# Quantitative Methods for Finance

Prof. Sanjiv R. Das
(My main web page)

February 11, 2010

## Contents

# 1 Overview

Here is this entire web site in a single pdf.

1. *Learning Objectives*: The goal of this course is for Finance majors to acquire knowledge of the most important quantitative tools they will need for the finance

curriculum. The students will

- Comprehend important concepts, techniques and tools in mathematics and statistics relevant for modern finance.
- Understand where these tools are applied in practice.
- Apply widely-used software to implement these techniques.

The course goal is to make sure that all Finance majors reach a baseline level of competence in various quantitative methods. The course is especially intended for those students who fear math, yet have a desire to come to grips with it.

2. *Pre-requisite*: OMIS 40 and OMIS 41.

3. *Class Format*: This is a regular 5-unit class and it will meet on the quarter schedule. Sessions will cover concepts and training in the use of software tools.

   There will be group homework assignments, because the only way to learn these concepts is by doing the homework. Group sizes will be determined after enrollment is final. There will be two mid-term exams and a final exam covering all the material as well. The homeworks, mid-terms and the final will count for 25% each.

4. *Software*: Software is an essential tool in Finance. While the primary focus will be on using Excel, there are a few concepts that are not implementable in Excel. In such cases, students will be introduced to free math/stat packages such as Octave and R. Also, for symbolic math students may access the free tool Maxima. See the following web links for these tools:

   - Octave
   - R
   - Maxima

   It is envisaged that these tools will be used in class, but formal training in the use of these packages will need to occur in extra sessions outside of class. To complement these tools it is also recommended that advanced students take a course in Stata offered by the Economics department.

5. *Course Materials*: The course materials will be posted on this web site and on Angel. All class notes will be provided either as slides, notes or links to sources on the web. No textbook is needed. The underlying philosophy is to use open-source materials as much as possible, for example, with the use of free software noted above.

   I will provide brief notes for each class. In addition, I will be using a tablet PC in class, and will save everything that is written and post it to Angel after class. I will also guide students to appropriate web sites that will have relevant material for each topic.

There are two good books that relate well to 50% of the material in this course, but do not have finance applications, nor do they have material on statistics. Hence, it is not necessary to buy a book. But these books are good for reference, especially for the more mathematically interested student.

- Mathematics for Economists, by Carl P. Simon & Lawrence E. Blume, (W.W. Norton & Co.).
- Fundamental Methods of Mathematical Economics, by Alpha Chiang & Kevin Wainwright, (McGraw Hill).

# 2 Software and Data

## 2.1 Using R and Excel

- R is a full-strength matrix language that may be used for numerical computing as well as econometrics.
- Download R here
- An example of a simple command and plot from R is one that generates several normal random numbers and plots them on a histogram:
  `hist(rnorm(100000),100)`
- Excel VBA is a programming language for use with Excel. We will see many applications in finance with Excel. For now, we will look at one useful feature, i.e., extracting data from the internet.

## 2.2 Sources of Financial Data

1. Using WRDS (Wharton Research Data Services) by accessing the same using the library server.
2. Downloading the data from a web source such as Yahoo! Finance.
3. The Federal Reserve's data repository. Also known as FRED. Here is the URL
4. Bloomberg.
5. There are many other sources that you may wish to access. Press reports and articles are available from `Factiva`, also accessible through the library server.

## 2.3 Representing Financial Data

1. Loading in an ascii data set into R:
   `read.table("filename",header=TRUE)`

2. Loading in a csv file into R:
   `read.csv("filename",header=TRUE,sep=",")`

3. Plotting a time series of stock prices:
   `plot(x,yseries,main="Title",xlab="x-axis title",ylab="y-axis title",color="red",`

4. Adding a line to this plot:
   `lines(x,yseries2,color="blue")`

For a superb collection of graph templates in R, check out this website. It comes with the source code for generating many cool types of graphs.

# 3 Basic Concepts

## 3.1 Interest Rates and Exchange Rates

1. What is the yield curve?

2. What is the nominal interest rate comprised of?

$$r = R + I, \quad R = E(R) + \pi(R), \quad I = E(I) + \pi(I)$$

3. Exchange rates. Example: 1 euro = 1.5 dollars.

4. Two-way quotes. Bid-Offer: 1.52–1.55 dollars/euro. Skewing quotes depending on trade preference.

## 3.2 Forward Exchange Rates

Let's suppose that today the exchange rate is 1.5 dollars/euro. The one-year interest rates are as follows: dollars 1%, euros 2%. What is the one-year forward exchange rate?

Answer: $F = 1.5 \times \frac{1+0.01}{1+0.02} = 1.4853$

Note: this is similar to finding the exchange ratio for two forward bonds.

We introduce the principle of covered interest-rate parity to justify this result. We proceed in three steps as follows:

1. Begin with 1 euro. Convert it at the spot exchange rate to get 1.5 dollars.
2. Invest the dollars at 1% to receive 1.515 dollars after one year.
3. Also contract forward to exchange the dollars back into euros at the end of the year at the forward exchange rate of 1.4853 dollars/euro.

At the end of the year, convert the 1.515 dollars into euros to get 1.02 euros (noting that 1.515/1.4853=1.02). This is exactly what one would have earned by simply investing in euros directly. Hence an investor is indifferent between investing in euros or dollars at the given exchange rate.

If the forward exchange rate was not 1.4853, then an investor would not be indifferent between the two alternatives (investing directly in euros versus the three-step dollar route above). In fact, he would be able to construct an arbitrage by shorting the approach that delievered fewer euros after one year, and going long the strategy that delivered more euros.

## 3.3 Descriptive Statistics

We now return to the features of financial time series and describe them more statistically. There are several summary or descriptive statistics that are used to describe data sets. The more common ones are:

1. *Mean*. The mean is the average of a list of numbers. It is computed as

$$\text{Mean} = \frac{1}{N} \sum_{j=1}^{N} x_j$$

   In R, the mean may be computed using the command `mean(x)`, where `x` is a vector of numbers. Just in case you are not familiar with the "summation" notation ($\Sigma$), and you should be, here it is

$$\sum_{j=1}^{N} x_j = x_1 + x_2 + \ldots + x_{N-1} + x_N$$

   More generally, the mean is the expected value of a random variable drawn from a statistical distribution. If $x$ takes $N$ distinct values, each with probability $p(x)$, then the expected or mean value of $x$ is written as

$$E(x) = \sum_{j=1}^{N} x\ p(x)\ \Delta x$$

where $\Delta x$ is the difference between the various values of $x$ in the histogram. It is essentially the width of each histogram bin. The expression above is the discrete form of the integral version of the expected value of $x$, i.e.,

$$E(x) = \int_x x\ p(x)\ dx$$

If we are given the histogram of $x$ in frequency distribution form, i.e., $p(x)$. then mean may be computed easily. Here is an example in R:

```
> x = rnorm(1000000)
> sol=hist(x,1000)
> summary(sol)
            Length Class  Mode
breaks      1038   -none- numeric
counts      1037   -none- numeric
intensities 1037   -none- numeric
density     1037   -none- numeric
mids        1037   -none- numeric
xname          1   -none- character
equidist       1   -none- logical
> mean(x)
[1] -0.001124247
> dx=sol$mids[2]-sol$mids[1]
> Ex = sum(sol$mids * sol$density)*dx
> Ex
[1] -0.00112431
```

We see that the mean is the same (up to minor discretization error) whether it is calculated using the number list or using the histogram. You should be able to see that the histogram is the discrete proxy for the continuous probability density function.

All else constant, investors prefer higher mean return on an investment.

2. *Median.* The median is the value of $x$ such that exactly half the values in an ordered list of $x$ lie to the right and left of the median. In R we compute the median as follows:

```
> x = runif(1000)
> median(x)
[1] 0.5218695
```

The function `runif` generates uniform random numbers. A more general way of generating the median is to find $M$ such that the following function is minimized:

$$\min_M \sum_{j=1}^{N} |x_j - M|$$

3. *Variance and Standard Deviation.* The variance is computed as follows:

$$Var(x) = \frac{1}{N} \sum_{j=1}^{N} (x_j - E(x))^2$$

The standard deviation of $x$ is just the square-root of the variance, i.e.,

$$\text{Std Dev} = \sigma_x = \sqrt{Var(x)}$$

A quicker calculation to get the variance is as follows:

$$Var(x) = E(x^2) - E(x)^2$$

where

$$E(x^2) = \frac{1}{N} \sum_{j=1}^{N} x_j^2$$

By now you should have become comfortable with the summation notation.

All else constant, investors prefer lower variance of returns on an investment.

4. *Skewness.* Skewness is a measure of the asymmetry in the distribution of $x$ values. When a distribution is left-skewed, it means that the left tail of the distribution is fatter than the right tail. It also states that the median is less than the mean.

When a distribution is right-skewed, it means that the right tail of the distribution is fatter than the left tail. The median will be greater than the mean.

The skewness is defined as

$$\text{Skewness}(x) = \frac{1}{\sigma_x^3} \frac{1}{N} \sum_{j=1}^{N} [x_j - E(x)]^3$$

For a symmetric distribution, skewness is zero.

All else constant, investors prefer positive skewness of returns on an investment.

5. *Kurtosis.* Kurtosis is a measure of how fat-tailed a distribution is. It tells you how large the outliers are (on the left and right side of the distribution) relative to the standard deviation. The greater the outliers relative to the standard deviation, the bigger the kurtosis.

   Kurtosis is defined as

$$\text{Kurtosis}(x) = \frac{1}{\sigma_x^4} \frac{1}{N} \sum_{j=1}^{N} [x_j - E(x)]^4$$

   For a normal distribution kurtosis is equal to 3. Therefore, the term "excess kurtosis" equals kurtosis minus 3.

   All else constant, investors prefer lower kurtosis of returns on an investment.

As we have seen, investors prefer higher mean returns, lower standard deviation, positive skewness (versus negative skewness), and lower kurtosis. The mean is a measure of return and the variance, skewness and kurtosis are all metrics for risk.

There is a common tendency in the investing world to consider risk in terms of variance only. One should be very careful to not ignore skewness and kurtosis. The recent financial crisis is an important reminder that negative skewness and high kurtosis are both indicators of extreme risk and can have devastating effects. The occurence of recent large losses has been attributed to what is known as the Black Swan effect.

# 4 Stochastic Processes and Dynamics

The normative modeling of financial series is undertaken using stochastic process specifications. We will specify the processes in discrete time here.

## 4.1 Equities

The process for the stock price ($S$) is usually specified as follows

$$\frac{\Delta S}{S} = \mu\, h + \sigma\, \epsilon \sqrt{h}$$

where $h$ is the time interval (specified in years) over which the stock return $\frac{\Delta S}{S}$ is computed. $\Delta S$ is the change in the stock price over the interval $h$. For instance, if the time interval is one month, then $h = 1/12$. The expected return per year is $\mu$ and the annual standard deviation of returns is $\sigma$. The variable $\epsilon$ is a standard normal variable, i.e., $\epsilon \sim N(0, 1)$.

This process gives the dynamics for the movement of the stock price. The following are its features:

1. Stock returns are normally distributed.
2. The dynamic movement of the stock price is given by the following equation:

$$S(t+h) \approx S(t) \exp\left[\mu h + \sigma \epsilon \sqrt{h}\right]$$

3. Stock prices are *not* normally distributed. From the previous equation, we infer that they are lognormally distributed. We will discuss this in more detail later on in the course.

Why is this a popular model for stock price movements?

1. Over the long-run, stock prices evidence exponential patterns of growth. To see this, look at the trend line of the S&P500 index over a century.
2. Markets are assumed to be efficient, and hence, the returns from one period to the next are independent of each other. Convince yourself why this must be so. Suppose returns in each period are given by $r = \mu h + \sigma \epsilon \sqrt{h}$, then the return over a year will be $r_1 = \mu + \sigma \epsilon$. The number of periods per year is $m = 1/h$. Hence, the variance of returns over one year is the sum of the variances of daily returns

$$Var(r + r + \ldots + r, m \text{ times}) = mVar(r) = m\sigma^2 h = \sigma^2 = Var(r_1)$$

3. The process may also be used to estimate the *volatility* of stock returns. Volatility is defined as the annualized standard deviation of stock returns. Suppose we have a time series of stock prices $S_0, S_1, S_2, \ldots, S_N$. We convert these into continuously-compounded returns as follows:

$$r(t+h) = \ln\left[\frac{S(t+h)}{S(t)}\right]$$

   Volatility is then given by the annualized standard deviation of these returns.

$$\sigma = \sqrt{\frac{1}{h} Var(r)}$$

We note here that it is also common to use the same stochastic process for exchange rates. However, the empirical evidence that exchange rates grow exponentially over time is weak.

## 4.2 Interest Rates

Interest rates, unlike stocks, do not grow exponentially over time. They tend to cycle up and down with the economy. In other words, they oscillate randomly around some average level. A simple specification that captures this oscillatory behavior is as follows:

$$\Delta r = k(\theta - r)h + \sigma\epsilon\sqrt{h}$$

where $k$ is known as the rate of "mean-reversion" and $\theta$ is the average level or "long-run mean" of the interest rate. To see why this produces oscillatory behavior, notice that when $r > \theta$, the term $k(\theta - r)h$ is negative, which produces a downward pull on the interest rate towards the mean interest-rate level $\theta$. An upward move is induced when $r < \theta$. Randomness around these movements is generated by the random term $\sigma\epsilon\sqrt{h}$.

We may write the dynamics of the interest rate as follows:

$$r(t+h) = r(t) + k(\theta - r(t))h + \sigma\epsilon\sqrt{h}$$

How do we estimate $k, \theta, \sigma$ from a time series of interest rates? First, let's write the previous equation as follows:

$$r(t+h) = k\theta h + [1 - kh]r(t) + \sigma\epsilon\sqrt{h}$$

which may be written as the following linear regression equation:

$$r(t+h) = a + b\, r(t) + e$$

where $a = k\theta h$, $b = 1 - kh$, and $e = \sigma\epsilon\sqrt{h}$. Running the regression, we get as output the values $\{a, b, \sigma_e^2\}$. Using these values we solve for $\{k, \theta, \sigma\}$.

- From $b = 1 - kh$, we get $k = (1 - b)/h$.
- From $a = k\theta h$, we get $\theta = a/(kh) = a/(1 - b)$.
- From $e = \sigma\epsilon\sqrt{h}$, we get that $\sigma = \sqrt{\frac{\sigma_e^2}{h}}$.

We demonstrate this by downloading data from the Fed web site and implementing the procedure in R.

# 5 Statistics

## 5.1 Arithmetic and Geometric Mean

- The arithmetic mean has already been defined previously as

$$E(x) = \frac{1}{N} \sum_{j=1}^{N} x_j$$

- The geometric mean is defined as

$$E_{\text{geometric}}(x) = \left[ \prod_{j=1}^{N} x_j \right]^{1/N}$$

  where the superscript $1/N$ means that we are taking the $n$-th root. We note here that in this case $x_j > 0$ is required to prevent ending up with complex roots.

- The geometric mean is smaller than the arithmetic mean.

- The arithmetic mean is used in finance in many areas, especially when trying to understand the distribution of returns of securities.

- The geometric mean is the concept needed when thinking about holding period returns.

Here is the sample code in R for computing the arithmetic and geometric mean of returns for a firm.

```
> arithmetic_mean = mean(returns)
> geometric_mean = prod((1+returns))^(1/length(returns)) - 1
```

## 5.2 Samples and Populations

The population is the entire universe of observations that one may sample from—a sample if therefore a subset of the population.

The population variance has been defined previously as

$$Var(x) = \frac{1}{N} \sum_{j=1}^{N} (x_j - E(x))^2$$

In contrast, the sample variance is obtained using the following formula

$$Var_{\text{sample}}(x) = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - E(x))^2$$

Note that the divisor is $N-1$, not $N$ as with the population variance. This is because one degree of freedom has been used up to calculate the mean.

For practical purposes, when samples are sufficiently large, this distinction between sample and population variance becomes immaterial. You should however, be aware of this when using software such as Excel. In this class, we assume that the divisor is $N$.

## 5.3 Covariance and Correlation

- Covariance measures how much two variables change together. It is defined as

$$Cov(x,y) = \frac{1}{N}\sum_{j=1}^{N}[x_j - E(x)]\cdot[y_j - E(y)]$$

- A simpler expression for the covariance is

$$Cov(x,y) = E(x \cdot y) - E(x)\cdot E(y)$$

- Covariance may be easily computed in R as follows:

```
> x = rnorm(100)
> y = runif(100)
> cov(x,y)
[1] -0.01699549
```

  The negative sign for covariance means that in this sample, $x$ and $y$ change inversely with each other *on average*.

- Correlation is a normalized measure of dependence. It is covariance normalized by the standard deviations of the two variables. Hence, it is defined as

$$Corr(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

- Correlation lies between $-1$ and $+1$.

- Correlation tells us how much two variables move in the same direction, regardless of the size of the movements.

- In R, we compute correlation as follows:

```
> cor(x,y)
[1] -0.05655692
```

- In finance, we care about correlations between securities in portfolios. The lower correlation is, the more diversification in the portfolio. For correlation less than 1, diversification is achieved. (Why not correlation less than zero?)

- Covariance also appears in the capital asset pricing model (CAPM). Recall that the beta of a security is defined as a covariance, i.e.,

$$\beta_i = \frac{Cov(r_i, r_m)}{\sigma_i \sigma_m}$$

where the $m$ subscript refers to the stock market.

## 5.4 Recap Basic Formulae

Here are some essential formulae we have looked at so far, and you should make sure you know these well:

- Expectations or means: $E(ax + by) = aE(x) + bE(y)$.

- Variance: $Var(x) = E(x^2) - [E(x)]^2$. The variance $Var(a \cdot x) = a^2 \cdot Var(x)$, where $a$ is a constant, and $x$ is a random variable.

- Covariance: $Cov(x, y) = E(x \cdot y) - E(x) \cdot E(y)$. Note that $Cov(x, x) = Var(x)$, the covariance of a variable with itself is just the variance.

- If $x = a + b\epsilon$, where $\epsilon \sim N(0, 1)$, then $x$ is also normally distributed with mean $a$ and variance $b^2$.

- Correlation: $Corr(x, y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$

- Variance of a sum: $Var(x_1 + x_2) = Var(x_1) + Var(x_2) + 2Cov(x_1, x_2)$. If the variables are independent and identically distributed, then the *variance of the sum* of random variables is equal to the *sum of the variances*. In particular, for our example here, you can see that $Var(x_1 + x_2) = Var(x_1) + Var(x_2)$.

## 5.5 Difference of Means

We are sometimes interested in whether two samples are different from each other. In finance, this arises in many settings:

- Categorizing firms into rating classes.
- Classifying firms into default and non-default types.
- Characteristic-based, i.e., are sets of firms sufficiently different in their characteristics such as size, leverage, profitability, operating features, etc.
- Distinguishing credit quality amongst credit card customers.

Suppose we have two samples with variables $x_1$ and $x_2$ of size $N_1$ and $N_2$, respectively. Each sample has mean

$$m_1 = E(x_1), \quad m_2 = E(x_2)$$

and variance

$$\sigma_1^2 = Var(x_1), \quad \sigma_2^2 = Var(x_2)$$

The difference of means is just

$$m_1 - m_2$$

Is this difference statistically significant? To figure this out, we need the standard deviation of the difference in means, i.e., we want to calculate $\sigma_{m_1-m_2}$. This is given by

$$\sigma_{m_1-m_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

At the 5% confidence level, we check whether

$$\frac{m_1 - m_2}{\sigma_{m_1-m_2}} > 1.96$$

As an example of using this technique, see the paper by Das, Jo, and Kim (Tables 2, 3, and 4) on differences in performance of syndicated versus non-syndicated new ventures.

# 6 Regression Analysis

Regression analysis is about the *estimation* of a mathematical relationship between a "dependent" variable and explanatory or "independent" variables. These are also known as the left-hand-side variable and right-hand-side variables, respectively.

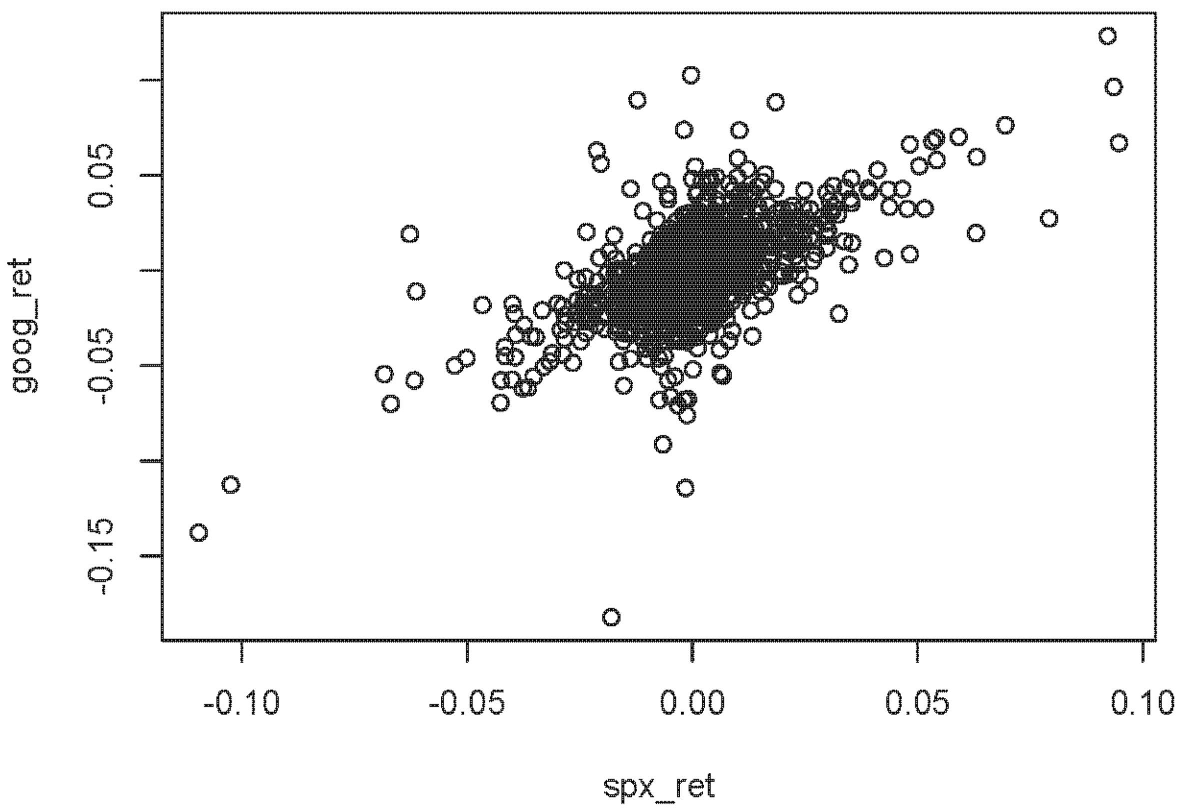
The relationship may be *linear* or *non-linear*. Can you think of financial relationships that may be linear? Non-linear?

As an example, let's explore the relationship between our favorite stock and the stock market. Using R, we will read in the data and then *run* a regression of the stock returns on the market's returns.

```
> data = read.csv("GOOG_SPX.csv",header=TRUE,sep=",")
> goog = data[,7]
> spx = data[,13]
> n = length(goog)
> goog_ret = log(goog[2:n]/goog[1:(n-1)])
> spx_ret = log(spx[2:n]/spx[1:(n-1)])
> res = lm(goog_ret ~ spx_ret)
> summary(res)

Call:
lm(formula = goog_ret ~ spx_ret)

Residuals:
       Min         1Q     Median         3Q        Max
-0.1648555 -0.0088648  0.0006444  0.0086946  0.1040159

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0009402  0.0005205  -1.806   0.0711 .
spx_ret      0.9153839  0.0342983  26.689   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01845 on 1255 degrees of freedom
Multiple R-squared: 0.3621,Adjusted R-squared: 0.3616
F-statistic: 712.3 on 1 and 1255 DF,  p-value: < 2.2e-16
```

First let's look at the data.

```
plot(spx_ret,goog_ret)
```




So, there are several terms we need to understand:

1. Linear regression: we may draw a line through the scatter plot that best fits the relationship between SPX and GOOG. The equation of this line is

$$r_{GOOG} = a + b \cdot r_{SPX} + e$$

   where $a$ is the "intercept" and $b$ is the "slope" in the regression.

2. Residuals ($e$): Unless the observation (a pair of GOOG and SPX) lies precisely on the regression line, there will be some error in fit, or residual. The average residual must be zero.

3. R-squared: the percentage of the variation in GOOG explained by SPX. It is the square of the correlation between GOOG and SPX.

4. F-statistic: a measure of whether the independent variables *jointly* explain the dependent variable in a statistically significant manner.

5. T-statistic: tells you if the individual independent variable is significant in explaining the variation in the dependent variable. Usually we want it to be greater than 2.

6. p-value: states the probability that the statistic in question (e.g., t-statistic, F-statistic, etc.) is non-significant statistically.

Look at the results of the regression. What interpretations might you draw from all the information in the output?

From a financial point of view, what can you say about the relationship of GOOG to the stock market?

Let's try an extended regression as follows.

```
> spx_ret2 = spx_ret^2
> res = lm(goog_ret ~ spx_ret+spx_ret2)
> summary(res)

Call:
lm(formula = goog_ret ~ spx_ret + spx_ret2)

Residuals:
       Min         1Q     Median         3Q        Max
-0.1648143 -0.0089160  0.0006191  0.0086994  0.1039573

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0008813  0.0005418  -1.626    0.104
spx_ret      0.9163462  0.0343972  26.640   <2e-16 ***
spx_ret2    -0.2560966  0.6521545  -0.393    0.695
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01846 on 1254 degrees of freedom
Multiple R-squared: 0.3621,Adjusted R-squared: 0.3611
F-statistic:   356 on 2 and 1254 DF,  p-value: < 2.2e-16
```

What does this regression tell you? Is it a non-linear regression?

## 6.1 Testing the CAPM

You will surely recall the Capital Asset Pricing Model from your first finance class. The model states that the return on a stock is related to the return on the market through the following linear relationship:

$$E(r_j) = r_f + \beta_j[E(r_m) - r_f]$$

This model may be estimated using a regression. i.e.,

$$r_{jt} - r_{ft} = \alpha + \beta_j[r_{mt} - r_{ft}] + e_t$$

where $e$ is a residual error. You can use regression to determine the intercept term $\alpha$ and the slope coefficient $\beta$. If the CAPM does explain the relation between the returns of the stock and that of the market, then the intercept $\alpha$ should be zero.

The coefficient $\beta$ is a measure of "systematic" or "non-diversifiable" risk. The higher it is, the more sensitive the stock is to the market. High beta stocks have more systematic risk than low beta stocks.

But what the CAPM implies economically is that since high beta stocks have more risk, they should also earn higher returns. Therefore, a *cross-sectional* regression of average stock returns on betas should result in a strong relationship between systematic risk and return. We need to run the following regression:

$$r_j = \gamma_0 + \gamma_1\beta_j + w_j$$

where $w$ is a residual error, and if the CAPM is valid, then $\gamma_1$ should be statistically significant, and $\gamma_1 > 0$.

# 7 Vector Algebra for Portfolios

Returns for each stock:

$$\mathbf{R} = \begin{pmatrix} R_2 \\ R_2 \\ : \\ : \\ R_n \end{pmatrix}$$

Unit vector:

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ : \\ : \\ 1 \end{pmatrix}$$

Portfolio weights:

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ : \\ : \\ w_n \end{pmatrix}$$

Riskless asset: $R_0 = r_f$, weight $w_0$

$$\sum_{i=0}^{n} w_i = 1, \quad \mathbf{w}'\mathbf{1} = 1$$

$$w_0 = 1 - \sum_{i=1}^{n} w_i$$

Portfolio return:

$$R_p = \mathbf{w}'\mathbf{R}$$

```
> w = c(0.3,0.4,0.2)
> R = c(0.10,0.10,-0.05)
> rf = 0.02
> w0 = 1-sum(w)
> w = c(w0,w)
> w
[1] 0.1 0.3 0.4 0.2
> t(w)
     [,1] [,2] [,3] [,4]
[1,]  0.1  0.3  0.4  0.2
> R = c(rf,R)
> R
[1]  0.02  0.10  0.10 -0.05
> Rp = t(w) %*% R
> Rp
      [,1]
[1,] 0.062
```

## 7.1 Portfolio Moments

Note that $\mathbf{R}$ is a random variable, as is $R_p$.

Suppose $\mathbf{R} \sim MVN[\mu; \mathbf{\Sigma}]$

Portfolio mean return:

$$E[\mathbf{w}'\mathbf{R}] = \mathbf{w}'E[\mathbf{R}] = \mathbf{w}'\mu$$

Portfolio return variance:

$$Var(\mathbf{w}'\mathbf{R}) = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$$

```
> w = c(0.3,0.4,0.3)
> SIGMA = c(0.02,0.01,0.01,0.01,0.03,0.01,0.01,0.01,0.04)
> SIGMA = matrix(SIGMA,3,3)
> SIGMA
     [,1] [,2] [,3]
[1,] 0.02 0.01 0.01
[2,] 0.01 0.03 0.01
[3,] 0.01 0.01 0.04
> w
[1] 0.3 0.4 0.3
> Variance = t(w) %*% SIGMA %*% w
> Variance
       [,1]
[1,] 0.0168
```

Multivariate normal density function:

$$f(\mathbf{R}) = \frac{1}{2\pi^{n/2}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(\mathbf{R}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{R}-\mu)\right]$$

## 7.2 Generating the portfolio frontier by simulation

Since we have the portfolio moments from the previous subsection, it is easy to generate the portfolio frontier which is just a plot of efficient portfolios in mean versus variance space (or mean versus standard deviation of the portfolio).

The R program to do this is as follows:

```
portf_mv = function(w,R,cv) {
    portf_mean = t(w) %*% R
    portf_sd = sqrt(t(w) %*% cv %*% w)
    result = c(portf_mean,portf_sd)
}
m = 1000
```

```
w1 = rnorm(m)
w2 = rnorm(m)
w3 = 1-w1-w2

R = matrix(c(0.01,0.20,-0.15),3,1)
cv = matrix(c(0.04,0.01,0.01,0.01,0.05,0.01,0.01,0.01,0.06),3,3)

mn = matrix(0,m,1)
vr = matrix(0,m,1)

for (j in 1:m) {
    w = matrix(c(w1[j],w2[j],w3[j]),3,1)
    res = portf_mv(w,R,cv)
    mn[j] = res[1]
    vr[j] = res[2]^2
}

plot(vr,mn)

sharpe = mn/sqrt(vr)
maxp = max(sharpe)
idx = which(sharpe==maxp)
print(c(w1[idx],w2[idx],w3[idx],maxp))
```

Running this program will generate the a plot of various portfolios for a three security choice-set. Many of these are "dominated", i.e., there are other portfolios that have the same mean but lower variance, or the same variance and higher mean. The "efficient" portfolio is the one on the upper envelope of the plot.

## 7.3 Regression

$$R_{it} = \sum_{j=0}^{k} \beta_{ik} R_{kt} + e_{it}, \quad \forall i.$$

$(k+1)$ independent variables, and usually $k=0$ is for the intercept.

Compactly, the same may be written as:

$$R_{it} = \beta_{\mathbf{i}}' \mathbf{R_t} + e_{it}$$

Each regression coefficient is:

$$\beta_{ik} = \frac{Cov(R_i, R_k)}{Var(R_k)}$$

In vector form, all coefficients may be calculated at once:

$$\beta_{\mathbf{i}} = (\mathbf{R'_{kt} R_{kt}})^{\mathbf{-1}} (\mathbf{R'_{kt} R_{it}})$$

We explain this in greater detail below.

## 7.4 Regression using Matrices

In a linear regression, we have

$$Y = X'\beta + e$$

where $X \in R^{t \times n}$ and the regression solution is (as is known from before), simply equal to $\beta = (X'X)^{-1}(X'Y)$.

To get this result we minimize the sum of squared errors.

$$\begin{aligned}
\min_{\beta} e'e &= (Y - X'\beta)'(Y - X'\beta) \\
&= (Y - X'\beta)'Y - (Y - X'\beta)'(X'\beta) \\
&= Y'Y - (X'\beta)'Y - Y'(X'\beta) + \beta^2(X'X) \\
&= Y'Y - 2(X'\beta)'Y + \beta^2(X'X)
\end{aligned}$$

Differentiating w.r.t. $\beta$ gives the following f.o.c:

$$\begin{aligned}
2\beta(X'X) - 2(X'Y) &= \mathbf{0} \\
&\Longrightarrow \\
\beta &= (X'X)^{-1}(X'Y)
\end{aligned}$$

```
> y = matrix(runif(250),250,1)
> x = matrix(1,250,1)
> x = cbind(x,y^2,y^3)
> dim(x)
[1] 250   3
> y = y + matrix(rnorm(250)*0.01,250,1)
> res = lm(y ~ x[,2:3])
> res$coefficients
(Intercept)   x[, 2:3]1   x[, 2:3]2
```

```
 0.08788948  2.34387214 -1.48498282
> coeffs = solve(t(x) %*% x) %*% (t(x) %*% y)
> coeffs
            [,1]
[1,]  0.08788948
[2,]  2.34387214
[3,] -1.48498282
```

## 7.5 Diversification – Independent Returns

$$Var(\mathbf{w'R}) = \mathbf{w'\Sigma w} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{w_i^2} \sigma_{\mathbf{i}}^{\mathbf{2}} + \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{j=1,i\neq j}}^{\mathbf{n}} \mathbf{w_i w_j} \sigma_{\mathbf{ij}}$$

If returns are independent,

$$Var(\mathbf{w'R}) = \mathbf{w'\Sigma w} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{w_i^2} \sigma_{\mathbf{i}}^{\mathbf{2}}$$

## 7.6 Diversification – Dependent Returns

If returns are dependent, and equal amounts are invested in each asset ($w_i = 1/n, \ \forall i$):

$$\begin{aligned} Var(\mathbf{w'R}) &= \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_i^2}{n} + \frac{n-1}{n} \sum_{i=1}^{n} \sum_{j=1,i\neq j}^{n} \frac{\sigma_{ij}}{n(n-1)} \\ &= \frac{1}{n} \bar{\sigma_i}^2 + \frac{n-1}{n} \bar{\sigma_{ij}} \\ &= \frac{1}{n} \bar{\sigma_i}^2 + \left(1 - \frac{1}{n}\right) \bar{\sigma_{ij}} \end{aligned}$$

The first term is the average variance, and the second is the average covariance.

As $n \to \infty$,

$$Var(\mathbf{w'R}) = \bar{\sigma_{\mathbf{ij}}}.$$

Exercise: Suppose all stocks have mean return of 10%, and a standard deviation of returns of 25%. Also assume that the correlation between all stock returns is an average of 40%.

- What is the return variance of each stock?

- What is the return covariance of each pair of stocks?
- How many stocks ($n$) are needed to ensure that the idiosyncratic component of stock return variance of an $n$-stock portfolio is less than 10%?
- Plot the total risk of the portfolio as $n$ increases.

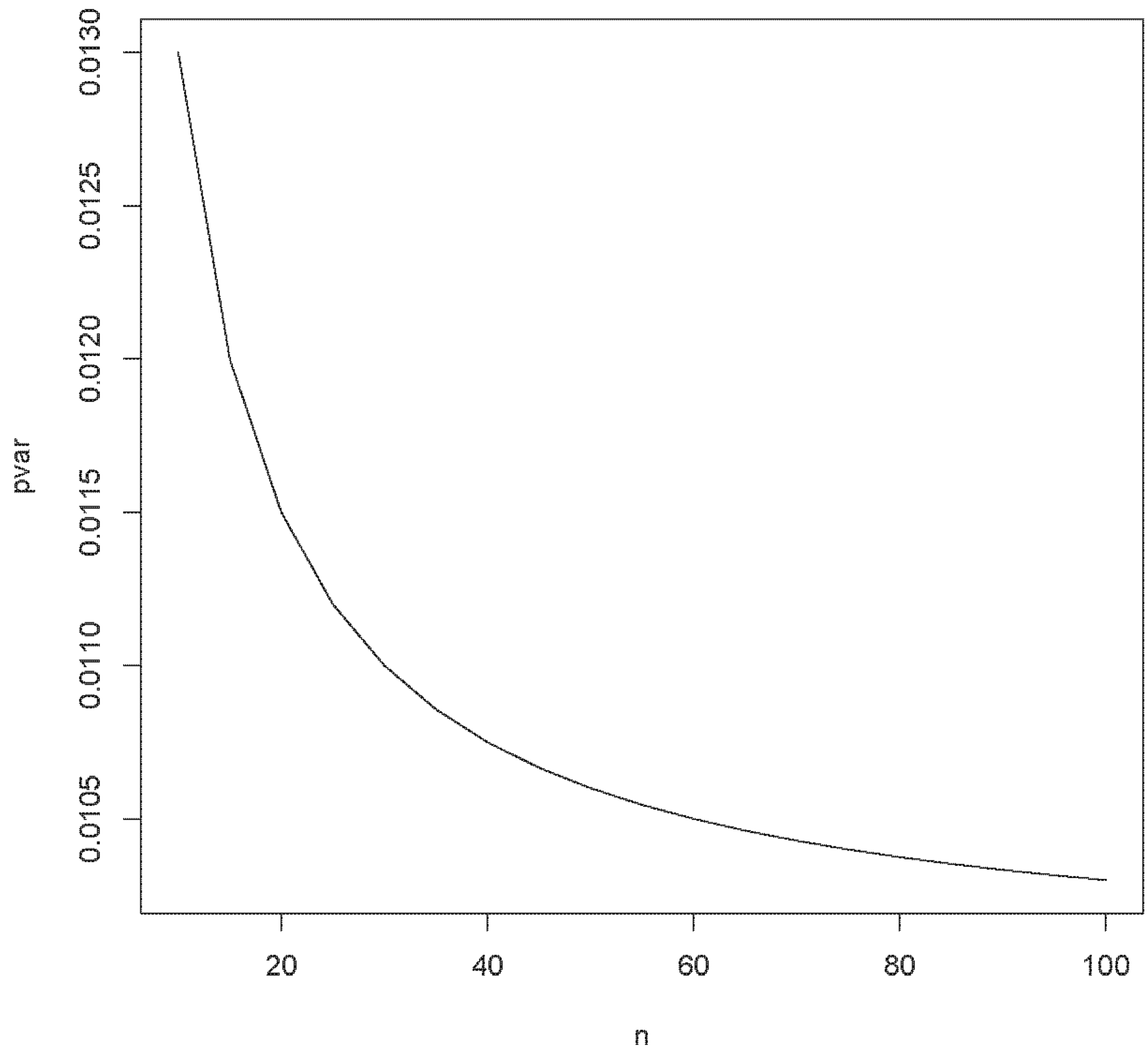
We will implement this model using a function, and thereby also learn about how to use functions in R.

First, we open a program editor window in R and type in the required function for the variance of the portfolio. We create a file called `qfin.R` in which we will build up many finance functions for repeated use. Here is the function that we will place in the file.

```
portfolio_variance = function(w,SIGMA) {
        Variance = t(w) %*% SIGMA %*% w
}
```

Now we use the function repeatedly by creating larger and larger portfolios and watching how the variance becomes smaller for the portfolio. To keep things simple, we assume that there are $n$ stocks each with the same variance equal to 0.04. The covariance between all pairs of stocks is taken to be 0.01. We will write a program loop that will keep on increasing $n$ and store the results in a vector and then we will plot the vector to see how the variance comes down as $n$ increases. Here is the R code.

```
> source("qfin.R")   #Read in the function
> n = seq(10,100,5)
> pvar = NULL
> for (j in n) {
+ SIGMA = matrix(0.01,j,j)
+ diag(SIGMA) = 0.04
+ w = matrix(1/j,j,1)
+ vr = t(w) %*% SIGMA %*% w
+ pvar = c(pvar,vr)
+ }
> pvar
 [1] 0.01300000 0.01200000 0.01150000 0.01120000 0.01100000 0.01085714 0.01075000
 [8] 0.01066667 0.01060000 0.01054545 0.01050000 0.01046154 0.01042857 0.01040000
[15] 0.01037500 0.01035294 0.01033333 0.01031579 0.01030000
> plot(n,pvar,type="l")
```




## 7.7 Recap: Basic Matrix Operations

### 7.7.1 Multidimensional arrays

```
> q = c(2,5,3,1)
> q
[1] 2 5 3 1
> x = array(1:3^4,dim=c(3,3,3,3))
> x[2,2,3,1]
[1] 23
> x = x*0
```

```
> x[2,2,3,1]
[1] 0

x = array(0,dim=c(2,2,2))
```

The last line gives an array of zeros.

### 7.7.2 Transpose

```
> x = c(1,2,3,4)
> x = array(x,dim=c(2,2))
> x
     [,1] [,2]
[1,]    1    3
[2,]    2    4
> t(x)
     [,1] [,2]
[1,]    1    2
[2,]    3    4
```

### 7.7.3 Inverse and Matrix Multiplication

```
> x = matrix(c(1,2,3,4),2,2)
> x
     [,1] [,2]
[1,]    1    3
[2,]    2    4
> x = t(x)
> x
     [,1] [,2]
[1,]    1    2
[2,]    3    4
> solve(x) %*% x
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

## 7.8 Solving Systems of Equations

Suppose we want to solve the following system of equations:

$$\begin{aligned} 3x_1 + 4x_2 - 5x_3 &= 20 \\ 4x_1 - 2x_2 - 11x_3 &= 30 \\ 5x_1 + 3x_2 + 6x_3 &= 40 \end{aligned}$$

This may be written in matrix form as

$$\begin{bmatrix} 3 & 4 & -5 \\ 4 & -2 & -11 \\ 5 & 3 & 6 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 20 \\ 30 \\ 40 \end{pmatrix}$$

The solution to this system of equations is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} 3 & 4 & -5 \\ 4 & -2 & -11 \\ 5 & 3 & 6 \end{bmatrix}^{-1} \begin{pmatrix} 20 \\ 30 \\ 40 \end{pmatrix}$$

We can implement the solution in R:

```
> A = matrix(c(3,4,-5,4,-2,-11,5,3,6),3,3)
> B = matrix(c(20,30,40),3,1)
> x = solve(A) %*% B
> x
          [,1]
[1,]  2.947658
[2,] -2.617080
[3,]  4.325069
```

## 7.9 Markov Chains

The Markov property of a stochastic process is easily understood by the following phrase: "where you will be tomorrow only depends on where you are today, and not on where you were before today."

This means that tomorrow's stock price depends only on where the stock is today, not on where it was any time preceding today.

An example will make this clear. The weather is a case in point. Every day it can be either hot or cold. When it is hot the probability that the next day is hot is

70%, and when it is cold the chance that the next day is cold is 60%. As long as these probabilities depend only on whether it is hot or cold today, and not on the temperature yesterday or any day before that, the Markov property is preserved.

The probabilities may be represented in a Markov *transition matrix*. In our weather example, we have the following matrix:

$$\begin{bmatrix} 0.70 & 0.30 \\ 0.40 & 0.60 \end{bmatrix}$$

The first row of the matrix represents "hot" weather and the second row represents "cold." Likewise for the columns. Hence, each cell in the matrix represents movement from one state to another. The [1,1] cell represents the probability of the weather being "hot" today and remaining "hot" tomorrow. The [1,2] cell represents the probability of the weather being "hot" today and becoming "cold" tomorrow. And so on, you get the point.

How is such a matrix calculated? Using data of course. You just go back some period of time, say two years, and count how many hot days were followed by hot days, and divide that by the total number of days to get cell [1,1]. Do the same for all cells.

## 7.10 Multiple Period Chains

The matrix we just looked at was for one day intervals. Suppose we want to know the probability of it being hot two days from now if it is hot today (or any of the other cases). The transition matrix makes this really simple to do: just multiply the transition matrix by itself. Let's do this in R:

```
> tr_mat = t(matrix(c(0.7,0.3,0.4,0.6),2,2))
> tr_mat
     [,1] [,2]
[1,]  0.7  0.3
[2,]  0.4  0.6
> tr_mat %*% tr_mat
     [,1] [,2]
[1,] 0.61 0.39
[2,] 0.52 0.48
```

We see that there is a 61% probability that it will be hot two days from now given that it is hot today. This should make sense—there are two ways in which we can get to hot weather after two days: (a) hot, hot; (b) cold, hot. The sum of probabilities from these sequences is $0.61 = 0.70 \times 0.70 + 0.3 \times 0.4$.

Think of the two sequences as "chains" and the transition matrix gives us the probability of different chains. Hence, the term "Markov chain."

Note also that the transition probability matrix after multiplying by itself remains a probability matrix, i.e., the rows add up to one.

What if we did this ten times? Here is the exercise in R:

```
> mat = tr_mat
> for (j in 1:10) { mat = mat %*% tr_mat }
> mat
          [,1]      [,2]
[1,] 0.5714293 0.4285707
[2,] 0.5714276 0.4285724
```

You can see that the rows still add up to one. That is good (phew)! But interestingly, the rows have begun to look almost the same. Indeed if we did the multiplication 25 times, we would get identical rows:

```
> mat = tr_mat
> for (j in 1:25) { mat = mat %*% tr_mat }
> mat
          [,1]      [,2]
[1,] 0.5714286 0.4285714
[2,] 0.5714286 0.4285714
```

What does this mean?

1. The Markov chain is "ergodic"—it simply means that it is stable and has a steady state. You could end up in either state after many periods or one state. This is a standard property of transition probability matrices.

2. The probabilities tell you what proportion of time you will spend in hot and cold in the long run, i.e., 57% of the time the weather will be hot, and 43% of the time it will be cold. Thats a pretty neat thing to get from the simple probabilities in the transition matrix.

## 7.11 Credit Transition Matrices

In finance, a good example of Markov chains is the credit rating transition matrix. This is published by credit rating agencies such as Moodys.

For example, we input into R the transition matrix from Moodys for 1920–1996 (suitably modified to make sure that the row probabilities add up to 1.

```
> print(rtm,digits=2)
        [,1]    [,2]    [,3]   [,4]    [,5]    [,6]    [,7]    [,8]
[1,] 0.92279 0.06426 0.01034 0.0024 0.00021 0.00000 0.00000 0.00000
[2,] 0.01279 0.91683 0.06087 0.0070 0.00169 0.00021 0.00000 0.00063
[3,] 0.00074 0.02447 0.91585 0.0497 0.00670 0.00106 0.00021 0.00128
[4,] 0.00033 0.00260 0.04193 0.8941 0.05071 0.00661 0.00065 0.00303
[5,] 0.00011 0.00088 0.00430 0.0509 0.87230 0.05475 0.00453 0.01225
[6,] 0.00000 0.00045 0.00145 0.0067 0.06472 0.85323 0.03443 0.03901
[7,] 0.00000 0.00022 0.00044 0.0037 0.01381 0.05799 0.78779 0.13603
[8,] 0.00000 0.00000 0.00000 0.0000 0.00000 0.00000 0.00000 1.00000
```

Each row is for the following rating categories: Aaa, Aa, A, Baa, Ba, B, Caa-C, Default. Note that the last row has all zeros and a final value of 1, meaning that a firm that is in default will remain in default. Therefore the default state is known as an "absorbing" state.

Let's look at the steady state of the matrix:

```
> rtmmat = rtm
> for (j in 1:25) { rtmmat = rtmmat %*% rtm}
> print(rtmmat,digits=1)
      [,1]  [,2] [,3] [,4] [,5] [,6]  [,7] [,8]
[1,] 0.169 0.288 0.28 0.14 0.06 0.02 0.004 0.03
[2,] 0.057 0.231 0.31 0.19 0.10 0.04 0.007 0.06
[3,] 0.022 0.127 0.29 0.24 0.14 0.06 0.012 0.11
[4,] 0.011 0.067 0.20 0.23 0.17 0.09 0.019 0.21
[5,] 0.005 0.034 0.11 0.17 0.17 0.11 0.025 0.37
[6,] 0.002 0.017 0.06 0.11 0.13 0.09 0.024 0.57
[7,] 0.001 0.007 0.03 0.05 0.06 0.05 0.013 0.80
[8,] 0.000 0.000 0.00 0.00 0.00 0.00 0.000 1.00
> for (j in 1:1000) { rtmmat = rtmmat %*% rtm}
> print(rtmmat,digits=2)
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7] [,8]
[1,] 7.2e-08 3.2e-07 6.6e-07 6.3e-07 4.6e-07 2.4e-07 5.5e-08    1
[2,] 6.4e-08 2.8e-07 5.9e-07 5.6e-07 4.1e-07 2.2e-07 4.9e-08    1
[3,] 5.5e-08 2.4e-07 5.1e-07 4.9e-07 3.5e-07 1.9e-07 4.2e-08    1
[4,] 4.5e-08 2.0e-07 4.1e-07 3.9e-07 2.8e-07 1.5e-07 3.4e-08    1
[5,] 3.2e-08 1.4e-07 2.9e-07 2.8e-07 2.0e-07 1.1e-07 2.4e-08    1
[6,] 2.0e-08 9.0e-08 1.9e-07 1.8e-07 1.3e-07 6.9e-08 1.5e-08    1
```

```
[7,] 9.1e-09 4.0e-08 8.4e-08 8.0e-08 5.8e-08 3.1e-08 6.9e-09    1
[8,] 0.0e+00 0.0e+00 0.0e+00 0.0e+00 0.0e+00 0.0e+00 0.0e+00    1
```

We see that in the long run all firms will be in default. Not surprising! As the economist John Maynard Keynes said, "In the long run we're all dead."

## 7.12 Input-Output Matrices

Wassily Leonteif developed the famous input-output for economic planning. It manipulates matrices to determine how changes in output in one industry will impact others.

The product of one industry is consumed by both, other industries and final consumers. Industry 2 uses the output of Industry 1 as an input, and everything is linked. Thus, Industry 3 might use the output of Industry 2, and is therefore, indirectly using the output of Industry 1. And then, Industry 1 might use the output of Industry 3, and hence, is using indirectly its own output! Our economies are essentially incestuous! This whole mess needs to be resolved if any planning is to be done, and matrix modeling is an easy way to do this.

Suppose we have three sectors in the economy: agriculture, manufacturing, and services. The consumption of each others' outputs is given by the following matrix:

$$C = \begin{bmatrix} 0.5 & 0.2 & 0.5 \\ 0.1 & 0.1 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$$

Each row tells you how much output value of each industry is needed to produce one unit of value in the given industry in that row. The external consumer demand is given by the demand vector

$$D = \begin{bmatrix} 40 \\ 50 \\ 60 \end{bmatrix}$$

Using this information, we can derive how much production is going to be needed in the economy. The production of each industry is given as an element in the production vector

$$P = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

The equation relating these matrices is

$$P = D + C \cdot P$$

We can re-arrange this equation to solve for $P$,

$$P = (I - C)^{-1} \cdot D$$

where $I$ is the identity matrix. Let's implement this in R:

```
> C = matrix(c(5,1,2,2,1,5,5,2,1),3,3)/10
> C
     [,1] [,2] [,3]
[1,]  0.5  0.2  0.5
[2,]  0.1  0.1  0.2
[3,]  0.2  0.5  0.1
> D = matrix(c(40,50,60),3,1)
> D
     [,1]
[1,]   40
[2,]   50
[3,]   60
> P = solve(Iden - C) %*% D
> P
         [,1]
[1,] 370.5607
[2,] 148.1308
[3,] 231.3084
```

This tells us the equilibrium output for each industry and the economy as a whole.

# 8 Basic Bond Math

In this section we will revisit some basic ideas that you surely covered in your first finance class. But we'll try and delve a little more deeply into the connection between time, money and mathematics!

## 8.1 Exponentials

The fundamental constant $e = 2.718281828459045....$ is one of five in mathematics. You should commit the value of $e$ to memory to about a 100 digits.

Euler's formula connects all five fundamental constants in math $\{0, 1, e, i, \pi\}$ through the following equation:

$$e^{i\pi} + 1 = 0$$

where $i = \sqrt{-1}$.

Think of $e$ as a function. Its arguments are powers of $e$. The function is called the "exponential" function, and is written as $\exp(x)$. For example

```
> exp(1)
[1] 2.718282
> exp(2)
[1] 7.389056
> exp(exp(1))
[1] 15.15426
```

As $x$ increases $\exp(x)$ increases exponentially. Not surprising is it? It's called the exponential after all.

$e$ may also be expressed as the sum of the following infinite series:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} + \ldots$$

where $n! = 1 \times 2 \times 3 \times \ldots \times n$. This is also known as "factorial" of $n$.

An important property of the function is that it shows up in many physical ways. The following property is one embodiment.

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$$

We use this relation in compounding money continuously.

## 8.2 Compounding and Discounting

**Discrete compounding:** $\left(1 + \frac{r}{n}\right)^{nt}$

$n$ = periods per year,

$t$ = years,

$r$ = annual interest rate.

**Continuous compounding:**

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{nt} = \lim_{n\to\infty}\left[\left(1+\frac{r}{n}\right)^{n/r}\right]^{tr} = e^{rt}$$

```
 (1+0.06/365)^365
ans = 1.0618
 exp(0.06)
ans = 1.0618
 (1+0.06/12)^12
ans = 1.0617
```

For discounting, we use the relation

$$\frac{1}{e^{rt}} = e^{-rt}$$

## 8.3 The physical world and $e$

There are many places where $e$ appears, it seems to be naturally connected to the world in which we live in a physical, visceral way.

1. As we have seen, it is used in continuously compounding and discounting money.
2. The rate of decay of radioactive material is exponential. The mass $M$ of uranium-238 may be described by the following equation: $M(t) = M(0)e^{-\Omega t}$, where $\Omega$ is the exponential rate of decay. You can see that this is almost like the initial amount of uranium $M(0)$ is being "discounted" at rate $\Omega$.
3. The cooling rate of a heated object is also exponential. Hence, the temperature follows the equation: $T(t) = T(0)e^{-Ct}$, where $C$ is the cooling rate.
4. Population grows exponentially.
5. The intensity of a noise decays exponentially with distance from the source of the noise.

In the case of bonds, the present value of money at time $t$ is given by

$$P(t) = Pe^{-rt}$$

where $P$ is the principal on the bond, and rate of interest it $r$. Also note that

$$e^{r_1} \times e^{r_2} = e^{r_1+r_2}$$

This will come in handy in finance, when we discuss forward interest rates.

## 8.4 Calculus and $e$

The derivative of a function tells you the rate at which that function is changing at any point.

We will now see how $e$ is related to calculus. First of all, the exponential function is the only function whose derivative equals itself, i.e.,

$$\frac{d}{dx}e^x = e^x$$

Intuitively, if a function is growing exponentially fast, then its rate of growth is also exponential.

The "chain rule" in calculus also applies here and so

$$\frac{d}{dx}e^{f(x)} = e^{f(x)} \cdot f'(x)$$

If you have forgotten the chain rule, you may review it here.

## 8.5 Pricing discount bonds

- The price of a dollar $t$ years from today:

$$P = e^{-rt}.$$

- Duration:

$$\frac{dP}{dr} = -te^{-rt} = -tP.$$

- Convexity:

$$\frac{d^2P}{dr^2} = -t^2P.$$

- With discrete compounding and discounting, we have extra "adjustment" terms that are messy.

## 8.6 Logarithms

The *natural* logarithm as the inverse of the exponential function. So

$$\ln[\exp(x)] = x, \quad \text{or} \quad \ln e^x = x$$

Likewise, the exponential function is the inverse of the natural logarithm

$$\exp[\ln(x)] = x, \quad \text{or} \quad e^{\ln x} = x$$

Let's review the following properties of logarithms:

1. $\ln(x \cdot y) = \ln(x) + \ln(y)$
2. $\ln(x/y) = \ln(x) - \ln(y)$
3. Derivative: $\frac{d}{dx} \ln(x) = \frac{1}{x}$
4. Chain rule: $\frac{d}{dx} \ln[f(x)] = \frac{1}{f(x)} \cdot f'(x)$

Therefore, the continuous return between two days stock prices, $S_1$ and $S_2$ maybe written as

$$\ln(S_2/S_1) = \ln(S_2) - \ln(S_1)$$

In finance, we often just call these continuously-comounded returns "log-returns."

*Example*: Suppose we have a two-year maturity bond that pays \$100 at maturity, and has a current price (present value) of \$90. What is the continuously-compounded discount rate on this bond?

We may use logarithms to answer this. Note that the price of the bond is given by the following equation:

$$\text{Price} = 100e^{-rt}, \quad t = 2.$$

This may be written as

$$90 = 100e^{-2r}$$

and now we wish to solve for $r$, the discount rate. To do this, divide by 100 on both sides, and take logarithms of both sides to get

$$\begin{aligned}
\ln(90/100) &= \ln(e^{-2r}) \\
\ln(90/100) &= -2r \\
\frac{\ln(90/100)}{-2} &= r \\
0.05268026 &= r
\end{aligned}$$

## 8.7 Discounted Cash Flow for Projects

You have already studies discounted cash-flow (DCF) techniques in your first finance course. So let's do a quick review here. Suppose you have a project that generated

$100 in cash flow for each of ten years, and the discount rate on this project is 5%. What is the present value (PV) of this project? If the initial investment is $500 what is the net present value (NPV)?

The calculations in R are:

```
> PV = 0
> for (j in 1:10) { PV = PV + 100/(1.10)^j }
> PV
[1] 614.4567
> NPV = PV - 500
> NPV
[1] 114.4567
```

Another way to do this is as follows:

```
> CF = matrix(100,10,1)
> t = seq(1,10)
> df = 1/(1.1)^t
> df
 [1] 0.9090909 0.8264463 0.7513148 0.6830135 0.6209213 0.5644739 0.5131581 0.4665074
 [9] 0.4240976 0.3855433
> PV = t(CF) %*% df     # Or, use   PV = sum(CF*df)
> PV
         [,1]
[1,] 614.4567
> NPV = PV - 500
> NPV
         [,1]
[1,] 114.4567
```

We now look at the key concepts related to DCF:

1. Internal rate of return: The IRR is the rate at which the NPV of the cash flows is zero. Hence, the IRR $k$ is the rate at which

$$\sum_{j=1}^{N} \frac{c_j}{(1+k)^j} - A = 0$$

$A$ is the initial investment, $\tau$ is the number of cash flows, and $k$ is the per period discount rate.

2. Break even analysis: Break-even is defined by the amount of time it takes to recoup th initial investment. This is also known as the payback period (PP). It is the time $T$ such that
$$\sum_{t=1}^{T} c_t = A$$
where $c_t$ is the cashflow at time $t$.

3. Benefit-Cost Ratio: This is defined as:
$$BCR = \frac{\text{PV of cash inflows}}{\text{PV of cash outflows}}$$

4. Project comparison: How do you choose one project over another? Should you compare NVP, IRR, PP, or BCR?

5. Capital constraints: How does your decision rule change if there are capital constraints, i.e., you have a limited amount of capital to allocate to projects.

*Example*: We have a project with an investment of 100, and has the following annual cash flows: $\{20, 20, -10, 20, 30, 40, 50\}$. What is the NPV at a discount rate of 10%? Find the IRR, PP, and BCR of these cashflows.

Here is the R code to undertake these calculations:

```
> CF = c(20,20,-10,20,30,40,50)
> PV = 0
> for (j in 1:length(CF)) { PV = PV + CF[j]/(1.1^j) }
> NPV = PV - 100
> NPV
[1] 7.722368

> PV_inflows = CF[1]/1.1+CF[2]/1.1^2+CF[4]/1.1^4+CF[5]/1.1^5+CF[6]/1.1^6+CF[7]/1.1^7
> PV_outflows = CF[3]/1.1^3+100
> BCR = PV_inflows/PV_outflows
> BCR
[1] 1.245966

> NPV = function(CF,k) {
+ PV=0
+ for (j in 1:length(CF)) { PV = PV + CF[j]/(1+k)^j }
+ NPV = PV - 100
+ }
> print(NPV(CF,0.10))
```

```
[1] 7.722368

> library(minpack.lm)
> sol = nls.lm(par=0.10,fn=NPV,CF=CF)
> sol
Nonlinear regression via the Levenberg-Marquardt algorithm
parameter estimates: 0.118242110159837
residual sum-of-squares: 0
reason terminated: Relative error between ‘par’ and the solution is at most ‘ptol’.
```

The IRR = 11.8242%. The payback period is 6 years. More precisely, assuming smooth cash flows, it is 5.5 years.

# 9 Term Structure of Interest Rates

## 9.1 Yield-to-Maturity

The "yield curve" is the relation between maturity and yield-to-maturity (YTM) of bonds in any market.

Recall that the YTM is the IRR of a bond.

US government bonds are semi-annual pay bonds, and their YTMs are based on a semi-annual compounding convention. For example, if a two-year bond has a coupon rate of 6% and a price of \$100, then the relation between the cash flows, price and YTM is as follows:

$$100 = \frac{3}{(1+y/2)} + \frac{3}{(1+y/2)^2} + \frac{3}{(1+y/2)^3} + \frac{103}{(1+y/2)^4}$$

If you solve this equation for $y$, you will find it equal to 0.06, or 6%.

## 9.2 Pricing by Replication

Pricing by replication is a widely-used approach in finance. It is also known as "relative" pricing.

Let's say we are interested in the price of security A. If this security is not traded in the market, then we do not have a directly observable price. However, if we can construct a portfolio of other securities to "replicate" security A, and these other securities have available prices, then we are able to infer security A's price.

*Example*: Suppose we wish to price a semi-annual pay bond A that has a coupon rate of 7% and is maturing in one year. We are not given it's price or YTM. But we can see two other bonds trading, a six-month maturity bond B with coupon 4% that is trading at par, and a one-year maturity bond C with a coupon of 6% trading at a price of $101. Can we find the price of A? (Assume that all compounding and discounting is on a continuous basis.)

First, note that it is tempting to take the YTM of bond C and apply it to the pricing of bond A since they both have the same maturity. Let's do that and see what we get.

```
> bondprice = function(cashflows,ytm) {
+ n = length(cashflows)
+ maturities = seq(1,n)/2
+ df = matrix(exp(-ytm*maturities),n,1)
+ cf = matrix(cashflows,n,1)
+ price = sum(cf*df)
+ }
>
> library(minpack.lm)

This product includes software developed by the University of Chicago, as
Operator of Argonne National Laboratory.

See the LICENSE file distributed with the minpack.lm source code or
http://www.netlib.org/minpack/disclaimer for the full license.

> CF = c(3,103)
> solveytm = function(cashflows,ytm,price) {
+ result = bondprice(cashflows,ytm)-price
+ }
>
> sol = nls.lm(par=0.06,fn=solveytm,cashflows=CF,price=101)
> sol
Nonlinear regression via the Levenberg-Marquardt algorithm
parameter estimates: 0.0490205895173565
residual sum-of-squares: 0
> sol$par
[1] 0.04902059
> ytmC = sol$par
```

So the YTM of bond C is 0.04902059. Now if we use this YTM to price bond A, we get a price of \$101.9640.

```
> CFa = c(3.5,103.5)
> priceA1 = bondprice(CFa,ytmC)
> priceA1
[1] 101.9640
```

Now, let's price this bond by *replication*. To replicate bond A, we form a portfolio of bond B and C that results in identical cashflows to bond A. We choose $b$ units of bond B and $c$ units of bond C.

$$
\begin{aligned}
3.5 &= 102\,b + 3\,c \\
103.5 &= 0\,b + 103\,c
\end{aligned}
$$

We use matrix math to solve this system of equations:

```
> cfmat = matrix(c(102,0,3,103),2,2)
> cfmat
     [,1] [,2]
[1,]  102    3
[2,]    0  103
> targetcf = matrix(c(3.5,103.5),2,1)
> bc = solve(cfmat) %*% targetcf
> bc
            [,1]
[1,] 0.004759185
[2,] 1.004854369
```

So our solution is to replicate bond A using a portfolio comprising 0.00476 units of bond B and 1.00485 units of bond C. What is the price of this portfolio?

```
> pricebc = matrix(c(100,101),2,1)
> t(bc) %*% pricebc
         [,1]
[1,] 101.9662
```

Therefore, the price by replication is \$101.9662, whereas the price using the YTM of C was \$101.9640. Even though these prices look similar, they are not equal because the pricing by replication is correct and the pricing using bond C is not.

The important lesson here is that you should never use the YTM of another bond (in this case C) to price any other bond (in this case A).

How would you construct an arbitrage if bond A were actually trading at \$101.9640 instead of its correct price of \$101.9662?

## 9.3 Discount Function

The discount function is denoted $d(t)$, and is the present value of \$1 received at time $t$. We can write the pricing function of the previous section as:

$$100 = 3d(0.5) + 3d(1) + 3d(1.5) + 103d(2)$$

How do we find these "d" values? We will consider this shortly. For now, just remember the definition of a discount function. But for starters, suppose we have a bond maturing in six months and paying \$103. If the current price (present value) of the bond is \$99, then the discount function may be implied from the following equation:

$$99 = 103d(0.5)$$

i.e., $d(0.5) = 99/103 = 0.961165$.

## 9.4 Zero-Coupon Rate (ZCR)

A zero-coupon bond (ZCB) is a discount bond that has just one final cash flow. The bond we just considered that was expiring in six months is an example of a ZCB, because it had only one cash flow.

The ZCR is defined as the IRR of a ZCB.

For the bond in the previous example, we would find the ZCR, $z(t)$ as follows:

$$99 = \frac{103}{(1 + z(0.5)/2)}$$

implying that

$$z(0.5) = \left(\frac{103}{99} - 1\right) \times 2 = 0.080808$$

We note that there is a simple relation between ZCRs and discount functions:

$$d(t) = \frac{1}{(1 + z(t)/2)^{2t}}$$

where this rule works for the semi-annual compounding convention. Of course, for other compounding conventions, you will need to make suitable modifications. Suppose we are using the continuous-compounding convention. Then, the applicable relation would be:

$$d(t) = e^{-z(t)t}$$

So in the example we have been using, $0.961165 = e^{-0.5z(0.5)}$, which means that

$$z(0.5) = -\ln(0.961165)/0.5 = 0.07921838$$

This is a little lower than the semi-annual rate we derived earlier that came to 0.0808.

## 9.5 Forward Rate

The forward rate is the rate applicable to some forward period of time. Suppose we are given that $d(1) = 0.94$ and $d(2) = 0.86$. Then the forward rate between years 1 and 2 must be such that money discounted back from the end of year two has present value with the following equivalence:

$$d(2) = d(1) \times \frac{1}{(1 + f/2)^2}$$

where $f$ is the forward rate between year 1 and 2. Note that we are using the semi-annual convention here. Solving we get

$$f = 2 \times \left[ \sqrt{\frac{d(1)}{d(2)}} - 1 \right] = 2 \times (\sqrt{0.94/0.86} - 1) = 0.090955$$

The forward rate may be easily derived from any close pair of discount functions. In general

$$f(t, t + 0.5) = 2 \times \left[ \frac{d(t)}{d(t + 0.5)} - 1 \right]$$

Try and understand these equations, do not try and apply them blindly. You will surely make an error by using the convention wrongly, or enter incorrect inputs.

## 9.6 Bootstrapping Yield Curves

Pricing bonds is simply a matter of computing the present value of the bond's cash flows. This may be done either by using the IRR of the bond, i.e., its YTM, or by multiplying the cash flows of the bond by the discount functions in the market.

What is observable are the prices of the bonds in the market and the cash flows. We may use these to derive the discount functions. An example is always best. Consider the following bonds:

| Maturity ($t$) | Bond Price | Bond Coupon |
|---|---|---|
| 0.5 | 99.00 | 6.0% |
| 1.0 | 101.11 | 7.0% |
| 1.5 | 100.00 | 7.5% |
| 2.0 | 98.25 | 8.0% |

Can we infer the discount functions from this data?

Sure, here is where matrix math comes in handy. First, let's write down the system of equations we will use to price these bonds in terms of the discount functions:

$$\begin{aligned}
99.00 &= 103\, d(0.5) \\
101.11 &= 3.5\, d(0.5) + 103.5\, d(1) \\
100.00 &= 3.75\, d(0.5) + 3.75\, d(1) + 103.75\, d(1.5) \\
98.25 &= 4\, d(0.5) + 4\, d(1) + 4\, d(1.5) + 104\, d(2)
\end{aligned}$$

This is a system of four equations and four unknowns: $\{d(0.5), d(1), d(1.5), d(2)\}$. It may be written in matrix form as follows:

$$\begin{bmatrix} 103 & 0 & 0 & 0 \\ 3.5 & 103.5 & 0 & 0 \\ 3.75 & 3.75 & 103.75 & 0 \\ 4 & 4 & 4 & 104 \end{bmatrix} \begin{bmatrix} d(0.5) \\ d(1) \\ d(1.5) \\ d(2) \end{bmatrix} = \begin{bmatrix} 99.00 \\ 101.11 \\ 100.00 \\ 98.25 \end{bmatrix}$$

We solve this in R as follows:

```
> C = matrix(c(103,3.5,3.75,4,0,103.5,3.75,4,0,0,103.75,4,0,0,0,104),4,4)
> C
       [,1]   [,2]   [,3] [,4]
[1,] 103.00   0.00   0.00    0
[2,]   3.50 103.50   0.00    0
[3,]   3.75   3.75 103.75    0
[4,]   4.00   4.00   4.00  104
> P = matrix(c(99,101.11,100,98.25),4,1)
> P
       [,1]
[1,]  99.00
[2,] 101.11
[3,] 100.00
[4,]  98.25
> d = solve(C) %*% P
> d
          [,1]
[1,] 0.9611650
[2,] 0.9444050
[3,] 0.8949794
[4,] 0.8369981
```

Note that we implemented the solution as $d = C^{-1}P$.

## 9.7 Duration and Bond Portfolio Risk

Risk in bonds comes from changes in interest rates, i.e. yields ($y$) or zero-coupon rates ($z$).

Price sensitivity is measured by the ratio of the change in price to the change in the zero-coupon rate, i.e.

$$\frac{\Delta P}{\Delta z}.$$

Therefore, if the price of $t$-period ZCB, semi-annual basis is

$$P = \frac{c}{[1+\frac{z}{2}]^t}$$

then, its price sensitivity is

$$\begin{aligned}\frac{dP}{dz} &= -\frac{t}{2} \times \frac{c}{[1+\frac{z}{2}]^t} \times \frac{1}{[1+\frac{z}{2}]} \\ &= -\frac{t}{2} \times P \times \frac{1}{1+\frac{z}{2}}\end{aligned}$$

which shows that price sensitivity is a function of the maturity of the bond, times its price, times an adjustment term.

The adjustment term comes from the fact that we have a semi-annual basis, and vanishes when we use continuous compounding.

With continuous compounding, we have

$$\begin{aligned}P &= e^{-zt} \\ \frac{dP}{dz} &= -t \times e^{-zt} \\ &= -t \times P\end{aligned}$$

which is a really simple expression!

Percentage price sensitivity is called "<u>modified duration</u>" ($D^*$), which is simply

$$D^* = -\frac{dP}{dz}\frac{1}{P} = \frac{t}{2} \times \frac{1}{1+\frac{z}{2}}$$

In continuous time

$$D^* = -\frac{dP}{dz}\frac{1}{P} = t$$

Note that $D^* > 0$ since bond prices have an inverse relationship to interest rates.

How would you represent duration graphically on the price-yield graph?

Modified duration is related to its cousin, Macaulay Duration ($D$), which was developed by Frederick Macaulay in 1938.

On a semi-annual basis

$$D = D^*(1 + \frac{z}{2}).$$

Duration is additive, hence for a bond with many cashflows

$$D = \frac{1}{P}\left(\sum_{t=1}^{T}\frac{t}{2}\times P_t\right)$$

where $P = \sum_{t=1}^{T} P_t$.

We find that this is just the same as

$$\begin{aligned} D &= \sum_{t=1}^{T}\left(\frac{t}{2}\times\frac{P_t}{P}\right) \\ &= \sum_{t=1}^{T}\left(\frac{t}{2}\times w_t\right) \end{aligned}$$

where $w_t$ is the weight of the $t^{th}$ cashflow in the total present value of the bond. Note that $\sum_{t=1}^{T} w_t = 1$.

Hence, there are 3 ways to compute price sensitivity:

1. Using Macaulay duration.
2. Analytically by the differentiating the price function.
3. Numerically by perturbation.

Example of Macaulay computation: FaceValue=100, Coupon=6%.

| Period ($t$) | Maturity ($t/2$) | Cashflow | ZCR | PV | PV $\times t/2$ | D=PV/totPV $\times t/2$ | $D^*$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.50 | 6.00 | 0.10 | 5.71 | 2.86 | 0.03 | 0.03 |
| 2 | 1.00 | 6.00 | 0.11 | 5.39 | 5.39 | 0.05 | 0.05 |
| 3 | 1.50 | 106.00 | 0.12 | 89.00 | 133.50 | 1.33 | 1.26 |
| TOTAL | | | | 100.10 | 141.75 | 1.42 | 1.34 |

Since

$$D^* = -\frac{dP}{dz}\frac{1}{P} = 1.34$$

the percentage change in price is given by

$$\frac{dP}{P} = -D^* \times dz = -1.34 \times dz.$$

So, if the change in zero-coupon rate is $dz = 0.01$, i.e. $+1\%$, then the percentage change in the bond price is

$$-1.34 \times 0.01 = -0.0134$$

i.e. negative 1.34%.

Lets check this in a different way, by numerical perturbation. In the table above, if we change each ZCR by increasing it by 0.001, then we get a new price for the bond which is 99.971, compared to the earlier price of 100.10. We can compute modified duration as follows

$$D^* = -\frac{\Delta P}{\Delta z}\frac{1}{P} = -\frac{99.971 - 100.10}{0.001 \times 100.10} = 1.34.$$

which checks the previous table.

*Generalizations*

- Inverse price to interest rate relationship.

$$\frac{dP}{dz} < 0.$$

- Price sensitivity increases with maturity.

$$\frac{dD^*}{dt} > 0.$$

- For the same maturity, price sensitivity declines with yield

$$\frac{dD^*}{dz} < 0$$

  Explan this graphically.

- Price sensitivity declines with increasing coupon rate.

$$\frac{dD^*}{dc} < 0$$

  Why?

*Convexity*

- Convexity is the rate of change of duration, i.e.

$$\gamma = \frac{d^2 P}{dz^2}$$

- In continuous time,

$$\gamma = \frac{d^2}{dz^2}[e^{-rt}] = t^2 P$$

  Prove this.

- In practice, convexity is normalized to be

$$\gamma' = \frac{1}{P}\gamma = t^2 > 0$$

- Show graphically that $\gamma > 0$. What implications does this have for the riskiness of bonds?

*Concept checks*:

1. Do corporate bonds or Treasury bonds have greater price sensitivity to changes in interest rates?
2. Is duration additive?
3. What is bond immunization?

## 9.8 Smoothing the Yield Curve using the Nelson-Siegel (NS) technique

This technique is based on the well-known paper:
Nelson, C.R., Siegel, A.F. (1987). Parsimonious modeling of yield curves, Journal of Business, 60(4), pp.473489.

A smooth function for fitting the yield curve is given by

$$y(t) = a + be^{-t/d} + c(t/d)e^{-t/d}$$

The parameters to be fitted are $\{a, b, c, d\}$.

*Example*: Suppose we are given the following maturities and yields. These are semi-annual for 20 years, i.e., 40 maturities and yields.

```
> t
 [1]  0.5  1.0  1.5  2.0  2.5  3.0  3.5  4.0  4.5  5.0  5.5  6.0  6.5  7.0
[15]  7.5  8.0  8.5  9.0  9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.0
[29] 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0 18.5 19.0 19.5 20.0

> y
 [1] 5.971905 6.664383 7.045291 7.272764 7.508568 7.683340 7.871456 7.979930
 [9] 8.098855 8.212605 8.297714 8.401490 8.477987 8.538548 8.625413 8.686280
[17] 8.742366 8.801806 8.851922 8.903571 8.954586 8.997979 9.051650 9.085412
[25] 9.130450 9.172797 9.205978 9.238384 9.280887 9.311143 9.347869 9.377006
[33] 9.407104 9.444717 9.473754 9.498263 9.522459 9.550073 9.574802 9.600665
```

We wish to fit the NS model to these yields. This will require a *non-linear* regression model.

```
> fitNS = function(params,t,y) {
+  a = params$a
+  b = params$b
+  c = params$c
+  d = params$d
+  yfitted = a + b*exp(-t/d) + c*(t/d)*exp(-t/d)
+  ydiff = yfitted - y
+ }
>
> t
 [1]  0.5  1.0  1.5  2.0  2.5  3.0  3.5  4.0  4.5  5.0  5.5  6.0  6.5  7.0
[15]  7.5  8.0  8.5  9.0  9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.0
[29] 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0 18.5 19.0 19.5 20.0
> y
 [1] 5.944214 6.668064 7.025993 7.252042 7.508365 7.708054 7.853181 7.981529
 [9] 8.119595 8.201952 8.312952 8.391955 8.472183 8.555156 8.617115 8.686458
[17] 8.743343 8.795703 8.850934 8.905995 8.954981 9.000728 9.048014 9.090464
[25] 9.128925 9.174937 9.204415 9.245608 9.280807 9.317517 9.345071 9.377303
[33] 9.411774 9.439694 9.469307 9.499912 9.521797 9.548254 9.577485 9.599345
>
> library(minpack.lm)
>
> parstart = list(a=2,b=1,c=1,d=2)
> sol = nls.lm(par=parstart,fn=fitNS,t=t,y=y)
> sol$par
$a
[1] 9.552814
```

```
$b
[1] -3.484138
$c
[1] 0.0001682015
$d
[1] 5.441153
```

It is worth taking a quick look at the fit of the model:

```
> a = sol$par$a
> a
[1] 9.552814
> b = sol$par$b
> c = sol$par$c
> d = sol$par$d
> yfitted = a + b*exp(-t/d)+c*(t/d)*exp(-t/d)
> yfitted
 [1] 6.374586 6.653636 6.908186 7.140386 7.352198 7.545414 7.721665 7.882441
 [9] 8.029101 8.162884 8.284920 8.396242 8.497790 8.590422 8.674920 8.752000
[17] 8.822312 8.886450 8.944958 8.998328 9.047012 9.091422 9.131932 9.168886
[25] 9.202595 9.233344 9.261394 9.286981 9.310321 9.331612 9.351034 9.368750
[33] 9.384911 9.399653 9.413101 9.425367 9.436557 9.446765 9.456076 9.464570
> diff = yfitted - y
> diff
 [1]  0.430371759 -0.014428424 -0.117806873 -0.111656026 -0.156166918
 [6] -0.162639720 -0.131516586 -0.099088450 -0.090493881 -0.039067946
[11] -0.028031195  0.004287669  0.025606483  0.035265586  0.057805826
[16]  0.065541656  0.078968718  0.090747786  0.094023217  0.092332933
[21]  0.092030984  0.090693132  0.083918400  0.078421572  0.073669865
[26]  0.058407734  0.056979432  0.041372573  0.029513882  0.014095240
[31]  0.005962852 -0.008553378 -0.026863284 -0.040040981 -0.056206116
[36] -0.074544875 -0.085239513 -0.101489203 -0.121409584 -0.134775033
```

The NS model is widely used. Central bankers use it a lot to smooth and model yield curves. The NS approach may also be used for modeling/fitting zero-coupon rate curves as well as forward rate curves.

# 10 Bayes' Rule

Bayes Rule is as follows:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A).Pr(A)}{Pr(B)}.$$

For $n$ possible outcomes of $A_i, i = 1...n$, where $A_i$ are disjoint, we have

$$Pr(A_i|B) = \frac{Pr(B|A_i).Pr(A_i)}{\sum_{i=1}^{n} Pr(B|A_i).Pr(A_i)}.$$

Here is a simple example of the application of Bayes' Rule.

Bayesian calculations often lead to surprising results. For example, when considering the risk of contracting HIV, we are interested in what it means to get a positive test result. Here are some brief calculations in R:

```
> prHIV = 0.003
> prTP_HIV = 0.9
> prTP_NoHIV = 0.20
> prNoHIV = 1-prHIV
> prTP = prTP_HIV*prHIV + prTP_NoHIV*prNoHIV
> prTP
[1] 0.2021
> prHIV_TP = prTP_HIV*prHIV/prTP
> prHIV_TP
[1] 0.01335972
```

So we can see that even though the probability of testing positive for HIV is high, the probability that one has HIV given a positive test is low. It is always important to condition correctly and Bayes' theorem does a great job in analyzing conditional likelihoods.

## 10.1 The NYSE

The New York Stock Exchange is one of the beacons of global capitalism. You should peruse their web site for the details of how it operates. In class we will consider the following features of the NYSE.

1. History (call auction). Here is a trading post.

2. How does the NYSE work (order routing, specialist, single market-maker, trading floor, superdot, ITS, etc)

3. What is a "good" market? [info efficiency, liquidity, fairness, technology, critical mass, manage performance risk]

4. Definitions: depth, liquidity, price efficiency, transparency, spread.

5. Mechanisms for efficiency: listing, ITS, order flow concentration, specialist, bid-ask (2 way quotes), floor trading (vs electronic), electronic ledgers.

6. Limit orders vs market orders

7. What determines bid-ask spreads?

8. Role of the market maker (sigma, adverse selection, info) Do we necessarily need a market maker?

9. Effect of tick sizes: [increase sigma, reduction in info flow, economical volume, limits or increases bid-ask spreads, effect on stock size - splits, liquidity, easier accounting]

10. Nasdaq tick crisis.

11. Difference between NYSE & NASDAQ: [n vs 1 market makers, no specialist, different type of stocks].

12. Comparison of exchanges: [volume, info efficiency, liquidity, fairness (manipulation), order concentration, txn costs].

13. Why not trade 24 hours??

The NYSE has huge market share. A paper by Marshall Blume and Michael Goldstein presents data on how many times NYSE has the best quote.

In 1994, William Christie and Paul Schultz uncovered collusion amongst market makers on the NASDAQ. They presented their evidence in two famous papers. The first, reported the collusion. The second, reported the aftermath. The Economist also reported on the incident.

## 10.2 Application to specialist's Bayesian updating of bid-ask spreads

1. We start by assuming that the only thing that impacts the price setting behavior of the specialist is information.

2. There are two types of trader: informed ($I$) and liquidity or noise type of trader ($N$). The specialist worries that he may be trading against a trader of type $I$. This is called the "adverse selection" problem. We assume an equal probability of trading against $I$ or $N$.

3. Assume the specialist is trading an asset of value $A$. This asset (for simplicity) may take a high value of $H = 1$ or a low value of $L = 0$. This is probabilistic, and we get $H$ with probability $p$, and $L$ with $(1 - p)$.

4. The specialist can set his expected value of the asset once he knows $p$, but he does not know this. He can only infer it from the trading behavior of others, and then synthesizes this information.

5. At the outset he must start with a Bayesian prior, which we set to be $p = Pr(A = H) = Pr(A = L) = 1/2$.

6. What the specialist sees is the order flow, which is simply a series of buy ($B$) or sell ($S$) trades. After each trade, depending on whether it is a buy or sell, he then is able to update his probability $p$, hence resulting in a new expected value of the asset, which in this case is $pH + (1 - p)L = p$.

7. We can work out the following *conditional* probabilities:

$$Pr(S|A = L) = Pr(S|I)Pr(I) + Pr(S|N)Pr(N) = 1.\frac{1}{2} + \frac{1}{2}.\frac{1}{2} = \frac{3}{4}.$$

   and

$$Pr(S|A = H) = Pr(S|I)Pr(I) + Pr(S|N)Pr(N) = 0.\frac{1}{2} + \frac{1}{2}.\frac{1}{2} = \frac{1}{4}.$$

8. Now lets see what happens as order flow arrives. Suppose the specialist experiences a sale transaction. He will compute the conditional probability that $A = L$ given that $S$ occurs:

$$\begin{aligned}
Pr(A = L|S) &= \frac{Pr(S|A = L)Pr(A = L)}{Pr(S|A = L)Pr(A = L) + Pr(S|A = H)Pr(A = H)} \\
&= \frac{3/4 \times 1/2}{3/4 \times 1/2 + 1/4 \times 1/2} \\
&= \frac{3}{4} \\
Pr(A = H|S) &= 1 - Pr(A = L|S) \\
&= 1 - \frac{3}{4} = \frac{1}{4}.
\end{aligned}$$

9. These values become the new priors for when the next trade arrives. Notice that the probability of the asset being of low value $L$ is now increased as a sale has occured.

10. Suppose instead, the first trade was a buy. Then, the Bayes updating is as follows:

$$\begin{aligned} Pr(A=L|B) &= \frac{Pr(B|A=L)Pr(A=L)}{Pr(B|A=L)Pr(A=L)+Pr(B|A=H)Pr(A=H)} \\ &= \frac{1/4 \times 1/2}{1/4 \times 1/2 + 3/4 \times 1/2} \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned} Pr(A=H|B) &= 1 - Pr(A=L|B) \\ &= 1 - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

These are now the new priors.

11. Suppose the first buy was followed by another buy. Bayesian updating continues with the new priors.

$$\begin{aligned} Pr(A=L|BB) &= \frac{Pr(B|A=L)Pr(A=L|B)}{Pr(B|A=L)Pr(A=L|B)+Pr(B|A=H)Pr(A=H|B)} \\ &= \frac{1/4 \times 1/4}{1/4 \times 1/4 + 3/4 \times 3/4} \\ &= \frac{1}{10} \end{aligned}$$

$$\begin{aligned} Pr(A=H|BB) &= 1 - Pr(A=L|BB) \\ &= 1 - \frac{1}{10} = \frac{9}{10}. \end{aligned}$$

12. We can express the above situation in somewhat more extensible form:

$$\begin{aligned} Pr(A=L|BB) &= \frac{Pr(B|A=L)Pr(A=L|B)}{Pr(B|A=L)Pr(A=L|B)+Pr(B|A=H)Pr(A=H|B)} \\ &= \frac{Pr(B|A=L)^2 Pr(A=L)}{Pr(B|A=L)^2 Pr(A=L)+Pr(B|A=H)^2 Pr(A=H)} \\ &= \frac{(1/4)^2 \times 1/2}{(1/4)^2 \times 1/2 + (3/4)^2 \times 1/2} \\ &= \frac{1}{10} \end{aligned}$$

$$
\begin{aligned}
Pr(A = H|BB) &= 1 - Pr(A = L|BB) \\
&= 1 - \frac{1}{10} = \frac{9}{10}.
\end{aligned}
$$

13. In general,

$$
\begin{aligned}
& Pr(A = L|\#B = m, \#S = n) \\
= \; & \frac{Pr(B|A = L)^m Pr(S|A = L)^n Pr(A = L)}{Pr(B|A = L)^m Pr(S|A = L)^n Pr(A = L) + Pr(B|A = H)^m Pr(S|A = H)^n Pr(A = H)}
\end{aligned}
$$

# 11 Revisiting Basic Calculus

In this section we will review some very basic calculus and try to obtain some feel for the use of these tools in finance.

## 11.1 Differentials

We list some basic formulae for derivatives of standard functions:

1. For $y = x^n$, $y' = \frac{dy}{dx} = nx^{n-1}$. For example, if $y = 4x^3$, then $y' = 12x^2$.
2. For $y = g[f(x)]$, $\frac{dy}{dx} = \frac{dg}{df} \cdot \frac{df}{dx}$. This is known as the "chain rule." For example, if $P = \frac{c}{(1+r/2)^t}$, then we have that $f(r) = (1 + r/2)$, and $g[f] = f^{-t}$, so that

$$
P' = \frac{dg}{df} \cdot \frac{df}{dy} = -tf^{-t-1} \times 1/2 = (-t/2)(1+r/2)^{-t}(1+r/2)^{-1} = (-t/2)P(1+r/2)^{-1}
$$

3. For $y = e^x$, $y' = e^x$. $e^x$ is the only function that has itself as its first derivative. By the chain rule, if $y = e^{f(x)}$, then $y' = e^{f(x)} \frac{dy}{dx}$. For example, if $y = e^{x^3}$, then $y' = e^{x^3}(3x^2)$.
4. For $y = \ln(x)$, $y' = \frac{1}{x}$. And likewise, for $y = \ln[f(x)]$, we have $y' = \frac{1}{f(x)} f'(x)$. For example, if $y = \ln(x^3)$, then $y' = \frac{1}{x^3}(3x^2)$.

A derivative may be approximated by taking small changes. In fact, a derivative $dy/dx$ may be loosely interpreted as "the change in $y$ for a very small change in $x$." To get a feel for this, let's take the last case we looked at above, i.e., $y = \ln(x^3)$ for which we the derivative is $y' = \frac{1}{x^3}(3x^2)$.

If we evaluate this derivative at $x = 2$, we get

$$y' = \frac{1}{x^3}(3x^2) = \frac{1}{2^3}(3 \times 2^2) = 12/8 = 1.5$$

We may also approximate this derivative by computing the function at $x = 2$ and then at $x = 2.0001$ (a small change in $x$), and then looking at the change in $y$ divided by the change in $x$.

When $x = 2$, $y = \ln(8) = 2.079442$. When $x = 2.0001$, $y = 2.079592$. Now, computing the change in $y$ divided by the change in $x$ is

$$\frac{\Delta y}{\Delta x} = \frac{2.079592 - 2.079442}{2.0001 - 2} = 1.499963$$

which is very close to the true derivative, which amounted to 1.5.

What happens to this derivative when $x = 5$? Let's calculate this one more time using R:

```
> (1/5^3)*(3*5^2)
[1] 0.6
> (log(5.0001^3)-log(5^3))/(0.0001)
[1] 0.599994
```

Again, we see how the derivatives, computed exactly and using a small change in $x$, resulted in the same values.

## 11.2 Integrals

Integrals are known as "anti-derivatives", i.e., the inverse of differentiating a function. Integrals are really just sums over regions. An integral may be approximated by a sum, and the following equation expresses it:

$$\int_a^b f(x)\, dx \approx \sum_{x=a}^{b} f(x)\, \Delta x$$

Such an integral, with a range of values from $a$ to $b$, is known as an "definite" integral. When no range is given, we denote it an "indefinite" integral, meaning that there is no defined interval.

An example helps. Since we already learnt that if $y = \ln(x)$, then $y' = \frac{1}{x}$, we know that the integral or anti-derivative of $1/x$ must be $\ln(x)$. Suppose we look at the definite integral $\int_2^3 \frac{1}{x}\ dx$. This is calculated as follows (note the way in which the math is written):

$$\int_2^3 \frac{1}{x}\ dx = [\ln(x)]_2^3 = \ln(3) - \ln(2) = 0.4054651$$

This is the same as writing the integral in approximate form where the interval of integration is broken down into small intervals of size $\Delta x = 0.01$:

$$\sum_{x=\{2,2.01,2.02,\ldots,2.99,3\}} \frac{1}{x}\Delta x = \sum_{x=\{2,2.01,2.02,\ldots,2.99,3\}} \frac{1}{x}(0.01) = 0.4096329$$

And we see that this is very close to the integral computed using the formula. In R we would have computed it both ways as follows:

```
> oneoverx = function(x) { result = 1/x }
> integrate(oneoverx,2,3)
0.4054651 with absolute error < 4.5e-15
> x = seq(2,3,0.01)
> sum(1/x*0.01)
[1] 0.4096329
```

We list just a few useful integrals here:

1. $\int x^n\ dx = \frac{1}{n+1}x^{n+1}$
2. $\int e^x\ dx = e^x$
3. $\int \frac{1}{x}\ dx = \ln(x)$

## 11.3 Symbolic Math using Mathematical Packages

It is useful to use Maxima for symbolic math and calculus. This system is free and is very easy to use. The most common commands are presented at the following reference page. You may download Maxima for windows and linux here.

The most widely-used commercial package for symbolic math is `Mathematica`, and it is obtained from Wolfram Research. If you like using software to do symbolic math, I strongly urge you to buy the student version of this product because it costs much more when you are not a student. There are several other commercial packages that are extremely good as well such as Maple.

A wonderful integrated web-based math system is freely available and does both numerical and symbolic math: It is called Sage.

## 11.4 Differential Equations: Simple Concepts

Differential equations are equations in which the variables of interest are expressed in changes. These are dynamic representations, and we often need to solve them to find the value of a function at some point in its dynamic evolution.

For example, examine the following representation of the evolution of the price of a zero-coupon bond.

$$dP(t) = rP(t)\ dt, \quad P(0) = 1$$

What does this mean. The LHS side represents the change in bond price $P(t)$, where $t$ represents the remaining time to maturity of the bond. The RHS says that this change is equal to the rate of interest multiplied by the current price for the small change in time $dt$. We are interested in solving this equation to find the function that represents $P(t)$, i.e., the price of the bond with $t$ years remaining until its maturity. The statement $P(0) = 1$ is the "boundary condition" that says that the bond pays \$1 at maturity. It is apparent that this is the model for a discount function.

How do we solve this differential equation to get the function $P(t)$?

We first re-arrange the differential equation as follows:

$$\frac{1}{P}\ dP = -r\ dt$$

This is known as "separation of variables"—we put $P$ on the LHS and we put $dt$ on the RHS, noting that $r$ is just a constant. Next we integrate both sides:

$$\int \frac{1}{P}\ dP = -r \int 1\ dt$$

We already know that the integral of $1/x$ is $\ln(x)$, and the integral of 1 is just the variable. So we get

$$\ln(P) = -rt + c$$

What's $c$? It is known as the "constant of integration" and is required because we need to assume that there may be a constant that dropped out when we took the differential. We will be able to determine the constant using the boundary condition. Solving further, we have:

$$P = e^{-rt}e^{c} = Ae^{-rt}, \quad A = e^{c}$$

This is the solution for the function $P(t)$. Since we know that when $t = 0$, $P = 1$, we put $t = 0$ in the solution above, and we get

$$P(0) = 1 = Ae^{-r\times 0} = A \quad \rightarrow \quad A = 1$$

which means that

$$P(t) = e^{-rt}$$

This is the solution to the differential equation. Does it look familiar?

Suppose $r = 0.05$ and $t = 3$. What is the price $P(3)$? We get $P(3) = e^{-0.05\times 3} = 0.860708$.

Now, lets solve this differential equation numerically. To do this, we choose $\Delta t = 1/12$, i.e., each time step is one month. We also discretize the differential equation as follows:

$$\Delta P = -rP\ \Delta t = -rP(1/12)$$

We may write this as

$$P(t + \Delta t) = P(t) - rP(t)(1/12)$$

We start at $t = 0$, when $P(0) = 1$. We then iterate backwards in time as follows:

$$\begin{aligned}
P(1/12) &= P(0) - 0.05P(0)(1/12) = 0.9958333 \\
P(2/12) &= P(1/12) - 0.05P(1/12)(1/12) = 0.9958333 - 0.05(0.9958333)(1/12) = 0.991684 \\
P(3/12) &= P(2/12) - 0.05P(2/12)(1/12) = 0.991684 - 0.05(0.991684)(1/12) = 0.987552 \\
&\vdots \\
P(3) &= P(35/12) - 0.05P(35/12)(1/12) = 0.8604383
\end{aligned}$$

We can do this quickly in R as follows:

```
> P=1
> for (j in 1:36) { P = P - 0.05*P*(1/12) }
> P
[1] 0.8604383
```

The solution is very close to the one we got from the closed-form solution.

# 12 Hypothesis Testing

Testing a hypothesis requires choosing a distribution that governs the statistical test. The most common ones we use are the normal and Poisson distributions.

## 12.1 Normal Distribution

If $x \sim N(\mu, \sigma^2)$, then

$$\text{density function:} f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$N(x) = 1 - N(-x)$$

$$F(x) = \int_{-\infty}^{x} f(u)du$$

Standard normal distirbution: $x \sim N(0, 1)$.

For the standard normal distribution: $F(0) = \frac{1}{2}$.

```
> x=rnorm(5)
> pnorm(x)
[1] 0.46288875 0.18025760 0.52793932 0.27681658 0.03142125
> dnorm(x)
[1] 0.39721491 0.26263569 0.39796354 0.33475284 0.07070038
> qnorm(seq(0.1,0.9,0.2))
[1] -1.2815516 -0.5244005  0.0000000  0.5244005  1.2815516
> qnorm(c(0.01,0.05,0.10))
[1] -2.326348 -1.644854 -1.281552
> qnorm(c(0.01,0.05,0.10)/2)
[1] -2.575829 -1.959964 -1.644854
```

## 12.2 Poisson Distribution

The rare-event distribution.

$$f(n; \lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$$

Mean = Variance = $\lambda$.

```
> x=rpois(100000,1)
> mean(x)
[1] 0.99802
> var(x)
[1] 0.999306
```

## 12.3 Null Hypothesis

A null hypothesis is an assertion of truth subject to statistical testing. It is inherently necessary that the hypothesis be falsifiable, else it is not useful.

For some years now, zoos have been using orangutans to predict the winner of the Super Bowl. Kutai, the youngest of the Oregon zoos orangutans has picked the winner correctly in 4 of the past 5 games. Is Kutai really good at predicting the Super Bowl winner? How would you set up the hypothesis to test this? What is the null hypothesis?

There are two broad choices for the null hypothesis:

(a) $H_0$: Kutai can predict superbowl winners.

(b) $H_0$: Kutai cannot predict winners and just picks winners randomly.

The null hypothesis is better set up with (b). We then try to falsify or reject the null, and if we do, then we can say with reasonable statistical confidence that Kutai has real skill. (a) is known as the alternate hypothesis.

One of the reasons (b) is easier is because it suggests the probabilities to be used in the test automatically. Suppose Kutai is just picking winners at random. Then the chance of picking a winner is 1/2. The probability of picking 4 of 5 winners is

$$\binom{5}{4}(0.5^4)(0.5^1) = 0.15625$$

The probability of picking all 5 correctly is

$$\binom{5}{5}(0.5^5)(0.5^0) = 0.03125$$

Hence, the probability of choosing 4 or more correctly is 0.1875.

If we want to be 95% sure that Kutai can pick 4 or more winners in 5 years then the null hypothesis that Kutai picks winners at random is not rejected because to be 95% sure, the probability of picking 4 or more winners should be less than 5%. It is actually 18.75%, so clearly, we cannot reject the null that Kutai is picking winners at random.

What if Kutai had picked all 5 winners in 5 years? The probability of doing this is only 3.125%, and we could be 96.875% ($> 95\%$) sure that this is not random.

What does this tell you about money managers?

## 12.4 Type I and Type II Errors

No test of a null hypothesis is perfect. It is prone to error. There are two types of errors we consider:

1. Type I error, or false positive, or $\alpha$ error. Reject the null when the null is true. For example, if we rejected the null that orangutans cannot predict the Super Bowl, when in fact they cannot. Note that $1 - \alpha$ is the same as the level of confidence we choose for our test. If we want to test so as to be 95% sure, then $\alpha = 5\%$. A test with high $\alpha$ has poor "specificity."

2. Type II error, or false negative, or $\beta$ error. This arises when we fail to reject the null when in fact it is not valid. For instance, if orangutans can really predict the Super Bowl, but we do not reject the null that they pick winners at random, then we are committing Type II error. A test with high Type II error is said to have poor "sensitivity." The $\beta$ is the probability that we failed to reject the null when it was indeed false.

In our criminal justice system, the null hypothesis is that we are innocent (until proven guilty). If the jury rejects the null, then they return a guilty verdict.

Type I error: the person is innocent but they find him guilty (false positive). False positive actually means a false rejection of the null. This is confusing, because the null fails falsely, yet we call is a positive.

Type II error: the person is guilty, but they find him innocent (false negative).

Which type of error do you want to minimize? Which type of error does the prosecutor prefer? Which type of error does the accused prefer? See this page for a very nice exposition of these errors in regard to the criminal justice system.

## 12.5 Power of a Test

The power or sensitivity of a test is $1 - \beta$. The lower the false negative rate of the test, the greater its power. In the case of the orangutans, a false negative is that we fail to reject the null when it is not true, i.e., orangutans did have ability to predict the Bowl game, but we did not reject the null that they had no ability.

Here is a terrific applet that relates Type I and Type II error to power of the test and the sample size.

We will use these ideas, plus matrix math, and calculus in computing Value-at-Risk (VaR).

# 13 Value-at-Risk (VaR)

What is Tail Risk?

1. Assume a portfolio which is currently worth \$100 and which will be worth \$$100 + X$ in one year's time, where $X \sim N(5, 10^2)$.
2. Estimate risk in terms of tail outcomes, i.e., how much can we lose and with what probability?
3. What is the probability $p$ that the one-year return on the portfolio will be less than some dollar amount $m$? For example, we may want to know the probability that the returns over the year will be less than $-10$, i.e., that the portfolio will lose at least 10% of its current value of 100.
4. The answer is 6.68%.

Definition of VaR:

1. Value-at-Risk poses this problem the other way around.
2. It fixes a probability $p$ and asks: what is the amount $m$ such that the likelihood of an outcome worse than $m$ is no more than $p$?
3. This number $m$ is called the $(1-p)$-VaR, because with probability at least $1-p$, the outcome will be better than $m$.
4. For instance, if $p$ is taken to be 1% (a typical value), $m$ is called the 99%-VaR.
5. VaR is usually reported with a positive sign even though it indicates a loss amount.

Example:

1. The 99%-VaR is the number $m$ such that the probability of an observation $x \leq m$ from a $N(5, 10^2)$-distribution is 0.01.
2. Since our attention is on the *left* tail of the portfolio, we compute the quantity
$$\text{mean} - (2.33 \times \text{standard deviation}).$$
This is the required number $m$. For the numbers in the example, the 99%-VaR works out to
$$m \;=\; 5 - (2.33 \times 10) \;=\; -18.30.$$
3. Three components to VaR: a probability $p$ (typically 1% or 5%), a dollar amount $m$, and a horizon $h$ (one day to one year).

Table 1: VaR and Returns Distributions: Example 1

| | Distribution 1 | Distribution 2 |
|---|---|---|
| Outcome | Probability | Probability |
| $-10$ | 0.01 | 0.01 |
| 0 | 0.90 | 0.09 |
| $+10$ | 0.09 | 0.90 |

Table 2: VaR and Returns Distributions: Example 2

| | Distribution 1 | Distribution 2 |
|---|---|---|
| Outcome | Probability | Probability |
| $-50$ | 0.025 | 0.000 |
| $-10$ | 0.035 | 0.060 |
| $+10$ | 0.940 | 0.940 |

## 13.1 Capital Adequacy

Capital Adequacy:

1. VaR provides a summary picture of *capital adequacy*.
2. Example: suppose a firm has equity capital of $10 million and has a 95% one-year VaR of $12 million. This implies that there is a 5% chance of the firm losing more money over the coming year than it holds in equity capital. If the firm wishes to reduce the probability of ruin, there are essentially only two options.
   - Hold more capital (say, increase the capital to $15 million), or
   - Lower the tail risk in its portfolio by altering the portfolio's composition.

## 13.2 Limitations

Limitations - I: VaR does not care about the shape of returns outside the tail.

Limitations - II: VaR does not care about the shape of losses in the tail.

## 13.3 Methods

Methods of Calculating VaR:

1. The Delta-Normal method.
2. Historical simulation.
3. Monte Carlo simulation.

Delta-Normal Method:

1. $n$ assets in the portfolio, indexed by $i = 1, \ldots, n$.
2. $w_i$ denote the dollar investment in asset $i$;
3. $\mu_i$ be the expected return on asset $i$; and
4. $\sigma_{ij}$ be the covariance of returns between assets $i$ and $j$.
5. Expected dollar return on the portfolio:
$$\mu_P = \sum_{i=1}^{n} \mu_i w_i,$$
6. Variance of dollar returns on the portfolio:
$$\sigma_P^2 = \sum_{i,j=1}^{n} w_i w_j \sigma_{ij}.$$
7. Suppose that we wish to calculate the 99% VaR of the portfolio. This is the quantity $-m$, where $m$ is defined by
$$m = \mu_P - (2.33 \times \sigma_P).$$

Example:

1. Two asset case (joint normal returns):
$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0.20 \\ 0.12 \end{bmatrix}, \qquad \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 0.04 & 0.02 \\ 0.02 & 0.03 \end{bmatrix}$$
2. Portfolio – investment of $w_1 = 5$ and $w_2 = 5$.

3. Expected return and variance of returns on the portfolio:

$$\mu_P \;=\; 1.60 \qquad \sigma_P^2 \;= 2.75.$$

4. The 95% VaR is

$$\mu_P - (1.645 \times \sigma_P) \;=\; 1.60 - (1.645 \times 1.658) \;=\; -1.128.$$

## 13.4 Risk Decomposition

Linear Homogeneity:

$$f(mx_1, \ldots, mx_k) \;=\; m \times f(x_1, \ldots, x_k).$$

For linear homogenous functions, we have that

$$f(x) \;=\; \left[x_1 \times \frac{\partial f(x)}{\partial x_1}\right] + \cdots + \left[x_n \times \frac{\partial f(x)}{\partial x_n}\right]$$

where, as usual, the term $\partial f(x)/\partial x_i$ refers to the partial derivative of the function $f(x)$ with respect to $x_i$.

VaR Decomposition: The VaR of a portfolio is linearly homogeneous in the portfolio weight vector $w$:

$$\mathrm{VaR}(m \cdot w) \;=\; m \times \mathrm{VaR}(w)$$

If we double all positions in the portfolio, dollar value in the 1% left tail of the distribution also doubles.

The standard deviation: $\sigma_P(m \cdot w) \;=\; m \times \sigma_P(w)$.

$$\mathrm{VaR}(w) \;=\; \left[w_1 \times \frac{\partial \mathrm{VaR}(w)}{\partial w_1}\right] + \cdots + \left[w_n \times \frac{\partial \mathrm{VaR}(w)}{\partial w_n}\right]$$

$$\%\ \text{Risk Contribution of } i \;=\; \frac{1}{\mathrm{VaR}(w)} \left[\frac{\partial \mathrm{VaR}(w)}{\partial w_i}\, w_i\right]$$

Example:

1. 99%-VaR of earlier portfolio is 2.264.

2. Portfolio value is normally distributed with a mean of $\mu_P(w)$ and a standard deviation of $\sigma_P(w)$, where

$$\mu_P(w) \;=\; \mu_1 w_1 + \mu_2 w_2$$

$$\sigma_P(w) \;=\; \left[\sigma_1^2 w_1^2 + \sigma_2^2 w_2^2 + 2\sigma_{12} w_1 w_2\right]^{1/2}$$

3. By definition, the 99%-VaR is given by

$$\mathrm{VaR}(w) = -\left[\mu_P(w) - 2.33 \times \sigma_P(w)\right].$$

Decomposition:

1. Differentiating $\mu_P$ and $\sigma_P$ with respect to $w_1$ and $w_2$, we obtain:

$$\frac{\partial \mu_P(w)}{\partial w_1} = \mu_1; \frac{\partial \sigma_P(w)}{\partial w_1} \;=\; \frac{1}{\sigma_P(w)}\left[w_1\sigma_1^2 + w_2\sigma_{12}\right]$$

$$\frac{\partial \mu_P(w)}{\partial w_2} = \mu_2; \frac{\partial \sigma_P(w)}{\partial w_2} \;=\; \frac{1}{\sigma_P(w)}\left[w_1\sigma_{12} + w_2\sigma_2^2\right].$$

2. By definition of the VaR, we have

$$\frac{\partial \mathrm{VaR}(w)}{\partial w_i} \;=\; -\left(\frac{\partial \mu_P(w)}{\partial w_i} - 2.33\,\frac{\partial \sigma_P(w)}{\partial w_i}\right).$$

3. Expanding this, we obtain the final expressions for the risk-contributions:

$$w_1\,\frac{\partial \mathrm{VaR}(w)}{\partial w_1} \;=\; w_1 \times \left[-\mu_1 + 2.33 \times \frac{1}{\sigma_P(w)}\left(w_1\sigma_1^2 + w_2\sigma_{12}\right)\right]$$

$$w_2\,\frac{\partial \mathrm{VaR}(w)}{\partial w_2} \;=\; w_2 \times \left[-\mu_2 + 2.33 \times \frac{1}{\sigma_P(w)}\left(w_1\sigma_{12} + w_2\sigma_2^2\right)\right]$$

4. Risk contributions are 1.1076 and 1.1563. Total = 2.264.

## 13.5 Hedging

Risk Decomposition and Hedging:

1. Let the initial dollar allocations be $w = (w_1, \ldots, w_n)$. Pick any $i$. Suppose we change the investment in asset $i$ from $w_i$ to $w_i(1 + \Delta_i)$, where $\Delta_i$ could be positive or negative. From the risk-decomposition, portfolio risk changes by approximately
$$\left[\frac{\partial \mathrm{VaR}(w)}{\partial w_i}\, w_i\right] \times \Delta_i.$$

2. If we wish to reduce portfolio risk by an amount $A$, the proportional change $\Delta_i$ that is required is
$$\Delta_i \;=\; A \Big/ \left[\frac{\partial \mathrm{VaR}(w)}{\partial w_i}\, w_i\right]$$

## 13.6 Optimization

Risk Decomposition and Portfolio Optimization:

1. We want the return of each asset to be commensurate to its risk-contribution.
$$\frac{\mu_i w_i}{\text{Risk-contribution of } i} \;=\; \frac{\mu_j w_j}{\text{Risk-contribution of } j}$$

2. If the left-hand side is greater than the right-hand side, then we reallocate a small amount from $j$ to $i$ and improve portfolio performance; if the left-hand side is smaller, then moving resources from $i$ to $j$ improves matters.

## 13.7 Coherence

Coherent Risk Measures: Let $\mathcal{R}$ denote a generic risk-measure. Given a portfolio $w$, we denote the risk of $w$ under the measure $\mathcal{R}$ by $\mathcal{R}(w)$. Four conditions that $\mathcal{R}$ should meet:

1. *Linear homogeneity*: If all positions are scaled by a factor of $m$, then the risk-measure should also scale by $m$: $\mathcal{R}(m \cdot w) = m\,\mathcal{R}(w)$ for any $m > 0$.

2. *Monotonicity*: Any portfolio that "dominates" another should result in a lower risk measure.

3. *Sub-Additivity*: Diversification should reduce risk.

$$\mathcal{R}(P1 + P2) \;\leq\; \mathcal{R}(P1) + \mathcal{R}(P2)$$

4. *Translation Invariance*: If we add a risk-free asset to the portfolio with an expected return of $r$, the risk of the portfolio should come down by the extent of this risk-free addition.

Failure of Sub-Additivity–Example: 100 corporate bonds: invest at 100, get back 102. Probability of default = 0.01, recovery rate on default = 0. We want the 95%-VaR.

(a) Borrow $1,000,000. Invest in a single bond. The VaR is $-\$20,000$.

(b) Invest $10,000 in 100 bonds instead. Prob of $k$ defaults is

$$\pi(k) = \frac{100!}{k!(100-k)!}(0.01)^k(0.99)^{100-k}$$

$\pi(k \leq 2) = 0.92$, $\pi(k \leq 3) = 0.98$. Hence the VaR is at $k \leq 2$.

Portfolio value for 2 defaults is:

$$10,200(100-k) - 1000,000 = -400$$

Hence VaR is 400.

## 13.8 Related Measures

Related Risk Measures:

1. *Worst-Case Scenario Analysis*: if we let $r_1, r_2, \ldots, r_N$ denote the returns on the portfolio over $N$ trading periods, then WCS looks at the distribution of $\min\{r_1, \ldots, r_N\}$.

2. *Expected Shortfall*: Known to be a coherent measure of risk, it is the one factor that has led to its popularity.

$$ES = \frac{\int_{-\infty}^{VaR} r\; f(r)\; dr}{\int_{-\infty}^{VaR} f(r)\; dr}$$

   where $f(r)$ is the probability density

# 14 CoVaR

What is CoVaR?

1. $CoVaR$ is a complementary measure that looks at the impact of a single firm on system-wide (i.e., systemic) risk. It was developed by Tobias Adrian and Markus Brunnermeier in the aftermath of the 2008 financial crisis. Hence $CoVaR(j)$ is the $VaR$ of the entire financial system conditional on institution $j$ being in financial distress.

2. $\Delta CoVaR$ is the difference between $CoVaR$ and the unconditional system-wide $VaR$. A bank that has a low $VaR$ but high $\Delta CoVaR$ is far more risky to the system than a bank with high $VaR$ but low $\Delta CoVaR$.

Three Measures:

1. $VaR$: Given by the implicit level $\alpha$ where
$$Pr[V_i \leq VaR_\alpha(i)] = \alpha$$

2. $CoVaR(i|j)$ is given implicitly by the quantile $\alpha$ such that
$$Pr[V_i \leq CoVaR_\alpha(i|j)|V_j = VaR_\alpha(j)] = \alpha$$
Hence, we are interested in the effect of $j$'s distress on the $VaR$ of $i$.

3. And then, $\Delta CoVaR(i|j)$ is given as the difference of these two risk measures, i.e.,
$$\Delta CoVaR_\alpha(i|j) = CoVaR_\alpha(i|j) - VaR_\alpha(i)$$

Properties of CoVaR:

1. *Cloning*: If we split a large financial institution $j$ into identical smaller ones $j'$, then the measure remains unchanged, i.e.,
$$CoVaR(\cdot|j) = CoVaR(\cdot|j')$$

2. *Causality*: The measure is agnostic as to the mechanics driving its value. It could be a causal relation or it could be driven by a factor common to all firms.

3. *Relative Risk Conditioning*: Since the conditioning is on $VaR_\alpha(j)$, i.e., a probability level ($\alpha$), the measure is independent of the risk strategy of $j$. If an absolute loss level were used instead of an $\alpha$ level, then the risk strategy of $j$ would matter.

More Properties of CoVaR:

1. *Directional*: That is, it is not symmetric.

$$CoVaR_\alpha(i|j) \neq CoVaR_\alpha(j|i)$$

2. *Endogeneity*: Risk-taking by any FI impacts the $CoVaR$ of the other FIs. By herding on the same type of trading strategy, all the FIs together increase the $CoVaR$s in the system, as they should. Hence, the measure picks up this endogenous connection between all FIs.

3. *Extensible*: The concept of $CoVaR$ is easily applied to other risk measures than $VaR$, such as expected shortfall or $ES$. Hence, we may think of measures such as $CoES$.

Median Regression:

- Median Regression: Say we have $n$ observations of $y$ (dependent variable) To find the median $M$ we solve the following optimization problem:

$$\min_M \sum_{i=1}^{n} |y_i - M|$$

- If $y$ may be conditioned on explanatory variables $x$, then we have the median regression as follows:

$$\min_b \sum_{i=1}^{n} |y_i - bx_i|, \quad \text{Median}(y|x_i) = bx_i$$

Quantile Regression:

- More generally,

$$\begin{aligned} \min_b & \quad \sum_{i=1}^{n} \rho(0.5)(y_i - bx_i) \\ \rho(\tau) &= \begin{cases} \tau & \text{if } y_i - bx_i > 0 \\ 1-\tau & \text{if } y_i - bx_i < 0 \end{cases} \end{aligned}$$

  Here $\tau = 0.5$ because we are interested in the median of $y$, conditional on $x$. We use the same equations to estimate any quantile we may like. For example, we might want the first quartile and in that case, $\tau = 0.25$.

- How do we use quantile regression to estimate $CoVaR(\cdot|j)$? Simple–run a quantile regression of the $VaR$ of the entire financial system on $VaR(j)$.

# 15 Extending the Regression Framework

See the manual `R-intro.pdf` on the R web site. Work through Appendix A, at least the first page. Also see Grant Farnsworth's document "Econometrics in R".

Issuing system commands:

```
system("<command>")
```

## 15.1 Loading in a file and doing a regression

Here is a simple regression run on some data from the 2005-06 NCAA basketball season for the March madness stats. We use the metric of performance to be the number of games played, and then see what variables explain it best. The data file is at ncaa.txt.

```
> ncaa = read.table("ncaa.txt",header=TRUE)
> y = ncaa[3]
> y = as.matrix(y)
> x = ncaa[4:14]
> x = as.matrix(x)
> fm = lm(y~x)
> summary(fm)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5075 -0.5527 -0.2454  0.6705  2.2344

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.194804   2.892203  -3.525 0.000893 ***
xPTS         -0.010442   0.025276  -0.413 0.681218
xREB          0.105048   0.036951   2.843 0.006375 **
xAST         -0.060798   0.091102  -0.667 0.507492
xTO          -0.034545   0.071393  -0.484 0.630513
xA.T          1.325402   1.110184   1.194 0.237951
xSTL          0.181015   0.068999   2.623 0.011397 *
```

```
xBLK          0.007185   0.075054   0.096 0.924106
xPF          -0.031705   0.044469  -0.713 0.479050
xFG          13.823190   3.981191   3.472 0.001048 **
xFT           2.694716   1.118595   2.409 0.019573 *
xX3P          2.526831   1.754038   1.441 0.155698
---
Signif. codes:  0  0.001  0.01  0.05  0.1  1

Residual standard error: 0.9619 on 52 degrees of freedom
Multiple R-Squared: 0.5418,     Adjusted R-squared: 0.4448
F-statistic: 5.589 on 11 and 52 DF,  p-value: 7.889e-06
```

An alternative approach using data frames is:

```
> ncaa_data_frame = data.frame(y=as.matrix(ncaa[3]),x=as.matrix(ncaa[4:14]))
> fm = lm(y~x,data=ncaa_data_frame)
> summary(fm)
```

Direct regression implementing the matrix form is as follows (we will derive this at some point later):

```
> wuns = matrix(1,64,1)
> z = cbind(wuns,x)
> b = inv(t(z) %*% z) %*% (t(z) %*% y)
> b
              GMS
    -10.194803524
PTS  -0.010441929
REB   0.105047705
AST  -0.060798192
TO   -0.034544881
A.T   1.325402061
STL   0.181014759
BLK   0.007184622
PF   -0.031705212
FG   13.823189660
FT    2.694716234
X3P   2.526830872
```

Note that this is exactly the same result as we had before, but it gave us a chance to look at some of the commands needed to work with matrices in R.

## 15.2 Heteroskedasticity

Simple linear regression assumes that the standard error of the residuals is the same for all observations. Many regressions suffer from the failure of this condition. The word for this is "heteroskedastic" errors. "Hetero" means different, and "skedastic" means dependent on type.

We can first test for the presence of heteroskedasticity using a standard Breusch-Pagan test available in R. This resides in the `lmtest` package which is loaded in before running the test.

```
> ncaa = read.table("ncaa.txt",header=TRUE)
> y = as.matrix(ncaa[3])
> x = as.matrix(ncaa[4:14])
> result = lm(y~x)
> library(lmtest)
Loading required package: zoo
> bptest(result)

        studentized Breusch-Pagan test

data:  result
BP = 15.5378, df = 11, p-value = 0.1592
```

We can see that there is very little evidence of heteroskedasticity in the standard errors as the p-value is not small. However, lets go ahead and correct the t-statistics for heteroskedasticity as follows, using the `hccm` function.

```
> wuns = matrix(1,64,1)
> z = cbind(wuns,x)
> b = inv(t(z) %*% z) %*% (t(z) %*% y)
> result = lm(y~x)
> library(car)
> vb = hccm(result)
> stdb = sqrt(diag(vb))
> tstats = b/stdb
> tstats
           GMS
   -2.68006069
PTS -0.38212818
REB  2.38342637
```

```
AST -0.40848721
TO  -0.28709450
A.T  0.65632053
STL  2.13627108
BLK  0.09548606
PF  -0.68036944
FG   3.52193532
FT   2.35677255
X3P  1.23897636
```

Compare these to the t-statistics in the original model

```
summary(result)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.194804   2.892203  -3.525 0.000893 ***
xPTS         -0.010442   0.025276  -0.413 0.681218
xREB          0.105048   0.036951   2.843 0.006375 **
xAST         -0.060798   0.091102  -0.667 0.507492
xTO          -0.034545   0.071393  -0.484 0.630513
xA.T          1.325402   1.110184   1.194 0.237951
xSTL          0.181015   0.068999   2.623 0.011397 *
xBLK          0.007185   0.075054   0.096 0.924106
xPF          -0.031705   0.044469  -0.713 0.479050
xFG          13.823190   3.981191   3.472 0.001048 **
xFT           2.694716   1.118595   2.409 0.019573 *
xX3P          2.526831   1.754038   1.441 0.155698
```

## 15.3 Auto-regressive models

Lets load in the Markowitz data and run tests on it.

```
> md = read.table("markowitz_data.txt",header=TRUE)
> y = as.matrix(md[2])
> x = as.matrix(md[7:9])
> rf = as.matrix(md[10])
> y = y-rf
> library(car)
> results = lm(y ~ x)
> durbin.watson(results,max.lag=6)
```

```
 lag Autocorrelation D-W Statistic p-value
   1     -0.07231926      2.144549   0.002
   2     -0.04595240      2.079356   0.146
   3      0.02958136      1.926791   0.162
   4     -0.01608143      2.017980   0.632
   5     -0.02360625      2.032176   0.432
   6     -0.01874952      2.021745   0.536
 Alternative hypothesis: rho[lag] != 0
```

Hence, there is one lag auto-correlation, but not more than that; markets are very efficient. Lets look at the regression:

```
> summary(results)

Call:
lm(formula = y ~ x)

Residuals:
       Min         1Q     Median         3Q        Max
-0.2136760 -0.0143564 -0.0007332  0.0144619  0.1910892

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.000197   0.000785  -0.251   0.8019
xmktrf       1.657968   0.085816  19.320   <2e-16 ***
xsmb         0.299735   0.146973   2.039   0.0416 *
xhml        -1.544633   0.176049  -8.774   <2e-16 ***
---
Signif. codes:  0  0.001  0.01  0.05  0.1  1

Residual standard error: 0.03028 on 1503 degrees of freedom
Multiple R-Squared: 0.3636,     Adjusted R-squared: 0.3623
F-statistic: 286.3 on 3 and 1503 DF,  p-value: < 2.2e-16
```

Lets correct the t-stats for autocorrelation using the Newey-West correction.

```
> res = lm(y~x)
> b = res$coefficients
> b
  (Intercept)        xmktrf           xsmb          xhml
-0.0001970164  1.6579682191  0.2997353765 -1.5446330690
```

```
> vb = NeweyWest(res,lag=1)
> stdb = sqrt(diag(vb))
> tstats = b/stdb
> tstats
(Intercept)      xmktrf        xsmb        xhml
 -0.2633665  15.5779184   1.8300340  -6.1036120
```

Compare these to the stats we had earlier. Notice how they have come down after correction for AR.

For fun, lets look at the autocorrelation in stock market indexes in this table.

## 15.4 Vector Auto-Regression

Also known as VAR (not the same thing as Value-at-Risk, denoted VaR). VAR is useful for estimating systems where there are simultaneous regression equations, and the variables influence each other.

For vector autoregressions (VARs), we do the following:

```
> md = read.table("markowitz_data.txt",header=TRUE)
> y = as.matrix(md[2:4])
> library(stats)
> var6 = ar(y,aic=TRUE,order=6)
> var6$order
[1] 1
> var6$ar
, , SUNW

         SUNW       MSFT          IBM
1 -0.00985635 0.02224093 0.002072782

, , MSFT

         SUNW       MSFT       IBM
1 0.008658304 -0.1369503 0.0306552

, , IBM

         SUNW      MSFT          IBM
1 -0.04517035 0.0975497 -0.01283037
```

```
> var6$aic
        0         1         2         3         4         5         6
23.950676  0.000000  2.762663  5.284709  5.164238 10.065300  8.924513
> var6$partialacf
, , SUNW

         SUNW         MSFT          IBM
1 -0.00985635  0.022240931  0.002072782
2 -0.07857841 -0.019721982 -0.006210487
3  0.03382375  0.003658121  0.032990758
4  0.02259522  0.030023132  0.020925226
5 -0.03944162 -0.030654949 -0.012384084
6 -0.03109748 -0.021612632 -0.003164879

, , MSFT

          SUNW        MSFT          IBM
1  0.008658304 -0.13695027  0.030655201
2 -0.053224374 -0.02396291 -0.047058278
3  0.080632420  0.03720952 -0.004353203
4 -0.038171317 -0.07573402 -0.004913021
5  0.002727220  0.05886752  0.050568308
6  0.242148823  0.03534206  0.062799122

, , IBM

         SUNW         MSFT         IBM
1 -0.04517035  0.097549700 -0.01283037
2  0.05436993  0.021189756  0.05430338
3 -0.08990973 -0.077140955 -0.03979962
4  0.06651063  0.056250866  0.05200459
5  0.03117548 -0.056192843 -0.06080490
6 -0.13131366 -0.003776726 -0.01502191
```

We can also run

```
> ar(y,method="ols",order=3)

Call:
ar(x = y, order.max = 3, method = "ols")

$ar
```

```
, , 1

        SUNW       MSFT       IBM
SUNW 0.01407 -0.0006952 -0.036839
MSFT 0.02693 -0.1440645  0.100557
IBM  0.01330  0.0211160 -0.009662

, , 2

          SUNW     MSFT     IBM
SUNW -0.082017 -0.04079 0.04812
MSFT -0.020668 -0.01722 0.01761
IBM  -0.006717 -0.04790 0.05537

, , 3

         SUNW      MSFT      IBM
SUNW 0.035412  0.081961 -0.09139
MSFT 0.003999  0.037252 -0.07719
IBM  0.033571 -0.003906 -0.04031


$x.intercept
      SUNW       MSFT       IBM
-9.623e-05 -7.366e-05 -6.257e-05

$var.pred
          SUNW      MSFT       IBM
SUNW 0.0013593 0.0003007 0.0002842
MSFT 0.0003007 0.0003511 0.0001888
IBM  0.0002842 0.0001888 0.0002881
```

For fun, let's look at the "cross"-autocorrelation in stock market data in this table.

## 15.5 ARCH and GARCH

Stands for "Generalized Auto- Regressive Conditional Heteroskedasticity" (!). Rob Engle invented ARCH (for which he just got the Nobel prize) and this was extended by Tim Bollerslev to GARCH.

ARCH models are based on the idea that volatility tends to cluster, i.e. volatility for period $t$, is auto-correlated with volatility from period $(t-1)$ or even more preceding periods. So if we had a time series of stock returns following a random walk, we would model it as follows

$$r_t = \mu + e_t, \quad e_t \sim N(0, \sigma_t^2)$$

If the variance were stationary then $\sigma_t^2$ would just be constant $\sigma^2$. But under ARCH it is auto-correlated with previous variances. Hence, we have

$$\sigma_t^2 = \sigma_0 + \sum_{k=1}^{p} e_{t-k}^2 + \sum_{k=1}^{q} \sigma_{t-k}^2$$

So current variance ($\sigma_t^2$) depends on past squared shocks ($e_t^2$) and past variances. The number of lags of past shocks is $p$, and that of lagged variances is $q$. The model is thus known as a $GARCH(p,q)$ model.

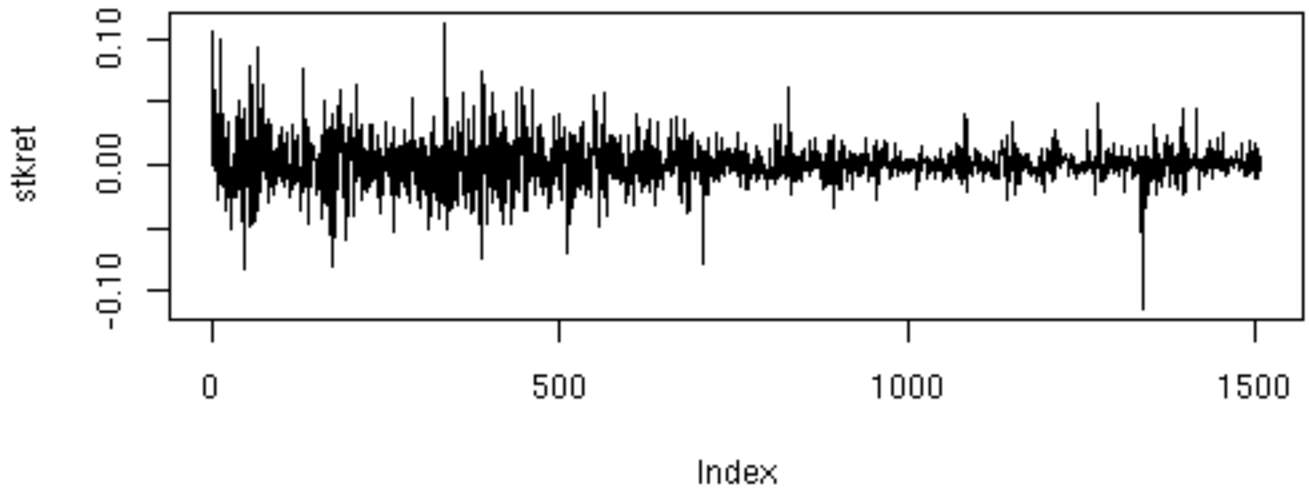
We will implement the model for stock returns using the data set on tech stocks.

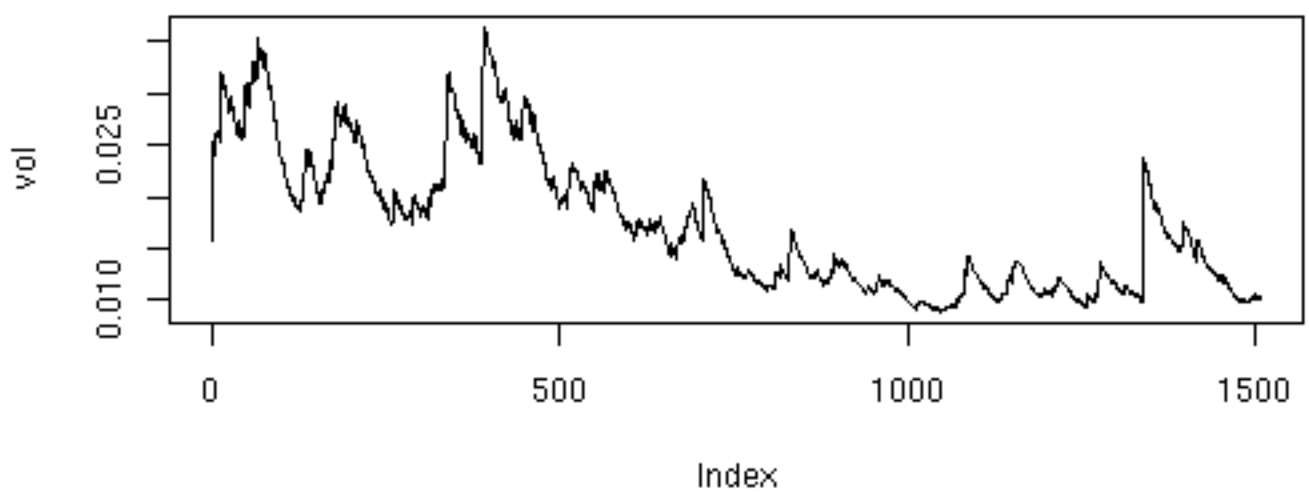
```
> md = read.table("markowitz_data.txt", header=TRUE)
> stkret = as.matrix(md[3])
> library(tseries)
Loading required package: quadprog
Loading required package: zoo
> res = garch(stkret,order=c(1,1))
> summary(res)

Call:
garch(x = stkret, order = c(1, 1))

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max
-11.4469  -0.5479   0.0000   0.5550   6.6422

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 2.428e-06   4.775e-07    5.085 3.68e-07 ***
a1 3.572e-02   5.111e-03    6.988 2.79e-12 ***
b1 9.549e-01   6.336e-03  150.721  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 11201.63, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 0.0042, df = 1, p-value = 0.9482
> res$coef
          a0           a1           b1
2.427728e-06 3.571891e-02 9.549465e-01
> sum(res$coef)
[1] 0.9906678
> vols = fitted(res)
> vol = vols[,1]
> plot(vol,type="l")
```

We can plot the stock series and the volatility series as follows:

```
> op = par(no.readonly=TRUE)
> par(mfrow=c(2,1))
> plot(stkret,type="l")
> plot(vol,type="l")
> par(op)
```




## 15.6 Maximum Likelihood Estimator

Also known as MLE. Suppose we have a sequence of $n$ observations $X$ drawn from the same probability distribution $F$ (with density $f$) which has a mean $\mu$ and variance $\sigma^2$. The likelihood of seeing these observations is given by

$$L = f(x_1) \times f(x_2) \times \ldots \times f(x_n)$$

We want to find $\mu$, $\sigma$ that maximizes this likelihood, i.e. the joint probability of seeing all the observations.

Since $f(x)$ is small, the product $L$ will be tiny. To avoid numerical problems we maximize the log-likelihood, i.e.

$$\max_{\mu,\sigma^2} \ln(L) = \max \sum_{i=0}^{n} \ln[f(x_i)]$$

Read in the data:

```
> md = read.table("markowitz_data.txt", header=TRUE)
> x = as.matrix(md[2])
> mean(x)
[1] -0.0003669222
> var(x)
            SUNW
SUNW 0.001438065
```

Create the likelihood function based on the normal distribution:

```
NormLL = function(p,x) {
        mu = p[1]
        sig2 = p[2]
        logf = -log(sqrt(2*pi*sig2)) - (x-mu)^2/(2*sig2)
        NormLL = -sum(logf)  #negate for minimizer
}
```

*Exercise* How do we get this function above?

Now load this function, and undertake MLE:

```
> p=c(0.01,0.01)
> source("NormLL.R")
> res=nlm(NormLL,p,x=x)
There were 50 or more warnings (use warnings() to see the first 50)
> res
$minimum
[1] -2793.408

$estimate
```

```
[1] -0.0003715541  0.0014368872

$gradient
[1]  -4.333487 100.815519

$code
[1] 2

$iterations
[1] 10
```

## 15.7 Logit

For the NCAA data, take the top 32 teams and make their dependent variable 1, and that of the bottom 32 teams zero.

```
> y1 = 1:32
> y1 = y1*0+1
> y1
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
> y2 = y1*0
> y2
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
> y = c(y1,y2)
> y
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
[39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

The running the model is pretty easy as follows:

```
> h = glm(y~x, family=binomial(link="logit"))
> logLik(h)
'log Lik.' -21.44779 (df=12)
> summary(h)

Call:
glm(formula = y ~ x, family = binomial(link = "logit"))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
```

```
-1.80174  -0.40502  -0.00238   0.37584   2.31767

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -45.83315   14.97564  -3.061  0.00221 **
xPTS         -0.06127    0.09549  -0.642  0.52108
xREB          0.49037    0.18089   2.711  0.00671 **
xAST          0.16422    0.26804   0.613  0.54010
xTO          -0.38405    0.23434  -1.639  0.10124
xA.T          1.56351    3.17091   0.493  0.62196
xSTL          0.78360    0.32605   2.403  0.01625 *
xBLK          0.07867    0.23482   0.335  0.73761
xPF           0.02602    0.13644   0.191  0.84874
xFG          46.21374   17.33685   2.666  0.00768 **
xFT          10.72992    4.47729   2.397  0.01655 *
xX3P          5.41985    5.77966   0.938  0.34838
---
Signif. codes:  0  0.001  0.01  0.05  0.1  1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 88.723  on 63  degrees of freedom
Residual deviance: 42.896  on 52  degrees of freedom
AIC: 66.896

Number of Fisher Scoring iterations: 6
```

Suppose we ran this just with linear regression:

```
> h = lm(y~x)
> summary(h)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-0.65982 -0.26830  0.03183  0.24712  0.83049

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.114185   1.174308  -3.503 0.000953 ***
```

```
xPTS        -0.005569   0.010263  -0.543 0.589709
xREB         0.046922   0.015003   3.128 0.002886 **
xAST         0.015391   0.036990   0.416 0.679055
xTO         -0.046479   0.028988  -1.603 0.114905
xA.T         0.103216   0.450763   0.229 0.819782
xSTL         0.063309   0.028015   2.260 0.028050 *
xBLK         0.023088   0.030474   0.758 0.452082
xPF          0.011492   0.018056   0.636 0.527253
xFG          4.842722   1.616465   2.996 0.004186 **
xFT          1.162177   0.454178   2.559 0.013452 *
xX3P         0.476283   0.712184   0.669 0.506604
---
Signif. codes:  0  0.001  0.01  0.05  0.1  1

Residual standard error: 0.3905 on 52 degrees of freedom
Multiple R-Squared: 0.5043,     Adjusted R-squared: 0.3995
F-statistic:  4.81 on 11 and 52 DF,  p-value: 4.514e-05
```

## 15.8 Probit

We can redo the same using a probit instead:

```
> h = glm(y~x, family=binomial(link="probit"))
> logLik(h)
'log Lik.' -21.27924 (df=12)
> summary(h)

Call:
glm(formula = y ~ x, family = binomial(link = "probit"))

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-1.7635295  -0.4121216  -0.0003102   0.3499560   2.2456825

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -26.28219    8.09608  -3.246  0.00117 **
xPTS         -0.03463    0.05385  -0.643  0.52020
xREB          0.28493    0.09939   2.867  0.00415 **
xAST          0.10894    0.15735   0.692  0.48874
xTO          -0.23742    0.13642  -1.740  0.08180 .
```

```
xA.T          0.71485    1.86701   0.383  0.70181
xSTL          0.45963    0.18414   2.496  0.01256 *
xBLK          0.03029    0.13631   0.222  0.82415
xPF           0.01041    0.07907   0.132  0.89529
xFG          26.58461    9.38711   2.832  0.00463 **
xFT           6.28278    2.51452   2.499  0.01247 *
xX3P          3.15824    3.37841   0.935  0.34988
---
Signif. codes:  0  0.001  0.01  0.05  0.1  1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 88.723  on 63  degrees of freedom
Residual deviance: 42.558  on 52  degrees of freedom
AIC: 66.558

Number of Fisher Scoring iterations: 8
```

# 16 Monte Carlo Simulation

## 16.1 Simulating Normal Random Variables

The normal distribution plays a major role in finance paradigms. Hence, when analytic tractability is not available, we need to resort to simulation, which means a need for a random number generator that provides standard normal $[N(0,1)]$ numbers.

Most spreadsheets provide a random uniform number generator. Recall that this is a number drawn uniformly between 0 and 1. Each element of this range may be drawn with equal probability. How do we convert $U(0,1)$ random numbers into a $N(0,1)$ random variate?

We do this by exploiting the central limit theorem, from which we know that the sum of numbers from any random distribution is normal. One common algorithm that exploits this idea is to generate a certain number of uniform random numbers and then compute their mean to obtain a normal random variate. Therefore, we generate 12 uniform random numbers, and take their mean. As follows:

$$
\begin{aligned}
x &= \frac{1}{12}\sum_{i=1}^{12}(w_i - 0.5), \\
y &\sim U(0,1),
\end{aligned}
$$

$$x \sim N(0,1).$$

The subtraction of 0.5 from each $w_i$ centers the mean of $x$ at zero. This is a crude approach to creating a random normal number.

A more sophisticated approach also begins with a uniform random number. Here, we exploit the inverse of the normal CDF, which we denote as $\Phi(z)$. If the normal PDF is denoted as $\phi(z)$, then

$$\Phi(a) = \int_{-\infty}^{a} \phi(z) dz.$$

We generate the random variate as follows:

$$\begin{aligned} x &= \Phi^{-1}(y), \\ y &\sim U(0,1), \\ x &\sim N(0,1). \end{aligned}$$

This is also more parsimonious, as only one uniform number is required, instead of 12 as in the previous method.

Note that this method, which relies on using the inverse CDF, is quite general and may be used for any probability distribution.

## 16.2 Bivariate Random Variables

- So far, we have considered univariate stochastic processes. Often, we would like to simulate two random variables, for example both the stock price and interest rate simultaneously. This requires generating two normal random variables jointly. These two random variables may be correlated. The correlation coefficient is denoted $\rho$. Of course, we require that $0 \leq \rho \leq 1$.
- How do we produce two random variables jointly? Denote these as $x_1$ and $x_2$. The procedure is as follows:
    1. First generate two *independent* random normal numbers. Denote these as $(e_1, e_2)$.
    2. Set $x_1 = e_1$.
    3. Set $x_2 = \rho e_1 + (\sqrt{1-\rho^2}) e_2$.
    4. The pair $(x_1, x_2)$ comprise two random normal numbers with correlation $\rho$.

## 16.3 Cholesky Decomposition

Given a symmetric positive definite matrix $A$, the Cholesky decomposition is a procedure that creates an upper triangular matrix $U$ such that

$$A = U^T U.$$

Therefore, $U$ is like the square-root of $A$. An equivalent exposition uses the lower triangular matrix $L$, i.e.

$$A = L\, L^T.$$

These are both embodiments of the "square-root" of $A$.

This decomposition is also known as the $LU$ decomposition because

$$A = L\, U.$$

which is a factorization procedure. The square-root connotation that is linked to the Cholesky decomposition is useful when considering covariance matrices.

Lets take an example we are already familiar with, i.e. the bivariate case from the previous section. If we want two standard normal random variates with correlation $\rho$, then the covariance matrix is

$$A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

If we apply Cholesky decomposition to this matrix $A$, we get

$$L = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$

This simplifies the mathematics in generating correlated random numbers. For example, if we generate two uncorrelated random normal variates, we can then use the $U$ or $L$ matrix to convert these numbers into correlated numbers. Let the two uncorrelated random normal numbers be

$$Z = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Then, we convert these into correlated numbers as follows:

$$X = L\, Z,$$

or in full form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Note that this is exactly the same result as we had in section 16.2.

Note also that instead of the correlation matrix, it is more usual to work with the Cholesky decomposition of the covariance matrix. In the example above, our use of the correlation matrix was correct since we assumed it to be based on (0,1) random variables, where the variance was 1, and means were zero. Therefore, the correlation matrix coincided with the covariance matrix.

**Note:** Cholesky decomposition is a standard feature in all packages for mathematical software. One of the best free languages for this is called **Octave**, distributed under the GPL. You should download (`www.octave.org`) it, and have it running on your PC. It runs on Unix, Linux, and Windows.

## 16.4 Stochastic Processes for Equity Prices

- The first step in derivative pricing comprises making an assumption about the mathematical process followed by the underlying stochastic variable in the model. When pricing equity options, we write down the stochastic process for the stock price. This is done in the form of a stochastic differential equation (SDE). For example, the Black-Scholes (1973) model is based on the following SDE:
$$dS(t) = \mu S(t) dt + \sigma S(t) dZ(t), \quad S(0) = S_0. \tag{1}$$
The parameter $\mu$ represents the mean return (continuously compounded) on the stock, and $\sigma$ is the volatility. The random shock is injected by the Brownian motion increment $dZ(t) \sim N(0, dt)$.

- The solution to the SDE for equity prices is as follows:
$$S(t) = S(0) \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma Z(t)\right] \tag{2}$$

- Thus, the stock price is lognormal, and the continuously compounded return is normal.

- The expected future value of the stock is
$$E[S(t)] = S(0) \exp(\mu t)$$
which follows directly from a simple calculation based on the lognormal distribution.

- For the purpose of implementing the random evolution of the stock via Monte Carlo simulation, we use the following equation:

$$S(t) = S(0) \exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma \epsilon \sqrt{t} \right] \tag{3}$$

  where $\epsilon \sim N(0,1)$. Simulation is undertaken using a series of draws of $\epsilon$ from a standard normal distribution.

- Hence, equity prices are characterized by a price process which grows exponentially.

## 16.5 ARCH Models

The Nobel Prize for Economics in 2003 was awarded to Robert Engle, for his innovation known as the ARCH model. ARCH stands for "AutoRegressive Conditional Heteroskedasticity", and has been found to characterize financial time series very well.

ARCH implies that financial data are (a) autocorrelated, and (b) their variance changes over time (heteroskedasticity).

The ARCH model is stipulated as follows:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 x_t + e_t & (4) \\ e_t &\sim N[0, h_t^2] & (5) \\ h_t^2 &= \beta_0 + \beta_1 e_{t-1}^2. & (6) \end{aligned}$$

The first equation is called the "mean" equation, and the second one is the "variance" equation. The variance equation is autoregressive, conditional on the previous innovation ($e_{t-1}$), and this injects heteroskedasticity.

In layman's terms, this states that the variable $y_t$ has a variance that depends on the previous variance. This conforms to a common feature in the financial markets, i.e. a large shock in one period causes volatility to rise and stay high for a while.

If we apply the ARCH model to the equity process (geometric Brownian motion) seen earlier, we will get the following system of equations:

$$\begin{aligned} S(t) &= S(0) \exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma Z(t) \right] & (7) \\ Var(Z_t) &= \beta_0 + \beta_1 Z_{t-1}^2. & (8) \end{aligned}$$

The GARCH model is a simple extension of the ARCH model and was developed by Tim Bollerslev. The system of equations is as follows:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 x_t + e_t & (9) \\ e_t &\sim N[0, h_t^2] & (10) \\ h_t^2 &= \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 h_{t-1}^2. & (11) \end{aligned}$$

## 16.6 Co-integration

The 2003 Nobel was shared by Rob Engle (for ARCH), and Clive Granger (for the concept of co-integration). Lets dwell on cointegration for a moment to close the loop on this topic.

Econometricians working with time series would prefer to assume that their data is stationary, which means "well-behaved" or stable, at least around a trend. The model that generates the data needs to be stable.

However, financial time series are "non-stationary" to a fault. This makes traditional approaches inapplicable, and calls into question usual statistics such as t-statistics, $R^2$, or the Durbin-Watson statistic. Sometimes, the non-stationarity may be fixed by differencing, to remove the time trend. But this is not always the only form of non-stationarity.

The notion of "integration" arises from the extent of differencing required to render a time series stationary. A series is said to be $I(d)$, integrated order $d$, if the number of times differencing is required to get a stationary time series is $d$. A stationary time series of data is order $I(0)$, i.e. it does not need any more differencing.

Take the example of a random walk, i.e. $x_t = x_{t-1} + e_1$. If the error term is stationary, then this series is $I(1)$, because differencing gives $\Delta x_t = e_t$, which is $I(0)$. Instead,

$$x_t = \beta x_{t-1} + e_1$$

is stationary if $|\beta| < 1$, because the series $y$ cannot randomly walk away, but this is not the case if $|\beta| \geq 1$. If $|\beta| >$, then the series would explode anyway. If $|\beta| = 1$, then the series is unstable. Therefore, testing for non-stationarity usually involves the hypothesis that $|\beta| = 1$. At an intuitive level, a variable that wanders incessantly is called "non-stationary".

Why do we care about this? First, the usual statistics we obtain from regressions are faulty. Second, is the issue of cointegration. If integrated time-series are the problem, cointegration is the panacea. Clive Granger is now rightly recognized for his work in providing and analysing this phenomenon.

If we try to explain a phenomenon by regressing nonstationary data on explanatory variables which are nonstationary, then we get nonsense results. However, if the variables are "cointegrated", we may still obtain meaningful results. What does this mean? Even if two time series wander around in an integrated manner, they may be related, so that we can find a function of the two series that provides stationarity. For example, the returns on stocks may be integrated $I(1)$, but regressing them on the index return which is also integrated, leads to stationary errors (residuals), making the coefficients in the regression acceptable.

## 16.7 Interest-rate Processes

- Interest rate process do not grow exponentially. Instead, they cycle with the economy. Technically speaking, they are *mean-reverting*.
- The Vasicek model is a popular choice for simple interest rate processes. It is based on the Ornstein-Uhlenbeck (OU) stochastic process, quite well-known in physics.

$$dr(t) = \kappa[\theta - r(t)]dt + \sigma dZ(t), \quad r(0) = r_0. \tag{12}$$

  Here, $\kappa$ is the speed of mean-reversion, $\theta$ is the long-run mean of the interest rate process.
- The solution to this SDE is as follows:

$$r(t) = r(0)e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dZ(s) \tag{13}$$

- The interest rate is normally distributed in this model.
- The expectation and variance of the interest rate are:

$$\begin{aligned} E[r(t)] &= r(0)e^{-\kappa t} + \theta(1 - e^{-\kappa t}) & (14) \\ Var[r(t)] &= \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}) & (15) \end{aligned}$$

- Simulation may be undertaken via the following equation:

$$r(t) = r(0)e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma e^{-0.5\kappa t}\epsilon\sqrt{t}, \tag{16}$$

  where $\epsilon \sim N(0, 1)$.
- The OU process permits negative interest rates. Hence, a popular alternative to the OU process is the square-root diffusion process, first used in finance by Cox, Ingersoll and Ross (CIR, 1985). This process is:

$$dr(t) = \kappa[\theta - r(t)]dt + \sigma\sqrt{r}dZ(t), \quad r(0) = r_0. \tag{17}$$

The only change is in the volatility function. Volatility is now dependent on the level of interest rates, and if $\sigma^2 < \kappa\theta$, then interest rates will never reach zero in this model.

- The CIR model may be simulated from the following equation, which is a small modification of the earlier OU model:

$$r(t) = |r(0)e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma e^{-0.5\kappa t}\epsilon\sqrt{r(t)\,t}|, \tag{18}$$

where $\epsilon \sim N(0,1)$. While in continuous-time the process never reaches zero, it may still do so during simulation in discrete-time. Hence, the absolute value of the interest rate is always taken during the Monte Carlo step. This correction was first pointed out by Beaglehole and Tenney (*Journal of Financial Economics*, 1992).

## 16.8 Estimating Historical Volatility for Equities

- Note that the volatility $\sigma$ is the *annualized* standard deviation of *continuously compounded* stock *returns*. It is NOT the volatility of the stock price itself. This follows from equation (2).

- Historical volatility is computed from a time series of continuously compounded stock returns. If the time interval between two consecutive stock prices is $h$, the continuously compounded return for that time interval is:

$$R_{t,t+h} = \ln\left(\frac{S_{t+h}}{S_t}\right).$$

For example, if $S_t = 101$ and $S_{t+h} = 105$, then $R_{t,t+h} = 0.03884$.

- Usually, we use daily data, and compute a return for each day, resulting in a time series of continuously compounded daily returns. From this time series, we compute the variance of daily returns, i.e. $Var[R_{t,t+h}]$.

- The annualized variance is obtained by multiplying the variance of daily returns by the number of trading days in the year:

$$\sigma^2 = Var[R_{t,t+h}] \times (\text{No. of trading days}).$$

And finally, the annualized volatility is the square-root of the the variance.

## 16.9 Estimating Historical Volatility for Interest Rates

- Regression is a simple approach to estimating interest rate process parameters. Recall the OU process used earlier:
$$dr(t) = \kappa[\theta - r(t)]dt + \sigma dZ(t), \quad r(0) = r_0. \tag{19}$$
- A discrete-time version of this process is as follows:
$$r(t+h) - r(t) = \kappa[\theta - r(t)]h + \sigma\epsilon(t)\sqrt{h} \tag{20}$$
where $h$ is the time interval between observations.
- This process may be rewritten as a linear regression model:
$$y(t) = \alpha + \beta r(t) + e(t) \tag{21}$$
where
$$\begin{aligned} y(t) &= r(t+h) - r(t) \\ \alpha &= \kappa\theta h \\ \beta &= -\kappa h \\ e(t) &= \sigma\epsilon(t)\sqrt{h} \end{aligned}$$
Note that this is a simple ordinary least squares (OLS) regression where $y(t)$ is the dependent variable and $r(t)$ is the independent variable. The intercept of the regression is $\alpha$ and the coefficient on the independent variable is $\beta$.
- The coefficients $[\alpha, \beta]$ are estimated from a time series of interest rates. If for example, daily interest rates are used, the time interval is $h = 1/262$, where 262 is the number of trading days in the year.
- From $\alpha$ and $\beta$, we can invert the values of the stochastic process parameters as follows:
$$\begin{aligned} \kappa &= -\frac{\beta}{h} \\ \theta &= \frac{\alpha}{\kappa h} = -\frac{\alpha}{\beta} \end{aligned}$$
- The parameter $\sigma$ is obtained from the standard error of the regression, i.e. from the value of $\sigma_e = \sqrt{Var[e(t)]}$, i.e. the variance of the residuals from the regression. Note that
$$Var[e(t)] = \sigma^2 h$$
Hence,
$$\sigma = \sqrt{\frac{1}{h} Var[e(t)]} = \frac{1}{\sqrt{h}}\,\sigma_e.$$
- Exercise: Can you use the OLS method for the CIR model? Suggest an alternative approach.

## 16.10 Path-dependent Options

- Simulation is especially useful for options where the payoff at maturity depends on the entire sample path of the process prior to maturity, not only on the value of the random variable at maturity.
- An example of a path-dependent options is the "Asian" option, based on the average of the stock price upto and including maturity. The period for which the average is computed is called the "averaging period", and it may be shorter than the life of the option itself. For example, the option may be written for a maturity of a year, but the averaging period may be a half year. Asian options may be puts or calls.
- Another example of path-dependence arises in the pricing of mortgage-backed securities (MBS). The prepayment function, a key ingredient of the pricing model for MBS, typically depends on the path of interest rates.
- Monte Carlo methods offer a seamless way of pricing path-dependent options. Pricing only involves keeping track of the path of stock prices or interest rates during the simulation.

## 16.11 Variance Reduction

- Monte Carlo simulation is an estimation technique based on random numbers. Hence it is not precise. The error in the estimation may be reduced by running more simulations.
- To estimate the price of a stock option, we simulate a path of the stock price from its starting value. We then compare its terminal value at maturity under the simulation with the strike price. If we simulated but one sample path, our estimate would have considerable error. Therefore, we usually simulate many sample paths. Our choice of the number of simulations depends on the degree of accuracy desired, which is a choice based on the practical requirements of the situation.
- Suppose we choose to run $n$ simulations. Each sample path is *independent* of every other path by construction of the Monte Carlo method. After generating $n$ sample paths, we average the values obtained from each sample path to estimate the expected price of the option. Under the assumption of independence, doubling the number of simulation to $2n$ will result in a decrease in the variance of the estimate by one half. In general, if we run $k \times n$ simulations, the variance will scale by $\frac{1}{k}$.

- Once we have run the Monte Carlo pricing algorithm for $n$ simulations and determined the variance, we can easily compute the number of simulations needed to obtain any desired level of accuracy.

- The estimation variance depends on all the features of the option, its moneyness, maturity, and volatility.

- The goal of variance-reduction techniques is to reduce the estimation variance without increasing the run time of the Monte Carlo algorithm. There are many methods that may be used. We outline a few below.

## 16.12 Antithetic variate Method

- Standard Monte Carlo imposes independence of sample paths. By modifying this feature variance reduction is attainable.

- What if each path were perfectly correlated with every other? This would mean all paths were identical. Increasing $n$ does not result in a reduction in estimation variance.

- Conversely, if paths were negatively correlated with each other then the estimation variance would decline. This is the idea behind the antithetic variate approach.

- Implementation of the antithetic variate method is very simple. Instead of generating one sample path at a time, we generate two paths, the primary path and its antithesis. Whenever we generate a random number $\epsilon$ for the primary path, we set the random number for the antithetic path to be $-\epsilon$.

- The method provides two paths for the computational price of one. It also reduces variance. Thus it offers doubled efficiency.

## 16.13 Control variate techniques

- A given set of random numbers used for Monte Carlo simulation results in the estimate being biased high or low. If the sign of the bias was known, we could skew our estimate accordingly in the right direction.

- Monte Carlo estimation is used when no closed-form pricing equation is available. Sometimes, a closed-form solution is known to a problem that is similar to, but not exactly the same as the one under consideration. This known solution, denoted a "control-variate", may be used to impose bias-correction on simulation estimates.

- Denote the Monte Carlo estimate of our pricing problem as $P$. The "look-alike" problem that admits a closed form solution has a Monte Carlo estimate denoted $Q$. The exact price from the closed-form equation is $Q_0$. By comparing $Q$ with $Q_0$, we can discover the direction of the bias from the simulation algorithm.

- After accounting for the bias, we correct the original estimate $P$ as follows:

$$P' = P + (Q_0 - Q)$$

- Variance reduction requires that

$$Var[P'] < Var[P].$$

  Let us compute the variance of the revised estimate $P'$:

$$\begin{aligned} Var[P'] &= Var[P + (Q_0 - Q)] \\ &= Var[P - Q] \\ &= Var[P] + Var[Q] - 2Cov(P, Q) \end{aligned}$$

  Therefore,

$$Var[P'] < Var[P] \quad \text{if} \quad Cov(P, Q) > \frac{1}{2}\, Var[Q]$$

  Simply put, the greater the similarity between the original problem and its control variate, the greater the benefit from the technique.

- **Caveat:** The random numbers used for computing $P$ and $Q$ must be exactly the same. Therefore, the simulation for $P$ must store the random numbers used, and reuse them for estimating $Q$.

# 17 Forwards and Futures

Forwards and futures are the simplest derivatives. What is a derivative security? It is a security whose payoffs (cashflows) are based on the movements of another security, usually termed the "underlying." This other security could be a "primary" security, i.e., it is priced based on fundamentals, or it could be based on another derivative security. Derivatives are therefore, "secondary" securiteies, i.e., their values are based on other securities, irrespective of fundamentals.

## 17.1 Arbitrage

1. Define the law of one price.

2. What is “risk” arbitrage?
3. What is a “riskless” arbitrage?
    - A deal that costs zero today, and generates positive payoffs somewhere in the future.
    - A deal that generates positive initial cashflow, and non-negative future payoffs.

## 17.2 Forward Contract

Define a forward contract. What are its features?

- Contract today to do something in future.
- Symmetric obligation.
- Net zero value at inception.
- Price vs Payoff.

## 17.3 Pricing Forwards

How do you determine the price of a forward contract for a financial security, such as the stock of IBM?

- IBM example, guess the forward price game.
- $F = S(1 + r)$, where $r$ = cost of carry
- No-arbitrage argument
- Pricing by replication: borrow cash, buy spot, 0 investment
- Payoff diagram.

## 17.4 Commodity Forward Prices

How do you determine the price of a forward contract for a commodity, such as gold or oil?

- Price-discovery mechanism vs cost-of-carry argument.
- $F = S(1 + r - \text{div/convenience yield} + \text{storage cost})$.
- $F > S \Longrightarrow$ contango
- $F < S \Longrightarrow$ backwardation

## 17.5 Replication of a Forward Contract

- Replicate a long position: borrow cash, buy spot, store asset. This gives $F_{long}$.
- Replicate a short position: borrow asset, sell spot, invest cash. This gives $F_{short}$.
- In general we get
$$S(1 + r - r_b) = F_{short} < F < F_{long} = S(1 + r + r_s)$$

## 17.6 The Need for forwards

Shown how forwards can be replicated using the asset and cash. If replication is possible, why do we need forwards?

- Credit risk - replace 100% risk with 10% risk.
- Convenience
- Replication affects the spot market
- Can do transactions off balance sheet
- Obviates short selling constraints
- Manages position in the presence of borrowing constraints
- Transactions costs lowered
- Transactions are collateralized
- Helps in repo deals
- Information revelation at lower cost

## 17.7 Futures

How are futures different from forwards? Why are they better then forwards in some situations?

- Daily settlement (MTM)
- Broker required
- Initial margin [volatility, credit risk, price limits]
- Variation margin
- Central clearing
- Indemnity fund
- Goes from 10% risk to 1% risk.[T day risk to 1 day risk]
- Forwards are non-linear, futures are linear.

## 17.8 Futures vs Forwards

When are futures less efficient than forwards?

- In standardized markets - FX
- Fwds have less paperwork
- When counterparties have good credit risk
- Futures have high commissions at times.

Should the forward price equal the futures price?

Only if interest rates are non-stochastic.

Proof: Each day enter into a futures contract and roll it over. Assume $r$ constant. We get for 10 days

| Day | Futures Price | Position | Gain/loss |
|---|---|---|---|
| 0 | $F_0$ | $(1+r)$ | 0 |
| 1 | $F_1$ | $(1+r)^2$ | $(F_1 - F_0)(1+r)$ |
| 2 | $F_2$ | $(1+r)^3$ | $(F_2 - F_1)(1+r)^2$ |
| : | : | : | : |
| 10 | $F_{10}$ | $(1+r)^{11}$ | $(F_{10} - F_9)(1+r)^{10}$ |

Compound all the gains to maturity, to get at $t = 10$

$$\begin{aligned}
\text{Gain} &= (F_1 - F_0)(1+r)(1+r)^9 + ... + (F_{10} - F_9)(1+r)^{10} \\
&= (F_{10} - F_0)(1+r)^{10} \\
&= (S_{10} - F_0)(1+r)^{10}
\end{aligned}$$

Now set up a portfolio of the gain plus an investment of $F_0$ in the riskfree bond. At maturity this portfolio is worth

$$(S_{10} - F_0)(1+r)^{10} + F_0(1+r)^{10} = S_{10}(1+r)^{10}$$

The initial cost of this portfolio is $F_0$. Now, suppose the forward price is $G_0$. Invest $G_0$ in a riskless bond and buy a long position in $(1+r)^{10}$ forward contracts. At $t = 10$, the portfolio is worth

$$(S_{10} - G_0)(1+r)^{10} + G_0(1+r)^{10} = S_{10}(1+r)^{10}$$

Since the final values of the futures portfolio and forwards portfolios are the same, their initial values must be the same by no-arbitrage. Hence, the forward price equals the futures price:

$$F_0 = G_0$$

# 18 Basic Options Concepts

## 18.1 Option Definitions

- What is an option? [right but not an obligation]
- What is the difference between American and European options?
- What are puts and calls? [in interest rate settings?]
- What are the different types of options available?
- Define the concept of a "derivative" security.

**Notation:**

- | | |
  |---|---|
  | $S$ | current stock price |
  | $S_T$ | terminal stock price |
  | $T$ | maturity |
  | $K$ | strike price |
  | $C, P$ | calls, puts |
  | $r$ | interest rate |
  | $\sigma$ | volatility |

## 18.2 Call Options

When a call is exercised, the buyer pays the writer the strike price, and gives up the call in exchange for the asset.

For the privilege of exercising only if desirable, the buyer of the call initially pays the writer a price called the *call price*.

From the *Wall Street Journal*:

Asset: GM Stock
Current asset price: $40\frac{1}{4}$
Strike Price: 35
Maturity date: February
Call price: $5\frac{3}{8}$

## 18.3 Put Options

When a put is exercised, the buyer gives up the put, and receives from the writer the exercise price in exchange for the asset.

For the privilege of exercising the put only if desirable, the buyer initially pays the writer a price called the *put price*.

From the *Wall Street Journal*:

Asset: GM Stock
Current asset price: $40\frac{1}{4}$
Strike Price: 35
Maturity date: February
Put price: $\frac{1}{8}$

## 18.4 Basic Trading Strategies

This segment examines the payoffs from:

1. Covered call.
2. Protective put.
3. Spreads:
   (a) Bullish vertical spread.
   (b) Bearish vertical spread.
   (c) Butterfly spread.
4. Combinations:
   (a) Straddles and strangles.
   (b) Strips and straps.

## 18.5 Covered Call: Long stock, short call

If $S_T < K$:

Value of stock: $S_T$
Value of option: 0
Total value: $S_T$

If $S_T \geq K$:

Value of stock: $S_T$
Value of option: $-(S_T - K)$
Total value: $S_T - S_T + K = K$

Long stock position "covers" writer of option from sharp increase in stock price.

## 18.6 Protective Put: Long stock, long put

If $S_T < K$:

Value of stock: $S_T$
Value of option: $K - S_T$
Total value: $S_T + K - S_T = K$

If $S_T \geq K$:

Value of stock: $S_T$
Value of option: 0
Total Value: $S_T$

A protective put portfolio protects the investor from a decrease in stock price.

## 18.7 Spreads

A *spread* involves taking positions in two or more options of the *same* type (i.e., all calls, or all puts).

For example:

- A *bullish vertical spread* involves buying a call option at exercise price $K_1$ and selling an otherwise identical call option at another exercise price $K_2 > K_1$.
- A *bearish vertical spread* involves selling a call option at exercise price $K_1$ and buying an otherwise identical call option at another exercise price $K_2 > K_1$.
- A *butterfly spread* is obtained by buying a call option each at strike prices $K_1$ and $K_3$, and selling two call options at a strike price of $K_2$, $K_1 < K_2 < K_3$.
- A *horizontal* or *calendar spread* is obtained by selling a call option with maturity $T_1$, and selling an otherwise identical call option with maturity $T_2 > T_1$.

## 18.8 Bullish Vertical Spread: Gross Payoffs

- If $S_T < K_1$: Neither option is exercised.
  Payoffs $= 0$
- If $K_1 < S_T < K_2$: Only first option is exercised.
  Payoffs $= S_T - K_1$
- If $K_2 < S_T$: Both options are exercised.
  Payoff from first option $= S_T - K_1$
  Payoff from second option $= -(S_T - K_2)$
  Total payoff $= K_2 - K_1$

## 18.9 Bearish Vertical Spread: Gross Payoffs

- If $S_T < K_1$: Neither option is exercised.
  Payoffs = 0
- If $K_1 < S_T < K_2$: Only first option is exercised.
  Payoffs $= -(S_T - K_1)$.
- If $K_2 < S_T$: Both options are exercised.
  Payoff from first option $= -(S_T - K_1)$
  Payoff from second option $= (S_T - K_2)$
  Total payoff $= K_1 - K_2$

## 18.10 Butterfly Spread: Gross Payoffs

Assume $K_2 = \frac{1}{2}(K_1 + K_3)$

- If $S_T < K_1$: No option is exercised.
  Payoffs: 0
- If $K_1 < S_T < K_2$: Only $K_1$-strike option is exercised. Payoffs: $S_T - K_1$
- If $K_2 < S_T < K_3$: Only options with strikes $K_1$ and $K_2$ are exercised.
  Payoffs: $S_T - K_1 - 2S_T + 2K_2 = K_3 - S_T$
- If $K_3 < S_T$: All options are exercised.
  Payoffs: $S_T - K_1 + S_T - K_3 - 2(S_T - K_2) = 0$

## 18.11 Combinations:

A *combination* is a trading strategy that involves taking a position in both calls and puts in the same stock. For example:

- A *straddle* is a put and a call with the same strike and maturity.
- A *strangle* is a call and a put with the same maturity, but different strikes.
- A *strip* is one call and two puts with the same strike and maturity.
- A *strap* is two calls and one put with the same strike and maturity.

### 18.12 Straddle: Gross Payoffs

If $S_T < K$:

Value of call $= 0$
Value of put $= K - S_T$
Total value $= K - S_T$

If $S_T \geq K$:

Value of call $= S_T - K$
Value of put $= 0$
Total value $= S_T - K$

### 18.13 Strip: Gross Payoffs

If $S_T < K$:

Value of call $= 0$
Value of puts $= 2(K - S_T)$
Total value $= 2(K - S_T)$

If $S_T \geq K$:

Value of call $= S_T - K$
Value of puts $= 0$
Total value $= S_T - K$

## 19 No-Arbitrage Restrictions on Option Values

Idea: To identify restrictions on prices of puts and calls based on the notion that a competitive market should not admit the possibility of a riskless profit.

Two portfolios which imply the presence of arbitrage:

1. A portfolio which involves no cash outflow today, but will lead to positive cash inflows in the future.
2. A portfolio which involves a net cash inflow today, and which will not involve cash outflows in the future.

## 19.1 Restrictions on Call-Option Pricing

We identify eight restrictions that "rational" call option prices must satisfy:

- The first four relate the price of the call to the price of the underlying asset.
- The next three relate the price of the call to the exercise price.
- The last relates the price of the call to the time to maturity.

**Notes**

- We will not distinguish between European and American options in this exercise.
- However, if a restriction holds only for American (or only for European) options, this is highlighted.
- The task of identifying the analogous restrictions for *put* option prices is left to the reader as an exercise.

## 19.2 Relationship between $C$ and $S$

1. $C \geq 0$.
2. $C < S$.
3. $C \geq S - K$.
4. $C \geq S - PV(K) - PV(D)$.

**Note** Restriction 3 holds only for American call options.

**Remark** In 4, $PV(K)$ is the present value of an amount $K$ received at time $T$, and $PV(D)$ is the present value of dividend inflows

**Proof of Restriction 1**

If $C < 0$, then an arbitrage results (buy the option and throw it away).

**Proof of Restriction 2**

A rational investor would not pay more than $S$ for right to purchase asset at a price of $S$.

**Proof of Restriction 3**

If $C < S - K$, then an arbitrage results (buy the option and exercise it immediately).

**Proof of Restriction 4**

Consider two portfolios:

1. Portfolio A: Long one stock.
2. Portfolio B: Long one call option, lend $PV(K)$, lend $PV(D)$.

At date $T$, if $S_T < K$:

Value of Portfolio A: $S_T + D$
Value of Portfolio B: $K + D$

At date $T$, if $S_T \geq K$:

Value of Portfolio A: $S_T$
Value of Portfolio B: $S_T - K + K + D$

So portfolio B always does as well as portfolio A, and in some cases does strictly better. Therefore, the cost of portfolio B must be higher than that of portfolio A. (Otherwise, an arbitrage results from selling portfolio A and buying portfolio B.) That is, we must have

$C + PV(K) + PV(D) \geq S$

which is the same as

$C \geq S - PV(K) - PV(D)$.

## 19.3 Relationship between $C$ and $K$

1. If $K_1 < K_2$, then $C(K_2) < C(K_1)$.
2. If $K_1 < K_2$, then $C(K_1) - C(K_2) < K_2 - K_1$.
3. If $K_1 < K_2 < K_3$, then
   $C(K_2) < wC(K_1) + (1 - w)C(K_3)$,
   where $w$ is defined by
   $$w = \frac{K_3 - K_2}{K_3 - K_1}.$$

**Note** The property described in Restriction 7 is called "convexity."

**Proof of Restriction 5**

If $K_1 < K_2$ and $C(K_1) < C(K_2)$, an arbitrage results (buy the option with strike $K_1$ and write the option with strike $K_2$). Note that Restriction 5 is simply the statement that a bullish vertical spread cannot be purchased "for free."

**Proof of Restriction 6**

We first prove the result assuming the options are European. Then we explain how the arguments may be extended to American options. Suppose $K_1 < K_2$, but $C(K_1) - C(K_2) > K_2 - K_1$.

Then, an arbitrage may be created as follows:

- Sell the call with strike $K_1$, buy the call with strike $K_2$, and lend $(K_2 - K_1)$ for maturity at $T$.
- At time $T$, exercise the $K_2$-strike option if and only if $S_T \geq K_2$.

The initial cash inflow from this strategy is

$$C(K_1) - C(K_2) - (K_2 - K_1) \ > \ 0.$$

There are three possible cases at time $T$: $S_T < K_1$, $K_1 < S_T < K_2$, and $S_T > K_2$.

We will show that the net cash inflow at $T$ is nonnegative in all three cases.

**Case 1:** $S_T < K_1$

In this case, both options lapse unexercised. The only cash inflow is from the loan made, which is positive.

**Case 2:** $K_1 < S_T < K_2$

Here, only the $K_1$ call is exercised leading to a cash outflow of $(S_T - K_1)$, and a cash inflow from the loan of $(K_2 - K_1) +$ int.

**Case 2: (cont'd)** $K_1 < S_T < K_2$

The net cash inflow that results is

$$K_2 - S_T + \text{ int} > 0.$$

**Case 3:** $S_T > K_2$

Now both options are exercised, leading to a cash outflow of $(S_T - K_1)$, and cash inflows of $(S_T - K_2)$ and $(K_2 - K_1) +$ int. The net cash flow is the interest, which is nonnegative. Thus, the suggested strategy provides an arbitrage.

**Modification for American options**

If the options are American, rather than European, an arbitrage can be created by setting up the same portfolio, but modifying the strategy as follows:

- Close out the loan at the time the $K_1$-strike option is exercised.
- Exercise the $K_2$ option at any this point if $S_t \geq K_2$.
- Otherwise exercise the $K_2$ option at $T$, provided $S_T \geq K_2$.

Note that at any time $t$, the value of the loan is equal to the principal $(K_2 - K_1)$ plus the interest accumulated upto that point.

**Proof of Restriction 7**

Suppose $K_1 < K_2 < K_3$, but

$C((K_2) > wC(K_1) + (1 - w)C(K_3),$

where

$$w = \frac{K_3 - K_2}{K_3 - K_1}.$$

We will show that an arbitrage profit may be created.

Once again, we will first do this for the case of European options, and then extend the arguments to cover American options also. Consider the following strategy:

- Buy $w$ calls with strike $K_1$, buy $(1 - w)$ calls with strike $K_3$, and sell one call with strike $K_2$.

**4 Cases**

The initial cash flow from this strategy is positive.

We will show that all future cash flows are nonnegative, so the strategy is an arbitrage.

There are four possible cases at time $T$:

**Case 1:** $S_T < K_1$

In this case, all options lapse unexercised, so there are no cash flows at $T$.

**Case 2:** $K_1 < S_T < K_2$

Here, only the $K_1$ option is in the money, so the cash inflow is positive, and there is no cash outflow.

**Case 3:** $K_2 < S_T < K_3$

Now, the $K_1$ and $K_2$ options are both in the money, and will be exercised. The cash inflow from the $K_1$ options is $w(S_T - K_1)$, while the outflow from the $K_2$ option is $S_T - K_2$. Substituting for $w$, some algebra shows that the net inflow is:

$$\left(\frac{K_2 - K_1}{K_3 - K_1}\right)(K_3 - S_T) \;>\; 0.$$

**Case 4:** $K_3 < S_T$

Now all the calls are in the money. The net cash inflow at $T$ is, therefore,

$$w(S_T - K_1) + (1 - w)(S_T - K_3) = -(S_T - K_2),$$

which equals

$$K_2 - wK_1 - (1 - w)K_3 \;=\; 0.$$

This completes the proof of restriction 7 for European options.

**Modification for American options**

If the options are American, rather than European, the same portfolio will result in an arbitrage, provided the exercise strategy is modified as follows:

- When the $K_2$-call is exercised, exercise all the calls that are in the money.
- At time $T$, exercise all unexercised calls that are in the money.

The initial cash inflow is still strictly positive.

Minor modifications of the arguments given above show that all cash inflows under the suggested strategy are nonnegative.

## 19.4 Relationship between $C$ and $T$

1. If $T_1 < T_2$, then $C(T_1) < C(T_2)$.

**Note:** Restriction 8 need not hold for European calls.

**Proof of Restriction 8:**

If $C(T_1) > C(T_2)$, an arbitrage opportunity can be created by adopting the following strategy:

- Buy the call with expiration $T_2$, and write the call with expiration $T_1$.
- Exercise the $T_2$-call when the holder of the $T_1$ call exercises.
- If the $T_2$-call is still unexercised at $T_1$, exercise it at any time that it is in the money.

This strategy leads to a positive cash inflow at time 0.

At any time prior to $T_1$, either both calls are exercised, or neither is, and in either case, there is no net cash flow.

After $T_1$, only the $T_2$ call is alive, so there is no cash outflow.

## 19.5 Restrictions on Put Option Pricing

Analogous to the restrictions for call, we can derive the following restrictions on put option prices (the proofs are omitted):

1. $P \geq 0$.
2. $P < K$.
3. $P \geq (K - S)$ (only American puts).
4. $P \geq PV(K) + PV(D) - S$.
5. If $K_1 < K_2$, then $P(K_1) < P(K_2)$.
6. If $K_1 < K_2$, then $P(K_2) - P(K_1) < K_2 - K_1$.
7. If $K_1 < K_2 < K_3$, then
   $$P(K_2) < wP(K_1) + (1 - w)P(K_3).$$
   where $w = (K_3 - K_2)/K_3 - K_1)$.
8. If $T_1 < T_2$, then $P(T_1) < P(T_2)$ (American puts only).

## 19.6 Comparing and Relating Options

We examine two issues in this segment:

1. European Option Prices vs. American Option Prices, when both options are of the same type (i.e., both calls, or both puts).
2. Put Option Prices vs. Call Option Prices, when both options are of the same style (i.e., both European, or both American).

The restrictions we derive will be based solely on no–arbitrage arguments, so they will have a very general validity.

## 19.7 European Option vs. American Option Prices

Only difference between American and European options is right to exercise early. We examine when this right is important (i.e., under what conditions it may be exercised). It is important to understanding relative prices of European and American options. In particular, if right to early-exercise is never used in some American options, then these options must trade at the same price as the corresponding European options.

We examine question of *early-exercise* in each of four cases:

1. No Interim Payouts on Asset.
   (a) American Call.
   (b) American Put.
2. Interim payouts exist.
   (a) American Call.
   (b) American Put.

**Case 1(a): American Call, No Dividends**

Recall no-arbitrage restriction that

$$C \geq S - PV(K).$$

Define $IV(C) = C - (S - PV(K))$.

$IV(C)$ is the *insurance value* of having the call. The no-arbitrage restriction may now be expressed as

$$C = S - PV(K) + IV(C).$$

Adding $K$ to both sides, and rearranging:

$$C - S + K = K - PV(K) + IV(C).$$

This is the same as:

$$C - (S - K) = (K - PV(K)) + IV(C).$$

Left-Hand Side (LHS): Loss/gain from immediate exercise.

Right-Hand Side (RHS): Expresses loss/gain as sum of two terms:

- First term $K - PV(K)$ is pure *time-value* of exercise price that is *lost* from early exercise.
- Second term $IV(C)$ is insurance value of call which is *lost* from early exercise.

There are no gains from early exercise.

**Conclusion:** American option on asset with no interim payouts is never exercised early.

**Case 1(b): American Put, No Dividends**

Recall no-arbitrage restriction that

$$P \geq PV(K) - S.$$

Define $IV(P) = P - (PV(K) - S)$.

$IV(P)$ is the *insurance value* of having the put. The no-arbitrage restriction may now be expressed as

$$P = PV(K) - S + IV(P).$$

Subtracting $K$ from both sides and rearranging,

$$P - K + S = -K + PV(K) + IV(P).$$

which is the same as:

$$P - (K - S) = -(K - PV(K)) + IV(P).$$

LHS: Loss/gain from immediate exercise.

RHS: Expresses loss/gain as sum of 2 terms:

- First term $(K - PV(K))$ is time-value of money *gained* by exercising put immediately.
- Second term $IV(P)$ is insurance value of put *lost* by exercising put immediately.

Sign of RHS is ambiguous.

**Conclusion:** If insurance value of put is larger than time value gained, do not exercise put. Otherwise exercise put.

**Example:** If interest rates are "high," and volatility of stock price is "low," put should be exercised.

**Case 2(a): American Call with Dividends**

If American Call is payout protected, then this is the same as the no-dividend case. Assume call is *not* payout protected.

Recall no-arbitrage restriction that

$$C \geq S - PV(K) - PV(D).$$

Let $IV(C) = C - (S - PV(K) - P(D))$.

$IV(C)$ is the *insurance value* of having the call.

Adding $K$ to both sides and rearranging, the no-arbitrage restriction may now be expressed as

$$C - (S - K) = (K - PV(K)) - PV(D) + IV(C).$$

LHS: Loss/gain from immediate exercise.

RHS: Expresses loss/gain as sum of 3 terms:

- The first term $(K - PV(K))$ is the time-value of exercise price *lost* by immediate exercise.

- The second term $PV(D)$ is the dividend amount *gained* by immediate exercise.
- The third term, $IV(C)$ is the insurance value of the call that is *lost* from immediate exercise.

RHS could be positive or negative.

**Conclusion 1:** Could be optimal to exercise American option on a dividend-paying stock early.

**Conclusion 2:** If early exercise is optimal, maximum gain occurs from exercising just before stock goes ex-dividend.

**Factors making early-exercise more likely:**

- High dividends.
- Low interest rates.
- Short period left to maturity.
- Low volatility of stock price.
- High depth-in-the-money.

**Case 2(b): American Put with Dividends**

Recall arbitrage restriction

$$P \geq PV(K) + PV(D) - S.$$

Define $IV(P) = P - (PV(K) + PV(D) - S)$.

$IV(P)$ is the *insurance value* of having the put.

Subtracting $K$ from both sides and rearranging, we get

$$P - (K - S) = -(K - PV(K)) + PV(D) + IV(P).$$

LHS: Loss/gain from immediate exercise.

RHS: Expresses loss/gain as sum of 3 terms:

- The first term $K - PV(K)$ is time-value of exercise price *gained* by immediate exercise.

- The second term $PV(D)$ is the value of dividends *lost* by immediate exercise.
- The third term $IV(P)$ is insurance value of option *lost* by immediate exercise.

RHS could be positive or negative.

**Conclusion 1:** Could be optimal to exercise put option on a dividend-paying stock early.

**Conclusion 2:** Early exercise could be optimal at any time, not just before stock goes ex-dividend.

**Factors making early-exercise more likely:**

- Low dividends.
- High interest rates.
- Low volatility of stock price.
- High depth-in-the-money.

## 19.8 Put–Call Parity

Consider put and call that are otherwise identical:

- Same underlying asset.
- Same exercise price.
- Same maturity date $T$.
- Same style (both American or both European).

How is put price $P$ related to call price $C$?

We examine the answer in four cases:

1. No interim payouts on stock.
   (a) Both options are European.
   (b) Both options are American.

2. Interim payouts exist.
    (a) Both options are European.
    (b) Both options are American.

**Case 1(a): Both European, No Dividends**

Consider the following portfolios:

**Portfolio A:** Buy 1 put, buy 1 share of stock.

**Portfolio B:** Buy 1 call, lend $PV(K)$.

Cost of setting up Portfolio A: $P + S$.

Cost of setting up Portfolio B: $C + PV(K)$.

Value of Portfolio A at maturity:

1. If $S_T < K$: Put option is worth $K - S_T$; stock is worth $S_T$. Total value: $K$.
2. If $S_T \geq K$: Put option expires worthless; stock is worth $S_T$. Total value: $S_T$.

Value of Portfolio B at maturity:

1. If $S_T < K$: Call option expires worthless; loan matures with face value $K$. Total value: $K$.
2. If $S_T \geq K$: Call option is worth $S_T - K$; loan matures with face value $K$. Total value: $S_T$.

Thus, the portfolios have the same terminal value whether $S_T < K$ or $S_T \geq K$. To avoid arbitrage, portfolios must have same initial value:

$$P + S \ = \ C + PV(K).$$

Therefore, the put and call prices of European options are related by:

$$P \ = \ C + PV(K) - S.$$

This formula is called the "put-call parity" formula for European options.

Put-call parity can be used to create *synthetic puts* using *traded calls*, and *synthetic calls* using *traded puts*

For example, suppose the aim is to create a European put option with strike price $K$ and maturity $T$. The following portfolio will mimic the payoff of such an option:

- Buy a European call with strike $K$ and maturity $T$.
- Short one unit of the stock.
- Lend $PV(K)$ (i.e., buy a T-bill with maturity $T$ and face value $K$).

The value of the portfolio at maturity is

$$\begin{array}{cc} K - S_T, & \text{if } S_T < K \\ 0, & \text{if } S_T \geq K \end{array}$$

which is exactly the payoff from the desired put.

**Case 1(b): Both American, No Dividends**

An American call on a stock that pays no dividends will not be exercised early. So, to prevent arbitrage, the price of an American call (denoted $C_A$) must coincide with the price of the corresponding European call (denoted $C_E$).

Since it may be optimal to exercise an American put early, the price of the American put (denoted $P_A$) must be at least as large as the price of the corresponding European put (denoted $P_E$).

Therefore, we have

$$C_A \ = \ C_E, \text{ and } P_A \ \geq \ P_E.$$

We have already shown that the following put-call parity formula must hold for European options:

$$P_E \ = \ C_E + PV(K) - S.$$

Since $P_A \geq P_E$, we must have

$$P_A \ \geq \ C_E + PV(K) - S.$$

Since $C_A = C_E$, we must finally have

$$P_A \ \geq \ C_A + PV(K) - S.$$

This formula may be regarded as the put-call parity formula for American options on stocks that do not pay any dividends.

**Case 2(a): Both European, Stock Pays Dividends**

This is similar to Case 1(a). Consider the following portfolios:

**Portfolio A:** Buy one put, buy one share of stock.

**Portfolio B:** Buy 1 call, lend an amount equal to $PV(K) + PV(D)$.

Cost of setting up Portfolio A: $P + S$.

Cost of setting up Portfolio B: $C + PV(K) + PV(D)$.

Value of Portfolio A at maturity:

1. If $S_T < K$: Put option is worth $K - S_T$; stock is worth $S_T$; dividends of $D$ have been received on stock. Total value: $K + D$.
2. If $S_T \geq K$: Put option expires worthless; stock is worth $S_T$; dividends of $D$ have been received on stock. Total value: $S_T + D$.

Value of Portfolio B at maturity:

1. If $S_T < K$: Call option expires worthless; loan matures with face value $K + D$. Total value: $K + D$.
2. If $S_T \geq K$: Call option is worth $S_T - K$; loan matures with face value $K + D$. Total value: $S_T + D$.

Thus, the portfolios have the same terminal value whether $S_T < K$ or $S_T \geq K$. To avoid arbitrage, portfolios must have same initial value:

$$P + S \;=\; C + PV(K) + PV(D).$$

Therefore, the put and call prices of European options on dividend-paying stocks are related by:

$$P \;=\; C + PV(K) + PV(D) - S.$$

This is the "put-call parity" formula for European options on dividend-paying stocks.

**Case 2(b): Both American, Stock Pays Dividends**

Once again, let $C_A$ and $P_A$ denote, respectively, the call and put option prices for American options, and let $C_E$ and $P_E$ denote the corresponding European option prices.

Since early exercise may be optimal for American put options on dividend-paying stocks, we must have

$$P_A \; \geq \; P_E.$$

Since the only reason to exercise the American call early is the presence of dividends, the *maximum* gain from early exercise is $PV(D)$. Therefore,

$$C_A \; < \; C_E + PV(D).$$

Finally, we have shown that

$$P_E \; = \; C_E + PV(D) + PV(K) - S.$$

Combining all these inequalities, we have:

$$\begin{aligned} P_A &\geq P_E \\ P_E &= C_E + PV(D) + PV(K) - S \\ C_A &< C_E + PV(D) \end{aligned}$$

And we finally obtain:

$$P_A \; \geq \; C_A + PV(K) - S.$$

This formula may be regarded as the put-call parity for American options on dividend paying stocks.

## 19.9 Synthesize forward contracts with options

- $F = C(K) - P(K), K = FV(S)$
- $C - P = S - PV(F) = F = 0.$
- Analogy to swaps and bond market.

### 19.10 Analogy to valuing assets in a firm

- $V = D + E$
- $E = \max(0, V - F) = C(V, F)$
- $D = \min(V, F) = F - \max(0, F - V) = F - P(V, F)$
- $D + E = C - P + F = PV(V)$ [put-call parity]
- Business risk (vary $\sigma$)
- Financial risk (vary $D/E$)

**How does this analogy help in understanding FDIC insurance and the S&L crisis?**

# 20 Replication and risk-neutral pricing

Let the current stock price be $S$. In the future it can take two values (up and down) of $uS$ and $dS$. The interest rate is $r$ per period, so that a one dollar bond today gives $R = 1 + r$ at the end. The exercise price is $K$.

Call payoff in the up state:

$$C_u = \max[0, uS - K]$$

Call payoff in the down state:

$$C_d = \max[0, dS - k]$$

**Replication equations**

Set up a system of 2 equations, holding $X$ shares of stock and $Y$ units of the one dollar bond such that this portfolio has the same payoffs as the option:

$$\begin{aligned} C_u &= X(uS) + YR \\ C_d &= X(dS) + YR \end{aligned}$$

Solving these 2 equations for the two unknowns $(X, Y)$ gives:

$$X = \frac{C_u - C_d}{S(u-d)}$$

and

$$Y = \frac{1}{R}\left[\frac{uC_d - dC_u}{u-d}\right]$$

which are the holdings of stock and bonds required to *replicate* the call option.

**Valuation**

Since the payoffs of the $(X, Y)$ portfolio pay off exactly the same as the option at maturity, it must have the same value today. So to get the option price, we just need to value the $(X, Y)$ portfolio, which is

$$\begin{aligned}
C_0 &= XS + Y \\
&= \frac{C_u - C_d}{S(u-d)}S + \frac{1}{R}\left[\frac{uC_d - dC_u}{u-d}\right] \\
&= \frac{C_u - C_d}{(u-d)} + \frac{1}{R}\left[\frac{uC_d - dC_u}{u-d}\right] \\
&= \frac{RC_u - RC_d}{R(u-d)} + \frac{1}{R}\left[\frac{uC_d - dC_u}{u-d}\right] \\
&= \frac{1}{R}\left[\frac{C_u(R-d) + C_d(u-R)}{u-d}\right] \\
&= \frac{1}{R}\left[C_u\left(\frac{R-d}{u-d}\right) + C_d\left(\frac{u-R}{u-d}\right)\right]
\end{aligned}$$

**Risk Neutral Probabilities**

Notice that we can write

$$\begin{aligned}
p &= \frac{R-d}{u-d} \\
1-p &= \frac{u-R}{u-d}
\end{aligned}$$

as a probability because

$$\begin{aligned}
0 &< p < 1 \\
0 &< 1-p < 1
\end{aligned}$$

which is only true if

$$d < R < u$$

which is true since if $R > u$ then there would be an arbitrage where you could buy the risky bond and do better than the risky stock in an expected return sense. Writing the expression for $p$ into our pricing equation above gives

$$C_0 = \frac{1}{R}\left[pC_u + (1-p)C_d\right]$$

which is the same as valuing a project by discounting expected cashflows using probability weights.

**Notes**

Two things to note here:

- the probabilities are "pseudo" probabilities and
- the discount rate is the riskless rate of interest. Hence we call this valuation method the "risk-neutral" valuation method. The final equation is used to value options on a binomial tree.

**Example**

Let current stock price = 10

It can go up to 14 and down to 5.

The interest rate for one month is 1%.

Price the call option on the stock for one month maturity with a strike price of 10.

$$\begin{aligned}
R &= 1.01 \\
u &= 1.4 \\
d &= 0.5 \\
p &= \frac{R-d}{u-d} = \frac{0.51}{0.9} = 0.5667 \\
1-p &= 0.4333 \\
C_u &= \max[0, 14-10] = 4 \\
C_d &= \max[0, 5-10] = 0 \\
C_0 &= \frac{1}{1.01}\left[0.5667(4) + 0.4333(0)\right] \\
&= 2.24
\end{aligned}$$

# 21 Binomial Model Implementation

- Combine long asset position with borrowing at risk–free rate to replicate the payoff of the option in each state.
- Calculate the probabilities under which the *expected* return of the asset is equal to the risk-free rate of return.
- Use these probabilities with the risk–free rate of interest to calculate the discounted expected payoff from the call.
- This is the insight underlying portfolio insurance.

## 21.1 Hedge Ratio

If you have written a call option, and want to protect yourself against a rise in price, the hedge ratio tells you how many units of the asset to go *long* in, in order to create a riskless position.

If you have written a put option, and want to protect yourself against a fall in price, the hedge ratio tells you how many units of the asset to go *short* in, in order to create a riskless position.

## 21.2 The Risk–Neutral Method

From a computational standpoint, the binomial model is by far the most useful method.

- It must be emphasized that the risk-neutral probabilities have nothing to do with the *true* probabilities of moving up or "down."

The general validity of this method relies on the Harrison–Pliska (1980) result that if an option can be priced by hedging (or replicating) arguments, the same price will result if we use risk–neutral methods.

**Notation:**

$S$: Current asset price.

$u$: size of up move.

$d$: size of down move.

$K$: Strike price of call option.

$r$: Risk–free rate of interest per period.

An important requirement is that $u > r > d$.

$r$ is the *gross* rate of interest, i.e., $r = 1+$ the net rate of interest.

## 21.3 The Two-Period Binomial Model

We begin our analysis with a *two*–period model.

There are 3 possible prices asset after two periods:

- $u^2 S$: Price after two up moves.
- $d^2 S$: Price after two down moves.
- $udS$: Price after one up and one down move.

Note that the order of up and down moves is irrelevant in the last case.

**Option Prices: Notation**

- Let $C$ be the initial price of a European call option with strike price $K$, exercisable after two periods.
- Let $C_u$ be the price of the call at the end of one period, if an up move in the asset price resulted in the first period.
- Define $C_d$ similarly.

- Let $C_{uu}$ be the call price after two periods following two up moves in the asset price.
- Define $C_{ud}$, $C_{du}$, and $C_{dd}$ similarly.

We know that:

- $C_{uu} = \max\{u^2S - K, 0\}$.
- $C_{ud} = C_{du} = \max\{udS - K, 0\}$.
- $C_{dd} = \max\{d^2S - K, 0\}$.

**Question:** What are $C$, $C_u$, and $C_d$?

**Calculating Option Values**

There are two ways of calculating $C$:

**Method 1:**

1. Use risk neutral valuation to determine the value of $C_u$ from $C_{uu}$ and $C_{ud}$.
2. Use risk neutral valuation to determine the value of $C_d$ from $C_{du}$ and $C_{dd}$.
3. Lastly, use risk-neutral valuation to determine the value of $C$ from $C_u$ and $C_d$.

**Method 2:**

1. Use the one-period risk–neutral probability to directly determine the risk-neutral probabilities of terminal nodes.
2. Use these terminal node probabilities to calculate the expected value of the option at the terminal nodes.
3. Discount back using the risk–free rate for two period $r^2$.

The first method is slower, but also gives enough information to calculate the option "deltas" at all points in the tree.

**The Risk–Neutral Probability**

The risk-neutral probability in this model is the value $p$ under which the expected return on the asset is equal to the risk-free rate of interest $r$, i.e., which satisfies

$$p\left(\frac{uS}{S}\right) + (1-p)\left(\frac{dS}{S}\right) \;=\; r.$$

This is the same as

$$pu + (1-p)d \; = \; r.$$

Therefore,

$$p = \left(\frac{r-d}{u-d}\right).$$

## 21.4 Method 1

Suppose $u$ occurs in the first period.

Then, the possible values for the call at the end of the second period are

- $C_{uu}$: if $u$ occurs again
- $C_{ud}$: if $d$ occurs

Therefore, the *expected value* of the call's payoffs under the risk–neutral probability $p$ is

$$pC_{uu} + (1-p)C_{ud}.$$

Discounting by $r$, we get

$$C_u \; = \; \frac{1}{r}\left[pC_u + (1-p)C_d\right]$$

Substituting for $p$, we finally obtain:

$$C_u \; = \; \frac{1}{r}\left[\left(\frac{r-d}{u-d}\right)C_{uu} + \left(\frac{u-r}{u-d}\right)C_{ud}\right]$$

In an analogous manner, we also obtain

$$C_d \; = \; \frac{1}{r}\left[\left(\frac{r-d}{u-d}\right)C_{ud} + \left(\frac{u-r}{u-d}\right)C_{dd}\right]$$

Using the risk–neutral probability $p$ again, the expected payoff of the call at date 0 is

$$pC_u + (1-p)C_d.$$

Therefore,

$$C = \frac{1}{r}\left[pC_u + (1-p)C_d\right].$$

Substituting for $C_u$ and $C_d$, some simplification yields the following value for $C$:

$$\frac{1}{r^2}[p^2C_{uu} + 2p(1-p)C_{ud} + (1-p)^2C_{dd}].$$

## 21.5 Option Deltas

Let $\Delta^c$, $\Delta^c_u$, and $\Delta^c_d$ represent the call options "deltas" initially, after an up move, and after a down move, respectively.

We have

$$\Delta^c = \frac{C_u - C_d}{uS - dS}.$$

$$\Delta^c_u = \frac{C_{uu} - C_{ud}}{u^2S - udS}.$$

$$\Delta^c_d = \frac{C_{ud} - C_{dd}}{udS - d^2S}.$$

### Replication

If the writer of the call wishes to create a riskless hedge portfolio over the two periods, he must

1. Go long $\Delta^c$ units of asset at the begining.
2. Change the long position to $\Delta^c_u$ units of asset at the end of period 1 if $u$ occurred in period 1.
3. Change the long position to $\Delta^c_u$ units of asset at the end of period 1 if $d$ occurred in period 1.

**More Delta**

There are some noteworthy points about these delta values:

- Firstly, and most obviously, the option delta changes as the price $S$ of the underlying changes.
- Secondly, the direction of change in the delta is intuitive. For example, as the asset price increases, the call gets more in-the-money. More of the underlying asset is now required to maintain a riskless hedge, and the delta increases.

## 21.6 Method 2:

As the first step, we calculate the risk-neutral probabilities of the terminal nodes:

- The probability of two up moves under $p$ is $p^2$.
- The probability of one up move and one down move under $p$ is $2p(1-p)$.
- The probability of two down moves is $(1-p)^2$.

**Note:** There are two ways in which one up move and one down move can occur. Hence, we multiply $p(1-p)$ by 2.

**Expected Payoffs**

As the second step, we compute the expected payoff of the call at the terminal nodes.

The call pays

- $C_{uu}$ if there are two up moves.
- $C_{ud}$ if there is one up and one down move.
- $C_{dd}$ if there are two down moves.

Therefore, the expected payoff under $p$ of the call after two periods is

$$p^2 C_{uu} + 2p(1-p)C_{ud} + (1-p)^2 C_{dd}.$$

**Discount payoffs**

As the last step, we discount this expected payoff from the call back to the present.

- The resulting value is the initial price $C$ of the call.

Since the one-period riskless rate is $r$, the two-period riskless rate is $r^2$.

Therefore, the initial call price is

$\frac{1}{r^2}[p^2 C_{uu} + 2p(1-p)C_{ud} + (1-p)^2 C_{dd}]$.

This is identical to the value we obtained by the first method.

## 21.7 Illustrative Example

Let $S = 100$, $K = 100$, $u = 1.05$, $d = 0.95$, and $r = 1.01$.

Consider a two–period European call option on the asset with the strike price $K$.

We have:

- $u^2 S \;=\; 110.25.$
- $udS \;=\; 99.75.$
- $d^2 S \;=\; 90.25.$

Therefore,

- $C_{uu} \;=\; 10.25.$
- $C_{ud} \;=\; 0.$
- $C_{dd} \;=\; 0.$

Lastly, the risk–neutral probability $p$ in this model is

$$p = \frac{r-d}{u-d} = \frac{1.01 - 0.95}{1.05 - 0.95} = \frac{3}{5}.$$

We now have the following:

$$\begin{aligned} C_u &= [pC_{uu} + (1-p)C_{ud}]/r \\ &= [(0.6)(10.25) + (0.4)(0)]/1.01 \\ &= 6.089 \end{aligned}$$

$$\begin{aligned} C_d &= [pC_{ud} + (1-p)C_{dd}]/r \\ &= [(0.6)(0) + (0.4)(0)]/1.01 \\ &= 0. \end{aligned}$$

Therefore:

$$\begin{aligned} C &= [pC_u + (1-p)C_d]/r \\ &= [(0.6)(6.089) + (0.4)(0)]/1.01 \\ &= 3.617 \end{aligned}$$

Finally, the deltas in this example are

$$\Delta^c = \frac{6.089 - 0}{10} = 0.6089.$$

$$\Delta^c_u = \frac{10.25 - 0}{10.50} = 0.976.$$

$$\Delta^c_d = \frac{0 - 0}{9.50} = 0.$$

Note that the delta changes sharply depending on the outcome of the first period, but in an entirely intuitive way.

- After an up move, the call is very deep in-the-money (and almost sure to finish in-the-money), so the new delta is almost unity.
- On the other hand, after a down move, the call is deep out-of-the-money (and definite to finish out-of-the-money), so the delta is zero.

## 21.8 The $n$-Period Problem

There are $(n+1)$ possible prices after $n$ periods for the asset:

$u^nS,\ u^{n-1}dS,\ u^{n-2}d^2S,\ \ldots, ud^{n-1}S,\ d^nS.$

The risk–neutral probability of $u^{n-j}d^jS$ is:

$$\binom{n}{j} p^{n-j}(1-p)^j$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}.$$

The call's value at the asset price $u^{n-j}d^j$, denoted (say) $C_{n-j}$ is

$$C_{n-j} = \max\{u^{n-j}d^jS - K, 0\}.$$

Therefore, the expected payoff from the call under $p$ is

$$\sum_{j=0}^{n}\left[\binom{n}{j} p^{n-j}(1-p)^j C_{n-j}\right]$$

## 21.9 European Calls

Define $m^*$ by

$$\begin{aligned} C_{n-j} &> 0, \quad \text{if } j \leq m^* \\ C_{n-j} &= 0, \quad \text{if } j > m^* \end{aligned}$$

In words, $m^*$ is the *maximum* number of down moves over $n$ periods after which the call still finishes in–the–money.

Then, discounting by $r$, the Binomial formula for the price $C$ of a European call option is

$$\frac{1}{r^n}\sum_{j=0}^{m^*}\left[\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j C_{n-j}\right]$$

**Call Decomposition**

Thus, the value of the call is the difference of two terms. The first term is:

$$S \cdot \textstyle\sum_{j=0}^{m^*}\left[\frac{1}{r^n}\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j u^{n-j}d^j\right]$$

and the second term is

$$\frac{K}{r^n}\left(\sum_{j=0}^{m^*}\left[\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j\right]\right).$$

The sum in the second term is the risk–neutral probability that the call finishes in the money. The expression $K/r^n$ is the present value of $K$.

## 21.10 European Put

The value of a European *put* option may be calculated similarly. Let $P_{n-j}$ represent the value of the put at the price $u^{n-j}d^jS$:

$$P_{n-j} = \max\{K - u^{n-j}d^jS, 0\}.$$

Let $\hat{m}$ be defined by

$$\begin{aligned} P_{n-j} &> 0, \quad \text{if } j \geq \hat{m} \\ P_{n-j} &= 0, \quad \text{if } j < \hat{m} \end{aligned}$$

That is, $\hat{m}$ is the *minimum* number of down moves over $n$ periods required for the put to finish in–the–money.

**Put Decomposition**

Thus, the Binomial formula for the price $P$ of a European put option is

$$\frac{1}{r^n}\sum_{j=\hat{m}}^{n}\left[\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j P_{n-j}\right]$$

Once again, this may be written as the difference of two terms. The first term is

$$\frac{K}{r^n}\left(\sum_{j=\hat{m}}^{n}\left[\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j\right]\right),$$

and the second term is

$$S\cdot\textstyle\sum_{j=\hat{m}}^{n}\displaystyle\left[\frac{1}{r^n}\frac{n!}{(n-j)!j!}p^{n-j}(1-p)^j u^{n-j}d^j\right]$$

## 21.11 Implementing the Binomial Model

So far, we have assumed $u$ and $d$ are given.

Question: In a "real world" model, how do we choose $u$ and $d$?

- We determine a *distribution* for asset prices.
- Then, we fit values for $u$ and $d$ so that as the number of steps in our Binomial model increases, we approximate the given distribution.

We will illustrate this method using the *log–normal* distribution for asset prices.

**The Log–Normal Distribution:**

Let

- $S$ be the current price of the asset.
- $S_t$ be price $t$ years from now.
- $\mu$ be the continuously compounded rate of return on the asset.
- $\sigma$ be the standard deviation of continuously compounded yearly return on the asset.

Then, the log–normal distribution assumption is that

$$\log\left(\frac{S_t}{S}\right) \sim N(\mu t, \sigma^2 t).$$

where "log" denotes natural logarithm.

Note that

- $\log(S_t/S)$ is the $t$-year return (in continuously compounded terms).
- Limited liability obtains in this model, since $S_t > 0$.

## 21.12 Constructing the Binomial Tree

How do we construct a Binomial tree to correspond to a given log–normal distribution?

Take an $n$ period model. Then, each period in this model will be of length $(t/n)$ years.

The Aim: As $n \to \infty$, the Binomial tree should converge to the given normal distribution.

Note that we are using a discrete distribution to approximate a continuous one. So we cannot hope for anything better than convergence in the limit.

In each period of the Binomial model, if $S$ is the start-of-period asset price, the end-of-period asset prices possible are $uS$ and $dS$, with probabilities $q$ and $1-q$.

We want to express the change in log terms.

Thus, if $\log S$ is the log of today's price, the next period price in log terms will be either

$$\log(uS) = \log u + \log S,$$

or

$$\log(dS) = \log d + \log S.$$

**Mean returns**

Thus, the mean return in log–terms is

$$q \log u + (1-q) \log d$$

and the variance of returns per period is

$$q(1-q) \log \left(\frac{u}{d}\right)^2 .$$

Therefore, the mean return over $n$ periods is

$$n[q \log u + (1-q) \log d]$$

which can be rewritten as

$$n \left[q \log \left(\frac{u}{d}\right) + \log d\right]$$

**Variance of returns**

The variance of return over this span is

$$nq(1-q) \log \left(\frac{u}{d}\right)^2$$

We want to choose $u$, $d$, and $q$ so that as $n \to \infty$,

1. $n[q \log(u/d) + \log d] \;\to\; \mu t$.
2. $nq(1-q)[\log(u/d)]^2 \;\to\; \sigma^2 t$.
3. The constructed Binomial distribution converges to the given log-normal distribution.

**Specific discrete distribution choices**

There are many ways to do this.

One, for instance, is to choose:

$$u = \exp\left\{\sigma\sqrt{\frac{t}{n}}\right\}.$$

$$d = \exp\left\{-\sigma\sqrt{\frac{t}{n}}\right\}.$$

$$q = \frac{1}{2} + \frac{1}{2}\left(\frac{\mu}{\sigma}\right)\sqrt{\frac{t}{n}}.$$

Note that $ud = 1$.

**Convergence**

It is readily checked that

$$n\left[q \log\left(\frac{u}{d}\right) + \log d\right] = \mu t,$$

and that

$$nq(1-q)\left[\log\left(\frac{u}{d}\right)\right]^2 = \sigma^2 t\left(1 - \left(\frac{\mu^2 t}{\sigma^2 n}\right)\right)$$

So as $n \to \infty$, the mean and variance of the Binomial model converge to the mean and variance of the log–normal distribution.

It can also be shown that the Binomial distribution itself converges to the log–normal distribution.

# 22 Homework and Solutions

## 22.1 HW1

1. Using WRDS or Yahoo! Finance, download daily stock returns for your favorite stock for five years. Also download the value of the S&P500 index for the same

period. Place the data in a comma-separated values (csv) file. Now answer the following questions using R:

(a) Plot the data for both the index and the stock on the same graph. Make sure that your graph has the following features: title, labels for the two axes, legends for the two data series, a grid.

(b) Compute the following statistics for both data series: (a) mean, (b) standard deviation, (c) skewness, (d) kurtosis.

(c) Is the skewness of the stock price positive or negative? Why do you think this is so?

(d) Is the excess kurtosis greater than zero or not? Why?

(e) Compute the covariance and correlation between the stock price and the index. Is this positive or negative? Why?

For each answer, provide the R code that was used to generate the answer (you will receive no credit without the code).

ANS: The R code that implements the solution is as follows:

```
#PLOT
min_s = min(c(goog,spx))
max_s = max(c(goog,spx))
plot(goog,ylim = c(min_s,max_s),type='l',xlab="Days",ylab="Prices")
lines(spx,col="red")
legend("topright",c("GOOG","SPX"),lty=1,col=c("black","red"))

#DESCRIPTIVE STATISTICS
> data = read.csv("GOOG_SPX.csv",header=TRUE,sep=",")  #Read in data
> dim(data)    #Check dimensions
[1] 1258   13
> goog = data[,7]
> spx = data[,13]
> print(mean(goog))
[1] 426.2466
> print(mean(spx))
[1] 1232.417
> print(sd(goog))
[1] 117.4463
> print(sd(spx))
[1] 201.6198
> print(skewness(goog))
[1] -0.0003097515
```

```
>
> library(moments)   #Load library
> print(skewness(goog))
[1] -0.0003097515
> print(skewness(spx))
[1] -0.626912
> print(kurtosis(goog))
[1] 2.832227
> print(kurtosis(spx))
[1] 2.641874
> print(cov(goog,spx))
[1] 11527.25
> print(cor(goog,spx))
[1] 0.4868029
```

2. What is the current exchange rate between the Japanese yen and US dollar? Suggest a reasonable bid-ask spread around this exchange rate. What would you do to the bid-ask spreads if you wanted to sell yen and buy dollars?

   ANS: The exchange rate is obtained from the x-rates.com website. The current rate is 92.43 yen/dollar. A reasonable bid-ask spread around this would be 92.33–92.53.

   If you want to sell yen, then it means you want to buy dollars. To attract sellers of dollars you will skew the prices upwards, i.e., make the exchange rate you offer something like 92.43–92.53.

3. If the Japanese interest rate for one year is 0.5% and the US one-year rate is 1%, what is the one-year forward exchange rate in yen/dollar? Use the spot exchange rate from the previous question as an input.

   ANS: The forward exchange rate is

$$92.43 \times (1 + 0.005)/(1 + 0.01) = 91.9724 \text{ yen/dollar}$$

4. What are the two broad components of the 10-year nominal interest rate? Using the Federal Reserve Web site explain what data you would access to obtain the two components of the nominal interest rate.

   ANS: The nominal interest rate may be obtained from the constant maturity data series on interest rates. See the series DGS10.

   The real rate may be inferred from the rates of Treasury-Inflation Protection Securities (also known as TIPS). These are in the series DTP10L18.

   The difference between the nominal and real rates is inflation. The consumer price index (CPI) may be obtained from the series CPIAUCSL.

5. Using the data from your favorite stock, compute the following, and report the R code used to do so.

   (a) Calculate the daily continuously-compounded return for the stock and plot the same. To compute the return on any day $t$, take the natural logarithm of the ratio of the day's stock price to the previous day's stock price, i.e., the return is $\ln[S(t)/S(t-1)]$.

   (b) Compute the mean and standard deviation of the daily return.

   (c) Compute the annualized mean and standard deviation of returns for the stock.

   ANS: Here is the code:

```
> n = length(goog)
> print(n)
[1] 1258
> ret = log(goog[2:n]/goog[1:(n-1)])
> print(mean(ret))
[1] -0.0008928693
> print(sd(ret))
[1] 0.02309461
> print(mean(ret)*n/5)   #Annual return
[1] -0.2246459
> print(sd(ret)*sqrt(n/5))    #Annualized standard deviation
[1] 0.3663245
```

## 22.2 HW2

1. Download the 10-year constant maturity series of data (DGS10) from the Federal Reserve since the inception of the series. Use this data (ignore missing values) to fit the mean-reverting stochastic process model of interest rates, and report the three parameters of that model that you estimated. Recall that these parameters are $\{k, \theta, \sigma\}$, i.e., the rate of mean-reversion, the long-run mean level, and the volatility, respectively. Assume that a year is 260 trading days. Report the R code used to complete this exercise. Interpret the coefficients you estimated.

   ANS: The code and results are as follows.

```
> #INTEREST RATE FITTING
> rdata = read.csv("DGS10.csv",header=TRUE,na.strings=".")
> r = rdata[,2]
```

```
> idx = which(r!="NA")
> r = r[idx]
> print(mean(r))
[1] 6.892419
> n = length(r)
> r1 = r[2:n]
> r0 = r[1:(n-1)]
> res = lm(r1~r0)
> print(summary(res))

Call:
lm(formula = r1 ~ r0)

Residuals:
       Min         1Q     Median         3Q        Max
-0.7488608 -0.0297893 -0.0003331  0.0294256  0.6518436

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 0.0023482  0.0017848    1.316    0.188
r0          0.9996564  0.0002426 4120.663   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.06829 on 11978 degrees of freedom
Multiple R-squared: 0.9993,     Adjusted R-squared: 0.9993
F-statistic: 1.698e+07 on 1 and 11978 DF,  p-value: < 2.2e-16

> a = res$coefficients[1]
> b = res$coefficients[2]
> h = 1/260
> k = (1-b)/h
> theta = a/(1-b)
> residuals = r1 - (a+b*r0)
> sigma = sqrt(var(residuals)/h)
> print(c(k,theta,sigma))
         r0 (Intercept)
  0.0893324   6.8343686   1.1011243
```

The rate of mean-reversion is $k = 0.089$, i.e., the rate reverts at a rate of $1/11$th of the way to the mean level per year. The long-run mean level is $\theta = 6.83\%$. The volatility is $\sigma = 1.1\%$ per annum.

2. In the previous question, compute the arithmetic and geometric mean of the interest rate series. Which one is greater?

   ANS: This is a little tricky. The arithmetic mean is easy to compute. The geometric mean is harder. The product of interest rates becomes too large for the computer to handle, hence a different approach is needed. What we require is the geometric mean, i.e.,

   $$\left[\prod_{j=1}^{N} r_j\right]^{1/N}$$

   We will rewrite this as follows and then compute it:

   $$\exp\left(\ln\left[\prod_{j=1}^{N} r_j\right]^{1/N}\right) = \exp\left(\frac{1}{N}\sum_{j=1}^{N}\ln(r)\right)$$

   The R code is as follows. You will see that the arithmetic mean is 6.89 whereas the geometric mean is 6.46. As we had learnt, the geometric mean is lower than the arithmetic mean.

```
> rdata = read.csv("DGS10.csv",header=TRUE,na.strings=".")
> r = rdata[,2]
> idx = which(r!="NA")
> r = r[idx]
> print(mean(r))
[1] 6.892419
> n = length(r)
> gmean = exp(sum(log(r))/n)
> print(gmean)
[1] 6.459122
```

3. Using the data series on your favorite stock and the S&P500 data, calculate the covariance of the two series using R. Next, calculate the covariance using the functional form $Cov(x, y) = E(xy) - E(x)E(y)$. Why are these two values different? Offer a reconcilement.

   ANS: The R function assumes that we are computing the covariance of a sample (not the population) and hence one degree of freedom is lost. The functional form does not assume loss of a degree of freedom. The following code presents the calculations and a reconcilement.

```
> #COVARIANCE TWO WAYS
> n = length(goog)
> cv1 = cov(goog,spx)
```

```
> print(cv1)
[1] 11527.25
> cv2 = mean(goog*spx)-mean(goog)*mean(spx)
> print(cv2)
[1] 11518.09
> print(cv2*n/(n-1))
[1] 11527.25
```

4. Compute the test statistic for the difference of means between the daily returns on your stock series and the returns on the index.

   ANS: The R code for this exercise is as follows.

```
> #DIFFERENCE OF MEANS
> m1 = mean(goog)
> m2 = mean(spx)
> v1 = var(goog)
> v2 = var(spx)
> n = length(goog)
> sig_diff = sqrt(v1/n+v2/n)
> test_stat = (m1-m2)/sig_diff
> print(test_stat)
[1] -122.5438
```

   The difference is highly significant.

5. Run a regression of the daily *returns* of your favorite stock against the returns of the S&P500 index for the five years you chose data for. Present your results highlighting the main statistics of interest and offering some interpretation of what you learned from this exercise.

   ANS: The R code for this exercise is as follows:

```
> data = read.csv("GOOG_SPX.csv",header=TRUE,sep=",")  #Read in data
> dim(data)    #Check dimensions
[1] 1258   13
> goog = data[,7]
> spx = data[,13]
> data = read.csv("GOOG_SPX.csv",header=TRUE,sep=",")
> goog = data[,7]
> spx = data[,13]
> n = length(goog)
> goog_ret = log(goog[2:n]/goog[1:(n-1)])
> spx_ret = log(spx[2:n]/spx[1:(n-1)])
```

```
> res = lm(goog_ret ~ spx_ret)
> summary(res)

Call:
lm(formula = goog_ret ~ spx_ret)

Residuals:
       Min         1Q     Median         3Q        Max
-0.1648555 -0.0088648  0.0006444  0.0086946  0.1040159

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0009402  0.0005205  -1.806   0.0711 .
spx_ret      0.9153839  0.0342983  26.689   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01845 on 1255 degrees of freedom
Multiple R-squared: 0.3621,Adjusted R-squared: 0.3616
F-statistic: 712.3 on 1 and 1255 DF,  p-value: < 2.2e-16
```

Google had a beta of 0.91, less than 1.0. This is unusual for a technology stock! The adjusted $R^2$ is 36% which is usual.

6. Take a set of 10 stocks each with mean return of 10%. Assume that the variance of returns of each stock is 0.05, and the covariance of all stock pairs is 0.01. (a) Compute the expected return of the portfolio of these ten stocks if you invest fraction 1/55 in the first stock, 2/55 in the second stock, and so on, until 10/55 in the tenth stock. (b) For the same portfolio weights compute the standard deviation of returns of the portfolio.

   ANS: The R code is as follows:

```
> mn = matrix(0.10,10,1)
> mn
      [,1]
 [1,]  0.1
 [2,]  0.1
 [3,]  0.1
 [4,]  0.1
 [5,]  0.1
 [6,]  0.1
 [7,]  0.1
```

```
 [8,]  0.1
 [9,]  0.1
[10,]  0.1
> w = matrix(seq(1,10)/55,10,1)
> w
            [,1]
 [1,] 0.01818182
 [2,] 0.03636364
 [3,] 0.05454545
 [4,] 0.07272727
 [5,] 0.09090909
 [6,] 0.10909091
 [7,] 0.12727273
 [8,] 0.14545455
 [9,] 0.16363636
[10,] 0.18181818
> mn_portfolio = t(w) %*% mn
> mn_portfolio
     [,1]
[1,]  0.1
> SIGMA = matrix(0.01,10,10)
> diag(SIGMA) = 0.05
> SIGMA
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,] 0.05 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01  0.01
 [2,] 0.01 0.05 0.01 0.01 0.01 0.01 0.01 0.01 0.01  0.01
 [3,] 0.01 0.01 0.05 0.01 0.01 0.01 0.01 0.01 0.01  0.01
 [4,] 0.01 0.01 0.01 0.05 0.01 0.01 0.01 0.01 0.01  0.01
 [5,] 0.01 0.01 0.01 0.01 0.05 0.01 0.01 0.01 0.01  0.01
 [6,] 0.01 0.01 0.01 0.01 0.01 0.05 0.01 0.01 0.01  0.01
 [7,] 0.01 0.01 0.01 0.01 0.01 0.01 0.05 0.01 0.01  0.01
 [8,] 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.05 0.01  0.01
 [9,] 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.05  0.01
[10,] 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01  0.05
> vr_portfolio = t(w) %*% SIGMA %*% w
> sd_portfolio = sqrt(vr_portfolio)
> sd_portfolio
          [,1]
[1,] 0.1228451
```

7. Assume that there are three credit ratings: investment-grade, junk, and default.

The rating transition matrix per period is given as

$$\begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0 & 1 \end{bmatrix}$$

(a) What is the probability of an investment grade firm ending up in default after three periods? (b) What is the probability of a junk grade firm ending up in default after three periods? (c) What is the probability of a junk grade firm returning to investment grade after three periods?

ANS: R code is as follows:

```
> R = matrix(c(0.8,0.1,0,0.15,0.8,0,0.05,0.1,1),3,3)
> R
     [,1] [,2] [,3]
[1,]  0.8 0.15 0.05
[2,]  0.1 0.80 0.10
[3,]  0.0 0.00 1.00
> R %*% R %*% R
       [,1]    [,2]    [,3]
[1,] 0.5480 0.29025 0.16175
[2,] 0.1935 0.54800 0.25850
[3,] 0.0000 0.00000 1.00000
```

The three required probabilities are: $0.16175, 0.2585, 0.1935$.

## 22.3 HW3

This homework is a simple one that asks you a series of questions and leads you to working out various analyses for portfolios using R. In this homework, there is no reading in data (phew!). You will learn to manipulate vectors and matrices, do matrix multiplication, write a function and call it from R.

1. Create a covariance matrix of returns of ten stocks where all the stocks have a variance of 0.05, and the covariances are all 0.02.

   ANS:

```
> cv = matrix(0.02,10,10)
> diag(cv) = 0.05
> cv
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,] 0.05 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02  0.02
 [2,] 0.02 0.05 0.02 0.02 0.02 0.02 0.02 0.02 0.02  0.02
 [3,] 0.02 0.02 0.05 0.02 0.02 0.02 0.02 0.02 0.02  0.02
 [4,] 0.02 0.02 0.02 0.05 0.02 0.02 0.02 0.02 0.02  0.02
 [5,] 0.02 0.02 0.02 0.02 0.05 0.02 0.02 0.02 0.02  0.02
 [6,] 0.02 0.02 0.02 0.02 0.02 0.05 0.02 0.02 0.02  0.02
 [7,] 0.02 0.02 0.02 0.02 0.02 0.02 0.05 0.02 0.02  0.02
 [8,] 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.05 0.02  0.02
 [9,] 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.05  0.02
[10,] 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02  0.05
```

2. Create a vector of returns for your ten stocks where each stock has a return equal to 13%.

   ANS:

```
> n = dim(cv)[1]
> n
[1] 10
> mn = matrix(0.13,n,1)
> mn
      [,1]
 [1,]  0.13
 [2,]  0.13
 [3,]  0.13
 [4,]  0.13
 [5,]  0.13
 [6,]  0.13
 [7,]  0.13
 [8,]  0.13
 [9,]  0.13
[10,]  0.13
```

3. Create a vector of portfolio weights for the ten stocks with equal weights in each stock.

   ANS:

```
> w = matrix(1/n,n,1)
> w
      [,1]
 [1,]  0.1
 [2,]  0.1
```

```
 [3,]  0.1
 [4,]  0.1
 [5,]  0.1
 [6,]  0.1
 [7,]  0.1
 [8,]  0.1
 [9,]  0.1
[10,]  0.1
```

4. Using R, calculate the average return of the portfolio given the weights in the previous question. Make sure you use a matrix multiplication to do this.

   ANS:

```
> portfolio_mean_return = t(w) %*% mn
> portfolio_mean_return
     [,1]
[1,] 0.13
```

5. Using R, calculate the standard deviation of return of the portfolio given the weights in the previous question. Make sure you use a matrix multiplication to do this.

   ANS:

```
> portfolio_variance = t(w) %*% cv %*% w
> portfolio_stdev = sqrt(portfolio_variance)
> portfolio_stdev
          [,1]
[1,] 0.1516575
```

6. Now, create a function called `portfolioRisk` and save it in a program file called `portfolio.R`. This function will do exactly the same calculations as was done in the previous questions to calculate the portfolio standard deviation using the same input values. Your function will accept the following inputs: (a) the number of stocks $n$; (b) the variance of return of each stock assuming they all have the same variance; (c) the covariance of all stocks assuming all pairs of stocks have the same covariance; (d) the portfolio weights for each stock, assuming all stocks are equally weighted in the portfolio. The output from the function will be the standard deviation of returns of the portfolio. Also present the R code you will use to call the function. Use the function to calculate the standard deviation of returns for $n = 100$ stocks. Compare this to the standard deviation you calculated earlier for ten stocks. Which one is greater and why?

(Note: Create a new script file for the function and make sure it ends in `.R` and then load your function by using the command

```
source("portfolio.R")
```

ANS: The function file may be typed in as follows:

```
portfolioRisk = function(n,variance,covariance) {
    cv = matrix(covariance,n,n)
    diag(cv) = variance
    weights = matrix(1/n,n,1)
    portfolio_variance = t(weights) %*% cv %*% weights
    portfolio_stdev = sqrt(portfolio_variance)
}
```

Now we can load the function as follows:

```
> source("portfolio.R")
> print(portfolioRisk(100,0.05,0.02))
          [,1]
[1,] 0.1424781
```

We see that this standard deviation is smaller than when the portfolio contained only ten stocks. Hence, diversification reduces portfolio risk.

7. Plot the portfolio standard deviation for values of $n$ going from 5 to 200 in steps of 5. Make sure you label the axes in the plot correctly.

   ANS: The R code to carry out this is as follows:

```
> m = seq(5,200,5)
> m
 [1]   5  10  15  20  25  30  35  40  45  50  55  60  65  70  75  80  85  90  95
[20] 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190
[39] 195 200
> sd = NULL
> for (n in m) {
+ portf_sd = portfolioRisk(n,0.05,0.02)
+ sd = c(sd,portf_sd)
+ }
> sd
 [1] 0.1612452 0.1516575 0.1483240 0.1466288 0.1456022 0.1449138 0.1444200
 [8] 0.1440486 0.1437591 0.1435270 0.1433369 0.1431782 0.1430438 0.1429286
[15] 0.1428286 0.1427410 0.1426637 0.1425950 0.1425335 0.1424781 0.1424279
[22] 0.1423823 0.1423407 0.1423025 0.1422674 0.1422349 0.1422049 0.1421770
```

```
[29] 0.1421510 0.1421267 0.1421040 0.1420827 0.1420627 0.1420439 0.1420262
[36] 0.1420094 0.1419935 0.1419785 0.1419642 0.1419507
> plot(m,sd,type="l",col="blue",xlab="No of Stocks",ylab="Portfolio SD")
```

The plot of portfolio standard deviation is declining and convex.

8. What concept(s) did you learn in an earlier finance course are being portrayed in the plot you created in the previous question?

   ANS: The plot shows two things: (a) Diversification reduces portfolio risk. (b) As $n$ increases, the standard deviation falls rapidly at first and then flattens out to a level of risk that no amount of diversification can eliminate. This level of risk is known as "systematic" or "non-diversifiable" risk.

## 22.4 Quiz 1

FNCE 115
Mid-Term Quiz 1 (20 questions, 20 points, 75 minutes)
(January 26, 2010)

Note: You do not need to calculate the final answer as long as you state the answer in a formula. For example, if your final answer is $0.25 \times (1/12)$ you may leave it that way and I will give you full credit.

Note: Detail your work. No credit will be given unless you have the working shown. It will also help getting you partial credit in case some of the answer is wrong.

1. What are the R commands to compute the mean, standard deviation and variance of a data series `x`?

   ANS:

```
mean(x)
sd(x)
var(x)
```

2. If the stock price on January 22 was \$10.35 and the stock price on January 21 was \$9.98, what is the continuously compounded daily return from January 21 to January 22?

   ANS:

```
> log(10.35/9.98)
[1] 0.03640343
```

3. Restate the return in the last question in annualized terms.

   ANS:

```
> log(10.35/9.98)*260
[1] 9.464892
```

   This is 946%

4. If the daily standard deviation of stock returns is 1%, (a) what is the annual standard deviation of returns? (b) What is the precise economic assumption you used to arrive at your answer?

   ANS:

```
> 1*sqrt(260)
[1] 16.12452
```

The assumption used is that markets are efficient. Under this assumption, the returns from one day are independent of those on other days. Hence, there is no covariance between the returns on one day and the next. This means that the variances are additive. For example, over two days returns $r_1, r_2$, we may write that $Var(r_1 + r_2) = Var(r_1) + Var(r_2)$.

---

———————-

5. Which of the following is the most accurate? In a highly diversified portfolio, the standard deviation of portfolio returns depends on: (a) Number of stocks in the portfolio; (b) The number of and variance of stocks in the portfolio; (c) The number of and covariance of stocks in the portfolio; (d) The variance and covariance of stocks in the portfolio. Explain your answer.

   ANS:

   ```
   (c)
   ```

---

———————-

6. The one-year interest rate $r$ is assumed to move daily based on the following dynamic equation:

   $$\Delta r = 0.25(0.10 - r)(1/260) + 0.10\epsilon\sqrt{1/260}$$

   What is the long-run average rate at which the one-year interest rate is going to be?

   ANS:

```
0.10 or 10%
```

---

7. In the previous question, what is the daily standard deviation of the interest rate?

   ANS:

```
> 0.1*sqrt(1/260)
[1] 0.006201737
```

---

8. If a stock has a mean daily return of 0.1% and a standard deviation of daily return of 0.5%, what are its annual mean return and standard deviation of return?

   ANS:

```
> 0.001*260
[1] 0.26
> 0.005*sqrt(260)
[1] 0.08062258
```

   That is, 26% and 8.06%, respectively.

---

9. If the nominal interest rate is 7% per year and the inflation rate is 3% per year, what is the real interest rate?

   ANS:

   ```
   4%
   ```

   ---

   ______________-

10. If the Japanese yen to US dollar exchange rate is 95 yen per dollar, and the one-year interest rates in Japan and the US are 1% and 2%, respectively, what is the forward one-year exchange rate between the yen and dollar. Make sure that you state your answer in the correct units.

    ANS:

    ```
    > 95*1.01/1.02
    [1] 94.06863
    ```

    ---

    ______________-

11. If two stocks have the same mean and variance of returns, but their skewness levels are 1.6 and 2.1, which one would investors prefer? Why?

    ANS:

    ```
    2.1, investors dislike negative skewness and like positive skewness.
    ```

12. If the kurtosis of returns of a stock is 5, what can you say about the risk of this stock relative to one with a kurtosis of 3? Explain clearly.

    ANS:

    ```
    It has fat tails and experiences greater outliers than a stock that has normally
    ```

13. A stock has the following prices on 3 consecutive days: \$10, \$11 \$10.5. What is the daily arithmetic mean of the stock's *simple* returns? (Not continuously-compounded returns.)

    ANS:

    ```
    > s = c(10,11,10.5)
    > returns = (s[2:3]-s[1:2])/s[1:2]
    > returns
    [1]  0.10000000 -0.04545455
    > mean(returns)
    [1] 0.02727273
    ```

14. For the same stock in the previous question, using the same daily returns, what is the geometric mean return?

ANS:

```
> sqrt(10.5/10)-1
[1] 0.02469508
```

---

15. Is the variance of a sample greater than or smaller than the variance of a population? Why?

    ANS:

```
Greater. We divide the squared deviations by (N-1) not N as in the case of the po
```

---

16. If the mean of a stock's return is $E(r) = 0.20$ and the mean of the square of the stock's return is $E(r^2) = 0.10$, then what is the standard deviation of the stock's return?

    ANS:

```
> sqrt(0.10-0.20^2)
[1] 0.2449490
```

That is, 24.49%.

---

17. In the previous question, what is the standard deviation of the stock's return if each and every return is doubled?

    ANS:

```
> sqrt(0.10-0.20^2)*2
[1] 0.4898979
```

---

18. Stock A's return is regressed on the S&P return and small stock returns, and we find that the t-statistic on the intercept and the coefficients are non-significant, but the F-statistic is significant. What would you infer from this?

    ANS:

```
We infer that the S\&P index and the small stock index do not explain stock A's
returns well individually, but that they do so jointly.
```

---

19. Three stocks in a portfolio have weights 0.4, 0.35, 0.25, respectively. Their mean returns are 10%, 12%, and 20% respectively. Write down the matrix calculation for the average return of the portfolio and calculate the mean return. (Make sure you construct the matrices and vectors precisely.)

ANS:

```
> w = matrix(c(0.4,0.35,0.25),3,1)
> R = matrix(c(0.1,0.12,0.2),3,1)
> t(w) %*% R
      [,1]
[1,] 0.132
```

---

20. The probability that a stock will go up today if it has gone up yesterday is 0.5, and that it will go down if it has gone down yesterday is 0.6. What is the probability that a stock that has gone up today will go up two days from now.

ANS:

```
> m = matrix(c(0.5,0.4,0.5,0.6),2,2)
> m
     [,1] [,2]
[1,]  0.5  0.5
[2,]  0.4  0.6
> m %*% m
     [,1] [,2]
[1,] 0.45 0.55
[2,] 0.44 0.56
```

The matrix above gives the two-period transitions. The answer is 0.45.

---