Math 576: Quantitative Risk Management

Haijun Li

lih@math.wsu.edu
Department of Mathematics
Washington State University

Week 6
Outline

1. Factor Models
2. Principal Component Analysis
Linear Factor Models

**Definition**

\( X = (X_1, \ldots, X_d) \) is said to follow a p-factor model if it can be decomposed as

\[
X = a + FB + \epsilon.
\]

**Basic Ingredients:**

- \( F = (F_1, \ldots, F_p) \) is a random vector of \( p \) factors, where \( p < d \) and its covariance matrix \( \text{Cov}(F) \) is positive definite.

- \( \epsilon = (\epsilon_1, \ldots, \epsilon_d) \) is a random vector of error terms, which are uncorrelated and have mean zero and \( \text{Cov}(\epsilon) = \rho I_d \), where \( I_d \) is the \( d \times d \) dimensional identity matrix.

- \( B \in \mathbb{R}^{p \times d} \) is the matrix of factor loading, and \( a \in \mathbb{R}^d \).

- Assume that \( F \) and \( \epsilon \) are uncorrelated; that is, \( \text{Cov}(F_i, \epsilon_j) = 0 \) for all \( 1 \leq i \leq p \) and all \( 1 \leq j \leq d \).
Remark

- If $F$ and $\epsilon$ are normally distributed, then $F$ and $\epsilon$ are independent.
- Since $\text{Cov}(\epsilon) = \rho I_d$, we have
  \[
  \text{Cov}(X) = B^\top \text{Cov}(F)B + \rho I_d.
  \]
- Using the Cholesky factorization,
  \[
  \text{Cov}(F) = A^\top A, \quad \text{where } A \in \mathbb{R}^{d \times d} \text{ is invertible.}
  \]

We have
\[
\text{Cov}(X) = B^\top A^\top AB + \rho I_d = C^\top C + \rho I_d,
\]
where $C = AB \in \mathbb{R}^{p \times d}$. Let $F^* = (F - E(F))A^{-1}$, a $p$-dimensional vector of factors. Then the factor model can be rewritten as
\[
X = E(X) + F^* C + \epsilon,
\]
with $E(F^*) = 0$ and $\text{Cov}(X) = C^\top C + \rho I_d$. 
Factor Decomposition Theorem

If a random vector $X$ has a covariance matrix which satisfies

$$\text{Cov}(X) = B^\top B + \rho I_d$$

for some $B \in \mathbb{R}^{p \times d}$ with $\text{rank}(B) = p < d$, then $X$ has a factor-model representation for some $p$-dimensional factor vector $F$ and $d$-dimensional error vector $\epsilon$:

$$X = a + FB + \epsilon.$$
Example: Equicorrelation Models

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with

$$E(X_i) = 0, \quad \text{Var}(X_i) = 1, \quad 1 \leq i \leq d$$

Assume that for $\rho > 0$,

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{pmatrix} = \mathbf{B}^\top \mathbf{B} + (1 - \rho) \mathbf{I}_d,$$

where

$$\mathbf{B} = (\sqrt{\rho}, \ldots, \sqrt{\rho}), \quad \mathbf{I}_d = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}.$$
Example (cont’d) : Equicorrelation = One-Factor Model

Let $Y$ be any random variable, independent of $X$, satisfying that

$E(Y) = 0, \ Var(Y) = 1.$

Let

$$F = \frac{\sqrt{\rho}}{1 + \rho(d - 1)} \sum_{i=1}^{d} X_i + \sqrt{\frac{1 - \rho}{1 + \rho(d - 1)}} Y$$

$$\epsilon = \left(X_1 - \sqrt{\rho}F, \ldots, X_d - \sqrt{\rho}F\right).$$

Verify that $X = FB + \epsilon$ where

$\text{Cov}(F, \epsilon_i) = 0, \ \text{Cov}(\epsilon_i, \epsilon_j) = 0, \ \forall \ i \neq j.$
Example (cont’d): Homogeneous Credit Portfolios

- If $X$ and $Y$ are Gaussian, then
  \[ \epsilon = \left( X_1 - \sqrt{\rho}F, \ldots, X_d - \sqrt{\rho}F \right) \]

  is a vector of independent Gaussian errors with variance $1 - \rho$.

- Let
  \[ Z_i := \frac{\epsilon_i}{\sqrt{1 - \rho}}, \quad 1 \leq i \leq d. \]

  Then $Z_1, \ldots, Z_d$ are iid standard Gaussian.

- Rewrite the one-factor model of risk scores for $d$ obligors:
  \[ X_i = \sqrt{\rho}F + \sqrt{1 - \rho}Z_i, \quad 1 \leq i \leq d. \]

  where $F$ is a common systematic risk factor affecting all obligors, and $Z_i$ is a specific risk factor affecting the $i$-th obligor.
Regression Analysis of Factor Models

Assume that the factors are observable at times \( t = 1, \ldots, n \) \((n = \text{sample size})\). That is, suppose that we have

- response data: \( X_t = (X_{t,1}, \ldots, X_{t,d}), \ t = 1, \ldots, n; \)
- factor data: \( F_t = (F_{t,1}, \ldots, F_{t,p}), \ t = 1, \ldots, n. \)

The goal is to estimate the parameters \( a = (a_1, \ldots, a_d), \)
\( B = (b_{ij})_{p \times d}, \) and \( \rho \) such that

\[
X_t = a + F_t B + \epsilon_t, \quad t = 1, \ldots, n.
\]

where

\[
\epsilon_t = (\epsilon_{t,1}, \ldots, \epsilon_{t,d}), \quad t = 1, \ldots, n.
\]

Rephrase it as a linear regression problem: for any \( 1 \leq j \leq d, \)

\[
X_{t,j} = a_j + \sum_{i=1}^{p} b_{ij} F_{t,i} + \epsilon_{t,j}, \quad t = 1, \ldots, n.
\]
Example: Single-index model for Stock Returns

- Consider 10 stock returns \((X_1, \ldots, X_{10})\) from 1992 to 1998: \((MO, KO, EK, HWP, INTC, MSFT, IBM, MCD, WMT, DIS)\).
  These companies are part of the Dow Jones 30 index at that time.
- The factor \(F\) is the corresponding return on the Dow Jones 30 index from 1992 to 1998.
- Fit the one-factor model to data:
  \[
  (X_1, \ldots, X_{10}) = F(B_1, \ldots, B_{10}) + (\epsilon_1, \ldots, \epsilon_{10}).
  \]
  \(\hat{B} = (0.87, 1.01, 0.77, 1.12, 1.12, 1.11, 1.07, 0.86, 1.02, 1.03).\) high beta
- \(r^2\) = the coefficient of determination, which measures the strength of the regression relationship between \(X_j\) and \(F\) and can be interpreted as the proportion of the variation of the stock return that is explained by variation in the market return \(F\):
  \[(0.17, 0.33, 0.14, 0.18, 0.17, 0.21, 0.22, 0.23, 0.24, 0.26)\]
Example (cont’d): Sample Correlation Matrices

<table>
<thead>
<tr>
<th></th>
<th>MO</th>
<th>KO</th>
<th>EK</th>
<th>HWP</th>
<th>INTC</th>
<th>MSFT</th>
<th>IBM</th>
<th>MCD</th>
<th>WMT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO</td>
<td>1.00</td>
<td>0.27</td>
<td>0.14</td>
<td>0.17</td>
<td>0.16</td>
<td>0.25</td>
<td>0.18</td>
<td>0.22</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>KO</td>
<td>0.27</td>
<td>1.00</td>
<td>0.17</td>
<td>0.22</td>
<td>0.21</td>
<td>0.25</td>
<td>0.18</td>
<td>0.36</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>EK</td>
<td>0.14</td>
<td>0.17</td>
<td>1.00</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>HWP</td>
<td>0.17</td>
<td>0.22</td>
<td>0.17</td>
<td>1.00</td>
<td>0.42</td>
<td>0.38</td>
<td>0.36</td>
<td>0.20</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>INTC</td>
<td>0.16</td>
<td>0.21</td>
<td>0.17</td>
<td>0.42</td>
<td>1.00</td>
<td>0.53</td>
<td>0.36</td>
<td>0.19</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.25</td>
<td>0.25</td>
<td>0.18</td>
<td>0.38</td>
<td>0.53</td>
<td>1.00</td>
<td>0.33</td>
<td>0.22</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>IBM</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.36</td>
<td>0.36</td>
<td>0.33</td>
<td>1.00</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>MCD</td>
<td>0.22</td>
<td>0.36</td>
<td>0.14</td>
<td>0.20</td>
<td>0.19</td>
<td>0.22</td>
<td>0.20</td>
<td>1.00</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>WMT</td>
<td>0.16</td>
<td>0.33</td>
<td>0.17</td>
<td>0.22</td>
<td>0.22</td>
<td>0.28</td>
<td>0.20</td>
<td>0.26</td>
<td>1.00</td>
<td>0.28</td>
</tr>
<tr>
<td>DIS</td>
<td>0.22</td>
<td>0.32</td>
<td>0.16</td>
<td>0.23</td>
<td>0.21</td>
<td>0.26</td>
<td>0.20</td>
<td>0.26</td>
<td>0.28</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MO</th>
<th>KO</th>
<th>EK</th>
<th>HWP</th>
<th>INTC</th>
<th>MSFT</th>
<th>IBM</th>
<th>MCD</th>
<th>WMT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO</td>
<td>1.00</td>
<td>0.24</td>
<td>0.16</td>
<td>0.18</td>
<td>0.17</td>
<td>0.19</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>KO</td>
<td>0.24</td>
<td>1.00</td>
<td>0.22</td>
<td>0.24</td>
<td>0.23</td>
<td>0.26</td>
<td>0.27</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>EK</td>
<td>0.16</td>
<td>0.22</td>
<td>1.00</td>
<td>0.16</td>
<td>0.15</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>HWP</td>
<td>0.18</td>
<td>0.24</td>
<td>0.16</td>
<td>1.00</td>
<td>0.17</td>
<td>0.19</td>
<td>0.20</td>
<td>0.20</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>INTC</td>
<td>0.17</td>
<td>0.23</td>
<td>0.15</td>
<td>0.17</td>
<td>1.00</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.19</td>
<td>0.26</td>
<td>0.17</td>
<td>0.19</td>
<td>0.19</td>
<td>1.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>IBM</td>
<td>0.20</td>
<td>0.27</td>
<td>0.18</td>
<td>0.20</td>
<td>0.19</td>
<td>0.22</td>
<td>1.00</td>
<td>0.23</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>MCD</td>
<td>0.20</td>
<td>0.28</td>
<td>0.18</td>
<td>0.20</td>
<td>0.19</td>
<td>0.22</td>
<td>0.23</td>
<td>1.00</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>WMT</td>
<td>0.20</td>
<td>0.28</td>
<td>0.18</td>
<td>0.21</td>
<td>0.20</td>
<td>0.22</td>
<td>0.23</td>
<td>0.23</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>DIS</td>
<td>0.21</td>
<td>0.29</td>
<td>0.19</td>
<td>0.22</td>
<td>0.21</td>
<td>0.23</td>
<td>0.24</td>
<td>0.24</td>
<td>0.25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Figure:** The sample correlation matrix (top) VS the sample correlation matrix implied by the factor model. The bottom matrix picks up much, but not all, of the structure of the former matrix.
**Example (cont’d): Residuals**

![Correlation Matrix](image)

**Figure:** The estimated correlation matrix of the residuals from the regression model, but only those elements which exceed 0.1 in absolute value. The residuals are indeed much less correlated than the original data.
Principal Component Analysis (PCA)

Goal: To reduce the dimensionality of highly correlated data by finding a small number of uncorrelated linear combinations that account for most of the variability of the original data.

Spectral Decomposition Theorem

Any symmetric matrix $A \in \mathbb{R}^{d \times d}$ can be written as

$$A = O^\top DO$$

where $O \in \mathbb{R}^{d \times d}$ is an orthogonal matrix satisfying $O^\top O = OO^\top = I_d$, and

$$D = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_d \\
\end{pmatrix}.$$

Here $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are all the eigenvalues.
Remark

- $O \in \mathbb{R}^{d \times d}$ is orthogonal iff
  \[ O^{-1} = O^\top. \]

- Let $\mathbf{v}_i$ denote the $i$-th row vector of matrix $O$, $1 \leq i \leq d$, then
  \[ \mathbf{v}_i A = \lambda_i \mathbf{v}_i, \quad 1 \leq i \leq d. \]

That is, $\mathbf{v}_i$ is an eigenvector of $A$ corresponding to $\lambda_i$. 
Principal Components

- Suppose the random vector $X$ has mean vector $\mu$ and covariance matrix $\Sigma$.
- Use spectral decomposition

\[ \Sigma = O^\top D O. \]

- Let $v_i$ denote an eigenvector of $\Sigma$ corresponding to $\lambda_i$, called the $i$-th vector of loadings.
- The $j$-th principal component of $X$ is defined as

\[ (X - \mu)v_i^\top \quad (= \text{linear combination of } X!), \]

which is the $j$-th component of the re-centered and rotated vector $Y = (X - \mu)O^\top$.
- Verify that $E(Y) = 0$ and

\[ \text{Cov}(Y) = O \text{ Cov}(X) O^\top = O\Sigma O^\top = D. \]
Figure: The estimated correlation matrix of the residuals from the regression model, but only those elements which exceed 0.1 in absolute value. The residuals are indeed much less correlated than the original data.
Explain the Variability in $\mathbf{X}$: An Approximation

$$\text{Cov}(\mathbf{Y}) = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_d
\end{pmatrix}, \quad \text{where } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d.$$

- $\text{Var}(Y_i) = \lambda_i, \ i = 1, 2, \ldots, d$. That is, the larger eigenvalues account for bigger portions of variability.
- The total variability
  $$\sum_{i=1}^{d} \text{Var}(X_i) = \text{trace}(\Sigma) = \text{trace}(D) = \sum_{i=1}^{d} \lambda_i = \sum_{i=1}^{d} \text{Var}(Y_i).$$
- The ratio
  $$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$
  represents the amount of this variability explained by the first $k$ principal components.
2-D Plot

Figure: Explain the Variability in $(X, Y)$: An Approximation Method
Sample Principal Components

Multivariate data observations: \( \mathbf{X}_t, \ t = 1, \ldots, n. \)

Goal: Find the \( k \) principal components that account for \((1 - \alpha)100\%\) variability of data.

Algorithm:

1. Estimate the covariance matrix

\[
\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} (\mathbf{X}_t - \overline{\mathbf{X}}) \mathbf{X}_t, \quad \text{where} \quad \overline{\mathbf{X}} = \sum_{t=1}^{n} \mathbf{X}_t.
\]

2. Find and rank the eigenvalues of \( \hat{\Sigma} \), with the corresponding eigenvectors:

\[
\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d, \quad (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d).
\]

3. Choose the first \( k \) empirical principal components

\[
(X_t - \overline{X})v_i^T, \quad i = 1, \ldots, k, \text{ such that } \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq 1 - \alpha.
\]
Remark

- The empirical spectral decomposition

\[ \hat{\Sigma} = O^\top DO, \]

where the \( i \)-row of \( O \) is \( v_i \).

- The empirical principal components transform

\[ Y_t = (X_t - \bar{X})O^\top \]

re-centers and rotates the original data vectors \( X_t, t = 1, \ldots, n \).

- The rotated vectors \( Y_t, t = 1, \ldots, n \), show no correlation between components and the components are ordered by their sample variances, from largest to smallest.

- Principal components as factors:

\[ X = \mu + YO + \epsilon \]

where \( Y \) is the first \( k \) principal components.
Example: Two-Factor Model for 10 Stock Returns

Consider 10 stock returns \( (X_1, \ldots, X_{10}) \) from 1992 to 1998:

\( (MO, KO, EK, HWP, INTC, MSFT, IBM, MCD, WMT, DIS) \).

These companies are part of the Dow Jones 30 index at that time.

The two largest eigenvalues of the sample covariance matrix explain almost 50% of the variation.

Use PCA to construct a two-factor model.

<table>
<thead>
<tr>
<th></th>
<th>MO</th>
<th>KO</th>
<th>EK</th>
<th>HWP</th>
<th>INTC</th>
<th>MSFT</th>
<th>IBM</th>
<th>MCD</th>
<th>WMT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>0.27</td>
<td>0.14</td>
<td>0.17</td>
<td>0.16</td>
<td>0.25</td>
<td>0.18</td>
<td>0.22</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>MO</td>
<td>0.27</td>
<td>1.00</td>
<td>0.17</td>
<td>0.22</td>
<td>0.21</td>
<td>0.25</td>
<td>0.18</td>
<td>0.36</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>KO</td>
<td>0.14</td>
<td>0.17</td>
<td>1.00</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>EK</td>
<td>0.17</td>
<td>0.22</td>
<td>0.17</td>
<td>1.00</td>
<td>0.42</td>
<td>0.38</td>
<td>0.36</td>
<td>0.20</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>HWP</td>
<td>0.16</td>
<td>0.21</td>
<td>0.17</td>
<td>0.42</td>
<td>1.00</td>
<td>0.53</td>
<td>0.36</td>
<td>0.19</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>INTC</td>
<td>0.25</td>
<td>0.25</td>
<td>0.18</td>
<td>0.38</td>
<td>0.53</td>
<td>1.00</td>
<td>0.33</td>
<td>0.22</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.36</td>
<td>0.36</td>
<td>0.33</td>
<td>1.00</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>IBM</td>
<td>0.22</td>
<td>0.36</td>
<td>0.14</td>
<td>0.20</td>
<td>0.19</td>
<td>0.22</td>
<td>0.20</td>
<td>1.00</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>MCD</td>
<td>0.16</td>
<td>0.33</td>
<td>0.17</td>
<td>0.22</td>
<td>0.22</td>
<td>0.28</td>
<td>0.20</td>
<td>0.26</td>
<td>1.00</td>
<td>0.28</td>
</tr>
<tr>
<td>WMT</td>
<td>0.22</td>
<td>0.32</td>
<td>0.16</td>
<td>0.23</td>
<td>0.21</td>
<td>0.26</td>
<td>0.20</td>
<td>0.26</td>
<td>0.28</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Example (cont’d): Largest Eigenvalues

**Figure:** Barplot of the sample variances of the first eight principal components; above each bar the cumulative proportion of the total variance explained by the components. The two largest eigenvalues of the sample covariance matrix explain almost 50% of the variation.
Example (cont’d): Residuals

**Figure:** Barplot summarizing the loadings vectors and defining the first two principal components: (a) factor 1 loadings; and (b) factor 2 loadings.
Example (cont’d): Two-Factor Model

\[
\hat{\beta}' = \begin{bmatrix} 0.20 \\ 0.39 \\ 0.35 \end{bmatrix}, \quad r^2 = \begin{bmatrix} 0.39 \\ 0.23 \\ 0.55 \end{bmatrix}
\]

\[
\begin{array}{cccccccccc}
& MO & KO & EK & HWP & INTC & MSFT & IBM & MCD & WMT & DIS \\
\hline
\hat{\beta}' & 0.20 & 0.19 & 0.16 & 0.45 & 0.51 & 0.44 & 0.32 & 0.18 & 0.24 & 0.22 \\
r^2 & 0.39 & 0.34 & 0.23 & -0.26 & -0.45 & -0.10 & -0.07 & 0.31 & 0.39 & 0.37 \\
& 0.35 & 0.42 & 0.18 & 0.55 & 0.75 & 0.56 & 0.35 & 0.34 & 0.42 & 0.41 \\
\end{array}
\]

**Figure:** Two-factor model: \( \mathbf{X}_{1 \times 10} = \mathbf{\mu}_{1 \times 10} + (F_1, F_2)\mathbf{B}_{2 \times 10} + \mathbf{\epsilon}_{1 \times 10} \).