STATS 206
Applied Multivariate Analysis
Lecture 7: Factor Analysis

Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Agenda

- The orthogonal factor model
- Methods of estimation
- A large sample test for the number of common factors
- Factor rotation
- Factor scores

The Factor Model

- Observable random vector $\mathbf{X}$: p components, mean $\boldsymbol{\mu}$, cov. matrix $\boldsymbol{\Sigma}$
- The factor model postulates that $\mathbf{X}$ is linearly dependent on:
  - m random variables $F_1, \ldots, F_m$: common factors;
  - and p additional sources of variation $\varepsilon_1, \ldots, \varepsilon_p$: errors;

$$\begin{aligned}
X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \ldots + l_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \ldots + l_{2m}F_m + \varepsilon_2 \\
&\vdots \qquad\qquad\qquad \vdots \\
X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pm}F_m + \varepsilon_p
\end{aligned}$$

In matrix notation:

$$\underbrace{\mathbf{X} - \boldsymbol{\mu}}_{p\times 1} = \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{F}}_{m\times 1} + \underbrace{\boldsymbol{\varepsilon}}_{p\times 1}$$

a. $F_1, \ldots, F_m$ and $\varepsilon_1, \ldots, \varepsilon_p$: unobservable $m + p$ random variables

b. $\mathbf{L} = \{l_{ij}\}$: matrix of factor loading

c. l_{ij}: loading of the i-th variable on the j-th common factor

The Orthogonal Factor Model (i)

- Orthogonal factor model with m common factors:

$$\underbrace{\mathbf{X}}_{p\times 1} = \underbrace{\boldsymbol{\mu}}_{p\times 1} + \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{F}}_{m\times 1} + \underbrace{\boldsymbol{\varepsilon}}_{p\times 1}$$

!! with additional assumptions on unobservable $\mathbf{F}$, $\boldsymbol{\varepsilon}$ below !!

a. $\mathbf{F}$, $\boldsymbol{\varepsilon}$: independent
b. $\mathrm{E}(\mathbf{F}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{F}) = \mathbf{I}$
c. $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$ ($\boldsymbol{\Psi}$: diagonal)

Here:

- μ_i: mean of variable i
- ε_i: the i-th specific factor (or the i-th error)
- F_j: the j-th common factor
- l_{ij}: loading of the i-th variable on the j-th factor

The Orthogonal Factor Model (ii): Covariance Structure

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} \quad [\mathbf{F}, \boldsymbol{\varepsilon} : \text{ independent}]$$

$$\mathrm{E}(\mathbf{F}) = \mathbf{0},\ \mathrm{Cov}(\mathbf{F}) = \mathbf{I};\ \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0},\ \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}(\boldsymbol{\Psi} : \text{diagonal})$$

1. $\mathrm{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{E}(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})'$$

$$= \mathbf{L}\underbrace{\mathrm{E}(\mathbf{F}\mathbf{F}')}_{=\mathbf{I}}\mathbf{L}' + \underbrace{\mathrm{E}(\boldsymbol{\varepsilon}\mathbf{F}')}_{=\mathbf{0}}\mathbf{L}' + \mathbf{L}\underbrace{\mathrm{E}(\mathbf{F}\boldsymbol{\varepsilon}')}_{=\mathbf{0}} + \underbrace{\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')}_{=\boldsymbol{\Psi}} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$$

or : $$\mathrm{Var}(X_i) = \sigma_{ii} = \underbrace{l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2}_{h_i^2\text{: the } i\text{-th communality}} + \underbrace{\psi_i}_{\substack{\text{specific variance} \\ \text{or uniqueness}}} = h_i^2 + \psi_i$$

$$\mathrm{Cov}(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \ldots + l_{im}l_{km}$$

2. $\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$ or $\mathrm{Cov}(X_i, F_j) = l_{ij}$

$$\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathrm{E}(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}' = \mathrm{E}(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{L}\mathrm{E}(\mathbf{F}\mathbf{F}') + \mathrm{E}(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{L}$$

Example 1: Verifying $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ for Two Factors

- It can be verified that $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$, where $\mathbf{\Sigma}, \mathbf{L}, \mathbf{\Psi}$ are given below:

$$\underbrace{\begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}}_{\mathbf{\Sigma}} = \underbrace{\begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix}}_{\mathbf{L}} \underbrace{\begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix}}_{\mathbf{L}'} + \underbrace{\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}}_{\mathbf{\Psi}}$$

Here $\mathbf{\Sigma}$ has the structure produced by an orthogonal factor model with $m = 2$. Since

$$\mathbf{L} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \\ l_{31} & l_{32} \\ l_{41} & l_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix}$$

$$\Longrightarrow \begin{cases} \text{Communality of } X_1: & h_1^2 = l_{11}^2 + l_{12}^2 = 4^2 + 1^2 = 17 \\ \text{Variance of } X_1: & \sigma_{11} = (l_{11}^2 + l_{12}^2) + \psi_1 = 17 + 2 = 19 \end{cases}$$

Nonexistence of a Proper Factor Model

- Most covariance matrices cannot be factored as $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$, where the number of factors m is much less than p.
- Example: see the textbook, p. 486

Ambiguity of Factor Models

- $m > 1$: inherent ambiguity with the factor model

  To see this, let $\mathbf{T}$ be any $m \times m$ orthogonal matrix, i.e., $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$

$$\Rightarrow \mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} = \underbrace{\mathbf{L}\mathbf{T}}_{\mathbf{L}^*}\underbrace{\mathbf{T}'\mathbf{F}}_{\mathbf{F}^*} + \boldsymbol{\varepsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon} \quad (\mathbf{L}^* = \mathbf{L}\mathbf{T}, \mathbf{F}^* = \mathbf{T}'\mathbf{F})$$

- If $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$ is an orthogonal factor model, so is $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon}$, since

$$\mathsf{E}(\mathbf{F}^*) = \mathbf{T}'\mathsf{E}(\mathbf{F}) = \mathbf{0}, \; \mathsf{Cov}(\mathbf{F}^*) = \mathbf{T}'\mathsf{Cov}(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

  Thus,

(1)	$\mathbf{L}$: determined only up to an orthogonal matrix $\mathbf{T}$
(2)	Loadings $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ and $\mathbf{L}$ $\Longrightarrow$ the same representation
(3)	Communalities: unaffected by the choice of $\mathbf{T}$
	Communalities are given by diagonal elements of $\mathbf{L}\mathbf{L}' = \mathbf{L}^*\mathbf{L}^{*\prime}$

Ambiguity of Factor Models

To proceed with the analysis of the factor model (knowing the ambiguity)

- Imposing conditions for uniquely determining $\mathbf{L}$ and $\mathbf{\Psi}$
- Rotating loading matrix $\mathbf{L}$ with an orthogonal matrix by some criterion (With $\mathbf{L}$ and $\mathbf{\Psi} \Longrightarrow$ factors identified)
- Constructing estimated values for the factors themselves (factor scores)

Methods of Estimation

- Estimating the orthogonal factor model ($\mathbf{L}$ and $\mathbf{\Psi}$)
- Two popular methods:
  - The principal component (PC) method
  - The maximum likelihood (ML) method

The Principal Component Method (i)

The PC method: using spectral decomposition of the cov. matrix $\mathbf{\Sigma}$

- Let $(\lambda_i, \mathbf{e}_i)$ be the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$ $(\lambda_1 \geq \ldots \geq \lambda_p \geq 0)$

$$
\begin{aligned}
\mathbf{\Sigma} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \ldots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \\
&= \underbrace{\begin{bmatrix} \sqrt{\lambda_1}\, \mathbf{e}_1 & \sqrt{\lambda_2}\, \mathbf{e}_2 & \ldots & \sqrt{\lambda_p}\, \mathbf{e}_p \end{bmatrix}}_{\mathbf{L}} \underbrace{\begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p}\mathbf{e}_p' \end{bmatrix}}_{\mathbf{L}'} = \mathbf{L}\mathbf{L}' + \mathbf{0}
\end{aligned}
$$

This is a special case of the orthogonal factor model with $m = p$ and special variance $\psi_i = 0$.

- An exact factor analysis
- Not particular useful/interesting, since number of common factors = number of variables ($m = p$)

The Principal Component Method (ii)

- For $m < p$: neglecting the contributions of the last $p - m$ eigenvalues (truncating the last $p - m$ columns) to obtain

$$\mathbf{L} = \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \quad \sqrt{\lambda_2}\,\mathbf{e}_2 \quad \dots \quad \sqrt{\lambda_m}\,\mathbf{e}_m \right]$$

In addition, let specific variances be the diagonal elements of $\mathbf{\Sigma} - \mathbf{L}\mathbf{L}'$:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \psi_p \end{bmatrix} = \begin{bmatrix} \sigma_{11} - \sum_{j=1}^{m} l_{1j}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{pp} - \sum_{j=1}^{m} l_{pj}^2 \end{bmatrix}$$

$$\psi_i = \sigma_{ii} - \sum_{j=1}^{m} l_{ij}^2$$

$$\text{Then :} \quad \mathbf{\Sigma} \approx \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$$

- For applications to data $\mathbf{x}_1, \dots, \mathbf{x}_n$, replace $\mathbf{\Sigma}$ by sample cov. matrix $\mathbf{S}$ (See next page!)

Principal Component Solution of the Factor Model

Principal Component Factor Analysis of the Sample Cov. Matrix $\mathbf{S}$

Let $(\widehat{\lambda}_1, \widehat{\mathbf{e}}_1)$, $(\widehat{\lambda}_2, \widehat{\mathbf{e}}_2)$, $\ldots$, $(\widehat{\lambda}_p, \widehat{\mathbf{e}}_p)$ be the eigenvalue-eigenvector pairs of $\mathbf{S}$ and $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq \widehat{\lambda}_p$. Let $m < p$ be the number of common factors. $\Longrightarrow$ The estimated factor loading matrix $\{\widetilde{l}_{ij}\}$ is

$$\widetilde{\mathbf{L}} = \left[\sqrt{\widehat{\lambda}_1}\, \widehat{\mathbf{e}}_1 \quad \sqrt{\widehat{\lambda}_2}\, \widehat{\mathbf{e}}_2 \quad \ldots \quad \sqrt{\widehat{\lambda}_m}\, \widehat{\mathbf{e}}_m \right]$$

The estimated specific variances are diagonal entries of $\mathbf{S} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'$:

$$\widetilde{\mathbf{\Psi}} = \begin{bmatrix} \widetilde{\psi}_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \widetilde{\psi}_p \end{bmatrix} \text{ with } \widetilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \widetilde{l}_{ij}^2, i = 1, \ldots, p$$

The communalities are estimated as: $\widetilde{h}_i^2 = \widetilde{l}_{i1}^2 + \widetilde{l}_{i2}^2 + \ldots + \widetilde{l}_{im}^2$.

The principal component factor analysis of the sample correlation matrix $\mathbf{R}$ is obtained by starting with $\mathbf{R}$ in place of the matrix $\mathbf{S}$.

Principal Component Solution of the Factor Model

Further Remarks (1)

- From the PC solution, the estimated loadings for a given factor do not change when the number of factors increases; e.g.,

$$m = 1: \quad \widetilde{\mathbf{L}} = \left[\underbrace{\sqrt{\widehat{\lambda}_1}\, \widehat{\mathbf{e}}_1}_{\text{For factor 1}} \right]; \qquad m = 2: \quad \widetilde{\mathbf{L}} = \left[\underbrace{\sqrt{\widehat{\lambda}_1}\, \widehat{\mathbf{e}}_1}_{\text{For factor 1}} \quad \underbrace{\sqrt{\widehat{\lambda}_2}\, \widehat{\mathbf{e}}_2}_{\text{For factor 2}} \right]$$

- The residual matrix is given by

$$\mathbf{S} - (\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' + \widetilde{\mathbf{\Psi}})$$

whose diagonal entries are all 0. It can be shown that its off-diagonal elements satisfy:

$$\text{Sum of squared entries of } \left(\mathbf{S} - (\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' + \widetilde{\mathbf{\Psi}})\right) \leq \widehat{\lambda}^2_{m+1} + \ldots + \widehat{\lambda}^2_p$$

Small sum of squares of the neglected eigenvalues $\Longrightarrow$ small sum of squared errors of approximation

[The above result can be used for selecting the number of factors m.]

Principal Component Solution of the Factor Model

Further Remarks (2)

- Contribution to sample variance s_{ii} from the 1st common factor $= \widetilde{l}_{i1}^2$

  Contribution to total sample variance $s_{11} + \ldots + s_{pp} = \text{trace}(\mathbf{S})$ from the 1st common factor:

$$\widetilde{l}_{11}^2 + \ldots + \widetilde{l}_{p1}^2 = \left(\sqrt{\widehat{\lambda}_1}\,\widehat{\mathbf{e}}_1\right)'\left(\sqrt{\widehat{\lambda}_1}\,\widehat{\mathbf{e}}_1\right) = \widehat{\lambda}_1$$

  In general,

$$\begin{array}{c}\text{Proportion of total}\\ \text{sample variance}\\ \text{due to } j\text{-th factor}\end{array} = \begin{cases} \dfrac{\widehat{\lambda}_j}{\text{trace}(\mathbf{S})} & \text{for a factor analysis of } \mathbf{S} \\[2ex] \dfrac{\widehat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}$$

  Note that $\text{trace}(\mathbf{R}) = p$.

Example 2: Factor Analysis of Stock-price Data – (1)

- $n = 103$ weekly rates of return of $p = 5$ stocks (Table 8.4 in the text)

Week	J P Morgan	Citibank	Wells Fargo	Shell	Mobil
1	0.013034	-0.00784	-0.00319	-0.04477	0.005215
2	0.008486	0.016689	-0.00621	0.011956	0.013489
3	-0.01792	-0.00864	0.010036	0	-0.00614
4	0.021559	-0.00349	0.017435	-0.02859	-0.00695
5	0.010823	0.003717	-0.01013	0.02919	0.040975
6	0.010171	-0.0122	-0.00838	0.013708	0.00299
...	...	...	...	...	...
98	0.021745	0.022965	0.029198	0.00844	0.03193
99	0.003374	-0.01531	-0.02382	-0.00167	-0.01723
100	0.003363	0.002902	-0.00305	-0.00122	-0.0097
101	0.017015	0.009506	0.018199	-0.01618	-0.00756
102	0.010393	-0.00266	0.004429	-0.00248	-0.01645
103	-0.01279	-0.01437	-0.01874	-0.00498	-0.01637

Example 2: Factor Analysis of Stock-price Data – (2)

- From the data, it can be obtained that

$$\mathbf{R} = \begin{bmatrix} 1 & 0.632 & 0.511 & 0.115 & 0.155 \\ 0.632 & 1 & 0.574 & 0.322 & 0.213 \\ 0.511 & 0.574 & 1 & 0.183 & 0.146 \\ 0.115 & 0.322 & 0.183 & 1 & 0.683 \\ 0.155 & 0.213 & 0.146 & 0.683 & 1 \end{bmatrix}$$

with 5 eigenvalue-eigenvector pairs ($\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq \widehat{\lambda}_5$):

$$\widehat{\lambda}_1 = 2.437, \quad \widehat{\mathbf{e}}_1 = [0.469 \quad 0.532 \quad 0.465 \quad 0.387 \quad 0.361]'$$

$$\widehat{\lambda}_2 = 1.407, \quad \widehat{\mathbf{e}}_2 = [-0.368 \quad -0.236 \quad -0.315 \quad 0.585 \quad 0.606]'$$

$$\widehat{\lambda}_3 = 0.501, \quad \widehat{\mathbf{e}}_3 = [-0.604 \quad -0.136 \quad 0.772 \quad 0.093 \quad -0.109]'$$

$$\widehat{\lambda}_4 = 0.400, \quad \widehat{\mathbf{e}}_4 = [0.363 \quad -0.629 \quad 0.289 \quad -0.381 \quad 0.493]'$$

$$\widehat{\lambda}_5 = 0.255, \quad \widehat{\mathbf{e}}_5 = [0.384 \quad -0.496 \quad 0.071 \quad 0.595 \quad -0.498]'$$

Example 2: Factor Analysis of Stock-price Data – (3)

- Factor analysis using the PC method: taking $m = 1$ and $m = 2$:

$$\sqrt{\widehat{\lambda}_1}\,\widehat{\mathbf{e}}_1 = [\underbrace{0.732}_{\widetilde{l}_{11}} \quad \underbrace{0.831}_{\widetilde{l}_{21}} \quad \underbrace{0.726}_{\widetilde{l}_{31}} \quad \underbrace{0.605}_{\widetilde{l}_{41}} \quad \underbrace{0.563}_{\widetilde{l}_{51}}]'$$

$$\sqrt{\widehat{\lambda}_2}\,\widehat{\mathbf{e}}_2 = [\underbrace{-0.437}_{\widetilde{l}_{12}} \quad \underbrace{-0.280}_{\widetilde{l}_{22}} \quad \underbrace{-0.374}_{\widetilde{l}_{32}} \quad \underbrace{0.694}_{\widetilde{l}_{42}} \quad \underbrace{0.719}_{\widetilde{l}_{52}}]'$$

$$\begin{aligned} m = 1: \quad \widetilde{\boldsymbol{\Psi}} &= \text{diag}\left\{1 - \widetilde{l}_{11}^2, 1 - \widetilde{l}_{21}^2, 1 - \widetilde{l}_{31}^2, 1 - \widetilde{l}_{41}^2, 1 - \widetilde{l}_{51}^2\right\} \\ &= \text{diag}\{0.46, 0.31, 0.47, 0.63, 0.68\} \quad (\widetilde{\mathbf{L}} = \sqrt{\widehat{\lambda}_1}\,\widehat{\mathbf{e}}_1) \\ m = 2: \quad \widetilde{\boldsymbol{\Psi}} &= \text{diag}\left\{1 - (\widetilde{l}_{11}^2 + \widetilde{l}_{12}^2), 1 - (\widetilde{l}_{21}^2 + \widetilde{l}_{22}^2), 1 - (\widetilde{l}_{31}^2 + \widetilde{l}_{32}^2), \right. \\ &\qquad \left. 1 - (\widetilde{l}_{41}^2 + \widetilde{l}_{42}^2), 1 - (\widetilde{l}_{51}^2 + \widetilde{l}_{52}^2)\right\} \\ &= \text{diag}\{0.27, 0.23, 0.33, 0.15, 0.17\} \quad (\widetilde{\mathbf{L}} = \left[\sqrt{\widehat{\lambda}_1}\widehat{\mathbf{e}}_1 \; \sqrt{\widehat{\lambda}_2}\,\widehat{\mathbf{e}}_2\right]) \end{aligned}$$

diag$\{\ldots\}$: diagonal matrix with entries given by the arguments

Example 2: Factor Analysis of Stock-price Data – (4)

- From previous page:

variables	1-factor solution		2-factor solution		
	Est. factor loading 1	$\widetilde{\psi}_i$	Est. factor loading 1	Est. factor loading 2	$\widetilde{\psi}_i$
1. J P Morgan	0.732	0.46	0.732	-0.437	0.27
2. Citibank	0.831	0.31	0.831	-0.280	0.23
3. Wells Fargo	0.726	0.47	0.726	-0.374	0.33
4. Shell	0.605	0.63	0.605	0.694	0.15
5. Mobil	0.563	0.68	0.563	0.719	0.17
Cum. prop.	0.487		0.487	0.769	

The last row shows the cumulative proportion of total (standardized) sample variance. For $m = 1$, it is $\widehat{\lambda}_1/p = 2.437/5 = 0.487$; for $m = 2$, with both factors, it is $(\widehat{\lambda}_1 + \widehat{\lambda}_2)/p = (2.437 + 1.407)/5 = 0.769$;

Example 2: Factor Analysis of Stock-price Data – (5)

- The residual matrix for the case with $m = 2$ is given by:

$$
\begin{aligned}
&\mathbf{R} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' - \widetilde{\boldsymbol{\Psi}} \\
&= \begin{bmatrix} 0 & -0.099 & -0.185 & -0.025 & 0.056 \\ -0.099 & 0 & -0.134 & 0.014 & -0.054 \\ -0.185 & -0.134 & 0 & 0.003 & 0.006 \\ -0.025 & 0.014 & 0.003 & 0 & -0.156 \\ 0.056 & -0.054 & 0.006 & -0.156 & 0 \end{bmatrix}
\end{aligned}
$$

Note: Throughout the analysis of stock-price data, we work on standardized variables, but omit the subscript z.

The Maximum Likelihood Method (i): Basic Idea

- Normal common factors $\mathbf{F}$ and specific factors $\boldsymbol{\varepsilon}$ $\Rightarrow$ ML estimation
  $\mathbf{F}_j$, $\boldsymbol{\varepsilon}_j$: jointly normal $\Longrightarrow$ random samples $\mathbf{X}_j = \boldsymbol{\mu} + \mathbf{L}\mathbf{F}_j + \boldsymbol{\varepsilon}_j, j = 1, \ldots, n$, i.i.d. normal

  1. The likelihood:

  $$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{\frac{-(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} e^{-\frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})'\right)\right]}$$

  $$\times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{n}{2}(\overline{\mathbf{x}}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})'}$$

     which depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$.
     (Here we obtain $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}$; from the above, we may get $\widehat{\mathbf{L}}$ and $\widehat{\boldsymbol{\Psi}}$.)

  2. But: ambiguity of $\mathbf{L} \Longrightarrow$ not a well-defined model!
     To make the choice of $\mathbf{L}$ unique, impose the following constraint

  $$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta} \text{ (a diagonal matrix)}$$

- The ML estimates $\widehat{\mathbf{L}}$ and $\widehat{\boldsymbol{\Psi}}$ (from the above procedure) are obtained through numerical maximization (computer-aided).

The Maximum Likelihood Method (ii): Result

$\mathbf{X}_1, \ldots, \mathbf{X}_n$:	random sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$:	the cov. matrix for the orthogonal factor model with m common factors.

Then, the ML estimators $\widehat{\mathbf{L}}$, $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$ maximizes $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ subject to $\widehat{\mathbf{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{L}}$ being diagonal.

The ML estimates of the communalities are:

$$\widehat{h}_i^2 = \widehat{l}_{i1}^2 + \widehat{l}_{i2}^2 + \ldots + \widehat{l}_{im}^2, \quad i = 1, 2, \ldots, p$$

and

$$\begin{array}{l}\text{Proportion of total sample} \\ \text{variance due to } j\text{-th factor}\end{array} = \frac{\widehat{l}_{1j}^2 + \widehat{l}_{2j}^2 + \ldots + \widehat{l}_{pj}^2}{s_{11} + s_{22} + \ldots + s_{pp}}$$

Invariance property of ML estimates has been used to obtain the estimates of communalities.

The Maximum Likelihood Method (iii)

Standardized Variables

- Normal random samples $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; recall standardized variables:

$$\mathbf{Z} = \mathbf{V}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}), \quad \mathbf{V}^{\frac{1}{2}} = \text{diag}\{\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{pp}}\}$$

- The covariance matrix $\boldsymbol{\rho}$ of $\mathbf{Z}$: [Factorization of $\boldsymbol{\rho}$]

$$\begin{aligned}\boldsymbol{\rho} =& \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Sigma}\mathbf{V}^{-\frac{1}{2}} = (\mathbf{V}^{-\frac{1}{2}}\mathbf{L})(\mathbf{V}^{-\frac{1}{2}}\mathbf{L})' + \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Psi}\mathbf{V}^{-\frac{1}{2}} \\ =& \mathbf{L}_z\mathbf{L}_z' + \boldsymbol{\Psi}_z \quad (\mathbf{L}_z = \mathbf{V}^{-\frac{1}{2}}\mathbf{L}, \boldsymbol{\Psi}_z = \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Psi}\mathbf{V}^{-\frac{1}{2}})\end{aligned}$$

- ML estimate of $\boldsymbol{\rho}$: (using invariance property of ML estimator)

$$\widehat{\boldsymbol{\rho}} = \widehat{\mathbf{L}}_z\widehat{\mathbf{L}}_z' + \widehat{\boldsymbol{\Psi}}_z, \text{ where } \widehat{\mathbf{L}}_z = \widehat{\mathbf{V}}^{-\frac{1}{2}}\widehat{\mathbf{L}}, \widehat{\boldsymbol{\Psi}}_z = \widehat{\mathbf{V}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Psi}}\widehat{\mathbf{V}}^{-\frac{1}{2}}$$

$\widehat{\mathbf{V}}^{-\frac{1}{2}}$ (or $\widehat{\mathbf{L}}$): ML estimator of $\mathbf{V}^{-\frac{1}{2}}$ (or $\mathbf{L}$)

diag(. . .): a diagonal matrix with diagonal entries given by the arguments

The Maximum Likelihood Method (iv)
Standardized Variables

- Working properly with sample correlation matrix $\mathbf{R}$ instead of $\mathbf{S}$ in the likelihood

  Below let $\widehat{l}_{ij}$'s denote the elements of $\widehat{\mathbf{L}}_z$
- ML estimates of the communalities:

$$\widehat{h}_i^2 = \widehat{l}_{i1}^2 + \widehat{l}_{i2}^2 + \ldots + \widehat{l}_{im}^2, \qquad i = 1, 2, \ldots, p$$

- Similarly:

$$\begin{array}{c}\text{Proportion of total (standardized)} \\ \text{sample variance due to } j\text{-th factor}\end{array} = \frac{\widehat{l}_{1j}^2 + \widehat{l}_{2j}^2 + \ldots + \widehat{l}_{pj}^2}{p}$$

Example 3: Factor Analysis of Stock-price Data – (1)

- Revisit the stock-price data and use the ML method for factor analysis

Week	J P Morgan	Citibank	Wells Fargo	Shell	Mobil
1	0.013034	-0.00784	-0.00319	-0.04477	0.005215
2	0.008486	0.016689	-0.00621	0.011956	0.013489
3	-0.01792	-0.00864	0.010036	0	-0.00614
4	0.021559	-0.00349	0.017435	-0.02859	-0.00695
5	0.010823	0.003717	-0.01013	0.02919	0.040975
6	0.010171	-0.0122	-0.00838	0.013708	0.00299
…	…	…	…	…	…
98	0.021745	0.022965	0.029198	0.00844	0.03193
99	0.003374	-0.01531	-0.02382	-0.00167	-0.01723
100	0.003363	0.002902	-0.00305	-0.00122	-0.0097
101	0.017015	0.009506	0.018199	-0.01618	-0.00756
102	0.010393	-0.00266	0.004429	-0.00248	-0.01645
103	-0.01279	-0.01437	-0.01874	-0.00498	-0.01637

Example 3: Factor Analysis of Stock-price Data – (2)

- As said, the ML method is usually computer aided and is built in many computing softwares.

  For example, we use the function "factoran.m" in the statistics toolbox in MATLAB, which takes the original data matrix as an input.

  For $m = 2$, we obtain:

$$\widehat{\mathbf{L}} = \begin{bmatrix} 0.1206 & 0.7543 \\ 0.3285 & 0.7857 \\ 0.1876 & 0.6502 \\ 0.9975 & -0.0071 \\ 0.6852 & 0.0263 \end{bmatrix}$$

$$\widehat{\mathbf{\Psi}} = \text{diag}\,\{0.4165, 0.2747, 0.5420, 0.0050, 0.5298\}$$

The default "factoran.m" gives rotated version of the estimated factor loadings; here we unrotate them.

Example 3: Factor Analysis of Stock-price Data – (3)

- The residual matrix for the case with $m = 2$ is given by:

$$
\begin{aligned}
&\mathbf{R} - \widehat{\mathbf{L}}\widehat{\mathbf{L}}' - \widehat{\mathbf{\Psi}} \\
&= \begin{bmatrix} 0 & 0.0000 & -0.0026 & -0.0003 & 0.0520 \\ 0.0000 & 0 & 0.0016 & 0.0002 & -0.0331 \\ -0.0026 & 0.0016 & 0 & -0.0000 & 0.0006 \\ -0.0003 & 0.0002 & -0.0000 & 0 & 0.0001 \\ 0.0520 & -0.0331 & 0.0006 & 0.0001 & 0 \end{bmatrix}
\end{aligned}
$$

Example 3: Factor Analysis of Stock-price Data – (4)

In the table below, we show the results for $m = 2$ using both PC and ML methods.

variables	ML Method			PC Method		
	Est. load. 1	Est. load. 2	$\widehat{\psi}_i$	Est. load. 1	Est. load. 2	$\widetilde{\psi}_i$
1. J P M.	0.121	0.754	0.4165	0.732	-0.437	0.27
2. Citi	0.329	0.786	0.2747	0.831	-0.280	0.23
3. W. F.	0.188	0.650	0.5420	0.726	-0.374	0.33
4. Shell	0.998	-0.007	0.0050	0.605	0.694	0.15
5. Mobil	0.685	0.026	0.5298	0.563	0.719	0.17
Cum. prop.	0.324	0.646		0.487	0.769	

The last row shows the cumulative proportion of total (standardized) sample variance. Using ML: $m = 1$, it is $\sum_{i=1}^{5} \widehat{l}_{i1}^2/5 = 0.324$; for $m = 2$, with both factors, it is $\sum_{i=1}^{5}(\widehat{l}_{i1}^2 + \widehat{l}_{i2}^2)/5 = 0.646$; $(p = 5)$.

A Large Sample Test for the Number of Common Factors

Part (1)

- Assuming normality enables testing the adequacy of the factor model;
- Testing the adequacy of the m common factor model is equivalent to

$$\text{testing } H_0 : \underbrace{\boldsymbol{\Sigma}}_{p\times p} = \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{L}'}_{m\times p} + \underbrace{\boldsymbol{\Psi}}_{p\times p}$$

$$\text{vs. } H_1 : \boldsymbol{\Sigma} \text{ any other positive definite (p. d.) matrix}$$

a. Under H_1: no specific form of $\boldsymbol{\Sigma}$; from previous results on maximum likelihood estimation (with both $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ unknown):

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S} \triangleq \mathbf{S}_n$$

$$\text{Maximized likelihood} \propto |\mathbf{S}_n|^{-\frac{n}{2}} e^{-\frac{np}{2}}$$

($\propto$ means proportional to)

A Large Sample Test for the Number of Common Factors

Part (2)

(Cont'd)

- Testing

$$H_0 : \underbrace{\mathbf{\Sigma}}_{p\times p} = \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{L}'}_{m\times p} + \underbrace{\mathbf{\Psi}}_{p\times p} \text{ vs. } H_1 : \mathbf{\Sigma} \text{ any other p. d. matrix}$$

b. Under H_0: (let $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{\Psi}}$ be the ML estimates of $\mathbf{L}$ and $\mathbf{\Psi}$, respectively)

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}, \quad \widehat{\mathbf{\Sigma}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}}$$

Maximized likelihood

$$\propto |\widehat{\mathbf{\Sigma}}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[\widehat{\mathbf{\Sigma}}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})'\right)\right]\right)$$

$$= |\widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}}|^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\text{tr}\left[(\widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}})^{-1}\mathbf{S}_n\right]\right)$$

A Large Sample Test for the Number of Common Factors

Part (3)

(Cont'd)

- Testing $H_0 : \underbrace{\mathbf{\Sigma}}_{p\times p} = \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{L}'}_{m\times p} + \underbrace{\mathbf{\Psi}}_{p\times p}$ vs. $H_1 : \mathbf{\Sigma}$ any other p. d. matrix

  c. The likelihood ratio statistic:

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$= -2\ln\left(\frac{|\widehat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-\frac{n}{2}} + n\left[\text{tr}(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p\right]$$

  with degrees of freedom

$$v - v_0 = \frac{1}{2}p(p+1) - [p(m+1) - \frac{1}{2}m(m-1)]$$

$$= \frac{1}{2}[(p-m)^2 - p - m]$$

A Large Sample Test for the Number of Common Factors

Part (4)

(Cont'd)

- Testing $H_0 : \underbrace{\mathbf{\Sigma}}_{p\times p} = \underbrace{\mathbf{L}}_{p\times m}\underbrace{\mathbf{L}'}_{m\times p} + \underbrace{\mathbf{\Psi}}_{p\times p}$ vs. $H_1 : \mathbf{\Sigma}$ any other p. d. matrix

d. It can be shown that: if $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}}$ is the ML estimate of $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$,

$$\text{then} \quad \text{tr}(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p = 0 \implies -2\ln\Lambda = n\ln(|\widehat{\mathbf{\Sigma}}|/|\mathbf{S}_n|)$$

Using Bartlett's correction: when p and $n-p$ are large

$\implies$ reject H_0 at significance level α, if

$$\left(n - 1 - \frac{2p+4m+5}{6}\right)\ln\frac{|\widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

To make the # of degrees of freedom $\frac{1}{2}[(p-m)^2 - p - m]$ positive,

$$m < \frac{1}{2}\left(2p + 1 - \sqrt{8p+1}\right)$$

Factor Rotation

- Factor rotation refers to: an orthogonal transformation of the factor loadings as well as the implied orthogonal transformation of the factors
- $\widehat{\mathbf{L}}_{p\times m}$: estimated factor loading matrix (e.g., from PC or ML methods)

  then : $\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\mathbf{T}$ (“rotated” loadings) where : $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$

  The following quantities remain unchanged:

  1) Estimated cov. matrix: $\widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}\mathbf{T}\mathbf{T}'\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*'} + \widehat{\mathbf{\Psi}}$
  2) Residual matrix: $\mathbf{S}_n - \widehat{\mathbf{L}}\widehat{\mathbf{L}}' - \widehat{\mathbf{\Psi}} = \mathbf{S}_n - \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*'} - \widehat{\mathbf{\Psi}}$
  3) Specific variances $\widehat{\psi}_i$ and communalities $\widehat{h}_i^2$
- When $m = 2$, use

$$\underbrace{\mathbf{T} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}}_{\text{clockwise rotation}}, \quad \text{or} \quad \underbrace{\mathbf{T} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}}_{\text{counterclockwise rotation}}$$

Factor Rotation: Illustration (1)

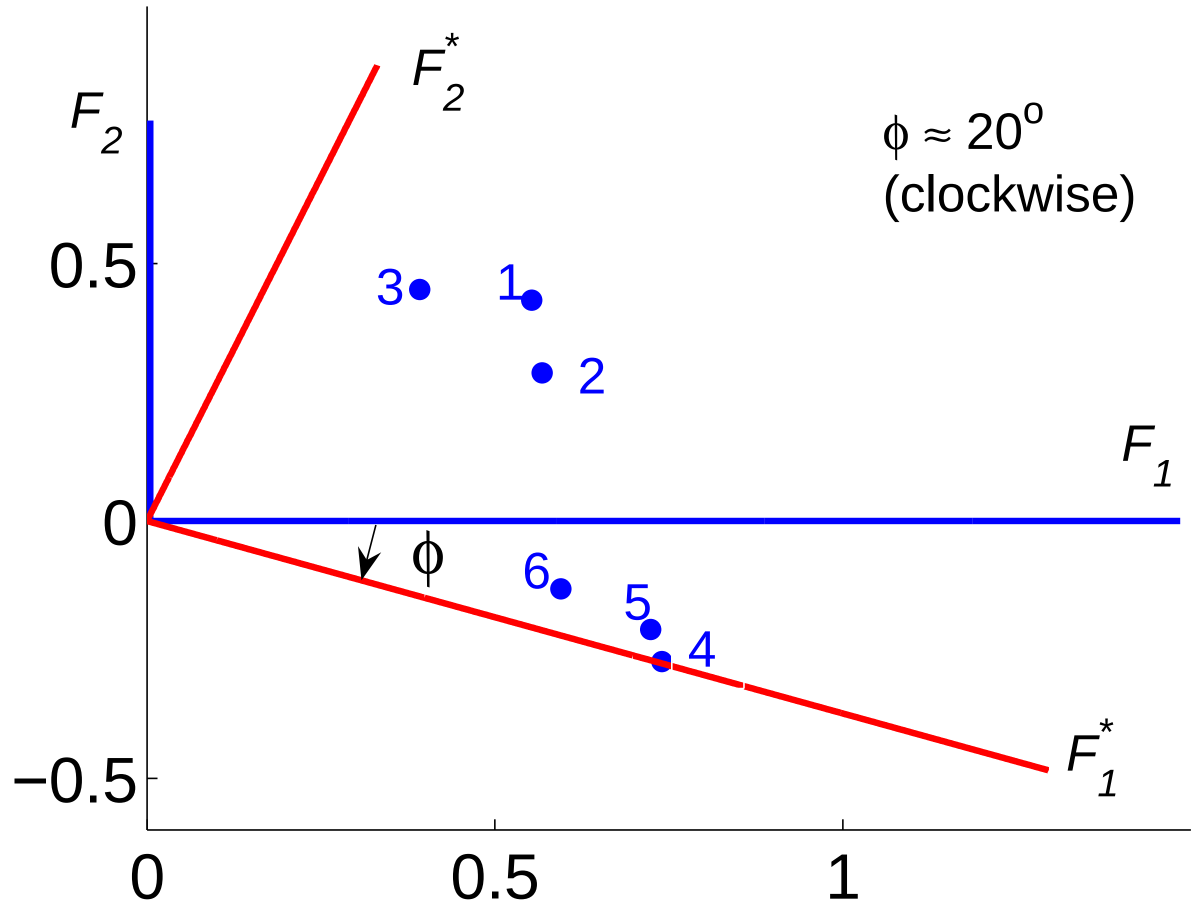
- (Example 9.8 in the text): Sample correlation matrix of exam scores in $p = 6$ subject areas for $n = 220$ male students:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.439 & 0.410 & 0.288 & 0.329 & 0.248 \\ & 1 & 0.351 & 0.354 & 0.320 & 0.329 \\ & & 1 & 0.164 & 0.190 & 0.181 \\ & & & 1 & 0.595 & 0.470 \\ & & & & 1 & 0.464 \\ & & & & & 1 \end{bmatrix}$$

- ML solution for $m = 2$

	Est. Factor	Loadings	Communalities
variables	F_1	F_2	$\widehat{h}_i^2$
1. Gaelic	0.553	0.429	0.490
2. English	0.568	0.288	0.406
3. History	0.392	0.450	0.356
4. Arithmetic	0.740	-0.273	0.623
5. Algebra	0.724	-0.211	0.569
6. Geometry	0.595	-0.132	0.372

Factor Rotation: Illustration (2)



Factor Rotation: Illustration (3)

- The rotated factor loadings with $\phi \approx 20^\circ$ (clockwise)

variables	Est. Rot. Factor F_1^*	Loadings F_2^*	Communalities $\widehat{h}_i^{2*} = \widehat{h}_i^2$
1. Gaelic	0.369	$\boxed{0.594}$	0.490
2. English	0.433	$\boxed{0.467}$	0.406
3. History	0.211	$\boxed{0.558}$	0.356
4. Arithmetic	$\boxed{0.789}$	0.001	0.623
5. Algebra	$\boxed{0.752}$	0.054	0.569
6. Geometry	$\boxed{0.604}$	0.083	0.372

Factor Rotation: Kaiser's "varimax"

- In applications, choose rotation to obtain a model with "simple structure"
- An **analytical** method: Kaiser's varimax criterion: choose $\mathbf{T}$ to make

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\widetilde{l}_{ij}^{*4} - \frac{1}{p}\left(\sum_{i=1}^{p}\widetilde{l}_{ij}^{*2}\right)^2\right]$$

as large as possible, where

$$\widetilde{l}_{ij}^{*} \triangleq \widehat{l}_{ij}^{*}/\widehat{h}_i$$

- varimax: available in many computing softwares

Example 4: Factor Rotation in Stock-price Data Analysis

- Previously: factor loadings using the ML method without factor rotation
- Below: we show the effect of factor rotation using varimax for $m = 2$

variables	ML Method		Rotated loadings		
	Est. load. 1	Est. load. 2	Rotated load. 1	Rotated load. 2	$\widehat{\psi}_i$
1. J P M.	0.121	0.754	0.7635	0.0242	0.4165
2. Citi	0.329	0.786	0.8210	0.2265	0.2747
3. W. F.	0.188	0.650	0.6687	0.1038	0.5420
4. Shell	0.998	-0.007	0.1191	0.9904	0.0050
5. Mobil	0.685	0.026	0.1128	0.6763	0.5298
Cum. prop.	0.324	0.646	0.346	0.646	

After rotation, stocks from companies 1, 2, 3 load highly on the first factor, while stocks from companies 4, 5 load highly on the second factor.

Factor Scores — i)

- Factor scores refer to: estimated values of the common factors
  i) Used, e.g., as input for further analysis;
  ii) Not the estimates of unknown parameters in the usual sense; they are estimates of the unobserved random factor vectors $\mathbf{F}_j, j = 1, \ldots, n$;
  iii) Unobserved quantities $\mathbf{f}_j$ and ε_j outnumbering the observed data $\mathbf{x}_j$ ($\mathbf{f}_j$: attained values of $\mathbf{F}_j$)
- Two methods here:
  1) Bartlett's weighted least squares method
  2) Regression method
- Common elements in the above two methods
  - Treating the estimated $\widehat{l}_{ij}$ and $\widehat{\psi}_i$ as true values
  - Linear transformations of original data (centered or standardized)

Factor Scores — ii)

Bartlett's Weighted Least Squares Method

- Assuming $\mathbf{L}, \boldsymbol{\mu}, \boldsymbol{\Psi}$ known in:

$$\mathbf{X}_{p\times 1} - \boldsymbol{\mu}_{p\times 1} = \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\varepsilon}_{p\times 1}$$

  and treating specific factors $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p]'$ as errors

  Bartlett's method: given observations $\mathbf{x}$, choose $\widehat{\mathbf{f}}$ of $\mathbf{f}$ and solve

$$\min_{\mathbf{f}} \sum_{i=1}^{p} \frac{\varepsilon_i^2}{\psi_i} \Longleftrightarrow \min_{\mathbf{f}} \boldsymbol{\varepsilon}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon}$$

$$\Longleftrightarrow \min_{\mathbf{f}} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})$$

  The solution: $\widehat{\mathbf{f}} = (\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu})$

- In applications: $\widehat{\mathbf{f}}_j = (\widehat{\mathbf{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{L}})^{-1}\widehat{\mathbf{L}}'\widehat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}), \forall j$

  With the ML method for $\widehat{\mathbf{L}}$ and $\widehat{\boldsymbol{\Psi}}$, $\widehat{\mathbf{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{L}} = \widehat{\boldsymbol{\Delta}}$ (diagonal)

Factor Scores — iii)
Bartlett's Weighted Least Squares Method

Factor Scores Using Weighted Least Squares from the ML Estimates

$$\begin{aligned}\widehat{\mathbf{f}}_j &= (\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}})^{-1}\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}) \\ &= \widehat{\mathbf{\Delta}}^{-1}\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}), \quad j = 1, 2, \ldots, n\end{aligned}$$

or, if the correlation matrix is factored

$$\begin{aligned}\widehat{\mathbf{f}}_j &= (\widehat{\mathbf{L}}'_z\widehat{\mathbf{\Psi}}_z^{-1}\widehat{\mathbf{L}}_z)^{-1}\widehat{\mathbf{L}}'_z\widehat{\mathbf{\Psi}}_z^{-1}\mathbf{z}_j \\ &= \widehat{\mathbf{\Delta}}_z^{-1}\widehat{\mathbf{L}}'_z\widehat{\mathbf{\Psi}}_z^{-1}\mathbf{z}_j, \quad j = 1, 2, \ldots, n\end{aligned}$$

where $\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \overline{\mathbf{x}})$, and $\widehat{\boldsymbol{\rho}} = \widehat{\mathbf{L}}_z\widehat{\mathbf{L}}'_z + \widehat{\mathbf{\Psi}}_z$.

Note: factor scores generated above have zero sample mean vector and zero sample correlations.

Factor Scores — iv)
The Regression Method

- Assuming that $\mathbf{L}, \boldsymbol{\Psi}$ are known in: $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$

  assuming the orthogonal factor model and joint normality of $\mathbf{F}$, $\boldsymbol{\varepsilon}$:

$$(\mathbf{X} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{LL}' + \boldsymbol{\Psi}) \qquad (\text{note}: \ \boldsymbol{\Sigma}_{p\times p} = \mathbf{LL}' + \boldsymbol{\Psi})$$

$$\begin{pmatrix} \mathbf{X} - \boldsymbol{\mu} \\ \mathbf{F} \end{pmatrix} \sim N_{m+p}\left(\mathbf{0}_{(m+p)\times 1}, \boldsymbol{\Sigma}^*_{(m+p)\times(m+p)}\right), \boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p} & \mathbf{L}_{p\times m} \\ \mathbf{L}' & \mathbf{I}_{m\times m} \end{bmatrix}$$

  then the conditional dist. of $\mathbf{F}|\mathbf{x}$ is normal with

$$\begin{aligned} \mathsf{E}(\mathbf{F}|\mathbf{x}) &= \mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ \mathsf{Cov}(\mathbf{F}|\mathbf{x}) &= \mathbf{I} - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}\mathbf{L} \end{aligned}$$

  $\mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}$: coefficients in a regression of the factors on the variables

Factor Scores — v)

The Regression Method

(Cont'd) Assuming that (1) $\mathbf{L}, \mathbf{\Psi}$ are known in: $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$, (2) the orthogonal factor model and (3) joint normality of $\mathbf{F}$, $\boldsymbol{\varepsilon}$, then the conditional dist. of $\mathbf{F}|\mathbf{x}$ is normal with

$$\mathsf{E}(\mathbf{F}|\mathbf{x}) = \mathbf{L}'(\mathbf{LL}' + \mathbf{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu}); \quad \mathsf{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}'(\mathbf{LL}' + \mathbf{\Psi})^{-1}\mathbf{L}$$

- Thus, taking the ML estimates $\widehat{\mathbf{L}}, \widehat{\mathbf{\Psi}}$ as true values, we obtain:

$$\widehat{\mathbf{f}}_j = \widehat{\mathbf{L}}'(\widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}})^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}) = \mathbf{L}'\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}})$$

In applications, we use $\mathbf{S}$ to replace $\widehat{\mathbf{\Sigma}}$

$$\widehat{\mathbf{f}}_j = \mathbf{L}'\mathbf{S}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}})$$

(See a summary next page)

Factor Scores — vi)
The Regression Method

Factor Scores Obtained by Regression

$$\widehat{\mathbf{f}}_j = \mathbf{L}'\mathbf{S}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}), \quad j = 1, 2, \ldots, n$$

or, if the correlation matrix is factored

$$\widehat{\mathbf{f}}_j = \widehat{\mathbf{L}}'_z\mathbf{R}^{-1}\mathbf{z}_j, \quad j = 1, 2, \ldots, n$$

where $\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \overline{\mathbf{x}})$, and $\widehat{\boldsymbol{\rho}} = \widehat{\mathbf{L}}_z\widehat{\mathbf{L}}'_z + \widehat{\boldsymbol{\Psi}}_z$.

Factor Scores — vii)

Relation between Bartlett's Method and the Regression Method

- Let the factor scores from Bartlett's method and the regression method be denoted as $\widehat{\mathbf{f}}_j^{\mathsf{WLS}}$ and $\widehat{\mathbf{f}}_j^{\mathsf{R}}$, respectively. It can be shown that (see the textbook):

$$\widehat{\mathbf{f}}_j^{\mathsf{WLS}} = \left[\mathbf{I} + \left(\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}}\right)^{-1}\right]\widehat{\mathbf{f}}_j^{\mathsf{R}}$$

- If ML estimates are used: $\left(\widehat{\mathbf{L}}'\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}}\right)^{-1} = \widehat{\mathbf{\Delta}}^{-1}$ (diagonal). If all the entries on this diagonal matrix are small enough, then the two methods yield similar factor scores.

Example 5: Computing Factor Scores – (1)

Stock-price Data

- In Example 4, using ML, the estimated rotated loadings and specific variances are: $(m = 2)$

$$\widehat{\mathbf{L}}^* = \begin{bmatrix} 0.7635 & 0.0242 \\ 0.8210 & 0.2265 \\ 0.6687 & 0.1038 \\ 0.1191 & 0.9904 \\ 0.1128 & 0.6763 \end{bmatrix}$$

$$\widehat{\mathbf{\Psi}} = \text{diag}\{0.4165, 0.2747, 0.5420, 0.0050, 0.5298\}$$

- The vector of standardized observations: (given)

$$\mathbf{z} = [0.50 \quad -1.40 \quad -0.20 \quad -0.70 \quad 1.40]'$$

Example 5: Computing Factor Scores – (2)

Stock-price Data

Scores on factors 1 and 2:

- Using Bartlett's weighted LS:

$$\widehat{\mathbf{f}}^{\mathsf{WLS}} = \left(\widehat{\mathbf{L}}^{*\prime}\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}}^{*}\right)^{-1}\widehat{\mathbf{L}}^{*}\widehat{\mathbf{\Psi}}^{-1}\mathbf{z} = \begin{bmatrix} -0.6004 \\ -0.6250 \end{bmatrix}$$

- Using the regression method: ($\mathbf{R}$ can be found from Example 2)

$$\widehat{\mathbf{f}}^{\mathsf{R}} = \widehat{\mathbf{L}}^{*\prime}\mathbf{R}\mathbf{z} = \begin{bmatrix} -0.5052 \\ -0.6336 \end{bmatrix}$$

In this case, both methods produce very similar result.

Example 5: Computing Factor Scores – (3)

Stock-price Data