Principal Factor Solution

• Recall the Orthogonal Factor Model
  \[ X = LF + \epsilon \]
  which implies
  \[ \Sigma = LL' + \Psi. \]

• The \( m \)-factor Principal Component solution is to approximate \( \Sigma \) (or, if we standardize the variables, \( R \)) by a rank-\( m \) matrix using the spectral decomposition
  \[ \Sigma = \lambda_1 e_1 e_1' + \cdots + \lambda_m e_m e_m' + \lambda_{m+1} e_{m+1} e_{m+1}' + \cdots + \lambda_p e_p e_p'. \]

• The first \( m \) terms give the best rank-\( m \) approximation to \( \Sigma \).
• We can sometimes achieve higher *communalities* (\(= \text{diag}(LL')\)) by either:
  
  – specifying an initial estimate of the communalities
  
  – iterating the solution
  
  or both.

• Suppose we are working with \(R\). Given initial communalities \(h_{i}^{*2}\), form the *reduced correlation matrix*

\[
R_r = \begin{bmatrix}
    h_{1}^{*2} & r_{1,2} & \cdots & r_{1,p} \\
    r_{2,1} & h_{2}^{*2} & \cdots & r_{2,p} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{p,1} & r_{p,2} & \cdots & h_{p}^{*2}
\end{bmatrix}.
\]
• Now use the spectral decomposition of $R_r$ to find its best rank-$m$ approximation

$$R_r \approx L_r^* L_r'^*.$$ 

• New communalities are

$$\tilde{h}_i^*{}^2 = \sum_{j=1}^{m} l_{i,j}^*{}^2.$$ 

• Find $\Psi$ by equating the diagonal terms:

$$\tilde{\psi}_i^* = 1 - \tilde{h}_i^*{}^2,$$

or

$$\tilde{\Psi}^* = I - \text{diag} \left( L_r^* L_r'^* \right).$$
• This is the Principal Factor solution.

• The Principal Component solution is the special case where the initial communalities are all 1.

• In proc factor, use method = prin as for the Principal Component solution, but also specify the initial communalities:
  
  – the priors = … option on the proc factor statement specifies a method, such as squared multiple correlations (priors = SMC);
  
  – the priors statement provides explicit numerical values.
SAS program and output:

```sas
proc factor data = all method = prin priors = smc;
  title 'Method = Principal Factors';
  var cvx -- xom;
```

In this case, the communalities are smaller than for the Principal Component solution.
• Other choices for the `priors` option include:

  - **MAX** ⇒ maximum absolute correlation with any other variable;

  - **ASMC** ⇒ *Adjusted SMC* (adjusted to make their sum equal to the sum of the maximum absolute correlations);

  - **ONE** ⇒ 1;

  - **RANDOM** ⇒ uniform on (0, 1).
Iterated Principal Factors

- One issue with both Principal Components and Principal Factors:
  - if $S$ or $R$ is exactly in the form $LL' + \Psi$ (or, more likely, approximately in that form), neither method produces $L$ and $\Psi$ (unless you specify the true communalities).

- Solution: iterate!
  - Use the new communalities as initial communalities to get another set of Principal Factors.
  - Repeat until nothing much changes.
In `proc factor`, use `method = prinit`; may also specify the initial communalities (default = `ONE`).

**SAS program and output:**

```sas
proc factor data = all method = prinit;
  title 'Method = Iterated Principal Factors';
  var cvx -- xom;
```

The communalities are still smaller than for the Principal Component solution, but larger than for Principal Factors.
Likelihood Methods

- If we assume that \( X \sim N_p(\mu, \Sigma) \) with \( \Sigma = LL' + \Psi \), we can fit by maximum likelihood:
  
  - \( \hat{\mu} = \bar{x} \);
  
  - \( L \) is not identified without a constraint (uniqueness condition) such as
    \[
    L'\Psi^{-1}L = \text{diagonal};
    \]
    
  - still no closed form equation for \( \hat{L} \); numerical optimization required.
• We can also test hypotheses about $m$ with the likelihood ratio test (Bartlett’s correction improves the $\chi^2$ approximation):

$- \ H_0 : m = m_0; \ H_A : m > m_0;$

$- \ -2 \times \log \text{ likelihood ratio } \sim \chi^2 \text{ with } \frac{1}{2} \left( (p - m_0)^2 - p - m_0 \right) \text{ degrees of freedom.}$

$- \ \text{Degrees of freedom } > 0 \iff m_0 < \frac{1}{2} \left( 2p + 1 - \sqrt{8p + 1} \right).$

• E.g. for $p = 5$, $m_0 < 2.298 \Rightarrow m_0 \leq 2$:

<table>
<thead>
<tr>
<th>$p$</th>
<th>$m_0$</th>
<th>degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
• In **proc factor**, use `method = ml`; may also specify the initial communalities (default = SMC); SAS program and **output**:

```sas
proc factor data = all method = ml;
  var cvx -- xom;
  title 'Method = Maximum Likelihood';

proc factor data = all method = ml heywood plot;
  var cvx -- xom;
  title 'Method = Maximum Likelihood with Heywood fixup';

proc factor data = all method = ml ultraheywood plot;
  var cvx -- xom;
  title 'Method = Maximum Likelihood with Ultra-Heywood fixup';
```
• Note that the iteration can produce communalities $> 1$!

• Two fixes:

  – use the Heywood option on the `proc factor` statement; caps the communalities at 1;

  – use the UltraHeywood option on the `proc factor` statement; allows the iteration to continue with communalities $> 1$. 
Scaling and the Likelihood

- If the maximum likelihood estimates for a data matrix $X$ are $\hat{L}$ and $\hat{\Psi}$, and

$$Y_{n \times p} = X_{n \times p} D_{n \times p \times p}$$

is a scaled data matrix, with the columns of $X$ scaled by the entries of the diagonal matrix $D$, then the maximum likelihood estimates for $Y$ are $D\hat{L}$ and $D^2\hat{\Psi}$.

- That is, the mle’s are invariant to scaling:

$$\hat{\Sigma}_Y = D\hat{\Sigma}_X D.$$
• Proof: $L_Y(\mu, \Sigma) = L_X(D^{-1}\mu, D^{-1}\Sigma D^{-1})$.

• No distinction between covariance and correlation matrices.
Weighting and the Likelihood

- Recall the uniqueness condition
  \[ L'\Psi^{-1}L = \Delta, \text{diagonal}. \]

- Write
  \[
  \Sigma^* = \Psi^{-\frac{1}{2}} \Sigma \Psi^{-\frac{1}{2}}
  = \Psi^{-\frac{1}{2}} (LL' + \Psi) \Psi^{-\frac{1}{2}}
  = \left( \Psi^{-\frac{1}{2}} L \right) \left( \Psi^{-\frac{1}{2}} L \right)' + I_p
  = L^* L^{*'} + I_p.
  \]

- \( \Sigma^* \) is the weighted covariance matrix.
• Here

\[ L^* = \Psi^{-\frac{1}{2}}L \]

and

\[ L^*'L^* = L'\Psi^{-1}L = \Delta. \]

• Note:

\[ \Sigma^*L^* = L^*L^*'L^* + L^* \]
\[ = L^*\Delta + L^* \]
\[ = L^*(\Delta + I_m) \]

so the columns of \( L^* \) are the (unnormalized) eigenvectors of \( \Sigma^* \), the weighted covariance matrix.
• Also

\[(\Sigma^* - I_p)L^* = L^* \Delta\]

so the columns of \(L^*\) are also the eigenvectors of

\[\Sigma^* - I_p = \Psi^{-\frac{1}{2}}(\Sigma - \Psi)\Psi^{-\frac{1}{2}},\]

the weighted reduced covariance matrix.

• Since the likelihood analysis is transparent to scaling, the weighted reduced covariance matrix is the same as the weighted reduced covariance matrix.