \\
 & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & a_{k,1} & a_{k,2} & \cdots & a_{k,m} & h_k^2 \\
\hline
\sum_{l=1}^{m} a_{kl}^2 & \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_m^2 &
\end{array}$$

- The factor analysis model can be described as a complicated regression problem where the observed variables are dependent variables and the factors are independent variables.

Communality

- Communality is the amount (proportion) of a variable's variance that it shares with the other variables. The common variance we are trying to explain.
- Let's think of a very simple factor model:

$$y = b_1 x_1 + \varepsilon$$

  - Here, we are not saying that all of the variance in y can be explained by x. Rather, we're saying that there is some part of y that varies systematically with x and some part that is unique to y and unrelated to x.
- Now, transform the above equation into the notation for our factor model:

$$x_1 = a_{11}F_1 + U_1$$

  Here, we're saying the same thing - that part of the variance of x_1 is unique to x_1 and has nothing to do with F_1.

A Short Digression

- To estimate the factor model, we will be basically trying to reproduce a correlation matrix. We will talk more about the implications of this later in the class.
- We will put our estimates of communality on the diagonal of the correlation matrix. Once we change these elements from 1's to something else, we will rename the matrix from $\mathbf{R}_{xx}$ to $\tilde{\mathbf{R}}_{xx}$.
- These new diagonal elements of $\tilde{\mathbf{R}}_{xx}$ will be less than one (we couldn't explain more than all of a variable's variance).

Communality (2)

- If you find that previous argument convincing, then we need to think of a way to estimate communality. There are a few common options:
  - $h_j^2 = 1$: This is called the "principal components factor model". Why is this? Well, when we did principal components analysis, we tried to account for all of the variance in each of the variables and so it sort of makes sense to call the factor model that attempts the same thing the "principal components factor model," but it is quite confusing. Also, this is almost certainly an *overestimate* of a variable's communality.
  - $h_j^2 = R_{-j}^2$: This is the "squared multiple correlation" - the R^2 of a regression with x_j as the dependent variable and all of the rest of the observed variables as the independent variables. The SMC is (in a properly specified factor model) the lower bound of the true, but unknown communality.

Communality (3)

- More Options
  - $h_j^2 =$ Reliability: If we know a variable's reliability, we could use this as the communality, the amount of variance shared with the other variables. Remember, Cronbach's α is not a variable's reliability, it is a unidimensional estimate of the variable's reliability. If we have reliability, it is theoretically the upper bound of the true, but unknown communality.
- Once we decide on a strategy, we can then decide if we want to iteratively make our estimates better or not, this applies to the $h_j^2 = R_{-j}^2$ strategy.
- The "Uniqueness" just equals $1 - h_j^2$. Remember, the Uniqueness and Communality sum to 1. All of a variable's variance is a function of 1) the part we can explain, plus 2) the part we can't explain.

What we Want

- $\mathbf{A}$ (Factor Pattern Matrix), the coefficients relating the unobserved factors to the observed variables.
- h_j^2 (Communality), we need to supply initial estimates and then we can either iteratively improve them or not.
- λ_m^2 (Eigenvalue), we will also want to know the amount of total variance explained by each of the factors.
- $\hat{F}_{im}$, estimates of the unobserved variables for each observation.

The Fundamental Theorem of Factor Analysis

$$
\begin{aligned}
\mathbf{X} &= \mathbf{AF} + \mathbf{U} \\
\mathbf{XX}' &= (\mathbf{AF} + \mathbf{U})(\mathbf{AF} + \mathbf{U})' \\
&= (\mathbf{AF} + \mathbf{U})(\mathbf{F}'\mathbf{A}' + \mathbf{U}') \\
&= \mathbf{AFF}'\mathbf{A}' + \mathbf{AFU}' + \mathbf{UF}'\mathbf{A}' + \mathbf{UU}' \\
E(\mathbf{XX}') &= \mathbf{A}E(\mathbf{FF}')\mathbf{A}' + \mathbf{A}E(\mathbf{FU}') + E(\mathbf{UF}')\mathbf{A}' + E(\mathbf{UU}') \\
\mathbf{R}_{XX} &= \mathbf{AR}_{FF}\mathbf{A}' + \mathbf{A}\underbrace{\mathbf{R}_{FU}}_{0} + \underbrace{\mathbf{R}_{UF}}_{0}\mathbf{A}' + \mathbf{R}_{UU} \\
&= \mathbf{AR}_{FF}\mathbf{A}' + \mathbf{R}_{UU}
\end{aligned}
$$

Here, $\mathbf{R}_{FF}$ is the correlation between the factors, which we'll assume for the moment is $\mathbf{I}$, a matrix with 1 on the diagonal and 0 elsewhere and $\mathbf{R}_{UU}$ we'll call the variance-covariance matrix of the uniquenesses, with unique variance on the diagonal and 0 elsewhere.

The Fundamental Theorem of Factor Analysis (2)

$$\begin{aligned}\mathbf{R}_{XX} &= \mathbf{AIA}' + \mathbf{R}_{UU}\\ \mathbf{R}_{XX} - \mathbf{R}_{UU} &= \mathbf{AIA}'\\ \tilde{\mathbf{R}}_{XX} &= \mathbf{AA}'\end{aligned}$$

What we're saying here is that we can break the adjusted correlation matrix $\tilde{\mathbf{R}}_{XX}$ into the product of something $\mathbf{A}$ and itself. This is where we get the term *Factor* analysis, because we're essentially *factoring* a matrix.

Factor Analysis and SVD

Let's go back to our old friend SVD, we know that we can decompose a data matrix into constituent parts:

$$\mathbf{X} = \mathbf{UDV}'$$

We also know that a correlation matrix is the product of two matrices scaled by $\frac{1}{n-1}$:

$$\mathbf{R}_{XX} = \mathbf{X}'\mathbf{X}(n-1)^{-1}$$

So, now we can substitute the first piece of information into the second:

$$\mathbf{R}_{XX} = (\mathbf{UDV}')'(\mathbf{UDV}')(n-1)^{-1}$$

Solving the Factor Model with SVD

$$\begin{aligned}\mathbf{R}_{XX} &= (\mathbf{UDV}')'(\mathbf{UDV}')(n-1)^{-1}\\ &= (\mathbf{VD}'\mathbf{U}'\mathbf{UDV}')(n-1)^{-1}\\ &= (\mathbf{VD}'\mathbf{IDV}')(n-1)^{-1}\\ &= \mathbf{VD}'(n-1)^{-1}\mathbf{DV}'\\ &= \left(\mathbf{VD}'(n-1)^{-\frac{1}{2}}\right)\left((n-1)^{-\frac{1}{2}}\mathbf{DV}'\right)\end{aligned}$$

Now, if we rename $\mathbf{D}'(n-1)^{-\frac{1}{2}} = \mathbf{\Lambda}$, then we have:

$$\mathbf{R}_{XX} = \underbrace{\mathbf{V\Lambda}}_{\mathbf{A}}\underbrace{\mathbf{\Lambda V}'}_{\mathbf{A}'}$$

Here, $\mathbf{\Lambda}^2$ is the diagonal matrix of eigenvalues, so if we want to create $\mathbf{\Lambda}$, we need to create a diagonal matrix of the square root of the eigenvalues.

Extraction Methods

- When we estimate a factor analysis, we have to choose an "extraction method" which basically relates to what we want to do with the communalities and how we want to estimate the the factor pattern matrix. In Stata, there are four options for this:
  1. Principal Factor (`pf`), the default - replaces diagonal elements of the correlation matrix with SMC's (R^2_{-j}'s) and factors that matrix to get the solution.
  2. Iterated Principal Factor (`ipf`), starts the same as `pf`, but iteratively improves communality estimates. Remember, if the model is appropriate / properly specified, then SMC's are the lower bound (and as such, almost certainly an underestimate) of the true communality.
  3. Principal Component Factors (`pcf`) also might be called a *nil uniqueness* model, leaves the ones on the main diagonal of the correlation matrix and factors that matrix.
  4. Maximum Likelihood (`ml`) does not employ the SVD, rather it maximizes $\Pr(\tilde{\mathbf{R}}_{XX}|\mathbf{A})$.

Rotation (1)

- Factor analysis is only *identified* up to a rotation. This means that any orientation of the factors in m dimensional space that preserves the lengths of and angles between all of the variable vectors will give an equally good reproduction of the correlation matrix.
- If we have a factor pattern matrix $\mathbf{\Lambda}$, then we can pick some matrix $\mathbf{T}$ such that $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}$. Where, in the two-factor solution:

$$\mathbf{T} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

  where α is the number of degrees you want to rotate the solution.

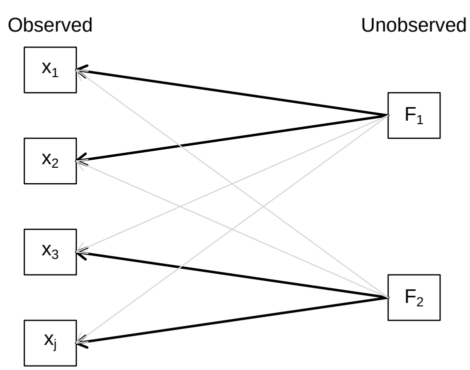
Rotation (2)

If every solution is equally "good" how do we choose discriminate between the solutions?

- We are looking for something called "simple structure". This mean that :
  1. As few factors as possible influence any one variable, and
  2. Each factor influences as few variables as possible.
- This means we want factor pattern coefficients to be as close to zero or one as possible.

Simple Structure Solution

Figure: Example of a Simple Structure Solution



Finding Simple Structure

- While this idea of simple structure makes intuitive sense, how do we find a rotation matrix that produces simple structure?
- Basically, for each column of the factor pattern matrix, we want to find the matrix $\mathbf{T}$ which maximizes the variance within the column. Why do this?

	$\mathbf{A_1}$	$\mathbf{A_2}$	$\mathbf{A_3}$
a_{11}	.5	.7	1
a_{21}	.5	.7	1
a_{31}	.5	.3	0
a_{41}	.5	.3	0
σ_A	0	0.23	0.58

- This idea is operationalized in the *varimax* rotation which is available in Stata and SPSS. This scheme is an *orthogonal* rotation, which means after rotating, the factors remain uncorrelated.

More on Rotation

- Both the original solution and the varimax rotation were orthogonal solutions, ones that maintain the zero correlation between all of the factors.
- We might want to relax this assumption. It is not the case in the real world that everything we would want to model with latent variables are uncorrelated.
  - This idea is operationalized in the *promax* solution in SPSS and Stata.
- The worked example will examine both types of rotation and offer some advice on choosing between them.

Communality in Obliquely Rotated Factor Solutions

- One difference that an oblique rotation induces is that the communality is no longer the sum of the squared factor pattern coefficients. Instead, it is the following:

$$h_j^2 = \sum_{l=1}^{m} a_{jl}^2 + \sum_{l=1}^{m-1} \sum_{k=l+1}^{m} 2a_{jl}a_{jk}r_{F_lF_k}$$

- For example, if we have the following model: $Z_1 = a_{11}F_1 + a_{12}F_2 + U_1$, then

$$h_1^2 = a_{11}^2 + a_{12}^2 + 2a_{11}a_{12}r_{F_1F_2}$$

Factor Scores

- Generally, we are doing this not only to understand the structure of the data, but also to get estimates of the m-dimensional representation of our k-dimensional data matrix.
- This was easy in PCA, but is less easy in Factor Analysis. Think back to the equation form of FA:

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m$$

- It's nice that we know $\mathbf{A}$, but what we really want is a matrix B with elements b_{jm} such that:

$$\hat{F}_m = X_1b_{1m} + X_2b_{2m} + \cdots + X_kb_{km}$$

  We need this because the X variables are the only information we have to identify the $\hat{F}$'s.

Factor Scores (2)

- This is a pretty complicated problem and in a radical departure from the previous slides, I won't bore you with the details, but we can obtain the matrix of scoring coefficients as follows:

$$\mathbf{B} = \mathbf{R}_{XX}^{-1}\mathbf{A}\mathbf{R}_{FF}$$

  where $\mathbf{R}_{XX}$ is the correlation matrix of the observed variables, $\mathbf{A}$ is the factor pattern matrix and $\mathbf{R}_{FF}$ is the correlation matrix between the factors.
- I think this gives us everything we need from our factor solution.

See Worked Examples

Why are PCA and PCF different?

- If PCA and PCF are both decomposing the correlation matrix with ones on the diagonal, why do they get different results?
  - Remember what each is doing:

$$\begin{aligned} \text{PCA}: &\quad C_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1k}X_1 \\ \text{PCF}: &\quad X_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1m}F_m \end{aligned}$$

  - In PCA $\mathbf{A} = \mathbf{V}$, but in PCF, we found that $\mathbf{A} = \mathbf{V\Lambda}$. So, if we multiply the PCF solution by $\mathbf{\Lambda}^{-1}$, we would get back to the PCA result.
- The correlation between the factor scores from PCF and the components from PCA is 1. So, while the models are different, the substantive solutions are similar.

Conclusions

- As much as I've tried to highlight the theoretical differences between these models, they all produce roughly similar results. You can take this one of two ways:
  - It doesn't matter which model I use, so I should use the easiest one.
  - It doesn't matter which model I use, so I should use the one that makes the most sense theoretically.
- I hope you'll gravitate to the latter of these two ideas. To me, the iterated principal factor model makes the most sense because it seems the most flexible with respect to communalities - that is, it doesn't make unrealistic assumptions.