Chapter 18: Factor Models and Principal Components

Professor: Hammou El Barmi
Columbia University

Introduction

- Many financial markets are characterized by a high degree of collinearity between returns.
- Variables are highly collinear when there are only a few important sources of information in the data that are common to many variables.
- High-dimensional data can be challenging to analyze.
- They are difficult to visualize, need extensive computer resources, and often require special statistical methodology.
- Fortunately, in many practical applications, high-dimensional data have most of their variation in a lower-dimensional space that can be found using dimension reduction techniques.
- There are many methods designed for dimension reduction, and in this chapter we will study two closely related techniques, factor analysis and principal components analysis, often called PCA.

Introduction

- PCA finds structure in the covariance or correlation matrix and uses this structure to locate low-dimensional subspaces containing most of the variation in the data.
  Idea: Extract the most important uncorrelated sources of variation in a multivariate system — principal component analysis (PCA)
- Factor analysis explains returns with a smaller number of fundamental variables called factors or risk factors.
- Factor analysis models can be classified by the types of variables used as factors, macroeconomic or fundamental, and by the estimation technique, time series regression, cross-sectional regression, or statistical factor analysis.

Example:

- A fund manager has 1000 stocks in their portfolio. If they analyze all the stocks quantitatively, they need a correlation or covariance matrix.
- This problem can get very cumbersome (the dimension of the matrix involved here is 1000×1000)
- If there are say, 10 factors which explain the movement of all these 1000 stocks, then by analyzing those 10 factors, the fund manger can get a handle on the dynamics of the entire 1000 stocks
- The 1000 stock portfolio has then been reduced to a 10 factor portfolio, where each of these 10 factors are independent of other factors and in some way explain the movement of all the 1000 stocks. This is what PCA does.
- These 10 factors are known as the "Principal Components" (PCs) of the asset correlation or covariance matrix.

Principal Components Analysis (PCA)

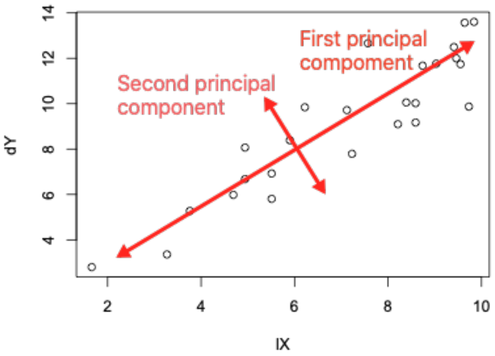
- The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, ie. by reducing the number of dimensions, without much loss of information
- PCA is concerned with explaining the variance covariance of a set of variables
- This explanation comes from a "few" linear combinations of the original variables
- Generally speaking, PCA has two objectives
  1. Data "reduction": moving from many original variables to a few linear combinations of the original variables
  2. Interpretation: which variables play a larger role in the explanation of the total variance

Principal Components Analysis (PCA)

- Principal components are ordered according to their variances.
- The first principal component is the linear combination that encapsulates most of the variability. In other words, the first principal component represents a rotation of the data along the axis representing the largest spread in the multidimensional cluster of data points.
- The second principal component is the linear combination that explains the most of the remaining variability while being uncorrelated (i.e. perpendicular) to the first principal component.
- If there was a third principal component, it would explains most of the remaining variability while being un- correlated to the first two principal components. This pattern continues for all consecutive principal components.

Directions of Most Variance

Two variables lX and dY are plotted against each other:



- The direction that has the largest variance is the first principal component
- The direction orthogonal to the first PC is the second component
- (In a 2-D problem, there is only one direction that can be orthogonal)

Principal Components Analysis (PCA)

- Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)^T$ be a random vector with mean $\boldsymbol{\mu}$ and covariance matrix Σ
- One way to measure the total variability in $\mathbf{Y}$ uses $|\Sigma|$, the determinant of Σ or the trace of V defined as

$$trace(\Sigma) = \sum_{i=1}^{d} \sigma_{ii}$$

  (We like to use one number as opposed to too many numbers)
- Notice that $trace(\Sigma) = 0$ if and only if $\sigma_{ii} = 0, \forall i$.

Principal Components Analysis (PCA)

- First principal component is $\mathbf{O}_1^T\mathbf{Y}$ where

$$\mathbf{O}_1 = argmax\{Var(\mathbf{O}^T\mathbf{Y}) = \mathbf{O}^T\Sigma\mathbf{O}\}$$

  subject $\mathbf{O}^T\mathbf{O} = 1$.
- The Lagrangian corresponding to this situation is

$$L(\mathbf{O}, \lambda) = \mathbf{O}^T\Sigma\mathbf{O} + \lambda(1 - \mathbf{O}^T\mathbf{O})$$

- To find **a** we need to solve

$$\begin{aligned} \frac{\partial}{\partial \mathbf{O}} L(\mathbf{O}, \lambda)) &= \Sigma\mathbf{O} - \lambda\mathbf{O} = \mathbf{0} \\ \frac{\partial}{\partial \lambda} L(\mathbf{O}, \lambda)) &= \mathbf{O}^T\mathbf{O} - 1 = 0 \end{aligned}$$

- The solution to these equation is $\mathbf{O}_1$ and λ_1
- Since $\Sigma\mathbf{O}_1 = \lambda_1\mathbf{O}_1, \mathbf{O}_1$is an eigen vector of V and λ_1 is its corresponding eigen value.

Principal Components Analysis (PCA)

- Second principal component

$$\mathbf{O}_2 = argmax\{Var(\mathbf{O}^T\mathbf{Y}) = \mathbf{O}^T\mathbf{\Sigma}\mathbf{O}\}$$

subject $\mathbf{O}_2^T\mathbf{O}_2 = 1$ and $\mathbf{O}_1^T\mathbf{O}_2 = 0$.

- The Lagrangian corresponding to this situation is

$$L(\mathbf{O}, \lambda, \gamma) = \mathbf{O}^T\mathbf{\Sigma}\mathbf{O} + \lambda(1 - \mathbf{O}^T\mathbf{O}) + \gamma\mathbf{O}^T\mathbf{O}_1$$

- To find $\mathbf{O}_2$ we need to solve

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{O}} L(\mathbf{O}, \lambda)) &= \mathbf{\Sigma}\mathbf{O} - \lambda\mathbf{O} = \mathbf{0} \\
\frac{\partial}{\partial \lambda} L(\mathbf{O}, \lambda)) &= \mathbf{O}^T\mathbf{O} - 1 = 0 \\
\frac{\partial}{\partial \gamma} L(\mathbf{O}, \lambda)) &= \mathbf{O}^T\mathbf{O}_1 = 0
\end{aligned}$$

- The solution to these equation is $\mathbf{O}_2$ and λ_2 and γ_2
- Since $\mathbf{\Sigma}\mathbf{O}_2 = \lambda_2\mathbf{O}_2$, $\mathbf{O}_2$is an eigen vector of $\mathbf{\Sigma}$ and λ_2 is its corresponding eigen value.

- We continue until we get d eigen-values and vectors $(\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_d)^T$ and $(\lambda_1, \lambda_2, \ldots, \lambda_d)$.
- Since Σ is symmetric, its eigenvalues (solutions of the polynomial equation $\det(\Sigma - \lambda I) = 0$) are real and can be ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$
- They are all nonnegative since Σ is nonnegative definite.
- Moreover

$$\sum_{i=1}^{d} Var(Y_i) \equiv \sum_{i=1}^{d} \sigma_{ii} = \sum_{i=1}^{d} \lambda_i \quad \text{and} \quad det(\Sigma) = \prod_{i=1}^{d} \lambda_i$$

- $\mathbf{O}_j^T Y$ is called the jth principal component

Principal Components Analysis (PCA)

It turns out that

- $\boldsymbol{\Sigma} = \lambda_1 \mathbf{O}_1 \mathbf{O}_1^T + \lambda_2 \mathbf{O}_2 \mathbf{O}_2^T + \ldots + \lambda_d \mathbf{O}_d \mathbf{O}_d^T$
- $Var(\mathbf{O}_j^T \mathbf{Y}) = \lambda_j$, $j = 1, 2, \ldots, d$.
- We hope that only a few principal components account for most of the overall variance. i.e.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

  is near 1 for small k
- Factor loadings are columns giving the elements of the column vectors $\mathbf{O}_i$s for the principal components $\mathbf{O}_i^T \mathbf{Y}$s
- The factor loading for the first principal component are $(O_{11}, O_{12}, \ldots, O_{1d})^T$

Principal Components Analysis (PCA)

- In practice Σ is not known and is estimated by

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1d} \\ s_{21} & s_{21} & \dots & s_{22} \\ \vdots & \vdots & \ddots & \vdots \\ s_{d1} & s_{d2} & \dots & s_{dd} \end{pmatrix}.$$

- If the data are not commensurate, we use the correlation matrix

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1d} \\ r_{21} & 1 & \dots & r_{22} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \dots & 1 \end{pmatrix}.$$

where

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}, \quad \forall (i,j)$$

- Let

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1d} \\ Y_{21} & Y_{22} & \dots & Y_{2d} \\ \vdots & \vdots & \dots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nd} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_d \end{pmatrix}$$

- The new data is

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & Z_{12} & \dots & Z_{1d} \\ Z_{21} & Z_{22} & \dots & Z_{2d} \\ \vdots & \vdots & \dots & \vdots \\ Z_{n1} & Z_{n2} & \dots & Z_{nd} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_d \end{pmatrix}$$

where $Z_{ij} = \mathbf{Y}_i^T \mathbf{O}_j$. That is

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Y}_1^T\mathbf{O}_1 & \mathbf{Y}_1^T\mathbf{O}_2 & \dots & \mathbf{Y}_1^T\mathbf{O}_d \\ \mathbf{Y}_2^T\mathbf{O}_1 & \mathbf{Y}_2^T\mathbf{O}_2 & \dots & \mathbf{Y}_2^T\mathbf{O}_d \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{Y}_n^T\mathbf{O}_1 & \mathbf{Y}_n^T\mathbf{O}_2 & \dots & \mathbf{Y}_n^T\mathbf{O}_d \end{pmatrix}$$

Principal Components Analysis (PCA)

- If the components of $\mathbf{Y}_i$ are comparable (e.g. are all daily returns on equities or are yield on bonds) then working with the original variables should be ok.
- If the components are not comparable (e.g. one in an unemployment rate and another is GDP in dollars), then some variables may many orders of magnitude larger than the others. In such a case, the large variables could completely dominate the PCA so that the first principal component is the direction if the variable with the largest standard deviation. To eliminate this, we should standardize the variables.

Principal Components Analysis (PCA)

- As a simple illustration of the difference between using standardized and unstandardized variables, suppose there are two variables ($d = 2$) with a correlation of 0.9. Then the correlation matrix is

$$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

- eigenvectors $(0.71, 0.71)^T$ and $(-0.71, 0.7)^T$ and eigenvalues 1.9 and 0.1. Most of the variation is in the direction (1, 1), which is consistent with the high correlation between the two variables.

Principal Components Analysis (PCA)

- However, suppose that the first variable has variance 1,000,000 and the second has variance 1. The covariance matrix is

$$\begin{pmatrix} 1,000,000 & 900 \\ 900 & 1 \end{pmatrix}$$

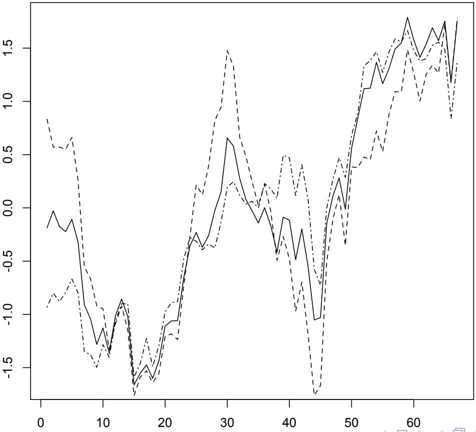
- Notice that the correlation coefficient is 0.9, just as was the case before.
- The eigenvectors, after rounding, equal to $(1.0000, 0.0009)^T$ and $(0.0009, 1)$ and eigenvalues 1000001 and 0.1899998
- The first variable dominates the principal components analysis based on the covariance matrix.
- This principal component analysis does correctly show that almost all of the variation is in the first variable, but this is true only with the original units.

Principal Components Analysis (PCA)

- Suppose that variable 1 had been in dollars and is now converted to millions of dollars.
- Then its variance is equal to 10^{-6}, so that the principal components analysis using the covariance matrix will now show most of the variation to be due to variable 2.
- In contrast, principal components analysis based on the correlation matrix does not change as the variables units change.

Here we look the the indices SP500, DOW JONES and NASDAQ (daily historical prices from 8/7/13-11/8/2013)

Figure: Indices SP500 Dow Jones and Nasdaq



Principal Components Analysis (PCA)

Standard deviations:

1.6506196 0.5151970 0.1001344

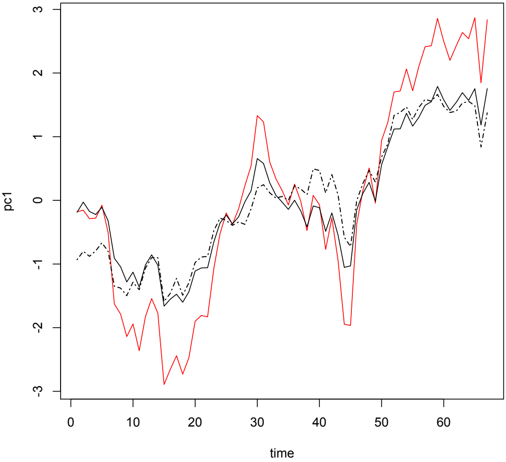
Rotation:

	PC1	PC2	PC3
sp500	0.6033967	-0.08039077	-0.7933787
dowjones	0.5579345	0.75339939	0.3479921
nasdaq	0.5697556	-0.65263058	0.4994515

The principal components are

$$\begin{aligned} PC1 &= 0.6033967\text{sp500} + 0.5579345\text{dowjones} + 0.5697556\text{nasdaq} \\ PC2 &= -0.08039077\text{sp500} + 0.75339939\text{dowjones} - 0.65263058\text{nasdaq} \\ PC3 &= -0.7933787\text{sp500} + 0.3479921\text{dowjones} + 0.4994515\text{nasdaq} \end{aligned}$$

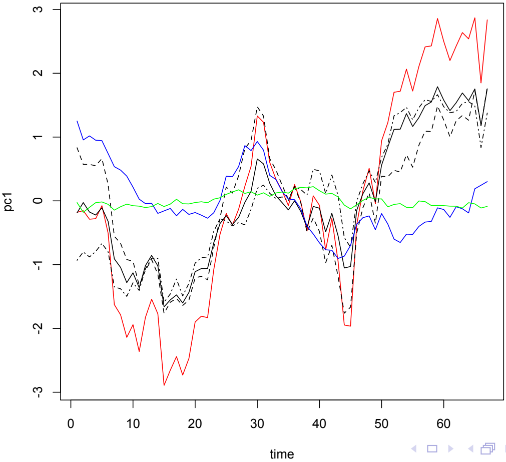
Figure: Indices SP500 Dow Jones and Nasdaq



Importance of components:

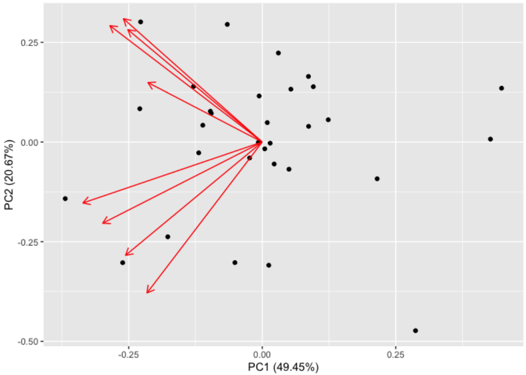
	PC1	PC2	PC3
Standard deviation	1.6506	0.51520	0.10013
Proportion of Variance	0.9082	0.08848	0.00334
Cumulative Proportion	0.9082	0.99666	1.00000

Figure: Indices SP500 Dow Jones and Nasdaq. PC1 (red), PC2 (blue) and PC3 (green)



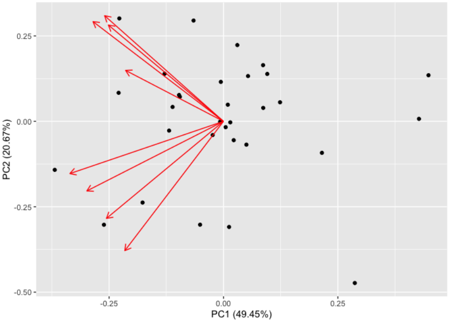
Stock Example: First 2 Components

- 8 stocks and 31 monthly returns, so the starting point is a 8-D space, and the coordinates of each point are the returns of the stocks that month
- Can you reduce the dimension of this space to just 2 or 3?
- Here is a projection of the points (black dots) on a 2-D plane:



- Each red arrow is a stock; the coordinates of the end points are the correlations to the corresponding principal components

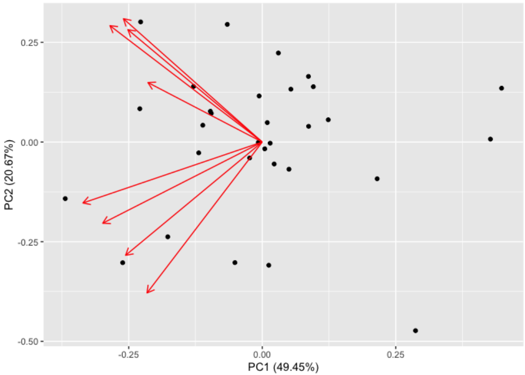
Stock Example: First 2 Components



- The first PC is the horizontal axis, 2nd PC is the y-axis
- Note how much of variance is explained by these 2 PCs. Half the stocks variance is driven by one factor, which is consistent with CAPM: this is the market (“beta”)
- The direction has no meaning
- There is no obvious interpretation of PC2. Maybe when we know the names of the stocks; but in general, this is a limitation of PCA
- Keep in mind that PCA knows nothing about the nature or name of these stocks. Yet two groups are clearly separated.

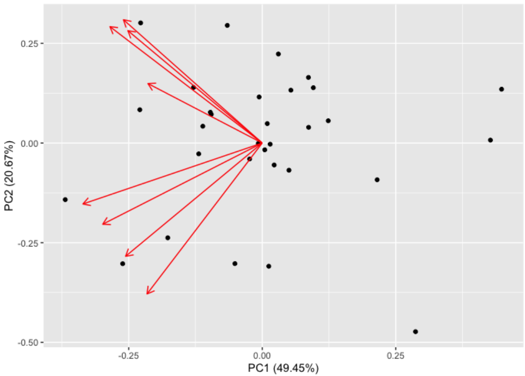
How to do it in R

```
library(ggfortify)
library(ggplot2)
stocks = read.csv(file = '/Users/user/Documents/stocks5.csv')
df=data.frame(stocks)
pca = prcomp(df, scale = TRUE)
autoplot(pca, data=stocks, loadings=TRUE)
```



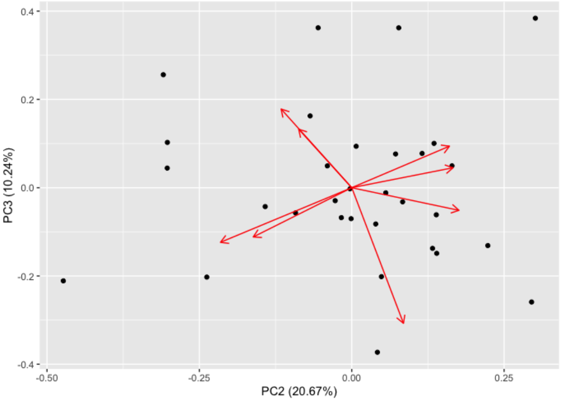
Stock Example: First 2 Components

- So, we saw that the stocks were clearly separated into two groups
- Can we separate these groups further?
- Also, can we explain more than 70% of variance?



Principal Components 2 and 3

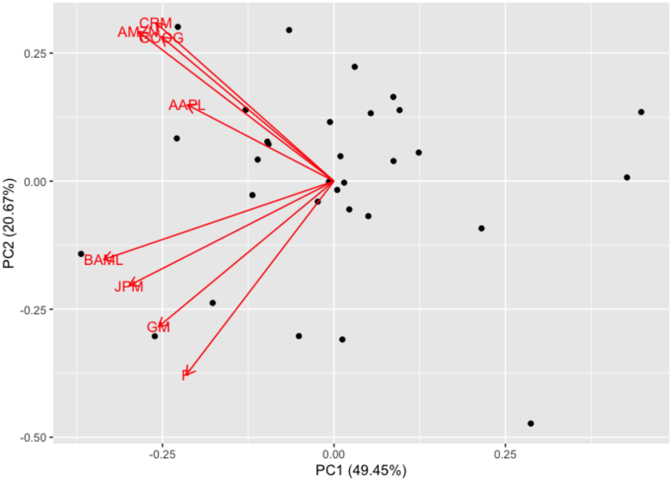
```
> autoplot(pca, data=stocks, loadings=TRUE, x=2, y=3)
```



- The 2nd and 3rd component separate the stocks into 4 groups
- We can't tell which of these stocks correspond to those on the previous slide but those with positive loadings on the PC2 are the same of course
- So PC3 separates the group with negative PC2 loading into 2 groups
- Remember that PCA knows nothing about these stocks

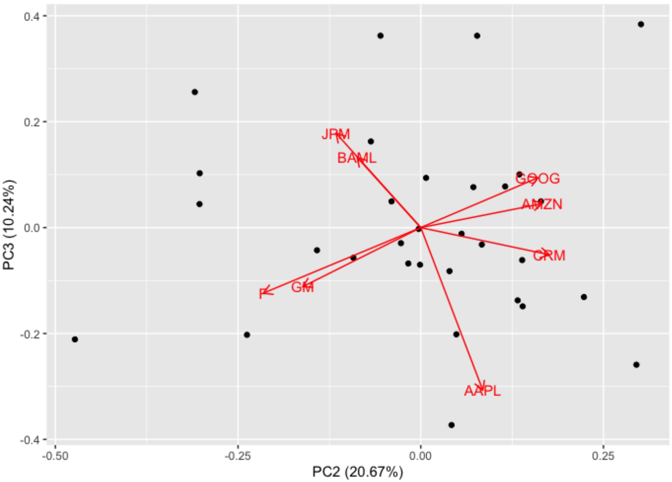
Example: Now With Stock Names

```
> autoplot(pca, data=stocks, loadings=TRUE, loadings.label=TRUE)
```



Example: Stocks

```
> autoplot(pca, data=stocks, loadings=TRUE,
               loadings.label=TRUE, x=2, y=3)
```



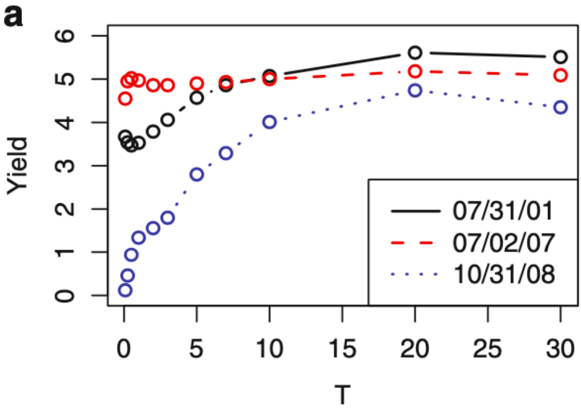
Example: Stocks

The data set used can be found on courseworks (stocks5.csv)

```
stocks<-read.csv("/Users/HElbarmi/Desktop/FINSTA2020/PCA/PCA/sto
header=TRUE, sep=",")
#Compute the covariance matrix Z
S <- cov(stocks)
# Compute the eigen values and vectors of S
s.eigen <- eigen(S)
s.eigen
#Make a scree graph
plot(s.eigen$values, xlab = 'Eigenvalue Number', ylab = 'Eigenva
main = 'Scree Graph')
lines(s.eigen$values)
# Find the principal componets
stocks.pca <- prcomp(stocks)
stocks.pca
#Compute the correlation matrix and the principal components bas
R <- cor(data)
stocks.pca.scaled <- prcomp(stocks, scale = TRUE)
stocks.pca.scaled
```

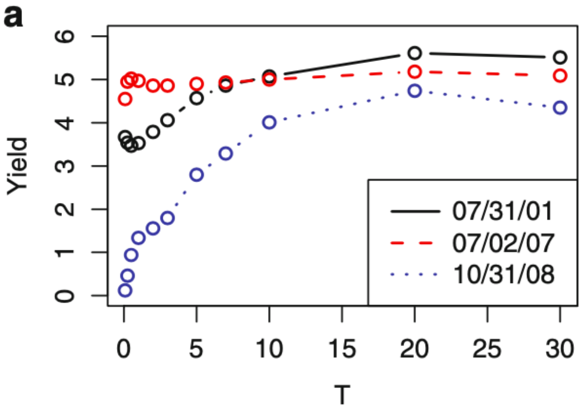
Application: Changes in Bond Yields (Example 18.2)

- This is certainly one of the most amazing achievements in financial statistics
- Consider the change in yield of bonds of different maturities issued by the U.S. Treasury – the book considers 11 maturities
- Imagine collecting these data every day; You can plot these 11 yields per day as one term structure per day (Figure 18.1 of textbook, reproduced below)



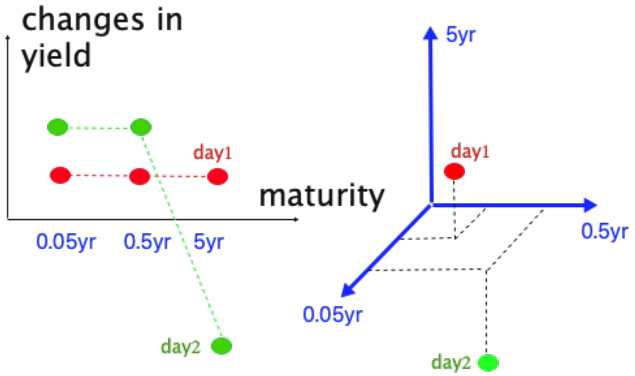
Application: Changes in Bond Yields (Example 18.2)

- You're interested in what drives daily **changes** of the yield curve; you can still plot a curve of 11 differences, at 11 maturities, for each day
- Alternatively, every day can be represented as one point in a 11-dimensional space
- We can't visualize these 11 dimensions. Do we need all of them, or can a few dimensions suffice to explain most of the daily changes in yields?



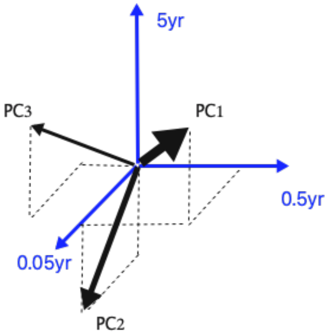
Application: Term Structure vs 11-D Space

- The "mapping" from one representation to the other is illustrated below
- Note that Categories "0.05yr," "0.5yr" and "5yr" do not necessarily have a natural order; in the case of term structures, they do: Categories are naturally ordered by the term (maturity)



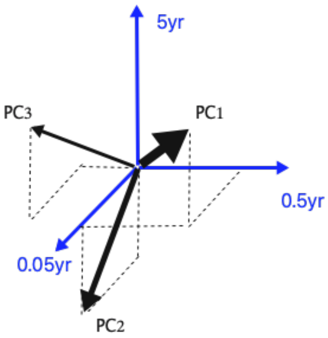
Application: Term Structure vs 11-D Space

- What the textbook's experiment verifies is that the direction of PC1 is very close to the unit vector (the unit vector is the vector from the origin to the red dot in this illustration)
- I.e., most of the variance of daily changes in yields, across maturities, comes from about-equal displacements in all dimensions, i.e. an almost constant shift across all rates



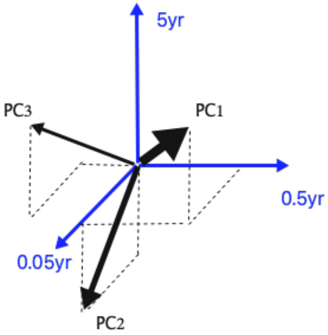
Application: Term Structure vs 11-D Space

- The 2nd PC is a decrease in short-term rates, flat mid-term rates, and an increase in longer-term rates
- This is called a **twist**. It corresponds to a steepening or flattening of the yield curve



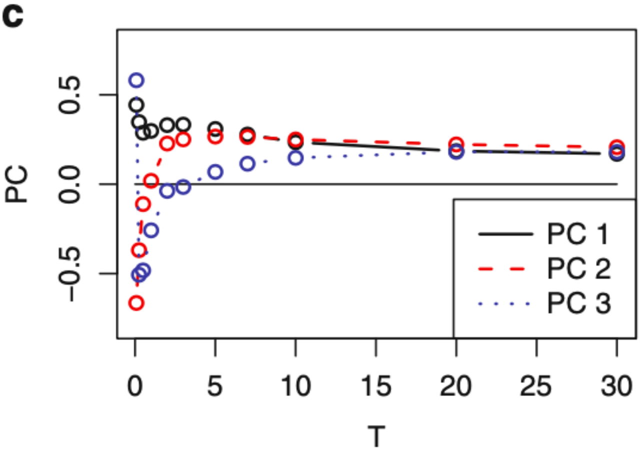
Application: Term Structure vs 11-D Space

- The 3rd PC is an increase in short and long-term rates but a decrease in intermediate rates
- This is a change in curvature of the yield curve and is often called the **butterfly**
- The textbook's experiment also confirms that 3 Principal Components are enough to capture 94.6% of the total variance



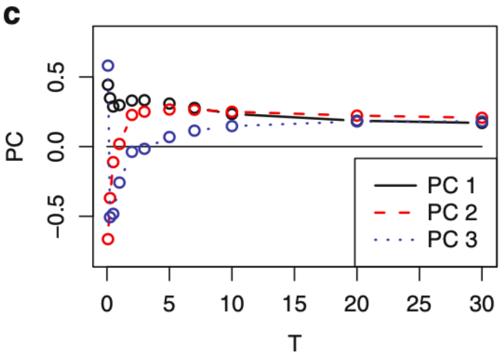
Application: Term Structure vs 11-D Space

- When you map back the first 3 PCs from the 11-D space to the term structure representation, you get the plot below (also from the textbook)
- PC1, in black, has all 11 components equaling about 0.2 to 0.5: this means most daily changes are almost-parallel shifts of the whole yield curve



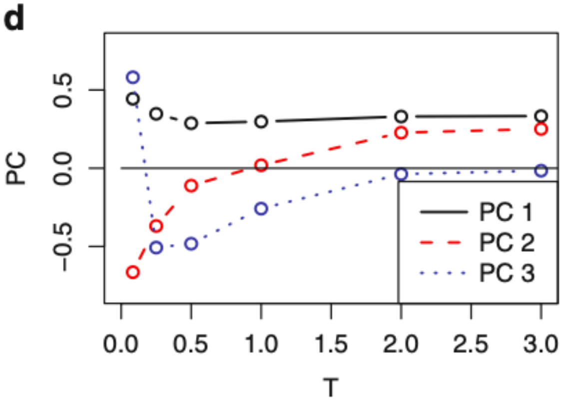
Application: Term Structure vs 11-D Space

- PC2, in red, has the first components negative, and the last ones are positive; this means a change in slope: the first / left points get lowered and the last / right points get lifted up
- Remember that we can talk of "first" and "left" and "last" only because there is a natural order due to maturities
- Also, PC1 and PC2 do not mean that most changes are upward shifts or steepenings of the slope: both PCs are just axes (without direction) that explain the variance



Application: Term Structure vs 11-D Space

- Interpreting PC3 requires us to zoom in, as the textbook does in Figure 18.1
- Its components are positive, then negative, then positive again
- This means that the daily changes of the yield curve tend to be bends of that curve (after parallel shifts and changes in slope)



Application to Risk Management of Fixed Income Portfolios

- For fixed income instruments that are not credit-sensitive, the main risk is changes to the yield curve
- And PCA just proved that most of these changes are parallel shifts, changes in slopes (**twist**) and bending (**butterfly**)

What about equities?

- PCA is not very helpful for equities. All PCA is telling us is that most of the returns comes from one component: the market
- And actually, the CAPM had reached that conclusion already: the expected return of a stock or a portfolio of stocks equals the risk-free return plus a multiplier times the market's return (or the market's excess return, depending on the equation we refer to)
- This opened the door to a different way toward "dimension reduction": one parameter (the beta to the market return) predicts a good part of each stock's return
- What other (hopefully few) factors could add to the explanatory power?

Factor Models)

- A factor model for excess equity returns is

$$Rj, t = \beta_{0,j} + \beta_{1,j} F_{1,t} + \ldots + \beta_{p,j} F_{p,t} + \epsilon_{j,t}$$

  where $R_{j,t}$ is either the return or the excess return on the *jth* asset at time t, $F_{1,t}, \ldots, F_{p,t}$ are variables, called factors or risk factors, that represent the Òstate of the financial markets and world economy at time t
- $\epsilon_{1,t}, \ldots, \epsilon_{n,t}$ are uncorrelated, mean-zero random variables called the unique risks of the individual stocks.
- The assumption that unique risks are uncorrelated means that all cross-correlation between the returns is due to the factors
- Notice that the factors do not depend on j since they are common to all returns
- The parameter β_{ij} is called a factor loading and specifies the sensitivity of the return of the *jth* asset to the *ith* factor
- Depending on the type of factor model, either the loadings, the factors, or both the factors and the loadings are unknown and must be estimated.

Factor Models)

- The CAPM is a factor model where $p = 1$ and $F_{1,t}$ is the excess return on the market portfolio.
- In the CAPM, the market risk factor is the only source of risk besides the unique risk of each asset.
- Because the market risk factor is the only risk that any two assets share, it is the sole source of correlation between asset returns.
- Factor models generalize the CAPM by allowing more factors than simply the market risk and the unique risk of each asset.

- A factor can be any variable thought to affect asset returns. Examples of factors include:
  1. returns on the market portfolio;
  2. growth rate of the GDP;
  3. interest rate on short term Treasury bills or changes in this rate;
  4. inflation rate or changes in this rate;
  5. interest rate spreads, for example, the difference between long-term Treasury bonds and long-term corporate bonds;
  6. return on some portfolio of stocks, for example, all U.S. stocks or all stocks with a high ratio of book equity to market equity Ñ this ratio is called BE/ME in Fama and French (1992, 1995, 1996);

This example uses the berndt Invest data set in Rs fEcofin package. This data set has monthly returns on 15 stocks over 10 years, 1978 to 1987. The 15 stocks were classified into three industries, Tech, Oil, and Other, as follows:

	tech	oil	other
CITCRP	0	0	1
CONED	0	0	1
CONTIL	0	1	0
DATGEN	1	0	0
DEC	1	0	0
DELTA	0	1	0
GENMIL	0	0	1
GERBER	0	0	1
IBM	1	0	0
MOBIL	0	1	0
PANAM	0	1	0
PSNH	0	0	1
TANDY	1	0	0
TEXACO	0	1	0
WEYER	0	0	1

Factor Models

- One possible model is a model that uses ÒtechÓ and ÒoilÓ as loadings and fit the model

$$R_j = \beta_0 + \beta_1 \text{tech}_j + \beta_2 \text{oil}_j + \epsilon_j$$

  where R_j is the return on the *jth* stock, tech_j equals 1 if the *jth* stock is a technology stock and equals 0 otherwise, and oil_j is defined similarly.
- the value of β_0, can be viewed as an overall market factor, since it affects all 15 returns. Factors 2 and 3 are the technology and oil factors. For example, if the value of β_2 is positive in any given month, then Tech stocks have better-than-market returns that month.

Fama-French 3-Factor Models

Fama and French (1995) developed a fundamental factor model with three risk factors, i.e. the market risk factor, the small minus large (SMB) factor, and the high minus low (HML) book-to-market factor.
The book-to-market value (BE/ME) is the ratio of the book value, which is the net worth of the firm according to its account balance sheet, to the market value, which is the market capitalization.

A high book-to-market value stock is called a value stock whose market value is relative cheap to its book value. A low book-to-market value stock is called a growth stock whose market value is relative expensive to its book value.

The second and third factors are the returns on two specific portfolios.

Fama-French 3-Factor Models

Fama and French 3-factor models:

$$R_{j,t} = \beta_{0,j} + \beta_{1,j} R_{M,t} + \beta_{2,j} SMB_t + \beta_{3,j} HML_t + \epsilon_{j,t}.$$

SMB is the average return on the three small portfolios minus the average return on the three big portfolios.

$$\begin{aligned} SMB &= 1/3(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) \\ &- 1/3(\text{Big Value} + \text{Big Neutral} + \text{Big Growth}) \end{aligned}$$

HML is the average return on the two value portfolios minus the average return on the two growth portfolios.

$$\begin{aligned} HML &= 1/2(\text{Small Value} + \text{Big Value}) \\ &- 1/2(\text{Small Growth} + \text{Big Growth}) \end{aligned}$$

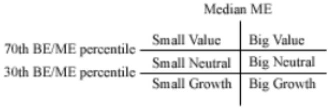
Fama-French 3-Factor Models

The above 6 portfolios are constructed by Fama and French at the end of each June. Every stock is classified into one of 6 categories based on the cutoff point for size/marketcap and two cutoff points for BE/ME ratio.

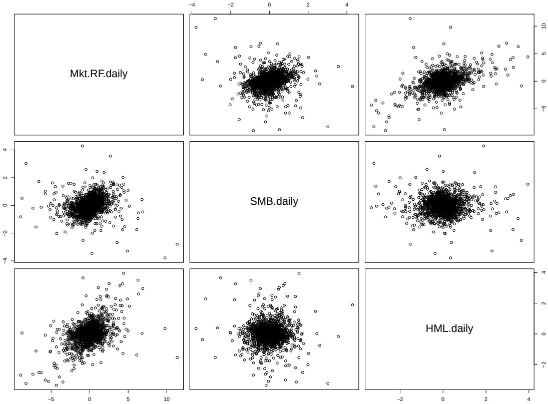
The cutoff point for size is the median of all market values on NYSE at the end of June of year t.

The two cutoff points for BE/ME are the 70th and 30th percentiles of BE/ME ratios on NYSE.

The weights for each portfolio are proportional to the market caps. So the returns for the following 6 portfolios are all value-weighted returns.



Then following pairs plot shows that factors SMB and HML are both correlated with the market risk factor. However, SMB seems to be orthogonal to HML.



Dimension Reduction for the Covariance by Factor Model

For a factor model for (excess) equity returns as follows,

$$R_{j,t} = \beta_{0,j} + \beta_{1,j} F_{1,t} + ... + \beta_{p,j} F_{p,t} + \epsilon_{j,t},$$

where $R_{j,t}$ is the (excess) return on the j^{th} asset at time t; $F_{1,t}, F_{2,t}, ..., F_{p,t}$ are factors, if $\epsilon_{1,t}, ..., \epsilon_{n,t}$ are indeed uncorrelated, mean-zero random variables, the the $n \times n$ covariance matrix of the asset returns is

$$\mathbf{\Sigma}_R = \boldsymbol{\beta}^T \mathbf{\Sigma}_F \boldsymbol{\beta} + \mathbf{\Sigma}_\epsilon,$$

where $\mathbf{\Sigma}_F$ is the $p \times p$ covariance matrix for the p factors, $\mathbf{\Sigma}_\epsilon$ is a $n \times n$ diagonal covariance matrix of ϵ.

Factor Models)

Factor Analysis is a method for investigating whether a number of variables of interest $R_1, R_2, \ldots, R_d$ are linearly related to a smaller number of unobservable factors $F_1, F_2, \ldots, F_p$.

The model is

$$R_{jt} = \beta_{0j} + \beta_{1j} F_{1t} + \ldots + \lambda_{pj} F_{pt} + \epsilon_{jt}, \quad j = 1, 2, \ldots, d, t = 1, 2, \ldots, T$$

or

$$\mathbf{R}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta} \mathbf{F}_t + \boldsymbol{\epsilon}_t, t = 1, 2, \ldots, T$$

where

- $\mathbf{R}_t = (R_{1t}, R_{2t}, \ldots, R_{dt})^T$
- $\mathbf{F}_t = (F_{1t}, F_{2t}, \ldots, F_{pt})^T$
- $\boldsymbol{\beta}_0$ is a $d \times 1$ vector

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0d} \end{pmatrix}.$$

and $\boldsymbol{\beta}$ is a $d \times p$ matrix (The betas are called the loading factors)

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{21} & \ldots & \beta_{p1} \\ \beta_{12} & \beta_{22} & \ldots & \beta_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1d} & \beta_{2d} & \ldots & \beta_{pd} \end{pmatrix}$$

In this model we assume that

- $$E(\mathbf{F}) = \mathbf{0}$$
- $$\Sigma_{\mathbf{F}} = I$$
- $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$
- $$\text{var}(\boldsymbol{\epsilon}_t) = \Sigma_{\epsilon} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_d^2 \end{pmatrix}$$

This implies that

$$\Sigma_{\mathbf{R}} = \boldsymbol{\beta}\boldsymbol{\beta}^T + \Sigma_{\epsilon}$$

or equivalently

$$\begin{aligned} \text{var}(R_{jt}) &= \sum_{i=1}^{p} \beta_{ji}^2 + \sigma_j^2, j = 1, 2, \ldots, d \\ &= h_{jj} + \sigma_j^2 \end{aligned}$$

where

- $h_{jj} = \sum_{i=1}^{p} \beta_{ji}^2$ is called communality
- σ_j^2 is called specific variance

Let

- S be the sample covariance matrix of $R_1, R_2, \ldots, R_2$.
- $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ be the largest eigenvalues of S and let $\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_p$ be the corresponding eigenvectors.
- $$D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_p \end{pmatrix}$$
- $\mathbf{O} = (\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_p)$

$\boldsymbol{\beta}$ is estimated by

$$\hat{\boldsymbol{\beta}} = \mathbf{O}D^{1/2} = (\sqrt{\lambda_1}\mathbf{O}_1, \sqrt{\lambda_2}\mathbf{O}_2, \ldots, \sqrt{\lambda_p}\mathbf{O}_p)$$

The specific variance for the jth return is estimated by

$$\hat{\sigma}_j^2 = s_{jj} - \sum_{i=1}^{p} \hat{\beta}_{ji}^2$$

Factor Models: Estimation of the loading factors

- The variance due to the ith factor is

$$\sum_{j=1}^{d} \beta_{ji}^2 = \lambda_i^2 \sum_{j=1}^{d} O_{ij}^2 = \lambda_j$$

which is the sum of the squares of the loadings in the jth column of $\hat{\beta}$

- The proportion of the total variance based on S due the jth factor is

$$\frac{\sum_{j=1}^{d} \beta_{ji}^2}{\mathrm{tr}(S)} = \frac{\lambda_j}{\mathrm{tr}(S)}$$

- The proportion of the total variance based on S due the p factors is

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{\mathrm{tr}(S)}$$

```
> fit

Call:
factanal(x = stocks, factors = 4)
Uniquenesses:
   GM     F   JPM  BAML  GOOG  AMZN  AAPL   CRM
0.037 0.426 0.005 0.090 0.005 0.266 0.730 0.117
Loadings:
     Factor1 Factor2 Factor3 Factor4
GM    0.126   0.216   0.932   0.177
F             0.360   0.642  -0.150
JPM   0.158   0.937   0.280   0.117
BAML  0.323   0.787   0.393   0.181
GOOG  0.473   0.186           0.858
AMZN  0.760   0.299           0.258
```

```
               Factor1 Factor2 Factor3 Factor4
SS loadings      2.002   1.814   1.569   0.938
Proportion Var   0.250   0.227   0.196   0.117
Cumulative Var   0.250   0.477   0.673   0.791

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 0.28 on 2 degrees of freedom.
The p-value is 0.867
```

Summary

The whole chapter discussed the extended Capital Asset Pricing Model. We learn

1. The Fama-French three factor models
2. The PCA method for statistical factor models