THE UNIVERSITY OF MINNESOTA

Statistics 5401 November 14, 2005

Some Methods of Factor Analysis Extraction

The factor analytic model

There are several methods used to estimate factor loadings and communalities in factor analysis. The assumed model for the observable random vector $\mathbf{x} = [X_1, X_2, ..., X_p]'$ is the so-called *factor analytic model*

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\,\mathbf{f} + \boldsymbol{\varepsilon},$$

where

$\mathbf{x}$, $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_p]'$, and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_p]'$ are p by 1 vectors;

$\mathbf{f} = [f_1, f_2, ..., f_m]'$ is a m by 1 vector of random unobservable *common factors* f_j, $j = 1, ..., m$;

$\mathbf{L} = [\ell_{ij}]$ is a p by m matrix, the *loading matrix* or matrix of *factor loadings*.

The elements ε_i of ε are the *unique factors* and are uncorrelated, that is, $V(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi} = \text{diag}[\psi_1,...,\psi_p]$ is diagonal. The ε_i's are also assumed to be independent of $\mathbf{f}$.

When $V(\mathbf{f}) = I_m$, the model is the *orthogonal factor* model. This is assumed below.

The quantity ψ_i is the *uniqueness* or *specific variance* associated with X_i. It measures the contribution to $\sigma_{ii} = V(X_i)$ of the unique factor ε_i "special" to X_i that is not shared by other variables.

The quantity $h_i^2 \equiv \sigma_{ii} - \psi_i$ is the *communality* associated with x_i. It measures the contribution to σ_{ii} that is attributable to or explained by the common factors.

Variance matrices with factor analytic form

The orthogonal factor analytic model implies that $\boldsymbol{\Sigma} = V(\mathbf{x})$ is

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} = \mathbf{V} + \boldsymbol{\Psi},$$

where the p by p matrix $\mathbf{V} = \mathbf{L}\mathbf{L}'$ has rank m. A variance matrix $\boldsymbol{\Sigma}$ that can be expressed in this way, that is as the sum of a *rank m matrix* $\mathbf{V}$ with $m < p$ and *a diagonal matrix* $\boldsymbol{\Psi}$, is said to have *factor analytic form* with m factors. When $V(\mathbf{f}) = \boldsymbol{\Gamma} \neq I_m$, $\mathbf{V} = \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}'$.

In summation notation, the factor analytic model is

$$X_i = \mu_i + \sum_{1 \le j \le m} \ell_{ij} f_j + \varepsilon_i,\ i = 1, \ldots, p.$$

This has the appearance of p multiple regressions of the elements of $\mathbf{x}$ as dependent variables on the m factors playing the role of independent variables. But it is profoundly different from the multiple regression model because the "independent variables" are not directly observable.

You can summarize the entire factor analytic form for $\mathbf{\Sigma}$ using subscript and summation notation as

$$V[X_i] = \sigma_{ii} = h_i^2 + \psi_i,$$

$$h_i^2 = \sum_{1 \le j \le m} \ell_{ij}^2 V[f_j] = \sum_{1 \le j \le m} \ell_{ij}^2,$$

$$Cov[X_{i_1}\ X_{i_2}] = \sum_{1 \le j \le m} \ell_{i_1 j} \ell_{i_2 j} V[f_j] = \sum_{1 \le j \le m} \ell_{i_1 j} \ell_{i_2 j},\ i_1 \ne i_2$$

Because of the last line, *all* the correlation among the variables results from the factors they have in common. This is closely related to one of the goals of factor analysis - to explain correlations between variables in terms of common influences. Moreover, the variance of each X_i is split into a part h_i^2 derived from the common factors and a part ψ_i that is unique to X_i. The larger h_i^2 is relative to ψ_i, the more completely the behavior of X_i can be explained in terms of the common factors.

Correlation matrices with factor analytic form

Let $\mathbf{\Delta} \equiv \text{diag}[1/\sqrt{\sigma_{11}}, \ldots, 1/\sqrt{\sigma_{pp}}]$. Then the correlation matrix of $\mathbf{x}$ is

$$\mathbf{\rho} = [\rho_{ij}] = \mathbf{\Delta \Sigma \Delta}.$$

When $\mathbf{\Sigma}$ has factor analytic form, so does $\mathbf{\rho}$:

$$\mathbf{\rho} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}} = \tilde{\mathbf{V}} + \tilde{\mathbf{\Psi}}$$
$$\tilde{\mathbf{V}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}',\ \tilde{\mathbf{\Psi}} = \text{diag}[\tilde{\psi}_1, \ldots, \tilde{\psi}_m]$$

where $\tilde{\mathbf{V}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$ has rank m and $\tilde{\mathbf{\Psi}} = \text{diag}[\tilde{\psi}_1, \ldots, \tilde{\psi}_m]$ is diagonal, with

$\tilde{\mathbf{L}} = [\tilde{\ell}_{ij}] = \mathbf{\Delta L}$ is p by m with $\tilde{\ell}_{ij} = \ell_{ij}/\sqrt{\sigma_{ii}}$, and $\tilde{\mathbf{\Psi}} = \mathbf{\Delta \Psi \Delta}$, $\tilde{\psi}_i = \psi_i/\sigma_{ii}$.

This displays a direct way to go from a factor analytic representation for $\mathbf{\Sigma}$ in terms of $\mathbf{L}$ and $\mathbf{\Psi}$ to a factor analytic representation for $\mathbf{\rho}$ in terms of $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{\Psi}}$.

Because

$$\mathbf{L} = \boldsymbol{\Delta}^{-1}\tilde{\mathbf{L}}, \text{ and } \boldsymbol{\Psi} = \boldsymbol{\Delta}^{-1}\tilde{\boldsymbol{\Psi}}\boldsymbol{\Delta}^{-1},$$

you can just as easily go in the other direction.

The quantities $\tilde{h}_i^2 = \sum_j \tilde{\ell}_{ij}^2 = h_i^2/\sigma_{ii}$ and $\tilde{\psi}_i = \psi_i/\sigma_{ii}$ based on the correlation matrix are also called communalities and uniquenesses.

Because $\tilde{\psi}_i + \tilde{h}_i^2 = 1$, the communality $\tilde{h}_i^2$ measures the *proportion* of the variance of X_i explainable by the common factors and is a little analogous to a multiple R^2 statistic in regression. The element $\tilde{\ell}_{ij}$ of $\tilde{\mathbf{L}}$ is the loading of standardized variable $X_i/\sqrt{\sigma_{ii}}$ on factor f_j.

Factor extraction

The first phase of factor analysis is *factor extraction*. This has the goal of finding some $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ such that $\hat{\boldsymbol{\Sigma}} \equiv \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ is a "good approximation" to the sample covariance matrix $\mathbf{S}$, or $\hat{\boldsymbol{\rho}} \equiv \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ is a "good approximation" to the sample correlation matrix $\mathbf{R}$.

It is more usual in factor analysis to work with $\mathbf{R}$. However, in the following I will phrase the development in terms of $\mathbf{S}$. The methods can be applied to approximating $\boldsymbol{\rho}$ by substituting $\boldsymbol{\rho}$ and $\mathbf{R}$ for $\boldsymbol{\Sigma}$ and $\mathbf{S}$ in what follows.

Non-uniqueness of factors or factor loadings

Suppose $\mathbf{L}^* \equiv \mathbf{LH}$, where $\mathbf{H}$ is a non-singular m by m matrix satisfying $\mathbf{H}'\mathbf{H} = \mathbf{HH}' = \mathbf{I}_m$ and $\det(\mathbf{H}) = 1$, (that is, $\mathbf{H}$ is *orthogonal*, and, because $\det(\mathbf{H}) = 1$, is a *rotation matrix*).

Then $\mathbf{L}^*\mathbf{L}^{*\prime} = \mathbf{LHH}'\mathbf{L}' = \mathbf{LL}' = \mathbf{V}$. This provides a factorization $\mathbf{L}^*\mathbf{L}^{*\prime}$ of $\mathbf{V} = \boldsymbol{\Sigma} - \boldsymbol{\Psi}$ that is different from $\mathbf{LL}'$.

There is a corresponding alternative representation of $\mathbf{x}$ in terms of factors:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\,\mathbf{f} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \mathbf{LHH}'\mathbf{f} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \mathbf{L}^*\mathbf{f}^* + \boldsymbol{\varepsilon},$$

where $\mathbf{f}^* = \mathbf{H}'\mathbf{f} \neq \mathbf{f}$ is a m by 1 vector of unobservable factors that is not the same as $\mathbf{f}$. The loading matrix for these factors is $\mathbf{L}^*$. A new factor f_j^* is linear combination of the original factors $f_1, \ldots, f_m$ with coefficients taken from column j of $\mathbf{H}$. Since $V[\mathbf{f}^*] = \mathbf{H}'V(\mathbf{f})\mathbf{H} = \mathbf{H}'\mathbf{I}_m\mathbf{H} = \mathbf{I}_m$ the new factors are also orthgonal. This means that the factor analytic decomposition of $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\Psi}$ and $\mathbf{L}$ (or of $\mathbf{x}$ in terms of $\mathbf{L}$, $\mathbf{f}$, and $\boldsymbol{\varepsilon}$) *is not unique*.

This lack of uniqueness of loadings and factors raises the question as to what, if anything, *is* unique about the factor analytic model. The answer, perhaps disappointingly, is that what is unique about the factor analytic model is the decomposition $\mathbf{\Sigma} = \mathbf{V} + \mathbf{\Psi}$ as the sum of a rank m matrix and a diagonal matrix. You can estimate $\mathbf{V}$ and $\mathbf{\Psi}$ from data in an unambiguous manner, but not $\mathbf{L}$, unless you introduce some further principles to eliminate the non-uniqueness.

The choice of which of the many factor loading matrices $\mathbf{L}$ that satisfy $\mathbf{V} = \mathbf{LL}'$ must be based on other considerations. If you consider the p rows of $\mathbf{L}$ as determining p points with coordinates $(\ell_{i1}, \ell_{i2},...,\ell_{im})$ in m-dimensional space, change from $\mathbf{L} \rightarrow \mathbf{LH}$ can be interpreted as a rigid rotation of points in that space. For this reason, the process of determining an $\mathbf{H}$ such that a new loading matrix $\mathbf{LH}$ has desirable properties is known as *rotation* of factor loadings.

For the purpose of an algorithm to do factor extraction, you can achieve uniqueness by imposing some purely *mathematical* constraints on $\mathbf{L}$. The most usual constraints are

(a) $\mathbf{L}'\mathbf{L}$ is diagonal (columns of $\mathbf{L}$ are orthogonal)
and
(b) $\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L}$ is diagonal (columns of $\mathbf{\Psi}^{1/2}\mathbf{L}$ orthogonal)

Neither restriction has any reasonable basis other than mathematical convenience, but you can use either to determine a specific p by m $\hat{\mathbf{L}}$ which in turn determines $\hat{\mathbf{V}} = \hat{\mathbf{L}}\hat{\mathbf{L}}'$. Once you have determined such a $\hat{\mathbf{L}}$ which satisfies either (a) or (b), you would normally rotate it by finding a rotation matrix $\mathbf{H}$ so that the estimated loading matrix $\hat{\mathbf{L}}^* \equiv \hat{\mathbf{L}}\mathbf{H}$ is "interpretable," a much less precise constraint than (a) or (b). The choice of $\mathbf{H}$ will not be further discussed in this handout.

Principal Component Factor Extraction

One option in many computer programs for factor analysis is to use principal component computations to do factor extraction.

Suppose $\mathbf{x}$ is a random vector with mean μ_x and covariance matrix $\mathbf{\Sigma}$ whose eigenvectors are $\mathbf{v}_1, ...,\mathbf{v}_p$, with associated eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Then the (population) principal components $z_j \equiv \mathbf{v}_j'(\mathbf{x} - \mu_x)$ have zero mean and covariance matrix $\mathbf{\Sigma}_z = \text{diag}[\lambda_1,...,\lambda_p]$.

You can exactly represent $\mathbf{x}$ in terms of the principal components as

$$\mathbf{x} = \mu_x + \sum_{1 \leq j \leq p} z_j \mathbf{v}_j = \mu_x + \sum_{1 \leq j \leq m} z_j \mathbf{v}_j + \sum_{m+1 \leq j \leq p} z_j \mathbf{v}_j.$$

An individual component X_k of $\mathbf{x}$ is then

$$X_k = \mu_k + \sum_{1 \le j \le m} z_j v_{kj} + \sum_{m+1 \le j \le p} z_j v_{kj}.$$

Similarly, you can decompose the variance matrix $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma} = \sum_{1 \le j \le p} \lambda_j \mathbf{v}_j \mathbf{v}_j' = \sum_{1 \le j \le m} \lambda_j \mathbf{v}_j \mathbf{v}_j' + \sum_{m+1 \le j \le p} \lambda_j \mathbf{v}_j \mathbf{v}_j'.$$

If you set $\tilde{\mathbf{L}} = [\sqrt{\lambda_1}\mathbf{v}_1, \ldots, \sqrt{\lambda_m}\mathbf{v}_m]$ and $\tilde{f}_j = z_j/\sqrt{\lambda_j}$, you get the model

$$\mathbf{x} = \boldsymbol{\mu}_x + \tilde{\mathbf{L}}\tilde{\mathbf{f}} + \boldsymbol{\varepsilon}, \quad \text{with } \tilde{\mathbf{f}} = [\tilde{f}_1, \ldots, \tilde{f}_m] \text{ and } \boldsymbol{\varepsilon} = \sum_{m+1 \le j \le p} z_j \mathbf{v}_j,$$

with $\mathbf{\Sigma}$ expressible as

$$\mathbf{\Sigma} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \mathbf{V}[\boldsymbol{\varepsilon}], \quad \mathbf{V}[\boldsymbol{\varepsilon}] = \sum_{m+1 \le j \le p} \lambda_j \mathbf{v}_j \mathbf{v}_j'.$$

Because $\tilde{\mathbf{V}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$ is a rank m matrix, this model bears a strong resemblance to the factor analytic model with factors $\tilde{\mathbf{f}}$. However, it is not an example of true factor analytic form because $\mathbf{V}[\boldsymbol{\varepsilon}]$ is *not* a diagonal matrix. Indeed it *cannot* be diagonal if $\tilde{\mathbf{L}}$ is defined in this way.

None the less, when $\lambda_{m+1}, \ldots, \lambda_p$ are all small relative to λ_m, this is a representation of $\mathbf{\Sigma}$ as a sum of a rank m matrix $\tilde{\mathbf{V}}$ and a *small* matrix $\mathbf{V}[\boldsymbol{\varepsilon}]$. If, in fact, $\mathbf{\Sigma}$ is of factor analytic form with $\mathbf{\Psi}$ small compared to $\mathbf{V} = \mathbf{L}\mathbf{L}'$, then you would expect that $\mathbf{V} \approx \tilde{\mathbf{V}}$. This is the rationale underlying the so called *principal components method of factor estimation*.

The *principal components method* first computes the eigenvectors $\hat{\mathbf{v}}_j$ and eigenvalues $\hat{\lambda}_j$, j = 1, ... ,p of the sample covariance matrix $\mathbf{S}$ (or more usually the correlation matrix $\mathbf{R}$). For a suitable choice of m (often chosen as the number of eigenvalues greater than some threshold such as1), the loading matrix $\mathbf{L}$ is estimated as

$$\hat{\mathbf{L}} = [\sqrt{\hat{\lambda}_1}\hat{\mathbf{v}}_1, \ldots, \sqrt{\hat{\lambda}_m}\hat{\mathbf{v}}_m] = [\hat{\boldsymbol{\ell}}_1, \ldots, \hat{\boldsymbol{\ell}}_m],$$

and then $\mathbf{\Psi}$ is estimated as $\hat{\mathbf{\Psi}} = \text{diag}[\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}']$[1], that is,

$$\hat{\psi}_i = s_{ii} - \sum_{1 \le j \le m} \hat{\ell}_{ij}^2 = \sum_{m+1 \le j \le p} \hat{\ell}_{ij}^2, \; i = 1, \ldots, p.$$

In other words, a diagonal matrix $\hat{\mathbf{\Psi}}$ is chosen so that the diagonal elements of $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ exactly match the diagonal elements of $\mathbf{S}$. The estimated communalities are $\hat{h}_i^2 = s_{ii} - \hat{\psi}_i = \sum_{1 \le j \le m} \hat{\ell}_{ij}^2$. Because of the of eigenvectors are orthogonal, the columns of $\hat{\mathbf{L}}$ are orthogonal. In fact, $\hat{\mathbf{L}}'\hat{\mathbf{L}} = \text{diag}[\hat{\lambda}_1, \ldots, \hat{\lambda}_p]$.

---

[1] *Notation*: If $\mathbf{A} = [a_{ij}]$ is square, the notation diag[$\mathbf{A}$] means the diagonal matrix $\text{diag}[a_{11}, \ldots, a_{pp}]$

The advantage of principal component factor extraction is that estimates of uniquenesses and the unrotated loadings are explicitly given in terms of the eigenvalues and eigenvectors of $\mathbf{S}$ or $\mathbf{R}$. You don't need any iteration or optimization.

The disadvantage is that, even in principle, the method is not actually estimating either $\mathbf{\Psi}$ or $\mathbf{V} = \mathbf{LL}'$. Even if $\mathbf{\Sigma} = \mathbf{V} + \mathbf{\Psi}$ and somehow $\mathbf{S}$ were exactly equal to $\mathbf{\Sigma}$, the principal component method would not reproduce $\mathbf{\Psi}$ and $\mathbf{V}$ although it might be close when $\sum_{m+1 \le i \le p}\lambda_i / \sum_{1 \le i \le p}\lambda_i$ is small.

Iterated Principal Factor Extraction

Iterated principal factor extraction is an iterative algorithm with each step rather similar to the principal component method. Here is a sketch of its rationale.

We first need some notation.

Let $\mathbf{\Sigma}$ be an arbitrary variance matrix (not necessarily of factor analytic form) and let $\mathbf{\Psi}$ be a diagonal matrix such that $\mathbf{\Sigma} - \mathbf{\Psi}$ is positive definite. Define a p by m "candidate" loading matrix $\mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma})$ associated with $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ as follows.

Suppose $\mathbf{u}_1,\ldots,\mathbf{u}_p$ are the (ordinary) eigenvalues and eigenvectors of $\mathbf{\Sigma} - \mathbf{\Psi}$ with associated eigenvalues $\delta_1 \ge \delta_2 \ge \ldots \ge \delta_p \ge 0$. As usual, assume the eigenvectors are scaled so as to be orthonormal ($\mathbf{u}_i'\mathbf{u}_i = 1$, $\mathbf{u}_i'\mathbf{u}_j = 0$, $i \ne j$).

Then $\mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma})$ is the p by m matrix

$$\mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma}) \equiv [\sqrt{\delta_1}\mathbf{u}_1, \ldots, \sqrt{\delta_m}\mathbf{u}_m].$$

The columns $\boldsymbol{\ell}_j$ of $\mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma})$ are orthogonal ($\boldsymbol{\ell}_j'\boldsymbol{\ell}_k = 0$, $j \ne k$), with squared length $\boldsymbol{\ell}_j'\boldsymbol{\ell}_j = \delta_j$, $j = 1,\ldots,m$.

Now suppose that, in fact, $\mathbf{\Sigma}$ does have factor analytic form with m factors and uniqueness matrix $\mathbf{\Psi}$. Then $\mathbf{V} = \mathbf{\Sigma} - \mathbf{\Psi}$ has rank m, $\delta_{m+1} = \ldots = \delta_p = 0$ and $\mathbf{V} = \sum_{1 \le j \le m}\delta_j\mathbf{u}_j\mathbf{u}'_j = \mathbf{LL}'$, where $\mathbf{L} = \mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma})$. In addition $\mathbf{L}'\mathbf{L} = \mathrm{diag}[\delta_1,\ldots,\delta_m]$ is diagonal. In fact, except for changes in the signs of its columns, $\mathbf{L} = \mathbf{L}_{m1}(\mathbf{\Psi},\mathbf{\Sigma})$ is the only p by m matrix satisfying $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$ such that $\mathbf{L}'\mathbf{L}$ is diagonal. Thus we have a method for finding $\mathbf{L}$ once we know $\mathbf{\Psi}$. It is unique only by virtue of satisfying the mathematical condition $\mathbf{L}'\mathbf{L}$ is diagonal (columns of $\mathbf{L}$ are orthogonal).

This suggests the following iterative approach to estimating $\mathbf{L}$:

Suppose $\hat{\mathbf{\Psi}}^{(i)}$ and $\hat{\mathbf{L}}^{(i)}$ are the current trial values of the estimated uniqueness

and loading matrices, respectively, at iteration i, so that $\hat{\mathbf{\Sigma}}^{(i)} = \hat{\mathbf{L}}^{(i)}\hat{\mathbf{L}}^{(i)\prime} + \hat{\mathbf{\Psi}}^{(i)}$ is the current approximation to $\mathbf{S}$ based on these matrices.

Then updated estimates at the $(i+1)^{th}$ iteration are

$$\hat{\mathbf{L}}^{(i+1)} = \mathbf{L}_{m1}(\hat{\mathbf{\Psi}}^{(i)}, \mathbf{S}), \qquad \hat{\mathbf{\Psi}}^{(i+1)} = \mathrm{diag}[\mathbf{S} - \hat{\mathbf{L}}^{(i+1)}\hat{\mathbf{L}}^{(i+1)\prime}].$$

That is $\hat{\psi}_j = s_{jj} - \sum_{1 \le k \le m} \hat{\ell}_{jm}^2$, j = 1, ..., p. The columns of $\hat{\mathbf{L}}^{(i+1)}$ are proportional to the eigenvectors of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(i)}$.

The updated factor analytic estimate of $\mathbf{\Sigma}$ is $\hat{\mathbf{\Sigma}}^{(i+1)} = \hat{\mathbf{L}}^{(i+1)}\hat{\mathbf{L}}^{(i+1)\prime} + \hat{\mathbf{\Psi}}^{(i+1)}$.

Thus at every stage we act as if $\hat{\mathbf{\Psi}}^{(i)}$ was the true matrix of uniquenesses, find the corresponding $\mathbf{L} = \hat{\mathbf{L}}^{(i+1)} = \mathbf{L}_{m1}(\hat{\mathbf{\Psi}}^{(i)}, \mathbf{S})$ from the eigen decomposition of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(i)}$, and then update the uniquenesses so that $\mathrm{diag}[\hat{\mathbf{\Sigma}}^{(i+1)}] = \mathrm{diag}[\mathbf{S}]$. Ideally you would continue the iteration until it converged, that is, $\hat{\mathbf{\Psi}}^{(i+1)} \approx \hat{\mathbf{\Psi}}^{(i)}$ or $\hat{\mathbf{\Sigma}}^{(i+1)} \approx \hat{\mathbf{\Sigma}}^{(i)}$. In practice, it may converge *very* slowly.

To get started, you must provide a <u>starting value</u> $\hat{\mathbf{\Psi}}^{(0)}$ for $\mathbf{\Psi}$. Because an element ψ_i of $\mathbf{\Psi}$ is the variance of the unique part of X_i, that is, the part that cannot be explained in terms of the common factors, a natural preliminary guess for ψ_i is the estimated residual variance of X_i in the multiple regression of X_i on *all* the other X_j's, $j \neq i$. It can be shown that the population residual variance is $1/\sigma^{ii}$, where $\mathbf{\Sigma}^{-1} = [\sigma^{ij}]$. Thus a popular choice for starting value is

$$\hat{\mathbf{\Psi}}^{(0)} = \mathrm{diag}[1/s^{11}, \ldots, 1/s^{pp}], \text{ where } \mathbf{S}^{-1} \equiv [s^{ij}].$$

When $\mathbf{x}$ is multivariate normal, $E[1/s^{ii}] = [(f_e - p + 1)/f_e](1/\sigma^{ii})$, where $\mathbf{\Sigma}$ is estimated on f_e degrees of freedom ($f_e = n-1$ for a random sample). This suggests that an alternative choice for $\hat{\mathbf{\Psi}}^{(0)}$ might be

$$\hat{\mathbf{\Psi}}^{(0)} = \{f_e/(f_e-p+1)\}\mathrm{diag}[1/s^{11}, \ldots, 1/s^{pp}],$$

although this extra level of sophistication is seldom used. Estimates obtained by iterating this process process to convergence are known as *iterated principal factor solutions*. The next section gives another rationale for them.

Unweighted least squares method (ULS)

Mathematics shows that at each iteration of the principal factor method, the quantity

$$U(\hat{\mathbf{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)}) \equiv \mathrm{tr}(\hat{\mathbf{\Sigma}}^{(i)} - \mathbf{S})^2 = \sum_k \sum_j (\hat{\sigma}_{kj}{}^{(i)} - s_{kj})^2$$

never <u>increases</u>, that is $U(\hat{\mathbf{\Psi}}^{(i+1)}, \hat{\mathbf{L}}^{(i+1)}) \le U(\hat{\mathbf{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)})$. In fact, $U(\hat{\mathbf{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)})$ will

decrease at each step unless it is already at a minimum.

Notation: tr(**A**) means trace(**A**) = $\sum_{1 \le i \le p} a_{ii}$ = sum of diagonal elements.

As you can see, $U(\hat{\boldsymbol{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)})$ is a sum of squared residuals $\hat{\sigma}_{kj}{}^{(i)} - s_{kj}$ between the elements of the factor analytic estimator $\hat{\boldsymbol{\Sigma}}^{(i)}$ of $\boldsymbol{\Sigma}$ and the elements of **S**, the unrestricted estimator of $\boldsymbol{\Sigma}$ so that $U(\hat{\boldsymbol{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)})$ is a possilbe measure of the goodness of fit of $\hat{\boldsymbol{\Sigma}}^{(i)}$ to **S**. This suggests the following objective approach to estimating **L** and $\boldsymbol{\Psi}$: Find $\hat{\boldsymbol{\Psi}}$ and $\hat{\mathbf{L}}$ so as to minimize $U(\boldsymbol{\Psi}, \mathbf{L})$. This is called the *unweighted least squares method* or *ULS* method.

For any fixed loading matrix **L**, $U(\boldsymbol{\Psi}, \mathbf{L})$ is smallest when $\boldsymbol{\Psi} = \text{diag}[s_{11}, \ldots, s_{pp}] - \text{diag}[\mathbf{LL}'] = \text{diag}[s_{ii} - \sum_{1 \le j \le m} \ell_{ij}{}^2]$, that is, the diagonal elements of $\mathbf{LL}' + \boldsymbol{\Psi}$ are the same as the diagonal elements of **S**.

Similarly, for any fixed diagonal uniqueness matrix $\boldsymbol{\Psi}$ such that $\boldsymbol{\Sigma} - \boldsymbol{\Psi}$ is positive definite, $U(\boldsymbol{\Psi}, \mathbf{L})$ is smallest when $\mathbf{L} = \mathbf{L}_{m1}(\boldsymbol{\Psi}, \mathbf{S})$, as defined above. Moreover, if $\delta_1 \ge \delta_2 \ge \ldots \ge \delta_p$ are the eigenvalues of $\mathbf{S} - \boldsymbol{\Psi}$

$$U(\boldsymbol{\Psi}) \equiv \min_{\mathbf{L}} U(\boldsymbol{\Psi}, \mathbf{L}) = U(\boldsymbol{\Psi}, \mathbf{L}_{m1}(\boldsymbol{\Psi}, \mathbf{S})) = \sum_{m+1 \le i \le p} \delta_i{}^2.$$

This quantity depends directly on on $\boldsymbol{\Psi}$ (and **S**) which ULS method attempts to minimize over all possible choices of $\boldsymbol{\Psi}$. Moreover you can find the δ_i's, the eigenvalues of $\mathbf{S} - \boldsymbol{\Psi}$. This means, the δ_i's and hence $U(\boldsymbol{\Psi}) = \sum_{m+1 \le i \le p} \delta_i{}^2$ are computable functions of the elements ψ_j of $\boldsymbol{\Psi}$. With proper software, you can determine $\hat{\boldsymbol{\Psi}}$ by numerically minimizing the quantity $U(\boldsymbol{\Psi})$ *as a function* of the ψ_j's. Although you cannot express the dependence of $\sum_{m+1 \le i \le p} \delta_i{}^2$ on the ψ_j's (and the elements of **S**) in a simple way, such a minimization can be routinely accomplished by a suitable computer algorithm for function minimization. Once $\hat{\boldsymbol{\Psi}}$ has been found, $(\hat{\boldsymbol{\Psi}}, \hat{\mathbf{L}})$ minimizes $U(\boldsymbol{\Psi}, \mathbf{L})$ where $\hat{\mathbf{L}} = \mathbf{L}_{m1}(\hat{\boldsymbol{\Psi}}, \mathbf{S})$.

Because the principal factor method decreases $U(\hat{\boldsymbol{\Psi}}^{(i)}) = U(\hat{\boldsymbol{\Psi}}^{(i)}, \hat{\mathbf{L}}^{(i)})$ at each iteration, you can view it as an iterative approach to determining the ULS estimates. It first holds $\hat{\boldsymbol{\Psi}}$ fixed and minimizes of $\hat{\mathbf{L}}$ (the eigenvector computation), and then holds $\hat{\mathbf{L}}$ fixed and minimizes over $\hat{\boldsymbol{\Psi}}$ by setting $\hat{\psi}_i = s_{ii} - \sum_{1 \le j \le m} \hat{\ell}_{ij}{}^2$. *If* the principal factor method converges it *may* reach the actual $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Lambda}}$ that minimize $U(\boldsymbol{\Psi}, \mathbf{L})$. However, convergence tends to be very slow and it may take very many iterations, each involving an eigenvector computation, to approach the actual minimum.

Another danger in using the principal factor method instead of a direct

minimization algorithm is the possibility that at some stage the m-th eigenvalue $\delta_m < 0$, thus making it impossible to compute $\sqrt{\delta_m}$ as required. This happens when $\mathbf{S} - \hat{\Psi}^{(i)}$ is not positive definite and is not particularly uncommon.

One disadvantage of the ULS method is that the ULS estimates obtained from the variance matrix $\mathbf{S}$ and ULS estimates obtained from the correlation matrix $\mathbf{R}$ are not simply related in the same way as their population counterparts are. That is, if $\hat{L}_S$ and $\hat{\Psi}_S$ minimize $\mathrm{tr}(\mathbf{S} - \hat{\Sigma})^2$ and $\hat{L}_R$ and $\hat{\Psi}_R$ minimize $\mathrm{tr}(\mathbf{R} - \hat{\rho})^2$ then, in general, $\hat{\Psi}_R \neq \hat{\Delta}\hat{\Psi}_S\hat{\Delta}$ and $\hat{L}_R \neq \hat{\Delta}\hat{L}_S$, where $\hat{\Delta} \equiv \mathrm{diag}[1/\sqrt{s_{11}},\ldots,1/\sqrt{s_{pp}}]$. Thus the factor structure you estimate depends on whether you start with $\mathbf{S}$ or $\mathbf{R}$.

Generalized least squares method (GLS)

You can overcome the problem of lack of correspondence between covariance-derived and correlation-derived estimates by minimizing the *generalized least squares* criterion

$$G(\Psi,L) = G_S(\Psi,L) \equiv \mathrm{tr}(S^{-1}(S - \Sigma^*))^2 = \mathrm{tr}(I_p - S^{-1}\Sigma^*)^2,\ \Sigma^* \equiv LL' + \Psi.$$

When Σ^* is "close" to $\mathbf{S}$, then $\mathbf{S}^{-1}\Sigma^*$ should be "close" to I_p and $G(\Psi,L)$ will be small. The quantity $G_S(\Psi,L)$ is a sum of squared residuals of the elements of $\mathbf{S}^{-1}\Sigma^*$ from the elements of I_p, and hence is a possible measure of the goodness of fit of $\Sigma^* = LL' + \Psi$ to $\mathbf{S}$. The method of estimating Ψ and $\mathbf{L}$ by minimizing $G(\Psi,L)$ is known as the *generalized least squares* or *GLS* method.

Suppose you rescale $\mathbf{x}$ by $\mathbf{x} \rightarrow \tilde{\mathbf{x}} = \mathbf{D}\mathbf{x}$, where $\mathbf{D}$ is *any* diagonal matrix with non-zero elements on the diagonal. Then $V[\tilde{\mathbf{x}}] = \tilde{\Sigma} \equiv \mathbf{D}\Sigma\mathbf{D}$ and $\hat{V}[\tilde{\mathbf{x}}] = \tilde{S} \equiv \mathbf{D}S\mathbf{D}$. If you define $\tilde{L} \equiv \mathbf{D}L$ and $\tilde{\Psi} \equiv \mathbf{D}\Psi\mathbf{D}$, then $\tilde{\Sigma} = \tilde{L}\tilde{L}' + \tilde{\Psi}$ is of factor analytic form since $\tilde{\Psi}$ is diagonal. The value of the GLS criterion doesn't change. Specifically

$$\begin{aligned} G_{\tilde{S}}(\tilde{\Psi},\tilde{L}) &= \mathrm{tr}(I_p - D^{-1}S^{-1}D^{-1}D\Sigma D)^2 = \mathrm{tr}(I_p - D^{-1}S^{-1}\Sigma D)^2 \\ &= \mathrm{tr}\{D^{-1}(I_p - S^{-1}\Sigma)D\}^2 = \mathrm{tr}\{D^{-1}(I_p - S^{-1}\Sigma)(I_p - S^{-1}\Sigma)\,D\} \\ &= \mathrm{tr}(I_p - S^{-1}\Sigma)^2 = G_S(\Psi,L). \end{aligned}$$

The next to last equality follows because $\mathrm{tr}\mathbf{AB} = \mathrm{tr}\mathbf{BA}$ for matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{AB}$ is square.

To summarize, $G(\Psi,L)$ is unchanged by rescaling $\mathbf{x}$, provided you replace $\mathbf{L}$ and Ψ by appropriately rescaled matrices $\tilde{L}$ and $\tilde{\Psi}$. This means that, when $\mathbf{L} = \hat{L}_S$

and $\Psi = \hat{\Psi}_S$ minimize $G_S(\Psi,L)$, then also $\tilde{L} \equiv DL_S$ and $\tilde{\Psi} \equiv D\Psi_S D$ minimize $G_{\tilde{S}}(\tilde{\Psi},\tilde{L})$.

In particular, this is true if $D = \hat{\Delta} = \text{diag}[1/\sqrt{s_{11}},...,1/\sqrt{s_{pp}}]$ so that $\tilde{S} = R$. This means that, if $L = \hat{L_S}$, and $\Psi = \hat{\Psi}_S$ minimize $G_S(\Psi,L)$ and $L = \hat{L_R}$, and $\Psi = \hat{\Psi}_R$ minimize $G_R(\Psi,L)$, then $\hat{L_R} = \hat{\Delta}L_S$ and $\hat{\Psi}_R = \hat{\Delta}\Psi_S\hat{\Delta}$. The conclusion is that for GLS estimates, the covariance- and correlation-derived estimates *are* related in the same way as their population counterparts.

For any fixed L, $G(\Psi,L)$ is smallest when Ψ is such that the diagonal elements of $S^{-1}\Sigma^*S^{-1} = S^{-1}(LL' + \Psi)S^{-1}$ are the same as the diagonal elements of S^{-1}, or equivalently, $\text{diag}[S^{-1}\Psi S^{-1}] = \text{diag}[S^{-1} - S^{-1}LL'S^{-1}]$. This implies that, given L, the minimizing elements $\psi_1,...,\psi_p$ of Ψ satisfy the simultaneous linear equations

$$\sum_{1\le j\le p}(s^{ij})^2\psi_j = (S^{-1} - S^{-1}LL'S^{-1})_{ii},\ i = 1,...,p,$$

This a system of p linear equations in the p unknown ψ_j's with coefficients that are the squares of the elements of S^{-1} and right hand sides the diagonal elements of $S^{-1} - S^{-1}LL'S^{-1}$.

Notation: $(S^{-1} - S^{-1}LL'S^{-1})_{ii}$ is the i^{th} diagonal element of $S^{-1} - S^{-1}LL'S^{-1}$, and $S^{-1} = [s^{ij}]_{1\le i,j\le p}$.

For given diagonal Ψ, define $U_\Psi \equiv [u_1,...,u_m,u_{m+1},...,u_p]$ and $\Gamma_\Psi = \text{diag}[\gamma_1,...,\gamma_p]$, with $\gamma_1 \ge \gamma_2 \ge ... \ge \gamma_p$, to be the matrices of eigenvectors and eigenvalues of S *relative* to Ψ (ordinary eigenvectors and eigenvalues of $\Psi^{-1}S$) As usual, these relative eigenvectors are normalized so as to satisfy $U_\Psi'\Psi U_\Psi = I_p$ and $U_\Psi'SU_\Psi = \Gamma_\Psi$. Then also U_Ψ and $\Gamma_\Psi - I_p = \text{diag}[\gamma_1-1,...,\gamma_p-1]$ are eigenvectors and eigenvalues of $S - \Psi$ relative to Ψ. Provided $\gamma_m \ge 1$, you can define the p by m matrix

$$L_{m2}(\Psi,S) \equiv [\sqrt{\{\gamma_1-1\}}\Psi u_1,...,\sqrt{\{\gamma_m-1\}}\Psi u_m]$$

For any fixed Ψ, $G(\Psi,L)$ is smallest when $L = L_{m2}(\Psi,S)$. Its minimum value is $G(\Psi) \equiv G(\Psi, L_{m2}(\Psi,S)) = \sum_{m+1\le j\le p}(1 - 1/\gamma_i)^2$. Since this quantity is determined from Ψ and S, just as was the case for the ULS method, we have reduced the problem to minimizing a function of the p variables $\psi_1,...,\psi_p$, a numerical problem solvable by a suitable computer algorithm. Also, it is easy to check by substitution

$$L'\Psi^{-1}L = L_{m2}(\Psi,S)'\Psi^{-1}L_{m2}(\Psi,S) = \text{diag}[\gamma_1-1,...,\gamma_p-1],$$

a diagonal matrix. Moreover, $\mathbf{L} = \mathbf{L}_{m2}(\mathbf{\Psi},\mathbf{S})$ is the only choice for $\mathbf{L}$ (except for trivial changes of signs of columns) such that this is true.

Formal GLS goodness-of-fit test

When $\mathbf{x}$ is $N_p(\boldsymbol{\mu},\mathbf{\Sigma})$, and if $\mathbf{\Sigma}$ is of factor analytic form, then for large f_e,

$$K \times G(\hat{\mathbf{\Psi}},\hat{\mathbf{L}}) = K \times G(\hat{\mathbf{\Psi}}) = K \times \sum_{m+1 \le j \le p}(1 - 1/\hat{\gamma}_j)^2 \approx \chi_f^2,$$

$$K = \{f_e - (2p+5)/6 - 2m/3\} \text{ and } f = \{(p-m)^2 - p - m\}/2$$

where $\hat{\gamma}_j$ are the eigenvalues of $\mathbf{S}$ relative to $\hat{\mathbf{\Psi}}$.

When $\mathbf{\Sigma}$ is *not* of factor analytic form, the $\hat{\gamma}$'s will tend to be much larger than 1 and $G(\hat{\mathbf{\Psi}})$ will be large. Hence comparing this statistic with $\chi^2_f(\alpha)$ provides a formal goodness of fit test H_0: $\mathbf{\Sigma}$ has factor analytic form with m factors against the alternative H_0: $\mathbf{\Sigma}$ does not have factor analytic form with m factors.

There is an easy to understand but slow iterative method for minimizing $G(\hat{\mathbf{\Psi}})$ that bears the same relationship to GLS as principal factor iteration bears to ULS.

Let $\hat{\mathbf{\Psi}}^{(i)}$ be the estimated uniqueness matrix at iteration i and define $\hat{\mathbf{L}}^{(i+1)} = \mathbf{L}_{m2}(\hat{\mathbf{\Psi}}^{(i)},\mathbf{S})$. This choice of $\mathbf{L}$ minimizes $G(\hat{\mathbf{\Psi}}^{(i)},\mathbf{L})$, holding $\hat{\mathbf{\Psi}}^{(i)}$ fixed. Then find $\hat{\mathbf{\Psi}}^{(i+1)}$ so as to minimize $G(\mathbf{\Psi},\hat{\mathbf{L}}^{(i+1)})$, holding $\hat{\mathbf{L}}^{(i+1)}$ fixed. But $\hat{\mathbf{\Psi}}^{(i+1)}$ can be found by solving the linear equations given above. Repeat this process until $\hat{\mathbf{\Psi}}^{(i+1)} \approx \hat{\mathbf{\Psi}}^{(i)}$ or $\hat{\mathbf{\Sigma}}^{(i+1)} \approx \hat{\mathbf{\Sigma}}^{(i)}$. It is again reasonable to take as starting value $\hat{\mathbf{\Psi}}^{(0)} = \text{diag}[1/s^{11},\ldots,1/s^{pp}]$. As with the iterated principal factor method, this method may converge slowly, if at all. However, at each step $G(\hat{\mathbf{\Psi}}^{(i+1)}) = G(\hat{\mathbf{\Psi}}^{(i+1)},\hat{\mathbf{L}}^{(i+1)})$ decreases. It may also happen that the iteration may fail because $\gamma_m^{(i)} < 1$, where $\gamma_m^{(i)}$ is the m-th eigenvalue of $\mathbf{S}$ relative to $\hat{\mathbf{\Psi}}^{(i)}$, thus making it impossible to compute $\sqrt{(\gamma_m^{(i)} - 1)}$ as required. In that case the iteration must be abandoned.

Because of the use of $\mathbf{L}_{m2}(\hat{\mathbf{\Psi}}^{(i)},\mathbf{S})$ at each stage, the iterated GLS estimate of $\mathbf{L}$ satisfies the condition that $\hat{\mathbf{L}}'\mathbf{\Psi}^{-1}\hat{\mathbf{L}}$ is diagonal.

Maximum likelihood estimation (ML)

When $\mathbf{x}$ is $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\mathbf{S}$ has a Wishart distribution. The general good properties of maximum likelihood estimates suggest that we attempt to find the maximum likelihood estimator $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ based on this distribution. This can be shown equivalent to minimizing the quantity

$$M(\hat{\boldsymbol{\Psi}},\hat{\mathbf{L}}) \equiv \mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \mathbf{I}_p) - \log(\det[\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}]) > 0,\ \hat{\boldsymbol{\Sigma}} \equiv \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$$

Note that if $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$, that is, $\mathbf{S}$ is exactly of the factor analytic form, then $M(\hat{\boldsymbol{\Psi}},\hat{\mathbf{L}}) = \mathrm{tr}\,\mathbf{0} - \log(\det(\mathbf{I}_p)) = 0 - \log(1) = 0$. Thus, at its minimum, the value $M(\hat{\boldsymbol{\Psi}},\hat{\mathbf{L}})$ can be viewed as a measure of the goodness of fit of $\mathbf{S}$ to factor analytic form.

For any fixed $\mathbf{L}$, you can show mathematically that $M(\boldsymbol{\Psi},\mathbf{L})$ is smallest when $\boldsymbol{\Psi}$ is such that the diagonal elements of $\mathbf{S}$ and $\mathbf{LL}' + \boldsymbol{\Psi}$ are the same, that is, $\boldsymbol{\Psi} = \mathrm{diag}[\mathbf{S} - \mathbf{LL}']$, just as in the case of the ULS method.

For any $\boldsymbol{\Psi}$ such that $\mathbf{S} - \boldsymbol{\Psi}$ is positive definite, let $\mathbf{U}_\Psi \equiv [\mathbf{u}_1,\ldots,\mathbf{u}_m,\mathbf{u}_{m+1},\ldots,\mathbf{u}_p]$ and $\boldsymbol{\Gamma}_\Psi = \mathrm{diag}[\gamma_1,\ldots,\gamma_p]$ be the matrices of eigenvectors and eigenvalues of $\mathbf{S}$ relative to $\boldsymbol{\Psi}$, with $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_p$, as in the GLS method. Provided $\gamma_m \geq 1$, again define (as for the GLS method)

$$\mathbf{L}_{m2}(\boldsymbol{\Psi},\mathbf{S}) \equiv [\sqrt{\{\gamma_1 - 1\}}\boldsymbol{\Psi}\mathbf{u}_1,\ldots,\sqrt{\{\gamma_m - 1\}}\boldsymbol{\Psi}\mathbf{u}_m].$$

Then for any fixed $\boldsymbol{\Psi}$, $M(\boldsymbol{\Psi},\mathbf{L})$ is smallest when $\mathbf{L} = \tilde{\mathbf{L}} \equiv \mathbf{L}_{m2}(\boldsymbol{\Psi},\mathbf{S})$. In addition, the eigenvectors and eigenvalues of $\tilde{\boldsymbol{\Sigma}} \equiv \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \boldsymbol{\Psi}$ relative to $\boldsymbol{\Psi}$ are the columns of $\mathbf{U}_\Psi$ and diagonal elements of $\tilde{\boldsymbol{\Gamma}}_\Psi = \mathrm{diag}[\gamma_1,\ldots,\gamma_m,1,\ldots,1]$ (same as $\boldsymbol{\Gamma}_\Psi$, with 1's replacing γ_j for $j > m$). Moreover, the minimized value is

$$M(\boldsymbol{\Psi}) \equiv M(\boldsymbol{\Psi},\mathbf{L}_{m2}(\boldsymbol{\Psi},\mathbf{S})) \equiv \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \mathbf{I}_p) - \ln\det(\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) = \sum_{m+1 \leq i \leq p}(\gamma_i - 1 - \ln\gamma_i) > 0.$$

Once again, since the value of $M(\boldsymbol{\Psi}) = M(\boldsymbol{\Psi},\mathbf{L}_{m2}(\boldsymbol{\Psi},\mathbf{S}))$ is determined by $\boldsymbol{\Psi}$, we have reduced the problem of minimizing a function $M(\boldsymbol{\Psi},\mathbf{L})$ of *both* $\mathbf{L}$ and $\boldsymbol{\Psi}$ to one of minimizing a function $M(\boldsymbol{\Psi})$ of the p variables $\psi_1,\ldots,\psi_p$.

Formal ML goodness-of-fit test

When $\mathbf{x}$ is multivariate normal, and $\boldsymbol{\Sigma}$ is of factor analytic form, then for large f_e,

$$K \times M(\hat{\boldsymbol{\Psi}},\hat{\mathbf{L}}) = K \times M(\hat{\boldsymbol{\Psi}}) = K \times \sum_{m+1 \leq j \leq p}(1 - 1/\hat{\gamma}_i)^2 \approx \chi_f^2,$$

$$K = \{f_e - (2p+5)/6 - 2m/3\} \text{ and } f = \{(p-m)^2 - p - m\}/2$$

where $\hat{\gamma}_j$ are the eigenvalues of $\mathbf{S}$ relative to $\hat{\boldsymbol{\Psi}}$. As for the GLS method, you can use this as a test of goodness of fit of the factor analytic model

itself.

There is a simple iteration method associated with ML estimation that is analogous to the iterated principal factor method. As before, if $\hat{\mathbf{L}}^{(i)} = [\hat{\ell}_{kj}{}^{(i)}]$ and $\hat{\boldsymbol{\Psi}}^{(i)}$ are the estimates at the i-th iteration,

$$\hat{\mathbf{L}}^{(i+1)} = \mathbf{L}_{m2}(\hat{\boldsymbol{\Psi}}^{(i)}, \mathbf{S}),$$
$$\hat{\boldsymbol{\Psi}}^{(i+1)} = \text{diag}[\mathbf{S} - \hat{\mathbf{L}}^{(i+1)}\hat{\mathbf{L}}^{(i+1)\prime}] = \text{diag}[s_{kk} - \textstyle\sum_{1 \le j \le m}(\hat{\ell}_{kj}{}^{(i+1)})^2].$$

That is, we update $\hat{\mathbf{L}}$ so as to minimize $M(\hat{\boldsymbol{\Psi}}^{(i)}, \hat{\mathbf{L}})$ and then choose $\hat{\boldsymbol{\Psi}}^{(i+1)}$ so that the $\text{diag}[\hat{\boldsymbol{\Sigma}}^{(i+1)}] = \text{diag}[\mathbf{S}]$. This iteration *may* converge, albeit often slowly, to the maximum likelihood solution since it can be shown that $M(\hat{\boldsymbol{\Psi}}^{(i+1)}) \le M(\hat{\boldsymbol{\Psi}}^{(i)})$.

As with the GLS method, the solutions using $\mathbf{R}$ and $\mathbf{S}$ are simply related by $\hat{\boldsymbol{\Psi}}_{\mathbf{R}} = \hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Psi}}_{\mathbf{S}}\hat{\boldsymbol{\Delta}}$ and $\hat{\mathbf{L}}_{\mathbf{R}} = \hat{\boldsymbol{\Delta}}\hat{\mathbf{L}}_{\mathbf{S}}$. Moreover, again $\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$ is diagonal.

Summary of factor extraction methods

The **principal components method** is based on a single eigenvalue/eigenvector computation. It is not consistent for the parameters of the factor analytic model, but is widely used. It is highly scale dependent

The other three basic methods, **unweighted least squares** (ULS), **generalized least squares** (GLS) and **maximum likelihood** (ML) are similar in several ways.

- Each seeks to minimize a criterion measuring how good the fit of the estimated variance matrix $\hat{\boldsymbol{\Sigma}}$ or correlation matrix $\hat{\boldsymbol{\rho}}$ is to the observed variance matrix $\mathbf{S}$ or correlation matrix $\mathbf{R}$.
- For a given uniqueness matrix $\boldsymbol{\Psi}$, it is easy to find an optimal loading matrix $\hat{\mathbf{L}}(\boldsymbol{\Psi})$. This allows the criterion to be expressed purely in terms of $\boldsymbol{\Psi}$ and effectively reduces the problem to finding $\hat{\boldsymbol{\Psi}}$ to minimize the criterion.

The three criteria are as follows:

Method	Criterion	Criterion expressed in terms of $\hat{\boldsymbol{\Psi}}$
ULS	$\mathrm{tr}\{(\mathbf{S}-\hat{\boldsymbol{\Sigma}})^2\}$ or $\mathrm{tr}((\mathbf{R}-\hat{\boldsymbol{\Sigma}})^2)$	$U(\hat{\boldsymbol{\Psi}}) = \sum_{m+1 \le i \le p} \delta_i^2$, where $\delta_1 \ge \delta_2 \ge \ldots \ge \delta_m \ge \delta_{m+1} \ge \ldots \ge \delta_p$ are eigenvalues of $\mathbf{S} - \hat{\boldsymbol{\Psi}}$ or $\mathbf{R} - \hat{\boldsymbol{\Psi}}$
GLS	$\mathrm{tr}\{(\mathbf{S}^{-1}(\mathbf{S}-\hat{\boldsymbol{\Sigma}}))^2\}$ or $\mathrm{tr}\{(\mathbf{R}^{-1}(\mathbf{R}-\hat{\boldsymbol{\Sigma}}))^2\}$	$G(\hat{\boldsymbol{\Psi}}) = \sum_{m+1 \le j \le p}(1 - 1/\gamma_i)^2$, where $\gamma_1 \ge \gamma_2 \ge \ldots \ge \gamma_m \ge \gamma_{m+1} \ge \ldots \ge \gamma_p$ are eigenvalues of $\mathbf{S}$ relative to $\hat{\boldsymbol{\Psi}}$ or $\mathbf{R}$ relative to $\hat{\boldsymbol{\Psi}}$
ML	$\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \mathbf{I}_p) - \log(\det[\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}])$ or $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{R} - \mathbf{I}_p) - \log(\det[\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{R}])$	$M(\hat{\boldsymbol{\Psi}}) = \sum_{m+1 \le i \le p}(\gamma_i - 1 - \ln\gamma_i)$, where the γ's are as for the GLS method.

- Iterative methods are required to do the minimization. The handout describes methods for each, all very similar and all likely to converge slowly if at all. The algorithm for ULS is sometimes considered to be a separate method, the **iterated principal factor method**.
- For all three, the preferred approach is to use a better minimization algorithm that uses derivatives of $U(\boldsymbol{\Psi})$, $G(\boldsymbol{\Psi})$ or $M(\boldsymbol{\Psi})$ with respect to ψ_1, ..., ψ_p. The derivatives can be calculated using an analytic formula or numerically. This is what MacAnova macro `facanal()` does.