Chapter 9. Factor Analysis

Factor analysis may be viewed as a refinement of the principal component analysis. The objective is, like the P.C. analysis, to describe the relevant variables in study in terms of a few underlying variables, called factors.

9.1 Orthogonal factor model.

Let $X = (X_1, \cdots, X_p)'$ be the variables in population with

$$\text{mean } E(X) = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \quad \text{and variance} \quad \text{var}(X) = \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}.$$

The orthogonal factor model is

$$X_{p\times 1} - \mu_{p\times 1} = L_{p\times m} F_{m\times 1} + \epsilon_{p\times 1}, \tag{9.1}$$

where $m \le p$, $\mu = E(X)$,

$$L \equiv \begin{pmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & \vdots & \vdots \\ l_{p1} & \cdots & l_{pm} \end{pmatrix}$$ is called the *factor loading matrix* (which is non-random),

$$F = \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}$$ are called the factors or common factors,

and $$\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{pmatrix}$$ are called errors or specific errors.

The model can be re-expressed as

$$X_i - \mu_i = \sum_{j=1}^{m} l_{ij} F_j + \epsilon_i, \qquad i = 1, ..., p. \tag{9.1$'$}$$

And l_{ij} is called the loading of X_i on the factor F_j.

The assumptions of the orthogonal model are:

(1). $E(F) = 0_{m\times 1}$ and $\text{var}(F) = I_m$.

(2). $E(\epsilon) = 0_{p\times 1}$ and $\text{var}(\epsilon) = \Psi$, a diagonal matrix with diagonal elements: $\psi_1, ..., \psi_p$.

(3). $\text{cov}(F, \epsilon) = 0_{m\times p}$.

Remark. The above model assumption implies that

$$\text{cov}(F_i, F_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \ne j \end{cases} \qquad \text{cov}(F_i, \epsilon_j) = 0 \text{ and } \text{cov}(\epsilon_i, \epsilon_j) = \begin{cases} \psi_i & \text{if } i = j \\ 0 & \text{if } i \ne j. \end{cases}$$

Moreover,

$$\text{cov}(X, F) = \text{cov}(LF, F) = L \qquad \text{cov}(X_i, F_j) = l_{ij}, \quad i = 1, ..., p; \ j = 1, ..., m.$$

Under the orthogonal factor model, the variance matrix of X, Σ, can be written as

$$\begin{aligned} \Sigma &= \text{var}(X) = \text{var}(LF + \epsilon) = \text{var}(LF) + \text{var}(\epsilon) \\ &= L\text{var}(F)L' + \Psi = LL' + \Psi. \end{aligned}$$

In particular,

$$\begin{aligned}
\sigma_{ii} &= \mathrm{var}(X_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i \equiv h_i^2 + \psi_i \qquad (9.2)\\
\sigma_{ij} &= \mathrm{cov}(X_i, X_j) = \sum_{k=1}^{m} l_{ik} l_{jk}, \qquad i \neq j.
\end{aligned}$$

Here $h_i^2 \equiv l_{i1}^2 + \cdots + l_{im}^2$ is called *communality*, which is the portion of the variance of X_i explained by common factors. ψ_i, called *specific variance* or *uniqueness*, is the portion of the variance of X_i explained by specific factor, the error pertained to the i-th variable X_i only.

In orthogonal factor model, the factors or common factors are supposed to be important underlying factors that significantly affect all variables. Besides these factors, the remaining ones are those only pertained to the relevant variables. Specifically, ϵ_i, the error pertained to the i-th variable X_i, explains the part of the variation of X_i that cannot be explained by common factors or by other errors.

Remark. The orthogonal factor model (9.1) is essentially different from linear regression model, although there is certain formality resemblance. The key difference is the common factor F, which seemingly plays the role of covariates in linear regression model, is not observable.

Remark. The orthogonal factor model (9.1) has unidentifiable L and F, up to a rotation, in the sense that

$$X - \mu = LF + \epsilon = L^* F^* + \epsilon,$$

where $L^* = LT$ and $F = T'F$ with T being any orthonormal $m \times m$ matrix. Then, (F^*, ϵ) still satisfies the assumptions (1)-(3).

9.2 Estimation by the principal component approach.

Recall the decomposition of Σ:

$$\Sigma = \mathbf{e}\Lambda\mathbf{e}' = (\mathbf{e}_1 \vdots \cdots \vdots \mathbf{e}_p) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}.$$

i.e., $(\lambda_k, \mathbf{e}_k)$, $k = 1, ..., p$ are the p eigenvalue-eigenvector pairs of Σ with $\lambda_1 \geq ... \geq \lambda_p > 0$. And the population P.C.s are

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} = \mathbf{e}'(X - \mu).$$

Then,

$$\begin{aligned}
X - \mu &= \mathbf{e}Y = (\mathbf{e}_1 \vdots \cdots \vdots \mathbf{e}_p) \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} = \sum_{j=1}^{m} \mathbf{e}_j Y_j + \sum_{j=m+1}^{p} \mathbf{e}_j Y_j.\\
&= \sum_{j=1}^{m} \sqrt{\lambda_j}\mathbf{e}_j (Y_j/\sqrt{\lambda_j}) + \sum_{j=m+1}^{p} \mathbf{e}_j Y_j.\\
&= (\sqrt{\lambda_1}\mathbf{e}_1 \vdots \cdots \vdots \sqrt{\lambda_m}\mathbf{e}_m) \begin{pmatrix} Y_1/\sqrt{\lambda_1} \\ \vdots \\ Y_m/\sqrt{\lambda_m} \end{pmatrix} + \sum_{j=m+1}^{p} \mathbf{e}_j Y_j.\\
&= LF + \epsilon, \qquad \text{say,}
\end{aligned}$$

where

$$L = (\sqrt{\lambda_1}\mathbf{e}_1 \vdots \cdots \vdots \sqrt{\lambda_m}\mathbf{e}_m), \qquad F = \begin{pmatrix} Y_1/\sqrt{\lambda_1} \\ \vdots \\ Y_m/\sqrt{\lambda_m} \end{pmatrix} \text{ and } \epsilon = \sum_{j=m+1}^{p} \mathbf{e}_j Y_j.$$

It can be verified that the assumptions (1) and (2) satisfied, but not necessarily (3). Nevertheless, the above derivation provides hint to using principal components as a possible solution to the orthogonal factor model. The common factors are simply the first m P.C.s standardized by their standard deviations: $F_j = Y_j/\sqrt{\lambda_j}$, $j = 1, ..., m$.

We next build the sample analogues of the above derivation. There are n observations, presented as

$$\mathbf{X} = (X_{(1)} \vdots \cdots \vdots X_{(p)}) = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \cdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

with $X_{(k)} = (x_{1k}, ..., x_{nk})'$ denoting the n observations of the k-th variable X_k.

Let $\mathbf{S}$ be the sample variance matrix, with $(\hat{\lambda}_k, \hat{\mathbf{e}}_k)$, $k = 1, ..., p$ as eigenvalue-eigenvector pairs and $\hat{\lambda}_1 \geq ... \geq \hat{\lambda}_p > 0$.

$$\mathbf{S} = \hat{\mathbf{e}}\hat{\Lambda}\hat{\mathbf{e}}' = (\hat{\mathbf{e}}_1 \vdots \cdots \vdots \hat{\mathbf{e}}_p) \begin{pmatrix} \hat{\lambda}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{e}}_1' \\ \vdots \\ \hat{\mathbf{e}}_p' \end{pmatrix}.$$

Then, with the P.C. approach, the factor loading matrix L is estimated by

$$\begin{aligned} \tilde{L}_{p\times m} &\equiv \begin{pmatrix} \tilde{l}_{11} & \cdots & \tilde{l}_{1m} \\ \vdots & \vdots & \vdots \\ \tilde{l}_{p1} & \cdots & \tilde{l}_{pm} \end{pmatrix} = (\tilde{l}_{(1)} \vdots \cdots \vdots \tilde{l}_{(m)}) \\ &= (\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \vdots \cdots \vdots \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m). \end{aligned}$$

Based on the fact that $\sigma_{ii} = \sum_{j=1}^{m} l_{ij}^2 + \psi_i$ and use s_{ii} to estimate σ_{ii}. Then, Ψ is estimated by

$$\tilde{\Psi} = \begin{pmatrix} \tilde{\psi}_1 & & \\ & \ddots & \\ & & \tilde{\psi}_p \end{pmatrix} \qquad \text{where} \qquad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{l}_{ij}^2.$$

Example 9.1 ANALYSIS OF WEEKLY STOCK RETURN DATA. (*Example 8.1 continued*) Suppose $m = 1$. $\hat{\lambda}_1 = 2.856$. $p = 5$. Then,

$$\tilde{L} = \sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 = \sqrt{2.856}\hat{\mathbf{e}}_1 = \begin{pmatrix} 0.783 \\ 0.773 \\ 0.793 \\ 0.713 \\ 0.712 \end{pmatrix}$$

and

$\tilde{h}_i^2 = \sum_{j=1}^{m} \tilde{l}_{ij}^2 = l_{i1}^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2 = 1 - \tilde{h}_i^2$
$\tilde{h}_1^2 = 0.783^2 = 0.61$	$\tilde{\psi}_1 = 1 - \tilde{h}_1^2 = 0.39$
$\vdots$	$\vdots$
$\tilde{h}_5^2 = 0.712^2 = 0.49$	$\tilde{\psi}_5 = 1 - \tilde{h}_5^2 = 0.51$

The proportion of total variation explained by the first (and only one) factor is

$$\frac{\sum_{i=1}^{p} \tilde{l}_{i1}^2}{\sum_{i=1}^{p} s_{ii}} = \frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_5} = \frac{2.856}{5} = 57.1\%.$$

9.3 Estimation by maximum likelihood approach.

Assume $X_1, ..., X_n$ are iid $\sim MN(\mu, \Sigma)$. Then, the likelihood is

$$lik(\mu, \Sigma) \equiv (2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)'\Sigma^{-1}(X_i - \mu)\}.$$

Under the orthogonal factor model (9.1), $\Sigma = LL' + \Psi$. Then, the likelihood becomes

$$lik(\mu, L, \Psi) \equiv (2\pi)^{-np/2}|LL' + \Psi|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)'(LL' + \Psi)^{-1}(X_i - \mu)\}.$$

With certain restriction, the MLE of L and Ψ can be computed. We denote them by $\hat{L}$ and $\hat{\Psi}$. (Actual computation of the MLE is not required.)

9.4 A test of the number of common factors.

The orthogonal factor model (9.1) pre-specifies m, the number of the common factors. In practice, m is often unknown. Here we consider, for a given m, a statistical test to test whether such an m is appropriate. Presented in terms of statistical hypothesis as

$$\begin{cases} H_0: & \Sigma = LL' + \Psi \quad \text{where } L \text{ is } p \times m \text{ matrix and } \Psi \text{ is diagonal} \\ H_a: & \text{otherwise.} \end{cases}$$

A generalized likelihood ratio test statistic is

$$\begin{aligned} & 2\log\Big(\frac{\max\{lik(\mu, \Sigma) : \mu, \Sigma\}}{\max\{lik(\mu, L, \Psi) : \mu, L, \Psi\}}\Big) = 2\log\Big(\frac{lik(\bar{X}, (n-1)/n\mathbf{S})}{lik(\bar{X}, \hat{L}, \hat{\Psi})}\Big) \\ \propto \quad & 2\log\Big(\frac{|(n-1)/n\mathbf{S}|^{-n/2}}{|\hat{L}\hat{L}' + \hat{\Psi}|^{-n/2}}\Big) = n\log\Big(\frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|(n-1)/n\mathbf{S}|}\Big) \end{aligned}$$

With some further refinement called Bartlett correction, the appropriate signficance level α test is

$$\text{reject } H_0 \text{ when } [n - 1 - (2p + 4m + 5)/6]\log\Big(\frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|(n-1)/n\mathbf{S}|}\Big) > \chi^2_{0.5[(p-m)^2 - p - m]}(\alpha).$$

9.5 Factor rotation.

As remarked in Section 9.1, the orthogonal factor model is not identifiable up to a rotation of the common factors or factor loading matrix. In other words,

$$X - \mu = LF + \epsilon = L^*F^* + \epsilon$$

for $L^* = LT$ and $F^* = T'F$ where T is any $m \times m$ orthonormal matrix. Therefore, up to a rotation, it is legitimate and often desirable to choose a pair (L, F) so that it may achieve better interpretability.

A criterion called *varimax criterion* can be applied to find an optimal rotation. Let $\hat{L}^*$ be the $p \times m$ *rotated* factor loading matrix with elements $\hat{l}_{ij}^*$. Define $l_{ij}^* = \hat{l}_{ij}^*/\hat{h}_i$, and

$$V = \sum_{j=1}^{m}\Big[\frac{1}{p}\sum_{i=1}^{p}(l_{ij}^*)^4 - \{\frac{1}{p}\sum_{i=1}^{p}(l_{ij}^*)^2\}^2\Big],$$

which is the sum of the column-wise variance of the squares of scaled factor loadings. Find the optimal l^*_{ij} such that V achieves maximum. Then, the optimal rotated factor loading matrix is $\hat{l}^*_{ij} = \hat{h}_i \times$ (the optimal l^*_{ij}).

9.6 Factor scores.

Let $x_1, ..., x_n$ be a sample of n observations of X that follows the orthogonal factor model (9.1). Write

$$x_i - \mu = Lf_i + e_i$$

Then, f_i and e_i may be regarded as the realized but unobserved values of the common factors and the errors that produced the i-th observation x_i. Note that x_i is $p \times 1$, f_i is $m \times 1$ and e_i is $p \times 1$. Factor scores refer to estimator of f_j, denoted by $\hat{f}_j$ or $\tilde{f}_j$. There are two commonly used methods of estimation.

(i). Method 1: weighted/unweighted least squares method.

Notice that, if one minimize

$$(x - \mu - Lf)'\Psi^{-1}(x - \mu - Lf)$$

over all m dimensional vector f. Then, the minimizer f is $(L'\Psi^{-1}L)^{-1}L'\Psi^{-1}(x - \mu)$.

With this minimization, we can obtain factor scores as

(1). Maximum likelihood approach:

$$\hat{f}_j = (\hat{L}'\hat{\Psi}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(x_j - \bar{x})$$

where $(\hat{L}, \hat{\Psi})$ are MLE of (L, Ψ). And $\hat{e}_j = x_j - \bar{x} - \hat{L}\hat{f}_j$ is the estimator of e_j.

(2). Principal component approach:

$$\tilde{f}_j = (\tilde{L}'\tilde{L})^{-1}\tilde{L}'(x_j - \bar{x})$$

where $\tilde{L}$ is the estimator of L based on the P.C. approach. And $\tilde{e}_j = x_j - \bar{x} - \tilde{L}\tilde{f}_j$ is the estimator of e_j.

(ii). Method 2: Regression method.

The motivation for this method comes from linear regression. The orthogonal factor model implies

$$\text{var}\begin{pmatrix} X \\ F \end{pmatrix} = \text{var}\begin{pmatrix} LF + \epsilon \\ F \end{pmatrix} = \begin{pmatrix} LL' + \Psi & L \\ L' & I_m \end{pmatrix}$$

and

$$E(F|X = x) = 0 + L'(LL' + \Psi)^{-1}(x - \mu) = L'\Sigma^{-1}(x - \mu)$$

citing from a proposition in Chapter 4. Then f_j is estimated by

$$\hat{f}_j = \hat{L}'\mathbf{S}^{-1}(x_j - \bar{x}).$$

9.7 A general guideline.

To perform a complete factor analysis, some guidelines are useful. The following steps are recommended.

1. Perform principal component factor analysis, with care of the issue of standardization.
2. Perform a maximum likelihood factor analysis.
3. Compare the results of step 1 and step 2.
4. Change number of common factors m and repeat steps 1-3.
5. For large set of data, split them in half, perform the above analysis on each half and compare the results.