# Lecture 10

## Factor Analysis

Reading: Johnson & Wichern 2007, Chapter 9; Zelterman Chapter 8.5; Izenman Chapter 15.4

*DSA 8070 Multivariate Analysis*

Whitney Huang
Clemson University

# Agenda

1. **Background**
2. **Factor Model Analysis**
3. **Stock Price Data Example**

# Introductory Example: Intelligence Tests [Smith & Stanley (1983)]

Six tests (general, picture, blocks, maze, reading, vocab) were given to 112 individuals. The resulting sample correlation matrix of these tests is as follows:

| | general | picture | blocks | maze | reading | vocab |
|---|---|---|---|---|---|---|
| general | 1.000 | 0.466 | 0.552 | 0.340 | 0.576 | 0.514 |
| picture | 0.466 | 1.000 | 0.572 | 0.193 | 0.263 | 0.239 |
| blocks | 0.552 | 0.572 | 1.000 | 0.445 | 0.354 | 0.356 |
| maze | 0.340 | 0.193 | 0.445 | 1.000 | 0.184 | 0.219 |
| reading | 0.576 | 0.263 | 0.354 | 0.184 | 1.000 | 0.791 |
| vocab | 0.514 | 0.239 | 0.356 | 0.219 | 0.791 | 1.000 |

Can the correlation between the six tests be explained by one or two variables describing some general concept of intelligence?

# Overview



Factor Analysis (FA) assumes the covariance structure among a set of variables, $\boldsymbol{X} = (X_1, \cdots, X_p)^T$, can be described via a linear combination of unobservable (latent) variables $\boldsymbol{F} = (F_1, \cdots, F_m)^T$, called factors.

There are three typical objectives of FA:

1. Data reduction: explain covariance between $p$ variables using $m < p$ latent factors
2. Data interpretation: find features (i.e., factors) that are important for explaining covariance ⇒ exploratory FA
3. Theory testing: determine if hypothesized factor strucuture fits observed data ⇒ confirmatory FA

# FA and PCA

FA and PCA have similar themes, i.e., to explain covariance between variables via linear combinations of other variables

However, there are distinctions between the two approaches:

- FA assumes a statistical model that describes covariation in observed variables via linear combinations of latent variables
- PCA finds uncorrelated linear combinations of observed variables that explain maximal variance

FA refers to a statistical model, whereas PCA refers to the eigenvalue decomposition of a covariance (or correlation) matrix

## Factor Model

Let $\boldsymbol{X}$ is a random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The factor model postulates that $\boldsymbol{X}$ can be written as a linear combination of a set of $m$ common factors $F_1, F_2, \cdots, F_m$:

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

where

- $\{\ell_{jk}\}_{p\times m}$ denotes the matrix of factor loadings, that is, $\ell_{jk}$ is the loading (importance) of the $j$-th variable on the $k$-th factor
- $(F_1, \cdots, F_m)^T$ denotes the vector of the latent factor scores, that is, $F_k$ is the score on the $k$-th factor
- $(\varepsilon_1, \cdots, \varepsilon_p)^T$ denotes the vector of latent error terms, which correspond to the random disturbances specific to each variable

# Factor Model in Matrix Notation







The factor model can be written in a matrix form:

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon},$$

where

- $\boldsymbol{L} = \{\ell_{jk}\}_{p\times m}$ is the factor loading matrix
- $\boldsymbol{F} = (F_1, \cdots, F_m)^T$ is the factor score vector
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)^T$ is the (latent) error vector

Unlike in linear model, we do not observe $\boldsymbol{F}$, therefore we need to impose some assumptions to facilitate the model identification

# Factor Model Assumptions

First, we assume:

$$\mathbb{E}(\boldsymbol{F}) = \mathbf{0}, \qquad \mathbb{Var}(\boldsymbol{F}) = \mathbb{E}(\boldsymbol{F}\boldsymbol{F}^T) = \boldsymbol{I}$$
$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \mathbb{Var}(\boldsymbol{\varepsilon}) = \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \boldsymbol{\Psi} = \mathrm{diag}(\psi_i), i = 1, \cdots, p$$

Moreover, we assume $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are independent, so that $\mathbb{Cov}(\boldsymbol{F}, \boldsymbol{\varepsilon}) = \mathbf{0}$

- The factors have variance one (i.e., $\mathbb{Var}(F_i) = 1$) and uncorrelated with one another
- The error vector are uncorrelated with one another with the specific variance $\mathbb{Var}(\varepsilon_i) = \psi_i$
- Under the model assumptions, we have

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon} \Leftrightarrow \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$$

# Variances and Covariances of Factor Models

Under the factor model, we have

$$\mathbb{Var}(X_i) = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 + \psi_i$$

$$\mathbb{Cov}(X_i, X_j) = \ell_{i1}\ell_{j1} + \ell_{i2}\ell_{j2} + \cdots + \ell_{im}\ell_{jm}$$

The portion of the variance that is contributed by the $m$ common factors is the communality:

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2,$$

and the portion that is not explained by the common factors is called the uniqueness (or the specific variance):

$$\mathbb{Var}(\varepsilon_i) = \psi_i = \mathbb{Var}(X_i) - h_i^2$$

To be determined: 1) number $m$ of common factors; 2) factor loadings $\boldsymbol{L}$; and 3) specific variances $\boldsymbol{\Psi}$

# Choosing the Number of Common Factors

- The factor model assumes that the $p(p+1)/2$ variances and covariances of $X$ can be reproduced from the $p(m+1)$ factor loadings and the variances of the $p$ unique factors
- Situations in which $m$, the number of common factors, is small relative to $p$ is when factor analysis works best. For example, if $p = 12$ and $m = 2$, then the $(12 \times 13)/2 = 78$ elements of $\mathbf{\Sigma}$ can be reproduced from $12 \times (2+1) = 36$ parameters in the factor model
- However, if $m$ is too small, the $p(m+1)$ parameters may not be adequate to describe $\mathbf{\Sigma}$

# Estimation in Factor Models

Given $m$, we consider two methods to estimate the parameters of a factor model:

- Principal Component Method

$$\text{PCA}: \quad \boldsymbol{\Sigma} = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^T + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2^T + \cdots + \lambda_p \boldsymbol{e}_p \boldsymbol{e}_p^T$$

$$\text{Factor Model:} \quad \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$$

  **Main idea**: Use the first $m$ PCs to form the factor loading matrix, then use the relationship $\boldsymbol{\Psi} = \boldsymbol{\Sigma} - \boldsymbol{L}\boldsymbol{L}^T$ to estimate the specific variances $\hat{\psi}_i = s_i^2 - \sum_{j=1}^m \lambda_j \hat{e}_{ji}^2$

- Maximum Likelihood Estimation: assuming data $\boldsymbol{X} \overset{i.i.d.}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi})$, maximizing the log-likelihood $\ell(\boldsymbol{\mu}, \boldsymbol{L}, \boldsymbol{\Psi}) \propto$ $-\frac{n}{2}\log|\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}| - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{X}_i - \boldsymbol{\mu})^T (\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi})^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu})$ to obtain the parameter estimates

# A Goodness-of-Fit Test for Factor Model

We wish to test whether the factor model (with a given $m$) appropriately describes the covariances among the $p$ variables. Specifically, we test

$$H_{0(m)} : \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$$

versus

$$H_1 : \boldsymbol{\Sigma} \text{ is an unconstrained covariance matrix}$$

- Bartlett-Corrected Likelihood Ratio Test Statistic

$$-2\log\Lambda = (n - 1 - (2p + 4m + 5)/6)\log\frac{|\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}}|}{|\hat{\boldsymbol{\Sigma}}|}$$

- Reject $H_0$ at level $\alpha$ if $-2\log\Lambda > \chi^2_{df=\frac{1}{2}[(p-m)^2-p-m]}$

Modelling strategy: Start with small value of $m$ and increase successively until some $H_{0(m)}$ is not rejected

## Scale Invariance of Factor Analysis

Suppose $Y_i = c_i X_i$ or in matrix notation $\boldsymbol{Y} = C\boldsymbol{X}$ ($C$ is a diagonal matrix), e.g., change of measurement units. Then,

$$\begin{aligned}
\mathrm{Cov}(Y) &= C\boldsymbol{\Sigma}C^T \\
&= C(\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}) \\
&= (C\boldsymbol{L})(C\boldsymbol{L})^T + C\boldsymbol{\Psi}C^T \\
&= \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T + \tilde{\boldsymbol{\Psi}}
\end{aligned}$$

That is, loadings and uniquenesses are the same if expressed in new units:

- Using covariance or correlation gives basically the same result
- The common practice is to use a correlation matrix or scale the input data

# Rotational Invariance of Factor Analysis



Assume $RR^T = I$ and transform $\boldsymbol{F}_* = R^T\boldsymbol{F}$, $\boldsymbol{L}_* = \boldsymbol{L}R$, then

$$\boldsymbol{X}_* = \boldsymbol{\mu} + \boldsymbol{L}_*\boldsymbol{F}_* + \boldsymbol{\varepsilon} = (\boldsymbol{L}R)(R^T\boldsymbol{F}) + \boldsymbol{\varepsilon} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon} = \boldsymbol{X};$$
$$\boldsymbol{\Sigma}_* = \boldsymbol{L}_*\boldsymbol{L}_*^T + \boldsymbol{\Psi} = (\boldsymbol{L}R)(\boldsymbol{L}R)^T + \boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} = \boldsymbol{\Sigma}.$$

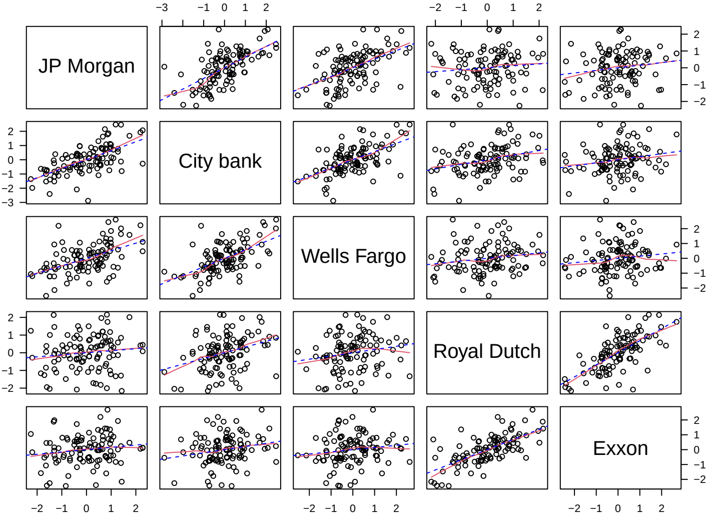
- Rotating the factors yields exactly the same model
- **Consequence**: Use rotation that makes interpretation of loadings easy
- Varimax rotation is the most popular rotation. Each factor should have a few large and many small loadings

# Example: Stock Price Data

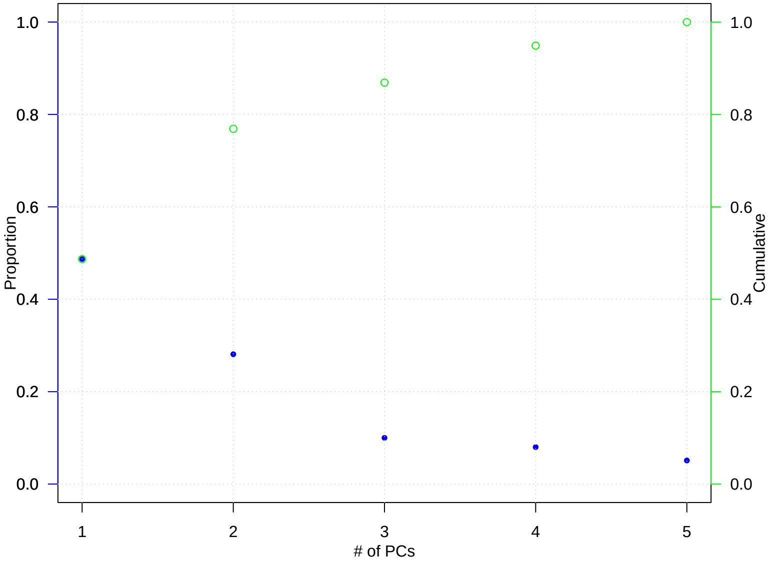
Data are weekly returns in stock prices for 103 consecutive weeks for five companies: JP Morgan, City bank, Wells Fargo, Royal Dutch (Shell), and Exxon

- The first three are banks and the last two are oil companies
- The data are first standardized and the sample correlation matrix is used for the analysis
- We will fit an $m = 2$ factor model

# Scatter Plot Matrix of the Standardized Data



# Screen Plot



# Factor Loadings, Specific Variances, and Residual Matrix

| Variable | Loadings 1 | Loadings 2 | Specific variances |
|---|---|---|---|
| JP Morgan | 0.732 | 0.437 | 0.273 |
| City bank | 0.831 | 0.280 | 0.230 |
| Wells Fargo | 0.726 | 0.374 | 0.333 |
| Royal Dutch | 0.605 | -0.694 | 0.153 |
| Exxon | 0.563 | -0.719 | 0.166 |

The residual matrix is $\boldsymbol{\Sigma} - (\tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^T + \tilde{\boldsymbol{\Psi}})$:

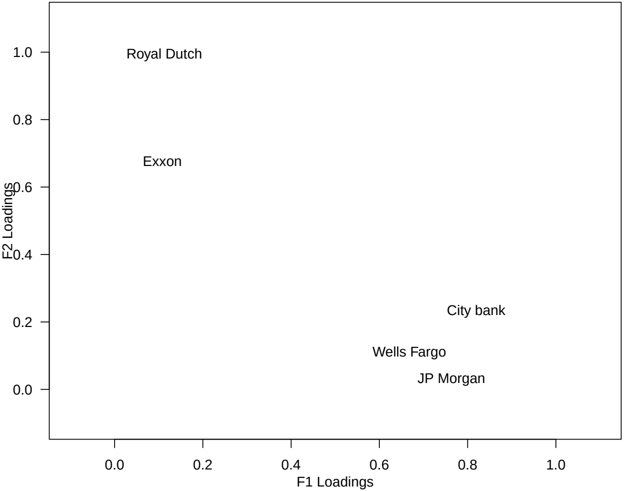
$$\begin{bmatrix} 0 & -0.10 & -0.18 & -0.03 & 0.06 \\ & 0 & -0.13 & 0.01 & -0.05 \\ & & 0 & 0.00 & 0.01 \\ & & & 0 & -0.16 \\ & & & & 0 \end{bmatrix}$$

**Question**: Are these off-diagonal elements small enough?

# Maximum Likelihood Estimation

```
> (stock.fac <- factanal(stock, factors = 2,
+ method = "mle", scale = T, center = T))

Call:
factanal(x = stock, factors = 2, method = "mle", scale = T, center = T)

Uniquenesses:
  JP Morgan   City bank Wells Fargo Royal Dutch
      0.417       0.275       0.542       0.005
      Exxon
      0.530

Loadings:
            Factor1 Factor2
JP Morgan   0.763
City bank   0.819   0.232
Wells Fargo 0.668   0.108
Royal Dutch 0.113   0.991
Exxon       0.108   0.677

               Factor1 Factor2
SS loadings      1.725   1.507
Proportion Var   0.345   0.301
Cumulative Var   0.345   0.646

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 1.97 on 1 degree of freedom.
The p-value is 0.16
```

# Factor Loading Plot




How to interpret these factors?

# PCA vs. FA Comparison Revisited

- PCA aims at explaining **variances**, while FA aims at explaining **correlations**
- PCA is exploratory and without assumptions FA is based on statistical model with assumptions
- First few PCs will be same regardless of $m$ First few factors of FA depend on $m$
- FA is scale and rotation invariant, while this property does not hold in PCA

# Summary

Concepts to know

- The form of the general Factor Model and its representation in terms of Covariance Matrix
- Scale and Rotation Invariance of Factor Model
- Interpretation of Factor Loadings

`R` functions to know

- `factanal`

In the next lecture, we will learn about Canonical Correlation Analysis