ETF5930 Financial Econometrics

Lecture 1

Monash Business School, Monash University, Australia

Unit Information

Education Team

- Chief Examiner and Lecturer: Professor Xibin Zhang
- Tutor: Ms Lu Wang

Teaching Activities

- Seminar: 2-hour face-to-face lecture which is recorded
  Tuesday 2pm–4pm
- Tutorial: 1-hour face-to-face tutorial
  Wednesday 11–12 and 12–13
- Workshop: 1-hour online learning to be supervised by the lecturer or tutor Thursday 2pm–3pm

Communication

- xibin.zhang@monash.edu
- lu.wang@monash.edu
- https://handbook.monash.edu/2025/units/ETF5930

Overview

Statistical and econometric tools to

- analyse and model key characteristics of empirical distributions of asset returns;
- model and estimate the simple capital asset pricing model and its extensions; and
- test for various financial market hypotheses.

This unit covers

- modelling, estimating and analysing properties of stationary and non-stationary financial time series;
- modelling and estimating simple and multivariate long-run relationships among financial variables; and
- modelling and estimating ARCH/GARCH volatilities, single- and multiple-factor capital asset pricing models.

Prohibition

ETF3300 (ETF5330) and ETC3460 (ETC5346)

Learning Outcomes

- Describe, interpret and critically analyse financial data.
- Apply the simple and multivariate models and theory to model the relationship among financial variables, interpret the results, and conduct reliable statistical inference.
- Test for stationary behaviour of financial time series.
- Model the long-run relationships among financial time series.
- Model and forecast the time-varying volatility of returns on financial assets.
- Be proficient at econometric modelling of financial data using R, which is widely used in statistics and econometrics.

Assessment

- ETF5930 does NOT contain a hurdle requirement that you must achieve to be able to pass the unit.
- Assignment 1 (15%) will be due by the 7th week.
- Assignment 2 (10%) will be due by the 11th week.
- Assignment 3 (15%) will be due by the 12th week.
- Final exam (60%) is an individual exam to be conducted via eExam under supervision.

Submission details

- Turnitin can help you discern when you are using sources fairly, citing properly, and paraphrasing effectively in accordance with University policy. These are skills essential to all academic work.
- More information about Turnitin can be found HERE

Assignments

- Each student will complete all assignments on their own. Detailed information on the contents of the assignments will be provided during the semester via the Moodle website.
- You are required to upload your completed assignment answers in PDF through Moodle submission.
- Your assignment can be either
  - entirely handwritten; or
  - entirely typed; or
  - a mixture of handwritten and typed answers.
- You may take photos of handwritten answers and paste them on a Word document, and save it as PDF.

Computing Tool: R

- We will be using R for computing purposes.
- Tutorials of the first two weeks will be focused R.
- It will be exciting to be able to learn R in 2 hours!

What is R?

- A programming language for statistical computing and graphics.
- Open-source and widely used in data science, statistics, and machine learning.
- Developed by Ross Ihaka and Robert Gentleman in 1993
- Supported by a large community with extensive packages and libraries.
- Install R via `https://cran.r-project.org/`

What is RStudio?

- RStudio is an integrated development environment (IDE) for R.
- Provides features such as:
  - Script editor
  - Console
  - Environment and History panel
  - Plots and Help panel
- Download RStudio at `https://posit.co/download/rstudio-desktop/`

Final Examination

- The Final Exam questions are all based on (a) lecture slides; and (ii) tutorial questions.
- Formula sheets will not be provided. Should you need to use formulae, they will be provided in the questions.
- You do NOT need to use R to answer exam questions.
- R commands are not examinable, but you are required to know how to interpret outputs obtained from R.

Materials

- Professional accreditation mandates that the Final Exam for ETF5930 must be a closed-book examination.
- You are not permitted to use any notes, texts, websites or other reference material in answering the questions.
- Students will not be allowed to access Moodle.
- Any physical calculators are permitted in the Final Exam.
- More information will be provided during the 12th week

Topic 1: Properties of Financial Data

- Asset Return Calculations:
  - Simple returns
  - Continuous compounded returns (log returns)
- R
- Descriptive Statistics: mean and variance
- Introduction to Portfolio Theory

Simple Returns

Consider purchasing an asset (for example, stock, bond, mutual fund, etc.).

- Let P_t denote the price of an asset that pays no dividends at time t:
- Let P_{t-1} denote the price at time $t-1$:
- Then the simple return or simple net return (denoted as R_t) on an investment in the asset between $t-1$ and t is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

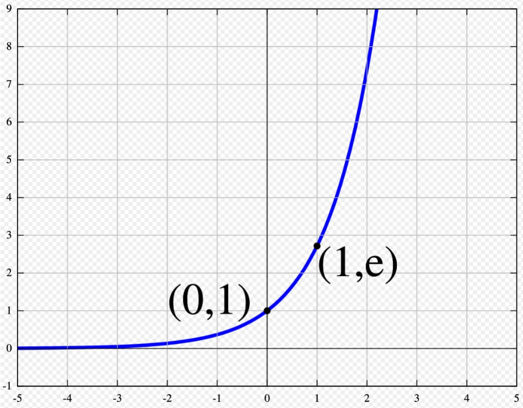
An Example

- Consider a one-month investment in Microsoft stock
- Suppose you buy the stock in the month $t-1$ at $P_{t-1} = \$85$ and sell the stock the next month for $P_t = \$90$
- Assume no dividend is paid during this month
- The one-month simple return is $(90-85)/85 \approx 5.88\%$

Continuously Compounded Returns

Exponential function $\exp(x)$ also written as e^x

The exponential function $y = \exp(x)$ is always positive and increasing in x.



Natural logarithm function $\ln(x)$ also written as $\log(x)$

- The computation of continuously compounded returns requires the use of natural logarithm functions.
- The natural logarithm function, $\ln(x)$, is the inverse function of $\exp(x)$.
- This is to say that $\ln(x)$ is defined such that $x = \ln(\exp(x))$.
- $\ln(x)$ is also written as $\log_e(x)$ which is often written as $\log(x)$ if there is no confusion.

Properties of $\ln(x)$

$$
\begin{aligned}
\ln(xy) &= \ln(x) + \ln(y) \\
\ln(x/y) &= \ln(x) - \ln(y)
\end{aligned}
$$

Continuously Compounded Returns

- Let R_t denote the simple monthly return on an investment.
- Continuously compounded monthly return is defined as

$$r_t = \ln(1 + R_t)$$

- Note that given R_t, we can calculate r_t because

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

  Therefore,

$$\begin{aligned} r_t &= \ln(1 + R_t) = \ln\left(1 + \frac{P_t}{P_{t-1}} - 1\right) = \ln\left(\frac{P_t}{P_{t-1}}\right) \\ &= \ln(P_t) - \ln(P_{t-1}) \end{aligned}$$

- r_t is the difference in log prices and usually is called the log return.

Continuously Compounded Returns

To summarise, the the log returns r_t is defined as

$$r_t = \ln(1 + R_t), \qquad \text{or} \qquad r_t = \ln(P_t) - \ln(P_{t-1}).$$

Example

- In the earlier example: P_{t-1}=85; P_t=90 and R_t=0.0588
- The monthly log return can be computed in two ways:

$$r_t = \ln(1 + R_t) = \ln(1 = 0.0588) = 0.0571$$
$$r_t = \ln(P_t) - \ln(P_{t-1}) = \ln(90) - \ln(85) = 0.0571$$

- The one-month investment in Microsoft stock yielded a 5.71% per month return.

Coding in R

Where to start?

```
> getwd()
[1] "\\\\ad.monash.edu/home/User069/xzhang/Documents"
> setwd("C:\\Users\\xzhang\\R")
> getwd()
[1] "C:/Users/xzhang/R"
```

It is my working directory, where I save coding script file and data

Choose an R Editor

- R Editor: Click on “File/New File/R script”
- RStudio
- Notepad++

Load Data

```
xdata<-read.csv(file="sp.vix.csv",header=T)
```

Excel file `sp.vix.csv` is in the same folder before reading it

Note that the symbol > is the command prompt

Fundamentals

- Any words behind # but in the same row are "comments"
- At this stage, do not worry about "[1]". It is the row number of the result.

Check some special functions

```
> abs(-0.64)
[1] 0.64
> log(10)
[1] 2.302528
> log(exp(0.1989))
[1] 0.1989
> log(exp(0.1989))
[1] 0.1989
> rep(1,10)
[1] 1 1 1 1 1 1 1 1 1 1
> matrix(0,nr=2,nc=3) # What is the output?
```

Working with the loaded data

- Check the data structure

```
> xdata[1:3,] #Print rows 1 to 3
```

	Date	SP500	VIX
1	10/23/2006	1377.02	100.00
2	10/24/2006	1377.38	99.41
3	10/25/2006	1382.22	98.96

- Summary statistics

```
mean(xdata[,2])      # arithmetic mean of SP500
sd(xdata[,2])        # standard deviation of SP500
median(xdata[,2])    # median of SP500
quantile(xdata[,2],0.25)   # 25 percent quantile
quantile(xdata[,2],0.5)    # 50 percent quantile
max(xdata[,2])   # maximum value of SP500
min(xdata[,2])   # minimum value of SP500
range(xdata[,2]) # the smallest and largest values
```

Time Series Data

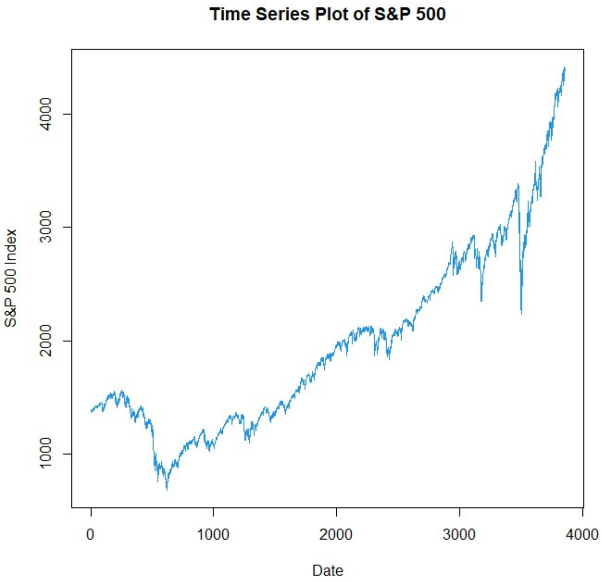
Time Series

- Time series data consists of observations collected across time.
- Data may collected:
  - annually (once a year)
  - quarterly (four times a year)
  - monthly (every month)
  - weekly
  - daily

Time Series Plot

```
xdata<-read.csv(file="sp.vix.csv",header=T)
plot(xdata[,2], typ='l', lty=1, col=4, xlab="Date",
ylab="S&P 500 Index", main="Time Series Plot of S&P
500")
```

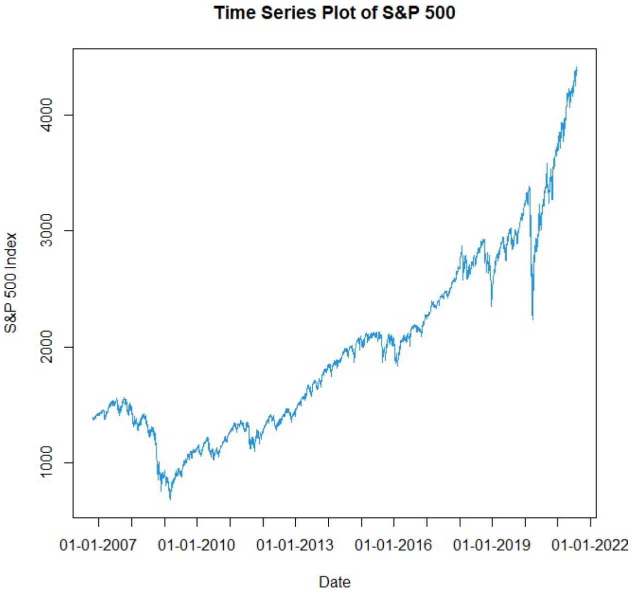
Time series plot of daily closing index of the Microsoft stock



How can we add the dates to the x-axis?

```
plot(xdata$Date, xdata[,2], typ='l', lty=1, col=4,
xlab="", ylab="SP 500 Index", main="Time Series Plot
of S&P 500")

xdata$Date <- as.Date(xdata[,1], format="%m/%d/%Y")
# Adjust date format if necessary

plot(xdata$Date, xdata[,2], type='l', lty=1, col=4,
xlab="Date", ylab="S&P 500 Index", main="Time Series
Plot of S&P 500", xaxt='n')

tem=pretty(xdata$Date, n=15)

axis(1, tem, format(tem,"%d-%m-%Y"))
```

Time series plot of daily closing index of the Microsoft stock



How to compute simple return series

```
sp <- xdata[,2] # 2nd column of xdata
n <- length(sp)
rt.sim <- (sp[2:n]-sp[1:(n-1)])/sp[1:(n-1)]*100
# element by element operations
```

How calculate continuously compounded returns

- If we have calculated simple returns, we can calculate continuously compounded returns as
  `rt.cc <- log(1+rt.sim)*100 #` $\ln(1+R_t)$
- We can directly calculate continuously compounded returns:
  `rt.cc <- (log(sp[2:n])-log(sp[1:(n-1)]))*100`

Summary Statistics

Population measures

Let Y_t denote an asset's return. As it can take any value, we treat Y_t as a random variable.

- The mean of Y_t denoted as $E(Y_t)$, is the expected value of Y_t, which is the average of Y_t.
- The variance of Y_t denoted as $\text{Var}(Y_t)$, is defined as

$$\text{Var}(Y_t) = E(Y_t - E(Y_t))^2,$$

  which measures the variation of Y_t.

Sample measures

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ denote a sample of observed asset returns.

- Sample mean is defined as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}(y_1 + y_2 + \cdots + y_n).$$

- Sample variance is defined as

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\left((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2\right).$$

- Sample standard deviation is

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

- Both mean and standard deviation have the same units as the original data

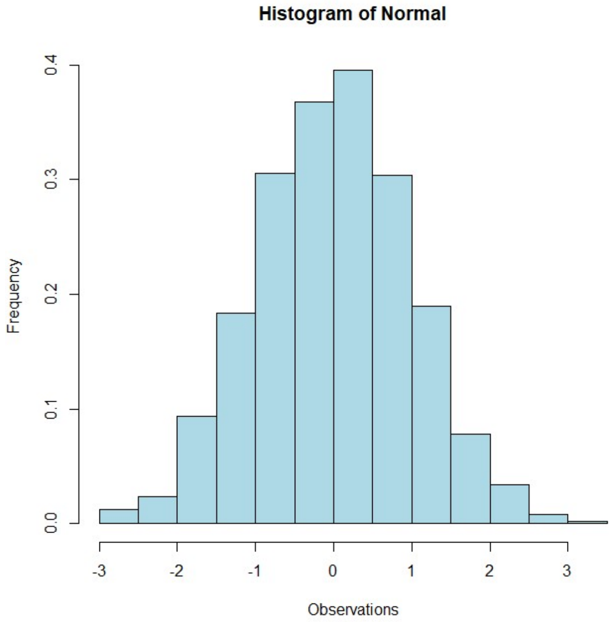
Histogram

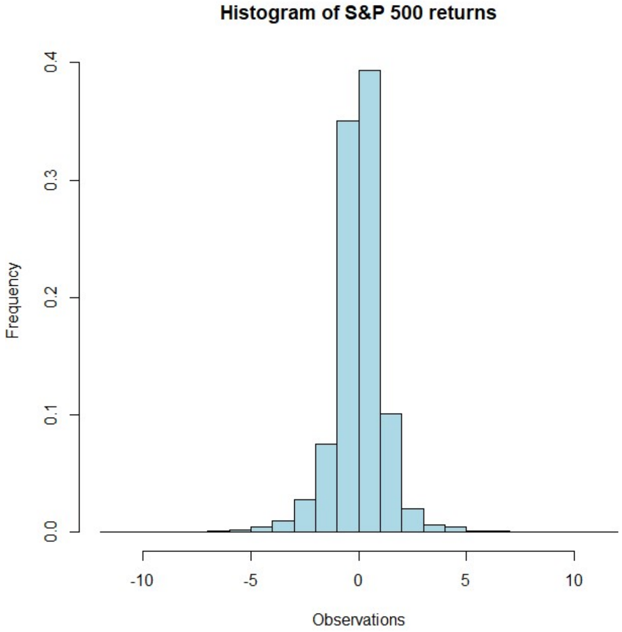
- A histogram describes the shape of the distribution of the data

```
x<-rnorm(1000,0,1) #generate random numbers~ N(0,1)
hist(x, breaks=16, col="lightblue", freq=FALSE)
```

- Can we plot the histogram of simple returns of S&P 500 Index?

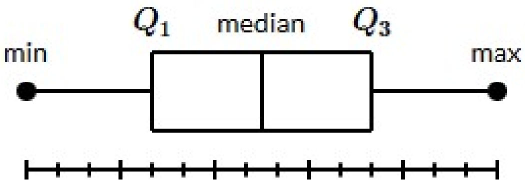
```
sp <- xdata[,2] # 2nd column of xdata
n <- length(sp)
rt.sim <- (sp[2:n]-sp[1:(n-1)])/sp[1:(n-1)]*100
hist(rt.sim, breaks=22, col="lightblue",
main="Histogram of S&P 500 returns", freq=FALSE)
```


Histogram of Normal
Frequency
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
Observations


Histogram of S&P 500 returns
Frequency
0.0
0.1
0.2
0.3
0.4
-10
-5
0
5
10
Observations

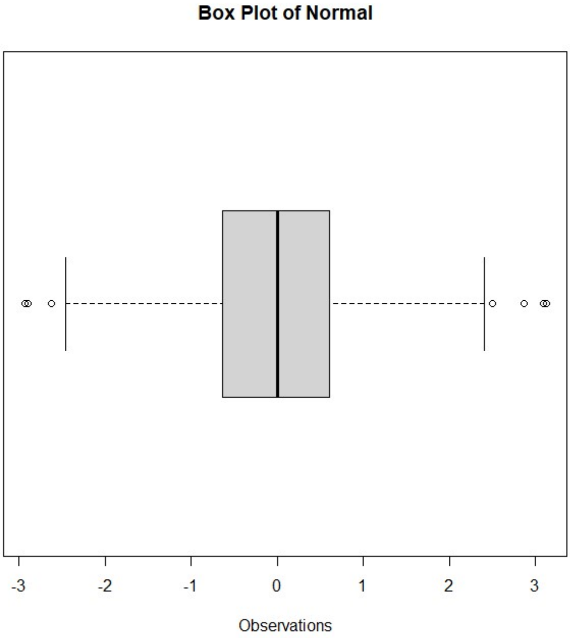
Box and Whisker plot

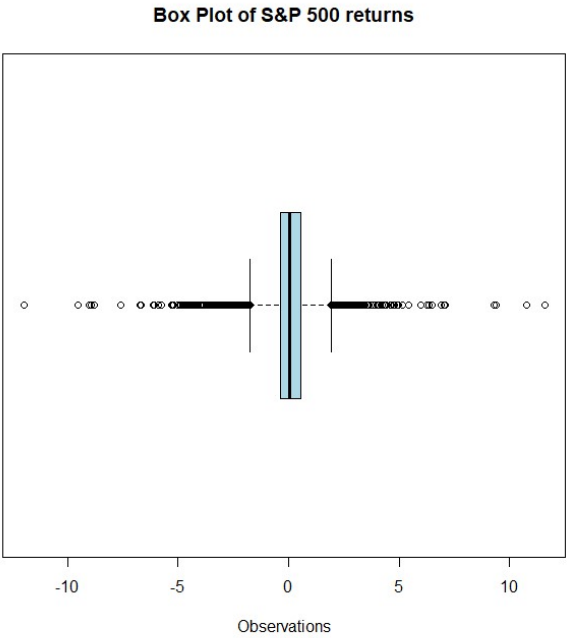
A box and whisker plot, also known as a box plot, displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.



R code

```
x<-rnorm(1000,0,1)
boxplot(x,xlab="Observations",main="Box Plot of
Normal",horizontal=TRUE)
```


Box Plot of Normal
-3
-2
-1
0
1
2
3
Observations


Box Plot of S&P 500 returns
-10
-5
0
5
10
Observations

Stationarity

A time series Y_t is defined to be stationary if

- $E(Y_t)$ does not depend on t (constant mean)
- $\text{Var}(Y_t)$ is finite and does not depend on t
- $\text{Cov}(Y_t, Y_{t-j})$ depends on j and not on t, for $j = 1, 2, \ldots,$

Are financial time series stationary?

- Asset price series (such as Microstft stock and S&P 500 index) do not seem to be stationary as they do not have a constant mean. These asset prices do not have the tendency to return to their means.
- Simple return series or log return series seem to be stationary as they hover around some constant average/mean value over the sample period.
- Unlike asset price series, asset return series has no obvious change in mean over the sample period.

Introduction to Portfolio Theory

Portfolio theory

- Professor Harry Markowitz developed portfolio theory, which looks at how investment returns can be optimized.
- Economists had long understood the common sense of diversification of a portfolio.
- The expression that "do not put all your eggs in one basket" is certainly not new.
- Markowitz showed how to measure the risk of various securities and how to combine them in a portfolio to get the maximum return for a given risk.

Portfolios of Two Risky Assets

- Imagine a portfolio made up of shares in just two companies: Amazon (denoted as A) and GM (denoted as B).
- Let r_A denote monthly log return on A
- Let r_B denote monthly log return on B.
- We assume to have information about the means, variances and covariances of these two returns.
- The mean of expected value of the returns:

$$\mu_A = E(r_A), \quad \text{and} \quad \mu_B = E(r_B).$$

  These are our best guess for the monthly returns.
- Variances

$$\sigma_A^2 = \text{Var}(r_A) = E(r_A - \mu_A)^2 \text{ is the variance of } r_A,$$
$$\sigma_B^2 = \text{Var}(r_B) = E(r_B - \mu_B)^2 \text{ is the variance of } r_B,$$

  while standard deviations are $\sigma_A = \sqrt{\sigma_A^2}$ and $\sigma_B = \sqrt{\sigma_B^2}$.

How risky is an asset?

- We say that the higher σ_A is, the riskier the asset A. Why?
- A higher σ_A implies that the asset's returns are more dispersed around the mean, indicating greater uncertainty and potential variability in returns.

Covariance

- Population measure

$$\sigma_{AB} = \text{Cov}(r_A, r_B) = E\left[(r_A - \mu_A)(r_B - \mu_B)\right].$$

- Sample measure

$$\widehat{\sigma}_{AB} = \widehat{\text{Cov}}(r_A, r_B) = \frac{1}{n}\sum_{i=1}^{n}(r_{i,A} - \overline{r}_A)(r_{i,B} - \overline{r}_B).$$

Co-movement between two assets

- If $\text{Cov}(r_A, r_B) > 0$: the two return series tend to move in the same direction.
- If $\text{Cov}(r_A, r_B) < 0$: the two return series tend to move in opposite directions.
- If $\text{Cov}(r_A, r_B) = 0$: the two return series tend to move independently.

Correlation

- It measures the strength of the dependence between the returns r_A and r_B

$$\rho_{AB} = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$$

- If ρ_{AB} is close to one, then the two return series mimic each other closely.
- If ρ_{AB} is close to zero, then the two return series show very little relationship.

Example

- This table provides information about two assets' returns

μ_A	μ_A	σ_A^2	σ_B^2	σ_A	σ_B	ρ_{AB}
0.175	0.055	0.067	0.013	0.26	0.11	−0.164

- Asset A is the higher risk asset with an annual return of μ_A=17.5% and standard deviation of σ_A=26%.
- Asset B is the lower risk asset with an annual return of μ_B=5.5% and standard deviation of σ_B=11%.
- The two assets are slightly negatively correlated with ρ_{AB}=−16.4%

Portfolios of Two Risky Assets

- Our investment in the two stocks forms a portfolio.
- The relative weights of the two stock holdings in the portfolio are w_A and w_B, which are assumed to be positive.
- Assume that $w_A + w_B = 1$ as all wealth is invested in A or B.

Portfolio Return

- The return on the portfolio is given by

$$r_p = w_A \times r_A + w_B \times r_B.$$

- The mean or expected return of r_p is

$$E(r_p) = w_A \times E(r_A) + w_B \times E(r_B) = w_A \times \mu_A + w_B \times \mu_B.$$

- The variance of r_p is

$$\begin{aligned} \mathrm{Var}(r_p) &= w_A^2 \times \mathrm{Var}(r_A) + w_B^2 \times \mathrm{Var}(r_B) + 2w_A \times w_B \times \mathrm{Cov}(r_A, r_B) \\ &= w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_A w_B \sigma_{AB} \end{aligned}$$

- The standard deviation of r_p is the square root of $\mathrm{Var}(r_p)$.
- A positive covariance ($\sigma_{AB} > 0$) will tend to increase the portfolio variance.
- A negative covariance ($\sigma_{AB} < 0$) will tend to reduce the portfolio variance.

Diversify risk

- Finding assets with negatively correlation (or negatively covariance) returns can be very beneficial, because risk, as measured by portfolio standard deviation, is reduced.
- If we cannot do that, at least avoid shares whose returns are very highly positively correlated with each other (don't put all eggs in one basket).

Variance of portfolio return

- Recall that

$$\rho_{AB} = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$$

- The covariance can be expressed as

$$\sigma_{AB} = \rho_{AB} \times \sigma_A \times \sigma_B$$

Variance of portfolio return

The variance of r_p can be expressed as

$$\begin{aligned}\text{Var}(r_p) &= w_A^2 \times \text{Var}(r_A) + w_B^2 \times \text{Var}(r_B) + 2w_A \times w_B \times \text{Cov}(r_A, r_B) \\ &= w_A^2 \times \sigma_A^2 + w_B^2 \times \sigma_B^2 + 2w_A \times w_B \times \rho_{AB} \times \sigma_A \times \sigma_B\end{aligned}$$

Example

- Consider creating some portfolios using the asset information given in an earlier Table.

μ_A	μ_A	σ_A^2	σ_B^2	σ_A	σ_B	ρ_{AB}
0.175	0.055	0.067	0.013	0.26	0.11	−0.164

- Assume an equally weighted portfolio with $w_a = w_B = 0.5$
- The Mean of the portfolio is

$$r_p = w_A \times r_A + w_B \times r_B = 0.5 \times 0.175 + 0.5 \times 0.055 = 0.115$$

Example: Portfolio variance

$$\begin{aligned}\text{Var}(r_p) &= w_A^2 \times \text{Var}(r_A) + w_B^2 \times \text{Var}(r_B) + 2 w_A \times w_B \times \text{Cov}(r_A, r_B) \\ &= 0.5^2 \times 0.067 + 0.5^2 \times 0.013^2 + 2 \times 0.5 \times 0.5 \times (-0.164) \times 0.26 \times 0.11 \\ &= 0.0177\end{aligned}$$

The standard deviation of r_p is

$$\sigma_p = \sqrt{\text{Var}(r_p)} = \sqrt{0.0177} = 0.133$$

Comparison among the standard deviations

- Portfolio standard deviation: $\sigma_p = \sqrt{\text{Var}(r_p)} = 0.133$
- Standard deviation of r_A: $\sigma_A = 0.26$
- Standard deviation of r_B: $\sigma_A = 0.11$

Notice that the portfolio risk σ_p is less than the risk of asset A. This reflects risk reduction via diversification.