Suppose that the common factors and the specific factors are jointly normally distributed.

Moreover, the joint distribution of \((\mathbf{X} - \mu)\) and \(\mathbf{F}\) is \(N_{m+p}(0, \Sigma + \Psi)\), where

\[
\begin{bmatrix}
\mathbf{X} - \mu \\
\mathbf{F}
\end{bmatrix}
\sim N_{m+p}(0, \Sigma + \Psi)
\]

the conditional distribution of \(\mathbf{F} | \mathbf{x}\) is multivariate normal with

\[
E(\mathbf{F} | \mathbf{x}) = \mathbf{L} \Sigma^{-1} (\mathbf{x} - \mu) = \mathbf{L}' (\mathbf{L} \Sigma^{-1} \mathbf{L} + \Psi)^{-1} (\mathbf{x} - \mu)
\]

covariance = \(\text{Cov}(\mathbf{F} | \mathbf{x}) = \mathbf{I} - \mathbf{L} \Sigma^{-1} \mathbf{L} = \mathbf{I} - \mathbf{L}' (\mathbf{L} \Sigma^{-1} \mathbf{L} + \Psi)^{-1} \mathbf{L}\)

the \(j^{\text{th}}\) factor score vector is given by

\[
\hat{f}_j = \mathbf{L}' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{x}) = \mathbf{L}' (\hat{\mathbf{L}} L + \hat{\Psi})^{-1} (\mathbf{x} - \bar{x})
\]

Denote the scores generated by the weighted least squares by \(\hat{f}_j^W\) and those by the regression method by \(\hat{f}_j^R\). Because

\[
\hat{f}_j^W = (\mathbf{I} + \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} \mathbf{L}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} (\mathbf{x} - \bar{x})
\]

For maximum likelihood estimates \((\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} = \hat{\Lambda}^{-1}\) and if the elements of this diagonal matrix are close to zero, the regression and generalized least squares methods will give nearly the same factor scores.

In an attempt to reduce the effect of a (possible) incorrect determination of the number of factors, \(\hat{\Psi}\) is often used for \(\hat{\Sigma}\), rather than \(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}\), i.e.,

\[
\hat{f}_j = \mathbf{L}' \hat{\Sigma}^{-1} (\mathbf{x} - \bar{x})
\]

may not be a good approximation.

or, if a correlation matrix is factored

\[
\hat{f}_j = \hat{\mathbf{L}}_{\chi} \mathbf{R}^{-1} (\mathbf{x} - \bar{x})
\]

where \(\mathbf{z}_j = \mathbf{D}^{-1/2} (\mathbf{x}_j - \bar{x})\)

If rotated loadings \(\hat{\mathbf{L}}_{\chi} = \hat{\mathbf{L}} \mathbf{T}\) are used in place of the original loadings, the subsequent factor scores, \(\hat{f}_j\), are related to \(\hat{f}_j\) by \(\hat{f}_j^R = \mathbf{T} \hat{f}_j^W\).

A strategy for factor analysis

1. perform a principal component factor analysis
2. perform a maximum likelihood factor analysis
3. compare the solutions obtained from the 2 factor analyses
4. repeat the 1st 3 steps for other number of common factors \(m\)
5. for large data sets, split them in half and perform a FA on each part

Reading: Textbook, 9.4, 9.5, 9.6